Human Navigational Intent Inference with Probabilistic and Optimal Approaches

Pedram Agand\textsuperscript{1*}, Mahdi Taherahmadi\textsuperscript{1*}, Angelica Lim\textsuperscript{1} and Mo Chen\textsuperscript{1}

Abstract—Although human navigational intent inference has been studied in the literature, none have adequately considered both the dynamics that describe human motion and internal human parameters that may affect human navigational behaviour. In this paper, we propose a general probabilistic framework to infer the probability distribution over future navigational states of a human. Our framework incorporates an extended Dubins car dynamics to model human movement, which captures differences in human navigational behaviour depending on their position, heading, and movement speed. We assume a noisily rational model of human behaviour that incorporates a) human navigational intent that may change over time, b) how optimal a person's actions are given the navigational intent, and c) how far ahead in time a person considers when choosing navigational actions. These parameters are recursively and continuously updated in a Bayesian fashion. To make the Bayesian update and inference tractable, we exploit properties of the time-to-reach value function from optimal control and the extended Dubins car dynamics to construct a utility function on which the human policy is based, and employ particle representations of probability distributions where necessary. We demonstrate the effectiveness of our method by comparing our results with a recent approach using synthetic data and validate it on real world data.

I. INTRODUCTION

Human navigational intent inference is essential for safe and smooth navigation of autonomous robots in the presence of humans. In many different indoor and outdoor scenarios, having an idea of where people may go in the next few seconds is of great importance in safe decision-making and autonomous navigation [1]. One key task in human-aware planning is to forecast future human states; in particular, due to the variability of human behaviour, inferring a distribution over future human states is important for safe navigation. One way to ensure safe and natural robot behavior is by maintaining a parametrized model of the environment that includes dynamic agents such as humans, and continuously updating these parameters based on observations [2], [3].

Methods for human navigational intent prediction can be grouped into different categories based on the available input modalities, the modeling approach, and the desired form of prediction which depends on the application domain [4]. Many recent works fall into the category of deep learning-based approaches. Some of these approaches use a sequence of scene frames and make predictions based on contextual/dynamical cues or social interactions in the scene [5]. A constant-velocity motion for humans is considered in [6] which forecasts trajectories of occluded pedestrians in crowded scenes. The authors in [7] incorporate environmental dynamics as scene graph representations to predict dynamically-feasible agent trajectories. Other approaches utilize advanced neural network architecture like GAN and LSTM to predict future human goals [8]–[13].

Previous works that used visual inputs tend to perform prediction in the image space; other works have considered predictions in real world coordinates. For example, some studies tried to employ probabilistic models based on observational data from humans. Given a goal and a hand-designed objective function, the authors of [14] proposed a maximum entropy-based probabilistic model for humans and introduces a belief parameter as an indicator of model confidence, which is updated in a Bayesian framework. The idea that a human can be seen as an agent acting based on a utility function was first introduced in [15]. The authors in [16] and [17] focused on Model Predictive Control and Inverse Reinforcement Learning (IRL) to model human behaviour. Also in [18] and [19], the authors investigated goal-directed human motion estimation from an Optimal Control perspective. Two of the first papers that combined optimal control and Bayesian filtering is [14], [20], which inspire our work. We adopt the same inference framework as [14], [20], and propose substantial improvements to the modeling of human dynamics and behaviour.

In this paper, we aim to predict the probability distribution over future human states, and propose an approach that more adequately accounts for the dynamics of human motion. We assume that humans navigate according to a four-dimensional (4D) extended Dubins car as in [21], and take actions in a noisily-rational manner [14], [20], [22]. The navigational model is parametrized by three internal human parameters: policy optimality, farsightedness, and navigational intent. We estimate the probability distribution of these parameters and predict future human states simultaneously.

In contrast to [14], [20], which used a simple 2D model of human dynamics, we account for the effects of human orientation and speed on future human states. However, the more realistic human dynamic model introduces two additional challenges. First, online prediction of 4D future states is intractable using a simple Bayes’ filter as was done in [14], [20]. Second, the human navigational cost function, assumed to be time in [14], [20] and our work, no longer leads to a simple, analytical policy; instead, the human policy now needs to be derived based on a value function that is also tractable to compute in real time.

