

Research Article

An Ecolevel Estimation Method of Individual Driver Performance Based on Driving Simulator Experiment

Yiping Wu , Xiaohua Zhao , Ying Yao, and Jian Rong

Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, 100 Ping Le Yuan, Beijing 100124, China

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

Received 4 January 2018; Revised 2 June 2018; Accepted 28 June 2018; Published 12 July 2018

Academic Editor: Richard S. Tay

Copyright © 2018 Yiping Wu 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.

Accurately acquiring the ecolevel of individual driver performance is the precondition for more targeted ecodriving behavior optimization. Because of obvious advantage in mining hidden relationship, machine learning was adopted to explore the complicated relationship between driver performance and vehicle fuel consumption and thus to predict the ecolevel of individual driver performance in this study. Based on driving simulator tests, data of driver performance and vehicle fuel consumption were collected. The ecolevel was indicated as the ecoscore corresponding to vehicle fuel consumption. The model input was designed as 10 feature indexes of driver performance (e.g., percentage number, mean value, standard deviation, and power of applying acceleration pedal). The output was treated as ecoscore. Taking a number of one hundred of data segments in vehicle starting process as training sample, the optimal structure, functions, and learning rate of a backpropagation neural network model with three layers were obtained, after repeated model simulation experiments. The validation test of 16 sample data items showed that the mean prediction accuracy of our developed model was 92.89%. In addition, comparative analysis displayed that the performance of backpropagation neural network based model was better than linear regression based model and random forest based model, from the aspects of elapsed time and prediction accuracy in estimating the ecolevel of driver performance. The study results provide an effective method to grasp the ecolevel of driver performance and further contribute to driving behavior optimization towards vehicle fuel consumption and emissions reduction.

1. Introduction

Previous researches have stated that driving behavior was a key factor influencing traffic accidents, congestion, and environment [1]. It is also indicated that driving behavior optimization has a big potential to improve traffic safety and efficiency and reduce vehicle fuel consumption and emissions [2, 3]. Specifically, ecodriving, the concept of changing driving behavior to improve fuel efficiency and reduce greenhouse (GHG) emissions of existing vehicles, has been widely reported to be effective in reducing fuel consumption and the reduction percent could even be 20 to 50% for excellent performers [4, 5]. Thus, comparing to vehicle manufacturing technology, new road materials, and other energy conservation and emission reduction measurements for motor vehicles, it seems to be more cost-effective and of higher efficiency to apply ecodriving behavior optimization for existing vehicle drivers.

As a primary condition for ecodriving behavior optimization, it is first necessary to know the ecolevel of individual driver's driving behavior exactly. Generally, driving behavior represents both driver performance and vehicle operating status. In the process of driving, vehicle operating status was mainly influenced by drivers' manipulation behavior of vehicle control units. Therefore, driver performance (e.g., pressing acceleration pedal, turning steering wheel, and shifting gear) seems to be more direct to represent drivers' ecodriving level. In the past years, a few studies tried to find methods about how to estimate individual driver performance's ecolevel reasonably. Most of the relative researches took driving experience or simple statistical analysis of hard accelerating or breaking times as the main reference for estimation criterion [6]. In addition, few quantitative evaluation approaches were based on the difference between the practical value and the desirable threshold of some specific driving performance parameters (e.g., average acceleration, deceleration,

and speed) in given condition [7, 8]. But, almost all of these quantitative estimation methods stated here analyzed these feature parameters separately. As driver performance is random, continuous, and instantaneous during vehicle operation, the ecolevel evaluation based on single parameter (e.g., only considering acceleration or deceleration pedal) is not practical. Thus, according to these limited evaluation methods above, on one hand, the estimation accuracy was not satisfactory; on the other hand, the estimation results had no small gap for supplying targeted ecodriving behavior optimization.

In fact, the core of ecolevel estimation for driver performance is to precisely explore the relationship between drivers' manipulating behavior and vehicle fuel consumption and emissions during driving. In the process of driving, driver performance is a dynamic and comprehensive result from driver perception and decision-making about road condition, traffic control, environment stimulation, and their interactions [9]. Therefore, it appears to be much challenging to estimate the ecolevel of driver performance exactly because of its multidimension, complication, randomness, and uncertainty. Furthermore, the relationship between driver performance and vehicle fuel consumption or emissions might be chaotic and hidden but not totally linear or obvious. So, the statistical analysis based traditional estimation and prediction models are not sufficient for accurate ecolevel estimation of driver performance because of limitation in expressing complicated relationship [10].

Fortunately, machine learning with excellent data processing ability has been widely applied in deep data mining in traffic areas [9]. Moreover, as with obvious advantage of mining hidden features, machine learning has been treated as an effective approach for driver behavior modeling, evaluation, and prediction. At present, shallow machine learning models including decision tree [11], Hidden Markov Model [12], Network of Neuron [13], and Support Vector Machine [14, 15] have been widely used for individual driving habit modeling, traffic flow prediction, and unsafe driving behavior (e.g., driving drowsiness and drunk driving) identification. Moreover, the prediction and identification accuracy were acceptable.

Comparing these shallow machine learning models commonly used, it is not difficult to find that artificial neural network could realize the highly nonlinear reflection from input to output [16]. Hence, aiming at reliability prediction and evaluation, the method based on artificial neural network would be better. Among these artificial neural networks, backpropagation (BP) neural network is the most commonly used, especially in practical application [17, 18]. Moreover, BP network has apparent advantages in self-learning, self-organizing, good fault tolerance, and excellent nonlinear approximation ability [19]. So, system with complicated structure and perfect performance could be established by a large number of neurons with simple performance, and further the complex tasks could be resolved [20].

Summarizing above, to grasp the ecolevel of driver performance more accurately, BP neural network was introduced in this study. As with obvious advantage in driver performance perception and convenient controlling for test

condition, interference factors, and driver safety, driving simulator technology was adopted in the current study for driver performance data collection in the process of driving. In general, vehicle fuel consumption reduction indicates individual driver's application of ecodriving behavior for a given vehicle in large degree, so vehicle fuel consumption is used to represent the ecolevel of driver performance. Based on our driving simulator experimental data with repeated tests, this paper aims to find the optimal structure, suitable network parameters and functions of the BP neural network based model for individual driver's ecolevel estimation. The study results established the relationship between driver performance and vehicle fuel consumption, which laid a foundation for more targeted ecodriving optimization and thus further contributed to reducing vehicle fuel consumption and emissions. The study results provide an effective method to grasp the ecolevel of driver performance and further contribute to driving behavior optimization towards vehicle fuel consumption and emissions reduction.

