Research Article

Vehicle Information Influence Degree Screening Method Based on GEP Optimized RBF Neural Network

Jingfeng Yang¹, Nanfeng Zhang,² Ming Li,^{3,4} Yanwei Zheng,¹ Li Wang⁵, Yong Li¹,
Ji Yang,¹ Yifei Xiang,⁶ and Lufeng Luo⁷

¹Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou Institute of Geography, Guangzhou 510070, China

²Guangzhou Customs, Guangzhou 510623, China

³South China Agricultural University, Guangzhou 510642, China

⁴Guangzhou Xingwei Mdt InfoTech Ltd, Guangzhou 510630, China

⁵School of Computer Software in Tianjin University, Tianjin 300072, China

⁶North China Electric Power University, Beijing 102206, China

⁷College of Mechanical and Electrical Engineering, Foshan University, Foshan 528000, China

Correspondence should be addressed to Lufeng Luo; luolufeng@fosu.edu.cn

Received 23 March 2018; Revised 27 June 2018; Accepted 8 July 2018; Published 14 October 2018

Academic Editor: Andy Annamalai

Copyright © 2018 Jingfeng Yang 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.

Due to the continuous progress in the field of vehicle hardware, the condition that a vehicle cannot load a complex algorithm no longer exists. At the same time, with the progress in the field of vehicle hardware, a number of studies have reported exponential growth in the actual operation. To solve the problem for a large number of data transmissions in an actual operation, wireless transmission is proposed for text information (including position information) on the basis of the principles of the maximum entropy probability and the neural network prediction model combined with the optimization of the Huffman encoding algorithm, from the exchange of data to the entire data extraction process. The test results showed that the text-type vehicle information based on a compressed algorithm to optimize the algorithm of data compression and transmission could effectively realize the data compression, achieve a higher compression rate and data transmission integrity, and after decompression guarantee no distortion. Therefore, it is important to improve the efficiency of vehicle information transmission, to ensure the integrity of information, to realize the vehicle monitoring and control, and to grasp the traffic situation in real time.

1. Introduction

With the continuous progress in the field of vehicle hardware technology, coupled with mobile Internet applications and other technology-driven applications, the amount of real-time vehicle information (particularly text information) has increased rapidly. A large amount of real-time information, in addition to a large number of storage and processing resources, occupies a large amount of transmission bandwidth, and practical applications often do not need to use all of the data generated. Therefore, the actual use of the data generated by the influence degree impact assessment is urgently required to solve one of the problems of the related

applications. Therefore, this topic has attracted considerable research attention thus far. Chuanqin and Yufeng [1], from the perspective of the influence of safe driving, analyzed the influencing factors using the survey data of AHP and a comparison of the effects of the most influential factors on the driving safety of the driver. Wang et al. [2] used factor analysis and principal component analysis to analyze the influencing factors of urban typical passenger transport mode selection. Xu et al. [3], on the basis of an analysis of the data sampling interval and the relationship between the traffic flow parameters and the other factors, proposed the methods of real-time dynamic traffic data screening of abnormal values and their recovery. Yulong and Ma [4]

proposed a method of screening and recovering real-time traffic data. Longhui et al. [5] proposed a screening and testing method of intelligent traffic sensor data. Yulong and Junying [6] sorted the influencing factors according to the size of the correlation degree and then screened the key factors affecting the evolution of the urban traffic structure. In addition, methods including traffic impact factor screening methods [7, 8] and neural network optimization methods [9–12] have been developed by scholars [13, 14]. However, for the screening of various factors related to traffic travel, most of the current methods are focused on trip data analysis, while few actual vehicle data applications have been considered. Therefore, in this study, we considered the normal operation of a purely electric bus in the process of running data through the GEP optimization of the RBF neural network, using the driving range of the purely electric bus as the research object, the influence of the main factor analysis, and the continued screening of various factors affecting the driving range of the electric buses, to provide decision support for the purely electric bus line planning and adjustment, vehicle scheduling, and charging station planning.

2. Vehicle Information and Impact Factor Data Description

Nowadays, a large number of public transport vehicles are new energy vehicles. In these new vehicles, the on-board information collection is more accurate and faster than earlier. Therefore, we considered the vehicle data of 26 purely electric buses running on 801 roads in Guangzhou as the object of this study and verified the accuracy of the algorithm proposed in this paper. The influence degree of the various influencing factors on the energy consumption was a result of the output. The influence of various types of vehicle information on energy consumption was thus analyzed.

The data sources included the following: the purely electric bus battery management system, time, SOC, total voltage, total current, maximum temperature, minimum temperature and maximum voltage, and the minimum voltage calculated using the historical and real-time data of eight parameters. We obtained the position, speed, direction, time, elevation, departure, shift, scheduling, vehicle entrance, vehicle departure, driver position information and management information from the intelligent bus dispatching system, and the satellite positioning vehicle travelling data recorder. The traffic card clearing and settlement system, combined with the traffic surveys, passenger flow analysis, and derivation, provided the passenger flow data. The data of the traffic performance index system were obtained from the traffic performance analysis system. Further, meteorological data were obtained from an open meteorological website.

