# Behavior and Interaction-aware Motion Planning for Autonomous Driving Vehicles based on Hierarchical Intention and Motion Prediction

Dachuan Li[1,2,3], Yunjiang Wu[1,2,3], Bing Bai[1,2,3] and Qi Hao[†1,2,3]

***Abstract*— Safe motion planning in complex and interactive environments is one of the major challenges for developing autonomous vehicles. In this paper, we propose an interaction and behavior-aware motion planning framework based on joint predictions of intentions and motions of surrounding vehicles. A multimodal hierarchical Inverse Reinforcement Learning (IRL) framework is designed to learn joint driving pattern-intention-motion models from real-world interactive driving trajectory data. Using the learned models, the proposed approach can probabilistically predict the continuous motions partitioned by discrete driving styles and intentions, while taking into account of the interactions between participants' actions. A chance-constrained POMDP strategy is utilized to generate risk-bounded motion plans based on those predictions. Simulation case studies based on challenging scenarios demonstrate that the proposed framework is effective in predicting surrounding vehicles' behaviors and generating safe motion plans in complex traffic environments.**

## I. Introduction

Among the criteria of autonomous vehicles (AVs), safety is of the highest priority for the design and development of AVs. As current transportation systems still lack the mass deployment of V2V (Vehicle-to-Vehicle) technologies, the safe and interactive motion planning poses significant challenges to AVs in the absence of explicit inter-vehicle communications. To safely and efficiently navigate in highly complex driving environments with various, interacting traffic participants, AVs should effectively infer and integrate the future intentions and behavior of surrounding vehicles to make decisions and motion plans. Despite the previous research efforts devoted to the domain of behavior prediction (e.g. [1], [2]) and behavior-aware motion planning (e.g. [3]), AVs are still facing many technical challenges to achieve safe motion planning in real-world traffic environments with uncertainties. Firstly, the future behavior of surrounding traffic participants depends not only on their previous actions and states, but they are also affected by the potential actions taken by the ego vehicle. Therefore, the behavior predictions should be interactive, i.e. such predictions should account for the influence of the ego vehicle's actions on the future behavior of other participants. Secondly, the behavior and motions of surrounding vehicles are probabilistic and related to drivers' driving patterns (e.g. aggressive, conservative, moderate) and maneuver intentions, thus simple assumptions involving behavior models with fixed likelihood are unrealistic in real world scenarios. Thirdly, it is still an open question how to effectively integrate intention and motion predictions into ego maneuver and motion planning scheme, and to account for highly diverse and dynamic driving behavior patterns.

In this paper, we aim at addressing these challenges by proposing an interactive and behavior-aware motion planning scheme. The primary contributions of this paper include: 1) We propose **an interactive driving behavior model** that formalizes the distribution of surrounding vehicles' future intentions and motions conditioned not only on the historical trajectories, current behavior observations and the drivers' driving patterns, but also on the ego-vehicles next-step actions; 2) **A hierarchical driving pattern-intention-motion prediction framework** is designed based on IRL. The predictions of surrounding vehicles' future trajectories are formulated as continuous distributions partitioned by driving patterns and discrete maneuver decisions. The IRL scheme allows for a straightforward integration of predictions into POMDP-base planning framework; 3) **A risk-bounded motion planning strategy** is designed based on a chance-constrained POMDP (Partially Observable Markov Decision Process) which integrates the intention and behavior predictions. The motion planning strategy is able to efficiently generate safe ego maneuver and motion plans that accounts for the surrounding vehicle's behaviors and interactions among ego and agent vehicles.

The remainder of this paper is organized as follows: Section II provides a brief literature review of related work; Section III formulates the behavior prediction and motion planning problem; Section IV presents details of the IRL-based hierarchical intention-behavior prediction with Chance-constrained POMDP-based motion planning strategy in Section V; Section VI provides evaluation results and the paper is finally concluded in Section VII.

## II. Related Work

Considering how future behavior of surrounding road participants are predicted and how these predictions are integrated, existing intention/behavior-aware decision/motion planning strategies for AVs can be divided into two major categories: reactive planning and interactive belief planning.

**Reactive planning** refers to planning strategies that generate maneuver/motion plans based on future predictions of road participants' behavior, without considering the effect of

*This work was supported by the Opening Project of Shanghai Trusted Industrial Control Platform

†Corresponding author. Email: haoq@sustech.edu.cn

[1]Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

[2]Sifakis Research Institute for Trustworthy Autonomous Systems, Shenzhen 518055, China

[3]SUSTech-Haylion Center for Intelligent Transportation, Shenzhen 518055, China

the ego vehicle's future actions. Such planners can be further categorized into deterministic and probabilistic approaches. Deterministic reactive motion planning approaches are based on the assumption that the surrounding vehicles follow a fixed trajectory patter, e.g. maintaining constant velocity [4], [10] or driving with known goal states [6]. Such approaches are inapplicable to real world scenarios that normally involve uncertainties. In contrast, probabilistic reactive planners integrate uncertainties by formulating surrounding vehicles' behavior predictions as probabilistic distributions. [7] and [8] utilize a Gaussian prediction model incorporating driving style, intentions and trajectories, and a POMDP model is used to generate risk-bounded maneuver plans using these probabilistic predictions. However, such approaches ignore interaction among traffic participants and their effectiveness in highly interactive scenarios remains to be evaluated.

**Interactive belief planning** represents a large body of research work that plan over probabilistic distributions of states and account for the interactions between the ego vehicles and surrounding participants. A widely adopted paradigm of such planning approaches is POMDP-based joint prediction and planning framework. For instance, [9] utilizes a POMDP framework to estimate the behavior of agent vehicles given the planned trajectory of the ego-vehicle, and a multi-policy model predictive control-based planner generates safe trajectories for the ego vehicle. Similarly, [2] and [10] propose a unified continuous POMDP framework to determine the optimal acceleration of ego-vehicle by interactively predicting the motions of other vehicles, and the algorithm can run in a online and anytime manner. The drawback of such frameworks is that the behavior prediction of surrounding vehicles is generally based on simplified models, which can lead to high inaccuracy in real-world problems with various driving styles and behaviors. In addition, existing interactive planners fail to ensure a bounded-risk of collision, which requires integrating chance-constraints to the planner.