2. Data Collection

Driving simulator platform in Beijing University of Technology was used to collect driver performance data and its corresponding fuel consumption (see Figure 1). This driving simulator is a fix-based type, which consists of a renovated real car, computers, and video and audio equipment. The road scenarios are projected onto three big screens providing a total of 130 degrees of the driver's visual field. There were two side mirrors and one rear view mirror to show the traffic behind the simulated vehicle. This simulator records the actions of drivers and vehicles, and the parameters that describe vehicle operations are saved 30 times per second. The data of drivers' actions collected by this simulator mainly include drivers' maneuver for acceleration and deceleration pedal, clutch pedal, steering wheel, and gear. The data of vehicle operations recorded includes the engine revolution as well as speed, acceleration, deceleration, etc. Moreover, the validation of this simulator in driving behavior tests has been conducted in our previous studies, including both subjective evaluation by questionnaires [21] and comparative experimental tests in simulator and field [22].

The output mode of vehicle energy and emission in our driving simulator depended on filed test data in Beijing. The microscopic emissions model was based on vehicle specific power (VSP) distribution. The parameters in the VSP based emission model was calibrated by a large number of taxis operating in realistic driving conditions in Beijing [23]. On this basis, vehicle fuel consumption was calculated by the carbon balance method [24]. After comparative and validation analysis, it has been verified that the vehicle fuel consumption per 100 kilometers outputted by our driving simulator is consistent with the field test data collected by On-Board Diagnostics mounted on vehicles [25].

As most common rules or guidance of ecodriving behavior was based on various driving events (e.g., vehicle starting event, vehicle stopping event, and vehicle cruising event) [2, 3], our current experiment is designed to record subjects' driving performance in the process of vehicle starting event.

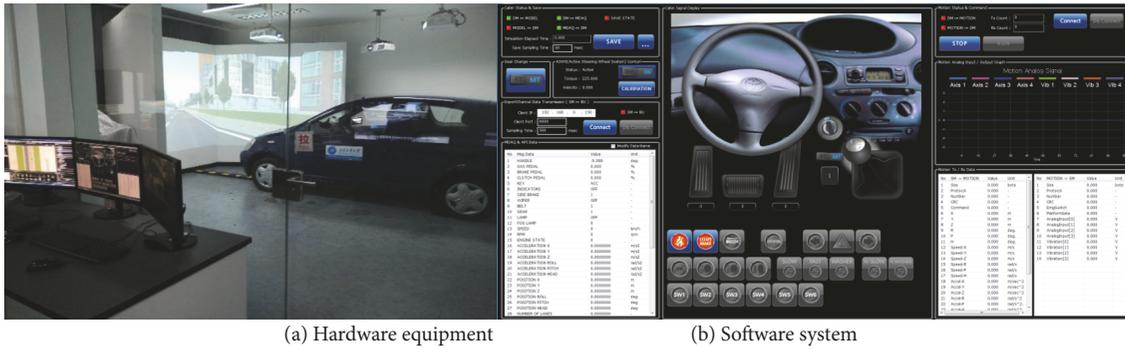


FIGURE 1: Driving simulator in Beijing University of Technology.

Driver performance during vehicle starting process requires more comprehensive maneuvers (manual transmission car) and it is direct reflection of drivers' ecodriving level. The scenario used in this experiment was an urban road with two lanes in each direction. The data collection segments were designed at four signal intersections with digital display. The interval between each intersection and the beginning and ending of this scenario were designed as urban roads with the length of 1,000 meters. Between each starting event, no other events were designed and participants were asked to drive routinely and obey the traffic laws. In order to eliminate the interference of confounding factors on the experimental data, the scenarios used for each vehicle starting event in different locations are uniformed. The road used for data collection was straight and flat. No other vehicles were visible in the participant's driving direction and no vehicles existed in front of the target vehicle.

Before the main driving, every participant took 3-5 minutes for a practice drive to become familiar with the simulator's steering, braking dynamics, and other driving manipulation. For each driving at these four signal intersections, drivers started the vehicle when the traffic signal turns green. Subjects were asked to drive straightly through each intersection. In data processing and analyzing, the starting process was defined as the vehicle speed increased from zero to the location where 100 meters passed by.

A total of thirty drivers (eighteen male and twelve female drivers) participated in this study. All participants were licensed drivers with an average age of 30 (between the ages of 22 and 50 years, with a standard deviation of 9.5 years) and 3 to 20 years of driving experience (average=7.6 years and standard deviation =6.1 years). According to the experimental design, each participant experienced four times of vehicle starting process and thus a total of one hundred and twenty data segments of vehicle starting would be collected.

As the road used for vehicle starting is straight and flat, only driver performance of accelerator pedal, clutch pedal, and gear was used for individual ecolevel estimation. Figure 2 displays two types of driver performance with different fuel consumption. As shown in Figure 2, the higher fuel consumption during vehicle starting process is 23.87L/100km (see Figures 2(a), 2(c), and 2(e)) and the lower is 18.90L/100km (shown in Figures 2(b), 2(d), and 2(f)). It

obviously indicates that the driver performance with high and low fuel consumption is significantly different.

3. Estimation Model Construction

Aiming at ecolevel estimation of driver performance, the input, output, and hidden layer of our BP neural network model should be driver performance parameters, ecolevel estimation results, and the relationship between driver performance and vehicle fuel consumption, respectively. In addition, as any continuous function in a closed interval could be approximated by BP neural network with single hidden layer, any reflection from n-dimension to m-dimension could be achieved through BP neural network with three layers [20]. Thus, this study focused on constructing an ecolevel estimation model of driver performance based on BP neural network with three layers.

According to the design of experimental data collected, a total of 120 data segments would be obtained. As one participant suffered from simulation sickness, data of twenty-nine successful candidates (seventeen male and twelve female drivers) were collected. Namely, the data of 116 segments was used in the following analysis. The age of these successful candidates was 29.5 ± 9.4 years. The driving experience was 7.5 ± 6.3 years. To construct ecolevel estimation model and analyze its applicability, one hundred data segments were randomly selected for model training and the remaining sixteen sample data were used for model accuracy test.

