

Probabilistic Prediction of Interactive Driving Behavior via Hierarchical Inverse Reinforcement Learning

Liting Sun¹, Wei Zhan¹, and Masayoshi Tomizuka¹

Abstract—Autonomous vehicles (AVs) are on the road. To safely and efficiently interact with other road participants, AVs have to accurately predict the behavior of surrounding vehicles and plan accordingly. Such prediction should be probabilistic, to address the uncertainties in human behavior. Such prediction should also be interactive, since the distribution over all possible trajectories of the predicted vehicle depends not only on historical information, but also on future plans of other vehicles that interact with it. To achieve such interaction-aware predictions, we propose a probabilistic prediction approach based on hierarchical inverse reinforcement learning (IRL). First, we explicitly consider the hierarchical trajectory-generation process of human drivers involving both discrete and continuous driving decisions. Based on this, the distribution over all future trajectories of the predicted vehicle is formulated as a mixture of distributions partitioned by the discrete decisions. Then we apply IRL hierarchically to learn the distributions from real human demonstrations. A case study for the ramp-merging driving scenario is provided. The quantitative results show that the proposed approach can accurately predict both the discrete driving decisions such as yield or pass as well as the continuous trajectories.

I. INTRODUCTION

Autonomous vehicles (AVs) are interacting with human on the road. To this end, it needs to predict and reason about possible future behavior of human, and plan its own trajectories accordingly. Wrong prediction can cause either too conservative motions such as unnecessary stops/yielding, or dangerous situations like emergency brakes and unavoidable collisions. Hence, accurate prediction is of crucial importance to enable a safe and efficient autonomous car.

Starting from as simple as assuming that the other drivers would maintain their current velocities within the planning horizon [1] [2] to as complicated as modelling the probability distributions over all possible trajectories/actions, many prediction approaches have been proposed. In terms of prediction output, they can be categorized into two groups: deterministic prediction [1]–[4] and probabilistic prediction [5]–[7]. Both categories can be implemented based on various models such as neural networks (NN) [8] [9], hidden Markov models (HMM) [10] and Bayes net [6]. Most of the work, however, formulated the future trajectories/actions as either deterministic functions or conditional probabilities of historical and current scene states. The influence of human drivers' beliefs about the other vehicles future actions are ignored in the prediction framework.

*This work was partially supported by the international Chair Drive for All, Foundation MINES ParisTech.

¹L. Sun, W. Zhan, and M. Tomizuka are with the Department of Mechanical Engineering, University of California, Berkeley, CA 94720 USA (e-mail: litingsun, wzhan, tomizuka@berkeley.edu).

In fact, in interactive driving scenarios, human drivers will actively anticipate and reason about surrounding vehicles' behavior when they decide the next-step actions. This means that besides historical and current scene states, the distribution over all possible trajectories is also influenced by their beliefs about other vehicles' plan. For example, given the same historical trajectories, a lane-keeping driver might be more probable to decelerate rather than maintaining current speed if the driver thinks another vehicle is about to merge into his lane. Such interaction-aware prediction is incorporated implicitly into a planning framework in [11], but the prediction output is deterministic.

Our insight is that in highly interactive driving scenarios, an accurate prediction should consider not only the time-domain dependency but also the interaction among vehicles. Namely, an autonomous car should anticipate the conditional probability over all possible trajectories of the human driver given not only historical and current states, but also its own future plans.

To obtain such a probabilistic and interactive prediction of human drivers' behavior, we need to approximate the "internal incentive" that drives human to generate specific behavior. Note that human's planning procedure is naturally hierarchical, involving both discrete and continuous driving decisions. The discrete driving decisions determine the game-theoretic outcomes of interaction such as to yield or to pass, whereas the continuous driving decisions influence details of the resulting trajectories in terms of smoothness, distances to other road participants and higher-order dynamics such as velocities, accelerations and jerks. Hence, to describe the influence of both discrete and continuous decisions, we propose a hierarchical inverse reinforcement learning (IRL) framework in this paper, to learn trajectory-generation process of human from observed demonstrations. Given the autonomous vehicle's future plan, the probability distribution over all future trajectories of the predicted human driver is modelled as a mixture of distributions, partitioned by the discrete driving decisions.

Contributions of this paper are the following two.

A formalism for probabilistic interactive prediction of human driving behavior. We model the prediction problem from the perspective of a two-agent game by explicitly considering the responses of one agent to another. We formalize the distribution of predicted vehicle's future trajectories as a probability distribution conditioned not only on historical and current information, but also on future plans of the ego vehicle.

A hierarchical IRL framework. We explicitly consider the

hierarchical planning procedure of human drivers, and formulate the influences of both discrete and continuous driving decisions. Via hierarchical IRL, the conditional distribution over all possible future trajectories is expressed as a mixture of probability distributions partitioned by different game-theoretic outcomes of interactive vehicles.

Related Work in IRL: Initially proposed by Kalman [12], the concept of Inverse Reinforcement Learning is first formulated in [13]. It aims to infer the reward/cost functions of agents from their observed behavior by assuming that the agents are rational. To deal with uncertainties and noisy observations, Ziebart *et al.* [14] extended the algorithm based on the principle of maximum entropy. It assumes that agents actions/behavior with lower cost are exponentially more probable, and thus an exponential distribution family can be established to approximate the distribution of actions/behavior. Building on this, Levine *et al.* [15] formulated the continuous IRL algorithm and used it to mimic and predict human driving behavior, and Liu *et al.* [16] used IRL to approximate the decision-making process of taxi drivers and passengers in public transportation. In [17], Shimomura *et al.* predicted human driving behavior in diverse environment by formulating IRL with multiple reward functions. Kretschmar *et al.* [18] considered the hierarchical trajectory-generation process of human navigation behavior, but they focus on cooperative agents that share the same reward function instead of interactive agents that have significantly different reward functions.

