Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8151720, 11 pages
https://doi.org/10.1155/2020/8151720

Research Article

Development of a Practical Method to Estimate the Eco-Level of Driver Performance

Jianbin Zheng[1] **and Yiping Wu**[2]

[1]*Faculty of Information Technology, Beijing University of Technology, No. 100 Ping Le Yuan, Beijing 100124, China*
[2]*Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, No. 100 Ping Le Yuan, Beijing 100124, China*

Correspondence should be addressed to Yiping Wu; wuyiping@bjut.edu.cn

Received 11 November 2019; Revised 16 September 2020; Accepted 15 October 2020; Published 27 October 2020

Academic Editor: Jose E. Naranjo

Motor vehicle's fuel consumption is one of the main sources of energy consumption in road transportation and is highly influenced by driver performance in the process of driving. Eco-driving behavior has been proved to be an effective way to improve the fuel efficiency of vehicles. Essential to the efforts towards saving vehicle fuel is the need to estimate the eco-level of driver performance accurately and practically. Depending on on-board diagnostics and Global Position devices, individual vehicle's instantaneous fuel consumption, engine revolution and torque, speed, acceleration, and dynamic location were collected. Back-propagation network was adopted to explore the relationship between vehicle fuel consumption and the parameters of driver performance. Taking 700 data samples in basic segments of urban expressways as our training set and 100 data samples as validation test, we found the optimal model structure and parameters through repeated simulation experiments. In addition to the average and standard deviation value, the fluctuation frequency of driver performance data was also viewed as influence factors in eco-level estimation model. The average estimation accuracy of our developed model has been tested to be 96.37%, which is quite higher than that of linear regression model. The study results provide a practical way to evaluate drivers' performance from the perspective of fuel consumption and thus give basis for rewarding best drivers within eco-driving programs.

1. Introduction

Fuel consumption of motor vehicles is one of the main sources of energy consumption in road transportation and has become one serious problem impacting the sustainable development of urban traffic system. As stated in the government statistics reports of Beijing, the energy consumption in road transport sector was 2,906,000 tons (standard coal) in the year of 2016, accounting for 22.7% of the total energy usage in the transportation field of this city [1]. To save natural resources, protect the environment, and improve the wellbeing of the general public, the Beijing government is making great efforts to reduce fuel consumption of vehicular traffic.

Till now, a series of measures have been taken to decrease vehicle fuel consumptions. Typical approaches have been adopted, including clean energy development, electric vehicle bringing, vehicle technology improvements, traffic flow and control optimization, and driver performance optimization. Among these measures, optimizing driver performance is one of the most important ones. Eco-driving behavior, which is also known as green driving, has been proved to be an effective way to improve the fuel efficiency of vehicles [2–4]. The concept of eco-driving refers to smooth and steady driving including starting smoothly, reducing the number of instances of sharp acceleration and deceleration, recommending the cruise mode, and anticipating the traffic flow to minimize the use of the brake and accelerator [5, 6].

It is first necessary to accurately estimate the eco-level of driver performance corresponding to various vehicle running status in the process of driving, as the primary support for improving driver performance more targetedly and effectively and reducing fuel consumption. Therefore, we need a practical way to evaluate drivers' performance from the perspective of fuel consumption. In fact, there have been several methods to calculate vehicle fuel consumption at

both macroscopic and microscopic levels. COPERT [7], EMFAC [8], and MOVES [9] were the three typical and mostly used macroscopic models in fuel consumption estimation. They could be used to calculate the total vehicle fuel consumption and predict future trends for a country or a city or at the project level. However, these models are based on general flow characteristics and might not be the optimal choices to calculate vehicle fuel consumption for individual drivers. Considering vehicle operation modes of deceleration, models of idle, and acceleration, microscopic models are often more accurate when estimating fuel consumption for driver performance. Generally, microscopic vehicle fuel consumption models could be classified as three types [10], which were, respectively, based on (1) the simulation of engine power, (2) the driving modes (e.g., acceleration, deceleration, cruise, and idling), and (3) statistical results of vehicle speed and acceleration. The calculation methods mainly included multiple regression, theoretical derivation, and engine bench tests [11].

Among fuel consumption models based on the simulation of engine power, ADVISOR [12], PSAT [13], and EVSIM [14] were the most commonly used models in automobile engineering field. They simulated vehicle operating status and its corresponding power flow with driving cycles. To obtain vehicle fuel consumption, parameters of vehicle features, engine types, driving conditions, and many other factors were needed. For the second type of microscopic vehicle fuel consumption model, the common operating status in different driving conditions was classified, and then the fuel consumption in each driving mode was measured and calculated. Thus, the total fuel consumption was the sum of the vehicle fuels used in all driving modes [15, 16]. MODEM [17] was a representative model to calculate fuel consumption based on the statistical results of vehicle speed and acceleration. In this model, data of vehicle fuel consumption were classified according to two indexes: speed and the product of speed multiplied by acceleration. Thus, the vehicles' instantaneous fuel usage could be figured out based on the combination of these two indexes.

Another popular approach to calculate vehicle fuel consumption was based on the carbon balance method. For example, the microscopic emissions model based on vehicle specific power (VSP) distribution could be used to estimate the vehicle emissions second by second [9, 18]. As gasoline is a kind of compound composed of carbon and hydrogen compounds, it will produce different amounts of CO, CO_2, HC, H_2O, and NO_X after burning. Regardless of the combustion degree, the carbon elements in CO, CO_2, and HC are always equal to the carbon in the gasoline consumption. Thus, based on the emissions, the fuel consumption could be obtained [4, 19].

The aforementioned microscopic vehicle fuel consumption models have been developed in the past. These models had a high accuracy in predicting vehicle fuel consumption; however, the data needed to calculate vehicle fuel consumption was intense. Particularly, the parameters in these models were either various or fine-grained, making it difficult for researchers to collect data in a real driving environment or at least the cost would be very high. Also, such models developed in one country may not work well for other countries because of differences in vehicle fleet and engine technologies. Therefore, it was inconvenient to evaluate the eco-level of driver performance accurately and practically. Owing to these limitations, it is insufficient to support rewarding best drivers within eco-driving programs and training drivers by more targeted and effective ways.