As the energy consumption of the purely electric buses is considerably affected by season, in this study, we chose the season with the maximum energy consumption for the analysis. The main reasons for this selection were as follows: first, when the energy consumption increases, the relationship between the driving range and the energy consumption is relatively easy to distinguish. Second, in the case of the purely electric bus, the replacement battery

FIGURE 1: Consumption of SOC value curve for 100 consecutive days.

must use specialized equipment, and this equipment is not movable. In the actual operation, it is unable to carry on the complete power consumption experiment; in the process of data collection, it can only be used for pure electric bus energy consumption, and for the influence factors of operating state data as the research object and the use of energy, more working process can make the prediction range closer to the true value. Third, the factors affecting the working life of a purely electric bus battery include the battery capacity, battery voltage, battery energy and specific energy of the battery, charging efficiency, and battery self-discharge rate. Generally, for experienced several charge-discharge processes and battery rated capacity in the first use of a period of time, the battery capacity will increase 5%–20%; subsequent use will remain unchanged after gradually reduced in a certain period of time. When the capacity of the battery is only 80% of the initial capacity, the battery life can be considered to be over, and the power batteries of the purely electric buses are in the stable state.

Therefore, first, the energy consumption data of electric buses in May 2013, five-month to 9-month high temperature based on the bus passenger flow, weather, operation, line, and station data samples, collected a total of 29113828 original data. According to the requirements of the adjacent section of the station, we integrated the data samples, on the basis of the relevant requirements of the station traffic energy index, data fusion, classification, information extraction, matching, removal of missing data, and other operations. Finally, 1023768 samples of complete data were obtained. The average section of each station contained 26941 attributes, complete stations, energy consumption data of the adjacent sections, and the actual operating driving range.

Figure 1 shows the actual value curve of the energy consumption for the 100 consecutive days from June 1, 2013, to September 8, 2013, at the University City station departure to the backyard. The energy consumption of SOC was the highest between 45% and 61%, the maximum energy consumption of SOC was 60.80%, the minimum

FIGURE 2: Line passenger flow curve (July 11).

FIGURE 3: Line overall passenger flow curve (July 1–15).

energy consumption of SOC was 45.20%, the average energy consumption of SOC was 53.67%, and the 69 operation energy consumption of SOC was 50%–57%. Figure 2 shows the passenger flow data of the entire route of route 801 on July 11, to 15 minutes for the statistical cycle, from 6:00 in the morning statistics until 23:45 classes closed. Figure 3 shows the average distribution of the passenger flow on days 1–15 of July with the statistical period of 15 min during the operational hours. Figure 4 shows the passenger flow diagram of the same condition.

With statistical intervals of 5 min, Figure 5 shows the traffic performance index of the 801 line from July 11, 2013, 6:00–23:59 in Guangzhou City, which reflects the changing trend of the regional traffic congestion and the temporal and spatial evolution laws of congestion. Similarly, with a 5-min statistical interval, Figure 6 shows the 801 lines on July 11, 2013, and July 14, 2013, from 6:00 to 23:59, representing the working days and weekends during the Guangzhou City road network integrated congestion. That is, the proportion

FIGURE 4: Line passenger flow curve per bus (July 1–15).

FIGURE 5: City road traffic congestion index.

FIGURE 6: City road network comprehensive congestion.

of the comprehensive congestion time and space to the total time and space range indicates the congestion proportion of the road network.

3. Influence Degree Description and Test Result of Energy Consumption Influence Factors

The influence degree of the energy consumption factors on the driving range of purely electric buses can be described by calculating the influence degree of the influencing factors of energy consumption. The RBF neural network is a type of artificial neural network, which uses local adjustment to perform function mapping [15–17]. It has strong input and output mapping functions and is the best network for a mapping function in a forward network. The RBF neural network has a strong nonlinear approximation ability, simple network structure, and fast learning speed. The output matrix of the hidden layer after iteration is linear with the output. Therefore, this is an ideal algorithm for calculating the influence degree.

In the two stages of the learning process of RBF neural networks, unsupervised learning was based on the input samples to determine the vectors and normalized parameters of the Gauss function of each node in the hidden layer. In the supervised learning phase, the hidden layer parameters were determined by using the least squares method, and the weights of the hidden layer and the output layer were calculated. The RBF neural network's hidden layer output was calculated as follows [15–17]:

$$h_i(t) = \phi_i(\|x(t) - c_i\|) = \exp\left(-\sum_{i=1}^n \frac{(x_i(t) - c_i)^2}{2s_i^2}\right), \quad (1)$$

where $x(t)$ is the t th time network input vector; c_i is the central vector of the i th cell in the hidden layer; s_i is the shape parameter of the Gauss function, besides $s_i > 0$, $1 \leq i < L$; and L is the number of hidden nodes.

For the RBF neural network, the overall network output can be expressed as follows:

$$y = \sum_{i=1}^L w_i h_i = \sum_{i=1}^L w_i \phi_i(\|x(t) - c_i\|) = H^T W, \quad (2)$$

where $H = [h_1, h_2, \dots, h_L]^T$ and $W = [w_1, w_2, \dots, w_L]$.