II. PROBLEM STATEMENT

We focus on predicting human drivers' interactive behavior within two vehicles: a host vehicle (denoted by $(\cdot)_H$) and a predicted vehicle (denoted by $(\cdot)_M$). All other in-scene vehicles are treated as surrounding vehicles, denoted by $(\cdot)_O^i$ ($i=1, 2, \dots, N$ is the index for surrounding vehicles). We use ξ to represent historical vehicle trajectories, and $\hat{\xi}$ for future trajectories¹.

It is obvious that the probability distribution over all possible trajectories of the predicted vehicle depends on his own historical trajectory and those of surrounding vehicles. Mathematically, such time-domain state dependency can be modelled as a conditional probability density function (PDF), as in [19]:

$$p(\hat{\xi}_M | \xi_O^{1:N}, \xi_H, \xi_M). \quad (1)$$

However, as discussed above, in interactive driving scenarios, the influence of human's beliefs about others' next-step actions cannot be ignored in order to get a good prediction. From this perspective, the probability distribution over all possible trajectories of the predicted vehicle should also be conditioned on potential plans of the host vehicle.

¹Note that trajectory is a sequence of states, i.e., $\xi = [x_1^T, x_2^T, \dots, x_L^T]^T$ where x_i is the vehicle state at i -th time step. L is the trajectory length. Depending on the representation, the vehicle state can be different. For instance, it can simply be the positions of vehicles, or it can include velocities, yaw angles, accelerations, etc.

Mathematically, such interaction-dependency further refines the conditional PDF in (1) as

$$p(\hat{\xi}_M | \xi_O^{1:N}, \xi_H, \xi_M, \hat{\xi}_H). \quad (2)$$

The key aspect of this formulation is to actively consider the influence of the host vehicle's actions when predicting the future trajectories of the predicted vehicle. In this sense, we are trying to enable the autonomous vehicle to "think" in the way that the predicted human driver thinks if he had known the autonomous vehicle's future plan.

III. MODELING HUMAN DRIVING BEHAVIOR

In order to predict human's interactive driving trajectories, we need to model the internal incentive of human that generates his behavior.

A. Probabilistic Hierarchical Trajectory-Generation Process

As addressed above, the trajectory-generation process of human drivers is naturally probabilistic and hierarchical. It involves both discrete and continuous driving decisions. The discrete driving decisions determine "rough" pattern (or homotopy class) of his future trajectory as a game-theoretic result (e.g., to yield or to pass), whereas the continuous driving decisions influence details of the trajectory such as velocities, accelerations and smoothness.

Figure 1 illustrates such probabilistic and hierarchical trajectory-generation process for a lane-changing driving scenario. The predicted vehicle (blue) is trying to merge into the lane occupied by the host vehicle (red). Given all observed historical trajectories $\xi = \{\xi_O^{1:N}, \xi_H, \xi_M\}$ and his belief about the host vehicle's future trajectory $\hat{\xi}_H$, he first decides whether to merge behind the host vehicle (d_M^1) or merge front (d_M^2). Such discrete driving decisions are outcomes of the first-layer probability distribution $P(d_M | \xi, \hat{\xi})$, and partition the space of all possible trajectories into two distinct homotopy classes², each of them can be described via a second-layer probability distribution $p(\hat{\xi}_M | d_M, \xi, \hat{\xi}_H)$. Note that among different homotopy classes, the distributions of the continuous trajectories can be significantly different since the driving preference under different discrete decisions may be different. For example, a vehicle decided to merge behind might care more about comfort and less about speed, while a vehicle merging front might care about completely the opposite.

Hence, the conditional distribution $p(\hat{\xi}_M | \xi, \hat{\xi}_H)$ in (2) is formulated as a mixture of distributions, which explicitly captures the influences of both discrete and continuous driving decisions of human drivers:

$$p(\hat{\xi}_M | \xi, \hat{\xi}_H) = \sum_{d_M^i \in \mathcal{D}_M} p(\hat{\xi}_M | d_M^i, \xi, \hat{\xi}_H) P(d_M^i | \xi, \hat{\xi}_H) \quad (3)$$

where \mathcal{D}_M represents the set of all possible discrete decisions for the predicted vehicle.

²Two trajectories belong to a same homotopy class if they can be continuously transformed to each other without collisions [18].

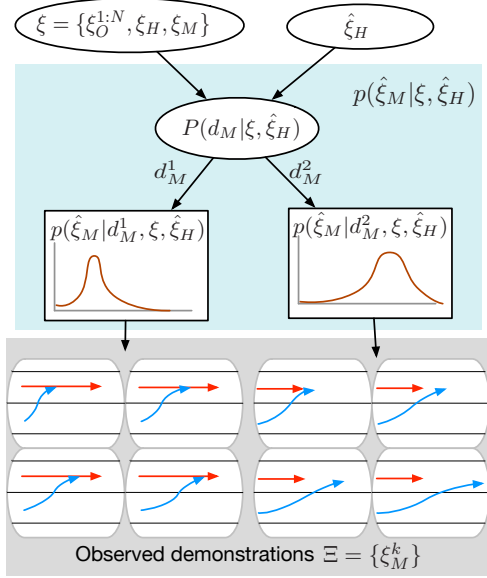


Fig. 1: The probabilistic and hierarchical trajectory-generation process for a lane changing scenario. The predicted vehicle (blue) is trying to merge into the lane of the host vehicle (red). Given all observed historical trajectories $\xi = \{\xi_O^{1:N}, \xi_H, \xi_M\}$ and his belief about the host vehicle's future trajectory $\hat{\xi}_H$, the trajectory distribution of the predicted vehicle over all the trajectory space is partitioned first by the discrete decisions: merge behind (d_M^1) and merge front (d_M^2). Under different discrete decisions, the distributions of continuous trajectories can be significantly different, and each of them is represented via a probability distribution model. The observed demonstrations are samples satisfying the distributions.