To address the first challenge stemming from the more complex human dynamic model, we use a particle represen-
tation of the probability distributions over human states and over the human navigational intent parameter, and estimate the distributions using an adapted particle filter. In addition to resolving the computational challenge, the use of particles also enables our approach to estimate arbitrary goal distributions, rather than assuming a set of fixed potential goal positions as was previously done.

To address the second challenge, we first pre-compute a time-to-reach (TTR) value function, assuming the goal is at the origin. Then, leveraging the position and orientation invariance of the extended Dubins car dynamics, we derive in real time the TTR function for goals at arbitrary positions and orientations via simple translation and rotation. Using properties of geometric series, this TTR function can then be converted into an action-value (Q) function, which also allows us to introduce the farsightedness factor, and using which the noisily rational policy can be computed. We validate our approach on a synthetic dataset of human-like trajectories created based on our model assumptions, as well as on real human trajectories from the SFU-Store-Nav [23] dataset.

This paper is organized as follows: we briefly explain preliminaries of our proposed method, including our problem statement in Section III. Then we describe our framework and each aspect of our method in detail in Section IV. After that, we present numerical simulation results in Section VI before concluding in Section VII.

II. BACKGROUND ON TIME-TO-REACH (TTR)

Consider a general dynamical model described by

$$\dot{z} = f(z, u),$$

where \(z, u\) are the state and action, respectively. In the TTR problem [24], one is given system dynamics in the form Eq. (1), and aims to find the minimum time that an agent requires to reach a target set of states \(\mathcal{G}\) from any initial state \(X\) while applying the optimal control policy:

$$\phi(z) := \min_{u(\cdot)} T_z[u(\cdot)],$$

where \(\phi\) is the TTR value function, \(T\) is the time it takes to reach the target, \(z\) is the initial state and \(u(\cdot)\) is the set of allowable control policies. The TTR function \(\phi\) can be obtained by solving the stationary Hamilton-Jacobi (HJ) partial differential equation (PDE) of the form of Eq. (3).

$$\min_u \{-\nabla \phi(z) \cdot f(z, u) - 1\} = 0 \forall z \notin \mathcal{G}$$

$$\phi(z) = 0 \forall z \in \mathcal{G}$$

For a more detailed description of TTR, we encourage the reader to refer to [24].

III. PROBLEM FORMULATION

Let \(z_H = (x, y, \theta, v) \in \mathbb{R}^4\) be the human state, consisting of the position \((x, y)\), heading (direction of travel) \(\theta\), and speed \(v\) in the direction of travel. We assume that \(z_H\) evolves according to the 4D extended Dubins car dynamics:

$$\dot{x} = v \cos \theta$$

$$\dot{y} = v \sin \theta$$

$$\dot{\theta} = \omega$$

$$\dot{v} = a$$

where the human action \(u_H := (\omega, a)\) consists of the angular speed \(\omega\) and the linear acceleration \(a\).

We assume that the human action space is discrete, and that the human adopts a state feedback policy \(u_H = u_H(z_H; \eta)\) parametrized by some hidden parameters \(\eta\). We further assume that the policy is stochastic as explained in Section IV. And finally we presume that the human generally aims to reach his/her goal quickly.

The goal of this paper is to tractably estimate the probability distribution \(P(\eta(t)|z_H)\) over possibly time-varying hidden parameters \(\eta(t)\) and the conditional probability distributions over future human states given the current human state, \(P(z_H(t+h)|z_H(t))\), where \(t\) represents the current time, and \(h\) represents the time horizon of prediction. The model parameters \(\eta\) includes 1) the goal \(g\), or the human navigational intent; 2) the policy optimality, \(\beta\), or how optimal the human policy is given the navigational intent and his/her actions; and 3) the farsightedness factor \(\gamma\), or how far ahead in time a person considers when choosing navigational actions.

IV. NAVIGATIONAL INTENT MODEL

In this section, we propose a probabilistic graphical model for human navigational behaviour, which is summarized in Fig. 1.

