Fine-Grained Driving Behavior Prediction via Context-Aware Multi-Task Inverse Reinforcement Learning

Kentaro Nishi[1] and Masamichi Shimosaka[2]

***Abstract*— Research on advanced driver assistance systems for reducing risks to vulnerable road users (VRUs) has recently gained popularity because the traffic accident reduction rate for VRUs is still small. Dealing with unexpected VRU movements on residential roads requires proficient acceleration and deceleration. Although fine-grained prediction of driving behavior through inverse reinforcement learning (IRL) has been reported with promising results in recent years, learning of a precise model fails when driving strategies vary with contextual factors, i.e., weather, time of day, road width, and traffic direction. In this work, we propose a novel multi-task IRL approach with a multilinear reward function to incorporate contextual information into the model. This approach can provide precise long-term prediction of fine-grained driving behavior while adjusting to context. Experimental results using actual driving data over 141 km with various contexts and roads confirm the success of this approach in terms of predicting defensive driving strategy even in unknown situations.**

I. Introduction

Thanks to recent advances in driver assistance systems and other vehicle technologies, the number of traffic accidents is decreasing. In particular, the number of in-vehicle driver deaths has fallen to a remarkable value with the worldwide popularization of improved seatbelt and air bag systems. However, the rate of traffic accident reduction has slowed over the past few years and the rate of critical pedestrian injuries and deaths is increasing. In Japan, the proportion of traffic deaths excluding in-vehicle drivers has become more than two-thirds of total traffic deaths, which is an alarming rate [2]. Also, the rate of accident reduction remains low on narrow streets in residential areas (residential roads) [15]. Therefore, a key issue is how to develop a next-generation safety system designed to reduce risks to vulnerable road users (VRUs) on residential roads.

Although there is recent progress in automatic emergency braking (AEB), these systems are not sufficient to save VRUs on residential roads because of certain limitations such as short braking distances and sensing accuracy. V2X communication [18] and early prediction of pedestrian crossing [8] are examples of VRU-centered active safety systems on residential roads. Those approaches, however, are currently difficult to implement because they require new infrastructures and devices and long-distance pedestrian detection.

Therefore, we suggest a challenging new approach to predict safe driving strategy for VRU-centered risk reduction. For the safety of VRUs on residential roads, proficient acceleration and deceleration with risk anticipation are important. This type of *fine-grained* driving approach based on superior knowledge and experience is called “risk anticipation and defensive driving” [26]. Predicting fine-grained driving behavior, specifically long-term acceleration and deceleration, would allow us to develop a novel, VRU-centered advanced driving assistance system (ADAS) for residential roads.

Long-term activity prediction using inverse reinforcement learning (IRL) has gained popularity in recent years [16]. It has been proved suitable also for fine-grained driving behavior modeling [23]. In modeling risk anticipation and defensive driving, it is important to account for “context” because driving strategy is likely to vary with contextual factors, i.e., weather, time of day, road width, and traffic direction (one- or two-way). To improve performance over such a variety of contexts, we have to design a model that can adjust to specific situations, rather than an averaged driving behavior model. In addition, more flexible, robust prediction could likely be achieved by sharing parameters when data have only some mutual contextual factors, and not only when data have exactly the same context, as illustrated in Figure 1. Therefore, prediction technology for *fine-grained*, *long-term* driving behavior according to *context* could be a key to developing a next-generation VRU-centered ADAS. Precise prediction while adjusting to context allows us to develop next stage of ADAS, where both unneeded alarms and risk oversights are avoided.

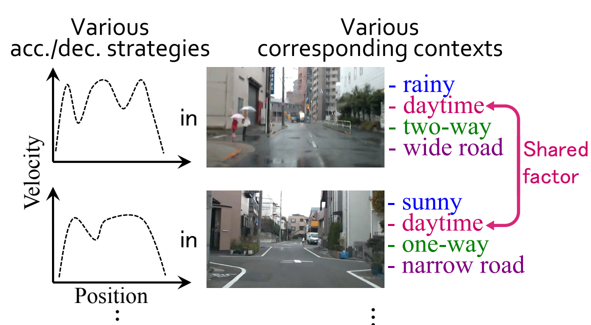


Fig. 1: Contextual driving data set and the concept of applying shared contextual factors. More flexible, robust prediction can be achieved by sharing parts of the same contexts when data contains such mutual factors.

In this paper, we present an new framework for predicting long-term acceleration and deceleration by learning different driving strategies adapting to various contexts. Since the model is based on multi-task learning, it can provide precise prediction even in rare or unknown contexts by using

[1]Kentaro Nishi with a former master course graduate student of the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo

[2]Masamichi Shimosaka with the faculty of Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan simosaka@miubiq.cs.titech.ac.jp

contextual information. Moreover, since our framework is implemented on a mobile phone connected to a vehicle's network to capture velocity and steering information, it has the potential to support powerful ubiquitous applications. We thus tackle with challenges of *fine-grained*, *long-term*, *contextual* driving behavior modeling for a next-generation ubiquitous ADAS.