B. Hierarchical Inverse Reinforcement Learning

Equation (3) suggests that in order to model the conditional PDF in (2) for interactive prediction, we need to model the hierarchical probabilistic models $P(d_M|\xi, \hat{\xi}_H)$ and $p(\hat{\xi}_M|d_M^i, \xi, \hat{\xi}_H)$ for each $d_M^i \in \mathcal{D}_M$.

We thus propose to apply inverse reinforcement learning hierarchically to learn all the models from observed demonstrations of human drivers. Based on the principle of maximum entropy [14], we assume that all drivers are exponentially more likely to make decisions (both discrete and continuous) that lead to a lower cost. This introduces a family of exponential distributions depending on the cost functions, and our interest is to find the optimal hierarchical cost functions that ultimately lead to trajectory distributions matching that of the observed trajectories in a given demonstration set Ξ .

C. Modeling Continuous Driving Decisions

Suppose that the demonstration set Ξ is partitioned into $|\mathcal{D}|$ subsets by the discrete decisions $d_M^i \in \mathcal{D}$. $|\mathcal{D}|$ is the dimension of \mathcal{D} . Each subset $\Xi_{d_M^i}$ contains trajectories that belong to the same homotopy class. We represent each demonstration in Ξ_i by a tuple $(\hat{\xi}_M, d_M^i, \xi, \hat{\xi}_H)$ where $\xi = [\xi_O^{1:N}, \xi_H, \xi_M]$ represents all historical information. Since the trajectory space is continuous and demonstrations are

noisily local-optimal, we use Continuous Inverse Optimal Control with Locally Optimal Examples [15].

1) *Continuous-Space IRL*: Under discrete decision d_M , we assume that the cost of each trajectory can be linearly parametrized by a group of selected features $\{\mathbf{f}_{d_M}\}$, i.e., $C(\boldsymbol{\theta}_{d_M}, \hat{\xi}_M, \xi, \hat{\xi}_H) = \boldsymbol{\theta}_{d_M}^T \mathbf{f}_{d_M}(\hat{\xi}_M, \xi, \hat{\xi}_H)$, where $\boldsymbol{\theta}_{d_M}$ is the parameter vector to determine the emphasis of each of the features. Then trajectories with higher cost are exponentially less likely based on the principle of maximum entropy :

$$P(\hat{\xi}_M|\boldsymbol{\theta}_{d_M}, d_M^i, \xi, \hat{\xi}_H) \propto e^{-C(\boldsymbol{\theta}_{d_M}, \hat{\xi}_M, \xi, \hat{\xi}_H)} \quad (4)$$

Hence, the log likelihood of the given demonstration subset $\{\Xi_{d_M}\}$ is given by

$$\log P(\Xi_{d_M}|\boldsymbol{\theta}_{d_M}) = \sum_{\hat{\xi}_M \in \Xi_{d_M}} \log \frac{e^{-C(\boldsymbol{\theta}_{d_M}, \hat{\xi}_M, \xi, \hat{\xi}_H)}}{\int e^{-C(\boldsymbol{\theta}_{d_M}, \tilde{\xi}_M, \xi, \hat{\xi}_H)} d\tilde{\xi}_M}. \quad (5)$$

Our goal is to find the optimal $\boldsymbol{\theta}_{d_M}$ such that the given demonstration set is most likely to happen:

$$\boldsymbol{\theta}_{d_M}^* = \arg \max_{\boldsymbol{\theta}_{d_M}} P(\Xi_{d_M}|\boldsymbol{\theta}_{d_M}) \quad (6)$$

To tackle the normalization factor in (5), we use Laplace approximation as in [15]. Namely, the cost along an arbitrary trajectory $\tilde{\xi}_M$ is approximated by its second-order Taylor expansion around the demonstrated trajectory $\hat{\xi}_M$:

$$C(\boldsymbol{\theta}_{d_M}, \tilde{\xi}_M, \xi, \hat{\xi}_H) \approx C(\boldsymbol{\theta}_{d_M}, \hat{\xi}_M, \xi, \hat{\xi}_H) + (\tilde{\xi}_M - \hat{\xi}_M)^T \frac{\partial C}{\partial \xi_M} + (\tilde{\xi}_M - \hat{\xi}_M)^T \frac{\partial^2 C}{\partial \xi_M^2} (\tilde{\xi}_M - \hat{\xi}_M)$$

This enables the normalization factor become a Gaussian integral and can be solved analytically. Define $g_{\hat{\xi}_M}(\boldsymbol{\theta}_{d_M}) = \frac{\partial C}{\partial \xi_M}|_{\hat{\xi}_M}$ and $H_{\hat{\xi}_M}(\boldsymbol{\theta}_{d_M}) = \frac{\partial^2 C}{\partial \xi_M^2}|_{\hat{\xi}_M}$ as, respectively, the gradient and Hessian of the cost along trajectory $\hat{\xi}_M$. Then (6) is translated to the optimization problem in (7), which intuitively means the optimal cost function should have small gradient and large positive Hessians along the demonstrated trajectories. For details, one can refer to [15].