According to previous studies [20–22], the parameters of driver performance lead to a vehicle power demand; that in turn leads to an engine power demand and then to the consumption of engine. In the process of driving, driver performance is the dynamic reflection of drivers' comprehensive decision about road geometry, traffic control strategy, environment stimulation factors, and their interactions [23]. Thus, the relationship between driver performance and fuel consumption would not be linear or obvious but chaotic and hidden. The traditional estimation and predictive methods based on statistic models might not be suitable for accurately evaluating vehicle fuel consumption based on driver performance because of limited expressing capability for complicated relationships [24]. In contrast, the method of machine learning with excellent data processing ability and hidden features mining would be effective in driver performance modeling, evaluation, and prediction. Currently, shallow machine learning models including Decision Tree [25], Hidden Markov Model [26], Gaussian Mixture Model [27], Support Vector Machine [28], and Network of Neuron [29] have been widely adopted for individual driving habit modeling, unsafe driving behavior (e.g., driving distraction, fatigue, and drunk driving) identification, and traffic flow prediction. More importantly, the predictive or identificative accuracy of these machine learning models was proved to be acceptable.

In addition to these estimation model construction methods of vehicle fuel consumption, the Internet and cloud computing technology have been changing and replacing traditional data sensing methods, which further accelerated the process of big data aggregation [30]. In the traffic areas, the data collection approaches of induction coil and microwave radar with fixed locations and movement detection based on probe vehicles are gradually disappearing. Instead, new data detection methods, such as Controller Area Network [31], satellite navigation system [32], and smartphone [33], have been widely applied for their obvious advantages of convenience and widespread implementation. In particular, the second-by-second vehicle operating status data during driving process in real environment could be gathered by on-board diagnostics (OBD) and Global Position System (GPS) devices, and these dynamic data might be more easily detected in the coming connected and automated vehicle (CAV) environment.

Thus, the current study aims at developing a practical model to precisely estimate the eco-level of individual driver performance during driving process. The database used in this study was the real-time data collected by on-board OBD + GPS devices in real driving environments. Because of apparent advantages in self-learning, self-organizing, good fault tolerance, and excellent nonlinear approximation ability, back-propagation (BP) neural network was finally

selected and applied in this study after comparison with those shallow machine learning models commonly used in traffic areas [34]. In particular, previous studies have illustrated that the BP network based model has good performance in vehicle fuel consumption prediction or estimation [29, 35, 36].

The rest of the paper is organized as follows. First, the database used in this study was introduced along with the data collection program. This was followed by a presentation of the eco-level estimation model construction including the model structure, parameters designing, and simulation test. After that, the test results of model accuracy were exhibited and we also discussed their reasonability and applicability. Finally, we summarized the main conclusions of the study and exhibited the limitation and future research needs of the current research.

2. Data Collection

The data used in this study were collected from our established driving behavior platform based on Internet + technology [37]. In this platform, the dynamic operating data of vehicles during driving were obtained by on-board diagnostics (OBD) and Global Position System (GPS) devices mounted on taxicabs. The real-time vehicle operating status in real driving condition was transmitted to the cloud for storage through 3G network. A local server was established to download and store the data needed for different applying purposes from the cloud platform.

The vehicle operation data were collected second by second. The data items include vehicle speed (km/h), engine speed (rad/min), instantaneous fuel consumption (0.01 L/h), and real-time locations (i.e., the latitude and longitude). Considering that vehicle operating status is instantaneous and dynamic, most of the basic data collected by OBD and GPS were transmitted to the cloud platform per second. In case of data missing, the basic data were also packaged and uploaded in a frequency of five minutes.

A total of 140 taxi drivers and their taxicabs were employed for data collection which lasted for four months from January to April 2016. Meanwhile, all taxicabs hired in the study are of the same vehicle type and are almost in the same working condition. All of these cabs were put into operation in 2013 and are Hyundai Elantra with 4 cylinders and 1.6-liter engine. They are certified by the National Level IV emission standard.

Using the latitude and longitude coordinates collected by the GPS mounted on vehicles and the existent map information, each vehicle's travelling path could be obtained. To build a basic model to estimate the eco-level of driver performance during driving, possible interference factors were excluded, such as roadway conditions, traffic control devices, traffic signal status, and many others. In data processing, basic sections of expressways were obtained by latitude and longitude coordinates screening through Matlab software. Except for the influence of entrance or exit, the vehicle operating data locating basic sections of Beijing expressways (i.e., straight and flat part) were used for developing our model. A sum of 3,709 data segments in expressway basic sections was acquired by matching the collection data and road network base map. After removing the invalid data from anomalous values, a total of 2,786 data segments were valid.

In the current study, the eco-level was represented by vehicle fuel consumption per 100 kilometers, which is calculated by the instantaneous fuel consumption and travel distance. It is true that the vehicle fuel consumption collected by OBD was not totally the same with standard procedure to measure fuel consumption via a calibrated tank. Actually, the instantaneous fuel recorded via OBD would be somewhat deviated from the true value. However, the total or average fuel consumption during a distance could be obtained with acceptable precision through OBD equipment. This conclusion has been tested and verified in our previous studies [37, 38]. Driver performance refers to vehicle speed, acceleration, and revolution and torque of vehicle engine. In order to construct the estimation model to analyze the eco-level of driver performance and test its validation, 700 data samples were randomly selected for model trials and 100 data samples were used to test model accuracy.

3. Estimation Model Development

To construct the BP network based eco-level estimation model, our main task is to obtain the optimal model structure and its corresponding parameters according to the features of driver performance and vehicle fuel consumption data. As BP network with a single hidden layer could approximate any continuous function in a closed interval, such a network with three layers could be adopted to realize any reflections from n-dimensional to m-dimensional [39]. Hence, our current study focuses on establishing eco-level estimation model of individual driver performance based on three-layer BP network. In this section, contents including model's input and output design, node number of hidden layer selection, function, and learning rate choice would be mainly discussed.

