Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2023) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

25th Euro Working Group on Transportation Meeting (EWGT 2023)

A Similarity-Based Data-Driven Car-Following Model Considering Driver Heterogeneity

Zi-Jian Liu[a], Qing-Long Lu[b]*, Jing Gao[c]

[a]*Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China*
[b]*Chair of Transportation Systems Engineering, Technical University of Munich, Munich, Germany*
[c]*Department of Civil and Environmental Engineering, The Hong Hong University of Science and Technology, Hong Kong SAR, China*

Abstract

Human drivers usually have distinct driving patterns and preferences. Driver heterogeneity is crucial for modeling driving behaviors. This paper incorporates driver heterogeneity with data-driven approaches to predict car-following behaviors. A bi-level similarity-based car-following model is proposed to predict the vehicle's moving distance. In the upper level, drivers with similar driving patterns as the ego vehicle are identified using k-nearest neighboring (kNN) search. In the lower level, leveraging kNN model, candidate records are selected from the identified vehicles' trajectories and applied to predict the ego vehicle's moving distance, combining the driving pattern similarity measured in the upper level. By taking into account the driver heterogeneity, the proposed model is capable of identifying the most relevant driving situations, which leads to an improvement of prediction accuracy. Furthermore, the established bi-level structure largely shrinks the searching space of candidate records, which reduces the searching complexity and enhances computational efficiency. We quantitatively evaluate and compare the performance of the proposed model in terms of both prediction accuracy and computational efficiency using real-world vehicle trajectory data.

© 2023 The Authors. Published by Elsevier BV.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 25th Euro Working Group on Transportation Meeting (EWGT 2023)

Keywords: Car-following; k-nearest neighbor; driver heterogeneity

* Corresponding author.
E-mail address: qinglong.lu@tum.de

2352-1465 © 2023 The Authors. Published by Elsevier BV.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 25th Euro Working Group on Transportation Meeting (EWGT 2023)

1. Introduction

Microscopic car-following models play a significant and fundamental role in simulating traffic dynamics. While many efforts have been contributed to calibrating conventional car-following models (e.g., Wiedemann model), the performance and accuracy are limited by the parsimonious parameter settings of these models. The availability of high-fidelity traffic data and the emergence of data-driven approaches have endowed modeling car-following behaviors with more possibilities.

Complex data-driven models such as different forms of neural networks are one of the most widely adopted models in this field. ANN in Khodayari et al. (2012), RNN in Wang et al. (2017), BPNN in Yang et al. (2018), and deep reinforcement learning in Zhu et al. (2018) are trained to establish car-following models. Isolated from the conventional kinematics frameworks and restrictions, these models are proved to be efficient by fully utilizing field data's inherent characteristics. However, their tendencies to be over-fitted and the lack in interpretability and flexibility have made other straight-forward statistic models a more practical option in application scenarios. For instance, the car-following behavior decision algorithm proposed in Hao et al. (2018) generates an optimal decision rule set from raw data and chooses appropriate rules to determine the follower output. The kNN-based non-parametric model established in He et al. (2015) predicts the follower's moving distance with the average of the top k most similar cases. Such similarity-based data-driven approaches have proved their efficiency and practicality in not only building car-following models, but also reproducing the dynamics and characteristics of traffic flow. Nevertheless, their dependence on the volume and quality of the field data tends to cause skewness by some outliers or abnormal driving behaviors in the dataset, restricting their applications to traffic simulators.

Meanwhile, driver heterogeneity, defined in Ossen and Hoogendoorn (2007) and Ossen and Hoogendoorn (2011) as differences between driving behaviors of driver/vehicle combination under comparable conditions, is one of the most influential factors in establishing and calibrating conventional kinematic models. The heterogeneity was proved to be evident and essential by the existing literature, such as the visual imaging model proposed in Wang et al. (2010). Despite the essential influence on car-following behaviors, driver heterogeneity was seldom considered in data-driven models.