**IRL-based behavior prediction** approaches model driving behavior as a reward function, which can be inferred from behavior observations by assuming the demonstration driving trajectories guarantee optimal rewards. [11] formulates the driving behavior prediction model as multiple reward functions in an IRL framework to account for environment diversity. Similarly, [12] proposes a hierarchical IRL framework for predicting discrete driving decisions and continuous motions in interactive scenarios. Taking advantage of data-driven paradigm, IRL-based prediction methods have shown strong potential in real-world applications with environment and driving behavior diversity.

Inspired by the interactive belief planning methods and the work in [12], we build our joint prediction and planning scheme based on an extension of hierarchical IRL framework which is trained using real-world driving data. The proposed framework can provide predictions of driving patterns, intentions and motions, while accounting for interactions. Our POMDP planner integrates chance-constraints to bound the risk of collisions of the generated trajectories.

## III. Problem Statement and System Overview

### A. Problem Statement

Our work focuses on the generation of motion plans for an ego vehicle, (denoted as $N_E$), by predicting the probabilistic behaviors of surrounding vehicles (denoted as $\mathcal{N} = \{N_0...N_k\}, k \in \mathbb{N}_0$). As discussed previously, the future behavior of a agent vehicle depends by its historical trajectories, current observed states and possible actions taken by the ego vehicle. Denoting the observed, historical and future state of a vehicle as $\mathbf{x}_o$, $\mathbf{x}$ and $\hat{\mathbf{x}}$ respectively, the driving behavior prediction problem is defined as follows:

**Definition 1 (interactive driving behavior prediction)**: Given the historical trajectory ($\mathbf{x}_{N_P}$) of an agent vehicle $N_P$, as well as current observations of its surrounding vehicles $\{\mathbf{x}_{o,i}\}, i = 1...k$, the historical trajectories $\mathbf{x}_{N_E}$ and the potential future states $\hat{\mathbf{x}}_{N_E}$ of an ego vehicle, predict the probabilistic distribution of $N_P$'s states:

$$p(\hat{\mathbf{x}}_{N_P}|\{\mathbf{x}_{o,i}\}, \mathbf{x}_{N_P}, \mathbf{x}_{N_E}, \hat{\mathbf{x}}_{N_E}), (i = 1...k). \tag{1}$$

where $i$ denotes the index of surrounding vehicles. As in [12], the interaction between $N_E$ and $N_P$ is integrated by modelling the future state of $N_P$ as probability distribution conditioned by future actions of $N_E$. Therefore, the future belief state of vehicles $\{\mathcal{B}_i\}, (i = 1...k)$ can be obtained using Eq. 1. Given the predicted future belief of surrounding vehicles, the risk-bounded optimal motion planning of the ego-vehicle aims at generating an optimal motion plan which guarantees the operation risk of near-collision is below a threshold, and it can be formulated as:

**Definition 2 (risk-bounded optimal motion planning)**: Given the predicted belief of surrounding vehicles' future states $\{\mathcal{B}_i\}, (i = 1...k)$ and a set of actions $\mathcal{A}$ of an ego vehicle $V_E$, generate a motion plan $\pi^*$ that minimizes the expected cost along the trajectory:

$$\pi^* = \arg\min_{\pi} E\left[\sum_{\tau}^{\tau+\Delta t} C(\pi_\tau)\Big|\pi, \{\mathcal{B}_i\}\right] \tag{2}$$

while satisfying the chance constraint:

$$\varrho\left(\{\mathcal{B}_i\}|\pi^*\right) = 1 - P\left(\bigwedge_{\tau}^{\tau+\Delta t} \neg\chi_{coll,\tau}\Big|\pi^*, \{\mathcal{B}_i\}\right) \leqslant \epsilon \tag{3}$$

where $C(\cdot)$ denotes the cost function associated with a motion plan, and $\Delta t$ is the temporal planning horizon. $\varrho$ represent the risk function corresponding to a given $\pi$, where $\chi_{coll}$ is a function indicating near-collision incidents, and $\epsilon$ is the upper-bound of operation risk.

### B. System Overview

Fig.1 illustrates the overview of the proposed framework, which primarily consists of three parts: offline learning and modeling, online driving behavior prediction and online risk-bounded motion planning.

The **offline learning and modeling** phase uses real-world demonstration trajectory data to establish a hierarchical driving behavior model consisting of driving pattern, discrete

maneuver and continuous motions (Fig.2). In the proposed framework, the top layer of the driving behavior is the driving pattern indicating the driver's operation style or preference (denoted by $\mathbf{x}_p, \mathbf{x}_p \in \mathcal{P}$, $\mathcal{P}$ is the set of semantic labels including aggressive, conservative and moderate). As the way a driver performs maneuver depends heavily on his/her driving pattern, each driving pattern is corresponded with several maneuver protocol models based on Probabilistic Hybrid Automaton (PHA) as in [7]. A maneuver protocol model is associated with a scenario-related maneuver sequence (e.g., lane changing, merging, overtaking, etc.) is specified by $\mathcal{P}_p = \langle \mathbf{x}_c, \mathbf{x}_d, \{\mathbf{x}_{M,i}\}, \{(\mathcal{T}_j, c_j)\}\rangle, (i, j \in \mathbb{N}_0)$, where $\mathbf{x}_c$ is the continuous states, $\{\mathbf{x}_{M,i}\}$ indicates the set of maneuvers involved in the protocol, and $\{(\mathcal{T}_j, c_j)\}$ is the transition probability function between different maneuvers. $\mathbf{x}_M \in \mathcal{M}$ where $\mathcal{M}$ is the set of discrete maneuvers that determines the homotopy pattern of continuous motions. Therefore, the continuous motion is formulated as distribution specified by $p(\hat{\mathbf{x}}_{N_P}|\mathbf{x}, \hat{\mathbf{x}}_{N_E})$, and partitioned by different discrete maneuvers: $p(\hat{\mathbf{x}}_{N_P}|\mathbf{x}, \hat{\mathbf{x}}_{N_E}) = \sum_{\mathbf{x}_{M,i}\in\mathcal{M}} p(\hat{\mathbf{x}}_{N_P}|\mathbf{x}_{M,i}, \mathbf{x}, \hat{\mathbf{x}}_{N_E}) P(\mathbf{x}_{M,i}|\mathbf{x}, \hat{\mathbf{x}}_{N_E})$ [12]. Therefore, the driving pattern model, the transition model of discrete maneuvers ($\mathcal{T}(\mathbf{x}_{M,j})$) and the continuous motion model ($p(\hat{\mathbf{x}}_{N_P}|\mathbf{x}, \hat{\mathbf{x}}_{N_E})$) are pre-learned offline via IRL using real-world demonstration driving data $\mathcal{D}$.

