Probabilistic And Optimal Human Navigational Intent Prediction

by

Mahdi Taherahmadi

B.Sc., AmirKabir University Of Technology, 2018

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
School of Computing Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Summer 2022

Declaration of Committee

Name:	**Mahdi Taherahmadi**
Degree:	**Master of Science**
Thesis title:	**Probabilistic And Optimal Human Navigational Intent Prediction**
Committee:	**Chair:** JC Liu Assistant Professor, Computing Science
	Mo Chen Supervisor Assistant Professor, Computing Science
	Angelica Lim Committee Member Assistant Professor, Computing Science
	Hang Ma Examiner Assistant Professor, Engineering Science

Abstract

The ability to predict human behaviour and choose the right action for a robot is crucial to safe and natural operation in an environment shared with humans. Assuming a probabilistic model for human navigational decision-making, in this thesis, we aim to tackle the problem of Human Navigational Intent Prediction. To this end, we propose a probabilistic framework for fast and accurate estimation of the probability distribution over future human states given the previous state. We introduce three internal parameters to our model: the human goal (g), the optimality of human actions (β), and farsightedness (γ). Our framework maintains and updates a belief over the parameters by observing human actions online. Also, in contrast to existing methods, we consider a 4D state for human dynamics. We overcome the challenges introduced by using a more complex model by precomputation of a Time-To-Reach based value function and exploiting particle filter sampling for human states and the goal distribution. We evaluated our method using synthetically generated and real-world data and have shown that our method outperforms the baseline in longer prediction horizons.

Keywords: Human-Robot Interaction; Human Trajectory Prediction; Optimal Control; Probabilistic Inference

Dedication

This is a dedication to all teachers. To those who plant seeds of hope, nurture minds and nourish souls. To those who inspire.

Acknowledgements

Words cannot express my gratitude towards my family for their unconditional love during all the ups and downs.

Wholeheartedly, I want to thank my supervisors Dr. Mo Chen and Dr. Angelica Lim for their encouragement, guidance, support and patience, without which getting through difficulties in this journey would have not been possible. I would like to especially thank Pedram Agand for his collaboration in this project, dedication and technical contributions to our publication. And I am thankful to all my friends and colleagues in MARS Lab and ROSIE Lab for creating such a wonderful place to learn and grow.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

With the emergence of robots, the safety of humans has become a real concern. In many scenarios in environments where humans and robots are operating, the ability to model human behavior is essential for robots to ensure the same qualities in interaction with humans to an acceptable degree. One of the important aspects of human behaviour modeling that we have put our focus on in this thesis is predicting human trajectory. In this thesis, we are investigating a Probabilistic and Optimal Control approach to address the problem of human navigational intent prediction.

1.1 Probabilistic Inference

In this project, we are estimating a probability distribution of the future human states, in contrast to point estimation of the final outcome. Probabilistic inference is the task of calculating the probability of one or more random variables taking a specific value or set of values. Probabilistic inference uses probabilistic models, in other words, models that employ probability theory and probability distribution principles to describe a statistical problem such as Confidence Intervals and Bayesian Inference [9]. On the other hand, point estimation is about using observational data samples to calculate a single best guess or a single final value for our estimation. The maximum likelihood (MLE) method is one of the examples of finding a point estimate of an unknown parameter [12].

The intuition behind using a probabilistic approach instead of point estimation is because probabilistic reasoning is a way of representing knowledge that allows for tools and measures for uncertainty in the knowledge. In our task, the navigational intent of a person involves uncertainty inherent in the human decision-making system and its mental model. Whereas point estimation we try to find a unique point in the parameter space which can be considered as the true value of the parameter. On the other hand, instead of a unique estimate of the parameter, we are interested in a probability distribution of the parameter in the parameter space. Using a probabilistic approach we are able to describe how reliable an estimate is.

In the task of Human Trajectory prediction, we can never predict one single destination or point of interest for human, since the intent of the human is constantly changing. Best we can do is to estimate the distribution of the data from which the parameters of the human model are sampled. Moreover, although many other works, such as neural network-based approaches, prefer to output a single final estimate, We are more interested in capturing the concept of confusion and the stochastic nature of the human decision-making process. Using a probabilistic approach enables us to reason about a set of possible values for model parameters and the underlying distribution they sampled from in a hypothetical parameter space.

1.2 Thesis Contributions

This work has been done in collaboration with Ph.D. candidate Pedram Agand and published in our paper 'Human Navigational Intent Inference with Probabilistic and Optimal Approaches' [2].

In our publication, while the implementation of the Probabilistic Inference framework, particle filter sampling and the goal particle evolution algorithm was mostly done by my colleague, I was responsible for:

- Literature review and coming up with the baseline work [11] due to its potential to be improved upon
- Development and incorporation of the idea of a Time-To-Reach base value function into the existing framework
- Re-implementation, adoption and visualization of the TTR from [19] into our specific problem
- Implementation of the discounted state-action value function and human policy
- Implementation of the Synthetic data generation based on human policy
- Pre-processing of the real world data from SFU-Store-Nav [32] dataset into dynamical data and control inputs
- Debugging, iterating and improving the implementation of the Probabilistic Inference framework
- Experimenting with real and synthetic data and analysis of the results
- The visualization of the probability density heat maps of the future human states
- The visualization of the goal candidates and human trajectory based on estimated future states against the ground truth data

1.3 Thesis Structure

In this thesis, we first go over the background of the problem of human navigational intent prediction, other similar tasks that have been worked on in the past and the roots of this problem in the literature. Briefly overview some of the related works then we further explain one of the most relevant ones that this works has been built upon, this hopefully helps the reader to understand our method more smoothly. Then in Preliminaries, we go over a summary of some of the topics discussed in this document to make it more understandable to a general reader. Next we introduce our method in detail, go over the dataset we gathered and used and then we discuss our experiments and results and the future works can be done continuing this path.

Chapter 2

Background

2.1 Related Works

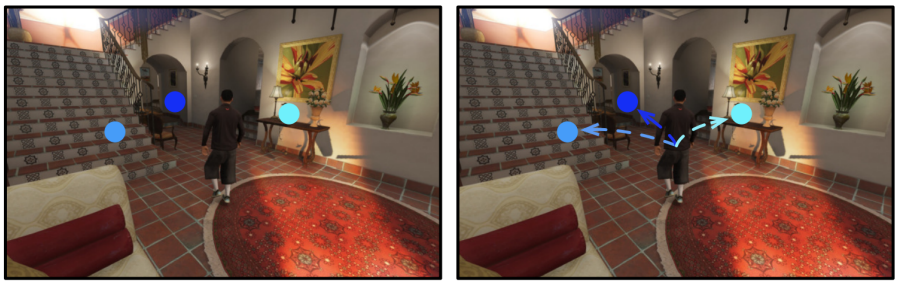
In this chapter, we go over some of the previous works on the Human Trajectory Prediction problem. Depending on the available input data, the modeling perspective, and the output format in each application, recent works in this area can be categorized into different groups. In our study, works that we focus on mostly fit into two main categories of Deep Learning based and Optimal Control based and Probabilistic approaches.

2.1.1 Deep Learning Based Approaches

Figure 2.1: Human trajectory prediction in image space; Social LSTM [3] performs predictions on a sequence of RGB images using Recurrent Neural Networks (RNNs).