Our contributions are threefold: (i) We propose a model based on multi-task IRL with a multilinear reward function in order to flexibly learn changing driver behaviors in various contexts. This model can achieve long-term, fine-grained driving behavior prediction even in unknown contexts. (ii) We demonstrate the effectiveness of the proposed approach through an experimental evaluation on real data captured from driving on 141 km of residential roads in various situations, such as day or night and sunny or rainy weather. (iii) We have developed a smartphone-based online driving behavior prediction system and driving skill scoring system to illustrate the potential of various future ubiquitous applications.

A. Related work

1) Driver behavior analysis: For developing ADAS, research on using sensed data such as GPS and velocity data, has gained popularity in recent years [24]. In the ubiquitous computing community, as well, automotive research is a hot topic, including innovative applications for navigation and recommendation systems based on real data [30], [14], driving activity monitoring and privacy protection [11], urban analysis using on-taxi GPS data [31], [29], and taxi allocation and route planning [28]. Most such research, however, does not consider the safety of VRUs on residential roads because of the difficulty of dealing with a variety of environmental factors such as mixing of pedestrians and vehicles.

Also, driving behavior modeling has become one of the most active research topics in active safety. Hermes *et al.* [13] proposed a history-based approach using past information to predict vehicle location and movement within a couple of seconds. Liebner *et al.* [17] leveraged an intelligent driver model, which is an explicit model simulating driver activities based on current velocity and inter-vehicle distance, to classify the behaviors of going straight, stopping, and turning. Neither the history-based approach nor direct use of an explicit model is suitable, however, in the case of an unknown environment.

In contrast, Gindele *et al.* [12] predicted vehicle trajectories and overtaking phenomena by means of dynamic Bayesian networks, and Berndt *et al.* [7] achieved lane change and turning prediction with a hidden Markov model (HMM) based on controller area network (CAN) data. Ziebart *et al.* [33] tried to model route preferences from taxi-mounted GPS data. They could predict route selection from requirements including safety, fuel efficiency, and duration, in contexts with varying time of day, speed limit, and traffic congestion. These conventional statistical driving behavior models frequently focus on macroscopic behaviors, such as turning at an intersection, overtaking, and early route prediction, as driving intentions. To develop a VRU-centered safety system for residential roads, however, will certainly require prediction of more fine-grained and long-term driving activity.

Long-term, fine-grained prediction of driving behavior on residential roads within a few tens of seconds by applying an IRL approach has been reported with promising results in recent years [23]. Although their framework deals with latent risks on roads, such as blind corners at intersections, it does not consider the diversity of driving behaviors, which are likely to change in different driving contexts. Instead, they only applied their approach to data with limited contexts: their experimental data is from only four courses driven in the daytime. Although that work [23] is beneficial in that it demonstrates the feasibility of long-term prediction of fine-grained driving behavior, it is far from practical for use in active safety applications.

2) Multi-task learning: To enhance performance in a variety of contexts, we have to design a model to adjust to specific contexts rather than simply an averaged driver behavior model. In our domain, learning of a precise model adjusting to a given context can be considered as a task. The number of tasks (that is, the variety of contexts) increases exponentially as we account for more contextual factors. Having too many tasks can decrease prediction accuracy because of the shortage of training data for each task. To solve this problem, multi-task learning techniques have recently been gaining attention in the machine learning field. Even though the multi-task learning is actively explored for the last decade, The main issue to be solved here is to combine multi-task learning and IRL problem.

As a great first step of multi-task learning in IRL, clustering-based multi-task learning approaches based on non-parametric Bayes [6], [27] have been proposed. In these approaches, tasks are divided into several clusters, and a model is learned for each cluster. Although these approaches enable parameter sharing among tasks, they do not make use of contextual information explicitly. Furthermore, this approach could not infer driving behavior precisely for unkown roads.

Multilinear model also employed in our model, in which the weight parameter is represented as a tensor, have recently gained attention in the machine learning field as a way to introduce external factors explicitly into the model [25], [22]. However, the application of multilinear models have been limited to simple regression and classification problems. Particulaly, they have not been considered for problems of structured prediction, such as IRL.

The study of multi-task learning in an IRL framework is thus still in its early development, and applications are few in this field. Although several approaches have been reported [10], [9], [5], no context-aware IRL framework has yet been discussed. In the following section, we therefore examine multi-task learning within the framework of IRL.