Given the pre-learned driving behavior model and the observed states of surrounding vehicles, the **online driving behavior prediction** function predicts each agent vehicle's driving pattern, discrete maneuver and continuous motion in a future time horizon. Firstly, the driving pattern $\mathbf{x}_p$ of each agent vehicle is identified, and the corresponding maneuver protocol model is selected based on the driving pattern, observed states and map data. Given the selected data, the future maneuver $\hat{\mathbf{x}}_M$ can be estimated using the transition model included $\mathcal{T}(\mathbf{x}_M)$ in the selected maneuver protocol model $\mathcal{P}_p$. The future continuous motions $\hat{\mathbf{x}}_{N_P}$ are then predicted using the pre-learned distribution model via Bayesian prediction, details of the procedure will be specified in the next section.

The **online risk-bounded motion planner** utilizes a POMDP framework to generate optimal, risk-bounded motion plans to navigate the ego vehicle to a goal state in a future horizon, given the predicted belief of surrounding vehicles' future states and static map data. Observations of ego vehicle and surrounding vehicles' state after executing the motion plans will be obtained and used in the next prediction and planning circle. Details of the motion planning strategy can be found in Section V.

## IV. Joint Intention and Behavior Prediction based on Hierarchical IRL

### A. Driving Behavior Modeling via Hierarchical Inverse Reinforcement Learning

As described in section III B, the proposed framework adopts a hierarchical driving behavior model consisting of driving patterns, discrete driving decisions and continuous trajectory (Fig.2). At the top layer, the motion of a vehicle

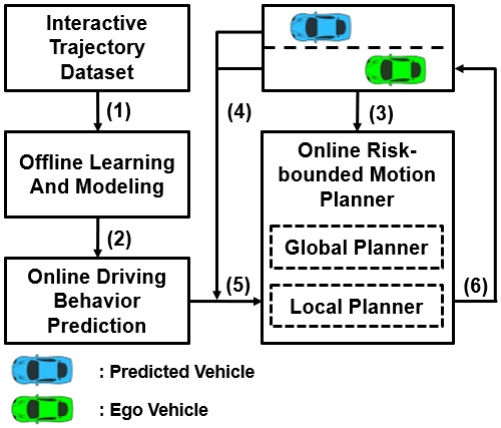


Fig. 1. The framework of interactive prediction and behavior-aware planning. (1) Collected trajectory sets. (2) Learned hierarchical driving behavior models. (3) Initial states, goal state and map information of the ego vehicle. (4) Historical trajectories and current behavior observations. (5) The predicted driving pattern, discrete maneuver decision and continuous trajectory for other vehicles. (6) Planned maneuver and motion for the ego vehicle in a future time horizon.

is largely determined by the manner its driver operates the vehicle, and it also affects how the driver makes decisions in the context of driving scenario. Once the intention is inferred, the corresponding future motions can be predicted accordingly. Given a demonstration trajectory set $\mathcal{D}$ ( $\mathcal{D} = \{\xi_1 ... \xi_N\}$), the task of driving behavior modeling is to approximate probabilistic models of driving behavior and model based on $\mathcal{D}$. Details of modeling each layer of the model are as follows:

**1) Driving Pattern Classification Model**

The objective of driving pattern classification modeling is to establish a model from $\mathcal{D}$ to generate probability over driving patterns adopted by the driver, given driving features extracted from recent observations. In this work, the driving patterns are divided into three categories: **aggressive, conservative and moderate**, e.g. $\mathcal{P} = \{\mathbf{AGG}, \mathbf{CON}, \mathbf{MOD}\}$ Since labeling all trajectories in $\mathcal{D}$ is time-consuming due to the volume of the dataset, we adopt a unsupervised clustering approach (EM, Expectation Maximization [13]) to approximate cluster labels of demonstration trajectories.

Denote $\mathbf{\Theta} = (\rho_1, \rho_2, \rho_3, \theta_{c,1}, \theta_{c,2}, \theta_{c,3})$ as the vector that needs to be estimated, where $\rho_i$ and $\theta_{c,j}, (i, j = 1, 2, 3)$ are the probability and reward weight corresponding to each driving pattern in $\mathcal{P}$, respectively. Denote $p_{i,j}$ as the probability that trajectory $\xi_i \in \mathcal{D}$ adopts driving pattern $\mathbf{x}_{p,j} \in \mathcal{P}$. Given $\mathcal{D}$ and $\mathcal{P}$, The EM approach generates estimate of $\mathbf{\Theta}$ by altering between Expectation step and Maximization step. In the Expectation step, the current probability $p^t_{i,j}$ is calculated by:

$$p^t_{i,j} = \prod_{s\in\xi_i} \pi_{\theta_{j,t}}(s)\rho_{t,j}/H \tag{4}$$

where $\pi_{\theta_{j,t}}(s)$ is the Boltzmann exploration policy [13] of trajectory $\xi_j$ with respect to $\theta_j$, and $H$ is the normalization factor. In the maximization step, the probability $\rho_l$ of each

cluster is updated by:

$$\rho_l^{t+1} = \sum_i p_{il}^t / N \tag{5}$$

and the reward weight can be obtained by maximizing the likelihood function:

$$\theta_l^{t+1} = \arg\max_{\theta} \sum_{i=1}^{N} \log P(\xi_i|\theta_l) p_{il}^t \tag{6}$$

where the probability $P(\xi_i|\theta_l)$ in Eq. 6 is specified by:

$$P(\xi_i|\theta_l) = \frac{e^{-\theta_l^T \phi(\xi_i)}}{\int e^{\theta_l^T \phi(\xi)} d\xi} \tag{7}$$

where $\phi(\xi)$ denotes the feature function, and we adopt the Maximum Entropy IRL approach to estimate reward weights that maximize likelihood of the demonstration trajectories as specified in Eq. 6, using gradient descent approach. In contrast to [12], the following features are selected to parameterize the driving pattern: acceleration $\psi_{acc}$, jerks $\psi_{jerk}$, clearance to surrounding participants $\psi_{dist}$, courtesy $\psi_{court}$ and an additional yaw rate feature ($\psi_{yaw_r} = \sum_t \dot{\psi}_t^2$) Therefore, given an observed trajectory $\xi_o^{1:k}$, we can infer the probability of driving pattern that this trajectory adopts based on Bayesian rules:

$$P(\theta_i|\xi_o) = P(\xi_o|\theta_i) P(\theta_i) / P(\xi_o) \tag{8}$$

where $P(\xi_o|\theta_i)$ can be calculated using Eq. 4, and $P(\theta_i) = \rho_i$. $P(\xi_o)$ is a normalization factor.