A. Human Navigational Behaviour

We use the 4D extended Dubins car model in Eq. (4) to describe human motion. For convenience of taking a Bayesian filtering approach, we discretize time, and derive a discrete time version Eq. (4) using Forward Euler:
\begin{align}
    x^{\tau+1} &= x^{\tau} + h v^{\tau} \cos \theta^{\tau} \\
y^{\tau+1} &= y^{\tau} + h v^{\tau} \sin \theta^{\tau} \\
    \theta^{\tau+1} &= \theta^{\tau} + h \omega^{\tau} \\
v^{\tau+1} &= v^{\tau} + h a^{\tau}
\end{align}

which we will abbreviate as $z^{\tau+1} = f_H(z_H^{\tau}, u_H^{\tau})$.

We denote the current discrete time step as $\tau$, and other time steps as $\tau + k$ where the conversion between discrete time steps and continuous time instants is given by $\tau + k = t + h k$, where $h$ is the length of time discretization. We will use superscripts to denote states $z_H^{\tau}$ or actions $u_H^{\tau}$ at time step $\tau$. We also assume that there is a discrete set of possible human actions.

Following [14], [20], [22], we assume that the human actions $u_H^{\tau} = (\omega^{\tau}, a^{\tau})$ are chosen stochastically in a noisily rational manner, conditioned on the state:

$$
P(u_H^{\tau} | z_H^{\tau}; \beta^{\tau}, g^{\tau}, \gamma^{\tau}) = \frac{e^{\beta Q(z_H^{\tau}, u_H^{\tau}; g^{\tau}, \gamma^{\tau})}}{\sum_a e^{\beta Q(z_H^{\tau}, a^{\tau}; g^{\tau}, \gamma^{\tau})}}
$$

where $\beta$ is the policy optimality, $g$ is the navigational intent, and $\gamma$ is the farsightedness factor. These parameters together form the parameters $\eta$ introduced in Section III. The policy optimality parameter reflects the degree of optimality of human actions given a predefined navigational intent, or goal. As $\beta$ approaches $\infty$, the action $u_H^{\tau}$ approaches optimality, which mathematically means that $u_H^{\tau}$ is chosen greedily in each time step with respect to $Q$. Smaller values for beta correspond to more exploration/a less optimal policy. $\beta \geq 0$ since we assume that the human in general tries to achieve his/her navigational intent quickly. $Q$ is the action value function, which represents the utility of applying action $u_H^{\tau}$ at state $z_H^{\tau}$; we will discuss it more detail in Section IV-B.

At every time step $\tau$, given the distribution of human actions $u_H^{\tau}$ from any state $z_H$, we can compute a distribution of future human states conditioned on the parameters $\eta^{\tau}$ as follows:

$$
P(z_H^{\tau+1} | z_H^{\tau}, \eta^{\tau}) = \sum_{\tilde{u}} P(z_H^{\tau+1} | \tilde{u}, z_H^{\tau}) P(u_H^{\tau} | z_H^{\tau}, \eta^{\tau})
$$

where according to the dynamics in Eq. (4),

$$
P(z_H^{\tau+1} | z_H^{\tau}, u_H^{\tau}) = I(z_H^{\tau+1} = f_H(z_H^{\tau}, u_H^{\tau}))
$$

where $I$ is the indicator function.

**B. Utility Functions: TTR and Q**

Without loss of generality, let us first consider the scenario in which a person aims to reach his/her goal, assumed to be the “pose origin” \{z_H | x = y = \theta = v = 0\}, as quickly as possible. One way to model this is to assume a constant reward function $r(z_H) = -h$ for all $z_H$, and that the human acts optimally to maximize the expected discounted sum of rewards over time. Taking inspiration from the action value function, or $Q$ function from reinforcement learning, we define

$$
Q(z_H, u_H; \eta) = Q(z_H, u_H; g, \gamma) = \mathbb{E}\left( \sum_{\tau=0}^{K} \gamma^{\tau} r(z_H^{\tau}) | u_H \right)
$$

where the expectation is taken over the probability distribution over state trajectory, and $K$ represent the optimal number of time steps to reach the navigational intent $g$, which again for now is the pose origin. Since the state space is 4D and the dynamics (5) are non-trivial, an analytic form of the $Q$ does not exist, and its online computation is also not tractable.