II. Problem setting and the base model

A. Modeling target

As mentioned in the introduction, we focus on "risk anticipation and defensive driving," key ideas for ensuring safety on residential roads. To this end, our modeling target is decision making for acceleration and deceleration. We formulate this problem setting under the assumption that the driving route is known in advance, that is, we are not concerned with route planning.

Under this assumption, we model driving behavior on each road segment which starts with a turn or stop and ends with the next turn or stop. In this paper we consider the driving behavior of acceleration and deceleration in such a segment as a path planning problem.

B. Driving behavior modeling with IRL

In this work, we model the path planning problem in position-velocity space as a Markov decision process (MDP) [20]. Since an MDP is a goal-oriented path planning framework with action selection, we can model a driver's behavior in terms of decision making while looking ahead to a goal. We define states and actions here in the same manner as in [23]. We represent state s in terms of position x and velocity v as $s = (x, v)$. Both the state s and action a are discretized in an adequate manner (described in detail in the experimental section of the paper). We represent the dynamics of driving behaviors with discrete states and actions, defining a state transition probability $P(s'|s, a)$. Given a reward function $R(s)$, the driver is assumed to drive so as to maximize total the obtained reward, incorporating future rewards.

With the maximum entropy principle in [32], the likelihood for state sequence $\zeta = \{(s_0, a_0), (s_1, a_1), ...\}$ is represented as

$$P(\zeta) \propto \exp\left(\sum_t (R(s) + \log P(s_{t+1}|s_t, a_t))\right). \quad (1)$$

When the MDP is deterministic, the likelihood is formulated as $p(\zeta) \propto \exp\left(\sum_t R(s)\right)$.

In an MDP, given initial states $P(s_0)$, transition probability $P(s'|s, a)$, and reward function $R(s)$, the distribution of a driver's maneuvers can be predicted through value iteration [32]. As a result of the iteration, the policy $\pi(a|s)$ is computed as $\pi(a|s) = \exp(Q^{\text{soft}}(s, a) - V^{\text{soft}}(s))$.

Once $\pi(a|s)$ is computed, the predicted state visitation distribution $D(s)$ is calculated via repetition of the state transition $s \to s'$ from the initial states according to $\pi(a|s)$.

Due to the difficulty in designing proper reward function $R(s)$, it is natural to use expert driving behavior data to obtain the proper reward function [19], [3]. In single task IRL, the reward function $R(s)$ parameterized by $\boldsymbol{\theta}$ with a feature vector. We then optimize $\boldsymbol{\theta}$ by minimizing the negative log likelihood with regularization term $\Omega(\boldsymbol{\theta})$ as $\boldsymbol{\theta}^* = \operatorname{argmin}_{\boldsymbol{\theta}} - \ln p(\boldsymbol{\theta}) + \Omega(\boldsymbol{\theta})$.

This simple IRL could be useful when the reward is consistent over road segments in each state; however, for handling contextual driving behavior where deceleration trends changes across road segments (See Fig. 1), it is not feasible for us to use reward function parameterized by single $\boldsymbol{\theta}$. In other words, the key issue here is to obtain a systematic design principle to adjust $\boldsymbol{\theta}$ across road segments via different contextual information such as road width, weather, and time zones.

III. Multi-task inverse reinforcement learning adapting to variable contexts

A. Multi-task model

If we ignore differences in driving strategy among contexts, a driving behavior model will predict an average strategy, giving less-accurate prediction despite context and strategy variations. However, it is not straight forward to incorporate such variation into a model. Such variation occurs in many situations and affects driving strategies in many ways, so as we account for more factors in the model, the number of tasks (that is, the variety of contexts) increases exponentially. Having too many tasks can decrease prediction accuracy because of the shortage of training data for each task. *Multi-task learning* is a promising approach to solve this problem. Given several tasks, absence of training data corresponding to specific contexts might result from the increase in the number of tasks. In such cases, it is desirable to use knowledge learned from other tasks having certain context factors in common with the target task.

One of the major approaches in multi-task IRL is based on cluster based IRL [21], [10]. Unfortunately, despite the literature on recent advances for multi-task learning in an IRL setting, none of the reported models is suitable for our application. This is because these approaches perform task clustering based on similarity among driving strategies, with learning of reward functions for the various clusters. For enhancing ubiquity in driver behavior prediction, it is natural to employ contextual information directly into reward function design instead of handling behavior similarity across roads. In other words, the application discussed in the paper requires to predict expert driving behavior from driving by a general driver in unknown situations. Therefore, multi-task IRL with a clustering-based approach is not appropriate in our case.

For this reason, we need a new multi-task IRL approach explicitly leveraging external information to predict appropriately for all tasks Moreover, the approach must enable prediction in unknown contexts by using context factors, even if no training data is available for the target task.