As described in Section III B, for general driving scenarios from real world, maneuver protocol models ($\mathcal{P}_p$) are established by associating driving pattern to PHA models. Take driving through a roundabout as an example, the protocol model can be specified by:

$$\begin{aligned}\mathcal{P}_{rndabt} =& \langle \mathbf{x}_c, \mathbf{x}_d, \{\mathbf{x}_{enter}, \mathbf{x}_{exit}, \mathbf{x}_{yield}, \mathbf{x}_{drivethr}\}, \\ & \{(\mathcal{T}_j, c_j)\}\rangle\end{aligned} \tag{9}$$

where $\{\mathbf{x}_{enter}, \mathbf{x}_{exit}, \mathbf{x}_{yield}, \mathbf{x}_{drivethr}\}$ denote the maneuvers that can be executed by the vehicle (enter/exit the roundabout, yield to a entering vehicle and keep driving through). $\{(\mathcal{T}_j, c_j)\}$ denote the probability transitions among discrete maneuvers, and each transition specifies the switch between different maneuvers with probability $p_{jk}$, once a guarded condition $c_j$ is met. These transition functions are determined by the driving pattern. For instance, a conservative driver in a roundabout will yield with high probability to a car entering the roundabout, while the probability of maintaining driving through may be high for an aggressive driver.

**2) Discrete driving maneuver model**

Maneuver models are corresponded with discrete driving decisions the driver may make, and they consequently determine the homotopy class of continuous trajectories at the lower level. Similar to [12], We also adopt the maximum entropy IRL approach to model the driving discrete maneuver. Under the maximum entropy principle and the assumption that drivers tend to make decisions with low cost, the objective of IRL is to find reward weights that maximize the probability of demonstration trajectories $\mathcal{D}$:

$$\begin{aligned}\theta_d^* &= \arg\max_{\phi} P(\mathcal{D}|\phi) \\ &= \arg\max_{\phi} \prod_{\mathbf{x}_M \in \mathcal{D}} \frac{e^{-\phi_d^T \mathbf{f}_d(\mathbf{x}_M)}}{\sum_{\tilde{\mathbf{x}}_M \in \mathcal{D}} e^{-\phi_d^T \mathbf{f}_d(\tilde{\mathbf{x}}_M)}}\end{aligned} \tag{10}$$

where $\phi_d^T \mathbf{f}_d$ is the cost function parameterized by selected feature vector $\mathbf{f}_d$, which consists of rotation angle $f_\angle$ and minimal cost of continuous trajectories $f_{cost}$ [12], as well as an additional acceleration feature $f_{acc}$, as some scenario (e.g. roundabout and ramp merging) involve description of yield or passing decisions. The gradient descent approach is utilized to solve the above optimization problem.

**3) Continuous motion model**

Once the discrete driving maneuver is recognized, trajectories in the demonstration set $\mathcal{D}$ can be grouped based on their driving maneuvers homotopy: $\mathcal{D} = \{\mathbf{D}_{\mathbf{x}_M}\}, \mathbf{x}_M \in M$, and each demonstration trajectory in $\mathbf{D}_{\mathbf{x}_M}$ is specified by $(\mathbf{x}_{i,M}, \xi_{N_E}, \xi_{N_P}, \{\xi_{o,j}\}, \hat{\xi}_{N_E})$. For each maneuver homotopy demonstration set, we utilize continuous maximum entropy IRL to compute the optimal reward weights that maximize the log likelihood of demonstration trajectory:

$$\begin{aligned}\theta^*_{c,\mathbf{x}_{i,M}} =& \arg\max_{\theta_{c,\mathbf{x}_{i,M}}} log P(\mathcal{D}_{\mathbf{x}_{i,M}}|\theta_{c,\mathbf{x}_{i,M}}) = \\ & \sum_{\hat{\xi}_{M,i} \in \mathcal{D}_{\mathbf{x}_{i,M}}} log \frac{e^{-\theta_{M,i}^T \mathbf{f}_{M,i}(\hat{\xi}_{N_P})}}{\int e^{-\theta_{M,i}^T \mathbf{f}_{M,i}(\tilde{\xi}_{N_P})} d\tilde{\xi}_{N_P}}\end{aligned} \tag{11}$$

selected features vector $\mathbf{f}_{M,i}$ in the cost function include speed $f_v$, traffic following $f_{IDM}$, comfortness $f_{jerk}$, and goal-reaching $f_g$ [12]. Again, we adopt gradient descent approach to solve the optimization problem specified by Eq. 11, details of the algorithm can be found in [14]. Note that interaction between the ego vehicle and predicted agent is integrated by including $\hat{\xi}_{N_E}$ in the modeling process, as discussed previously.

### *B. Interactive Intention and Behavior Prediction*

Given observed trajectory $\xi_{o,N_P}^{1:k}$ of a surrounding target vehicle $N_P$ in a previous time horizon $[1, k]$, its probabilistic driving behavior can be predicted based on the learned hierarchical driving behavior model, using Bayesian prediction approach.

Firstly, the probability $P(\mathbf{x}_{p,N_E}|\xi_{o,N_P}^{1:k})$ of the driving pattern adopted by $\xi_{o,N_P}^{1:k}$ is identified using Eq. 8, where $P(\mathbf{x}_{p,N_E}|\xi_{o,N_P}^{1:k}) = P(\theta_p|\xi_{o,N_P}^{1:k})$. The identified driving pattern is then combined with the current driving scenario to select the current maneuver protocol based on PHA (the current driving scenario can be identified using map information and observations of surrounding vehicles).