$$\min_{\boldsymbol{\theta}_{d_M}} \sum_{\hat{\xi}_M \in \Xi_{d_M}} g_{\hat{\xi}_M}^T(\boldsymbol{\theta}_{d_M}) H_{\hat{\xi}_M}^{-1}(\boldsymbol{\theta}_{d_M}) g_{\hat{\xi}_M}(\boldsymbol{\theta}_{d_M}) - \log |H_{\hat{\xi}_M}(\boldsymbol{\theta}_{d_M})| \quad (7)$$

2) *Features*: The features we selected to parametrize the continuous trajectories can be grouped as follows:

- **Speed** - The incentive of the human driver to reach a certain speed limit v_{lim} is captured by the feature

$$f_v(\hat{\xi}_M) = \sum_{t=0}^L (v_t - v_{\text{lim}})^2 \quad (8)$$

v_t is the speed at time t along trajectory $\hat{\xi}_M$ and L is the length of the trajectory.

- **Traffic** - In dense traffic environment, human drivers tend to follow the traffic. Hence, we introduce a feature based on the intelligent driver model (IDM) [20]

$$f_{\text{IDM}}(\hat{\xi}_M) = \sum_{t=0}^L (s_t - s_t^{\text{IDM}})^2 \quad (9)$$

where s_t is the actual spatial headway between the front vehicle and predicted vehicle at time t along trajectory $\hat{\xi}_M$, and s_t^{IDM} is the spatial headway suggested by IDM.

- Control effort and smoothness - Human drivers typically prefer to drive efficiently and smoothly, avoiding unnecessary accelerations and jerks. To address such preference, we introduce a set of kinematics-related features:

$$f_{\text{acc}}(\hat{\xi}_M) = \sum_{t=0}^L a_t^2, \quad f_{\text{jerk}}(\hat{\xi}_M) = \sum_{t=1}^L \left(\frac{a_t - a_{t-1}}{\Delta t} \right)^2 \quad (10)$$

where a_t represents the acceleration at time t along the trajectory $\hat{\xi}_M$. Δt is the sampling time.

- Clearance to other road participants - Human drivers care about their distances to other road participants when they drive since distance is crucially related to safety. Hence, we introduce a distance-related feature

$$f_{\text{dist}}(\hat{\xi}_M) = \sum_{t=0}^L \sum_{k=1}^{N+1} e^{-\frac{(x_t - x_t^k)^2}{l^2} - \frac{(y_t - y_t^k)^2}{w^2}} \quad (11)$$

where (x_t, y_t) and (x_t^k, y_t^k) represent, respectively, the coordinates of the predicted vehicle along $\hat{\xi}_M$ and those of the k -th surrounding vehicle. Parameters l and w are the length and width of the predicted vehicle. We use coordinates in Frenet Frame to deal with curvy roads, i.e., x denotes the travelled distance along the road and y is the lateral deviation from the lane center.

- Goal - This feature describes the short-term goals of human drivers. Typically, goals are determined by the discrete driving decisions. For instance, if a lane-changing vehicle decides to merge in front of a host vehicle on his target lane, he will set his short-term goals to be ahead of the host vehicle. The goal-related feature is given by

$$f_g(\hat{\xi}_M) = \sum_{t=0}^L \|(x_t, y_t) - (x_t^g, y_t^g)\|_2^2 \quad (12)$$

- Courtesy - Most of human drivers view driving as a social behavior, meaning that they not only care about their own cost, but also care about others' cost, particularly when they are merging into others' lanes [21]. Suppose that the cost of the host vehicle is $C_H(\hat{\xi}_M, \xi, \hat{\xi}_H)$, then to address the influence of courtesy to the interaction, we introduce the feature

$$f_{\text{court}}(\hat{\xi}_M) = \max \left\{ C_H(\hat{\xi}_M, \xi, \hat{\xi}_H) - C_H^{\text{default}}, 0 \right\}. \quad (13)$$

This feature describes the possible extra cost brought by the trajectory $\hat{\xi}_M$ of the predicted vehicle to the host vehicle, compared to the default host vehicle's cost C_H^{default} . We can learn about $C_H(\cdot)$ also from demonstrations. Details about this will be covered in the case study.

For vehicles in different driving scenarios, or under different discrete driving decisions, the features we used to

parametrize their costs are different subsets of the above listed ones. For instance, drivers decided to merge behind or front would set different goals, and drivers with right of way would most likely care less about courtesy than those without.

Remark 1: Note that all variables v_t, a_t, δ_t, x_t and y_t in (8)-(13) can be expressed as functions of trajectory $\hat{\xi}_M$. For instance, if we define $\hat{\xi}_M = [x_0, y_0, \dots, x_L, y_L]^T$ where x and y are the coordinates in Frenet Frame, then we can obtain all variables via numerical differentiation. Details are omitted.