Some of the recent work that approached the task of Human Trajectory Prediction in the image space lay in the domain of Computer Vision and Neural Networks. Some of these studies benefited from the sequential nature of available visual datasets [16], [25], these approaches perform predictions on a sequence of RGB images and use information present in the images such as contextual cues, social interactions, or dynamics between humans in the crowd in the scene [3], [7]. in [13] T. Khurana et al. suggested to forecasts the trajectories of occluded pedestrians in crowded scenes assuming a constant velocity for the motion of humans. One of the missing aspects of previous works was not accounting for environmental dynamics, which was later addressed in [26] by incorporating a scene graph representation of environmental dynamics to predict dynamically feasible trajectories of agents. In [8] the

authors implemented a Generative Adversarial Network (GAN) responsible for generating a probability distribution of possible future goal positions of a human and then in a separate stage estimated the set of plausible trajectories to the candidate goals by using a Long Short-Term Memory (LSTM) network and leveraging the information of the past trajectory and visual context of the surroundings. Furthermore, [31] exploited an LSTM based architecture along with Convolutional Neural Networks (CNN) in a hybrid framework consisting of both a Deep Learning module and a Classical Control based module for prediction of future human trajectories. A Variational Auto-Encoder (VAE) based architecture conditioned on end-points in [21] and [20] was used to estimate socially compliant trajectories of humans in the crowd.



(a) predicted goals (b) planned paths

Figure 2.2: In [7] the authors exploited the context of the scene to predict long-term human motions in terms of the goal (a) and trajectory (b) in the image coordinates using a two-branch Conditional Variational Auto-Encoder (CVAE) architecture.

While using visual data for human trajectory prediction has shown various benefits such as providing contextual cues about the human goal and surrounding environment, other approaches predict in real world coordinates and use dynamical data due to being more compact and efficient in many robotics applications.

2.1.2 Probabilistic and Optimal Control Based approaches

Some of the recent work approached this problem using probabilistic models and based on observational data from humans. The authors of [22] considered the human as an agent acting based on a utility function. Similarly [15] and [28] used a Model Predictive Control approach. D Vasquez et al. employed Inverse Reinforcement Learning (IRL) to model human behaviour. Optimal Control methods were used in other work such as [23] and [24] as a goal-directed approach for Human Motion Estimation.

Based on the observation that modeling the human as an approximately rational agent with respect to an objective function learned from its previous behaviour [15, 34], does not adequately capture all human trajectories, due to unpredictability of the human, unknown internal parameters and randomness inherent in human motion, and computational restrictions in the recalculation of the human model in real-time, [10, 11] were among the first

papers that used both optimal control and Bayesian filtering in the estimation of the future state of the human for safe trajectory planning. In [11] Fridovich-Keil et al. propose a method for the robot to continuously estimate its confidence in the human model and its predictions, then adapt its planning according to this confidence. Assuming a set of goals is known, and by using a hand-crafted objective function, the authors estimated the probability distribution over the future states of the human. Their probabilistic model for humans is based on maximum entropy assumptions and incorporates a belief parameter as an indicator of model confidence, which is updated in a Bayesian framework.

We adopt the same inference framework as [10, 11], and propose substantial improvements to the modeling of human dynamics and behaviour.

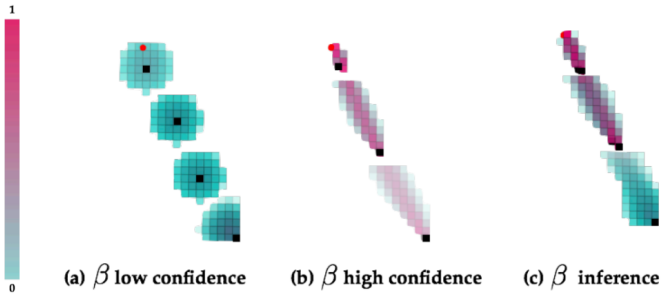


Figure 2.3: Probability distribution of the future states of the agent based on the probabilistic model introduced in [11] bottom right is the initial state and the red point in top left is the goal. (a) and (b) show the predictions of the model based on low and high values of confidence parameter; (c) shows the model predictions where the confidence parameter updated online in a Bayesian framework. Higher values of confidence parameter results in more deterministic predictions due to the nature of humans, sometimes human acts quite non-deterministically.

Confidence-Aware Human Motion Prediction

Since our work is inspired and built on [11], in this section, we specifically explain the basics of their method, as it helps to gain a better understanding of our proposed method and the following chapters of the thesis.

The authors in this work, in order to address the problem of safe trajectory planning for a robot, employ a predictive model of the human to be able to adjust their planning algorithm according to human motion. As the modelling part of their approach they propose a Bayesian framework for reasoning about the uncertainty inherent in the predictions of their model of the human motion.

Problem Formulation and Simplifying Assumptions

In this work, the authors consider a single human moving in a space with a set of known possible goal locations g. The human state is $z_H \in R^2$ and is modeled by the dynamics $\dot{z}_H = f(z_H, u_H)$ in which u_H is the human control action. Let us assume that, human either walks with the constant speed ($v = 1m/s$) or remains in its place ($v = 0$). The robot's goal is to estimate the probability distribution of the future human states in a $2D$ occupancy grid, as shown in Fig. 2.3.

Approximately Rational Human Behavior Model

Based on previous works in cognitive science and econometrics, Utility-Driven Optimization is an efficient way of modeling human behavior [4, 22, 29]. Respectively, the authors in this work assume that the human is optimizing a reward function, $r_H(z_H, u_H; g)$, that depends on the human's state and action, as well as the goal parameter g.

Thus, we can model the human's choice of control action as a probability distribution over actions conditioned on state. Let $U(.)$ be the utility of the human actions calculated based on a given reward function $r_H(.)$, then the probability of human control action is as follow:

$$P(u_H|z_H, g) = \frac{U(u_H|z_H, g)}{\sum_{\bar{u}_H} U(\bar{u}_H|z_H, g)} \tag{2.1}$$

Under maximum entropy assumptions [33] and inspired by noisy-rationality decision-making models [17,18] the human is more likely to choose the action with maximum expected utility. In this case, assuming that we have the state-action value function for the human as $Q(.)$ and that the human's choice of action is associated with the exponential function of this state-action value function $U(u_H|z_H, g) = e^{Q(z_H, u_H; g)}$, we can re-write the formula 2.1 as:

$$P(u_H|z_H; g) = \frac{e^{Q(z_H, u_H; g)}}{\sum_{\tilde{u}_H} e^{Q(z_H, \tilde{u}_H; g)}} \tag{2.2}$$

In this paper, for simplicity, the reward function is assumed to be based on the distance traveled by human:

$$r_H(z_H, u_H; g_H) = -||v_H||_2 = -v_H \Delta t \tag{2.3}$$

According to this reward function, the human is penalized for each step taken to reach the goal g in the time step Δt. Based on this, we can formulate the state-action value function as:

$$Q_H(z_H, u_H; g_H) = -||v_H||_2 - ||z_H + u_H - g||_2 \tag{2.4}$$

Now, by calculation of the Eq. (2.2) given Q in Eq. (2.4) for each goal, we can draw the most probable human action for the future state.

We can conclude that according to this model, the human as a rational agent seeks to reach the goal as soon as possible and maximize its state-value function. Now in order to account for uncertainties in the human model, for example, when the human is not taking action toward any of the predefined goals, we incorporate the model confidence coefficient β into the Eq. (2.2):

$$P(u_H|z_H;\beta,g) = \frac{e^{\beta Q(z_H,u_H;g)}}{\sum_{\tilde{u}_H} e^{\beta Q(z_H,\tilde{u}_H;g)}} \tag{2.5}$$

The role of the β parameter is to determine how much the human actions are aligned with the robots model of the human which directly translates to the state-value function Q. When the values of $\beta \to 0$, the human acts irrational, as if choosing its actions regardless of the state-action function, so the probability distribution of its actions start to shape like a uniform distribution. While $\beta \to \infty$ corresponds to a situation in which human action is completely rational and choosing its actions to optimize its objective according to our model.