For the RBF network with k input nodes, M output nodes, and n learning samples, the error objective function can be expressed as follows:

$$\begin{aligned} E(n) &= \frac{1}{2} \sum_{k=1}^n \lambda^{n-k} \sum_{i=1}^M \delta_i(n) = \frac{1}{2} \sum_{k=1}^n \lambda^{n-k} \sum_{i=1}^M [\bar{y}_i(n) - y_i(n)]^2 \\ &= \lambda E(n-1) + \frac{1}{2} \sum_{i=1}^M [\bar{y}_i(n) - H^T(n) W_i(n-1)]^2. \end{aligned} \quad (3)$$

FIGURE 8: Influencing factors' prediction error (total voltage).

FIGURE 7: Influencing factors' prediction error (braking mileage proportion).

Among them, λ is called the forgetting factor, and $\delta_i(n)$, $\bar{y}_i(n)$, and $y_i(n)$ represent the output node error, expected output, and actual output, respectively.

The sample data of the adjacent sections of the same bus at the same time of departure in July 2013 were selected. The following figure (Figures 7 and 8) are based on these sample data of the impact factors, braking mileage proportion, and the total voltage according to the RBF neural network calculation of the prediction error map.

Since the number of hidden nodes in RBF networks and the centers of the hidden nodes were difficult to determine, the accuracy of the entire network was affected. The above figure shows that when the training sample selection was more random, the error reached more than 33%, and these errors directly led to the calculation of the influence degree of the influencing factors. In addition, as the input samples existed in various forms, including discrete values, continuous values, and missing values, the training samples were usually formed by random sampling. The center of the hidden layer basis function of the RBF neural network was selected from

the input sample set, which had a considerable dependence on the training samples, and in many cases it was difficult to reflect the real input-output relationship of the system. Moreover, if the initial center points too many cases caused by the optimization process, there will be data phenomenon, which is the key problem for RBF neural network modeling to solve the nonlinear system and is the sample selection problem. To solve this problem, in this study, all of the training samples were normalized, and the introduction of the gene expression programming (GEP) algorithm [15–17], the other algorithm [12, 18–31], and the method of influence degree calculation were based on RBF neural network optimization.

For the normalization, we used the maximum and minimum method to normalize the sample data to the range of $[0, 1]$:

$$y(k) = \frac{x(k) - \min(x(n))}{\max(x(n)) - \min(x(n))}, \quad k = 1, 2, \dots, N, \quad (4)$$

where $\min(x(n))$ is the minimum in the data sample of a specific influencing factor and $\max(x(n))$ is the maximum value in the data sample of a specific influencing factor.

After normalization, GEP was optimized and the RBF neural network was computed. GEP combines the advantages of genetic algorithm (GA) and genetic programming (GP) with the advantages of simple coding, strong local refinement, and broad adaptability of GP. The algorithm could genotype and phenotype to solve complex application problems through simple compact encoding; has strong global search ability; is a highly parallel, randomized, adaptive search algorithm; and overcomes the two types of GA operation with a low survival rate and semantic richness. The GEP optimization of the RBF neural network can be expressed as follows [16].

First, the chromosome was encoded as an expression tree, the initial cluster centers of the gene tail were automatically segmented and merged, and the new clustering centers and the number of centers were obtained. Second, the new clustering centers as the center vectors of the RBF neural network with the chromosome and the weight were used to construct an RBF neural network structure, and the sample data were input to the neural network to obtain the actual output. Then, we used the $J = (1/2) \sum_{p=1}^N \sum_{k=1}^L k = 1(\xi_k - y_k^p)^2$ (where N is the sample logarithm and L is the output node number of the network. ξ_k represents the expected output of the k neurons under the action of the sample p , and y_k^p represents the actual output of the first k neurons under the action of the sample p) formula for calculating the total network error and according to the fitness function for calculating the fitness value. The chromosomes with a large fitness value were retained in the next generation, and the GEP operator was used to perform the genetic manipulation of the chromosomes in the population to obtain the next-generation chromosomes. Then, the same was true for the next generation of chromosomes, with the evolution of chromosomes, the center vector, and the weights of the RBF neural network minimizing the total

error of the network (i.e., the maximum fitness value) in the direction of gradually closer, so as to achieve the purpose of optimizing the entire network.

Considering the actual situation of purely electric buses, in this study, we adjusted the input randomness of the GEP optimized RBF neural network; the specific steps follow: input: normalized dataset, GEP parameter, and its input power range array w , and output: trained RBF neural network (including influence degree of influencing factor).

Step 1. Initialize the RBF neural network, and select a number of weights from the input power range array w as initial weights.

Step 2. Select the fuzzy cluster center of the adjacent section of the station as the initial cluster center, form the individual chromosomes of GEP, and initialize the population.

Step 3. Segment and merge the initial cluster centers formed by the fuzzy clustering in the adjacent section of the station, and form a new clustering center according to the individuals' expressions.