D. Modeling Discrete Driving Decisions

1) *Features:* Different from the continuous driving decisions that influence the higher-order dynamics of the trajectories, the discrete decisions determine the homotopy of the trajectories. To capture this, we selected the following two features to parametrize the cost function that induces the discrete decisions:

- Rotation angle - To describe the interactive driving behavior such as overtaking from left or right sides, merging in from front or back, we compute the rotation angle from (x_t, y_t) to $(x_{t,H}, y_{t,H})$ along trajectory $\hat{\xi}_M \in \Xi_{d_M}$. Define the angle as ω_t , and then the rotation angle feature is given by

$$f_{\angle}(d_M) = \sum_{t=0}^L \omega_t \quad (14)$$

where d_M is a discrete decision and L is the length of the trajectory $\hat{\xi}_M$.

- Minimum cost - It is also possible that human drivers make discrete decisions by evaluating the cost of trajectories under each decision, and select the one leading to the minimum-cost trajectory. To address this factor, we consider the feature

$$f_{\text{cost}}(d_M) = \min_{\hat{\xi}_M} \theta_{d_M}^T \mathbf{f}_{d_M}(\hat{\xi}_M, \xi, \hat{\xi}_H). \quad (15)$$

where θ_{d_M} and \mathbf{f}_{d_M} , respectively, represent the learned parameters and selected features for the continuous trajectory distribution under discrete decision d_M .

2) *Discrete-Space IRL:* Similarly, we assume that the decisions with lower cost are exponentially more probable. We also assume that the cost function is linearly parametrized by ψ and feature vector $\mathbf{f}^d = [f_{\angle}, f_{\text{cost}}]^T$, i.e., $C^d(d_M, \psi) = \psi^T \mathbf{f}^d(d_M)$. Again, our goal is to find the optimal ψ^* such that the likelihood of the demonstration set Ξ is maximized:

$$\max_{\psi} P(\Xi|\psi) = \max_{\psi} \prod_{d_M \in \Xi} \frac{e^{-\psi^T \mathbf{f}^d(d_M)}}{\sum_{\tilde{d}_M \in \mathcal{D}} e^{-\psi^T \mathbf{f}^d(\tilde{d}_M)}} \quad (16)$$

By taking the log probability and gradient-descent approach, the parameter ψ is updated via

$$\psi_{s+1} = \psi_s - \alpha \left(\frac{1}{N} \sum_{\Xi} \mathbf{f}^d(d_M) - \sum_{\tilde{d}_M} P(\tilde{d}_M|\psi_s) \mathbf{f}^d(\tilde{d}_M) \right) \quad (17)$$

$$P(\tilde{d}_M|\psi_s) = \frac{e^{-\psi_s^T \mathbf{f}^d(\tilde{d}_M)}}{\sum_{\tilde{d}_M \in \mathcal{D}} e^{-\psi_s^T \mathbf{f}^d(\tilde{d}_M)}} \quad (18)$$

where α is the step size and N is the number of demonstrated trajectories in set Ξ . To estimate the probability $P(\tilde{d}_M|\psi_s)$, we use sampling-based method. Detailed implementation will be covered later in the case study.

IV. CASE STUDY

In this section, we apply the proposed hierarchical IRL approach to model and predict the interactive human driving behavior in a ramp-merging scenario.

A. Data Collection

We collect human driving data from the Next Generation SIMulation (NGSIM) dataset [22]. It captures the highway driving behaviors/trajectories by cameras mounted on top of surrounding buildings. The sampling time of the trajectories is $\Delta t=0.01s$. We choose 134 ramp-merging trajectories on Interstate 80 (near Emeryville, California), and separated them into two sets: a training set of size 80 (denoted by Ξ , i.e., the human demonstrations), and the other 54 trajectories as the test set.

Figure 2 shows the road map (See [22] for detailed geometry information) and an example group of trajectories. There are four cars in scene, one merging vehicle (red), one lane-keeping vehicle (blue) and two surrounding vehicles (black), with one ahead of the blue car and the other behind. Our interest focuses on the interactive driving behavior of both the merging vehicle and the lane-keeping vehicle.



Fig. 2: The merging map on Interstate 80 near Emeryville, California. Red: merging vehicle; Blue: lane-keeping vehicle; Black: other surrounding vehicles

B. Driving Decisions and Feature Selection

We use the same hierarchical IRL approach to model the conditional probability distributions for both the merging vehicle and the lane-keeping vehicle.

1) *Driving Decisions*: In the ramp-merging scenario, the driving decisions are listed as in Table I. As mentioned above, $\mathbf{x}=[x_1, \dots, x_L]^T$ and $\mathbf{y}=[y_1, \dots, y_L]^T$ are, respectively, the coordinate vectors in Frenet Frame along the longitudinal and lateral directions. L is set to be 50, i.e., in each demonstration, 5s trajectories are collected.

	Discrete Decisions	Continuous Decisions
merging-in vehicle	$\mathcal{D} = \{\text{merge front, merge back}\}$	trajectory $\xi=[x_1, y_1, \dots, x_L, y_L]^T$
lane-keeping vehicle	$\mathcal{D} = \{\text{yield, pass}\}$	trajectory $\xi=[x_1, y_1, \dots, x_L, y_L]^T$

TABLE I: Driving decisions for the interactive vehicles

2) *Feature Selection*: Since the right of way for the merging vehicle and the lane-keeping vehicles are different, we define different features for them.