3.1. Input Design. Based on our driving simulator experimental data collected and combining with these influencing factors on vehicle fuel consumption found in our previous studies [6, 7, 25], the driver performance features in the process of vehicle starting was deeply analyzed. Finally, ten indexes were selected as the input parameters for BP neural network model, which mainly consists of drivers' maneuver for vehicle acceleration pedal, clutch pedal, and gear. As the data segment was limited to vehicle starting process and the road condition was straight and flat, drivers' maneuver for vehicle steering and deceleration pedal hardly occurred. Thus, the index of applying deceleration pedal was not included in the current research, although drivers' action behavior of deceleration pedal is usually an important

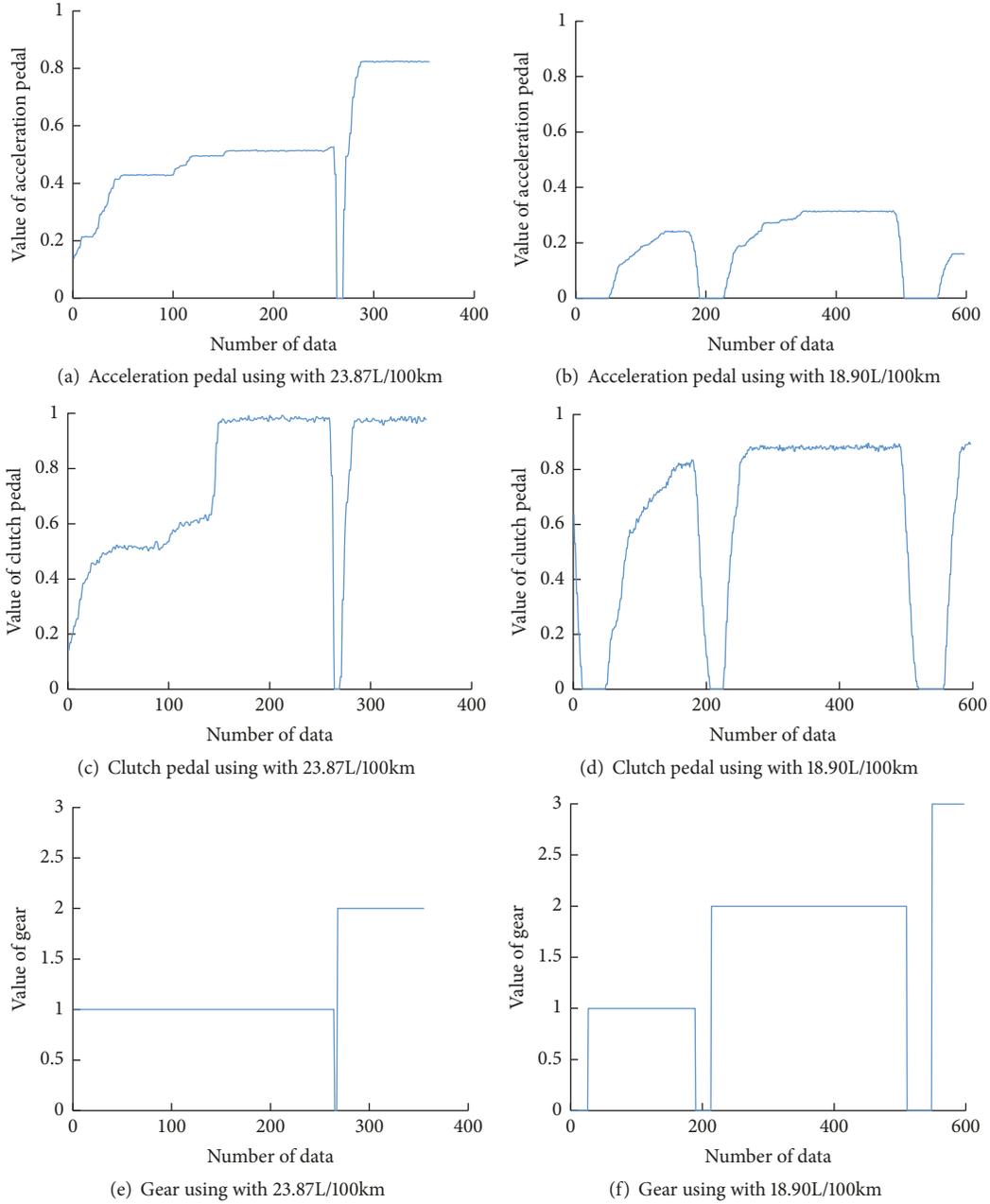


FIGURE 2: Driver performance in vehicle starting process with different fuel consumption.

representation of driver performance during other driving events. The ten indexes used for model input were described below.

(1) Percentage of number when applying acceleration pedal (A_number).

A_number is used to reflect drivers' operating frequency of acceleration pedal. A_number with higher value means more frequent maneuver of acceleration pedal, which might lead to higher fuel consumption. The calculation method of A_number is described as

$$A_number = \frac{N_A}{N} \quad (1)$$

$$N_A = \sum_{i=1}^N n_i \quad (2)$$

$$n_i = \begin{cases} 1, & a_i \neq 0 \\ 0, & a_i = 0 \end{cases} \quad (3)$$

where

a_i is the value of acceleration pedal applying at time i ;

n_i is the statistic value of acceleration pedal applying at time i . The value of n_i is 1 if the acceleration pedal is pressed at time i , otherwise n_i is zero;

N_A is sum of n_i during the given time or journey;

N is the total number of data recorded during the given time or journey.

(2) Mean value of applying acceleration pedal (A_mean)

A_mean is adopted to represent drivers' average pressing intensity of acceleration pedal, which is defined as (4). Vehicle fuel consumption would be increased when A_mean becomes larger.

$$A_mean = \frac{\sum_{i=1}^N a_i (a_i \neq 0)}{N_A} \quad (4)$$

(3) Standard deviation of acceleration pedal applying value (A_SD)

In addition to A_mean , A_SD indicates the stability of drivers' maneuver on acceleration pedal. A_SD with larger value would have negative impact on reducing vehicle fuel consumption. The calculation method of A_SD is shown in

$$A_SD = \sqrt{\frac{1}{N_A - 1} \sum_{i=1}^{N_A} (a_i - A_mean)^2 (a_i \neq 0)} \quad (5)$$

(4) Percentage of number when acceleration pedal applying value is more than 2/3 of the whole pedal travel (A_hard)

A_hard represents the percentage of drivers' rapid acceleration behavior, which is defined as (6) to (8). Higher A_hard will increase vehicle fuel consumption significantly.

$$A_hard = \frac{N'_A}{N_A} \quad (6)$$

$$N'_A = \sum_{i=1}^N m_i \quad (7)$$

$$m_i = \begin{cases} 1, & a_i \geq \frac{2}{3} \\ 0, & a_i < \frac{2}{3} \end{cases} \quad (8)$$

where m_i is the statistic value of applying acceleration pedal at time i . The value of m_i is 1 if the acceleration pedal applying is equal to or more than 2/3 of its whole pedal travel at time i , otherwise m_i is zero and N'_A is sum of m_i during the given time or journey.