3.1. Input Design. The input of our developing BP network model was set as the characteristic parameters of driver performance collected from on-board OBD and GPS devices. Driver performance data items highly related to fuel consumption were selected as independent variables, including velocity (V), acceleration (A), engine revolution (R), and torque (T) [40]. In general, under the same road, traffic, and environment conditions, the variation of fuel consumption for a fixed vehicle in the process of driving was predominantly resulting from the change of driver performance. It is assumed that the frequent change of driver performance would lead to high vehicle fuel consumption. Thus, in addition to the basic statistical values like average and standard deviation, the fluctuation frequency of driver performance was also treated as a prominent influence factor in fuel consumption.

In this study, the fluctuation of driver performance was shown as the performance data (i.e., V, A, R, and T) was significantly changed at one moment when compared to the whole driving process. Accordingly, the fluctuation of driver performance could be classified to one, two, three, or four fluctuation items at the same time. Specifically, one item fluctuation represents that only one parameter of driver performance changed obviously at one given moment, while several item fluctuations indicate that at least two driver performance items simultaneously varied significantly at this moment. Based on this, the percentage of driver performance fluctuation during whole driving process was obtained as an influence factor in vehicle fuel consumption (i.e., eco-level). The calculation methods for the percentage of driver performance fluctuation were stated as equations (1) to (16).

3.1.1. Fluctuation Percentage of One Driver Performance Item (P1).

$$P_1 = \frac{M_V + M_A + M_R + M_T}{M}, \tag{1}$$

where M_V is the number of velocity fluctuations at a given moment during whole driving period; M_A is the number of acceleration fluctuations at a given moment during whole driving period; M_R is the number of engine revolution fluctuations at a given moment during whole driving period; M_T is the number of engine torque fluctuations at a given moment during whole driving period; M is the data recorded number of the whole driving period.

Taking velocity as example, M_V could be obtained by equations (2) to (4). Similarly, the fluctuations number of the other three indexes (i.e., acceleration, engine revolution, and torque) could be calculated:

$$M_V = \sum_{j=1}^{M} w_j, \tag{2}$$

$$w_j = \begin{cases} 1, & \left|\Delta V_j\right| > \Delta V_{85} = \text{Percentile}\left(\left|\Delta V_1\right|, \left|\Delta V_2\right|, \left|\Delta V_3\right|, \ldots, \left|\Delta V_M\right|, 0.85\right), \\ 0, \end{cases} \tag{3}$$

$$\Delta V_j = V_{j+1} - V_j, \tag{4}$$

where V_j is the velocity value at time j; ΔV_{85} is the 85th percentile value of velocity variation during a given time period; w_j is the statistical value of velocity fluctuation at time j. The value of w_j would be increased to one when the change of velocity exceeded ΔV_{85} at time j.

Figure 1 shows the moment of velocity fluctuation when taking the data of velocity during 200 meters as an example. The blue line, red-dotted line, and green straight line represent the original value, absolute change value, and 85th percentile of absolute change value, respectively. When the red-dotted line exceeds the green line, it is the moment when the vehicle velocity changes significantly.

3.1.2. Fluctuation Percentage of Two Driver Performance Items (P2). The calculation method of P2 is shown in equations (5) to (8):

$$P_2 = \frac{M_{VA} + M_{VR} + M_{VT} + M_{AR} + M_{AT} + M_{RT}}{M}, \tag{5}$$

where M_{VA} is the number of both velocity and acceleration fluctuations at a given moment during whole driving period; M_{VR} is the number of both velocity and revolution fluctuations at a given moment during whole driving period; M_{VT} is the number of both velocity and torque fluctuations at a given moment during whole driving period; M_{AR} is the number of both acceleration and revolution fluctuations at a given moment during whole driving period; M_{AT} is the number of both acceleration and torque fluctuations at a given moment during whole driving period; M_{RT} is the number of both revolution and torque fluctuations at a given moment during whole driving period. The calculation methods of MVA are shown as equations (6) to (8), which are similar for M_{VT}, M_{AR}, M_{AT}, and M_{RT}:

$$M_{VA} = \sum_{j=1}^{M} h_j, \tag{6}$$

$$h_j = \begin{cases} 1, \left|\Delta V_j\right| > \Delta V_{85} \& \left|\Delta A_j\right| > \Delta A_{85}, \\ 0, \end{cases} \tag{7}$$

$$\Delta A_j = A_{j+1} - A_j, \tag{8}$$

where A_j is the acceleration value at time j; ΔA_{85} is the 85th percentile value of acceleration variation during a given time period; h_j is the statistical value of both velocity and acceleration fluctuations at time j. The value of h_j would be increased to one, when the change of velocity exceeded ΔV_{85} and the change of acceleration exceeded ΔA_{85} at time j.

3.1.3. Fluctuation Percentage of Three Driver Performance Items (P3). Equations (5) to (8) illustrate the calculation method of $P3$:

$$P_3 = \frac{M_{VAR} + M_{VAT} + M_{VRT} + M_{ART}}{M}, \tag{9}$$

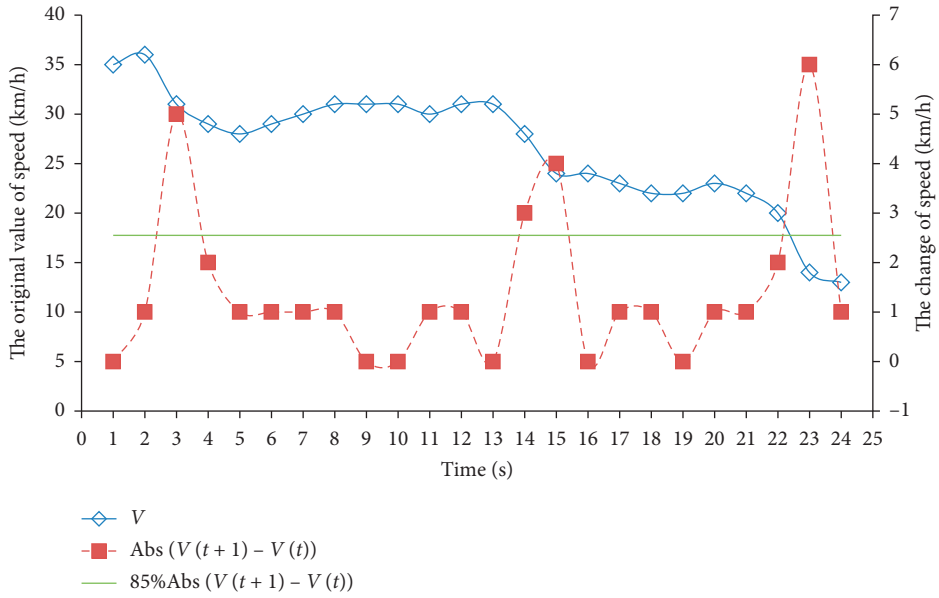

Figure 1: Illustration of the fluctuation of vehicle velocity (taking 200 meters as an example).