For the lane-keeping vehicle, the feature vectors related to the continuous driving decisions are as follows:

- **Yield**: $\mathbf{f}_{\text{yield}}=[f_v, f_{\text{acc}}, f_{\text{jerk}}, f_{\text{dist}}, f_g]^T$. We exclude the feature f_{IDM} because once the lane-keeping driver decides to yield to the merging vehicle, it is very likely that he cares more about the relative positions to the merging vehicle instead of the heading space to the front vehicle when he plans his continuous trajectories. The goal position in f_g is set to be $[x_{\text{current lane center}}, y_{t, \text{merging vehicle}} - s_0]$, i.e., s_0 behind the merging vehicle along longitudinal direction.
- **Pass**: $\mathbf{f}_{\text{pass}}=[f_v, f_{\text{IDM}}, f_{\text{acc}}, f_{\text{jerk}}, f_{\text{dist}}, f_g]^T$. In this case, the goal position in f_g is set to be ahead of the merging vehicle along longitudinal direction, i.e., $[x_{\text{current lane center}}, y_{t, \text{merging vehicle}} + s_0]$. Also, if the driver decides to pass, it is more probable that the heading space to the front vehicle will influence the distribution of his continuous trajectories.

For the merging vehicle, the feature vectors for the continuous driving models are:

- **Merge back**: $\mathbf{f}_{\text{back}}=[f_v, f_{\text{acc}}, f_{\text{jerk}}, f_{\text{dist}}, f_g]^T$. The goal position is set to be $[x_{\text{target lane center}}, y_{t, \text{on-lane vehicle}} - s_0]$.
- **Merge front**: $\mathbf{f}_{\text{front}}=[f_v, f_{\text{IDM}}, f_{\text{acc}}, f_{\text{jerk}}, f_{\text{dist}}, f_{\text{court}}, f_g]^T$. Once the merging driver decides to merge in front of the lane-keeping vehicle, his heading space to the front lane-keeping vehicle is crucial to the distribution of his possible trajectories. Hence, we include feature f_{IDM} . Moreover, to respect the right of way of the lane-keeping vehicle, merging drivers might care about the extra cost they bring to the lane-keeping vehicle and prefer trajectories that induce less extra cost. To capture such effect, we add f_{court} . Given demonstrated trajectory group $\Xi=[\xi_{\text{merging}}, \xi_{\text{surroundings}}, \xi_{\text{lane-keeping}}]$, f_{court} is computed via (13), i.e.,

$$f_{\text{court}}(\xi_{\text{merging}}) = C_{\text{lane-keeping}}(\theta_{\text{yield}}, \xi) - C_{\text{lane-keeping}}^{\text{default}} \quad (19)$$

The cost function $C_{\text{lane-keeping}}(\theta_{\text{yield}})=\theta_{\text{yield}}^T \mathbf{f}_{\text{yield}}$ is learned with the features selected above. Regarding to C_H^{default} , we assume that the lane-keeping vehicle is by default following IDM.

C. Implementation Details and Training Performance

We use Tensorflow to implement the hierarchical IRL algorithm. Figure 3 gives the training curves regarding to both the continuous and discrete driving decisions. Due to the hierarchical structure, we first learn all four continuous distribution models for both the merging vehicle and the lane-keeping vehicle under different discrete decisions. We randomly sample subsets of trajectories from the training set and perform multiple trains. As seen from Fig. 3, the parameters in each continuous model converge quite consistently with small variance.

With the converged parameter vectors θ_{yield} , θ_{pass} , θ_{front} and θ_{back} , the discrete feature vectors \mathbf{f}^d s are then calculated

and thus the optimal parameter vectors ψ s are learned via discrete IRL. To efficiently sample continuous trajectories under different discrete decisions, we first obtain the most-likely trajectories by optimizing the learned cost functions under each decision and then randomize them with additive Gaussian noise. The training curves (blue) are also shown in Fig. 3.

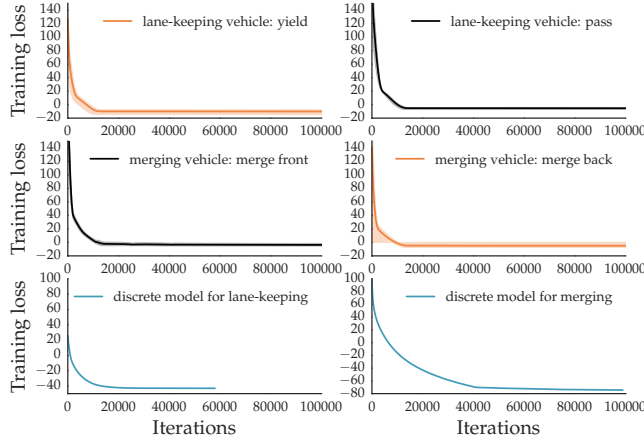


Fig. 3: The training curves of both the lane-keeping vehicle and the merging vehicle under different discrete driving decisions

D. Test Results

Once θ_{yield} , θ_{pass} , θ_{front} , θ_{back} and ψ_{merging} , $\psi_{\text{lane-keeping}}$ are acquired, we can obtain the conditional PDF defined in (2) via (3). With this PDF, probabilistic and interactive prediction of human drivers' behavior can be obtained. The prediction horizon is 3s, i.e., 30 points with a sampling period of 0.1s.

1) *Accuracy of the probabilistic prediction of discrete decisions:* To measure the accuracy of the probabilistic prediction of discrete decisions, we extract $N=2000$ short trajectories in the test set with a horizon length of 10 from the 54 long trajectories. For each short trajectory, starting from the same initial condition, we sample $M=4$ trajectories with different motion patterns (discrete decisions) in the spatiotemporal domain, and one of them is set the same as the ground truth. Hence, the ground truth probabilities for all trajectories are either 0 or 1.

