

Regular Paper

Accurate and Efficient Driving Intention Inference Based on Traffic Environment Information and FES-XGB Framework

SHUO WANG^{1,a)} HIDEKI FUJII^{1,b)} SHINOBU YOSHIMURA^{1,c)}

Received: April 11, 2021, Accepted: October 8, 2021

Abstract: Reliable driving intention inference is an essential issue in the mixed automation traffic system. To improve traffic safety and efficiency, this study develops an accurate and efficient driving intention inference framework named FES-XGB, which is short for Feature Extraction and Selection based eXtreme Gradient Boosting (XGBoost) algorithm. In contrast with conventional approaches, which only consider motion information of the subject and neighboring vehicles, this study includes a new kind of decision variables into driving intention inference for the first time, i.e., the local and global traffic environment information assumed to be obtained from vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) technology. The high-precision NGSim trajectory dataset is employed to learn the relationship between traffic environment information and driving intentions and evaluate the proposed framework. According to the experiment results, by taking the environment information as additional input, the accuracy of the conventional XGBoost model can increase from 89.42% to 92.86%, indicating the environment information has a close relationship with the driving intention. By employing the proposed FES-XGB framework, the accuracy can be further increased to 94.09%, while the training and online inference cost can be reduced by 94.03% and 65.25% respectively. With the traffic environment information as additional input, the proposed FES-XGB framework can be integrated into advanced driver-assistance systems (ADAS) for a safer and more efficient traffic system.

Keywords: driving intention inference framework, traffic environment information, feature extraction and selection, XGBoost algorithm

1. Introduction

Nowadays, autonomous vehicles are gradually being introduced into the traffic system [1]. A mixed automation traffic system consisting of autonomous and human-driving vehicles has become challenging for both sides [2]. The appearance of autonomous vehicles will disturb the cooperative driving among humans and cause inconvenience [3]. At the same time, once autonomous vehicles can't properly understand human driving intentions, they will get involved in dangerous occasions even accidents [4].

To improve the safety and efficiency of the mixed automation traffic system, a unified ADAS should be developed [5], [6], [7], [8], [9]. It should be able to be implemented in both HDVs and CAVs, providing driving assistance to human driving vehicles (HDVs) and taking full control of connected autonomous vehicles (CAVs) to realize autonomous driving [5], [6], [7]. From the human driver perspective, the ADAS should recognize and predict the driving intention of human drivers in an accurate and timely manner, thus providing lane departure warning or lane-keep assistance without disturbing the normal driving [6], [9]. From the autonomous control perspective, the trained ADAS model should

well understand and imitate human behaviors, so as to generate more social and natural autonomous driving maneuvers [3], [8].

The emerging connected vehicle technology can provide unprecedented abundant information to vehicles. With the aid of V2V and V2I technology, the environment information of the traffic system can be surely obtained by vehicles [10], [11]. On the other hand, although traditional HDVs have no access to explicit environment information, the traffic condition can be a latent perception of human drivers and influence the driving intention. Therefore, it is reasonable to assume that there is a mapping between traffic environment information and human driving intentions. In other words, taking traffic environment information as an additional input, the driving intention inference model can better assist humans to make decisions, and can navigate CAVs to behave more human-like at the same time.

This study designs a driving intention inference framework named FES-XGB, which combines feature extraction, feature selection and XGBoost algorithm. First, in addition to conventional vehicle information, the traffic environment information, including local environment and global environment information, are employed as an additional input of driving intention inference model. Second, a statistical-based feature extraction method is employed to identify the key influential factors on driving intentions to obtain more concise and efficient model inputs. Third, a feature selection method based on Shapley values is employed,

¹ The University of Tokyo, Bunkyo, Tokyo 113-8656, Japan

a) sh_wang@save.sys.t.u-tokyo.ac.jp

b) fujii@sys.t.u-tokyo.ac.jp

c) yoshi@sys.t.u-tokyo.ac.jp

which can select the most useful features and abandon those with minimal contributions, so as to strike a balance between model accuracy and complexity. At last, by using the above features as model input, the XGBoost model is trained to conduct the driving intention inference.

The main aim of this study is to achieve accurate and efficient driving intention inference with less training time needed. The experiment results prove that the aim can be realized by either adding traffic environment variables to the model input or employing a feature extraction and selection process, and a combination can produce the best result.

The remainder of the paper is organized as follows. Section 2 reviews the existing research on driving intention prediction; Section 3 provides a detailed description of the methodologies to realize the proposed intention inference framework, including feature extraction, feature selection and XGBoost; In Section 4, the experiment settings, experiment results and analysis on results are presented. The final section is the discussion and conclusion.

2. Literature Review

Recent studies have studied the driving intention inference from various perspectives. Some researches directly focus on the driver's physical behaviors, e.g., electroencephalography (EEG), head pose, and eye gaze [12], [13], [14], [15], [16]. Kim et al. studied the neural correlates in different driving situations and found them useful in detecting a driver's intention before body response [12]. Wang et al. analyzed the time-frequency of EEG signals and clarified the difference in emergency and soft braking intentions [13]. Doshi et al. found the mapping between physical gestures and driving intentions. It is figured out that compared with eye gaze, head motion plays a more important role in recognizing and predicting driving intentions [15], [16]. To find early cues of driver's intention, some researchers turn to driving simulation or vehicle test method to investigate driver behaviors [6], [17], [18], [19], [20], while some others obtain the relationship between driving intention and vehicle maneuvers from vehicle trajectory data [21], [22], [23], [24], [25], [26], [27]. The real vehicle test and simulation data usually include the internal information of the tested vehicle, such as steering wheel angles, throttle and brake pedal pressure, while the trajectory data can include external information of a large number of vehicles, such as position, speed, acceleration, heading angle and so on Refs. [20], [27]. Compared with real vehicle test and simulation data, the trajectory data can provide temporal and spatial information of a large number of vehicles. They can offer a full view of interactions between vehicles, as well as vehicle and infrastructure when combined with the road topology data [21], [22], [23], [24], [25], [26], [27]. Therefore, this study focuses on the construction of intention inference models based on trajectory data.

On the other hand, machine learning (ML) and deep learning (DL) based methods have been proved effective in various branches of driving intention inference studies [13], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]. A binary logistic regression (BLR) model is proposed to establish lane-changing decision models in Ref. [22]. Artificial neural net-

Table 1 Summary of recent intention inference studies based on ML and trajectory data.

	Decision Variables			Evaluation	
	Ego vehicle motion	Neighbor vehicle motion	Environment information	Accuracy models/inputs	Efficiency models/inputs
[21]	✓	✓	×	✓/×	×/×
[22]	✓	✓	×	✓/×	×/×
[23]	✓	✓	×	✓/×	×/×
[24]	✓	✓	×	✓/×	×/×
[25]	✓	×	×	✓/×	×/×
[26]	✓	×	×	✓/×	×/×
[27]	✓	✓	×	✓/✓	×/×
This study	✓	✓	✓	✓/✓	✓/✓

work (ANN) and support vector machine (SVM) are used to augment vehicle state information and detect driving intentions in Ref. [17]. In Ref. [24], the driving behavior is learned by a two-layer LSTM model. In Ref. [26], the motion information of the studied vehicle is predicted using the online ARIMA algorithm, then the predicted results are combined with current motion information and fed into a Bi-LSTM model to detect lane-changing from lane-keeping. Although in many cases, deep learning (DL) methods outperform ML methods [18], [24], [25], [26], the training cost of DL methods could be relatively high since there are numerous parameters to be tuned. Therefore, ML based models are more applicable in real implementations of driving intention inference. In Ref. [28], several ML methods were compared, the tree-based algorithms, especially eXtreme Gradient Boosting (XGBoost) algorithm [29], were found to have the optimal performances in driving intention inference.