(5) Power of pressing acceleration pedal (A_power)

The power of pressing acceleration pedal was defined by drivers' press intensity and duration, as shown in (9). The power of acceleration applying could be intuitively represented by the area surrounded by the x-axis (time), y-axis (value of acceleration pedal), and press intensity profile in a graph (e.g., Figures 2(a) and 2(b)). With the area increasing, it means that a driver presses on the acceleration pedal with more power.

$$A_power = \int_{i=1}^N a_i di \quad (9)$$

(6) Percentage of number when applying clutch pedal (C_number)

In the same way with A_number , C_number is used to reflect drivers' operating frequency of clutch pedal. The calculation method of C_number is described as

$$C_number = \frac{N_C}{N} \quad (10)$$

$$N_C = \sum_{i=1}^N z_i \quad (11)$$

$$z_i = \begin{cases} 1, & c_i \neq 0 \\ 0, & c_i = 0 \end{cases} \quad (12)$$

where

c_i is the value of clutch pedal pressing at time i ;

z_i is the statistic value of clutch pedal applying at time i . The value of z_i is 1 if the clutch pedal is used at time i , otherwise z_i is zero;

N_C is sum of z_i during the given time or journey.

(7) Power of applying clutch pedal (C_power)

In the same way with A_power , C_power represents drivers' pressing intensity and duration of vehicle clutch. The definition of C_power is described as

$$C_power = \int_{i=1}^N c_i di \quad (13)$$

(8) Percentage of number when gear is equal to 1 (G_number)

Vehicle fuel consumption would be significantly enlarged when drivers adopt a lower gear but higher driving speed. Shifting up the gear timely and accurately is one important rule guided in ecodriving behavior. Thus, G_number is adopted to indicate the percentage of using the first gear during vehicle starting process, which was illustrated in (14) to (15).

$$G_number = \frac{N_G}{N} \quad (14)$$

$$N_G = \sum_{i=1}^N u_i \quad (15)$$

$$u_i = \begin{cases} 1, & g_i = 1 \\ 0, & g_i \neq 1 \end{cases} \quad (16)$$

where

g_i is the value of gear using at time i ;

u_i is the statistic value of gear using at time i . The value of u_i is 1 if the gear locates in the first level at time i , otherwise u_i is zero;

N_G is sum of u_i during the given time or journey.

(9) Mode of gear (G_mode)

G_mode means the gear value with the highest appearance frequency, which would be highly related to vehicle fuel consumption during vehicle starting process. On one hand, it reflects drivers' preference of gear using; on the other hand, it indicates whether drivers shifted up the gear timely in the process of vehicle starting.

TABLE 1: Description Statistic results of ten feature indexes of model input.

	Maximum	Minimum	Average	Median	Standard deviation
<i>A_number</i>	0.99	0.55	0.86	0.87	0.09
<i>A_mean</i>	0.84	0.10	0.36	0.32	0.18
<i>A_SD</i>	0.32	0.05	0.13	0.11	0.07
<i>A_hard</i>	0.85	0.00	0.12	0.00	0.24
<i>A_power</i>	313.87	57.04	142.48	127.20	55.22
<i>C_number</i>	1.00	0.72	0.94	0.95	0.06
<i>C_power</i>	661.25	175.65	366.81	356.35	89.71
<i>G_number</i>	1.00	0.19	0.54	0.52	0.21
<i>G_mode</i>	2.00	1.00	1.35	1.00	0.48
<i>ACG_percent</i>	0.25	0.00	0.05	0.04	0.04

(10) Percentage of number when pressing acceleration pedal with no work (*ACG_percent*)

ACG_percent means that driver has pressed the acceleration pedal but no work was produced or exported. Higher *ACG_percent* would waste vehicle energy directly. As shown in (17) to (19), *ACG_percent* is defined as two aspects: (1) drivers pressed the acceleration pedal but the gear is zero or (2) acceleration pedal is pressed and the pedal travel of clutch is equal to or more than 2/3 at the same time.

$$ACG_percent = \frac{N_{ACG}}{N} \quad (17)$$

$$N_{ACG} = \sum_{i=1}^N o_i \quad (18)$$

$$o_i = \begin{cases} 1, & a_i \neq 0 \& g_i = 0, \text{ or, } a_i \neq 0 \& c_i \geq \frac{2}{3} \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

where o_i is the statistic value of acceleration pedal, gear and clutch pedal at time i . The value of o_i is 1 if the acceleration pedal is pressed and the gear is zero or the pedal travel of clutch is equal to or more than 2/3 at time i , otherwise o_i is zero and N_{ACG} is sum of o_i during the given time or journey.

Summarizing above, the description statistic results of these ten indexes of all one hundred and sixteen data segments were calculated and displayed in Table 1.

3.2. Output Design. The output of BP neural network should be the ecollevel corresponding to various driver performance. The vehicle fuel consumption was used to represent the ecollevel. For perceptual intuition, the ecollevel should be designed as ecorank (e.g., excellent, good, fair, and poor) or ecoscore (e.g., 73.2 out of a perfect 100 points). In this study, ecoscore was selected to reflect the ecollevel of driver performance. Thus, the core issue of ecollevel estimation for individual driver performance was to predict drivers' ecoscore based on his or her driver performance.