Metric: We adopt a fatality-aware Brier metric [23] to evaluate the prediction accuracy. The fatality-aware metric refines the Brier score [24] by formulating the prediction errors in three different aspects: ground-truth accuracy (\mathcal{G}) measuring the prediction error of the ground-truth motion pattern (Ξ_g), conservatism accuracy (\mathcal{C}) measuring the false alarm of aggressive motion patterns (Ξ_a), and non-defensiveness accuracy (\mathcal{D}) measuring the miss detection of dangerous motion patterns (Ξ_d). Let $P_{i,j}$ be the predicted probability of motion pattern j for i -th test example ($j=1, 2, \dots, M$ and $i=1, 2, \dots, N$), then the three scores

and the overall score are given by

$$\mathcal{G} = \frac{1}{|\Xi_g|} \sum_{(i,j) \in \Xi_g} (P_{i,j} - 1)^2, \quad (20)$$

$$\mathcal{C} = \frac{1}{|\Xi_a|} \sum_{(i,j) \in \Xi_a} W_c(i,j) (P_{i,j} - 0)^2, \quad (21)$$

$$\mathcal{D} = \frac{1}{|\Xi_d|} \sum_{(i,j) \in \Xi_d} W_d(i,j) (P_{i,j} - 0)^2, \quad (22)$$

$$\mathcal{B}_c = \mathcal{G} + \mathcal{C} + \mathcal{D}. \quad (23)$$

$W_c(i,j)$ and $W_d(i,j)$ are, respectively, the weights penalizing the conservatism and non-defensiveness of the motion pattern (i,j) . For more details, one can refer to [23].

Comparison: We compare the prediction among three different approaches: the proposed hierarchical IRL method, a neural-network (NN) based method [25] and a hidden markov models (HMM) based method [26].

Results: The scores for all three methods are shown in Table II. We can see that the proposed method can yield better overall prediction performance than HMM and NN based methods, particularly in terms of conservatism and non-defensiveness. This means that the prediction generated by the proposed method has similar criticality as the ground truth in the interaction process.

	HMM	NN	Hierarchical IRL
\mathcal{G}	0.0701	0.0563	0.1117
\mathcal{C}	0.0476	0.0493	0.0178
\mathcal{D}	0.1356	0.1303	0.0698
$\mathcal{B}_c = \mathcal{G} + \mathcal{C} + \mathcal{D}$	0.2551	0.2361	0.2053

TABLE II: Scores of different probabilistic prediction approaches

2) *Accuracy of the probabilistic prediction of continuous trajectories:* In this test, we generate the most probable trajectories under different discrete driving decisions by solving a finite horizon Model Predictive Control (MPC) problem using the above learned continuous cost functions. We show three illustrative examples in Fig. 4. The red dotted lines and blue solid lines represent, respectively, the predicted most-likely trajectories and the ground truth trajectories. The thick black dash-dot lines are trajectories of other vehicles. We can see that the predicted trajectories are very close to the ground truth ones.

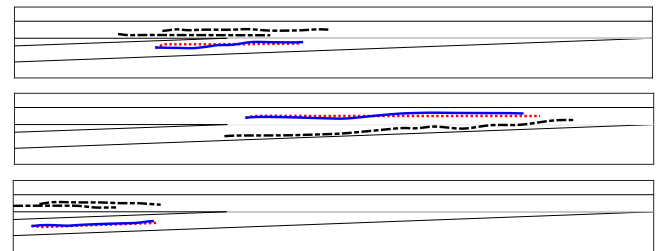


Fig. 4: Three illustrative examples of the predicted most probable trajectories (red dotted line) compared with the ground truth trajectories (blue solid line). Thick black dash-dot lines represent the trajectories of other vehicles except for the predicted one.

Metric: We also adopt Mean Euclidean Distance (MED) [27] to quantitatively evaluate the accuracy of the prediction for continuous trajectories. Given the ground truth trajectory $\xi_{\text{ground}} = [x_{g,1}, y_{g,1}, \dots, x_{g,L}, y_{g,L}]^T$ and predicted trajectory $\xi_{\text{prediction}} = [x_{p,1}, y_{p,1}, \dots, x_{p,L}, y_{p,L}]^T$ of same length L and same sampling time ΔT , the trajectory similarity is calculated as follows:

$$\mathcal{S}_{\text{MED}} = \frac{1}{L} \sum_{i=1}^L \| [x_{p,i}, y_{p,i}]^T - [x_{g,i}, y_{g,i}]^T \|_2 \quad (24)$$

Results: We test on 20 long trajectories in the test set, and results are summarized in Table III: the proposed hierarchical IRL method can achieve trajectory prediction with a mean MED of 0.6172m with standard deviation of 0.2473m.

	Mean (m)	Max (m)	Min (m)	Std (m)
MED	0.6172	1.0146	0.3644	0.2473

TABLE III: Trajectory similarities in terms of MED

V. CONCLUSION

In this work, we have proposed a probabilistic and interactive prediction approach via hierarchical inverse reinforcement learning. Instead of directly making predictions based on historical information, we formulate the prediction problem from a game-theoretic view: the distribution of future trajectories of the predicted vehicle strongly depends on the future plans of the host vehicle. To address such interactive behavior, we design a hierarchical inverse reinforcement learning method to capture the hierarchical trajectory-generation process of human drivers. Influences of both discrete and continuous driving decisions are explicitly modelled. We also have applied the proposed method on a ramp-merging driving scenario. The quantitative results verified the effectiveness of the proposed approach in terms of accurate prediction to both discrete decisions and continuous trajectories.