Table 1 is the summary and comparison of some recent driving intention inference models using ML and trajectory data, with two major limitations. First, traffic environment information can be latently perceived by human drivers and influence human intention, but most studies only take the vehicle motions as decision variables and ignore the environment information. Second, although finding the most relative decision variables from trajectory data is of vital importance for driving intention inference, neither of the existing literature has comprehensively analyzed the influence of input variables on the intention inference models, none of them have specifically discussed how to improve the training efficiency of the models by extracting and selecting decision variables.

The main contributions of this study are as follows:

- To the best of our knowledge, traffic environment information are introduced into decision variables of intention inference model for the first time. The newly introduced variables can increase the accuracy of driving intention inference.
- We propose a hybrid framework combining feature extraction, feature selection and XGBoost method, which can infer driving intention with a high accuracy and low training cost.

3. Methodology

The flowchart of the proposed FES-XGB framework is shown in **Fig. 1**, which can be divided into six parts: (1) decision variable extraction; (2) feature extraction (FE); (3) feature selection (FS); (4) model training; (5) online model inference; (6) model

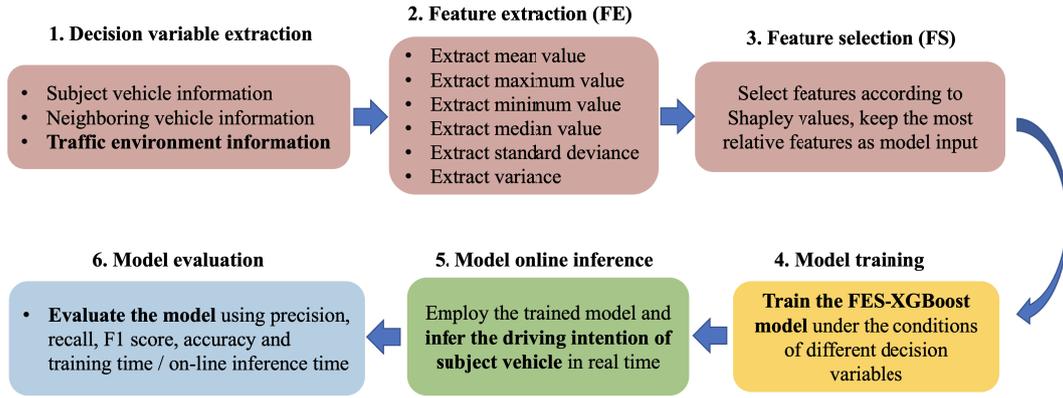


Fig. 1 Flowchart of the framework.

Table 2 Description of four kinds of decision variables.

Category	Variables	Symbol	Description
Subject vehicle information	Velocity	$v_i(t)$	Velocity of subject vehicle i at time t
	Acceleration	$a_i(t)$	Acceleration of subject vehicle i at time t
Neighboring vehicle information	Relative Position	$\Delta x_{ij}(t), \Delta y_{ij}(t)$	Longitudinal, lateral distance between vehicle i and j at time t
	Relative Velocity	$\Delta v_{ij}(t)$	Velocity difference between vehicle i and j at time t
	Relative Acceleration	$\Delta a_{ij}(t)$	Acceleration difference between vehicle i and j at time t
	Inverse Time To Collision	$iTTC_{ij}(t)$	$\Delta v_{ij}(t) / \Delta y_{ij}(t)$
	Time Headway	$THW_{ij}(t)$	$\Delta y_{ij}(t) / v_i(t)$
Local environment information	Local traffic velocity	$Local_{vel}(t)$	The lane-wise average velocity in 50 meters area centered by vehicle i at time t
	Local traffic volume	$Local_{vol}(t)$	The lane-wise average volume in 50 meters area centered by vehicle i at time t
Global environment information	Global traffic velocity	$Global_{vel}(t)$	The lane-wise average velocity in the entire study area at time t
	Global traffic volume	$Global_{vol}(t)$	The lane-wise average volume in the entire study area at time t

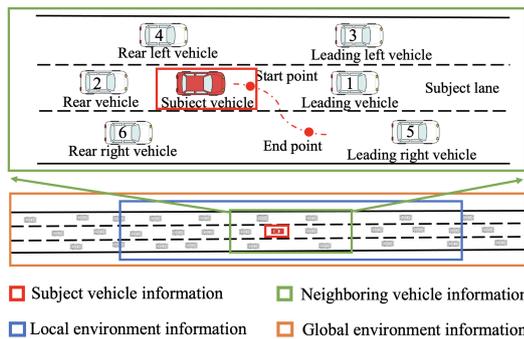


Fig. 2 The collection range of four kinds of decision variables.

evaluation.

Methodologies including decision variable extraction, feature extraction, feature selection and the theoretical background of XGBoost will be presented in this section.

3.1 Decision Variable Extraction

The trajectory data contains information on position, velocity, acceleration and lane number of each vehicle in every recorded time step. As can be seen in Fig. 2, from the trajectory data, by scaling the data collection range, four kinds of decision variables can be extracted, namely the subject vehicle information, the neighboring vehicle information, the local environment information and the global environment information. It should be noted that the neighboring vehicle information refers to the relationships between the subject vehicle and its six most adjacent neighbors. In case of the absence of neighboring vehicles, a pre-defined default value will be used correspondingly [30].

Table 2 is the detailed description of these four kinds of decision variables. The subject vehicle information includes the ve-

locity and acceleration of the subject vehicle in every time step. The neighboring vehicle information refers to the relative position, relative velocity, relative acceleration, inverse time to collision and time headway between the subject vehicle and its most adjacent neighbors within the 50 meters area centered by the subject vehicle. The local environment information is the lane-wise traffic velocity and volume within the 50 meters area. The global environment information is the lane-wise traffic velocity and volume of the entire study area. Among the above four categories of decision variables, the subject vehicle information and neighboring vehicle information are commonly employed in literatures, while the local and global environment information has not been explored. However, with the development of communication and data storage technology, the V2V, V2I and in-vehicle sensors can provide real-time global and local traffic environment information to vehicles. Therefore, it is essential to find the relationship between the driving intention and the traffic environment information, so as to better infer driving intention.

3.2 Feature Extraction and Selection

With the increase of historical sequence length and numbers of input variables, the input dimension of the driving intention inference models will become larger. Among the input features, some may have strong correlations with each other and carry redundant information. As a result, it will take longer time to train the model and the model accuracy will be reduced.

To avoid the aforementioned problem, some operations to capture and select the core information from the input features should be defined. In this study, feature extraction and selection operations have been introduced to produce an accurate and efficient intention inference process.



Fig. 3 Modeling process of XGBoost algorithm.

The feature extraction is realized through statistical analysis of each sample. In other words, the mean, maximum, minimum, median, standard deviation, sum and variance of each decision variable sequence are calculated, and these extracted features are set as input of the intention inference algorithm instead of the original decision variables. Such a feature extraction operation can reduce the dimensions of the model input, but will not reduce the accuracy of the model output if the extracted features can well represent the original decision variables.

The feature selection is realized through introducing a model explanation method named SHAP, short for SHapley Additive ex-Planations [33]. The SHAP method defines an explanation model to evaluate the contribution of each input feature. The explanation model is a linear function of binary variables [34]:

$$g(z') = \rho_0 + \sum_i^M (\rho_i, z'_i) \quad (1)$$

Where $z'_i \in \{0, 1\}^M$, $z'_i = 1$ when the i^{th} feature is presented, otherwise, $z'_i = 0$; M is the number of input features; ρ_i is the feature contribution of the i^{th} feature and $\rho_i \in R$.