In this study, we assumed that the ecoscore was from 40 (i.e., the highest fuel consumption) to 100 points (the lowest fuel consumption). Based on the data in the process

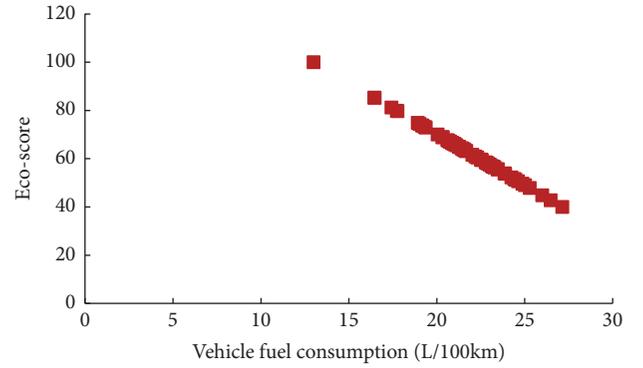


FIGURE 3: Relationship between fuel consumption and ecoscore during vehicle starting process.

of vehicle starting, the method of transforming vehicle fuel consumption to ecoscore is described as

$$Eco - score_i = 40 + \left(1 - \frac{FC_i - FC_{min}}{FC_{max} - FC_{min}}\right) * 60 \quad (20)$$

where

$Eco - score_i$ is the ecoscore in vehicle starting process for driver i ;

FC_i is the mean fuel consumption in vehicle starting process for driver i , L/100km;

FC_{max} is the maximum value of all FC_i ;

FC_{min} is the minimum value of all FC_i .

According to (20) and the experimental data collected, the relationship between fuel consumption per 100 kilometers and ecoscore in the process of vehicle starting could be obtained, as shown in Figure 3.

3.3. Node Number of Hidden Layer Selection. The node number of hidden layer selection is a key issue for BP neural network model construction. However, there is still a lack of certain methods about how to determine the suitable number of nodes in hidden layer for different tasks. According to our literature review, we found that BP neural network model has been widely applied in traffic areas and there were many empirical equations to select the suitable number of hidden layer size. Generally, the range of possible neuron numbers

TABLE 2: Prediction error and training speed corresponding to different training functions.

Training function	Prediction error	Training speed (Second)
trainr ()	0.089	8.133
trainbfg ()	0.086	0.738
trainingd ()	0.082	0.545
trainingdm ()	0.093	0.544
trainlm ()	0.072	0.286
traincgp ()	0.081	0.963
traincgb ()	0.074	1.070

would be obtained based on these empirical equations. To find the best number of hidden layer size, experimental test was conducted and the smaller prediction error was treated as control objective [19, 20, 26, 27]. So, previous study experience and experimental test were jointly used to find the optimal node number of hidden layer for the BP neural network model towards ecolevel estimation of driver performance.

Firstly, based on the number of input and output indexes, the experiential equations (21) to (23) were used to calculate the range of node number [19, 20, 26, 27].

$$n_{node} = \sqrt{n_{input} + n_{output}} + \alpha \quad (21)$$

$$n_{node} = \sqrt{n_{input} \times n_{output}} \quad (22)$$

$$n_{node} = \log_2 n_{input} \quad (23)$$

where

- n_{node} is the number of nodes in hidden layer;
- n_{input} is the number of input indexes;
- n_{output} is the number of output indexes;
- α is constant, and the value is from 0 to 10.

As stated above, the number of input parameters is ten and the output number is one. Hence, the node number should be set as from 4 to 14 according to (21) to (23). Based on this, the node number range was designed as an arithmetic progression with a tolerance of 2 for optimal value test. To find the most suitable node number, the prediction error was treated as control objective. Namely, we should find the optimal value of node number from 4, 6, 8, 10, 12, and 14 under the restriction of the smallest prediction error.

In order to get a relatively stable prediction error corresponding to different node numbers, each BP neural network model with a given node number (i.e., 4, 6, 8, 10, 12, and 14) ran 10 times. The relationship between the average prediction error and the number of nodes in hidden layer was displayed in Figure 4. It revealed that, at the beginning, the prediction error decreased slightly with the node number increasing, while the prediction error increased when the node number further enlarged. Obviously, the mean prediction error is the smallest when the node number is 8. Therefore, the optimal number of the nodes in hidden layer of our established BP neural network should be 8.

3.4. Function Selection. In the same way with other typical BP neural network models [26, 28], the transfer function

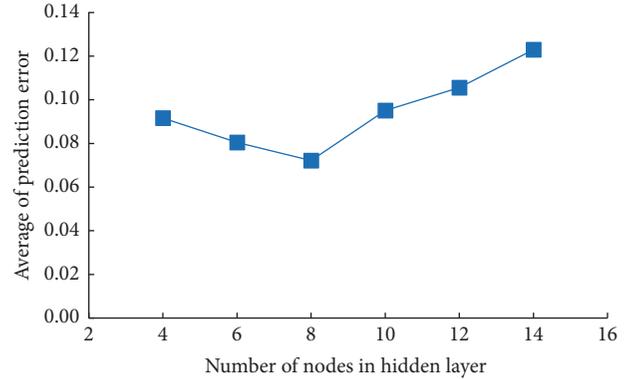


FIGURE 4: Relationship between neuron node number and average prediction error.

from input to output layer was set as S-type tangent function “tansig” in this study. Similarly, the linear function “purelin” was set as the transfer function from hidden to output layer. Considering that different training functions have apparent effects on training time and might also influence prediction accuracy, experimental test was applied to find the most suitable training function from these commonly used. A total of 7 types of training functions were treated as candidates. The prediction accuracy and training speed were the evaluation indexes. To get steady prediction results, the BP neural network model with each training function was tested for ten times. The average prediction error and training speed corresponding to different training functions were calculated and illustrated in Table 2.

According to Table 2, the function of “trainlm” was finally selected as the optimal training function in our developed BP neural network model, because of both lower prediction error and more rapid training speed when comparing to other training functions.

3.5. Learning Rate Selection. Learning rate determines the weight variation for each loop iteration in the BP neural network. The learning rate with smaller value could make the prediction error the smallest at the end of iteration but the learning time might increase accordingly, and thus the model convergence rate would be slower. Generally, the range of learning rate was set as from 0.01 to 0.8 [28].