Step 4. Use the new clustering center as the central vector of the RBF neural network, together with the weights in the chromosome, to construct the RBF network structure.

Step 5. Calculate the total network error $J = (1/2) \sum_{p=1}^N \sum_{k=1}^L k = 1(\xi_k - y_k^p)^2$ and the fitness of chromosomes according to the network structure constructed, and retain the chromosomes with the highest fitness in the next generation.

Step 6. Select the chromosomes by using the roulette strategy.

Step 7. Perform the cross operation on chromosomes by using the cross probability P_c .

Step 8. Mutate chromosomes according to mutation probability P_m .

Step 9. If the maximum number of iterations or the maximum fitness value convergence is performed to choose the optimal chromosome, otherwise the third step of individual expression was segmented and combined to form a new cluster center on the site of adjacent sections by fuzzy clustering in the initial cluster center.

Step 10. Decode the selected optimal chromosome and construct the final network structure.

Step 11. Output the RBF neural network optimized by GEP.

Simultaneously, we selected the sample data of the adjacent sections of the station for July 2013 to predict the impact factors. After the GEP optimization, the error maps were compared with the optimized ones, which are shown by Figures 9 and 10.

FIGURE 9: Influencing factor prediction error of RBF neural network before and after optimization (brake mileage proportion).

FIGURE 10: Influencing factors prediction error of RBF neural network before and after optimization (total voltage).

After the optimization of the RBF neural network by GEP, the influence of the influencing factor input on the i th hidden layer node can be expressed as follows:

$$E_{ij} = \frac{\text{cov}(h_i(t), x_j(t)) v_{ij}}{\text{var}(h_i(t)) \text{var}(x_j(t))}, \quad i = 1, 2, \dots, M, j = 1, 2, \dots, k, \quad (5)$$

where $\text{cov}(h_i(t), x_j(t))$ is the i th hidden layer node that outputs the covariance between $h_i(t)$ and the j th input influencing factor $x_j(t)$; $\text{var}(h_i(t))$ and $\text{var}(x_j(t))$ indicate the variance of $h_i(t)$ and $x_j(t)$, respectively; and v_{ij} is the correlation coefficient.

The influence of the i th hidden layer node on the output can be expressed as follows:

$$e_i = \frac{\text{cov}(h_i(t), y) w_i}{\text{var}(h_i(t)) \text{var}(y)}, \quad i = 1, 2, \dots, M, \quad (6)$$

where $\text{cov}(H_i, y)$ is the i th hidden layer node that outputs the covariance between the $h_i(t)$ and the network output y ; $\text{var}(h_i(t))$ and $\text{var}(y)$ indicate the variance of $h_i(t)$ and y , respectively; and w_i is the output weight.

The impact of the j th input factor on the output quantity can be expressed as follows:

$$\text{inf}_j = E_{ij} \times e_i. \quad (7)$$

All of the factors affecting the total influence degree can be expressed as follows:

$$\text{inf} = \sum_{j=1}^k |\text{inf}_j|. \quad (8)$$

Generally, for the influence degree of the influence factor in the order from large to small, from the maximum influence degree of the beginning of the year, if the cumulative results reach the influence degree sum of 85% or more, the factors affecting the influence degree are the cumulative corresponding results that can be regarded as important influence factors; these factors will play an important influence on the prediction results.

Calculated for the May 2012–April 2013 data, the impact factor' influence degree values of all the available quantitative information for the impact calculations are shown in Tables 1–5.

The ranking of the influencing factors was as follows: residual SOC value, energy feedback mileage proportion, season (air conditioning use), time and direction of travel, braking time proportion, uniform time proportion, acceleration/deceleration time proportion, minimum temperature, slipping mileage proportion, braking mileage proportion, energy feedback proportion, traffic congestion index, maximum temperature, slipping time proportion, congestion duration, uniform mileage proportion, average speed on main road, average deceleration, average speed of secondary trunk roads, average deceleration mileage proportion, average acceleration, weather, acceleration and deceleration mileage proportion, average accelerated mileage proportion, section weight coefficient, and congestion mileage proportion. The total percentage of factors reached 86.11%, and the following factors were considered the important factors: passenger number (2.33%), network stability index (2.06%), and number of congested sections of adjacent stations (2.00%). The proportion of the three influencing factors was more than 2%, and it also occupies a large proportion: outside lighting (light), voltage difference, total current, total voltage, ceiling voltage, and minimum voltage. Only 7.5% of the total number of the factors was affected. These factors were, in general, non-important factors.

TABLE 1: The influence degree of energy consumption influence factors (battery).

Influence factor	Residual SOC value	Total voltage	Ceiling voltage	Minimum voltage	Maximum temperature	Minimum temperature	Total current	Voltage difference
Influence degree value	172.42	39.65	26.78	23.34	102.34	123.29	42.98	56.63
Percentage	5.24%	1.20%	0.81%	0.71%	3.11%	3.74%	1.31%	1.72%

TABLE 2: The influence degree of energy consumption influence factors (traffic performance index).