The ρ_i is calculated by the game theory concept Shapley value, given a model f and the entire feature set F , the explanation model g will test a feature set S , which is a subset of F , and observe the model result. A model $f_{S \cup \{i\}}$ represents the i^{th} feature is presented, another model f_S represents the i^{th} feature is absent. The difference between $f_{S \cup \{i\}}(S \cup \{i\})$ and $f_S(S)$ can be used to evaluate the contribution of the i^{th} feature. The Shapley value can be represented as:

$$\rho_i = \sum_{S \in F/\{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(S \cup \{i\}) - f_S(S)] \quad (2)$$

SHAP can help explain which feature contributes more in obtaining the model output, features with high SHAP values are highly relevant to the model output, while features with low SHAP values only carry minimal useful information.

3.3 Extreme Gradient Boosting Algorithm

XGBoost, short for eXtreme Gradient Boosting algorithm, is a scalable ensemble learning method with many algorithmic innovations, such as approximate greedy search and parallel training to reduce time cost. It also has the advantage of effective tree pruning and the ability to handle missing values [29].

As shown in **Fig. 3**, XGBoost follows the gradient boosting framework proposed by Friedman [35], completing the learning task by building and combining multiple weak learners. It adds new decision trees to fit the residual of previous decision trees, so the accuracy can be gradually increased in the training process.

The target function is designed as follows.

$$L(\theta) = \sum_i l(\hat{y}_i, y_i) + \sum_k \psi(f_k) \quad (3)$$

$$\psi(f_k) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (4)$$

The first term in Eq. (3) is the cross-entropy loss function evaluating model output value \hat{y}_i and ground-truth y_i , which aims to increase the accuracy. The second term in Eq. (3) is a regulation term to reduce the complexity of the tree model. The composition of $\psi(f_k)$ is shown in Eq. (4), where the first term is the punishment of number of leaves T in the tree; the second term is to smooth the learnt weight of the leaf point and reduce over-fitting problems.

4. Numerical Experiment

The experiments are implemented on a ubuntu system computer with Intel Core i9 CPU, 3.60 GHz, 64 GB memory. All the models are established based on python 3.8.

4.1 Evaluation Criterion and Model Settings

To evaluate the prediction results of different models, evaluation criteria including precision, recall, F1 score and accuracy are utilized, as shown in Eqs. (5)–(8).

$$\text{precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1 score} = \frac{2 * \text{recall} * \text{precision}}{\text{recall} + \text{precision}} \quad (7)$$

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

Where TP is True Positive, representing a positive sample predicted as positive by the model; TN is True Negative, representing a negative sample predicted as negative by the model; FP is False Positive, representing a negative sample predicted as positive by the model; FN is False Negative, representing a positive sample predicted as negative by the model. In addition to the model accuracy, another evaluation criterion is the training time. Under conditions of large-scale problems, models with the short training time will have advantages in the deployment. In this study, the training time is recorded using the CPU time, which specifically refers to the time after determining the hyper-parameters.

The settings of XGBoost and other baseline models are listed in **Table 3**. XGBoost employs the xgboost package as backend and baseline models use the scikit-learn package as backend. Since SVM and tree based models are sensitive to hyper-parameters and the hyper-parameter values must be re-determined in different input cases, a parameter search range is defined, in which the grid search and 3-fold cross validation are employed to find the suitable parameters.

4.2 Dataset Description

The dataset used in this study comes from the Next Generation Simulation (NGSim) project of the U.S. federal highway administration (<https://ops.fhwa.dot.gov/trafficanalysisstools/ngsim.html>). Two datasets named I-80 and US-101 are employed and the location of the study area can be seen in **Fig. 4**.

As shown in Fig. 4, the study areas of the I-80 and the US-101 dataset are 503 and 640 meters long respectively and cover 6 lanes. The datasets provide high precision trajectory information

Table 3 Model settings.

Model	Parameter search range	Training stop criterion
LR	max_iter: 1e7 solver: 'lbfgs'	The loss function differs less than threshold/ reach the max_iter
MLP	max_iter: 1e4 solver: 'sgd' activation: 'relu' hidden_units: (200, 200)	The loss function differs less than threshold/ reach the max_iter
SVM	C: 1e-3, 1e-2, 1e-1, 1, 1e2, 1e3 γ : 1e-3 1e-4	The loss function differs less than threshold
DT	criterion: 'entropy', 'gini', max_depth: 10, 20, 1e2, 1e3, 2e3, min_samples_split: 2, 3	The tree depth reach max_depth
GBDT	max_depth: 3, 4, 5, n_estimators: 10, 50, 100	The tree depth reaches max_depth and the number of regression trees reach n_estimators
XGBoost	max_depth: 3, 4, 5, n_estimators: 10, 50, 100, min_child_weight: 1, 5	

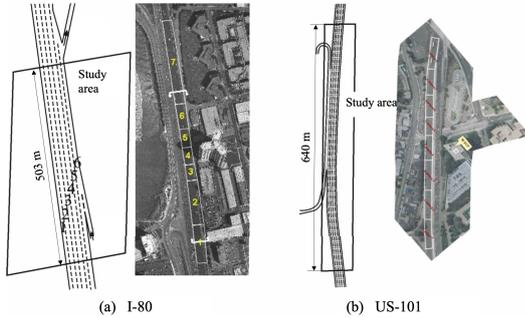


Fig. 4 Study area of NGSim dataset.

of all the vehicles in the study area, including the vehicle coordinates, velocity, acceleration, vehicle type and lane number every 0.1 seconds.

The extracted trajectory data need to be divided into left turn, right turn and keep straight with corresponding labels. In this study, the start and end points of lane change intention are determined by the heading angle of the vehicle, which is calculated from the change of vehicle position by using 3 successive coordinates. Figure 5 is the examples of labelled trajectories, where the blue dots are the start of lane change intention and the purple dots indicate the end point of lane change intention (start point of keep straight intention as well). The data points between start and end points are labelled as lane changing, others are labelled as keep straight.

To get the full use of the dataset, the sliding window method is employed to obtain samples from the dataset. As shown in Fig. 6 (a), depending on the label of the last time step, the entire sample will be labelled as turn left, turn right or keep straight. To determine the length of the sliding window, different window lengths are tested by using different models. To ensure fairness, all the test cases have employed the same input. As shown in Fig. 6 (b), most models have the best performance with a time window length of 15 time steps. Therefore, in the following experiments, the time window length of 15 time steps is employed.

Due to the nature of the trajectory dataset, the number of extracted samples in each class is inconsistent: most are keep straight, and the number of turn right is also different with turn

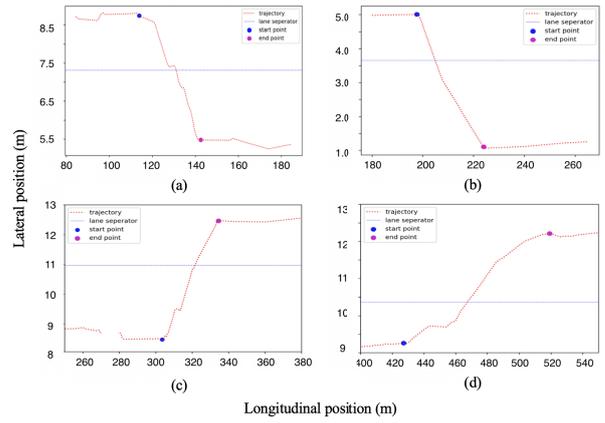


Fig. 5 Examples of labelled samples.