In this study, according to these typical applications of BP neural network model in traffic areas [19, 20, 26, 27], the

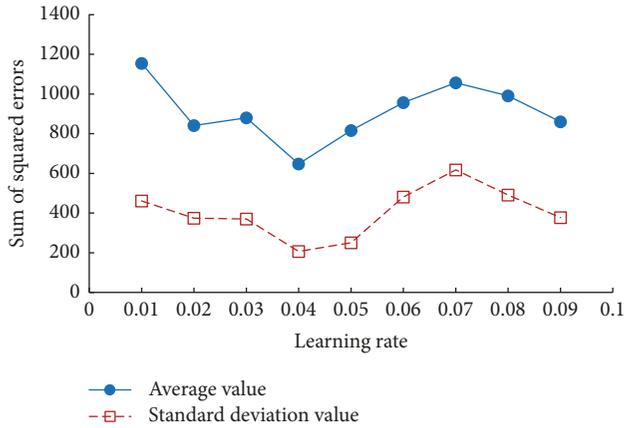


FIGURE 5: Relationship between learning rate and sum of squared errors.

range of learning rate was set as from 0.01 to 0.09. Then, the method of repeated trials was conducted, and the learning rate with the smallest sum of squared errors was treated as the optimal value. The learning rate range was designed as an arithmetic progression with a tolerance of 0.01 for candidate test. Each BP neural network model with a given learning rate ran 10 times. The relationships between learning rate and the average (AVG) and standard deviation (SD) of sum of squared errors were illustrated in Figure 5. It is obvious that both the AVG and SD are the smallest when the learning rate is 0.04. Therefore, the optimal value of learning rate of our established BP network should be 0.04.

3.6. Results of Model Construction. Summarizing the above analysis and discussion, the ecolevel estimation model of driver performance including input and output, number of hidden layers and its corresponding nodes, transfer and training functions, and learning rate could be obtained. Based on these test results, we also attempted to find the most suitable model format by changing the number of neurons in hidden layer, learning rate, and function at the same time. After repeated tests, the optimal structure of our BP neural network model could be described as Figure 6.

In this model, there are ten input indexes of driver performance and the output index is the ecolevel. This model has one hidden layer with eight neuron nodes. The transfer function from input to hidden layer and from hidden to output layer is “tansig” and “purelin,” respectively. The suitable training function is “trainlm” and the optimal learning rate is 0.04. The accuracy rate calculated by the difference between prediction value based on training data and the train label is 96.81%.

4. Model Validation Test and Discussion

According to our experimental data collected and the data distribution for model training and validation test, model validation test results show that our established model has relatively high operating efficiency with the average elapsed time of 0.238 seconds. The prediction results were displayed in Figure 7.

As shown in Figure 7(a), it is clearly indicated that the fitting degree of our developed model is pretty good. The prediction value is around the original value for each vehicle starting process. Taking the absolute value of the difference between the original and predictive values divided by the original value as a forecasting performance measure, this average percentage error was 7.11%. Namely, the mean prediction accuracy of our developed model is 92.89%.

The prediction error for each driving process was shown in Figure 7(b). It revealed that most of the absolute prediction error was less than 5 points and the biggest absolute prediction error was not more than 10 points. Furthermore, to evaluate the performance of our developed model more completely, another five evaluating indexes about model prediction error were calculated and illustrated in Table 3. Summarizing these evaluating indicators, it demonstrated that the ecolevel prediction model based on BP neural network was effective. Thus, it is practicable to precisely estimate the ecolevel of driver performance based on BP neural network models.

Besides its own validation test, we also compared the estimation results based on our developed model (i.e., backpropagation neural network) with the linear regression model and random forest model. The comparing analysis was displayed below.

(1) *Linear Regression Model.* According to our experimental data collected and the data distribution for model training (the data base was the same with BP neural network model training), the linear regression model was obtained by using the stepwise method. Table 4 displays the model form, estimated coefficients, the probability of t-test for regression coefficients, the probability of F-test, and goodness-of-fit for the regression equations. The accuracy rate from the training data is 89.18%.

Using the remaining 16 data samples, the prediction accuracy of this estimation model was tested. Figure 8 displays the measured values and the predictive values calculated by regression model. Taking the absolute value of the difference between the measured and predictive values divided by the measured value as a forecasting performance measure, this average percentage error was 12.81%.

(2) *Random Forest Model.* Using the same data base with the BP neural network model training, the prediction model for ecolevel of driver performance based on random forest was obtained. Figure 9 shows the relationship between number of grown trees and out-of bag (OOB) regression error. The optimal number of grown trees was selected as 60, by analyzing the relationship between prediction accuracy and the number of grown trees and also the relationship between OOB error and the number of grown trees. After repeated tests, the elapsed time of this random forest based model is 0.521seconds. The accuracy rate from the training data is 94.56%.

Similarly, using the remaining 16 data samples, the prediction accuracy of this estimation model was also tested. Figure 10 displays the measured values and the predictive values calculated by random forest model. Taking the absolute value

TABLE 3: Statistical results of prediction error for ecoscore.

Mean absolute error (MAE)	Mean percentage error (MPE)	Mean absolute percentage error (MAPE)	Mean square error (MSE)	Standard deviation of error (SDE)
4.61	-4.47%	7.02%	35.59	5.97

TABLE 4: Linear regression model to predict ecoscore of individual driver performance.

Models	F-test	Goodness-of-fit (R^2)
$Eco_score = 124.716^{***} - 0.188A_power^{***} - 22.233G_number^{***} + 13.406A_hard^{**} - 5.122G_mode^{**} - 19.636A_number^*$	$F_{(5,44)} = 52.572$ $p < 0.001$	0.84

Significance codes: 0.001 ‘***’; 0.01 ‘**’; 0.05 ‘*’.

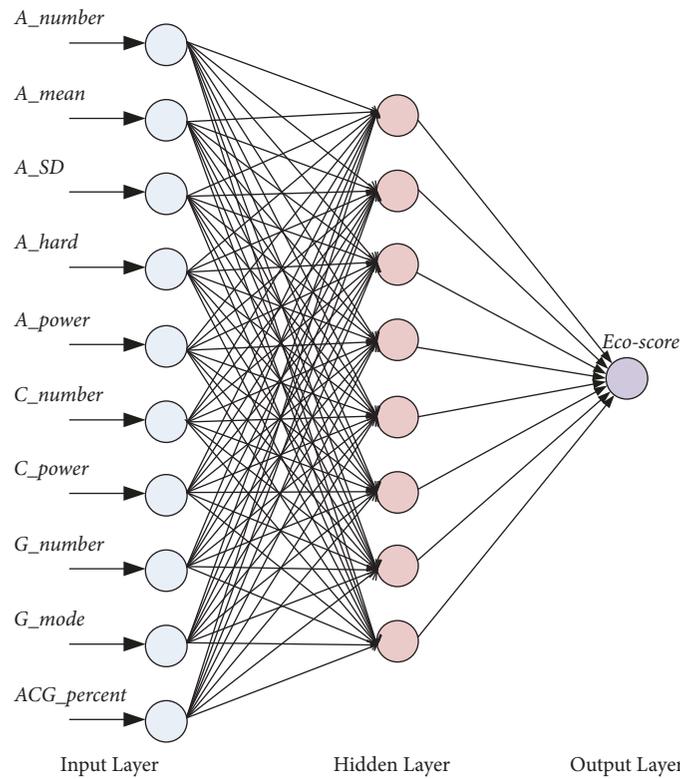


FIGURE 6: BP neural network structure for ecoscore estimation of driver performance.