where M_{VAR} is the number of simultaneous fluctuations of velocity, acceleration, and revolution at a given moment during whole driving period; M_{VAT} is the number of simultaneous fluctuations of velocity, acceleration, and torque at a given moment during whole driving period; M_{VRT} is the number of simultaneous fluctuations of velocity, revolution, and torque at a given moment during whole driving period; M_{ART} is the number of simultaneous fluctuations of acceleration, revolution, and torque at a given moment during whole driving period. M_{VRT} is the number of simultaneous fluctuations of velocity, revolution, and torque at a given moment during whole driving period.

Equations (10) to (12) displayed the calculation methods of M_{VAR}, which are similar to three other indexes: M_{VAT}, M_{VRT}, and M_{ART}:

$$M_{VAR} = \sum_{j=1}^{M} k_j, \tag{10}$$

$$k_j = \begin{cases} 1, & \left|\Delta V_j\right| > \Delta V_{85} \& \left|\Delta A_j\right| > \Delta A_{85} \& \left|\Delta R_j\right| > \Delta R_{85}, \\ 0, \end{cases} \tag{11}$$

$$\Delta R_j = R_{j+1} - R_j, \tag{12}$$

where R_j is the engine revolution value at time j; ΔR_{85} is the 85th percentile value of revolution variation during a given time period; k_j is the statistical value of simultaneous fluctuation of velocity, acceleration, and revolution at time j. The value of k_j would be increased to one, when the change of velocity, acceleration, and revolution exceeded ΔV_{85}, ΔA_{85}, and ΔR_{85}, respectively, at the same time j.

3.1.4. Fluctuation Percentage of Four Driver Performance Items (P4). The calculation method of $P4$ is shown in equations (13) to (16):

$$P_4 = \frac{M_{VART}}{M}, \tag{13}$$

$$M_{VART} = \sum_{j=1}^{M} q_j, \tag{14}$$

$$q_j = \begin{cases} 1, & \left|\Delta V_j\right| > \Delta V_{85} \& \left|\Delta A_j\right| > \Delta A_{85} \& \left|\Delta R_j\right| > \Delta R_{85} \& \left|\Delta T_j\right| > \Delta T_{85}, \\ 0, \end{cases} \tag{15}$$

$$\Delta T_j = T_{j+1} - T_j, \tag{16}$$

where M_{VART} is the number of simultaneous fluctuations of velocity, acceleration, revolution, and torque at a given moment during whole driving period; T_j is the torque value at time j; ΔT_{85} is the 85th percentile value of torque variation during a given time period; q_j is the statistical value of simultaneous fluctuation of velocity, acceleration, revolution, and torque at time j. The value of q_j would be increased to one, when the change of velocity, acceleration, revolution, and torque exceeded ΔV_{85}, ΔA_{85}, ΔR_{85}, and ΔT_{85}, respectively, at the same time j.

3.2. Output Design. The output of our establishing estimation model was set as the eco-level. In order to facilitate comparative analysis of various drivers with different running time or distances, the vehicle fuel consumption was uniformed as liters per 100 kilometers. For perceptual intuition purpose and, more importantly, to practically qualify drivers' performance in terms of fuel consumption, vehicle fuel consumption should be translated to fuel ranks (e.g.,

excellent, good, fair, and poor) or fuel scores (e.g., 75 out of a perfect 100 points). The fuel scores were finally used to reflect the eco-level in this study. The highest fuel consumption got the lowest score; correspondingly, the lowest got the full score. In this study, the range of fuel consumption scores was from 40 to 100 points, as shown in the following equation:

$$\text{Eco} - \text{level} = 40 + \left(1 - \frac{(FC_i - FC_{\min})}{FC_{\max} - FC_{\min}}\right) * 60, \quad (17)$$

where eco-level is the fuel consumption level of vehicle i in one basic expressway segment; FC_{i} is the fuel consumption of vehicle i in one basic expressway segment, $L/100$ km; $FC_{\max}$ is the maximum value of FC_{i} in all basic expressway segments; FC_{min} is the minimum value of FC_{i} in all basic expressway segments.

According to equation (17) and the experimental data collected, the eco-level (i.e., fuel score) in the basic segments of expressways was obtained, as shown in Figure 2.

3.3. Node Number of Hidden Layer Selection. One of the key issues of BP network based estimation model construction is to design an appropriate node number of hidden layer. However, there is still no certain approach existing to determine the suitable number of nodes in hidden layer corresponding to various tasks. Thus, according to previous studies, repeated experimental tests were applied to find out the optimal node number of hidden layer in developing our BP network based eco-level estimation model.

Firstly, the experiential equations (18) to (20) were used to calculate the possible range of node number [39–41], taking the numbers of both input and output indexes into account:

$$n_{\text{node}} = \sqrt{n_{\text{input}} + n_{\text{output}}} + \alpha, \quad (18)$$

$$n_{\text{node}} = \sqrt{n_{\text{input}} \times n_{\text{output}}}, \quad (19)$$

$$n_{\text{node}} = \log_2 n_{\text{input}}, \quad (20)$$

where n_{node} is the suitable node number in hidden layer; n_{input} is the number of input indexes; n_{ouput} is the number of output indexes; α is constant, and the value is from zero to ten.

According to our input and output design stated above, the numbers of input and output indexes are twelve and one, respectively. Therefore, according to equations (18) to (20), the node number should be selected from the value range of 4 to 14. In order to find the most suitable node number from the possible values, the node number was obtained from an arithmetic progression range with a tolerance of 2 for optimal value test. Taking the prediction error as a control objective, we should find the optimal value of node number from 4, 6, 8, 10, 12, and 14 in line with the smallest prediction error.