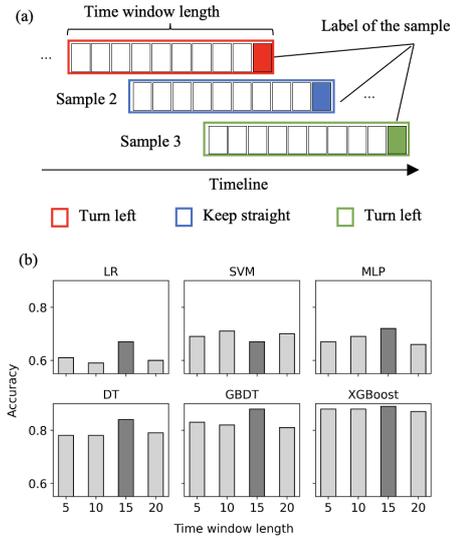


Fig. 6 Diagrammatic of samples.

Table 4 Detail of dataset.

No. of KS	No. of TR	No. of TL	Train test ratio
1,200	1,200	1,200	4 : 1

left. Such a sample imbalance problem will pose difficulty for driving intention inference [31]. To solve this problem, SMOTE, short for Synthetic Minority Oversampling TEchnique, is introduced to increase the pieces of minority samples and make a balance dataset. As an effective oversampling method, SMOTE can produce synthetic samples through interpolation. For a better understanding one can refer to literature [32]. The detail of the pre-processed dataset can be summarized in Table 4.

KS refers to keep straight, TL is turn to left and TR is turn to right. Overall, 2,880 pieces of samples are used to train the model and the rest 720 pieces are used for on-line inference and validation.

4.3 Experiment Results of FES-XGB Framework

To find the relationship between environment information and driving intention, at the same time, to prove the effectiveness of the proposed FES-XGB framework, experiments are conducted under five different input cases and three different models. Each of the five different input cases utilizes one or more kinds of input variables belonging to subject vehicle information, neighboring

vehicle information, local environment information, and global environment information. The detailed description of these four kinds of input variables can be found in Table 2. More specifically, case 1 only takes subject vehicle information as model input; case 2 takes subject and neighboring vehicle information as input; case 3 adds local environment information to case 2; case 4 adds global environment information to case 2; in case 5, all the decision variables including subject vehicle information, neighboring vehicle information, local and global traffic environment information are employed. The three different models include XGBoost, Feature Extraction based XGBoost (FE-XGB), as well as the proposed FES-XGB. As compared to FES-XGB model, the FE-XGB model, which has not employed the feature selection module, can be used to prove the effectiveness of feature selection operation.

The input dimensions of different models under different cases are presented in Table 5. From Table 5, it can be observed that the feature extraction and feature selection operations can greatly reduce the input dimensions of the intention inference models. In particular, when in case 5, the number of input features can be reduced from 750 to 15. With less input features, the efficiency and generalization abilities of the inference model are expected to be improved. At the same time, the inference accuracy is hoped not

Table 5 The input dimensions of different models.

Model \ Input	Case 1	Case 2	Case 3	Case 4	Case 5
XGBoost	30	570	660	660	750
FE-XGB	14	266	272	272	278
FES-XGB	13	13	13	15	15

to be reduced.

To simultaneously realize the high accuracy and efficiency, the input features should be carefully selected. In this study, the feature selection method named SHAP is introduced. The SHAP values of each feature are calculated: features with higher SHAP values are considered to have larger impacts on the model output, and those with lower SHAP values usually have minimal contributions. Therefore, selecting important features and pruning the unnecessary ones will produce a more efficient model and will not reduce the accuracy.

Taking the input case 5 as an example, Fig. 7 and Fig. 8 show the SHAP values of features and the corresponding feature selection process.

Figure 7 (a) displays the ranks of feature importance according to SHAP values, note that only the top 20 important features are shown here. It shows the overall contribution of each feature, as well as the feature’s impact on the model output. It is well illustrated that traffic environment information have relatively large impacts on the model output, together with the lateral distance between the subject vehicle and neighboring vehicles. Figure 7 (b)–(d) present which features are important when the model gives an output of keep straight, turn left, and turn right respectively. The colors of the feature indicate how changes in the feature value influence the model output.

Figure 8 illustrates how the inference accuracy and training time change when the number of input features gradually increases according to the SHAP values. It can be observed that with the increment of input features, the inference accuracy increases while the training efficiency decreases. The model can

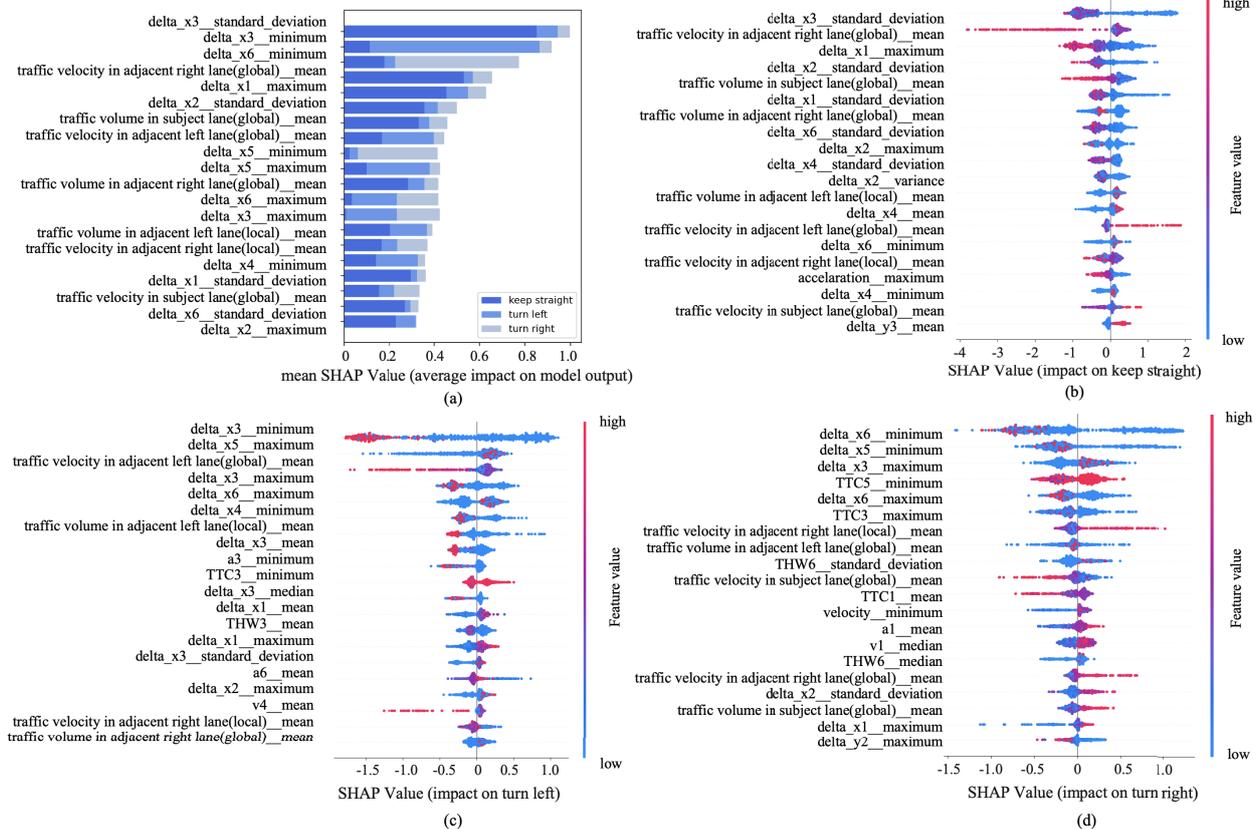


Fig. 7 Impact of important features on model output.