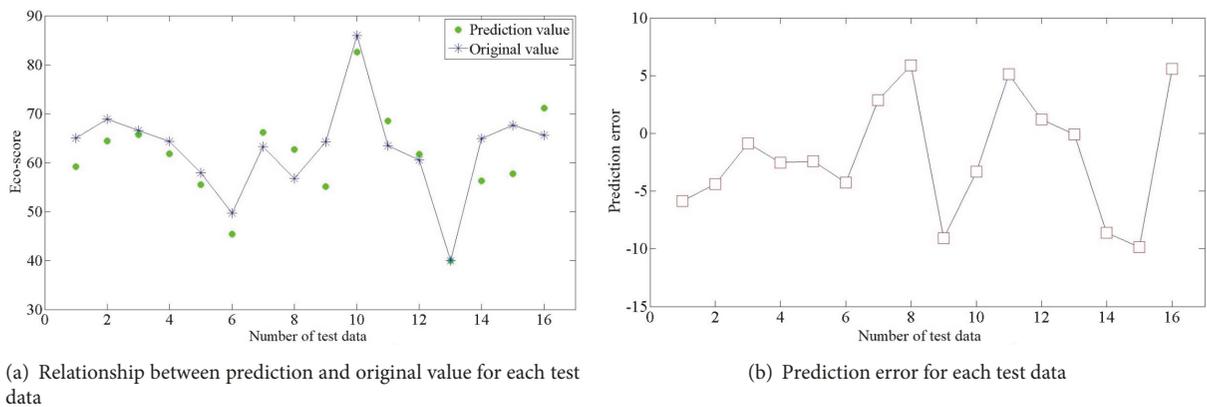


FIGURE 7: Accuracy test results of ecollevel estimation model.

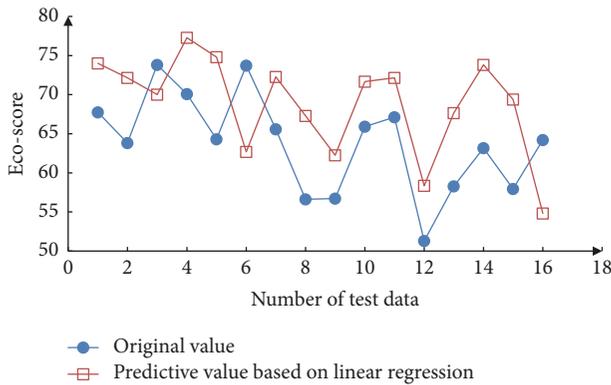


FIGURE 8: Relationship between original values and predictive values based on linear regression model.

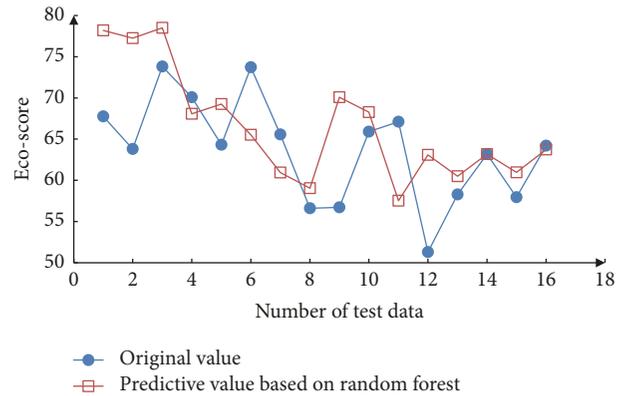


FIGURE 10: Relationship between original and predictive values based on random forest model.

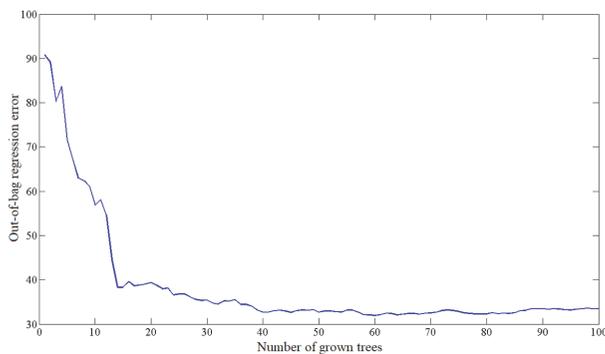


FIGURE 9: Relationship between number of grown trees and out-of-bag regression error.

of the difference between the measured and predictive values divided by the measured value as a forecasting performance measure, this average percentage of prediction error was 9.37%.

Summarizing above, the backpropagation neural network based model is better to estimate the ecolevel of driver performance, after comparing the elapsed time (except for linear regression model) and prediction accuracy of these models developed above. Thus, our study results indicated that BP neural network model in estimating drivers' ecolevel of driver performance was applicable and the model performance was good. The model structure, functions, and parameters used in our developed model were according to the repeated trial tests based on experimental data and thus increased the model reliability markedly.

The current developed model was focused on the ecolevel estimation of driver performance based on driving simulator experimental data. On one hand, the study results support simulator-based ecodriving training, such as dynamic ecodriving evaluation and real-time feedback system development. On the other hand, although this driving simulator has been approved to be effective in studying driving behaviors in many of our previous researches [21, 22], the current study results should be further validated through realistic driving data in the field. In fact, we have planned to establish the ecolevel estimation model based on vehicle operating status

data on-road. With the rapid development of vehicle sensor data collection technology such as On-Board Diagnostics (OBD) + Global Position System (GPS), it is desirable and practicable to estimate the real-time ecolevel of driver behaviors in naturalistic driving conditions.

Besides, although the prediction accuracy of our developed model seems to be acceptable and improved more than 5% when comparing to linear regression model, only shallow learning model was used in this study. The deep learning model with stronger learning ability is expected to be more effective in driving behavior prediction and estimation. Meanwhile, we failed to introduce self-adaptive adjustment algorithm for learning rate selection in model training process. This more improved method for learning rate selection would be integrated into deep BP neural network model in future studies. In addition, our established model should be further compared with other more machine learning approaches (e.g., Support Vector Machine). Furthermore, a challenge that always existed for machine learning is the black-box effect of the model itself. Although the ecolevel of driver performance could be estimated accurately by BP neural network model, the difference of driver performance between drivers with perfect or poor ecoscore was still unknown. Much deeper researches to explore the hidden relationship between input parameters and output results is needed.