Influence factor	Section weight coefficient	Network stability index	Average speed of main road	Average speed of secondary trunk roads	Traffic congestion index	Congestion mileage proportion	Congestion duration	The number of congested sections of adjacent stations
Influence degree value	84.12	67.76	97.54	89.87	102.52	78.98	99.81	65.82
Percentage	2.55%	2.06%	2.96%	2.73%	3.11%	2.40%	3.03%	2.00%

TABLE 3: The influence degree of energy consumption influence factors (driving characters).

Influence factor	Acceleration/deceleration time proportion	Uniform time proportion	Slipping time proportion	Braking time proportion	Average acceleration	Average deceleration	Energy feedback proportion
Influence degree value	123.73	130.23	102.30	132.76	87.98	94.06	148.98
Percentage	3.76%	3.96%	3.11%	4.03%	2.67%	2.86%	4.52%

TABLE 4: The influence degree of energy consumption influence factors (driving actual service life characteristics).

Influence factor	Acceleration and deceleration mileage proportion	Uniform mileage proportion	Slipping mileage proportion	Braking mileage proportion	Average accelerated mileage proportion	Average deceleration mileage proportion	Energy feedback mileage proportion
Influence degree value	87.67	98.84	114.74	112.93	84.69	88.79	108.67
Percentage	2.66%	3.00%	3.49%	3.43%	2.57%	2.70%	3.30%

TABLE 5: The influence degree of energy consumption influence factors (other).

Influence factor	Passenger number	Time and direction of travel	Outside lighting (light)	Weather	Season (air conditioning use)
Influence degree value	76.87	139.05	57.56	87.90	140.78
Percentage	2.33%	4.22%	1.75%	2.67%	4.28%

4. Influence Degree Application and Test Result

Using Guangzhou's 801 line of the purely electric bus as the research object, according to the influence degree of screening factors, the establishment of pure electric bus operating driving range forecasting model operates on the line of pure electric bus driving operation status prediction to verify. What needs to be explained is that some influencing factors have certain regularity, such as traffic data on the same day the same weekly class time has certain regularity, so this paper uses the same day the same departure time as the data sample under the same condition, prediction of operating driving range.

In general, 801 lines are maintained daily between 52–58 shifts; the working days and nonworking days are different, and the departure is adjusted according to the actual

situation. The buses on the 801 line departed between 6:00 and 23:00 on July 11, 2013, according to the half-hour statistics, the number of times, and departure time distribution diagram, which is shown by Figure 11.

The data from May 2012 to August 2013 were used as the training samples, and the actual data from September and October 2013 were used as the actual values to compare with the predicted values. The error values of the predicted values during the period were statistically analyzed. During the festival, there are 11 days (Mid-Autumn Festival, National Day holiday, and September 30 count as holidays), 11 days of weekend, and 39 days of work. In the removal of incomplete data or the frequency of information obvious data errors, for comparison with the predicted value of holidays, there are a total of 613 departures, a total of 597 weekend departures, and a total of 2105 work day departures.

FIGURE 11: Departure curve for bus line 801.

Table 6 shows the statistical results of the relative error and the absolute value of the partial energy consumption influencing factors. The factors considered were those that were very influential and could be quantified, such as minimum temperature, energy feedback mileage proportion, traffic congestion index, sliding time proportion, energy feedback proportion, braking mileage proportion, and uniform mileage proportion data.

Table 6 lists the prediction results of some influential factors. The total percentage of the seven influencing factors was 23.58%, which has a certain significance. Seven factors in the “number of sample errors less than 20%” statistics and the absolute value of error statistics mostly greater than 70%, basically can achieve more accurate prediction of the value of the impact factor requirements. Most of the absolute values for “number of sample errors less than 5%” were between 10% and 15%, and the accurate estimate of the running mileage was realized. Among the three categories of work days, weekends, and holidays, the influencing factors of the working days and the number of sample error less than 20%” statistics, in addition to the proportion of energy feedback proportion and brake mileage proportion of two influence factors, have reached more than 80%. The average value of the seven factors for the working days and the “number of sample errors less than 20%,” was 81.65%, which was considerably larger than the average of 74.11% and 69.99% on the weekends and on holidays, respectively. This implied that the influencing factors had strong regularity, but the regularity of weekends and holidays was relatively weak, which was consistent with the actual flow of traffic, passenger flow, and so on.

To visually express the difference between the predicted and the actual values of the influencing factors, the predicted values and the actual values were displayed on the coordinates in the form of scatterplots. If the predicted value was the same as the actual value, the data point fell on the $y = x$ line. If the predicted value was greater than the actual value, the data point fell below the straight line; otherwise, it fell over. The actual and the predicted values of all the energy

feedback proportions during the same operating period of September and October 2013 were used as examples to illustrate the deviations between the predicted and the actual values. Figure 12 shows the scatter diagram for the predicted values and the actual values of the factor of energy feedback proportion on the workdays.