To overcome this challenge, we make the crucial observation that since we have assumed for now that the optimal policy is taken, $Kh$ is exactly the TTR value introduced in Section II. We can pre-compute the TTR value function by solving (2) with the dynamics given in Eq. (4) — that is, by setting $f$ in (2) to that given by Eq. (4) — and setting $G = \{z_H : x = y = \theta = v = 0\}$.

Observing that the TTR is the undiscounted sum of reward introduced in Eq. (9), we can derive the $Q$ function from the TTR function, now denoted $K$.

Since $r(z_H) = -h$ for all $z_H$, we have

$$
Q(z_H, u_H, g, \gamma) = \sum_{\tau=0}^{K} \gamma^{\tau} r(z_H^{\tau}) = -h \sum_{\tau=0}^{K} \gamma^{\tau} = h \left( \frac{\gamma^{K+1} - 1}{\gamma - 1} \right)
$$

**C. Mapping Q Function for Arbitrary Goals**

Since the dynamics in (4) is invariant in translation and rotation, we can transform the $Q$ function corresponding to the navigational intent being at the pose origin to a $Q$ function corresponding to the navigational intent being at any position $(x, y)$ and heading $\theta$. Equivalently, we can also just assume that the navigational intent is always at the origin, define the human pose in the reference frame of the navigational intent. We denote the human pose relative to some navigational intent $g = (x_g, y_g, \theta_g)$ as $z_{H,m}$, given as follows:

$$
z_{H,m} := \begin{pmatrix} x_m \\ y_m \\ \theta_m \end{pmatrix} = \begin{pmatrix} \cos \theta_g & -\sin \theta_g & 0 \\ \sin \theta_g & \cos \theta_g & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x - x_g \\ y - y_g \\ \theta - \theta_g \end{pmatrix}
$$

In practice we allow some small deviation from the origin due to numerical errors in solving the HJ PDE.
Hence, the $Q$ function (corresponding to any $g$) can now be obtained as follows:

$$Q(z_H, u_H; g) = Q(z_{H,m}, u_H, g = (0, 0, 0); \gamma)$$  \hspace{1cm} (12)

Recall that this $Q$ value determines the human control policy in Eq. (6).

V. INFERENCE FRAMEWORK

In this section, we propose a framework to estimate and update the parameters $\eta$. We assume that $\gamma$ and $\beta$ are time-invariant, whereas $g$ may change over time as a human may alter its goal as he/she navigates. Thus, $\eta = (g^\tau, \beta, \text{gamma})$ For implementation convenience, we assume the following factorization of $g$:

$$P(\eta^\tau) = P(g^\tau, \beta, \gamma) = P(g^\tau)P(\beta, \gamma)$$  \hspace{1cm} (13)

A. Estimating $\gamma$ and $\beta$

At every time step $\tau$, we use a new measurement of the human’s action $u_H$ as evidence to update $\gamma$ and $\beta$ using Bayes’ rule:

$$P(\gamma, \beta | z_H^0, u_H) = \frac{P(u_H | z_H^0, \hat{\eta}^{\tau-1})P(\gamma, \beta | \hat{z}_{H}^{\tau-1})}{Z}$$  \hspace{1cm} (14)

where $P(u_H | z_H^0, \hat{\eta}^\tau)$ is obtained from the noisy rational policy in Eq. (6), and $Z = \sum_{\gamma, \beta} P(u_H | z_H^0, \hat{\eta}^\tau, \gamma, \beta)P(\gamma, \beta | \hat{z}_{H}^{\tau-1})$ is a normalizing constant. The notations $\hat{\eta}$ and $\hat{g}$ denote the expected value of the latest estimate of the variables.

B. Estimating $g^\tau$

The measurement update step for $g^\tau$, the human navigational intent, which can be interpreted as the distribution over possible goal states, is similar.

$$P(g^\tau | z_H^0, u_H) = \frac{P(u_H | z_H^0, \hat{\eta}^{\tau-1})P(g^{\tau-1} | z_H^{\tau-1})}{Z}$$  \hspace{1cm} (15)

where $Z$ again normalizes the distribution.