To obtain relatively stable prediction results under the influence of node numbers, every BP network model with a given node number (i.e., [4, 14] with a tolerance of 2) was run 10 times. Figure 3 displays the relationship between the node number in hidden layer and the mean prediction error. It indicated that the average prediction error is the smallest when the number of nodes was 10. Therefore, the hidden layer of the developed BP network based estimation model should be designed as 10 neuron nodes.

3.4. Function Selection. The same as other typical BP network models [39, 42, 43], the transfer functions selected in this study were general types. Namely, the transfer function from input to output layer was set as "tansig" (i.e., a S-type tangent function) and the function from hidden to output layer was "purelin" (i.e., a linear function).

Since training functions would apparently affect mode training speed and might further influence the accuracy of predictive results, the most suitable training function used in this study was also confirmed from repeated experiment tests. Five common training functions were selected as candidates by referring to other similar models developed in previous researches [44]. The prediction accuracy and training speed were controlled as evaluation indexes for comparison and selection of model training function. In order to make predictive result more reliable and steady, we tested the BP network based estimation model with every candidate training function ten times. The testing results of mean prediction accuracy and training speed of each training function were obtained and are displayed in Table 1.

As illustrated in Table 1, the function of "traingdm" should be the best one when taking both the prediction accuracy and training speed into account. For our developed BP network based estimation model, the "traingdm" function has both higher prediction accuracy and faster training speed when compared to other training functions.

3.5. Learning Rate Selection. In the BP network, the weight variation for each loop iteration was determined by learning rate. The prediction error would be smaller at the end of iteration if setting the learning rate as smaller value; however, the model convergence rate would be slower accordingly because of increasing learning time. In general, the method of learning rate selection was based on previous studies and repeated trials [39–44]. Usually, the optimal learning rate should have the smallest sum of squared errors through comparative analysis.

In this study, the value range of learning rate was set from 0.01 to 0.09. Then, repeated experimental tests were went through to find out the optimal value of learning rate. With a tolerance of 0.01, the candidate learning rate was designed as an arithmetic progression. Every BP network model was run ten times with a given learning rate. Figure 4 exhibits the relationships between the average and standard deviation of sum of squared errors and different learning rates. Obviously, both the average and standard deviation are the smallest when selecting learning rate as 0.03. Thus, the optimal learning rate should be 0.03 for our developed BP network based estimation model.

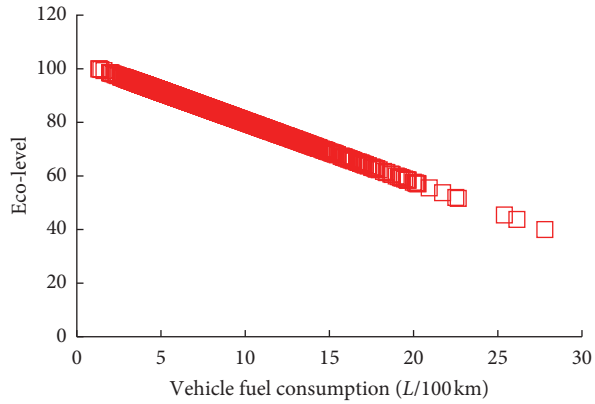

Figure 2: The relationship between eco-level and fuel consumption in the basic segments of urban expressways.

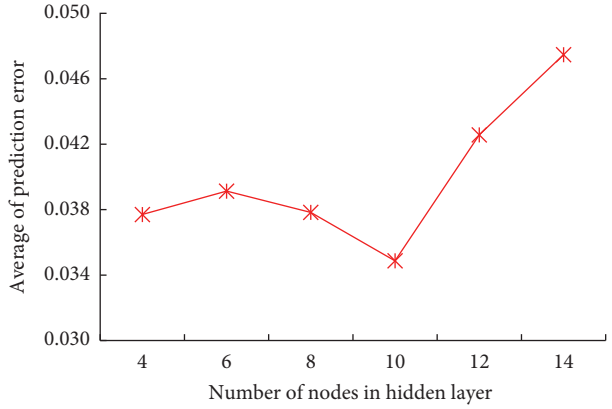

Figure 3: Relationship between model prediction error and the number of nodes in hidden layer.

Table 1: Prediction accuracy and training speed of each candidate training function.

Training function	Prediction accuracy (%)	Training speed (s)
Trainr ( )	96.8	109.368
Trainbfg ( )	96.1	2.0199
Traingd ( )	96.3	0.7010
Traingdm ( )	96.5	0.636
Trainlm ( )	95.6	1.588

3.6. Results of Model Construction. Summarizing the above, the structure, parameters, and functions of our eco-level estimation model in accordance with driver performance were developed. According to these separate test results, we also tried to find the most suitable model format by adjusting the node number in hidden layer, learning rate, and functions simultaneously. After repeated trials, the most appropriate structure of our constructed BP neural network model was obtained. In our developed model, twelve characteristic indexes of vehicle operating performance are the input parameters. Eco-level (i.e., score) is the output. This BP network based estimation model has one hidden layer and the number of neuron nodes in this layer is ten. The transfer function from input to hidden layer is "tansig" and that from hidden to output layer is "purelin." The appropriate training function is "traingdm" and the optimal learning rate is 0.03.

4. Model Accuracy Test and Discussion

Depending on our distribution of experimental data for model accuracy test, it was calculated that the average training time of our established model was 0.732 seconds. The model operating efficiency was relatively high. Taking the absolute value of the difference between original and predictive values divided by the original value as model forecasting performance measurement, the average prediction accuracy of this BP network based estimation model was 96.37%. Besides, the most absolute prediction error was less than 5 points and the biggest absolute prediction error was less than 10 points. In addition, five other evaluating indexes about model prediction error were calculated to evaluate the model performance (shown in Table 2). In an overall view, the model evaluation results tested and verified that the BP network based eco-level estimation model was effective.