The scatter plot of the energy feedback proportion was between 0.05% and 0.2%. The predicted and the actual values were larger than those of the other regions, and the mean square error of all the data was 0.0786. The Pearson coefficients $r = (\sum(x_i - \bar{x})(y_i - \bar{y})) / (\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2})$ were considered the basis of the correlation measurement, and the calculated value was $r = 0.9238$.

The scatter diagrams for the traffic congestion index, acceleration and deceleration proportion, braking time proportion, uniform mileage proportion, and the slipping mileage proportion are shown in Figures 13–17. The mean square errors between the predicted values and the actual values of the slipping mileage proportion, braking mileage proportion, braking time proportion, uniform time proportion, and acceleration/deceleration time proportion were 0.0657, 0.0877, 0.0723, 0.0802, and 0.0902, respectively, and the Pearson coefficients were 0.9109, 0.9321, 0.9103, 0.9331, and 0.9213, respectively.

The error of the operating mileage values of the purely electric bus was compared, by the ways of expressing the same and main factors and by comparing the operating mileage errors between actual and predicted values, indicating the degree of accuracy of the algorithm. Figure 18 shows the scatter diagram of the actual values and the predicted values of the driving range; the abscissa shows the predicted values, and the ordinate shows the actual values. Similar to the expression of the influencing factors, if the predicted value was the same as the actual value, the data point fell on the straight line. If the predicted value was greater than the actual value, the data point fell below the straight line; otherwise, it fell over the top.

The operating driving range’s test samples were concentrated in the mileage between 49 km and 56 km, in line with the actual operation of the law, the proportion of 81.45%. The mean square error of the predicted value was $MSE = 1.9673$, and the Pearson coefficient was $r = 0.9313$.

We extracted the actual mileage value of the test sample and analyzed the error between the data of 49–56 km. In Figure 19, the abscissa denotes the actual values of the operating driving range, and the ordinate represents the difference between the predicted value of the operating driving range and the actual value.

From the operating driving range’s predicted value and the actual value, the prediction error distribution of the two scatter diagrams was observed. The proposed algorithm achieved the operating driving range accurately, and the error control was in the acceptable range. The statistics of the overall results (including working days, weekends, and holidays) showed that the operating driving range’s prediction error value was less than or equal to 3% and less than or equal to 5% and less than or equal to 8%, respectively.

TABLE 6: The operating driving range influence factors forecasting results of statistics for pure electric bus (relative error values).

Influence factor	Prediction object	Sample size	Number of sample error less than 5%	Proportion of the sample error less than 5%	Number of sample error less than 10%	Proportion of the sample error less than 10%	Number of sample error less than 20%	Proportion of the sample error less than 20%	Number of sample error more than 20%	Proportion of the sample error more than 20%
Minimum temperature	Workdays	79990	14344	17.93%	38920	48.66%	65876	82.36%	14114	17.64%
	Weekends	22686	2988	13.17%	10098	44.51%	18004	79.36%	4682	20.64%
	Holidays	23294	2780	11.93%	9871	42.38%	18232	78.27%	5062	21.73%
Energy feedback mileage proportion	Workdays	79990	15021	18.78%	40987	51.24%	67213	84.03%	12777	15.97%
	Weekends	22686	1742	7.68%	10844	47.80%	15980	70.44%	6706	29.56%
	Holidays	23294	1892	8.12%	8109	34.81%	16231	69.68%	7063	30.32%
Traffic congestion index	Workdays	79990	18002	22.51%	41,984	52.49%	70024	87.54%	9966	12.46%
	Weekends	22686	3674	16.20%	14091	62.11%	20125	88.71%	2561	11.29%
	Holidays	23294	4402	18.90%	12990	55.77%	17034	73.13%	6260	26.87%
Sliding time proportion	Workdays	79990	12208	15.26%	39995	50.00%	64029	80.05%	15,961	19.95%
	Weekends	22686	3011	13.27%	15091	66.52%	19221	84.73%	3465	15.27%
	Holidays	23294	2677	11.49%	13044	56.00%	17302	74.28%	5992	25.72%
Energy feedback proportion	Workdays	79990	17201	21.50%	42201	52.76%	62981	78.74%	17,009	21.26%
	Weekends	22686	2312	10.19%	9981	44.00%	15472	68.20%	7214	31.80%
	Holidays	23294	2508	10.77%	14647	62.88%	14006	60.13%	9288	39.87%
Braking mileage proportion	Workdays	79990	9748	12.19%	37784	47.24%	59027	73.79%	20963	26.21%
	Weekends	22686	1670	7.36%	12231	53.91%	13098	57.74%	9588	42.26%
	Holidays	23294	1732	7.44%	14352	61.61%	15007	64.42%	8287	35.58%
Uniform mileage proportion	Workdays	79990	18009	22.51%	44998	56.25%	68039	85.06%	11951	14.94%
	Weekends	22686	4098	18.06%	12022	52.99%	15780	69.56%	6906	30.44%
	Holidays	23294	3119	13.39%	13005	55.83%	16424	70.51%	6870	29.49%

Note: the “number of sample error less than 10%” containing the “number of sample error less than 5%” and “number less than 5% error value is less than or equal to 10%” can be calculated like “number of sample error less than 10%” minus “number of sample error less than 5%.” Other similar.