B. Multi-task IRL with multilinear reward function

In IRL, it is common to formulate the reward function with a linear model–that is, the reward function is represented as the inner product of the feature vector and the weight parameter [33]. To predict acceleration and deceleration, we must use features related to the positions of acceleration and deceleration according to road configuration factors (e.g., intersection and stop positions). These features depend on state s. In addition, we use a feature vector encompassing various context factors (e.g., weather, time of day, road width, and traffic direction) to represent the differences in driving strategy among contexts. Note that the context-related features do not depend on s. Let us consider a naive approach to incorporate additional feature vectors in a reward function. In a linear model, it would be common to append new features to the existing feature vector. In terms of maximizing the total reward in an MDP framework, however, the optimal path remains unaffected by the new context-related features, since they do not depend on s. Therefore, a linear reward function would not meet our requirement.

As an alternative to the linear formulation, we introduce a multilinear reward function. We define a "context" as a combination of multiple contextual factors. We assume that the model accounts for E types of factors and each factor is represented as $\boldsymbol{c}_e$ $(e = 1, ..., E)$ in a "1-of-K" coding scheme (i.e., a vector with one element 1 and all other elements 0). We extend the vector $\boldsymbol{c}_e$ as $\boldsymbol{c}'_e = [1, \boldsymbol{c}_e^\top]^\top$. This extension helps with sharing parameters when certain context factors match among tasks.

The weight parameter is represented as a tensor of order $E+1$, but we can formulate the context as a vector $\boldsymbol{d}$ by using the Kronecker product: $\boldsymbol{d} = \boldsymbol{c}'_1 \otimes ... \otimes \boldsymbol{c}'_E$. Thereby we obtain a simple form for the reward function as a bilinear model of the feature vector $\boldsymbol{f}(s) \in \mathbb{R}^{\dim(\boldsymbol{f})}$ related to s and the feature vector $\boldsymbol{d} \in \mathbb{R}^{\dim(\boldsymbol{d})}$ related to the contexts:

$$R(s|\boldsymbol{\Theta}, \boldsymbol{d}) = \boldsymbol{d}^\top \boldsymbol{\Theta} \boldsymbol{f}(s), \tag{2}$$

where $\boldsymbol{\Theta} \in \mathbb{R}^{\dim(\boldsymbol{d}) \times \dim(\boldsymbol{f})}$ is a weight matrix, with $\dim(\cdot)$ denoting the dimension of a vector. The concrete design of $\boldsymbol{f}(s)$ (features related to acceleration and deceleration positions) and $\boldsymbol{c}_e$ (context factors) is explained in the experimental section.

Given the i-th driving behavior as a T_i-step state sequence $\zeta_i = \{(s_{i,1}, a_{i,1}), ..., (s_{i,T_i}, a_{i,T_i})\}$ and the feature vectors $\boldsymbol{d}_i$ for the driving context, the likelihood for the state sequence is formulated as $p(\zeta_i|\boldsymbol{\Theta}, \boldsymbol{d}_i) \propto \exp\left(\sum_t \boldsymbol{d}_i^\top \boldsymbol{\Theta} \boldsymbol{f}(s_{i,t})\right)$..

For optimization, we minimize the sum of the negative log-likelihood $-\ln p(\boldsymbol{\Theta})$ and the regularization term $\Omega(\boldsymbol{\Theta})$ akin to the IRL in [32] except the regularization term in our model leverages Frobenius norm. Since this objective function could provide its gradient from path planning result, simple gradient methods could be available to obatin optimal parameter $\boldsymbol{\Theta}$.

IV. Experimental results

A. Dataset

1) Experimental vehicle: We employ vehicle with sensing capability then drived this vehicle on residential roads.

During the driving, we collected experimental data consisting of the vehicle position on a road and the corresponding velocity. The velocity data was obtained via a controller area network (CAN) bus. GPS data and steering data obtained via CAN were also collected for vehicle localization. Additionally, a LIDAR sensor, inertial sensor, and cameras were attached to the vehicle to provide reference data[1].

2) Course selection: The experimental data was acquired by driving on residential roads in Tokyo, Japan. Data was collected in various contexts, including both sunny and rainy days and daytime and night. The total travel distance was about 141 km. The collected data included 72 road segments, which can be modeling targets. Each road segment starts with a turn or stop and ends with the next turn or stop. To exclude possible effects on prediction accuracy and evaluation caused by velocity differences at the end of a road segment, we only tested with road segments ending with a stop rather than a turn. There were 26 such road segments as including the subset shown in Figure 2[2]. It should be noted that each

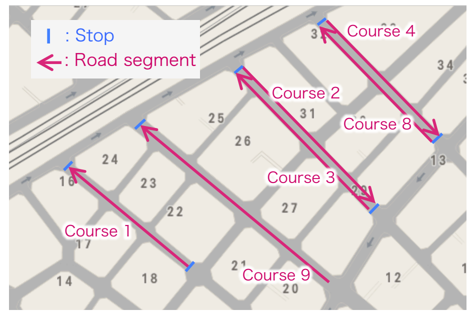


Fig. 2: Experimental course examples.

course includes 1–4 unsignalized intersections, with a variety of lengths, crossroad widths, and so forth, and both one-way and two-way roads. We conducted 1–21 trials on each course.