Inference framework

The humans actions can be used as evidence about the underlying objective function to infer the parameters of our model. Assuming that the human seeks to maximize their objective in the future, at each time step t, we make a new observation of the human action u_H^t. Using this measurement, we can update a probability distribution over the different values of the parameter, called the belief of the parameter. This can be done for each parameter using a Bayesian update rule. Let us write the update rule for parameter β:

$$P(\beta|z_H^{0:t}) = \frac{P(u_H^t|z_H^t,\beta,g)P(\beta|z_H^{0:t-1})}{\sum_{\tilde{\beta}} P(u_H^t|z_H^t,\tilde{\beta},g)P(\tilde{\beta}|z_H^{0:t-1})} \tag{2.6}$$

Similarly, the belief over parameter g is calculated in the same way. In this paper, a relatively small set of values for both β and g parameters are considered because with higher number of values for each parameter, the computation load increases accordingly. As shown in Fig. 2.3 column (c) shows the flexibility of the model predictions when the confidence parameter is updated in run time compared to (a) and (b) where β has a fixed value.

Now, leveraging the parameter β updated in the run time, we can recursively propagate the human's state over time and calculate a probabilistic prediction of the future states at any number of time steps. To be specific, at every time step in the future, the likelihood of the human taking action u_H from any state x_H by using the formula Eq. (2.5) can be estimated, then in combination with the human dynamics (that we assume to be deterministic here) we

can recursively generate a probability distribution over a 2D occupancy grid of the human in space over time:

$$P(z_H^{t+1}|z_H^t, \beta, g) = \sum_{u_H^t} P(z_H^{t+1}|z_H^t, u_H^t, \beta, g) P(u_H^t|z_H^t; \beta, g) \tag{2.7}$$

in which $P(z_H^{t+1}|z_H^t, u_H^t, \beta, g)$ refers to the deterministic dynamics of the human, meaning that given human action from any state z_H, the next state can be calculated using $\dot{z}_H = f(z_H, u_H)$.

2.2 Challenges And Motivations

Although in the previous work the human model was useful for the task of safe trajectory planning, it was overly simple and missing some nuances of human motion. For example, the previous work only considers x, y position of the human as human state, assumes a constant speed for the human movement and does not consider the changes in the speed, which is very important in inferring the goal of the human. Imagine the scenario in which two people start from the same location but at different speeds, as shown in Fig. 2.4. In the previous model, the probability of the first goal location $g1$ is higher than second goal location $g2$ which is further. In other words, the observation that a human with higher speed is more likely to pass $g1$ does not affect the probability distribution of the goals.

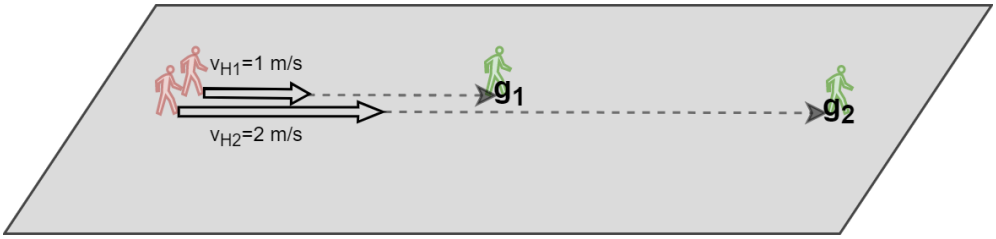


Figure 2.4: In a scenario that two people H_1 and H_2 have the same initial x, y location but have different initial velocity v values, intuitively the person with more pace is more likely to aim for a goal further than a person with a slower initial speed.

Similar settings can be imagined for the difference in the initial direction of travel of two humans in inferring the probability of their destination. Obviously, one could include velocity and direction of travel to the human state, but increasing the dimension of the human and goal states will introduce curse of the dimensionality problem. Using a 4D state space is computationally intractable in real-time since the cost of calculation of the probability distribution of future human states, estimated using Eq. (2.7), will increase exponentially.

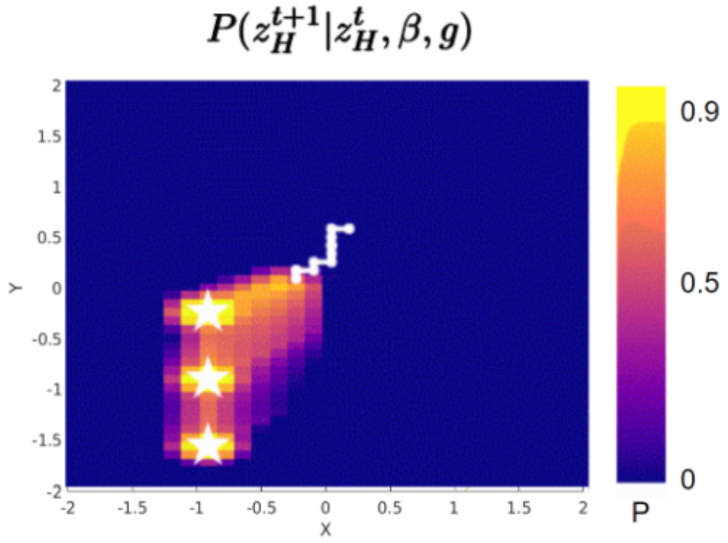


Figure 2.5: A probability distribution heat map representation of the predictions of the [11] in a scenario in which there are three predefined goal locations shown in stars and a sequence of human locations shown in white dots. In this case the goals are aligned in a line in front of the person, given the constant speed assumption about the human, the model makes predictions based on the distance to the goal locations regardless of the speed of the person.

Additionally, using a more complex human state generally makes the calculation of the reward function for human policy more challenging. Furthermore, design of a reward function no longer leads to a closed form and can not be tractably computed in real-time and needs an analytical solution which is tractable in real-time.

The other flaw in the previous approach is the use of a fixed set of known goal locations for the human as a parameter of the model. Having the knowledge of the possible final states of the human is a naive assumption, since the human can change its mind at any moment and choose a new destination. In this case, the previous work tends to give the uniform distribution for all future human states and is no longer able to make useful predictions about the future states.

In summary, three main challenges that were not addressed in the previous work that motivated us to address in this thesis are as follows.

- Overly simplified model of human dynamics and not considering the velocity and heading of the person
- Using a more sophisticated reward function for human that is also tractable for higher dimensions of the human state space
- Using a fixed set of goal locations instead of arbitrary goal states for the human

2.3 Our Approach

This thesis is motivated by the fact that predicting the future state of the human is essential for safe and smooth operation of autonomous robots around humans, and previous work that modeled human motion did not adequately account for the nuances of the human motion which affects its predictability. We take the same framework as [11] and make improvements upon it. We aim to model human motion and propose a real-time and accurate framework to estimate the probability distribution of the future human navigational state that is based on more realistic human motion dynamics and accounts for internal parameters of the human that may affect its navigational behaviour and also the use of arbitrary goal locations.

2.4 Contributions

In summary the contributions of this work are including:

- Proposing a probabilistic framework for online human navigational intent prediction, estimation, and update of the parameter's probability distribution
- Introducing Navigational Intent and Farsightedness parameters into the model
- Incorporating Time-To-Reach (TTR) as a value function in the human policy
- Offline computation of the TTR-based value function and transformation to the state-action value function in the run-time
- Using a Particle Filter representation of the human states and parameters to facilitate the computation
- Introducing arbitrary goal positions with the use of Particle Filter Sampling

Chapter 3

Preliminaries

In this chapter, we briefly go over some of the concepts used in the context of this thesis to facilitate the understanding of the rest of this manuscript.