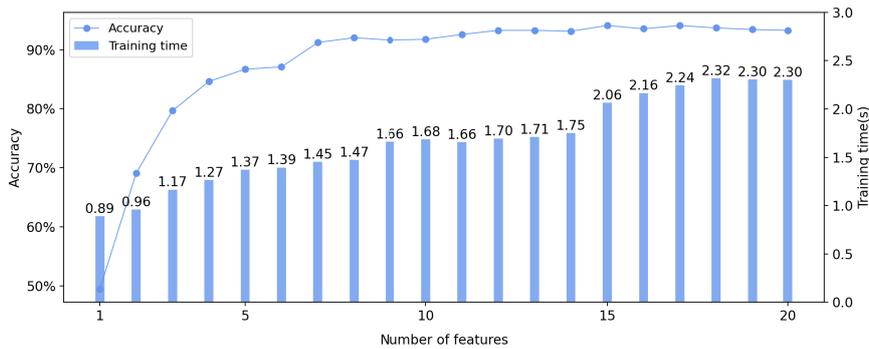


Fig. 8 Feature selection process.

Table 6 Comparison of different models with different input.

Case 1									
Method	XGBoost			FE-XGB			FES-XGB		
Criteria	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
KS	65.78%	66.28%	66.03%	73.28%	71.43%	72.34%	73.25%	74.79%	74.01%
TL	43.03%	47.16%	45.00%	68.80%	63.89%	66.26%	71.18%	64.68%	67.78%
TR	46.73%	42.02%	44.25%	69.85%	76.89%	73.20%	72.66%	78.15%	75.30%

Case 2									
Method	XGBoost			FE-XGB			FES-XGB		
Criteria	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
KS	93.70%	91.19%	92.43%	94.44%	92.86%	93.64%	93.70%	93.70%	93.70%
TL	85.71%	86.46%	86.09%	86.96%	87.30%	87.13%	85.99%	87.70%	86.84%
TR	88.48%	90.34%	89.40%	90.04%	91.18%	90.61%	87.98%	86.13%	87.05%

Case 3									
Method	XGBoost			FE-XGB			FES-XGB		
Criteria	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
KS	93.39%	91.95%	92.66%	97.06%	97.06%	97.06%	95.83%	96.64%	96.23%
TL	86.40%	86.03%	86.21%	90.48%	90.48%	90.48%	87.84%	88.89%	88.36%
TR	90.95%	92.86%	91.89%	92.86%	92.86%	92.86%	90.56%	88.66%	89.60%

Case 4									
Method	XGBoost			FE-XGB			FES-XGB		
Criteria	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
KS	93.23%	95.02%	94.12%	97.48%	97.48%	97.48%	97.81%	93.70%	95.71%
TL	87.17%	86.03%	86.59%	91.20%	90.48%	90.84%	88.33%	90.08%	89.19%
TR	91.10%	90.34%	90.72%	92.50%	93.28%	92.89%	90.12%	92.02%	91.06%

Case 5									
Method	XGBoost			FE-XGB			FES-XGB		
Criteria	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
KS	95.90%	98.47%	97.16%	97.90%	97.90%	97.90%	100.00%	97.06%	98.51%
TL	90.18%	88.21%	89.18%	92.00%	91.27%	91.63%	91.02%	92.46%	91.73%
TR	91.95%	91.18%	91.56%	92.92%	93.70%	93.31%	91.29%	92.44%	91.86%

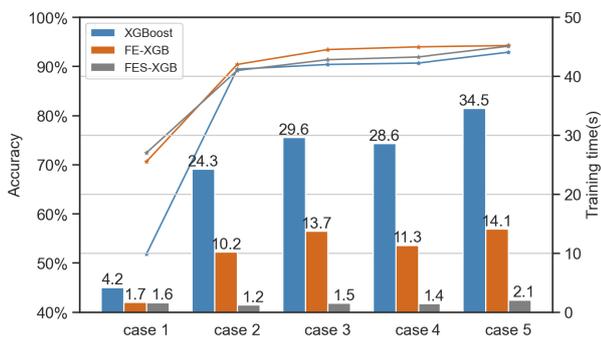


Fig. 9 Comparison on model accuracy and training time.

reach a satisfying prediction accuracy with only several important features as input. According to the changing trend in Fig. 8, the FES-XGB model is established by using the first 15 features as input.

Table 6 and Fig. 9 present the driving intention inference results of XGBoost, FE-XGB and FES-XGB with different kinds of input cases. From Table 6, it can be observed that for XGBoost,

FE-XGB and FES-XGB models, the Precision, Recall, F1-score of KS, TL, TR gradually increase when the model input changes from case 1 to case 5, illustrating that not only with the subject vehicle information, the driving intention is also connected with the neighboring vehicle information as well as the traffic environment information. By comparing the criteria in case 5 to that in case 2, it can be found that the environment information plays an important role in the intention inference. Therefore, it is necessary to take traffic environment into consideration when establishing driving intention inference models. Besides, it can be found that in each input cases, the proposed FES-XGB model outperforms the baseline XGBoost model, and is comparable with the FE-XGB model in each evaluation criterion.

Figure 9 displays the overall accuracy of each model in each input case, as well as the training time. The following phenomena can be observed from the results shown in Fig. 9: (1) The XGBoost, FE-XGB and FES-XGB models all have a better inference accuracy when provided with neighboring vehicle information and traffic environment information. However, the base-

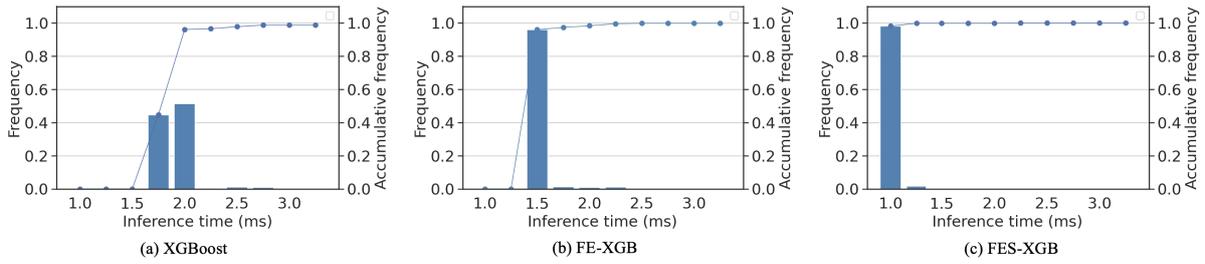


Fig. 10 The distribution of on-line inference time.

line XGBoost model has an obvious rise in training time when added these information, while the FE-XGB and FES-XGB models can keep a relatively stable training time, especially the proposed FES-XGB model, which can maintain a very low training time in all input cases. Under the condition of case 5, the proposed FES-XGB model can save around 95% of training time as compared to the baseline XGBoost model. (2) Although the FES-XGB model is low in computation cost, the model accuracy is not inferior to the FE-XGB model. In different input cases, the performances of the FES-XGB model outperform the corresponding baseline XGBoost model.

Compared to the efficiency of the off-line training, the on-line inference efficiency is more important for intention inference models. By recording the inference time of each sample in the test dataset respectively, Fig. 10 shows the distribution of intention inference time of the XGBoost, FE-XGB and FES-XGB models respectively. From Fig. 10, it can be observed that when given the subject vehicle information, neighboring vehicle information and environment information, the proposed FES-XGB model needs 1 to 1.25 milliseconds to output the driving decision. While under the same input condition, the baseline XGBoost and FE-XGB models need 1.75 to 2.25 milliseconds and 1.5 to 1.75 milliseconds respectively. At the same time, it can be found the proposed FES-XGB model is more stable and reliable regarding on-line inference speed, i.e., compared with the baseline XGBoost model and the FE-XGB model, the on-line inference time of the FES-XGB model is strictly within 1.5 milliseconds.