Given the current maneuver protocol model and $\xi_{o,N_P}^{1:k}$, the probability distribution of discrete maneuver of $N_P$ in

the next time step $(k+1)$ can be estimated by:

$$p(\mathbf{x}_{M,N_P}^{k+1}|\mathbf{x}_{p,N_P},\xi_{o,N_P}^{1:k}) = \sum_{\mathbf{x}_{M,N_E}^{k}} p(\mathbf{x}_{M,N_P}^{k+1}|\mathbf{x}_{M,N_P}^{k},\xi_{o,N_P}^{1:k},\mathbf{x}_{p,N_E})p(\mathbf{x}_{M,N_P}^{k}|\mathbf{x}_{p,N_E},\xi_{o,N_P}^{1:k}) \tag{12}$$

where $p(\mathbf{x}_{M,N_P}^{k+1}|\mathbf{x}_{M,N_P}^{k},\xi_{o,N_P}^{1:k},\mathbf{x}_{p,N_E})$ is the transition probability specified in the PHA maneuver protocol model, and $p(\mathbf{x}_{M,N_P}^{k}|\mathbf{x}_{p,N_E},\xi_{o,N_P}^{1:k})$ is the estimation of the previous step. Therefore, the conditional distribution of continuous motion of vehicle can be predicted by:

$$p(\mathbf{x}_{c,N_P}^{k+1}|\xi_{o,N_P}^{1:k},\{\xi_{o,N_i}^{1:k}\},\xi_{N_E}^{k+1}) = \sum_{\mathbf{x}_{M,i}\in M}\sum_{\mathbf{x}_{p,j}\in\mathcal{P}} [P(\mathbf{x}_{p,j,N_P}|\xi_{o,N_P}^{1:k})p(\mathbf{x}_{M,i,N_P}^{k+1}|\mathbf{x}_{p,N_P},\xi_{o,N_P}^{1:k}) p(\mathbf{x}_{c,N_P}^{k+1}|\mathbf{x}_{M,i,N_P}^{k+1},\{\xi_{o,N_i}^{1:k}\},\xi_{N_E}^{k+1})] \tag{13}$$

where $p(\mathbf{x}_{c,N_P}^{k+1}|\mathbf{x}_{M,i,N_P}^{k+1},\{\xi_{o,N_i}^{1:k}\},\xi_{N_E}^{k+1})$ is the distribution of continuous motion model learned offline.

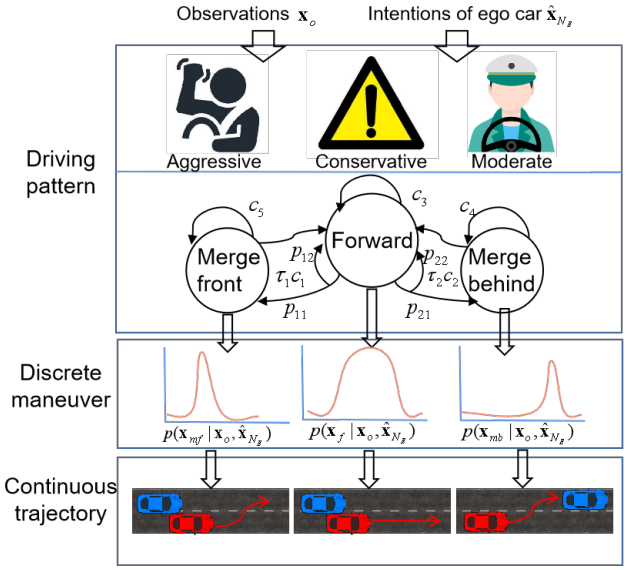


Fig. 2. Hierarchical driving behavior model.

## V. Risk-bounded Motion Planning

### A. Chance-Constrained POMDP Model

The risk-bounded motion planning problem for a ego vehicle $N_E$ defined by definition 2 in Section III A can be further modeled as tuple $\Pi = \langle \mathbf{S}, \mathbf{A}, \mathbf{O}, T, O, R, \epsilon \rangle$, where:

- $\mathbf{S}$ is the state space, and the states of the target vehicle include the position, heading and discrete driving decisions.
- $\mathbf{A}$ is the action space consisting of driving maneuvers that can be taken by $N_E$.
- $\mathbf{O}$ is the observation space which includes information captured by sensors. In this framework, the position, heading and velocity of both $N_E$ and surrounding vehicles within the sensing range are assumed to be observable, while the driving intentions of other vehicles are not directly observable.
- $T$ denote the transition model: $T = p(\mathbf{x}_{k+1}|\mathbf{a}_k,\mathbf{x}_k)$, and it is updated in the POMDP framework using the driving behavior predictions.
- $O$ is the observation model $(O = p(\mathbf{o}_k|\mathbf{x}_{k+1},\mathbf{a}_k))$ and it is specified by the driving behavior prediction discussed in Section IV.
- $R$ is the reward function which aims at navigating the ego vehicle towards the goal efficiently, ensuring passenger comfortness (by punishing unnecessary acceleration) and near-avoiding collision.
- $\epsilon$ is the upper bound of near-collision risk along the planned policy, as specified in Eq. 3.

### B. Solving the POMDP

In the proposed framework, the chance-constrained POMDP is solved based on an extension of the Adaptive Belief Tree (ABT) [15] (termed as CC-ABT). The basic idea of the ABT is to approximate the solution of POMDP by sampling episodes to construct a tree of reachable belief.

