Simulation Model of Driving Behavior Based on Nonparametric Kernel Density Estimate

Yaqi Liu, Zhenxue Liu, Wei Tian, Haibo Wang, Xiaoyuan Wang

***Abstract*— The research on car-following behavior is an important part of microscopic research in traffic flow theory. With the continuous development of Intelligent Transportation System (ITS) and the deepening of big data research, the car-following simulation model based on a series of assumptions and the single control rule has been difficult to meet the precision requirement. In order to overcome the shortcomings of previous car-following models, a real vehicle driving test on a city road was designed and a series of data acquisition equipment, mainly including two experiment vehicles, speedometer and range sensor, were used to collect a variety of driving behavior data. The information mining technology was used to extract the valuable driving behavior information from measured data in this paper. The noise was eliminated by the Nonparametric Kernel Density Estimate (NKDE), and a new simulation model of driving behavior was established based on nonparametric regression. A simulation experiment was designed to verify the validity of the car-following model. The main contents of model validation included acceleration, velocity, relative velocity and relative distance. The measured running state was compared with the simulated result, and the comparison between the measured value and the simulated value can fit well. The simulation results showed that the model can effectively reflect the driving behavior in the car-following process. The research results of this paper can provide new ideas and methods for the study of microscopic traffic flow in the era of big data. The research results are of great significance to improve traffic safety.**

***Index Terms*— Car-following behavior, Kernel Density Estimate, Nonparametric regression, Microscopic traffic flow simulation.**

I. INTRODUCTION

The car-following model which is used to describe human-vehicle behavior is one of the most important dynamic models in traffic system simulation. To a great extent, the quality of the model determines the reliability of the simulation results. The research of car-following model is of great significance to understand the traffic flow characteristics and solve the traffic problems [1].

This study was supported by the State Key Laboratory of Automotive Safety and Energy under Project No.KF16232, Natural Science Foundation of Shandong Province (Grant No.ZR2014FM027, ZR2016EL19), Social Science Planning Project of Shandong Province (Grant No. 14CGLJ27), Project of Shandong Province Higher Educational Science and Technology Program (Grant No.J15LB07), and the National Natural Science Foundation of China (Grant Nos. 61074140, 61573009, 51508315, 51608313).

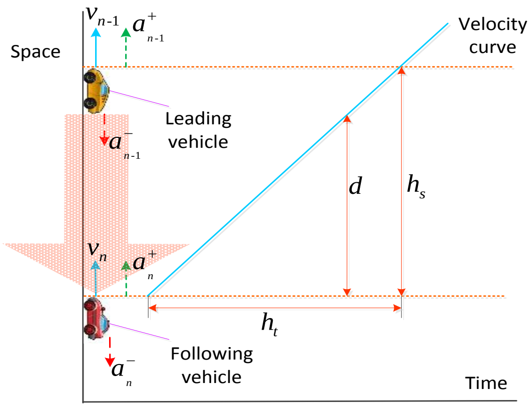

Fig.1 Temporal and spatial variation of car-following process

The Fig. 1 was used to describe the temporal and spatial variation of car-following process, in which h_s was used to represent the space headway of leading and following vehicle, h_t was used to express the time headway, d was the relative distance, v_{n-1} and v_n represented the velocities of leading and following vehicle respectively, a^+_{n-1} and a^+_n were the forward acceleration of leading and following vehicle severally, a^-_{n-1} and a^-_n were the backward acceleration of leading and following vehicle. The car-following model was used to describe the dynamic process of the above parameters. In a steady state, the macroscopic flow parameters such as flow rate, velocity and density can be derived through simple calculation of the above microcosmic parameters. Therefore, the analysis and modeling of car-following behavior not only contribute to the in-depth analysis of the microscopic driving behavior parameters, but also contribute to the profound understanding of macroscopic traffic characteristics [2].