In addition to the accurate and real-time inference, the FES-XGB model should predict the driving intention as earlier as possible. By defining the anticipation time as the driving time interval from the current position to the lane change or lane keep decision point, the original testing samples are divided into six subintervals, indicating anticipation time of more than 3 s, 2.5 s, 2 s, 1.5 s, 1 s, and 0.5 s respectively. The samples of each subinterval are employed to test the intention prediction ability of the proposed FES-XGB model. Besides, the XGBoost model without environment input, the XGBoost model with environment input and the FE-XGB model with environment input, are also tested to be compared with the FES-XGB model. The prediction performances of the four models under different anticipation time are shown in Fig. 11.

Figure 11 reveals the intention prediction accuracy increases when the vehicle is approaching the lane change or lane keep decision point. By comparing the XGBoost models without and with the environment information input, it can be observed that with the environment information, the model can give a better in-

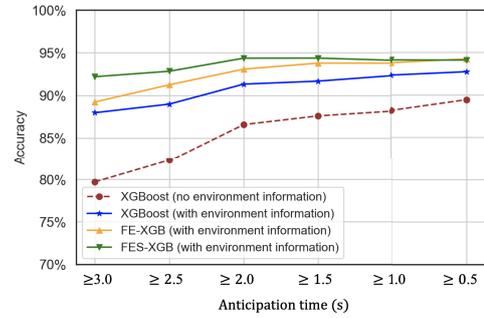


Fig. 11 Prediction accuracy under different anticipation time.

tion prediction performance in different anticipation periods. The reason is that the traffic environment information can help the vehicle get aware of the traffic states from the view of the subject vehicle, to better predict its driving intention. By comparing the XGBoost, FE-XGB and FES-XGB models with the environment information as input, it can be proved that the proposed FES-XGB model has the best performance in terms of prediction accuracy and anticipation time. In the time point three seconds before the lane change point, the FES-XGB model is capable to predict the driving intention with the accuracy of more than 90%.

4.4 Analysis on Environment Information

In this section, concerning the influences of the traffic environment information, we further analyze the performance of the proposed FES-XGB framework. There are mainly three perspectives investigated, including the delay of the environment information, the perception range of the environment information and the category of the environment information.

4.4.1 Delay of Environmental Information

In the previous discussions, we have assumed the traffic environment information coming from V2V or V2I communication can be obtained in real-time. However, in real traffic environment, the information transmission might be influenced by many factors, such as the block from architectures or heavy trucks, or the delay of wireless communication itself. It is usually hard and expensive to obtain high fidelity traffic information in real time. Therefore, analyzing how the communication delays influence the driving intention inference accuracy of the proposed FES-XGB framework is necessary and practical.

Three kinds of environment information delay are tested in the experiments, including: (1) delay of global environment information, where the local environment information are assumed to be obtained in real time; (2) delay of local environment information, where the global environment information are assumed to

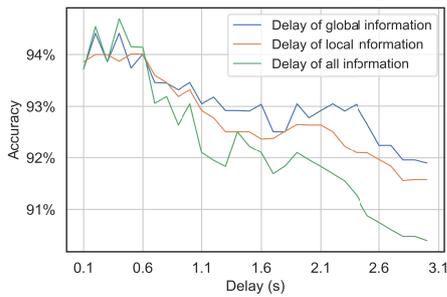


Fig. 12 Influence of environment information delay on intention inference.

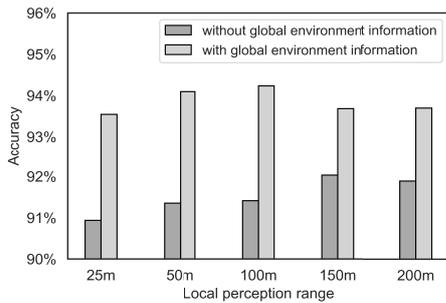


Fig. 13 Influence of local perception range on intention inference.

be obtained in real time; (3) Delay of all environment information, where neither global nor local environment information can be obtained in real time.

Figure 12 shows how the intention inference accuracy of the FES-XGB framework changes with three different environment information delays. The delay is designed to increase from 0.1 to 3 seconds with a step of 0.1 seconds. From Fig. 12, it can be observed that with the increase of the environment information delay, the intention inference accuracies of the FES-XGB framework under three delay conditions all show a decrease trend. Besides, compared with the condition where both global and local environment information are delayed, the inference accuracy will suffer from less reduction if the framework can obtain one kind of environment information in real-time (global or local). Overall, the FES-XGB framework can maintain an intention inference accuracy higher than 90% when the traffic environment information delay is within three seconds.

4.4.2 Perception Range of Environment Information

Another important influential factor on the inference accuracy is the perception range of environmental information. As the global environment information is the lane-wise traffic volume and velocity in the entire research area, its range has already been determined. In this regard, we only research the influence from the range of the local environmental information, i.e., the length of the concerned area centered by the subject vehicle.

Two kinds of conditions are involved in the experiments. In one condition, the global environment information is available to the subject vehicle, above which the local environment information in a different range is provided. In another condition, no global environment information can be obtained and the subject vehicle can only access the traffic environment information in a local area.

Figure 13 displays the FES-XGB performances with different local perception range. By comparing the intention inference ac-

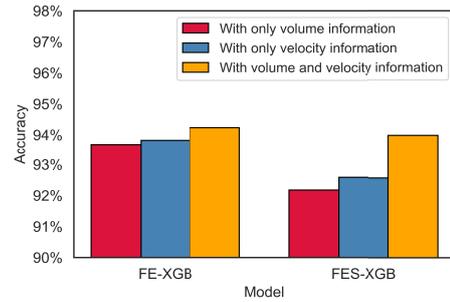


Fig. 14 Influence of traffic environment information category on intention inference.

curacy with and without global environment information, it can be observed that the FES-XGB framework has higher accuracies over all the local perception ranges when the additional global environment information is available. Besides, with global environment information provided, the FES-XGB framework can give the best performance with a shorter local traffic perception need: 100 meters centered by the subject vehicle is sufficient. On the other hand, when the global environment information is not available, the intention inference accuracy shows an increasing trend with the increase of the local perception range. In other words, without the global environment information, the FES-XGB framework needs longer local perception range to give accurate intention inference results.

4.4.3 Category of Environmental Information

As the lane-wise traffic volume and velocity are both employed as environment information input in the previous discussions, we further invest the influence of single traffic volume or velocity on the performance of intention inference frameworks respectively.

The FE-XGB and FES-XGB frameworks are tested over three conditions, i.e., provided with only traffic volume information, provided with only traffic velocity information and provided with both volume and velocity information. The experiment results are shown in **Fig. 14**.

From Fig. 14, it can be observed that for both FE-XGB and FES-XGB frameworks, traffic environment information containing both volume and velocity information will lead to better intention inference results. When providing only one category of traffic environment information, the velocity is slightly more preferable according to the experiment results. This result is consistent with the data characteristics of the employed trajectory data extracted from video records: the traffic volume will be large when in both free flow and congestion flow, while the traffic velocity can better represent the traffic states in these conditions.

4.5 Analysis on Framework Generalization and Sensitivity

The experiments results in previous discussions have proved the effectiveness of the FES-XGB framework and the importance of traffic environment information. In this subsection, we further discuss the generalization ability of the framework and conduct sensitivity analysis on the framework subcomponent.