To this end, this paper aims to integrate the factor of driver heterogeneity into similarity-based data-driven approaches, as well as balance prediction accuracy and computational efficiency in modeling car-following behaviors, to develop a practical model that can be applied in both analyses and simulations. A data-driven model with a bi-level excavation rule is established in the following section. Then, the field data for case study and preprocessing procedure are presented. Finally, the driver heterogeneity, traffic dynamics, and model performance are validated and analyzed before the conclusions are drawn.

2. Methodology

Fig. 1 displays the structure of the proposed bi-level model. Historical driving trajectories and vehicle/driver (referred to as vehicle hereafter) characteristics will be processed and stored in two separate datasets, i.e., trajectory set and vehicle set. At the upper level, vehicles with similar driving patterns to the ego vehicle are selected from the vehicle set based on predefined similarity measures. At the lower level, the trajectories of the selected vehicles are then retrieved from the trajectory set, making up the candidate set for the record selection model. Analogously, records under a similar driving environment (which will be evaluated by record features) as the ego vehicle faces are drawn from the candidate set. The corresponding responses (here the moving distance) of those records are then used to estimate the action of the ego vehicle at the following time step. The underlying assumption of this model is that vehicles with similar driving patterns and characteristics will behave similarly in the face of alike situations.

In this study, the k-nearest neighbor (kNN) algorithm is applied to search for vehicles and records in the upper layer and low layer, respectively. kNN is simple to understand but very effective in many estimation and prediction problems. Its potential in modeling and predicting car-following behaviors has been proved in He et al. (2015). The experiment results therein showed that it is capable of replicating the platoon traffic, rubbernecking scenario, etc. The resulted fundamental diagrams were also consistent with the empirical ones. kNN has also been used to distinguish driving patterns in Lu et al. (2021). The similarity measures used to select vehicles and records in this study are elaborated in the following text of this section.

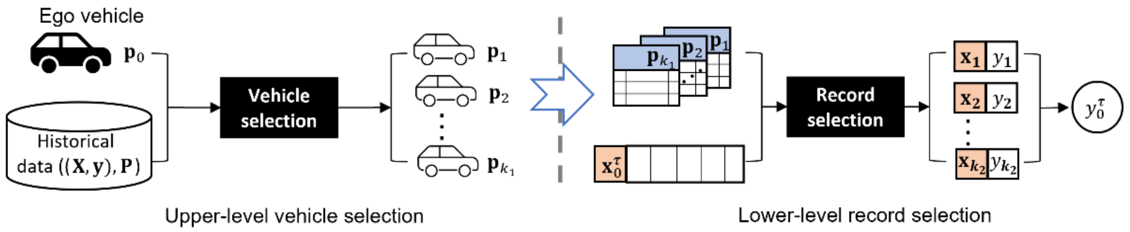


Fig. 1. Hierarchy of the bi-level car-following model.

2.1. Upper-level vehicle selection

Let $\mathbf{P} \in \mathbb{R}^{n_v \times n_u}$ be the vehicle set with n_v the number of vehicles and n_u the number of features describing driving patterns. A vehicle instance is then denoted by $\mathbf{p}_i \in \mathbb{R}^{n_u}$ (a row of $\mathbf{P}$). In particular, we denote the ego vehicle as $\mathbf{p}_0$. The Euclidean distance is applied to measure the difference between vehicle i and the ego vehicle as below.

$$d_i^u = \sqrt{(\mathbf{p}_i - \mathbf{p}_0)^T (\mathbf{p}_i - \mathbf{p}_0)} \tag{1}$$

where the superscript u of d_i^u indicates *upper level*. Noteworthy, $\mathbf{P}$ is expected to include the important statistical moments of variables that portray vehicle trajectories, such as the mean and variance of velocity and headway. $k_1 (k_1 \le n_v)$ vehicles with the smallest d^u's to the ego vehicle will be selected. k_1 is a hyper-parameter of the model.