In addition to testing the model accuracy by itself, we also want to compare the estimation results of our established model with those of traditional linear regression approaches. Taking the same model input parameters as independent factors and treating eco-level as dependent variable, we adopted the method of stepwise regression to construct the linear regression model using the same database stated above. But, unfortunately, we failed to establish this linear regression model because no variables were applied into the linear equation. This comparative result illustrates and verifies that the relationship between driver performance and fuel consumption is not linear or obvious but chaotic and hidden. Driver performance leads to a vehicle power requirement, which in turn leads to an engine power requirement and then to an engine fuel usage [20–22]. Since machine learning has advantage in mining hidden and complicated features, it is effective to develop a machine learning model to estimate the eco-level of driver performance precisely. In one of our previous studies [44], we found that the performance of BP network based model was better than that of the random forest based model, from the aspects of elapsed time and prediction accuracy in estimating the eco-level of driver performance. Thus, we developed a BP network based model to qualify drivers' performance in terms of fuel consumption.

Besides the common feature parameters (e.g., average and standard deviation) used to describe the influence on eco-level in driving process, fluctuation percentage of driver performance was newly adopted as an independent variable to estimate eco-level in the current study. The test results of mode estimation accuracy (96.37%) indicate that these input parameters selected could highly interpret the difference of eco-level corresponding to various driver performance. It verified that the change of driver performance might be a key influence factor decreasing eco-level of vehicle operating. Although these independent variables were of no physical meaning, it was implied that the fluctuation of driver performance would be

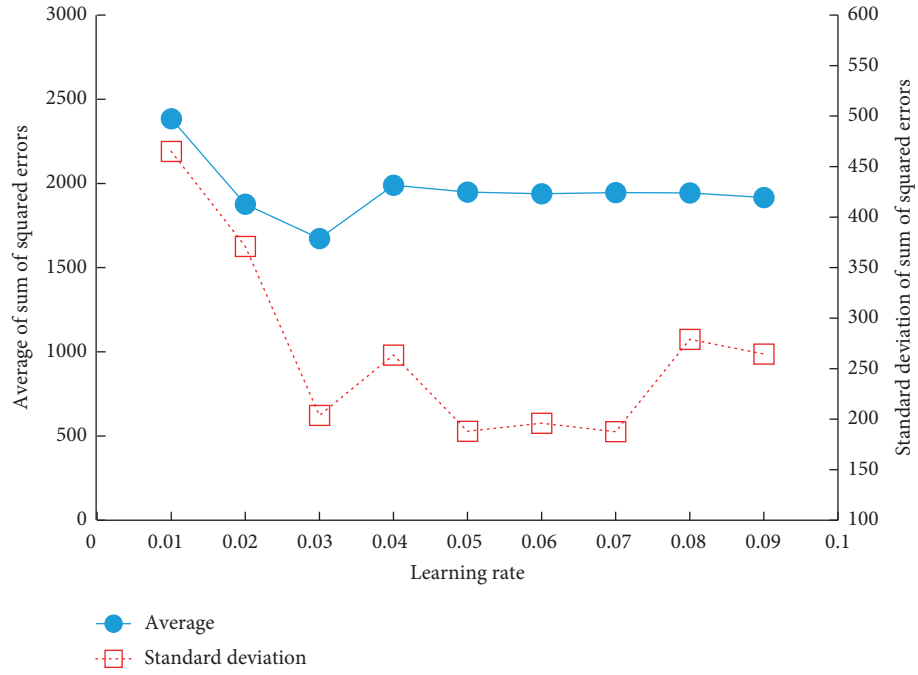

Figure 4: Relationship between sum of squared errors and different learning rate.

Table 2: Prediction error for eco-level of driver performance.

Mean absolute error (MAE)	Mean percentage error (MPE)	Mean absolute percentage error (MAPE) (%)	Mean square error (MSE)	Standard deviation of error (SDE)
2.92	−0.81%	3.11	13.48	3.67

highly related to vehicle fuel consumption. Drivers should maintain their performance as stable as possible to achieve a higher eco-level of vehicle operation.

Apart from directly taking default parameters and structures of existing BP network for model construction, the main work of this study focused on finding the suitable number of neuronal nodes, functions, and learning rate through securing the estimation error under an acceptable level. Undoubtedly, the rule of thumb of establishing BP network model was also referred to in our study. Thus, the reasonability, practicability, and robustness would be highly enhanced. The prediction accuracy of the BP network model proposed in our current study is much higher than that of traditional methods based on linear regression or other statistic models [45, 46]. More importantly, the data demand is significantly less than that of most existing microscopic vehicle fuel consumption models that often require data on trajectories along with engine power, vehicle features, engine types and driving conditions [12, 13], driving models, the acceleration, deceleration, cruise, and idling parameters [15, 16]. Our developed model provided a practical way for government and companies to evaluate and reward drivers with eco-driving behavior. Meanwhile, the drivers could also know well their driving ability from the perspective of eco-driving behavior.

For the inputs of our developed model, the data used could all be collected in various road, traffic, and environmental conditions with no interventions to drivers, based on on-board OBD and GPS devices. As vehicle operation information would be more easily detected with rapid development of detection and communication technology, the current study results would be a foundation for vehicle fuel using level estimation and further optimization in CAV environment. Therefore, the eco-level estimation model proposed in this study is more valuable for further applications.

Overall, this research proposed a new and practical method to estimate the eco-level of driver performance based on OBD + GPS data in naturalistic driving conditions. Combined with our previously developed model to estimate vehicle fuel consumption by driver manipulating data (e.g., controlling the steering wheel, accelerator pedal, and decelerator pedal) in driving simulator [44], it tested and verified that BP network based model did have an advantage and applicability in exploring the relationship between vehicle fuel consumption and driver behaviors, thus contributing to qualifying the eco-level of drivers' performance from the perspective of fuel consumption. Moreover, the difference between our previously developed model and the current proposed model was obvious, namely, in terms of the input parameters, node number of hidden layer, training function, and learning rate.

5. Summary and Conclusions

In order to find a practical method to accurately estimate the eco-level of driver performance in naturalistic driving conditions, a back-propagation network based estimation

model was developed in our current study. Depending on the database of taxicab's instantaneous running data (e.g., second-by-second fuel consumption, engine revolution and torque, speed, and acceleration) collected by on-board diagnostics and Global Position System devices, the optimal model structure and parameters were built and obtained from repeated tests. The model accuracy and performance were tested to be acceptable.