To test the generalization ability of the framework, experiments are designed by combining the proposed framework with multiple ML models, including logistic regression (LR), support vector machine (SVM), multiple layer perception (MLP), decision tree

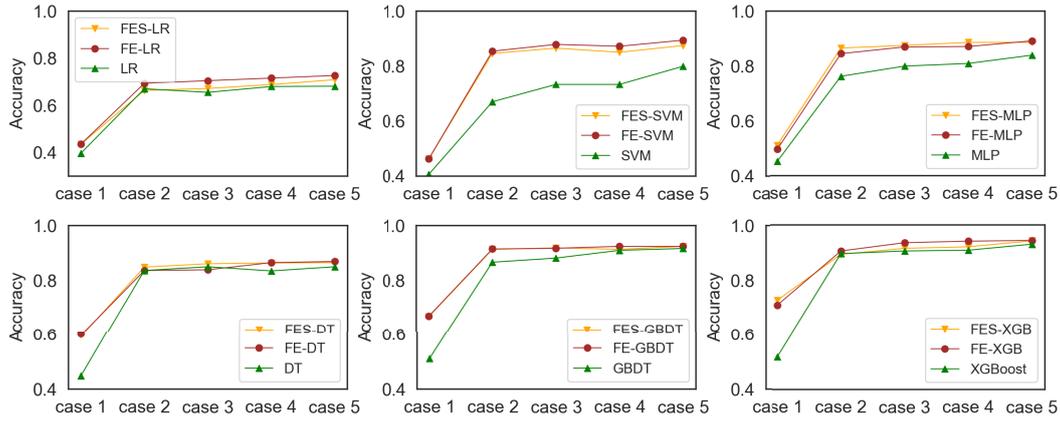


Fig. 15 Comparison over multiple ML models.

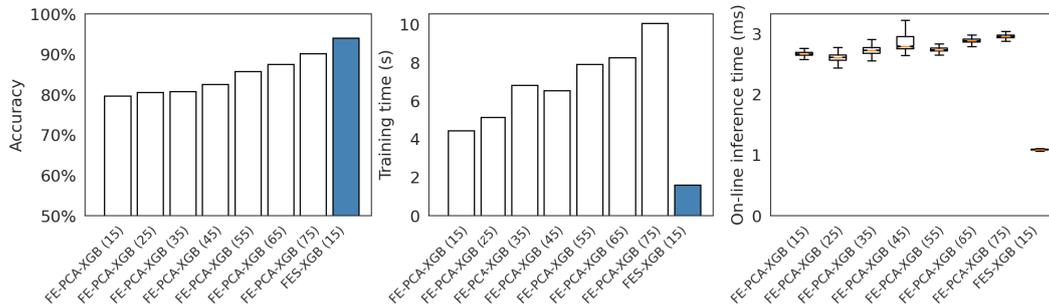


Fig. 16 Comparison between FE-PCA-XGB and FES-XGB.

(DT) and gradient boosting decision tree (GBDT). All the comparison experiments are conducted with five input cases (case 1 to case 5). The experiments results are shown in Fig. 15.

As can be observed in Fig. 15, compared with the plain ML models, the ones combined with the FE and FES operations have better intention inference performances, proving the good generalization ability of the FE and FES based framework. At the same time, the traffic environment information is found essential in the inputs of multiple models, as the intention inference accuracy gets improved when the environment information is provided in most of the test cases.

To illustrate the superiority and indispensability of the proposed FES-XGB framework, sensitivity analysis on the framework subcomponent is conducted. The SHAP feature selection module is an important subcomponent of the FES-XGB framework and it is mainly used to reduce the feature dimensions. Since principal component analysis (PCA) can also realize dimension reduction, we establish the FE-PCA-XGB framework by replacing SHAP module with PCA. The FE-PCA-XGB framework is compared with the proposed FES-XGB framework. The comparison result is shown in Fig. 16.

As can be seen in Fig. 16, when employing the same number of features (the top 15) as input, the FES-XGB framework outperforms FE-PCA-XGB in terms of intention inference accuracy, training efficiency, and on-line inference speed. With more features employed, the accuracy of FE-PCA-XGB will increase, but the training and inference time cost increase as well. Compared with the FE-PCA-XGB, the FES-XGB framework using SHAP subcomponent has higher accuracy and inference speed, as well as lower training cost. Therefore, the FES-XGB framework is

more applicable in performing driving intention inference in real traffic environment.

5. Conclusion

In this study, the FES-XGB framework combining feature extraction, feature selection and XGBoost algorithm is proposed to infer driving intention in the V2V and V2I background. Considering the emerging communication and connected vehicle technology, the relationship between the traffic environment information and driving intention is discussed.

The NGSIM dataset is employed to validate the proposed framework and learn the relationship between the environment information and driving intention. When adding the traffic environment information into input, the intention inference accuracy of the conventional XGBoost model can be improved from 89.42% to 92.86%, but the training time also has an increment of 29.56% at the same time. To strike a delicate balance between accuracy and efficiency, the feature extraction and selection operations are introduced and the FES-XGB framework is established. It is found that the proposed framework can achieve a high inference accuracy and low computation cost simultaneously. With the environment information as input, the intention inference accuracy of the FES-XGB framework can further reach 94.09%, at the same time, the training and on-line inference costs are 94.03% and 65.25% smaller than the conventional XGBoost model as well.

Moreover, influences coming from the transmission delay, perception range, and category of the traffic environment information have been analyzed. The following findings can be obtained: (1) When the delay of the traffic environment information is smaller

than three seconds, the FES-XGB framework can maintain an intention inference accuracy higher than 90%. (2) When the global environment information is provided, a relatively short local traffic perception range is sufficient for the FES-XGB framework to give accurate intention inference results. (3) When both traffic volume and velocity information are provided to the FE-XGB and FES-XGB frameworks, the intention inference accuracy will be higher than those provided with a single kind of information.

The effectiveness of the environment information and the FES-XGB framework structure has been tested employing multiple ML models. As they can universally increase the accuracy, the environment information and the structure of the framework can be applied to a wide range of areas, such as the design of ADAS and autonomous driving.

The limitation of this study is that the field data employed comes from highway segments, which only includes lane changing and lane-keeping maneuvers. By using an enhanced driving dictionary containing more driving interacting scenarios including U-turn, intersection, lane-drop and others, a pervasive model could be trained which can be applied in more general scenarios. The data obtained from new technology like autonomous vehicles, in-vehicle communication devices and cameras can be used to validate the proposed intention detection method as well. Future studies will also focus on digging deeper into the influential factors of driving behaviors and providing a generalized means to recognize and predict driving intention.

Acknowledgments This work was supported by JSPS KAKENHI Grant Number JP19H02377. Shuo Wang also wants to thank China Scholarship Council for providing financial support to study in Japan.