2.2. Lower-level vehicle selection

Let $\mathbf{X}(\mathbf{p_0}) \in \mathbb{R}^{n_r(\mathbf{p}_0) \times n_l}$ be the trajectory set of the vehicles selected for the ego vehicle $\mathbf{p}_0$ in the upper-level procedure. $n_r(\mathbf{p}_0)$ is the number of records and n_l is the number of features reflecting the traffic situation. A record instance is then given by $\mathbf{x}_i \in \mathbb{R}^{n_l}$ (a row of $\mathbf{X}$). We denote the trajectory record of the ego vehicle at time τ as $\mathbf{x}_0^\tau$. We apply the weighted Euclidean distance to calculate the difference between the candidate record j and the record in focal.

$$d_j^l = \sqrt{(\mathbf{x}_j - \mathbf{x}_0^\tau)^T \mathbf{W} (\mathbf{x}_j - \mathbf{x}_0^\tau)} \tag{2}$$

where the superscript l of d_j^l indicates *lower level*, and $\mathbf{W} = \text{diag}(w_1, ...w_m, ...w_{n_l})$ is a weighting matrix with the m-th diagonal entry proportional to the absolute Pearson correlation between the m-th column in $\mathbf{X}$ (noted by $\mathbf{X}_m$) and the response vector (i.e., the moving distance at the next time step, $\mathbf{y}$), i.e., $w_m = | corr(\mathbf{X}_m, \mathbf{y}) | / \sum_{k=1}^{n_l} | corr(\mathbf{X}_k, \mathbf{y}) |$

Equation (2) implies that the contribution of feature difference to the total difference, d_j^l, are weighted by the features' correlation with the moving distance. The introduction of the weighting matrix emphasizes more on the significant features while allowing more disturbance in the insignificant features. It is worth pointing out, the trajectory set $\mathbf{X}$ should include the important indicators that can represent the state changes of a leader-follower pair, such as headway, velocity, and acceleration. $k_2 (k_2 \le n_r(\mathbf{p}_0))$ records with the smallest d^l's will be selected and respective moving distance of these records will be used to predict the moving distance under $\mathbf{x}_0$. k_2 is another hyper-parameter.

2.3. Car-following behavior prediction

Let $\hat{\mathbf{y}}$ be the vector of moving distances of the selected records. Denote y_0^τ as the moving distance between τ and $\tau+1$ (we set 1 s as the length of a time step) to be predicted. Note, rather than simply averaging the values in $\hat{\mathbf{y}}$, a weighting vector based on vehicle difference will be applied to $\hat{\mathbf{y}}$ to predict y_0^τ, i.e., $y_0^\tau = \hat{\Lambda}^T \hat{\mathbf{y}}$, where $\hat{\Lambda}$ is a weighting vector with each entry equals the inverse of the vehicle difference $d_{\mathbf{v}(j)}^u$, so that larger weights indicate greater similarity. $\mathbf{v}(j)$ indicates the vehicle to which record j belongs. Denote λ_j as the j-th ($j \le k_1$) entry in $\hat{\Lambda}$. λ_j can then be computed by

$$\lambda_j = \left(1/d^u_{\mathbf{v}(j)}\right) / \left(\sum_{k=1}^{k_2} 1/d^u_{\mathbf{v}(k)}\right) \tag{4}$$

One can also apply appropriate kernel functions to transform the distances to weights, such as the heat kernel which has been widely applied in graph-based data mining algorithms to decide if an edge should be added between two nodes when constructing a neighborhood graph.

Noteworthy, the proposed model is designed for modeling car-following behaviors (the time headway should be within a limited range, e.g., 10s (Treiber and Kesting, 2013)). In specific circumstances where the car-following state terminates (e.g., an increasingly large headway or when a preceding vehicle has left the network, etc.), the follower is assumed to be driving at a desired velocity, referring to conventional kinematic car-following models.