At present, the research of car-following model is relatively mature. The typical models include stimulus-response model, safe distance model, physiological- psychological model, fuzzy inference model and cellular automata model and so on[2]. These models were constructed based on a series of assumptions:

- Drivers strived to achieve optimal driving performance.
- The following process was equivalent to the continuous use of a single control rule.
- Some input variables which the driver cannot perceive but can be calculated by some analytical method were used
- The model attributed all the problems that could not be explained to the limitations of perception and control.

Based on these assumptions, some problems concerned the variability of driver behavior were ignored:

- Drivers usually perform multiple driving tasks at the same time, and only intermittent attention and control strategy are needed.
- The high information perception variable rather than

Newtonian kinematic variables were used by drivers.

- Drivers used satisfaction index rather than the best performance evaluation index to meet the driving requirements [3].

In view of this, the practices of using assumption in the construction of the previous car-following model were abandoned in this paper. The data mining technology was used directly to extract the valuable information from measured data, and a car-following model based on nonparametric regression was established.

II. Driving Behavior Simulation Model

A. Multivariate nonparametric regression model

The information was obtained directly from historical data in multivariate nonparametric regression [4]. Y was set as explained variable, and it is a random variable. X were some important factors that affecting Y. Assumed that $\{Y_i\}$ were independent and identically distributed, and the multivariate nonparametric regression model can be constructed using sample observations $(X_1,Y_1),\cdots(X_n,Y_n)$.

$$Y = m(X) + \mu \tag{1}$$

In which, $m(\bullet)$ was regression function, μ was random error. μ reflects the influence of other observable or unobservable factors, except explanatory variables, the explained variables and the setting error of the model. The conditional estimate regression function [5] was introduced to obtained the minimum estimated mean square error of regression function($\min\left|E(\hat{m}(x)-m(x))^2\right|$). That is:

$$\begin{aligned}\hat{m}(x) &= E[Y \mid X = x] \\ &= \int y f(y \mid x) dy \\ &= \int y f(x,y) dy / \int f(x,y) dy\end{aligned} \tag{2}$$

Kernel Density Estimate was used to calculate the density function in formula 2.

B. Multivariate nonparametric regression model

KDE is a method to study data characteristics from the data sample itself [5]. KDE do not attach any assumption to the data distribution, and the prior knowledge of the data distribution was not needed.

Assumed that the density function $f(x) = f(x_1, x_2, \cdots, x_m)$ of m-dimensional random vector is unknown, and $X_1, X_2, \cdots, X_n$ are the independent identically distributed samples of X. And then, the definition of $f(x)$'s kernel estimation is:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \tag{3}$$

In the formula (3), m is the dimension, n is the sample number, h is window width or smooth parameter, $K(\bullet)$ is multiple kernel function. The $K(\bullet)$ meet:

$$\begin{aligned}&K(u) \geq 0, \int K(u) du = 1, \int K(u) u du = 1, \\ &\int K(u) u u^T du = \mu_2(K) I\end{aligned} \tag{4}$$

In the formula (4), 0 is the zero vector and I is the unit matrix.

(1) Selection of kernel function

The kernel problem of density function kernel estimation is the choice of kernel function and window width. The kernel function plays a smooth role in the kernel estimation of density function. That is to eliminate the random factors of disturbance to make the curve smooth. Gauss kernel which has convenient mathematical properties is a commonly used kernel function. The definition of Gauss kernel is:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{5}$$

For the Gauss function, the closer the distance between x_i and x is, the closer the value of $\frac{x - x_i}{h}$ get to zero and the bigger the density value $\varphi(\frac{x - x_i}{h})$ is. For the range of normal density is the whole real axis, all the data are used to estimate $\hat{f}_h(x)$. And the closer the distance between x_i and x is, the greater the impact of x_i on the estimation is. When the h is very small, only a point which close to the x can play a bigger role. With the increase of h, the role of some distant point will increase.