Since $g^\tau$ is a 3D continuous random variable, we use a set of particles $\{g^\tau(i)\}$ to represent its distribution; thus, the above update would be implemented via the standard measurement update step of a particle filter [25].

To model the way $g^\tau$ may change over time, we need an equation for the prediction step of the particle filter that captures the way a person’s navigational goals may change over time. We consider two cases for the evolution of the goal particles $g^\tau(i)$ into two cases, depending on whether a uniform random variable $\zeta(i)$ is above or below the threshold:

$$g^\tau(i) = \begin{cases} g^{\tau-1}(i) + \nu & \zeta(i) < \zeta^{th} \\ G^{(i)} & \zeta(i) \geq \zeta^{th} \end{cases}$$  \hspace{1cm} (16)

where $g^\tau(i)$ is the $i$-th particle at time step $\tau$, $\nu \sim N(0, \sigma)$ and $\sigma$ controls the random walk variance, $G$ is a set of random poses chosen from a fixed set of entropy-maximizing particles.

When $\zeta < \zeta^{th}$, the observed human action is likely given the current model parameter estimates, so $g^\tau(i)$ is adjusted by a small amount to model minor adjustments to a person’s goal as the person navigates in an environment. When $\zeta \geq \zeta^{th}$, the observed human action is unlikely given the current model parameter estimates, so $g^\tau(i)$ gets “reset”, which models the scenario that people may sometimes abruptly change their navigational intent.

C. Human state prediction

Putting everything together, we have

$$P(z^{\tau+1} | z^\tau, u^\tau) = P(u^\tau | z^\tau, \eta^\tau)P(\eta^\tau | z_H^{\tau+1})$$  \hspace{1cm} (17)

The direct computation on a grid of (17) relation is intractable, especially in real time, due to the 4D state space. Hence, we again use a particle representation to represent distributions in the 4D state space. This starts with drawing samples from the action probability distribution in Eq. (6), and then propagating these particles via the system dynamics in Eq. (5).

VI. RESULTS

We now demonstrate our method using synthetic and real-world data. Initially, we assume that $\beta \in \{0.1, 10\}$, $\gamma \in \{0.9, 0.99\}$, and $\beta, \gamma$ have a uniform distribution in all experiments. We consider three different approaches, which we call the particle filter approach (PF, our approach), the brute force approach (BF, based on probability distribution updates from our theory), and the approach in [20] (WA CA). To reduce clutter with our approach, PF, we will often present the average of the surviving particles after resampling, which represents the empirical mean of the distribution that the particles represent. For BF, we compute probability distributions such as $P(z^{\tau+1} | z^\tau)$ for the states $z^{\tau+1}$ that can result from iterating over all control input combinations. Other related distributions are derived in an analogous fashion. Observe that while BF gives more accurate result in theory, predictions for one time step quickly becomes intractable since the number of probability computations grow exponentially with the number of time steps.

A. Synthetic Data

For the numerical simulation involving synthetic data, we assumed the ground truth parameters are $h = 0.2, T_j = 20$, where $h$ is the sampling interval and $T_j$ is the maximum time horizon. We also assumed that the input bounds of human controls to be $|a| \leq 1$, $|\omega| \leq 1.5$. For the simulation, we discretized the control space with a resolution of 0.5, resulting in 35 combinations: $a \in \{-1, -0.5, 0, 0.5, 1\}$ and $\omega \in \{-1.5, -1, -0.5, \ldots, 1, 1.5\}$. These parameters are used to generate synthetic trajectories on which we applied
Fig. 2: One-step ahead (0.2 seconds) prediction results. The particle representation of $P(z^H_\tau)$ is shown as green dots, that of $P(g^\tau)$ as light-blue arrows, that of $P(z^{\tau+1}_H | z^H_\tau)$ as blue dotted arrows, and the ground truth path of human as solid red line.

In the prediction stage, 50 particles have been used to represent $P(g^\tau | z_0: \tau^H)$ and 200 particles for $P(z^{\tau+1}_H | z^H_\tau)$.