2.4. Relaxation on k_1 and k_2

k_1 and k_2 are two hyper-parameters in this model. However, they can be relaxed with two thresholds, ϵ_1 and ϵ_2, to control the similarity of homogeneous vehicles and similar records. Instead of selecting a fixed number of vehicles and records, ϵ_1 and ϵ_2 can be applied to select vehicles and records in a soft and flexible way, namely, only vehicles with d^u less than ϵ_1 and records d^l less than ϵ_2 will be selected.

While the kNN-based car-following model present in He et al. (2015) illustrates its viability in replicating typical car-following flows, the proposed bi-level framework can improve its robustness from two aspects: (i) adding another layer for evaluating the driver heterogeneity which is also used to weight the moving distance of the selected records in prediction (i.e., $\hat{\Lambda}$ in Section 2.3); (ii) considering the variable correlation via the weighting matrix $\mathbf{W}$ in Equation (2). Furthermore, the consideration of driver heterogeneity can also promote the computational efficiency of its ancestor. Specifically, the efficiency is increased by $n_r / (n_v + n_r(\mathbf{p}_0))$ times, where n_r is the total number of records in the entire trajectory set. This significantly increases the possibility of incorporating it into traffic simulators, which used to be a primary limitation of data-driven car-following models.

3. Experiment Setup

3.1. Data description and processing

The proposed model are trained and validated with the NGSIM dataset. The trajectories adopted in this paper were sampled from south-bound US-101 highway, spreading over a 45-minute period during a congested morning peak hour(from 7:50 a.m. to 8:35 a.m. on June 15, 2005), with majority of the speeds are less than 60 km/h. The data was aggregated from 0.1s to 1s, to smooth out the detection errors and simulate the drivers' reaction time.

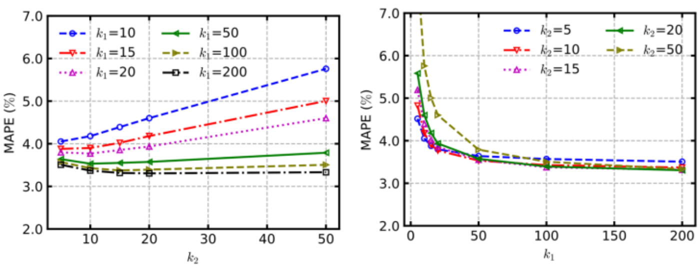
A vehicle set is first extracted in advance to support the driver-layer filtering, which identifies each car-following pair as [vehicle ID, preceding vehicle ID], resulting in 3,740 pairs. This not only recognizes the unique drivers, but also distinguishes their following behaviors under different contexts (led by different vehicles). As mentioned in Section 2.1 the characteristics of each car-following pair would be described by the statistical moments of the corresponding trajectory. Here, we use the mean and standard deviation of velocity, acceleration, absolute acceleration, space headway and time headway as vehicle features. Namely, each row in $\mathbf{P}$ will represent a car-following pair instance composed of those features which reflect their historical driving patterns under different situations. It should be noted that we excluded the invalid pairs by a one-minute threshold, i.e., only trajectories last longer than one minute will be considered. This, however, is necessary to avoid frequent lane-changing and over-taking behaviors, which may introduce noise into the dataset.

141,650 records belonging to the selected car-following pairs are available in the trajectory set. In the experiment, the moving distance of records will be predicted. For the lower-level model, partially referring to the correlation matrix in Fig. 3, we selected velocity v, acceleration a and time headway T as model features, which are listed as the first three variables. Obviously, they are with a rather weak correlation to one another. It is worth noticing that the selected features are entirely different from inputs of the plain kNN model in He et al. (2015): the space headway of the vehicle H and the moving distance y_p of its leader in both current and previous time step (denoted by $\tau-1$). However, referring to Fig. 3, H is highly correlated to $H^{\tau-1}$, and y_p is highly correlated to $y_p^{\tau-1}$. Therefore, two variables are statistically redundant and contribute little information, and our following experiments show that this redundancy will significantly impede the model performance.