In addition to the common feature parameters (e.g., average and standard deviation) used to describe the influence factors in eco-level of driving performance during driving, fluctuation percentage of driver performance during driving segments was newly adopted as the independent variable to estimate the eco-level in the current study. A total of twelve feature indexes of driver performance were set as model input parameters. The eco-level was used as model output. Based on 700 data samples in basic segments of urban expressways as training set and 100 data samples as validation test, model structure and parameters were obtained through controlling estimation error and training speed through repeated simulation tests. In particular, our established model has three layers and the number of neuron nodes in the hidden layer is ten. The transfer function from input to hidden layer is "tansig" and that from hidden to output layer is "purelin". The suitable training function is "traingdm" and the optimal learning rate is 0.03. Validation test shows that the average estimation of our developed model is 96.37%.

Besides, the result of the comparison between our constructed model and traditional linear regression analysis demonstrated that the relationship between driver performance and fuel consumption should be chaotic and hidden but not linear or obvious. Different from most previous studies based on experiments under restrictive condition and in smaller sample size, this study gives some new insights about data mining of natural driving characteristics in oncoming traffic big data era. The study results provide a practical approach to accurately qualify drivers' performance in terms of fuel consumption in naturalistic conditions and thus give basis for rewarding best drivers within eco-driving programs. This study also provides evaluation supports for more targeted and effective driver behavior training towards vehicle energy conservation.

Although our established BP network based model has relatively high accuracy in estimating the eco-level of driver performance, only taxicabs' running data in basic segments of expressway were processed in our eco-level estimation model. The data used was processed by a single vehicle technology, with the vehicles restricted to operating on a single mode of operation. Different roadway and vehicle types should also be considered in the future researches to enhance the robustness of our developed eco-level estimation model. In addition, other machine learning models with different structures or algorithms (e.g., deep learning model or convolutional neural network) should be further employed to establish the eco-level estimation method of driver performance and thus get the most optimal estimation methods with more effective calculation performance.

Data Availability

The data used in the current study was from the authors' established driving behavior platform based on Internet + technology in Beijing University of Technology. The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by Beijing Municipal Education Commission Foundation (no. KM201910005002).

References

[1] Beijing Municipal Statistics Bureau, *Beijing Statistical Yearbook*, China Statistics Press, Beijing, China, 2017.

[2] A. Sanguinetti, E. Queen, C. Yee, and K. Akanesuvan, "Average impact and important features of onboard eco-driving feedback: A meta-analysis," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 70, pp. 1–14, 2020.

[3] C. Andrieu and G. S. Pierre, "Comparing effects of eco-driving training and simple advices on driving behavior," *Procedia - Social and Behavioral Sciences*, vol. 54, pp. 211–220, 2012.

[4] Y. Wu, X. Zhao, J. Rong, Y. Liu, and L. Xu, "The potential of eco-driving in reducing fuel consumption and emissions: A driving simulator study," *Journal of Beijing University of Technology*, vol. 41, no. 8, pp. 1212–1218, 2015.

[5] Y. Wu, X. Zhao, J. Rong, and Y. Zhang, "The effectiveness of eco-driving training for male professional and non-professional drivers," *Transportation Research Part D: Transport and Environment*, vol. 59, pp. 121–133, 2018.

[6] M. Sivak and B. Schoettle, "Eco-driving: Strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy," *Transport Policy*, vol. 22, pp. 96–99, 2012.

[7] L. Ntziachristos and Z. Samaras, *COPERT III Computer Programme to Calculate Emissions from Road Transport, Methodology and Emission Factors*, European Energy Agency (EEA), Copenhagen, Denmark, 2000.

[8] California Air Resources Board (CARB), *EMFAC2014 Volume III – Technical Documentation*, California Air Resource Board, Sacramento, CA, USA, 2015.

[9] Environmental Protection Agency (EPA), "Transportation and climate division, office of transportation and air quality," Technical Report EPA-420-B-15-007, U.S. Environmental Protection Agency, Washington, DC, USA, 2015.

[10] G. Song, "A review of road traffic fuel consumption and emissions microcosmic models," *Energy Conservation & Environmental Protection in Transportation*, vol. 1, pp. 173–179, 2010.

[11] H. Zhang, Y. Pang, S. Cheng, and Z. Cheng, "Establishing theoretical fuel consumption models for four typical vehicles," *East China Highway*, vol. 6, pp. 62–66, 2002.

[12] T. Markel, A. Brooker, T. Hendricks et al., "ADVISOR: A systems analysis tool for advanced vehicle modeling," *Journal of Power Sources*, vol. 110, no. 2, pp. 255–266, 2002.

[13] D. W. Gao, C. Mi, and A. Emadi, "Modeling and simulation of electric and hybrid vehicles," *Proceedings of the IEEE*, vol. 95, no. 4, pp. 729–745, 2007.

[14] K. T. Chau, Y. S. Wong, and C. C. Chan, "EVSIM - A PC-based simulation tool for an electric vehicle technology course," *The International Journal of Electrical Engineering & Education*, vol. 37, no. 2, pp. 167–179, 2000.

[15] K. Ahn, H. Rakha, A. Trani, and M. Van Aerde, "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels," *Journal of Transportation Engineering*, vol. 128, no. 2, pp. 182–190, 2002.

[16] A. Cappiello, I. Chabini, E. K Nam, A. Lue, and M. Abou Zeid, "A statistical model of vehicle emissions and fuel consumption," in *Proceedings of the IEEE 5th International Conference*, pp. 801–809, Amsterdam, Netherlands, February 2002.

[17] J. Park, R. Noland, and J. Polak, "Microscopic model of air pollutant concentrations: Comparison of simulated results with measured and macroscopic estimates," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1750, pp. 64–73, 2001.

[18] H. Zhai, H. C. Frey, and N. M. Rouphail, "A vehicle-specific power approach to speed- and facility-specific emissions estimates for diesel transit buses," *Environmental Science & Technology*, vol. 42, no. 21, pp. 7985–7991, 2008.