Figs. 2 and 3 show how the particles representing $P(g^\tau | z_0: \tau^H)$ and $P(z^{\tau+1}_H | z^H_\tau)$ for a horizon of one and three time steps ahead, respectively. The states change over time as more information is obtained from observing the action taken by the human. As it can be seen, the goal particles eventually converge to their true values. Fig. 4 shows the distribution over predicted future states in each time step as a heat map for 4 different time points. Fig. 5 shows the estimated probability distribution of $\gamma, \beta$ respectively over time. They both start as uniform distributions and eventually converge to the ground truth.

Fig. 6 shows all components of the $g^{\tau,(i)}$ particles in cyan and their average in black. Again, they converge to the ground truth over time. Fig. 7 shows the predicted future of the human states using the PF (ours) and BF methods. Due to the number of probability mass evaluations increasing exponentially, BF is intractable for horizons greater than one. The PF approach, however, is scalable to longer time horizons.

Table I quantitatively compares the final result of this paper with [20]. As it can be seen, using the 4D state of the human increases the source of knowledge to the inference model and enhances the accuracy of the prediction accordingly. This difference become obvious as the horizon increases. In each consecutive step of prediction to reach desired horizon, we have a distribution over internal parameters and set of possible actions. Using PF, our approach, we are able to draw samples according to their probability, instead of just working on point estimation which could result in accumulated errors. By uniformly sampling from the latest available information, we can obtain better prediction results.

In addition, WA CA results do not provide any predictions for $\theta, v$ due to the definition of states. Even thought its result outperforms our PF approach for one-step-ahead prediction, since there is no simplification in probabilistic model, the performance for three-step-ahead is worse compare to PF. This could potentially be because as the horizon increases, WA CA does not adequately account for the system dynamics and internal model parameters, which result in less accurate predictions in $x, y$. 
TABLE I: Comparison of our approach with [20], predicting $k$ steps ahead.

<table>
<thead>
<tr>
<th>Errors</th>
<th>$x$ (m)</th>
<th>$y$ (m)</th>
<th>$\theta$ (rad)</th>
<th>$v$ (m/sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF (ours) ($k = 1$)</td>
<td>0.3796</td>
<td>0.3738</td>
<td>0.7980</td>
<td>0.7193</td>
</tr>
<tr>
<td>BF (ours) ($k = 1$)</td>
<td>0.1462</td>
<td>0.1261</td>
<td>0.7043</td>
<td>0.5903</td>
</tr>
<tr>
<td>WA CA ($k = 1$)</td>
<td>0.2813</td>
<td>0.2408</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PF (ours) ($k = 3$)</td>
<td>0.8491</td>
<td>0.8552</td>
<td>1.9848</td>
<td>1.6163</td>
</tr>
<tr>
<td>BF (ours) ($k = 3$)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>WA CA ($k = 3$)</td>
<td>1.5234</td>
<td>1.4256</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

B. Real-World Data

We evaluated our method on the SFU-Store-Nav [23] dataset. This dataset contains translational data of humans moving in an indoor setting in real world coordinates, recorded by a motion capture system. We used a Kalman filter to reduce the noise derivatives and to fix the inconsistency in time intervals in the data. We then calculated the speed and heading angle by integrating over positional data. Finally, we estimated the controls (acceleration and turn rate) through the same process. Evaluation on real data, confirms our assumptions about human behaviour model based on which our synthetic data was generated and as the quantitative analysis for horizon one is suggesting in Table II, both BF and PF approaches outperform WA CA.

VII. CONCLUSION

We propose a probabilistic model for predicting the probability distribution over human future states. Compared to previous literature, we use a more sophisticated 4D dynamical system to model human movement, and more precisely model hidden navigational parameters, especially the human navigational intent, by taking into account that navigational goals may change over time.

Using a 4D state space for human dynamics introduces computational challenges for deriving the $Q$ function, which determines the likelihood of each human action, in a principled way. To alleviate these computational challenges, we exploit the position and rotation invariance of the system dynamics, and draw a connection between the TTR value function from optimal control with the $Q$ function.

Future work includes a more nuanced way to update goal particles, ideally given the context of the environment to account for the presence of other people and objects, the room layout, and other real-world factors.
REFERENCES