In the experiment we excluded dynamic factor data, such as the presence of pedestrians, bicycles, and other cars, because the detection accuracy may influence the results. Thereby, we could focus the experiment on defensive driving with potential risk anticipation. Since the experimental vehicle has a LIDAR sensor and cameras, in the future we could use these sensors to deal with such dynamic factors. To illustrate the feasibility of our powerful new smartphone-based application, however, we omitted such additional (often expensive) sensing in this experiment.

3) Driver selection: We recruited a single expert driver from a taxi company as a test subject. We should note that defensive driving styles vary among individuals; there is no unique perfect style. Furthermore, in constructing a reliable active safety system, the model need only be learned from a single reliable driver. We therefore have not concerned ourselves with the diversity of driving strategies among individuals[3].

B. Implementation with state space discretization

In this paper, we formulate the route prediction problem in position-velocity space as an MDP in a discrete state space. Borrowing from the procedure presented in [23], we discretize velocity at $0.5-\mathrm{m/s}$ intervals in 17 steps, ranging from $0.5\ \mathrm{m/s} = 1.8\ \mathrm{km/h}$ to $8.5\ \mathrm{m/s} = 30.6\ \mathrm{km/h}$. This

[1]The data obtained from these additional sensors was never used for predicting driving behavior.

[2]The background maps are from Google Maps [1].

[3]The model can also learn with driving data from multiple drivers.

range of velocities covers speeds slower than 4.0 km/h the speed of a human being taking a leisurely walk as well as the legal speed limit of 30.0 km/h. In addition, we discretize time into $5-\text{Hz}$ intervals, or 0.2 s, considering that a human being takes about one second to brake after recognizing some sort of danger and that prediction should be performed with even finer granularity. We also discretized the behaviors of acceleration, speed maintenance, and deceleration, i.e. transition rule across states by focusing on the driving behavior data.

C. Designing feature descriptors

As described previously, the reward function is represented with two different sets of features: one set related to the positions of acceleration and deceleration, and the other set related to context.

1) State-space-driven features: To generate features representing acceleration and deceleration, we use four environmental factors: intersections, the corners of intersections, and the start and goal positions. Since the data for these environmental factors is embedded in digital road maps for common car navigation systems, therefore, such data is easy to obtain and suitable for practical application.

Specifically, we employ the feature descriptor design from [23] as follows. Given these environmental factors, we represent five kinds of descriptors: (i) reduced velocity at the start and goal, (ii) reduced velocity at corners near intersections, (iii) features related to the velocity upper limit, (iv) features related to acceleration and deceleration from start to goal, and (v) features related to acceleration and deceleration at intersections. These features are represented as potential fields with Gaussian kernels in position-velocity space. The potential fields is varied w.r.t. the position and shape (i.e., the mean and covariance matrix) of the Gaussian kernels so that the features can represent varying degrees and positions of acceleration and deceleration. The resulting $\boldsymbol{f}(s)$ is a 145-dimensional vector containing the features related to the positions of acceleration and deceleration.

2) Context-driven features: To represent varying driving strategies depending on context, we also use features related to context. To generate these features, we use four types of contexts: weather, time of day, road width, and traffic direction. These contexts are represented in binary: sunny or rainy for weather, before sunset or after for time of day, greater or less than the mean of all courses for road width, and one- or two-way for traffic direction. This gives 16 alternative combinations, so $\boldsymbol{d}$ is an 81-dimensional vector. It should be noted that these features do not depend on s.

D. Comparison methods

1) Single-task IRL: In single-task IRL, the same approach proposed in [23], the reward function is represented as the inner product of the weight parameter and the feature vector for the positions of acceleration and deceleration; that is, a single-task reward function is learned. The reward function $R(s)$ is thus formulated as the inner product of the weight vector $\boldsymbol{\theta}$ and feature vector $\boldsymbol{f}(s)$: $R(s|\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \boldsymbol{f}(s)$. For comparative evaluation, we used two different single-task IRL approaches: one learned the reward function from all data (ST-IRL-a), while the other learned multiple single-task reward functions for different combinations of contexts (ST-IRL-b).