3.1 Dynamical Models

3.1.1 Dubins Car Dynamical Model

The Dubins car dynamical model which is widely used in robotics and control is as follows:

$$\dot{x} = V \cos\theta \tag{3.1a}$$
$$\dot{y} = V \sin\theta \tag{3.1b}$$
$$\dot{\theta} = \omega \tag{3.1c}$$

where (x, y) is the position of the car, θ is the heading or direction of travel, and we assume that the car in this model is moving at a constant speed V, and the angular speed ω is bounded.

3.1.2 Extended Dubins Car Dynamical Model

An extended version of the Dubins car also considers variable speed, and the relations are as follows:

$$\dot{x} = v \cos\theta \tag{3.2a}$$
$$\dot{y} = v \sin\theta \tag{3.2b}$$
$$\dot{\theta} = \omega \tag{3.2c}$$
$$\dot{v} = a \tag{3.2d}$$

Similarly, (x, y) is the car's position, θ is the heading or direction of the travel, ω is the angular speed, and a is the linear acceleration.

3.2 Boltzmann Noisily-Rational Decision Model

According to Luce's axiom of choice [17, 18] in mathematical psychology about human decision-making, we have a set of options O and want to quantify the likelihood that a human will choose any arbitrary option $o \in O$. Then function $v : O \to R^+$ is the desirability of each option, where more desirable options are resulting in higher v values. According to the Luce's choice axiom, the probability of selecting an option o is given by

$$P(o) = \frac{v(o)}{\sum_{\bar{o} \in O} v(\bar{o})} \tag{3.3}$$

Now if we assume that each option o is associated with some underlying reward $R(o) \in \mathbb{R}$, and then letting the desirability to be an exponential function of this reward, then we have the Luce-Shepard choice rule [27]:

$$P(o) = \frac{e^R(o)}{\sum_{\bar{o} \in O} e^R(\bar{o})} \tag{3.4}$$

When the options $o \in O$ chosen by the human are trajectories or sequences of actions that are potentially continuous, we refer to 3.4 as the Boltzmann model of noisily-rational behavior [6].

3.3 Optimal Control

In Optimal Control theory, we aim to find an optimal control policy for a dynamical system. This is achieved by optimizing predefined cost functions that are designed to serve a specific goal [14]. One way of finding a global optimality is by solving the Hamilton–Jacobi–Bellman (HJB) equation. Dynamic programming leads to necessary and sufficient conditions for optimality expressed in Hamilton-Jacobi-Bellman (HJB) partial differential equation (PDE) for the optimal cost [5].

In classical Optimal Control, methods need a full knowledge over system dynamics and used offline. When there are changes in dynamics or uncertainty is involved, they do not perform well. Let us assume the following dynamical system in $\mathbb{R}^n$:

$$\dot{X} = f(X(t), u(t), d(t)) \tag{3.5}$$

where $X(.)$ denotes the state, $u(.)$ is control and $d(.)$ is disturbance of the system. $u(.)$ and $d(.)$ are two competing inputs and can be considered as controls for player I and Player II. Assuming that f is continuous and that there exists $L > 0$ such that $||f(x_1, a, b) - f(x_2, a, b)|| < L||x_1 - x_2||$ for all $x_1, x_2 \in \mathbb{R}^n$, then the dynamical system in Eq. (3.5) has a unique solution [30].

3.4 Time-To-Reach Function

The goal of the Time-To-Reach (TTR) [19] function is to find the shortest time it takes an agent to reach a target set of states form any initial state X while applying the optimal policy. Assume Eq. (3.5) in which player II wants to maximize the disturbance and player I wants to minimize the control. This problem can be formulated as:

$$\phi(X) := \min_{\theta \in U} \max_{d \in D} T_X[d, \theta[d]] \tag{3.6}$$

where X is the initial state, T is the time and ϕ is the TTR value function. In this case, Player II is trying to maximize the time, and Player I wants to minimize the time by employing the strategy $\theta(d)$ and has knowledge of the current and past decisions of Player II. Now, the TTR function (θ) can be obtained by solving the following stationary Hamilton-Jacobi (HJ) PDE:

$$H(X, \nabla\phi(X) = \min_{d \in D} \max_{u \in U} \{-\nabla\phi(X)^T f(X, u, d) - 1\} \tag{3.7}$$

Where H is the Hamiltonian and f is the system dynamics [19].

3.5 Notations

Throughout this document, we use discretized time and denote the current discrete time step as τ and the other time steps as $\tau + k$, where k is the number of steps. The discrete time steps can be converted to continuous time using the relation $\tau + k = t + hk$, where h is the length of the discretized time step. We will use superscripts to denote time steps τ such as in states z_H^τ or actions u_H^τ.

Chapter 4

Human Navigational Intent prediction

In this chapter we will explain more about our approach for tackling human navigational intent prediction problem. First, we describe our problem setting and the limitations we consider for the scope of this matter. Second, we delve into the framework and explain pieces of the framework one at a time; starting with underlying assumptions necessary for solving the problem, we introduce our model inner parameters and their interplay in our mental model of the human. Next, we explain human policy step by step and at the end, by putting all the pieces together, we explain the inference framework used to estimate the probability distribution of the human navigational states in the future.

4.1 Problem Statement

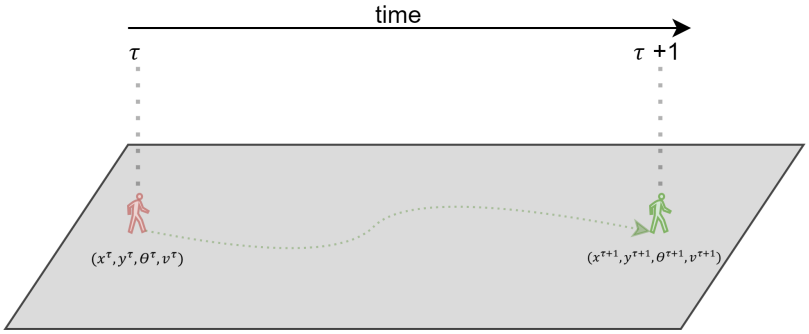


Figure 4.1: This is a general schema of the problem. A human intends to go from an initial state to an unknown final state. The human in the initial state of $(x^{\tau}, y^{\tau}, \theta^{\tau}, v^{\tau})$ at time τ after applying control $u_H^{\tau} := (\omega, a)$ in the given dynamical system and after one step in time will reach the state $(x^{\tau+1}, y^{\tau+1}, \theta^{\tau+1}, v^{\tau+1})$. Our goal is to estimate the probability distribution over the future state of the human, referred to as their navigational intent.

Assuming that we have human states as $z_H = (x, y, \theta, v) \in \mathbb{R}^4$ in which x, y is the position of the human, θ is the heading or direction of travel of the human, and v is the linear speed of travel in the direction of travel. And the human has the set of actions $u_H := (\omega, a)$ in which ω is the angular speed and a is the linear acceleration. We then assume a dynamical model for human movement, using which we are able to evolve human states in time as shown in Fig. 4.1.

At the highest level, the goal of this thesis is to estimate the probability distribution of the future states of the human given the previous states without having the knowledge of his goal locations, $P(z_H(t+h)|z_H(t))$, where t is the current time and h is the time horizon of prediction. To solve this problem, we consider a model for human movement consisting of a set of internal parameters and propose a probabilistic framework using which we are able to estimate the probability distribution of the model parameters as well as the future human navigational states.