(2) Selection of window width

Window width is an important parameter to control the accuracy of kernel estimation. The estimation bias will increase when the window width is too large, and the data characteristics cannot be estimated well. While the variance will increase when the window width is too small, and the shape will be irregular. The mean integrated square error (MISE) is used as the evaluation index of density estimation.

$$MISH(h) = AMISH(h) + o\left(\frac{1}{nh} + h^4\right) \tag{6}$$

$$AMISH(h) = \frac{\int K^2(u) du}{nh} + \frac{h^4 \sigma^4 \int [f''(u)]^2 dx}{4} \tag{7}$$

$AMISH(h)$ is called as asymptotic mean integrated square error. The h needs to be set to an intermediate value to minimize the $AMISH(h)$, and it can avoids excessive deviation of $\hat{f}_h(x)$. The optimal window width is:

$$h = \left(\frac{\int K^2(u) du}{n\sigma^4 \int [f''(u)]^2 dx}\right)^{1/5} \tag{8}$$

Setting $R(g) = \int g^2(x) dx$, $R(\varphi'') / \hat{\sigma}^5$ is used to estimate the unknown quantity $R(f'')$. In which, φ is the standard normal density function, K is the Gauss density function, $\hat{\sigma}$ is the sample variance. And then the optimal window width can be obtained:

$$h = \left(\frac{4}{3n}\right)^{1/5} \hat{\sigma} \tag{9}$$

C. Car-following model based on NKDE

Based on the sample observation value $\{a_{n,i}, x_i\}_{i=1}^{t}$ and the basic principle of nonparametric regression, the car-following model was established as:

$$a_{n,i} = m(x_i) + \varepsilon_i \quad i = 1, 2, \cdots t; t \in Z^+ \tag{10}$$

In formula(10), a_n is the acceleration of following vehicle, v_n is the velocity of following vehicle, v_{n-1} is the velocity of leading vehicle, d is the relative distance, $x_i = (v_{n-1,i} - v_{n,i}, v_{n,i}, d_i)$ [6].

The estimation of regression function is:

$$\begin{aligned}\hat{m}(x) &= E[a_n \mid v_{n-1}-v_n, v_n, d] \\ &= \int a_n f(a_n \mid v_{n-1}-v_n, v_n, d) da_n \\ &= \frac{\int a_n f(a_n, v_{n-1}-v_n, v_n, d) da_n}{\int f(a_n, v_{n-1}-v_n, v_n, d) da_n}\end{aligned} \tag{11}$$

III. DATA ACQUISITION

A. Experiment condition

The sample size was 21, including 12 male drivers and 9 female drivers. The age distribution of the drivers was between 20-40 years old, and the driving experience distribution was between 1-7 years. The sections of Zhangzhou Road in Zhangdian district, Zibo city was selected as the experiment route under good weather and road conditions. The whole length of the route including the straight line and curve segment is about 9.3km. Two experiment vehicles were used to simulate the process of car-following. One of the experiment vehicles, as the leading vehicle, ran normally along the experiment route. The other experiment vehicle, as the following one, which was drove by an object driver followed the front vehicle.

B. Experiment equipment

The experiment equipment mainly include two experiment vehicles, speedometer and range sensor.

(1) Speedometer

The SSG299 GPS non-contact multi-function speedometer (see Fig. 2) was used to test the speed of experiment vehicles. The speedometer based on GPS Doppler effect principle is equipped with standard RS232 interface. The test results can be stored and analyzed on computers conveniently.

Fig. 2 SSG299 GPS non-contact multi-function speedometer

(2) Range sensor

The BTM300-905-200 laser range sensor (see Fig. 3) was used to test the relative distance of leading and following vehicles. The range sensor is characterized by quick response, large measuring range, high resolution and small error.