2) IRL based on Dirichlet process mixtures: One alternative approach to achieve multi-task IRL is Dirichlet process mixture IRL (DPM-IRL) [9], which performs clustering of tasks and learns multiple reward functions for the respective clusters with a Dirichlet process prior. DPM-IRL makes it possible to automatically segment the training data while simultaneously estimating the number of course clusters.

In DPM-IRL, the model is learned under the assumption that the driving behavior on identical road segments was obtained from the same reward function. Such data is forced to belong to the same cluster during the learning process. The weight vector of the reward function $R_k(s)$ for the k-th cluster is written as $\boldsymbol{\theta}_k$ and represented as $R_k(s) = \boldsymbol{\theta}_k^\top \boldsymbol{f}(s)$. In addition, $z_c = k$ expresses the assignment variable for course c to cluster k, which can be written as $\boldsymbol{z} = \{z_c\}$. The blend ratio for each cluster has a Dirichlet process prior, and the base measure is the model based on the likelihood given by Equation 1. Inference is performed by alternately sampling $\boldsymbol{z}$ and $\{\boldsymbol{\theta}_k\}$. Because it is impossible to draw a sample directly from the likelihood so that the posterior distribution of $\boldsymbol{\theta}_k$ is non-conjugate, the Metropolis-Hastings (MH) algorithm is applied as a sampler for $\boldsymbol{\theta}_k$. For the prior distribution of weights, we used the Laplace distribution with an average of 0.

This approach requires no prior information and makes it possible to improve the prediction accuracy over a wide variety of driving strategies. Appropriate prediction in unknown contexts is problematic, however, because we have no way of estimating which cluster has the most appropriate reward function in the current task. In the experimental evaluation, we assumed that data from one or more runs had already been obtained as evaluation data on the course targeted for prediction, and using this data, we selected the model with the highest likelihood.

E. Evaluation metric

We used the modified Hausdorff distance (MHD) [4] as the evaluation metric. The MHD evaluates the similarity between the state sequence in actual driving and the state sequence generated using learned reward functions in position-velocity space.

The MHD parameter β was set to $\beta = 0.5, 0.9$. Note here that $\beta = 0.5$ represents the median of the MHD between the two point sets, while $\beta = 0.9$ represents the 90th percentile of the MHD with values arranged in ascending order. In the following, we refer to the former as MHD_{50} and the latter as MHD_{90}.

F. Experimental results

In the evaluation, we applied cross-validation to all combinations; that is, we used 1 of the 26 courses for test data and the remaining courses for training data to learn the

TABLE I: Experimental results.

Method	MHD_{50}	MHD_{90}
ST-IRL-a	1.199 ± 0.213	2.765 ± 0.533
ST-IRL-b	1.290 ± 0.399	2.926 ± 0.697
DPM-IRL	1.122 ± 0.097	2.619 ± 0.391
Proposed	**1.072 ± 0.200**	**2.295 ± 0.482**

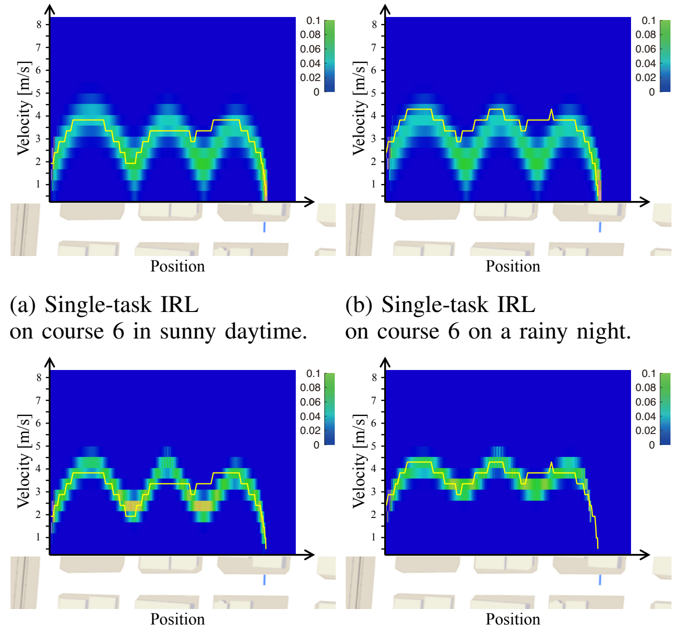
reward function. As noted before, DPM-IRL is not suitable for prediction in unknown contexts. Therefore, we used one trial in the test data to assess model selection in evaluating DPM-IRL. Note that we used data obtained only from the same course in the evaluation of DPM-IRL, meaning that it was performed on extremely favorable terms for DPM-IRL. We excluded course 14 from the DPM-IRL evaluation because there was only one trial on that course.