As limited to the high resource demands of driving simulator experiment, only one hundred and sixteen data segments were used for model construction and test (i.e., one hundred samples for model training and sixteen for model accuracy test). Meanwhile, only vehicle starting process was used to estimate the ecolevel of driver performance because starting a vehicle is more typical to reflect drivers' ecodriving level. In future studies, more driving events and bigger sample size are needed to further enhance the stability and applicability of our developed model.

5. Conclusions

Aiming at accurately estimating the ecolevel of individual driver performance, a BP neural network model with three layers was developed in our current study. Based on repeated

experimental tests with the driver performance and vehicle fuel consumption data collected in a driving simulator, the optimal structure, functions, and learning rate of BP neural network models in the process of vehicle starting was obtained. The model accuracy was tested to be satisfactory. The study results provide an effective method to estimate the ecolevel of individual driver performance and further contribute to driving behavior optimization towards vehicle fuel consumption and emissions reduction.

Specifically, ten feature indexes of driver performance (e.g., percentage number, mean value, standard deviation, and power of applying acceleration pedal) were selected as model input parameters. The output is designed as the ecoscore to reflect individual's ecodriving level. This model has one hidden layer with eight neuron nodes. The transfer function from input to hidden layer and from hidden to output layer is "tansig" and "purelin," respectively. The suitable training function is "trainlm" and the optimal learning rate is 0.04. Taking 100 segments' data in starting process as training samples and 16 samples as model accuracy test, the mean prediction accuracy of our developed model is 92.89%. The performance of backpropagation neural network based model was better than linear regression based model (i.e., 87.19%) and random forest based model (i.e., 90.63%), from the aspects of elapsed time and prediction accuracy in estimating the ecolevel of driver performance.

Data Availability

The data used in the current study was based on our experimental tests in the driving simulator in Beijing University of Technology. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

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

Acknowledgments

The authors would like to show great appreciation for China Postdoctoral Science Foundation (no. 2017M620562) and the support from the National Natural Science Foundation of China (no. 61672067).

References

- [1] Y. Bai, *Research on driving energy-saving technology*. [Master, thesis], Changan University, Changan, 2011.
- [2] R-Y. Lin, *The international situation for promoting eco-driving*, Industrial Technology Research Institute, Xinzhu, Taiwan, 2011.
- [3] R. Luther and P. Bass, "Eco-driving scoping study," Tech. Rep., AA Research Foundation, Wellington: US, 2011.
- [4] E. Martin, N. Chan, and S. Shaheen, "How public education on ecodriving can reduce both fuel use and greenhouse gas emissions," *Transportation Research Record*, no. 2287, pp. 163–173, 2012.
- [5] 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.
- [6] 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.
- [7] C. Chen, *Research on evaluation method of drivers eco-driving behavior on urban road*. [Master, thesis], Beijing University of Technology, Beijing, China, 2016.
- [8] C. Ma, *Study on the evaluation method of driving behavior habits based on OBD technology*. [Master, thesis], Nanjing University, Nanjing, 2016.
- [9] 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.
- [10] 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.
- [11] 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. 415–420, 2008.
- [12] 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.
- [13] Z-H. Yao, Y-S. Jiang, P. Han, X-L. Luo, and T. Xu, "Traffic Flow Prediction Model Based on Neural Network in Small Time Granularity," *Journal of Transportation Systems Engineering and Information Technology*, vol. 17, no. 1, pp. 67–73, 2017.
- [14] C. Xu, S.-J. Pei, and X.-S. Wang, "Driver drowsiness detection based on non-intrusive metrics considering individual difference," *Zhongguo Gonglu Xuebao/China Journal of Highway and Transport*, vol. 29, no. 10, pp. 118–125, 2016.
- [15] Z-L. Li, J-L. Han, X-H. Zhao, M-H. Zhu, and W-H. Dong, "Comparison of Drunk Driving Recognizing Methods Based on KNN and SVM," *Journal of Transportation Systems Engineering and Information Technology*, vol. 15, no. 5, pp. 246–251, 2015.
- [16] 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.
- [17] K. Aghabayk, M. Sarvi, N. Forouzideh, and W. Young, "Modelling heavy vehicle car-following behaviour in congested traffic conditions," *Journal of Advanced Transportation*, vol. 48, no. 8, pp. 1017–1029, 2014.
- [18] M. Badhrudeen, J. Raj, and L. D. Vanajakshi, "Short-term prediction of traffic parameters - Performance comparison of a data-driven and less-data-required approaches," *Journal of Advanced Transportation*, vol. 50, no. 4, pp. 647–666, 2016.
- [19] 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.
- [20] D. Kang, S-F. Ma, and S-Q. Zhong, "A microcosmic forecasting model for traffic safety based on BP neural network," *Journal of Transport Information*, vol. 29, no. 3, pp. 79–83, 2011.
- [21] H. Ding, X. Zhao, J. Rong, and J. Ma, "Experimental research on the effectiveness of speed reduction markings based on driving simulation: a case study," *Accident Analysis & Prevention*, vol. 60, pp. 211–218, 2013.

- [22] X. Zhao, Y. Wu, J. Rong, and J. Ma, "The effect of chevron alignment signs on driver performance on horizontal curves with different roadway geometries," *Accident Analysis & Prevention*, vol. 75, pp. 226–235, 2015.
- [23] L. Xu, *Comparative analysis and optimization of car-following models for emission estimation*. [Master, thesis], Beijing Jiaotong University, Beijing, 2012.
- [24] 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.
- [25] Y.-P. Wu, X.-H. Zhao, J. Rong, Y. Liu, and L. Xu, "Potential of eco-driving in reducing fuel consumption and emissions based on a driving simulator," *Beijing Gongye Daxue Xuebao/Journal of Beijing University of Technology*, vol. 41, no. 8, pp. 1212–1218, 2015.
- [26] 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.
- [27] 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.
- [28] B. Liu and H-X. Guo, *MATLAB Neural Network Super Learning Manual*, Posts & Telecom Press, Beijing, China, 2014, MATLAB neural network super learning manual. Posts Telecom Press.