4.2 Human Navigational Intent Model

In this section, we will describe our proposed model for human movement. First, we name the simplifying assumptions about human movement, which makes this model useful while being computationally tractable. Then we introduce the internal parameters we considered for our model and their benefits. Next, we explain how we define the value function based on TTR, calculate the value function offline, and transform it to the state-action value function in the run-time, before we form the action policy for the human. Then we explain the Bayesian update rules for the parameters, our inference framework, and how we exploit particle filter sampling to compute the probability distributions of human states and the goal parameter.

4.2.1 Underlying Assumptions

To further simplify the problem at hand, we have made a set of assumptions about the human navigational model.

First our model assumes a Noisily-Rational Model for the human (introduced in 3.2) According to this model, the human is likely to choose the action from a stochastic policy $\pi_H(.)$ which maximizes an underlying reward function but has some inherent randomness involved that causes the human to sometimes act irrational or sub-optimal. This model is useful because in the real world humans are not always acting optimal or "rational" in other sense. This allows us to use a Boltzmann distribution for the human actions conditioned on the previous state as in Eq. (3.4). According to this formulation, the likelihood that the human will choose an action is associated with the exponential function of the state-action value function. This facilitates the use of a coefficient of the state-action function as an

indicator of the optimality of the state value function with respect to the observed actions of the human.

Second, we assume that the human is trying to reach its goal in the shortest amount of time. The value function is designed to have a penalty for each time step.

Third, we assume the 4D Extended Dubins car dynamic model for human (introduced in Section 3.1.2).

Fourth, we assume a discrete set of possible actions for the human.

4.2.2 4D State Space

We consider a 4D extended Dubins car model as in Eq. (3.2) for the human movement. In order to benefit from a Bayesian filtering approach, we discretize time and derive a discrete-time version Eq. (3.2) through Forward Euler discretization:

$$x^{\tau+1} = x^{\tau} + hv^{\tau}\cos\theta^{\tau} \tag{4.1a}$$

$$y^{\tau+1} = y^{\tau} + hv^{\tau}\sin\theta^{\tau} \tag{4.1b}$$

$$\theta^{\tau+1} = \theta^{\tau} + h\omega^{\tau} \tag{4.1c}$$

$$v^{\tau+1} = v^{\tau} + ha^{\tau} \tag{4.1d}$$

We will use $z_H^{\tau+1} = f_H(z_H^{\tau}, u_H^{\tau})$ for simplicity.

4.2.3 Internal Parameters

We introduce three parameters in our model, Policy Optimality β, Navigational Intent g, and Farsightedness γ. Their interpretation, the advantage of each, and the way they affect the human navigational intent prediction is explained in this section.

Policy Optimality β

There are unknown internal parameters that affect decision-making process of the human and because of that, human may seem not to act according to our model of their behaviour. Based on the Noisily-Rational Model assumption about the human model we can say that, in cases in which human is not acting as expected by our model, the model is less optimal. We employ this interpretation for the use of a policy optimality β in our model. This parameter allows us to capture when our model is not optimal regarding the actions observed from human.

Navigational Intent g

Initially we don't make an assumption about the human goal and we don't have a set of predefined goals, so we introduce the goal locations as parameters of the human navigational

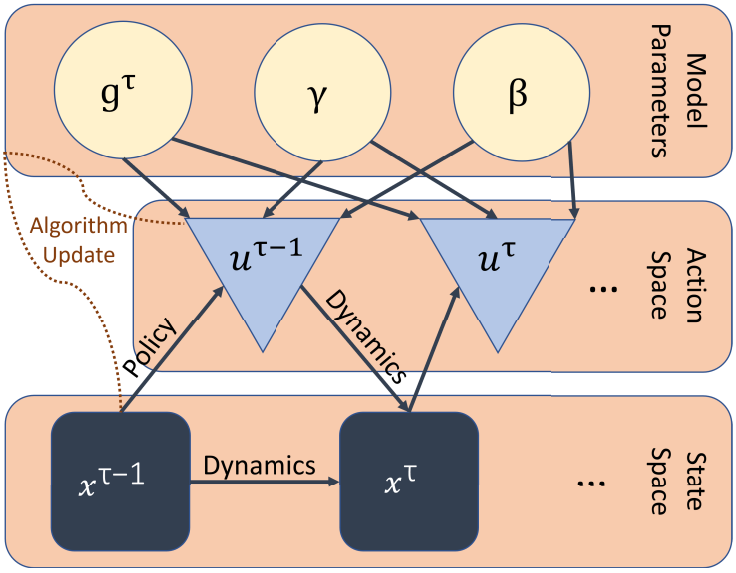


Figure 4.2: A graphical model of the relationship between model parameters, human actions and human states in time. Recursively, as human starts taking action, we compare the observation of their actions with our model expectation and update the belief over model parameters. Then again use the parameters to make prediction for the next time step. Note that β and γ are time-invariant, while g may change over time.

model. Initially we sample a fixed number of random goals distributed evenly from the human state space, and then as we make observations about human action we update the probability of each goal parameter and re-sample our goals in a way that finally the samples closer to the actual goal of the human have higher probability and the goal particles converge to the actual goal location of the human.

Farsightedness γ

The farsightedness parameter is a measure to emphasize the preference of the human about closer goals compared to the further goal locations. This parameters is inspired by the discount factor in the Reinforcement Learning (RL) literature. Since when computing the discounted state-action value function we get to incorporate a discount factor to differ the weight between immediate rewards versus far in the future rewards, or in our case reaching the goals in the closer proximity versus goals that are far and take more time to reach.

We consider a fixed set for β and γ so these two parameters do not change over time, while the size of the g is fixed, its elements are changing over time. In our model, we let the goal parameter to evolve over time, because human may change its navigational intent over time as they navigate through the environment.

4.2.4 State-Action Value Function

Let us first assume that the human wants to reach a goal that is defined at the pose origin as: $\{z_H \mid x = y = \theta = v = 0\}$, from any initial state z_H in the shortest amount of time. As mentioned in Section 4.2.1 according to the second assumption about the human model, we assume that the human acts optimally to maximize the expected discounted sum of rewards over time. For all z_H, considering a constant reward function:

$$r(z_H) = -h \tag{4.2}$$

Inspired by Reinforcement Learning, an state-action value function in Eq. (4.3) indicates how good an action is for the human in a state z_H with a policy π. Let us say that the Q function is denoted by Q(s). It specifies the value of taking the action u_H in the state $z - H$ following a policy π by calculation of the expected sum of rewards $r(z_H)$ at each time step τ assuming that it takes K steps to terminate at the goal.

$$Q(z_H, u_H; g) = \mathbb{E}_\pi \left(\sum_{\tau=0}^{K} r(z_H^\tau) \mid u_H \right) \tag{4.3}$$

Note that the human choice of action depends on the g parameter.

Now we can define the state-action value function Q as a discounted sum of rewards by injecting the coefficient γ:

$$Q(z_H, u_H; g, \gamma) = \mathbb{E} \left(\sum_{\tau=0}^{K} \gamma^\tau r(z_H^\tau) \mid u_H \right) \tag{4.4}$$

4.2.5 Time-To-Reach Based Action-State Value Function

Since we know that it takes K time steps for the human to reach the navigational intent g, the k is the optimal number of steps from initial state z_h to the goal g at the pose origin. Also, by definition, we know that the TTR function (introduced in Section 3.4), represents the time it takes for the human to reach the target set in the origin from any state z_H, and is obtained by solving (3.7). We can make the important observation that the TTR value important is exactly the same value as K.