Table I summarizes the experimental results. The t-test revealed that both MHD_{50} and MHD_{90} for the proposed method were significantly lower than those for ST-IRL-a (p=0.003 for MHD_{50}, 0.003 for MHD_{90}); that is, prediction based on the proposed method was more precise than with ST-IRL-a. For MHD_{50}, which is the median of the prediction error, the prediction accuracy improved by approximately 10.6%. Furthermore, the prediction accuracy for MHD_{90} improved by approximately 17.0%. These results indicate that prediction based on our approach can avoid especially huge mistakes by leveraging contextual information. On the other hand, the prediction accuracy for ST-IRL-b was higher than that for ST-IRL-a in contexts with a large amount of data, but the ST-IRL-b accuracy was the worst overall because of the decreased prediction accuracy due to the shortage of training data for some tasks. The proposed method also exhibited superior performance to DPM-IRL for both MHD_{50} and MHD_{90}. This indicates that even though the proposed method can be applied in unknown contexts, its performance can be made significantly greater than that of DPM-IRL, which requires evaluation data from the target course for model selection, by adding contextual information.

Figure 3 shows the prediction results for driving behavior by the model learned via ST-IRL-a and the proposed method in two different contexts. In each graph the background color indicates the probability of driver behavior prediction at the initial state, when the vehicle was located at the start of the course. That is, the background color is a visualization of $D(s)$, so a lighter color indicates a higher $D(s)$. The yellow lines show both the actual state sequence and the response data. When a yellow line is closer to regions of lighter color, the model's prediction is more precise. In each graph the position on the horizontal axis corresponds to the map below.

Figures 3 (a) and (b) show the predicted and actual driving behavior, respectively, for the case of ST-IRL-a. In this case, the same driving behavior was predicted regardless of the context. With ST-IRL-a, therefore, the model could not predict the lower speed reduction on a rainy night, as shown in Figure 3 (b). In contrast, Figures 3 (c) and (d) show the predicted results with the proposed method and the actual state sequences, respectively. These graphs show that the model using the proposed method could predict the driving behavior more precisely, because in accounted for differences in driving strategies caused by contexts such as weather or the time of day.

In conclusion, the results confirmed the validity of explicitly including context, in comparison with ST-IRL-a. Furthermore, the proposed method was more robust than ST-IRL-b for prediction in novel contexts, because it truncates unnecessary features and shares parameters when certain context factors match among tasks.



(a) Single-task IRL on course 6 in sunny daytime.
(b) Single-task IRL on course 6 on a rainy night.
(c) Propoesd method on course 6 in sunny daytime.
(d) Proposed method on course 6 on a rainy night.

Fig. 3: Prediction results on course 6 in two different contexts. The yellow lines indicate actual behaviors, and the background color indicates the predicted state visitation expectation ($D(s)$). The horizontal position corresponds to the maps below.

V. CONCLUSION

In this paper, we proposed a fine-grained, long-term driving behavior prediction framework. In particular, our framework focuses on how driving strategies vary with context, i.e., with weather, time of day, road width, and traffic direction. We formulated acceleration and deceleration behaviors with inverse reinforcement learning adjusting to multiple contexts with a novel multi-task IRL framework. We conducted an experiment with an expert driver, collecting data over 141 km for various contexts and roads. The results confirmed the effectiveness of the proposed framework by comparing it with conventional IRL and clustering-based multi-task IRL. The results also showed the effectiveness of our approach even in the case of insufficient data and unknown environments, since the model explicitly uses contextual information. Our future work will extend the proposed approach to develop improved applications that include specific driving support, expand the state space for more fine-grained driving activity prediction, and use dynamic factors such as pedestrian locations.

References

[1] "Google maps," https://www.google.com/maps/.

[2] "Statistics of traffic accidents in 2013 (in Japanese)," https://www.npa.go.jp/toukei/koutuu48/toukei.htm.

[3] P. Abbeel and A. Y. Ng, "Apprenticeship Learning via Inverse Reinforcement Learning," in *Proc. of ICML*, 2004.

[4] S. Atev *et al.*, "Learning traffic patterns at intersections by spectral clustering of motion trajectories," in *Proc. of IROS*, 2006, pp. 4851–4856.

[5] M. Babes, V. Marivate, K. Subramanian, and M. L. Littman, "Apprenticeship learning about multiple intentions," in *Proc. of ICML*, 2011, pp. 897–904.

[6] B. Bakker and T. Heskes, "Task clustering and gating for Bayesian multitask learning," *The Journal of Machine Learning Research*, vol. 4, pp. 83–99, 2003.

[7] H. Berndt and K. Dietmayer, "Driver intention inference with vehicle onboard sensors," in *Proc. of ICVES*, 2009, pp. 102–107.