3.2. Determination of k_1 and k_2

The selection of k in kNN models is usually done by empirical estimation, or performing grid search on the dataset. In this paper, we determine k_1 and k_2 by a grid search based on the prediction results of the records of 144 randomly sampled car-following pairs. k_1 is selected from [10, 15, 20, 50, 100, 200] and k_2 is selected from [5, 10, 15, 20, 50]. Mean absolute percentage error (MAPE) is used to measure the model performance under different combinations of k_1 and k_2.

As shown in Fig. 2a a larger k_1 indicates a more precise prediction, by involving more similar records at the lower level. However, outliers or abnormal driving behaviors might also be introduced in this case. It's worth noticing that the increasing of k_1 can only contribute to a limited improvement in MAPE after k_1 reaches above 50. Considering the trade-off between computation costs and output accuracy, and to avoid the inclusion of more noise, it would be a more suitable practice to settle k_1 at 50. In terms of k_2, Fig. 2b tells that smaller values (e.g., $k_2 \leq 20$) suggest more accurate output. The sole increase of k_2 will involve more irrelevant records, especially with a smaller k_1 narrowing down the candidates within a more relevant scope. However, the line of k_2 =5 shows that the performance improvement will also be restricted by a very small k_2 when increasing k_1 (e.g., at k_1=200, k_2=5 results in a bigger MAPE than k_2=10). Therefore, we specify k_2=10 in the following model evaluations.



(a) k_1 (b) k_2

Fig. 2. Sensitivity analysis on model parameters.

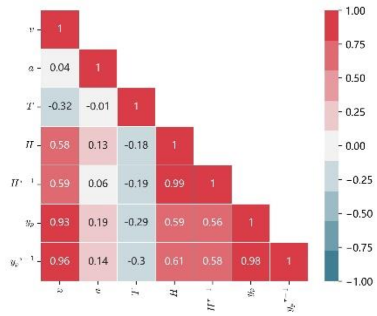


Fig. 3. Correlation matrix heatmap of key attributes.

4. Model evaluation

This section first applies a hypothesis testing method to validate the existence of driver heterogeneity among car-following pairs and shows that the kNN in the upper level can effectively construct homogeneous vehicle sets. The capability of the bi-level model of simulating traffic dynamics of the study highway is then tested. Finally, models with different input features are compared, showing the superiority of the proposed model in prediction accuracy and computational efficiency.

4.1. Driver heterogeneity evaluation

Zhang et al. (2016) applied a hypothesis testing on the mean speeds of car-following-car trajectories and car-following-truck trajectories to check if they are significantly different, namely, if they are from two different populations. Specifically, we assume two population means are μ_1 and μ_2, respectively, and sample means are $\overline{Y}_1$ and $\overline{Y}_2$. The null hypothesis then can be stated as H_0 : $\mu_1 - \mu_2 = 0$, while the alternative hypothesis is $H_1 : \mu_1 - \mu_2 \neq 0$. The test statistics can be written as: $Z_0 = (\overline{Y}_1 - \overline{Y}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2}$, where s_1^2 and s_2^2 are the sample variance, n_1 and n_2 are the number of observations in respective samples. Considering a significance level α, if Z_0 is out of the range of [- $Z_{\alpha/2}$, $Z_{\alpha/2}$], the null hypothesis H_0 will then be rejected, indicating two samples are drawn from different populations.