Now we can generalize the Eq. (4.4) using TTR in place of K.

$$Q(z_H, u_H; g, \gamma) = \mathbb{E} \left(\sum_{\tau=0}^{TTR} \gamma^\tau r(z_H^\tau) \mid u_H \right) \tag{4.5}$$

From now on, K and TTR can be used interchangeably. Since $r(z_H) = -h$ for all z_H, where h is the time step length, we have the following:

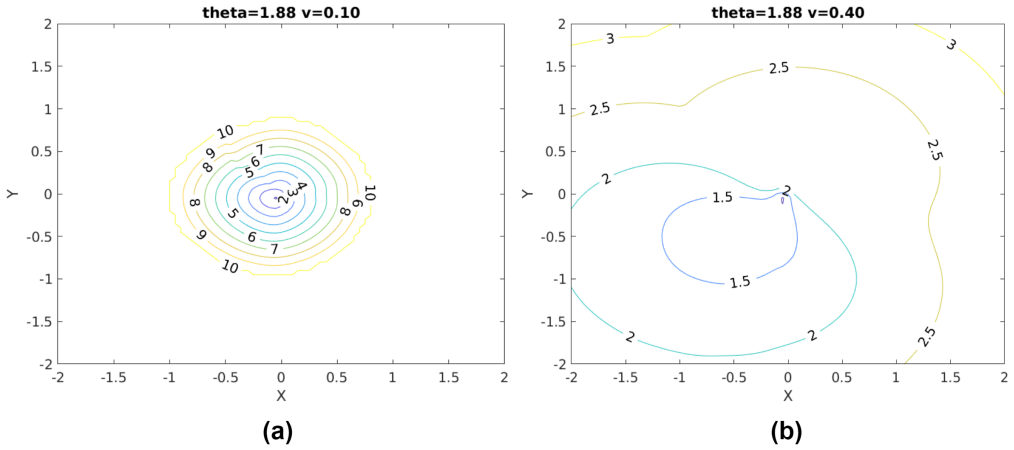


Figure 4.3: Illustration of two example level sets of TTR function values for initial states (x, y, θ, v) and goal at pose origin $\mathcal{G} = \{z_H : x = y = \theta = v = 0\}$. In both figures, θ is considered to have fixed value of $\theta = 1.88$. (a) shows the TTR values for the $v = 0.10m/s$ and (b) shows the TTR values for the $v = 0.40m/s$. As can be seen here, TTR values for different velocities are different which means the human with the grater initial velocity needs fewer steps in time to reach their goal or in other words tend to choose more distance goals.

$$Q(z_H, u_H, g, \gamma) = \sum_{\tau=0}^{K} \gamma^\tau r(z_H^\tau) \tag{4.6a}$$

$$= -h \sum_{\tau=0}^{K} \gamma^\tau \tag{4.6b}$$

$$= h \left(\frac{\gamma^{K+1} - 1}{\gamma - 1} \right) \tag{4.6c}$$

4.2.6 Offline Computation Of TTR

Since the state space is 4D and human dynamics is assumed to be (4.1), a closed form of Q does not exist and its online computation is intractable. A useful trick here is to pre-compute the TTR value function by solving (3.7) given the dynamics in Eq. (3.2). This can be done by replacing system dynamics f in (3.7) with the dynamics of the human in our problem given in Eq. (3.2). During computation of the TTR we need to assume that the goal is at the origin and is stationary $\mathcal{G} = \{z_H : x = y = \theta = v = 0\}$, however, in the implementation we consider some deviation from the exact values of origin due to the numerical errors in solving the HJ PDE.

4.2.7 State-Action Value Function Transformation

Up to now, we have pre-computed the state-action value function Q from any initial state z_H to a goal g at the pose origin, if we intend to use this state-action value function during inference, we need to use this Q function for any arbitrary goal in the (x, y, θ) space. Note that we defined the goal state to be stationary, meaning $v = 0$ and the human always approaches a goal while facing it.

Given that the dynamics in (3.2) are invariant in translation and rotation, we can simply transform the Q function corresponding to the goal being at the pose origin, to a Q function corresponding to the new arbitrary goal locations (x, y) and heading θ by translation and rotation.

In the same way, we can just define the human pose in reference to its goal so that its goal is always at the origin using the transformations $z_{H,m}$ of the human pose relative to the navigational intent $g = (x_g, y_g, \theta_g)$:

$$z_{H,m} := \begin{pmatrix} x_m \\ y_m \\ \theta_m \end{pmatrix} = \begin{pmatrix} \cos\theta_g & -\sin\theta_g & 0 \\ \sin\theta_g & \cos\theta_g & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x - x_g \\ y - y_g \\ \theta - \theta_g \end{pmatrix} \tag{4.7}$$

Now, by using the transformations $z_{H,m}$ in place of z_H, we can drive the Q function stated in Eq. (4.5) for any arbitrary navigational intent g as follows:

$$Q(z_H, u_H; \gamma, g) = Q(z_{H,m}, u_H, g = (0, 0, 0); \gamma) \tag{4.8}$$

4.2.8 Human Policy

Inspired by [4, 22, 29] we consider that the human choice of control action comes from a probability distribution conditioned on state. Now that we have the state-action value function Q for the human, we can model the probability distribution of human action conditioned on state and set of parameters at time τ as a Boltzmann distribution as follows:

$$P(u_H^\tau | z_H^\tau; \beta, \gamma, g^\tau) = \frac{e^{\beta Q(z_H^\tau, u_H^\tau; g^\tau, \gamma)}}{\sum_{\tilde{u}} e^{\beta Q(z_H^\tau, \tilde{u}; g^\tau, \gamma)}} \tag{4.9}$$

where β is the *policy optimality*, g is the *navigational intent*, and γ is the *farsightedness factor*.

Q is the transformed state-action value function for goals at arbitrary locations, as in Eq. (4.8), discussed in Section 4.2.7. According to Q, optimally human is likely to choose the action u_H^τ with maximum utility form state z_H^τ.

The policy optimality parameter used as the coefficient of Q to control how optimal actions of the human are compared to our models expectation of humans action, given a

predefined navigational intent, or goal. As discussed in Section 4.2.3, the greater values of β tend to allow the model to choose the more optimal action u_H^τ, which mathematically translates to the greedy choice of u_H^τ at each time step with respect to Q. Smaller values of β can account for more exploration or a less optimal policy. Since we assume that the human intends to reach their goal as quickly as possible, only positive values for β are accepted.

4.3 Online Inference Framework

In this section, we explain the framework using which we can estimate the probability distribution of the human navigational states at any number of time steps in the future. Also in this framework we are able to simultaneously maintain a Bayesian belief over model parameters in real-time.

4.3.1 Future State Prediction

For now, assume that we have a probability distribution over our model parameters β, γ, and g. The probability of human control actions given in Eq. (4.9) enables us to calculate the following probability:

$$P(z_H^{\tau+1}|z_H^\tau, g^\tau, \beta, \gamma) = \sum_{\tilde{u}} P(z_H^{\tau+1}|z_H^\tau, \tilde{u}) P(u_H^\tau|z_H^\tau, g^\tau, \beta, \gamma) \tag{4.10}$$

Where $P(z_H^{\tau+1}|z_H^\tau, u_H)$ is deterministic and given by the human dynamics in Eq. (4.1):

$$P(z_H^{\tau+1}|z_H^\tau, u_H) = I(z_H^{\tau+1} = f_H(z_H^\tau, u_H^\tau)) \tag{4.11}$$

I is the indicator function here.

4.3.2 Parameter Update

We observed that the degree of optimality of our model with respect to human actions varies over time, for example, due to a sudden change in the behaviour of the human. Consequently, the accuracy of our prediction varies over time. That is the root cause of the need for an online estimation of the probability distribution of this parameter. Based on this important observation, the same principle applies to all three parameters of our model. We need to update the probability distribution of these parameters on the go. To do this, we recursively update the probability distribution of the set of values of the parameters. We introduced three parameters in our model and we have a set of values for each parameter which initially have a uniform distribution, and over time these distributions evolve by

making observations of the human's state and actions. We do this using a Bayesian update rule and call it the belief of the parameter.