[19] X. Zhao, Y. Wu, J. Rong, and Y. Zhang, "Development of a driving simulator based eco-driving support system," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 631–641, 2015.

[20] C. Bach and P. Soltic, "CO_2 reduction and cost efficiency potential of natural gas hybrid passenger cars," *SAE International Journal of Engines*, vol. 4, no. 2, pp. 2395–2404, 2011.

[21] G. Pannone, B. Betz, M. Reale, and J. Thomas, "Decomposing fuel economy and greenhouse gas regulatory standards in the energy conversion efficiency and tractive energy domain," *SAE International Journal of Fuels and Lubricants*, vol. 10, no. 1, pp. 202–216, 2017.

[22] P. Phlips, "Analytic engine and transmission models for vehicle fuel consumption estimation," *SAE International Journal of Fuels and Lubricants*, vol. 8, no. 2, pp. 423–440, 2015.

[23] C. Miyajima, Y. Nishiwaki, K. Ozawa et al., "Driver modeling based on driving behavior and its evaluation in driver identification," *Proceedings of the IEEE*, vol. 95, no. 2, pp. 427–437, 2007.

[24] J. Y. Zhang and G. Y. Pan, "Comparison and application of multiple regression and BP neural network prediction model," *Journal of Kunming University of Science and Technology (Natural Science Edition)*, vol. 38, no. 6, pp. 61–67, 2013.

[25] X. Y. Wang and X. Y. Yang, "Study on decision mechanism of driving behavior based on decision tree," *Journal of System Simulation*, vol. 20, no. 2, pp. 405–414, 2008.

[26] S. S. Seyhan, F. N. Alpaslan, and M. Yavaş, "Simple and complex behavior learning using behavior hidden Markov model and CobART," *Neurocomputing*, vol. 103, pp. 121–131, 2013.

[27] P. Velarde-Alvarado, A. F. Martínez-Herrera, and A. Iriarte-Solis, "Using entropy spaces and mixtures of Gaussian distributions to characterize traffic anomalies," *Procedia Technology*, vol. 3, pp. 97–108, 2012.

[28] G. Li and W.-Y. Chung, "Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier," *Sensors*, vol. 13, no. 12, pp. 16494–16511, 2013.

[29] X. H. Zhao, Y. Yao, Y. P. Wu, C. Chen, and J. Rong, "Prediction model of driving energy consumption based on PCA and BP network," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 5, pp. 185–191, 2016.

[30] Z. T. Duan, J. Kang, L. Tang, L. Fan, Y. Liu, and J. T. Dai, "Traffic information service cooperation architecture based on vehicular network big data," *Journal of Chang'an University (Natural Science Edition)*, vol. 34, no. 2, pp. 108–114, 2014.

[31] S. Mubeen, J. Mäki-Turja, and M. Sjödin, "Integrating mixed transmission and practical limitations with the worst-case response-time analysis for controller area network," *Journal of Systems and Software*, vol. 99, pp. 66–84, 2015.

[32] Y. Yu, L. J. Lei, Z. T. Jing, Z. H. Wang, and W. L. Wang, "Brief probe into the status and trend of compass navigation's application in domestic intelligent transportation systems," *Journal of Engineering Studies*, vol. 6, no. 1, pp. 86–91, 2014.

[33] R. Ahas, A. Aasa, S. Silm, and M. Tiru, "Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 1, pp. 45–54, 2010.

[34] W. Wang, W. Zhang, H. Guo, H. Bubb, and K. Ikeuchi, "A safety-based approaching behavioural model with various driving characteristics," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1202–1214, 2011.

[35] J. D. Yang, *Research on the Test Method of Fuel Consumption Based on BP Neural Network*, Jilin University, Jilin, China, 2010.

[36] Y. Du, J. Wu, S. Yang, and L. Zhou, "Predicting vehicle fuel consumption patterns using floating vehicle data," *Journal of Environmental Sciences*, vol. 59, pp. 24–29, 2017.

[37] Y. Wu, X. Zhao, C. Chen, J. Rong, and Y. Zhang, "Development and application of an ecodriving support platform based on Internet+: Case study in Beijing taxicabs," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2645, no. 1, pp. 57–66, 2017.

[38] Y. Yao, X. Zhao, Y. Zhang, C. Chen, and J. Rong, "Modeling of individual vehicle safety and fuel consumption under comprehensive external conditions," *Transportation Research Part D: Transport and Environment*, vol. 79, Article ID 102224, 2020.

[39] S. H. Dong and Z. S. Shi, "Study on road traffic accident forecast based on BP neural network," *China Safety Science Journal*, vol. 20, no. 9, pp. 15–20, 2010.

[40] D. Lois, Y. Wang, A. Boggio-Marzet, and A. Monzon, "Multivariate analysis of fuel consumption related to eco-driving: Interaction of driving patterns and external factors," *Transportation Research Part D: Transport and Environment*, vol. 72, pp. 232–242, 2019.

[41] J. Li, X. Cheng, Y. Qin, Y. Zhang, and Z. Xing, "Reliability prediction of urban rail transit vehicle based on BP neural network," *Journal of Central South University (Science and Technology)*, vol. 44, no. 1, pp. 42–46, 2013.

[42] B. Liu and H. X. Guo, *MATLAB Neural Network Super Learning Manual*, Posts & Telecom Press, Beijing, China, 2014.

[43] X. M. Zhang, Y. S. Ge, Y. Zhang, X. K. Han, and L. P. Gao, "Research on fuel consumption measurement using carbon balance method," *Vehicle Engine*, vol. 3, pp. 56–58, 2005.

[44] Y. Wu, X. Zhao, Y. Yao, and J. Rong, "An ecolevel estimation method of individual driver performance based on driving simulator experiment," *Journal of Advanced Transportation*, vol. 2018, pp. 1–12, Article ID 9058674, 2018.

[45] C. Chen, "*Research on evaluation method of drivers eco-driving behavior on urban road*," M.S. Thesis, Beijing University of Technology, Beijing, China, 2016.

[46] C. Ma, "*Study on the evaluation method of driving behavior habits based on OBD technology*," M.S. Thesis, Nanjing University, Nanjing, China, 2016.