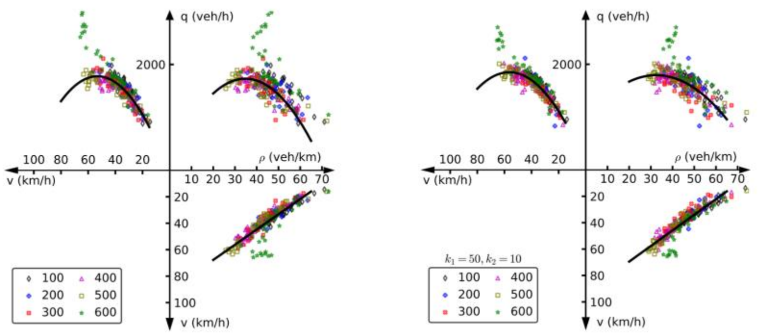
Analogously, we apply this measure to verify the existence of driver heterogeneity so as to show the necessity of adding the upper level (driver layer). We perform the tests for four variables separately, including the velocity, acceleration, time headway, and moving distance. The former three are used in the lower-level selection, while the latter is the response variable of the model. The trajectories of the vehicles (car-following pairs) selected from the upper level will be treated as the first sample, and the rest of the trajectories will be the second. For each vehicle, k_1 vehicles will be deemed to have a similar driving pattern as per the proposed bi-level model. The test is conducted for each driver particularly. As a result, 3740 tests will be carried out for each variable. Table 1 summarizes the test results. The first column lists the variables, the second and the third column provides the percentage of tests rejecting H_0 at the respective significance level. Obviously, except the tests on acceleration, almost all tests on other variables approve the driver heterogeneity. Rigorously, with $\alpha = 0.1$, 69.5% of tests on acceleration also reject H_0. Thus, it is plausible to claim that driver heterogeneity is common in car-following behaviors and should be explicitly considered in modeling. Furthermore, kNN can efficiently find out homogeneous drivers and thus guarantee the conformity of car-following behaviors.

Table 1. Summary of test statistics of selected vehicles.

Variable	Velocity	Acceleration	Time Headway	Moving Distance
$\alpha = 0.1$	97.9%	69.5%	100%	97.7%
$\alpha = 0.01$	96.8%	55.6%	100%	96.7%

4.2. Simulation evaluation

Fig. 4 compares the fundamental diagrams of empirical traffic and simulated traffic on the study highway segment, divided into 6 sub-segments as illustrated in Fig. 5. Calculation are performed with the trajectory records tapped within the 30-meter area from the end of each sub-segment. Specifically, the sub-segment flow q_s (s is the sub-segment ID) is estimated by the inverse of the mean headway of the records, i.e., $q_s = 3600/\bar{T}_s$. Also, density ρ_s is estimated by the inverse of the mean space headway, i.e., $\rho_s = 1000/\bar{H}_s$, while speed $v_s = 3.6\bar{v}_s$. They are used to explain the traffic dynamics of the respective sub-segment. These traffic characteristics are calculated on a minute basis. That is to say, each point in Fig. 4 represents the traffic status of a sub-segment within one specific minute. Considering the dataset is collected within a 45-minute interval, each sub-segment should have 45 flow-speed-density tuples (thus 45 points in the each subplot). Moreover, as shown in the figure, the data of flow-speed and flow-density are fitted with polynomial functions, and speed-density data is fitted with a linear function.



(a) Empirical traffic (b) Estimated traffic

Fig. 4. Fundamental diagrams.

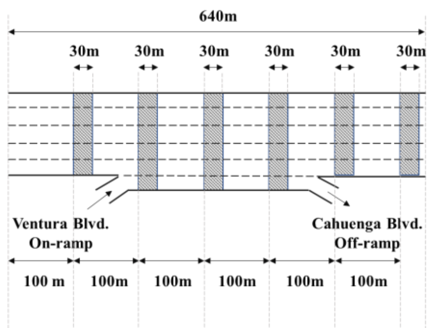


Fig. 5. Cross-sections for measuring traffic dynamics.