For simplicity, let us say we have an arbitrary parameter η in our model. At every time step τ, we make a new measurement of the human's action u_H^τ, this measurement is used as evidence to update the belief $b^\tau(.)$ of the parameter η over time via a Bayesian update:

$$b^\tau(\eta) = P(\eta | z_H^{0:\tau}, u_H^\tau) = \frac{P(u_H^\tau | z_H^\tau, \hat{\eta}^{\tau-1}) b^{\tau-1}(\eta)}{\sum_{\tilde{\eta}} P(u_H^\tau | z_H^\tau, \tilde{\eta}) b^{\tau-1}(\tilde{\eta})} \tag{4.12}$$

where $P(u_H^\tau | z_H^\tau, \hat{\eta}^\tau)$ is obtained from the human policy in Eq. (4.9). The notation $\hat{\eta}$ denotes the expected value of the latest estimate of the variables.

For the Bayesian update for each of the model parameters β, γ, and g we just have to replace η with the parameter in Eq. (4.12). Also, we can estimate the joint probability distribution of the parameters $P(g^\tau, \beta, \gamma) = P(g^\tau | \beta, \gamma) P(\beta | \gamma) P(\gamma)$ as follows.

$$b^\tau(g, \beta, \gamma) = \frac{P(u_H^\tau | z_H^\tau, \hat{g}^{\tau-1}, \hat{\beta}, \hat{\gamma}) b^{\tau-1}(g, \beta, \gamma)}{\sum_{\tilde{g}, \tilde{\beta}, \tilde{\gamma}} P(u_H^\tau | z_H^\tau, \tilde{g}, \tilde{\beta}, \tilde{\gamma}) b^{\tau-1}(\tilde{g}, \tilde{\beta}, \tilde{\gamma})} \tag{4.13}$$

4.3.3 Overall Occupancy Probability Distribution

Now that we have a belief over model parameters at time step τ, we are able to recursively propagate the human state distribution forward to any future time. For example, assume that we know the probability of the human in state z_H^τ at the time step τ. Also, the probability of the human choosing the control action u_H^τ is given by Eq. (4.9) and having Eq. (4.11) the deterministic dynamics model, the human state distribution at the time step $\tau + 1$ can be obtained by marginalizing over model parameters according to our belief at the current step time τ:

$$P(z_H^{\tau+1} | z_H^\tau) = \sum_{\tilde{u}, \tilde{g}, \tilde{\beta}, \tilde{\gamma}} P(z_H^{\tau+1} | z_H^\tau, \tilde{u}, \beta, \gamma) P(u_H^\tau | z_H^\tau, g^\tau, \beta, \gamma) \tag{4.14}$$

The computation on a grid of the Eq. (4.10) for 4D state space is intractable in real-time. Therefore, we use particle filter sampling to represent distributions in the 4D state space. To do this, we start by drawing samples from the set of actions u_H in each state z_H to estimate the action probability distribution according to Eq. (4.9), and then calculate (4.10) for a subset of actions and propagating these particles via the system dynamics in Eq. (4.1).

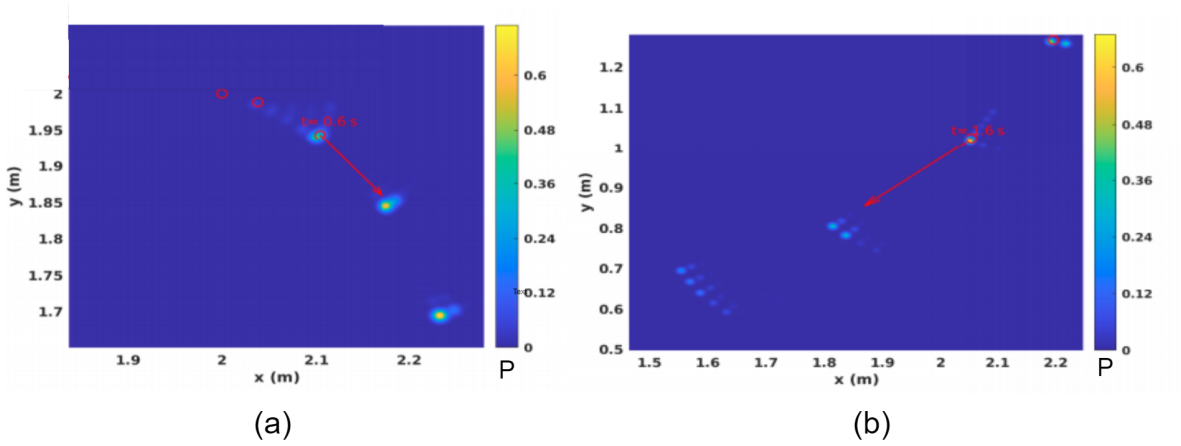


Figure 4.4: Heat map representation of the probability distribution of occupancy in future states (4.14) at time τ. The red circle and the red arrow show the human location x, y on a 2D grid and the human heading θ, respectively. The velocity v is is associated with the length of the arrow. (a) is at time $\tau = 0.6s$ and (b) is at $\tau = 1.6s$.

4.3.4 Goal Particles Evolution

As discussed in Section 4.2.6, g^τ is in 3D space (x, y, θ), for ease of computation, we use a set of particles $\{g^{\tau,(i)}\}$ to represent its distribution, then the update formula for parameters (4.12) is employed using the standard measurement update step of a particle filter. These particles evolve over time to account for changes in the navigational intent of the person. We do this using a model mismatch index $\zeta^{th} \in [0, 1]$ [1]. ζ^{th} is a threshold based on which we decide whether to evolve each target particle via a random walk or choose from a fixed set of entropy-maximizing particles.

$$g^{\tau,(i)} = \begin{cases} g^{\tau-1,(i)} + \nu & \zeta < \zeta^{th} \\ G^{(i)} & \zeta \geq \zeta^{th} \end{cases} \tag{4.15}$$

where $g^{\tau,(i)}$ is the i-th particle at time step τ, $\nu \sim N(0, \sigma)$ is the random walk variable. σ controls the random walk variance.

In the first case, when $\zeta < \zeta^{th}$, the goal particle $g^{\tau,(i)}$ is only minimally adjusted since the human is likely to choose from the current set of goals and in the second case when $\zeta \geq \zeta^{th}$, $g^{\tau,(i)}$ will be reset to new random location of the goal, which translates to when the human makes a sudden change in their movement.

We assume that the particle weights are modified in accordance with Eq. (4.12) for the g^τ parameter.

Chapter 5

Datasets And Experiments

We have experimented our method using real-world and synthetically generated data. We have implemented three different approaches for comparison; the particle filter approach (PF), the brute force approach (BF), which is based on probability distribution updates from our method but by trying all possible cases in the state space grid without using particle filter, and the approach in the base line work [11] (WA CA). For the PF approach, we present the average of the particles after each update step, which is equal to the empirical mean of the distribution that the particles represent. For BF, we compute $P(z^{\tau+1}|z^{\tau})$ for the states $z^{\tau+1}$ using all available control input combinations.

5.1 Experiment Settings

In our experiments, we assumed that $\beta \in \{0.1, 10\}$, $\gamma \in \{0.9, 0.99\}$ and they initially have a uniform distribution. The grid boundaries in the TTR computation were:

$$|x| < 8, |y| < 8, |\theta| < \pi, |v| < 2.2 \tag{5.1}$$

In order to take gradients of TTR, the grid has to include negative values of v. Since the negative value is not defined for v, we defined the states with $v < 0$ as obstacles. Control constraints are:

$$a \in [-1, 1], \omega \in [-1.5, 1.5] \tag{5.2}$$

The target set is defined as $\mathcal{G} = \{z_H : x = y = \theta = v = 0\}$ with radius 0.2 and the obstacle radius is 0.15.