[8] S. Bonnin, T. H. Weisswange, F. Kummert, and J. Schmuedderich, "Pedestrian crossing prediction using multiple context-based models," in *Proc. of ITSC*, 2014, pp. 378–384.

[9] J. Choi and Kee-Eung Kim, "Nonparametric Bayesian inverse reinforcement learning for multiple reward functions," in *Proc. of NIPS*, 2012, pp. 314–322.

[10] C. Dimitrakakis and C. A. Rothkopf, "Bayesian multitask inverse reinforcement learning," in *Proc. of EWRL*, 2012, pp. 273–284.

[11] X. Gao, B. Firner, S. Sugrim, V. Kaiser-Pendergrast, Y. Yang, and J. Lindqvist, "Elastic pathing: Your speed is enough to track you," in *Proc. of UbiComp*, 2014, pp. 975–986.

[12] T. Gindele *et al.*, "A Probabilistic Model for Estimation Driver Behaviors and Vehicle Trajectories in Traffic Environments," in *Proc. of ITSC*, 2010, pp. 1625–1631.

[13] C. Hermes *et al.*, "Long-term Vehicle Motion Prediction," in *Proc. of IV*, 2009, pp. 652–657.

[14] E. Horvitz and J. Krumm, "Some help on the way: Opportunistic routing under uncertainty," in *Proc. of UbiComp*, 2012, pp. 371–380.

[15] *ITARDA INFORMATION (in Japanese)*, No. 98, Institute for Traffic Accident Research and Data Analysis, 2013. [Online]. Available: http://www.itarda.or.jp/itardainfomation/info98.pdf

[16] K. M. Kitani *et al.*, "Activity forecasting," in *Proc. of ECCV*, 2012, pp. 201–214.

[17] M. Liebner *et al.*, "Driver Intent Inference at Urban Intersections using the Intelligent Driver Model," in *Proc. of IV*, 2012, pp. 1162–1167.

[18] ——, "Active safety for vulnerable road users based on smartphone position data," in *Proc. of IV*, 2013, pp. 256–261.

[19] A. Y. Ng and S. J. Russell, "Algorithms for Inverse Reinforcement Learning," in *Proc. of ICML*, 2000.

[20] M. L. Puterman, *Markov Decision Process: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.

[21] D. Ramachandran and E. Amir, "Bayesian inverse reinforcement learning," in *Proc. of IJCAI*, 2007, pp. 2586–2591.

[22] B. Romera-Paredes, H. Aung, N. Bianchi-Berthouze, and M. Pontil, "Multilinear multitask learning," in *Proc. of ICML*, 2013, pp. 1444–1452.

[23] M. Shimosaka, T. Kaneko, and K. Nishi, "Modeling risk anticipation and defensive driving on residential roads with inverse reinforcement learning," in *Proc. of ITSC*, 2014.

[24] J. Straub, S. Zheng, and J. W. F. Iii, "Bayesian nonparametric modeling of driver behavior," in *Proc. of IV*, 2014, pp. 932–938.

[25] J. B. Tenenbaum and W. T. Freeman, "Separating style and content with bilinear models," *Neural computation*, vol. 12, no. 6, pp. 1247–1283, 2000.

[26] Traffic Accident Analysis System, "Driver's manual - risk anticipation and defensive driving," https://taas.koroad.or.kr/Eng/dm/dm_indexMain.jsp?link=Risk.

[27] Y. Xue, X. Liao, L. Carin, and B. Krishnapuram, "Multi-task learning for classification with dirichlet process priors," *The Journal of Machine Learning Research*, vol. 8, pp. 35–63, 2007.

[28] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger," in *Proc. of UbiComp*, 2011, pp. 109–118.

[29] D. Zhang, N. Li, Z.-H. Zhou, C. Chen, L. Sun, and S. Li, "iBAT: Detecting anomalous taxi trajectories from gps traces," in *Proc. of UbiComp*, 2011, pp. 99–108.

[30] J. Zheng and L. M. Ni, "Modeling heterogeneous routing decisions in trajectories for driving experience learning," in *Proc. of UbiComp*, 2014, pp. 951–961.

[31] Y. Zheng, Y. Liu, J. Yuan, and X. Xie, "Urban computing with taxicabs," in *Proc. of UbiComp*, 2011, pp. 89–98.

[32] B. D. Ziebart *et al.*, "Maximum Entropy Inverse Reinforcement Learning," in *Proc. of AAAI*, 2008, pp. 1433–1438.

[33] B. D. Ziebart, A. L. Maas, A. K. Dey, and J. A. Bagnell, "Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior," in *Proc. of UbiComp*, 2008, pp. 322–331.