It can be easily seen from Fig. 4 that the two plots share high similarity, implying the bi-level model's capability in reproducing the characteristics of the empirical traffic. More specifically, it highly recurrents the congested status on the first four sub-segments, and the congestion dispersion from fifth sub-segment to the last one. The congestion release is captured by the non-congested status on the two sub-segments and a sudden increase on the last one. One should notice that this traffic propagation pattern illustrated by both diagrams corresponds to the ramp settings on the studied US-101 segment: the on-ramp merged with the highway near the tail of the second sub-segment and the off-ramp diverged from the highway on the fifth sub-segment. In general, the merging traffic from the on-ramp and diverging traffic to the off-ramp indicate more potential lane-changing conflicts on the main lanes in-between. As a result, a traffic jam is easily generated within this area and its immediate upstream, which is illustrated by the points of the first four sub-segments gathering around the right half of the q - v curve and q - ρ curve. However, the congestion obviously dissipated over the downstream of the off-ramp, and the traffic moved at a higher speed according to the v - ρ diagram. The high-fidelity reproduction of the above details proves the capability of the proposed model in capturing the major characteristics and propagation pattern in the traffic flow.

4.3. Model comparison

This subsection aims at demonstrating the superiority and efficiency of the proposed model by comparing against other car-following models, including both conventional and data-driven ones, in the aspects of prediction accuracy and computational costs. In terms of data-driven models, we compare our model against another kNN-based model proposed in He et al. (2015), which shares a similar model architecture but with different model features. In terms of conventional models, we calibrated two representative conventional car-following models, i.e., the Intelligent Driver Model (IDM, Treiber et al., 2000) and the Krauss model (Krauß et al., 1997), with the same NGSIM dataset utilizing the Finite Difference Stochastic Approximation (FDSA, Spall, 1992) algorithm.

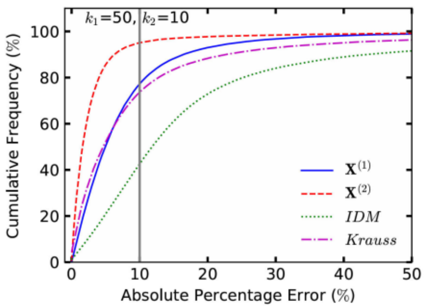


Fig. 6. Model Comparison.

The blue and red lines in Fig. 6 show the cumulative distribution of the absolute percentage errors of predictions by applying the proposed model with two different feature sets respectively. $\mathbf{X}^{(1)}$ denotes the features constructed in He et al. (2015) (i.e., the space headway of the ego vehicle to its leader and the moving distance of its leader, in two successive time steps). $\mathbf{X}^{(2)}$ denotes the features adopted in our model, composed of velocity, acceleration and headway. It can be easily seen that the model using the features constructed in He et al. (2015) failed to capture the essential representatives of the record attributes, resulting in a relatively poor accuracy performance with more than 20% of the predictions have a deviation bigger than 10% from the true values. In contrast, by using the features selected in this study, nearly 90% of the model outputs make an error of less than 10%. Furthermore, referring to the green dashed line and purple dotted-dashed line in Fig. 6 conventional kinematics models result in less accurate estimations, especially the IDM model, only with less than 40% of its prediction make an error of less than 10%.

Considering the computational costs, similarity-based data-driven models are intrinsically less efficient than traditional kinematic models due to the process of mining similar records of every target, lacking universality compared with the calibrated conventional models simulation. However, the bi-level architecture of our model largely shrinks the searching range within a reasonable threshold, and it demonstrates an outstanding efficiency edge over the plain kNN model proposed in He et al. (2015) As discussed in Section 2, *average* query records of the bi-level model can be estimated by $n_v + k_1 n_r / n_v$, while the query of the single-layer kNN driven by the same dataset will be executed n_r times. Thus, the improvement in efficiency is dependent on the dataset used and k_1. For example, in the study dataset, $n_r = 141,650$, $n_v = 3740$. If we set $k_1 = 50$ for the bi-level model, then the improvement in computational efficiency will increase by $n_r / (n_v + k_1 n_r / n_v) \approx 26$ times in average. Such an improvement makes assembling it into traffic simulators possible.