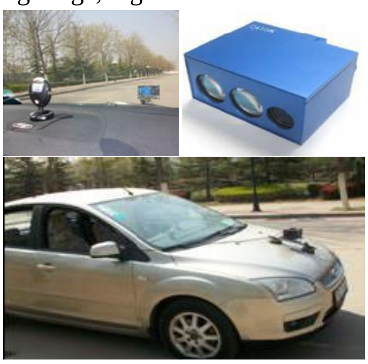

Fig. 3 BTM300-905-200 laser range sensor

C. Experiment data

204600 groups of original data were obtained from driving experiment. The original data were sorted out and 2018 groups of invalid data were eliminated. The 150000 sets of data were used to calibrate the model parameters, and the remaining 52582 sets of data were used to validate the model. Due to space limitation, only a few data of one driver were enumerated, shown as Table 1.

Table 1 Part of experiment data

Time(s)	v_{n-1}	v_n	a_n	d
⋮	⋮	⋮	⋮	⋮
21	12.24	11.85	1.28	35.84
22	12.16	13.13	0.66	35.09
23	12.33	13.79	-2.58	34.27
24	12.17	11.21	1.24	35.10
25	12.19	12.45	0.01	35.01
26	12.86	12.46	-1.69	35.33
27	13.01	10.77	0.87	35.42
28	12.39	11.64	1.26	35.81
29	12.64	12.91	-0.15	35.55
30	12.31	12.76	-1.94	35.33
31	12.44	10.82	2.22	36.29
32	13.05	13.04	0.54	36.30
33	12.63	13.59	0.45	35.73
34	12.14	14.04	-1.88	35.33
35	12.08	12.15	1.34	35.29
⋮	⋮	⋮	⋮	⋮
81	11.64	12.58	-1.26	33.98
82	12.09	11.32	1.51	34.45
83	12.37	12.83	0.13	34.06
84	12.51	12.97	-0.28	33.89
85	12.15	12.69	-0.02	33.85
86	12.54	12.67	0.07	33.80
87	12.75	12.73	0.68	33.81
88	12.52	13.41	-0.92	33.55
89	12.82	12.49	-1.39	33.69
90	13.10	11.10	0.69	34.73
91	13.24	11.79	0.76	35.08
92	13.11	12.55	-1.72	35.59
93	12.41	10.83	0.62	35.98
94	12.26	11.44	1.58	36.48
95	12.47	13.02	0.61	35.95
⋮	⋮	⋮	⋮	⋮
184	12.26	12.38	1.06	32.88
185	12.29	13.43	-0.55	32.40
186	9.39	12.89	0.48	29.59
187	8.74	13.36	-0.69	29.00
188	6.31	12.68	-0.05	25.66
189	4.22	12.63	0.14	24.34
190	2.16	12.76	-1.70	20.62
191	1.37	11.06	-1.78	12.87
192	0.63	9.28	-1.93	11.87
193	0.00	7.35	-1.30	7.90
194	0.00	6.05	-1.44	2.88
195	0.00	4.61	-2.47	2.10
196	0.00	2.14	-1.28	2.08
197	0.00	0.86	-0.86	1.56
198	0.00	0.00	0.00	1.56

IV. MODEL VALIDATION

The measured data stored in hard disk was read in the simulation process, and these data was seen the running state of following car in each scan time. The measured running state was compared with the simulated result, and the comparison between the measured acceleration and the simulated acceleration was obtained. The contrast result of an experiment sample was shown in Fig. 4.