Given a initial belief $\mathcal{B}_0$ generated by the driving behavior prediction step, the CC-ABT incrementally builds a belief tree by rejection sampling. Each node $G_i$ of the tree represents a belief reachable from its parent node, and the edge between nodes represent the process of taking an action and generating the observations. For each node $G_i$, we estimate its Q-value:

$$Q(\mathcal{B}_k,\mathbf{a}_k) = \sum_{\mathbf{s}_k} R(\mathbf{s}_k,\mathbf{a}_k)\mathcal{B}_{\mathbf{s}_k} + \sum_{o_{k+1}} P(o_k|a_k,\mathcal{B}_k)\tilde{Q}(\mathcal{B}_{k+1}) \tag{14}$$

which is the summation of the current reward and the expected reward by execution the current optimal policy. Therefore, the approximation of optimal policy $\hat{\pi}$ can be estimated by maximizing the cumulative Q function:

$$\hat{\pi}(\mathcal{B},\mathbf{a}) := \arg\max_{\mathbf{a}\in\mathbf{A}} \hat{Q}(\mathcal{B},a) \tag{15}$$

Therefore the optimal policy $\hat{\pi}$ can be generated by performing heuristic tree search on the belief tree, using the Q-value as the heuristic. In addition, to satisfy the chance constraint specified in Eq. 3, the operation risk is integrated as a heuristic in the tree search. Given the belief $\mathcal{B}_k$ and policy $\pi$ at a certain step, the operation risk can be propagated recursively [7] by:

$$\varrho(\mathbf{B}_k|\pi) = r_{\mathcal{B}}(\mathcal{B}_k) + (1-r_{\mathcal{B}}(\mathcal{B}_k)) \sum_{\mathbf{o}_{k+1}\in\mathbf{O}} P_{safe}(\mathbf{o}_{k+1}|\pi(\mathcal{B}_k),\mathcal{B}_k)\varrho(\mathcal{B}_{k+1}|\pi) \tag{16}$$

where the first term is the risk (probability of near-collision) at $\mathbf{B}_k$, and the second term is the risk of the horizon planning starting from $k+1$, by following policy $\pi$, calculation of the risk can be found in [7]. The above operation risk works as an important heuristic in the CC-ABT solver. Firstly, the operation risk is evaluated along the edges during the belief

construction, and child nodes leading to high operation risk are pruned. Secondly, the operation risk is used during policy generation for selecting actions to bound the risk of near-collisions

## VI. Evaluation

### A. Implementation Details and Case Study

**1) Dataset**

We collect driving data from the INTERACTION dataset [16], which contains naturalistic motions of various traffic participants in a variety of highly interactive and complex driving scenarios such as roundabouts, multi-lane highway and unprotected intersections. The sampling time of each trajectory is 0.1s. To evaluate the performance of our driving behavior prediction approach, we select 182 and 108 trajectories from two scenarios, the multi-lane highway and roundabout, respectively, as the training dataset for IRL-based driving behavior modeling. 46 and 30 trajectories are selected from the two scenarios as the testing dataset.

Fig.3 shows the semantic map and the RGB pixel map for multi-lane highway driving and roundabout scenarios. Fig.4 shows two groups example of trajectory in multi-lane highway. The upper figure is a trajectory set of passing maneuver (in red lines) and merging back maneuver (in blue lines). The lower figure is a trajectory set of yielding maneuver (in red lines) and merging front maneuver (in blue lines).

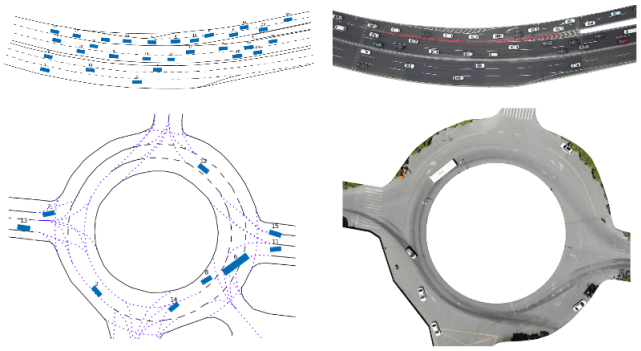

Fig. 3. Examples of driving data from highway and roundabout scenarios in the training dataset.

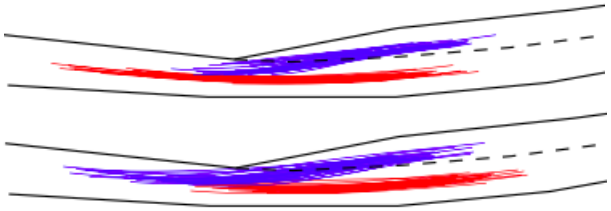

Fig. 4. Selected interactive trajectory set in the highway merging scenario. Upper figure: a set of trajectories of passing maneuver (in red) and rear merging maneuver (in blue). Lower figure: a set of trajectories of yielding maneuver (in red) and front merging maneuver (in blue).

**2) Simulation scenario setup** In order to evaluate the proposed risk-bounded motion planning approach in complex traffic environments, we setup a highway ramp-merging scenario and a roundabout scenario based on the open-source CARLA simulator. As shown in Fig.7 and Fig.8, each scenario involves three vehicles. The surrounding vehicles (in red vehicle and yellow) are driving in the same lane, while the ego vehicle (in black) is attempting to merge into the lane the surrounding vehicles are driving in. Therefore, to generate a safe decision and motion planning, the ego vehicle must predict the future driving pattern, discrete maneuver decision and continuous motion trajectories of surrounding vehicles and take account of the interactions among them. The state of surrounding vehicles can be obtained via simulated sensors from CARLA, and we have full access to the control input (steering wheel angle and throttle) of the agent vehicles.

### B. Testing Results

**1) Driving behavior prediction results** In the highway merging scenario, the proposed hierarchical IRL approach is applied to model the driving pattern, intention and behavior for both lane-keeping vehicles and merging vehicles. The time horizon of each trajectory for training is 6s, i.e. both the x-coordinate and y-coordinate vectors extracted from the dataset contain 60 datapoints.

Fig.5 illustrates examples of the trajectory prediction results in the highway merging scenario. The green and red dotted line, represent the ground truth and predicted trajectory, respectively. For the upper figure, the predicted driving pattern is moderate and discrete driving decision is passing. For the lower figure, the driving pattern is moderate while discrete driving decision is merging from behind.

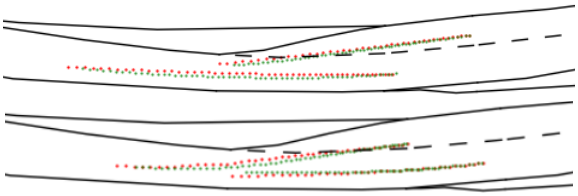

Fig. 5. Examples of the predicted trajectories (red dotted line) versus ground truth (green dotted line) in the highway merging scenario.