FIGURE 12: Scatter diagram for predicted value and actual value (energy feedback proportion).

FIGURE 15: Scatter diagram of predicted values and actual values (braking time proportion).

FIGURE 13: Scatter diagram of predicted values and actual values (traffic congestion index).

FIGURE 16: Scatter diagram of predicted values and actual values (uniform range proportion).

FIGURE 14: Scatter diagram of predicted values and actual values (acceleration and deceleration time proportion).

FIGURE 17: Scatter diagram of predicted values and actual values (sliding range proportion).

FIGURE 18: Scatter diagram of predicted values and actual values (operating driving range, km).

FIGURE 19: Prediction error distribution (operating driving range, km).

According to the proportion of the operating lines that approved the longest mileage of 60 km, corresponding to the error values of less than 1.80 km and mileage of 3 km and 4.80 km, the average mileage was 54.13 km for the test samples, and the corresponding error values were less than 1.08 km, 2.17 km, and 3.25 km (they were 14.48%, 44.43%, and 81.87%, respectively), which meet the basic requirements of the pure electric bus operation operating driving range and the actual operation safety requirements.

5. Conclusions

This collection of the purely electric bus's location data and energy consumption data as the basic data, combined with the factors affecting the quantitative expression of the bus operation process, was used for analyzing the operating driving range's influencing factors and establishing a database for the purely electric bus. The RBF neural network algorithm was optimized by GEP, and the influencing factors of the operating driving range of the pure electric bus were

calculated. The influence degree of all of the factors was analyzed, and the influential factors were obtained. Based on the fuzzy clustering algorithm and the fuzzy time series algorithm, a prediction model of the purely electric bus's operating driving range was established in this study. The experimental results showed that this method could simplify the mathematical model and reduce the computational complexity and did not have a greater effect on the accuracy of the prediction results. Although the vehicle hardware continued to progress, the amount of data required by the application continued to increase, but this required a certain amount of mobile bandwidth, computing capabilities of mobile devices, and so on. In this study, through the optimization algorithm, we attempted to improve the efficiency of the bandwidth and the application data; to realize the safe operation of electric buses, battery protection, battery energy management, line adjustment, and estimation of the purely electric bus's emergency scheduling mileage; and to provide decision-support charging station planning. The further promotion of the use of the purely electric buses, hybrid buses, and other new energy vehicles is of considerable significance. At the same time, it provides experience for other applications that need to be screened and analyzed.

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 they have no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (no. 51705365), Special Fund of Frontier and Key Technology Innovation of Guangdong Province (provincial major science and technology project) (nos. 2014B010112008, 2015B010106001, 2015B010129003, and 2016B010109007), Pearl River S&T Nova Program of Guangzhou (no. 201610010034), and Guangdong Science and Technology Program (nos. 2013B01010201, 2013B010102018, and 2013B090600152). We thank Let-Pub (<http://www.letpub.com>) for its linguistic assistance during the preparation of this manuscript.

References

- [1] P. Chuanqin and B. Yufeng, "Study on influence factors screening of highway traffic safety evaluation," *Highway Traffic Science and Technology (Applied Technology Edition)*, vol. 22, no. 11, pp. 246–248, 2015.
- [2] S. Wang, F. Ma, and C. Jun, "Influencing factor analysis on typical urban passenger transport mode choice," *Journal of Transportation Systems Engineering and Information Technology*, vol. 10, no. 3, pp. 93–98, 2010.
- [3] C. Xu, Z. Qu, P. Tao, and S. Jin, "Methods of real-time screening and reconstruction for dynamic traffic abnormal data,"