References

- [1] Kyriakidis, M., Happee, R. and de Winter, J.C.F.: Public opinion on automated driving: Results of an international questionnaire among 5000 respondents, *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol.32, pp.127–140 (2015).
- [2] Duffield, T.J. and Krupenia, S.: Drivers' Interaction Preferences in Autonomous Vehicle Multimodal Interactive Systems, *Proc. Human Factors and Ergonomics Society Annual Meeting*, Vol.59, No.1 (2015).
- [3] Bellet, T. et al.: From semi to fully autonomous vehicles: New emerging risks and ethico-legal challenges for human-machine interactions, *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol.63, pp.153–164 (2019).
- [4] Favaro, F.M. et al.: Examining accident reports involving autonomous vehicles in California, *PLoS One*, Vol.12, No.9, e0184952 (2017).
- [5] Goñi-Ros, B. et al.: Using advanced adaptive cruise control systems to reduce congestion at sags: An evaluation based on microscopic traffic simulation, *Transportation Research Part C: Emerging Technologies*, Vol.102, pp.411–426 (2019).
- [6] Li, K. et al.: Lane changing intention recognition based on speech recognition models, *Transportation Research Part C: Emerging Technologies*, Vol.69, pp.497–514 (2016).
- [7] Benloucif, A. et al.: Cooperative trajectory planning for haptic shared control between driver and automation in highway driving, *IEEE Trans. Industrial Electronics*, Vol.66, No.12, pp.9846–9857 (2019).
- [8] Okuda, R., Kajiwar, Y. and Terashima, K.: A survey of technical trend of ADAS and autonomous driving, Technical Papers of 2014 International Symposium on VLSI Design, Automation and Test, IEEE (2014).
- [9] Song, W. et al.: Lane detection and classification for forward collision warning system based on stereo vision, *IEEE Sensors Journal*, Vol.18, No.12, pp.5151–5163 (2018).
- [10] Rahman, M.S. et al.: Safety benefits of arterials' crash risk under connected and automated vehicles, *Transportation Research Part C: Emerging Technologies*, Vol.100, pp.354–371 (2019).
- [11] Rahman, M.S. et al.: Understanding the highway safety benefits of different approaches of connected vehicles in reduced visibility conditions, *Transportation Research Record*, Vol.2672, No.19, pp.91–101 (2018).
- [12] Kim, I.-H. et al.: Detection of braking intention in diverse situations during simulated driving based on EEG feature combination, *Journal of Neural Engineering*, Vol.12, No.1, 016001 (2014).
- [13] Wang, H. et al.: An EEG-Based Multi-Classification Method of Braking Intentions for Driver-Vehicle Interaction, *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, IEEE (2019).
- [14] Chapman, P.R. and Underwood, G.: Visual search of driving situations: Danger and experience, *Perception*, Vol.27, No.8, pp.951–964 (1998).
- [15] Doshi, A. and Trivedi, M.: Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions, *2009 IEEE Intelligent Vehicles Symposium*, IEEE (2009).
- [16] Doshi, A. and Trivedi, M.M.: On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes, *IEEE Trans. Intelligent Transportation Systems*, Vol.10, No.3, pp.453–462 (2009).
- [17] Kim, I.-H. et al.: Prediction of driver's intention of lane change by augmenting sensor information using machine learning techniques, *Sensors*, Vol.17, No.6, 1350 (2017).
- [18] Xing, Y. et al.: An ensemble deep learning approach for driver lane change intention inference, *Transportation Research Part C: Emerging Technologies*, Vol.115, 102615 (2020).
- [19] Liu, Q. et al.: Early recognition of driving intention for lane change based on recurrent hidden semi-Markov model, *IEEE Trans. Vehicular Technology*, Vol.69, No.10, pp.10545–10557 (2020).
- [20] Li, G. et al.: Lane change maneuver recognition via vehicle state and driver operation signals—Results from naturalistic driving data, *2015 IEEE Intelligent Vehicles Symposium (IV)*, IEEE (2015).
- [21] Dong, C., Zhang, Y. and Dolan, J.M.: Lane-change social behavior generator for autonomous driving car by non-parametric regression in reproducing kernel hilbert space, *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE (2017).
- [22] Oh, C., Choi, J. and Park, S.: In-depth understanding of lane changing interactions for in-vehicle driving assistance systems, *International Journal of Automotive Technology*, Vol.18, No.2, pp.357–363 (2017).
- [23] Hou, Y., Edara, P. and Sun, C.: Modeling mandatory lane changing using Bayes classifier and decision trees, *IEEE Trans. Intelligent Transportation Systems*, Vol.15, No.2, pp.647–655 (2013).
- [24] Tang, L. et al.: Driver Lane Change Intention Recognition of Intelligent Vehicle Based on Long Short-Term Memory Network, *IEEE Access*, Vol.8, pp.136898–136905 (2020).
- [25] Mahajan, V., Katrakazas, C. and Antoniou, C.: Prediction of Lane-Changing Maneuvers with Automatic Labeling and Deep Learning, *Transportation Research Record*, 0361198120922210 (2020).
- [26] Zhang, H. and Fu, R.: A Hybrid Approach for Turning Intention Prediction Based on Time Series Forecasting and Deep Learning, *Sensors*, Vol.20, No.17, 4887 (2020).
- [27] Zhang, Y. et al.: Lane-change intention estimation for car-following control in autonomous driving, *IEEE Trans. Intelligent Vehicles*, Vol.3, No.3, pp.276–286 (2018).
- [28] Das, A., Ahmed, M.M. and Ghasemzadeh, A.: Using trajectory-level SHRP2 naturalistic driving data for investigating driver lane-keeping ability in fog: An association rules mining approach, *Accident Analysis & Prevention*, Vol.129, pp.250–262 (2019).
- [29] Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016).
- [30] Wolf, P. et al.: Adaptive behavior generation for autonomous driving using deep reinforcement learning with compact semantic states, *2018 IEEE Intelligent Vehicles Symposium (IV)*, IEEE (2018).
- [31] Krawczyk, B.: Learning from imbalanced data: Open challenges and future directions, *Progress in Artificial Intelligence*, Vol.5, No.4, pp.221–232 (2016).
- [32] Chawla, N.V. et al.: SMOTE: Synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, Vol.16, pp.321–357 (2002).
- [33] Lundberg, S.M. et al.: From local explanations to global understanding with explainable AI for trees, *Nature Machine Intelligence*, Vol.2, No.1, pp.2522–5839 (2020).
- [34] Lundberg, S.M. and Lee, S.-I.: A unified approach to interpreting model predictions, *Proc. 31st International Conference on Neural Information Processing Systems* (2017).
- [35] Friedman, J.H.: Greedy function approximation: A gradient boosting machine, *Annals of Statistics*, pp.1189–1232 (2001).



Shuo Wang received her B.E. degree in transportation engineering from China University of Mining and Technology, Jiangsu, China, in 2016, and her M.S. degree in transportation planning and management from Beijing Jiaotong University, Beijing, China, in 2019. She is currently pursuing her Ph.D. degree in the

School of Engineering, the University of Tokyo. Her current research interests include traffic simulation, connected and automated vehicles, traffic flow analysis, and intelligent transportation systems.



Hideki Fujii received his Ph.D. degree in Environmental Studies from the Graduate School of Frontier Sciences, the University of Tokyo, in 2009. He is an Associate Professor at the Graduate School of Engineering, the University of Tokyo. He has been working on multi-agent-based traffic and social system simulation. His profes-

sional membership includes the World Conference on Transportation Research Society, the Japan Society of Traffic Engineers, the Japanese Society for Artificial Intelligence, etc.



Shinobu Yoshimura is currently Vice President and Professor, School of Engineering, the University of Tokyo (UTokyo). In 1987, he completed his studies in School of Engineering, UTokyo (specialized in nuclear engineering) as Doctor of Engineering. He became Lecturer, Associate Professor of School

of Engineering in UTokyo, then Professor in Graduate School of Frontier Sciences, and since 2005, Professor of School of Engineering. His specialties are High-performance and Intelligent Computational Mechanics with Real World's Applications. He serves as a Vice President of IACM and the President of APACM. He has received numerous awards, including ICCES Distinguished Achievement Medal (2015), IACM Fellow Award (2014), APACM Computational Mechanics Award (2013), AIAA Liquid Propulsion Best Paper Award (2009), IEEE/ACM Supercomputing 06 Gordon Bell Award finalist (2006).