Fig.6 illustrates examples of behavior prediction results in the roundabout scenario. Table I shows the accuracy of motion prediction of the two scenarios. The accuracy is evaluated in terms of the mean Euclidean distance between the predictions and the ground truth. As can be seen from Table I, the mean errors in highway merging and roundabout are 0.46m and 0.58m, respectively.

TABLE I

Statistics of trajectory prediction errors in two testing scenarios.

| Driving Scenarios | Mean (m) | Max (m) | Min (m) | Std (m) |
|---|---|---|---|---|
| Highway Merging | 0.4602 | 0.9502 | 0.1333 | 0.5527 |
| Roundabout | 0.5804 | 1.2002 | 0.1873 | 0.5868 |

We compare the proposed hierarchical IRL-based prediction approach with the baseline approach in [7]. The

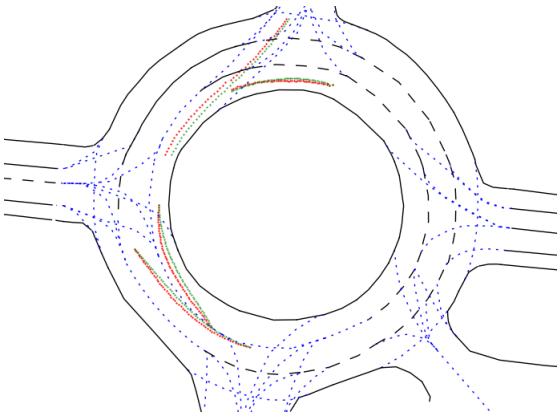

Fig. 6. Examples of the predicted trajectories (in red dotted line) versus the ground truth (in green dotted line) in the roundabout scenario.

TABLE II

COMPARISON OF PREDICTION ACCURACY RATES OF DRIVING PATTERNS, DISCRETE MANEUVER DECISIONS AND CONTINUOUS TRAJECTORIES

| Metric | Baseline | Hierarchical IRL |
|---|---|---|
| Driving Patterns (Prediction Accuracy) | 64.07% | **80.31%** |
| Discrete Maneuver Decision (Prediction Accuracy) | 68.72% | **90.87%** |
| Continuous Motion Trajectory (Prediction Mean Error) | 1.3112m | **0.5203m** |

results are shown in Table II . Our approach outperforms the baseline approach in terms of prediction accuracy of driving pattern, discrete maneuver decision and continuous motion trajectory (in terms of mean error).

**2) Risk-bounded Planning results** The proposed risk-bounded planning approach is evaluated in the CARLA simulation environment, the evaluation is based on two typical complex scenarios involving interactions: ramp-merging and roundabout. Fig.7 and Fig.8 illustrate examples of planning results in the two scenarios. For the ramp-merging case, the ego vehicle (in black) attempts to merging from a ramp to a lane where two surrounding vehicles (in yellow and red) are driving in. The ego vehicle firstly predicts that the intention of the leading vehicle (in yellow) is "passing" (Fig.7 (2)), so it decelerates and tries to merge behind the leading vehicle. Given the observation of the following vehicle (in red), the ego vehicle predicts that the red vehicle will decelerate to yield based on the interaction (Fig.7 (4)), and it finally generate a safe maneuver plan and merged between the two vehicles successfully (Fig.7 (5),(6)). Similarly, in the roundabout scenario, the ego vehicle (in black) tries to enter the roundabout where two agent vehicles (in yellow and red) are already in. The ego vehicle predicts that the leading vehicle will not yield but the following vehicle will decelerate, so the ego vehicle generates a safe plan to first yield to the yellow vehicle and then enter the roundabout between the two vehicles (Fig.8 (2), Fig.8 (3)).

To further evaluate the effect of the choice of risk bound values $\epsilon$ on the success rate and completion time of maneuvers, we conduct simulation experiment with different choice

TABLE III

STATISTICS OF PLANNING PERFORMANCE WITH DIFFERENT RISK BOUND SETTINGS IN HIGHWAY MERGING SCENARIO.

| Methods | Success Rate (%) | Completion Time (s) |
|---|---|---|
| Conservative | 98.62 | 8.46 |
| Risky | 26.40 | **4.52** |
| Risk-bounded ($\epsilon$ = 1%) | 98.14 | 7.26 |
| Risk-bounded ($\epsilon$ = 0.1%) | **99.57** | 7.54 |

TABLE IV

STATISTICS OF PLANNING PERFORMANCE WITH DIFFERENT RISK BOUND SETTINGS IN ROUNDABOUT SCENARIO.

| Methods | Success Rate (%) | Completion Time (s) |
|---|---|---|
| Conservative | 98.64 | 5.15 |
| Risky | 28.81 | **3.06** |
| Risk-bounded ($\epsilon$ = 1%) | 98.25 | 4.37 |
| Risk-bounded ($\epsilon$ = 0.1%) | **99.62** | 4.52 |

of risk bound based on the above mentioned scenarios. Two corner cases (conservative) and two risk bound settings (i.e. $\epsilon = 1\%$ and $0.1\%$) are considered for the evaluation, and 100 trials are conducted for each setting. Table III and Table IV summarizes the averaged statistics of the highway-merging and roundabout scenario, respectively. The conservative setting refers to cases where the ego-vehicle assume equal probability for each maneuver prediction of the predicted vehicle, while risky cases represent the ego vehicle assumes surrounding vehicles will yield to it with $100\%$ probability. As shown in the tables, the extreme cases indicate that more risky policies result in faster completion of maneuvers, but they also lead to high risk of accidents. Statistics of different risk bound settings show that strict risk tolerance (i.e. smaller $\epsilon$) can guarantee safety (i.e. high success rate) at the cost of increased time to complete the maneuver, but the proposed approach still achieves a desirable trade-off between safety and efficiency.