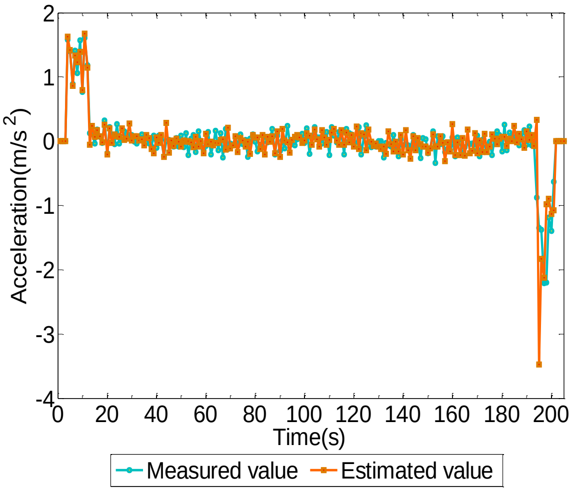

Fig. 4 Comparison of measured acceleration and simulated acceleration

The $v_{n,t}$, $v_{n-1,t}-v_{n,t}$ and d_t were verified respectively using $\hat{a}_{n,t}$ obtained from regression, and the comparison between simulated and measured values were obtained, shown in Fig. 5, Fig. 6 and Fig. 7.