- Journal of Harbin Engineering University*, vol. 37, no. 2, pp. 211–217, 2016.
- [4] P. Yulong and J. Ma, “Real-time traffic data screening and reconstruction,” *China Civil Engineering Journal*, vol. 36, no. 7, pp. 78–83, 2003.
 - [5] L. Gang, G. Jiang, X. Zhang, and J. Wang, “Screening and checking for ITS traffic sensor data,” *Journal of Jilin University: Engineering and Technology Edition*, vol. 34, no. 1, pp. 122–126, 2004.
 - [6] P. Yulong and B. Junying, “Selection of the key influencing factors of urban traffic structure evolution,” *Journal of Transport Science and Engineering*, vol. 33, no. 1, pp. 66–71, 2017.
 - [7] H. Li, X. Peng, Z. Zhong, Z. Bai, and H. Hong, “Influence factors and control countermeasures of traffic noise based on grey relational analysis,” *Noise and Vibration Control*, vol. 32, no. 1, pp. 93–95, 2012.
 - [8] Z. Yang and L. Xiaoming, “The effect of urban travel cost on travel structure,” *Journal of Transportation Systems Engineering and Information Technology*, vol. 12, no. 2, pp. 1–26, 2012.
 - [9] C. Yang, Z. Li, R. Cui, and B. Xu, “Neural network-based motion control of an under-actuated wheeled inverted pendulum model,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 11, pp. 2004–2016, 2014.
 - [10] C. Yang, Y. Jiang, Z. Li, W. He, and C. Y. Su, “Neural control of bimanual robots with guaranteed global stability and motion precision,” *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1162–1171, 2017.
 - [11] C. Yang, X. Wang, L. Cheng, and H. Ma, “Neural-learning based telerobot control with guaranteed performance,” *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3148–3159, 2017.
 - [12] C. Yang, X. Wang, Z. Li, Y. Li, and C. Y. Su, “Teleoperation control based on combination of wave variable and neural networks,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 2125–2136, 2017.
 - [13] B. Xu, Z. Shi, C. Yang, and F. Sun, “Composite neural dynamic surface control of a class of uncertain nonlinear systems in strict-feedback form,” *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2626–2634, 2014.
 - [14] B. Xu and F. Sun, “Composite intelligent learning control of strict-feedback systems with disturbance,” *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 730–741, 2018.
 - [15] T. Junjian, Y. Changan, and C. Hongguo, “Enhanced RBF neural network algorithm based on gene expression programming,” *Journal of Chinese Computer Systems*, vol. 31, no. 5, pp. 950–954, 2010.
 - [16] C. Ferreira, *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*, Springer-Verlag, Berlin Heidelberg, 2nd edition, 2006.
 - [17] S. Deng, C. Yuan, B. Zhao et al., “Attribution reduction classification algorithms based on GEP and neural network,” *Computer Engineering and Applications*, vol. 23, pp. 154–157, 2006.
 - [18] W. Li, C. Yang, Y. Jiang, X. Liu, and C. Y. Su, “Motion planning for omnidirectional wheeled mobile robot by potential field method,” *Journal of Advanced Transportation*, vol. 2017, Article ID 4961383, 11 pages, 2017.
 - [19] Z. Li, C. Yang, C. Y. Su, J. Deng, and W. Zhang, “Vision-based model predictive control for steering of a nonholonomic mobile robot,” *IEEE Transactions on Control Systems Technology*, vol. 24, no. 2, pp. 553–564, 2015.
 - [20] C. Yang, C. Zeng, P. Liang, Z. Li, R. Li, and C.-Y. Su, “Interface design of a physical human–robot interaction system for human impedance adaptive skill transfer,” *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, pp. 329–340, 2018.
 - [21] C. Yang, K. Huang, H. Cheng, Y. Li, and C. Y. Su, “Haptic identification by ELM-controlled uncertain manipulator,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 2398–2409, 2017.
 - [22] C. Yang, H. Wu, Z. Li, W. He, N. Wang, and C.-Y. Su, “Mind control of a robotic arm with visual fusion technology,” *IEEE Transactions on Industrial Informatics*, vol. 13, p. 1, 2017.
 - [23] Z. Zhao, Z. Liu, Z. Li, N. Wang, and J. Yang, “Control design for a vibrating flexible marine riser system,” *Journal of the Franklin Institute*, vol. 354, no. 18, pp. 8117–8133, 2017.
 - [24] Z. Zhao, X. Wang, C. Zhang, Z. Liu, and J. Yang, “Neural network based boundary control of a vibrating string system with input deadzone,” *Neurocomputing*, vol. 275, pp. 1021–1027, 2018.
 - [25] Z. Zhao, Y. Liu, and F. Luo, “Output feedback boundary control of an axially moving system with input saturation constraint,” *ISA Transactions*, vol. 68, pp. 22–32, 2017.
 - [26] Z. Zhao, J. Shi, X. Lan, X. Wang, and J. Yang, “Adaptive neural network control of a flexible string system with non-symmetric dead-zone and output constraint,” *Neurocomputing*, vol. 283, pp. 1–8, 2017.
 - [27] C. Yang, J. Luo, Y. Pan, Z. Liu, and C. Y. Su, “Personalized variable gain control with tremor attenuation for robot teleoperation,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 99, pp. 1–12, 2017.
 - [28] Z. Zhao, Y. Liu, F. Luo, and W. He, “Adaptive boundary control of an axially moving belt system with high acceleration/deceleration,” *IET Control Theory & Applications*, vol. 10, no. 11, pp. 1299–1306, 2016.
 - [29] L. Wang, J. Yang, N. Zhang et al., “A spatial-temporal estimation model of residual energy for pure electric buses based on traffic performance index,” *Tehnicki vjesnik - Technical Gazette*, vol. 24, no. 6, pp. 1803–1811, 2017.
 - [30] L. Wang, J. Yang, N. Zhang et al., “Time-space relationship analysis model on the bus driving characteristics of different drivers based on the traffic performance index system,” *Tehnicki vjesnik - Technical Gazette*, vol. 25, no. 1, pp. 236–244, 2018.
 - [31] L. Wang, S. Lin, J. Yang et al., “Dynamic traffic congestion simulation and dissipation control based on traffic flow theory model and neural network data calibration algorithm,” *Complexity*, vol. 2017, Article ID 5067145, 11 pages, 2017.

Submit your manuscripts at
www.hindawi.com