ACKNOWLEDGMENT

We thank Yeping Hu and Jiachen Li for their assistance with the implementation of the neural network (NN) and hidden Markov model (HMM) based prediction approaches in Section IV-D.

REFERENCES

- [1] Y. Kuwata, J. Teo, G. Fiore, S. Karaman, E. Frazzoli, and J. P. How, "Real-time motion planning with applications to autonomous urban driving," *IEEE Transactions on Control Systems Technology*, vol. 17, no. 5, pp. 1105–1118, 2009.
- [2] Z. Liang, G. Zheng, and J. Li, "Automatic parking path optimization based on bezier curve fitting," in *Automation and Logistics (ICAL)*, 2012 *IEEE International Conference on*. IEEE, 2012, pp. 583–587.
- [3] J. Ziegmann, J. Shi, T. Schnrer, and C. Endisch, "Analysis of individual driver velocity prediction using data-driven driver models with environmental features," in *Intelligent Vehicles Symposium (IV)*, 2017 *IEEE*. IEEE, 2017, pp. 517–522.
- [4] R. Graf, H. Deusch, F. Seeliger, M. Fritzsche, and K. Dietmayer, "A learning concept for behavior prediction at intersections," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, June 2014, pp. 939–945.

- [5] S. Lefèvre, C. Laugier, and J. Ibañez Guzmán, "Intention-aware risk estimation for general traffic situations, and application to intersection safety," *Inria Research Report*, no. 8379.
- [6] M. Schreier, V. Willert, and J. Adamy, "An Integrated Approach to Maneuver-Based Trajectory Prediction and Criticality Assessment in Arbitrary Road Environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2751–2766, Oct. 2016.
- [7] X. Geng, H. Liang, B. Yu, P. Zhao, L. He, and R. Huang, "A Scenario-Adaptive Driving Behavior Prediction Approach to Urban Autonomous Driving," *Applied Sciences*, vol. 7, no. 4, p. 426, Apr. 2017.
- [8] D. J. Phillips, T. A. Wheeler, and M. J. Kochenderfer, "Generalizable intention prediction of human drivers at intersections," in *Intelligent Vehicles Symposium (IV)*, 2017 *IEEE*. IEEE, 2017, pp. 1665–1670.
- [9] Y. Hu, W. Zhan, and M. Tomizuka, "Probabilistic prediction of vehicle semantic intention and motion," *arXiv preprint arXiv:1804.03629*, 2018.
- [10] C. Dong, J. M. Dolan, and B. Litkouhi, "Intention estimation for ramp merging control in autonomous driving," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, June 2017, pp. 1584–1589.
- [11] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions," in *Robotics: Science and Systems*, 2016.
- [12] R. E. Kalman, "When is a linear control system optimal?" *Journal of Basic Engineering*, vol. 86, no. 1, pp. 51–60, 1964.
- [13] A. Y. Ng, S. J. Russell, et al., "Algorithms for inverse reinforcement learning," in *ICML*, 2000, pp. 663–670.
- [14] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning," in *AAAI*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.
- [15] S. Levine and V. Koltun, "Continuous inverse optimal control with locally optimal examples," 2012.
- [16] S. Liu, M. Araujo, E. Brunskill, R. Rossetti, J. Barros, and R. Krishnan, "Understanding sequential decisions via inverse reinforcement learning," in *Mobile Data Management (MDM)*, 2013 *IEEE 14th International Conference on*, vol. 1. IEEE, 2013, pp. 177–186.
- [17] M. Shimosaka, K. Nishi, J. Sato, and H. Kataoka, "Predicting driving behavior using inverse reinforcement learning with multiple reward functions towards environmental diversity," in *Intelligent Vehicles Symposium (IV)*, 2015 *IEEE*. IEEE, 2015, pp. 567–572.
- [18] H. Kretschmar, M. Spies, C. Sprunk, and W. Burgard, "Socially compliant mobile robot navigation via inverse reinforcement learning," *The International Journal of Robotics Research*, vol. 35, no. 11, pp. 1289–1307, 2016.
- [19] D. Lenz, F. Diehl, M. T. Le, and A. Knoll, "Deep neural networks for markovian interactive scene prediction in highway scenarios," in *Intelligent Vehicles Symposium (IV)*, 2017 *IEEE*. IEEE, 2017, pp. 685–692.
- [20] A. Kesting, M. Treiber, and D. Helbing, "Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4585–4605, 2010.
- [21] L. Sun, W. Zhan, M. Tomizuka, and A. Dragan, "Courteous autonomous cars," in *2018 IEEE/RSJ International Conference on Intelligent Robots (IROS)*, to appear.
- [22] V. Alexiadis, J. Colyar, J. Halkias, R. Hranac, and G. McHale, "The Next Generation Simulation Program," *Institute of Transportation Engineers. ITE Journal; Washington*, vol. 74, no. 8, pp. 22–26, Aug. 2004.
- [23] W. Zhan, L. Sun, Y. Hu, J. Li, and M. Tomizuka, "Towards a fatality-aware benchmark of probabilistic reaction prediction in highly interactive driving scenarios," in *the 21st IEEE International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018.
- [24] G. W. Brier, "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, vol. 78, no. 1, pp. 1–3, 1950.
- [25] C. M. Bishop, "Mixture density networks," 1994.
- [26] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [27] J. Quehl, H. Hu, O. S. Tas, E. Rehder, and M. Lauer, "How good is my prediction? finding a similarity measure for trajectory prediction evaluation," in *2017 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)*, 2017, pp. 120–125.