## VII. CONCLUSION

This paper presents a joint driving behavior prediction and motion planning framework for autonomous vehicles in complex and interactive environments. To integrate the influence of ego vehicle's action on the future behaviors of surrounding vehicles, we propose a hierarchical IRL-based framework to model the driving behavior at three levels: driving pattern, discrete maneuver decision and continuous motion trajectory. The predicted driving behavior is integrated in a chance-constrained POMDP scheme to generate maneuver and motion plans that can bound the overall operation risk. Simulation results show that the proposed approach can achieve high accuracy in predicting driving behaviors involving interactions, and can generate safe plans in complex scenarios. Future work includes addressing noisy observations and validating the approach in more complex scenarios. We will also evaluate the performance of the proposed approach in crowded driving scenarios involving larger number of traffic participants. In addition, a more comprehensive comparison with other counterpart prediction

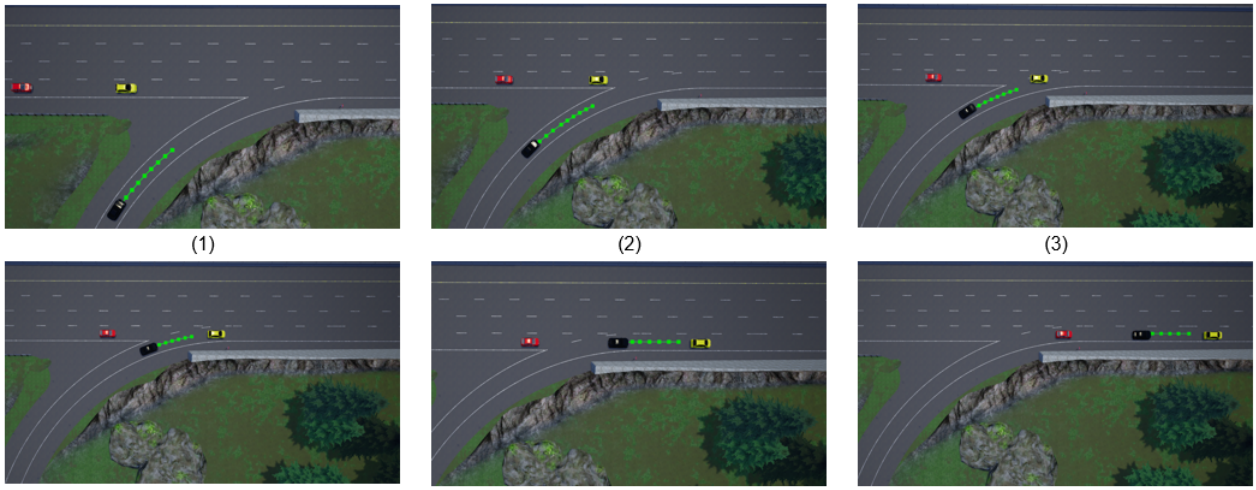


Fig. 7. Example of risk-bounded planning results in a highway merging driving scenario. The ego vehicle (in black) predicts the driving behaviors of two surrounding vehicles (in yellow and red) and generates a safe plan to merge successfully between the two vehicles.


Fig. 8. Example of risk-bounded planning results in a roundabout driving scenario. The ego vehicle (in black) predicts the driving behaviors of two surrounding vehicles (in yellow and red) and generates a safe plan to enter the roundabout between the two vehicles.

and planning approaches will be conducted.

## REFERENCES

[1] LEFÈVRE, Stéphanie; VASQUEZ, Dizan; LAUGIER, Christian. A survey on motion prediction and risk assessment for intelligent vehicles. ROBOMECH journal, 2014, 1.1: 1. 2

[2] Brechtel, Sebastian, Tobias Gindele, and Rüdiger Dillmann. "Probabilistic decision-making under uncertainty for autonomous driving using continuous POMDPs." 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, 2014.

[3] Paden, B., Cap, M., Yong, S. Z., Yershov, D., Frazzoli, E. A survey of motion planning and control techniques for self-driving urban vehicles. IEEE Transactions on Intelligent Vehicles. 2016, 1(1):33–55.

[4] Ajanovic, Z., Lacevic, B., Shyrokau, B., Stolz, M., Horn, M. Search-based optimal motion planning for automated driving. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018. p. 4523-4530. 4

[5] C. Hubmann, M. Aeberhard, and C. Stiller, A generic driving strategy for urban environments, in Proc. IEEE Int. Conf. Intell. Transp. Syst., 2016, pp. 1010–1016.

[6] Bansal, S., Cosgun, A., Nakhaei, A., Fujimura, K. Collaborative planning for mixed-autonomy lane merging. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018. p. 4449-4455. 6

[7] Huang, X., Hong, S., Hofmann, A., Williams, B. C. Online Risk-Bounded Motion Planning for Autonomous Vehicles in Dynamic Environments. In: Proceedings of the International Conference on Automated Planning and Scheduling. 2019. p. 214-222. 7

[8] Huang, Xin Xin Cyrus. Safe intention-aware maneuvering of autonomous vehicles. 2019. PhD Thesis. Massachusetts Institute of Technology.

[9] Zhou, B., Schwarting, W., Rus, D., Alonso-Mora, J. Joint multi-policy behavior estimation and receding-horizon trajectory planning for automated urban driving. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018. p. 2388-2394.

[10] Hubmann, C., Schulz, J., Becker, M., Althoff, D., Stiller, C. Automated driving in uncertain environments: Planning with interaction and uncertain maneuver prediction. IEEE Transactions on Intelligent Vehicles, 2018, 3.1: 5-17.

[11] Shimosaka, M., Nishi, K., Sato, J., Kataoka, H. Predicting driving behavior using inverse reinforcement learning with multiple reward functions towards environmental diversity. In: 2015 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2015. p. 567-572.

[12] SUN, Liting, ZHAN, Wei, TOMIZUKA, Masayoshi. Probabilistic prediction of interactive driving behavior via hierarchical inverse reinforcement learning. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018. p. 2111-2117.

[13] Babes, M., Marivate, V., Subramanian, K., Littman, M. L. Apprenticeship learning about multiple intentions. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11). 2011. p. 897-904.

[14] Levine, Sergey; Koltun, Vladlen. Continuous inverse optimal control with locally optimal examples. arXiv preprint arXiv:1206.4617, 2012.

[15] H. Kurniawati and V. Yadav, An online POMDP solver for uncertainty planning in dynamic environment in Proc. Int. Symp. Robot. Res., 2013, pp. 611–629.

[16] Zhan, W., Sun, L., Wang, D., Shi, H., Clausse, A., Naumann, M., Tomizuka, M.INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps. arXiv preprint arXiv:1910.03088, 2019.