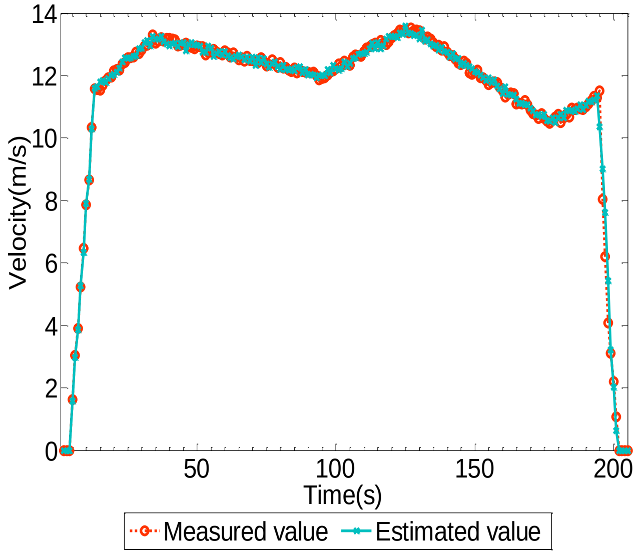

Fig. 5 Comparison of measured velocity and simulated velocity

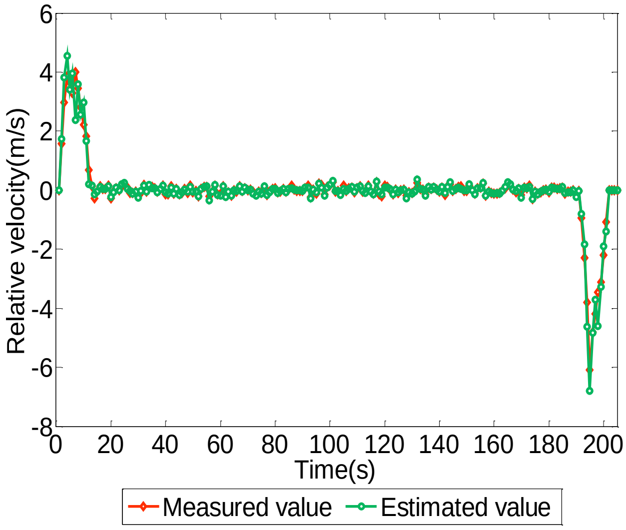

Fig. 6 Comparison of measured relative velocity and simulated relative velocity

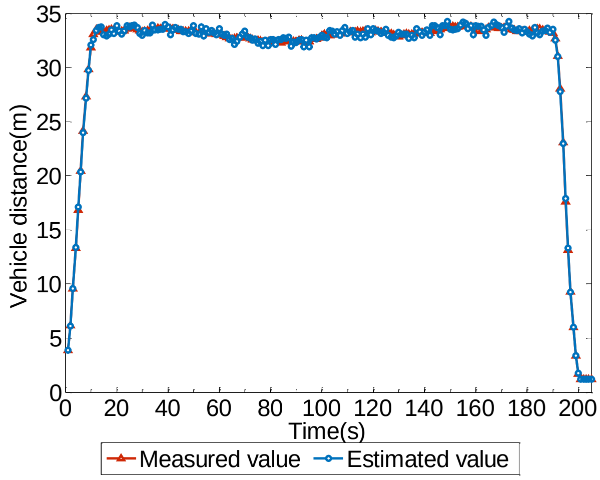

Fig. 7 Comparison of measured distance and simulated distance

According to the results of the model validation, the measured acceleration and the mean acceleration obtained by nonparametric kernel density estimation model were well fitted. The measured acceleration was small swing with the axis of regression acceleration, and other parameters also had good fitting. In the car following process, as long as the driving behavior data was very close to the spatial domain with a set of data in the sample database, the driving behavior can be estimated through the average acceleration obtained from the regression model.

V. CONCLUSION

A car-following model based on nonparametric regression was presented in this paper for the deficiency of existing vehicle following model. The required assumption of nonparametric regression method is much weaker than the parameter regression method. The nonparametric regression method, with a wider range of applications and a strong ability that adapt to the data changes, is suitable for arbitrary distribution data. The error caused by model assumptions can be avoided in nonparametric regression [7, 8]. The research results of this paper provide a theoretical basis for the microscopic study of traffic theory under the background of big data. It should be pointed out that the phenomenon of "high dimension disaster" will appear with the increase of dimension in the process of nonparametric regression and a further research is needed. In addition, the model calibration and verification techniques need to be explored from the point of micro and macro view.

REFERENCES

[1] Y. T. Liu, *Traffic system simulation technology*. Beijing: China Communications Press, 2002.

[2] D. H. Wang, S. Jin, "Review and outlook of modeling of car following behavior" in *China Journal of Highway and Transport*, 2012, vol.25, no. 1, pp: 115-127.

[3] E. R. Boer, "Car following from the driver's perspective" in *Transportation Research Part F: Traffic Psychology and Behavior*, 1999, vol.2, no. 4, pp: 201-206.

[4] A. Z. Ye, *Nonparametric Econometrics*. Tianjin: Nankai University Press, 2003.

[5] J. R. Trapero. "Calculation of solar irradiation prediction intervals combining volatility and kernel density estimates" in *Energy*, 2016, vol.114, pp:266-274.

[6] X. Y. Wang, B. C. Su, Z. W. Meng, "Factor analysis to choose the input variable of car-following model for microscopic traffic simulation" in *Soft Science*, 2004, vol.18, no. 2, pp:16-19.

[7] SAS Institute. *SAS INSIGHT user's guide*. Cary: SAS Institute, 1993.

[8] P. J. Green, B. W. Silverman, *Nonparametric regression and generalized linear models*. London: Chapman and Hall, 1994.

Yaqi Liu
He was born on 12th December, 1989 in Shandong province, China. He is currently working toward the M.S. degree in traffic and transportation engineering. His research interests include analysis of driving behavior, affective computing, controlling and cooperative intelligence of human-vehicle-environment.

Zhenxue Liu
She was born on 16th July, 1991 in Shandong province, China. She is a graduate student at the Shandong University of Technology, and major in transportation engineering. Her research interest is controlling and cooperative intelligence of human-vehicle-environment.

Wei Tian
He was born on 26th January, 1991 in Shanxi province, China. He is a graduate student at the Shandong University of Technology, and major in transportation engineering. His research interests include Urban Transportation Planning, smart city, Intelligent Transportation.

Haibo Wang
He was born on 13th September, 1990 in Shandong province, China. He is currently working toward the M.S. degree in traffic and transportation engineering. His research interests include analysis of cyclist behavior, controlling and cooperative intelligence of human-vehicle-environment.

Xiaoyuan Wang
He was born on November, 1970 in Shandong province, China. He works as a professor with the School of Transportation and Vehicle Engineering, Shandong University of Technology. He also served as a visiting research fellow in State Key Laboratory of Automotive Safety and Energy, Tsinghua University. He is specialized in intelligent transportation systems, especially in cooperative intelligence of human-vehicle-environment. He has authored or coauthored more than 100 publications in international journals.