5. Conclusions

This paper proposes a bi-level kNN model that incorporates driver heterogeneity into data-driven approaches in predicting car-following behaviors, i.e., vehicles' moving distances. In the upper level, vehicles with similar driving patterns as the ego vehicle are identified based on the driving features including the mean and standard deviation of velocity, acceleration, absolute acceleration, space headway and time headway. In the lower level, using the Euclidean distance weighted by driving features, candidate records are selected from the trajectory sets of vehicles generated from the upper level. Finally, the moving distance of the ego vehicle is predicted by a weighted average of moving distances of the selected candidate records based on the driving pattern similarity measured in the upper level.

The proposed prediction model is calibrated and evaluated based on the real-world vehicle trajectory dataset. The evaluation results demonstrate the necessity of considering the driver heterogeneity, as well as the model's capability of accurately predicting drivers' behaviors/vehicles' movements. The predicted results can reproduce the microscopic fundamental diagrams extracted from the real-world traffic flow and capture the generation, propagation, and release of the congestion on the highway. Besides, the bi-level prediction model considering driver heterogeneity largely reduces the computational complexity and improves the computational efficiency. This implies considerable potential in practical use, such as the integration into traffic simulators.

References

Hao, S., Yang, L., Shi, Y., 2018. Data-driven car-following model based on rough set theory. IET Intelligent Transport Systems 12, 49–57.

He, Z., Zheng, L., Guan, W., 2015. A simple nonparametric car-following model driven by field data. Transportation Research Part B: Methodological 80, 185–201.

Khodayari, A., Ghaffari, A., Kazemi, R. and Braunstingl, R., 2012. A modified car-following model based on a neural network model of the human driver effects. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 42(6), pp.1440-1449.

Krauß, S., Wagner, P., Gawron, C., 1997. Metastable states in a microscopic model of traffic flow. Physical Review E 55, 5597.

Lu, Q.L., Qurashi, M., Varesanovic, D., Sodnik, J., Antoniou, C., 2021. Exploring the influence of automated driving styles on network efficiency. Transportation research procedia 52, 380–387.

Ossen, S., Hoogendoorn, S.P., 2007. Driver heterogeneity in car following and its impact on modeling traffic dynamics. Transportation Research Record 1999, 95–103.

Ossen, S., Hoogendoorn, S.P., 2011. Heterogeneity in car-following behavior: Theory and empirics. Transportation research part C: emerging technologies 19, 182–195.

Spall, J.C., 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE transactions on automatic control 37, 332–341.

Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Physical review E 62, 1805.

Treiber, M., Kesting, A., 2013. Traffic flow dynamics. Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg, 983–1000.

Wang, H., Wang, W., Chen, J., Jing, M., 2010. Using trajectory data to analyze intradriver heterogeneity in car-following. Transportation Research Record 2188, 85–95.

Wang, X., Jiang, R., Li, L., Lin, Y., Zheng, X., Wang, F.Y., 2017. Capturing car-following behaviors by deep learning. IEEE Transactions on Intelligent Transportation Systems 19, 910–920.

Yang, D., Zhu, L., Liu, Y., Wu, D., Ran, B., 2018. A novel car-following control model combining machine learning and kinematics models for automated vehicles. IEEE Transactions on Intelligent Transportation Systems 20, 1991–2000.

Zheng, L., Jin, P.J., Huang, H., Gao, M., Ran, B., 2016. A vehicle type-dependent visual imaging model for analysing the heterogeneous car-following dynamics. Transportmetrica B: Transport Dynamics 4, 68–85..

Zhu, M., Wang, X., Wang, Y., 2018. Human-like autonomous car-following model with deep reinforcement learning. Transportation research part C: emerging technologies 97, 348–368.

NGSIM, 2006. The Next Generation Simulation Program. http://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.