In the particle filter representation, 50 particles have been used to represent $P(g^{\tau}|z_H^{0:\tau})$ and 200 particles for $P(z^{\tau+1}|z^{\tau})$ for prediction.

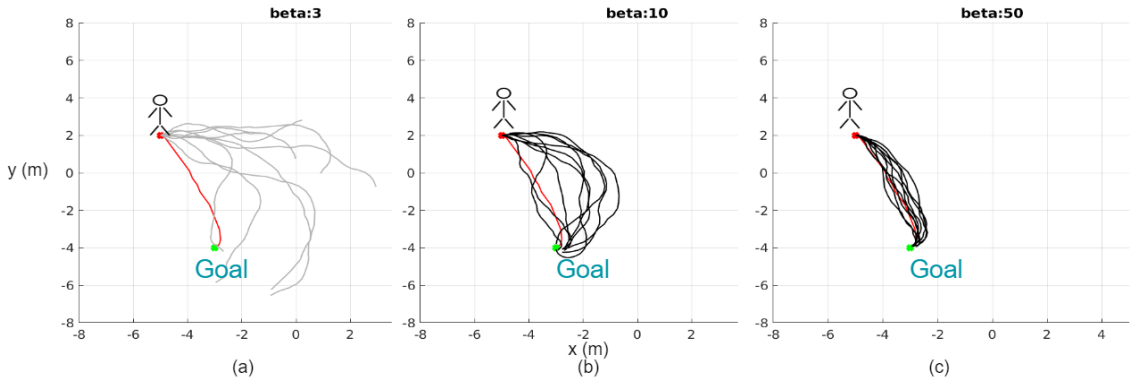


Figure 5.1: The effect of optimality factor β in the synthetically generated trajectories based on the introduced human policy can be seen here, Red is the optimal trajectory given initial and final states. You can see that as the β values go toward infinity, the human tends to choose more optimal actions.

5.2 Synthetic Data

We generated trajectories from different initial state and goal locations in a 2D grid of 8×8, the same as the boundaries of the TTR function. We generated sequences of states by choosing actions according to human policy as stated in Eq. 4.9 in order to reach the goal. During the generation of synthetic data, for the numerical simulation, we assumed the maximum time horizon $T_f = 20$ and the sampling interval $h = 0.2$.

We used human controls within the limits of $|a| \leq 1, |\omega| \leq 1.5$. For the simulation, we used a resolution of 0.5 to discretize the control space, resulting in 35 combinations: $a \in \{-1, -0.5, 0, 0.5, 1\}$ and $\omega \in \{-1.5, -1, -0.5, \ldots, 1, 1.5\}$. After generating synthetic trajectories using these parameters, we applied our prediction algorithm. Some of the examples of generated trajectories and the result of trying different values of β in human policy are demonstrated in Fig. 5.1. It can be observed that as we increase the β values, the human trajectories start to become more optimal. This also confirmed our assumptions about human policy that human actions come from a noisy-rationality decision-making model.

Next, we compare our method with the baseline [11] using the synthetically generated dataset. The quantitative results of these experiments are in table 5.2. It is observed that, using the 4D state, we managed to better model the human and increase the information of the model, and as a result, our inference model achieved a lower displacement error on the predictions. The qualitative results of predictions for one step ahead $k = 1$ and three step ahead $k = 3$ are demonstrated in Fig. 5.2 and Fig. 5.3 respectively.

While in the shorter horizons the difference in performance is negligible, especially for BF since it is considering all possible outcomes, as the horizon increases, this difference

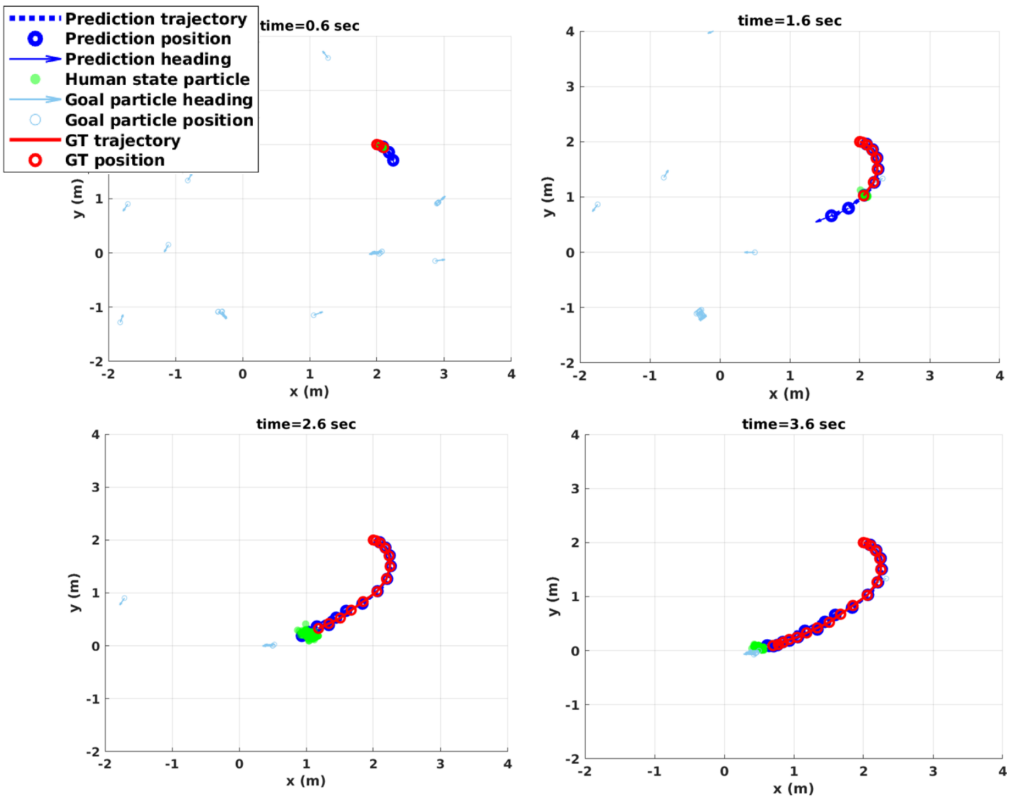


Figure 5.2: Prediction results for one step ahead (0.2 seconds). The particle representation of $P(z_H^\tau)$ is shown as green dots, that of $P(g^\tau)$ as light-blue arrows, that of $P(z_H^{\tau+1}|z_H^\tau)$ as blue dotted arrows, and the ground truth path of the human as solid red line.

becomes more obvious. It should be mentioned that making predictions for longer horizons for the BF approach is intractable in real-time.

Note that the baseline (WA CA) results for θ, v are not available in result tables due to the difference in the definition of state space. Other observation in Tables 5.1 and 5.2 is that, while the baseline in one-step ahead performs better than our PF method, in three-step ahead the baseline due to a simpler model for human, results in higher displacement error.

5.3 Real-World Data

We also evaluated our method on the SFU-Store-Nav dataset [32]. This dataset contains dynamical data of humans moving in an indoor setting in real word coordinates, captured by a motion capture system. In order to process this data, we used a Kalman filter to reduce the noise of the derivatives and to fix the inconsistency in the time intervals in the data. We calculated the speed and heading angle by integrating the positional data and also estimated the human controls $u_H = (\omega, a)$ by integration. Evaluation on real data confirms

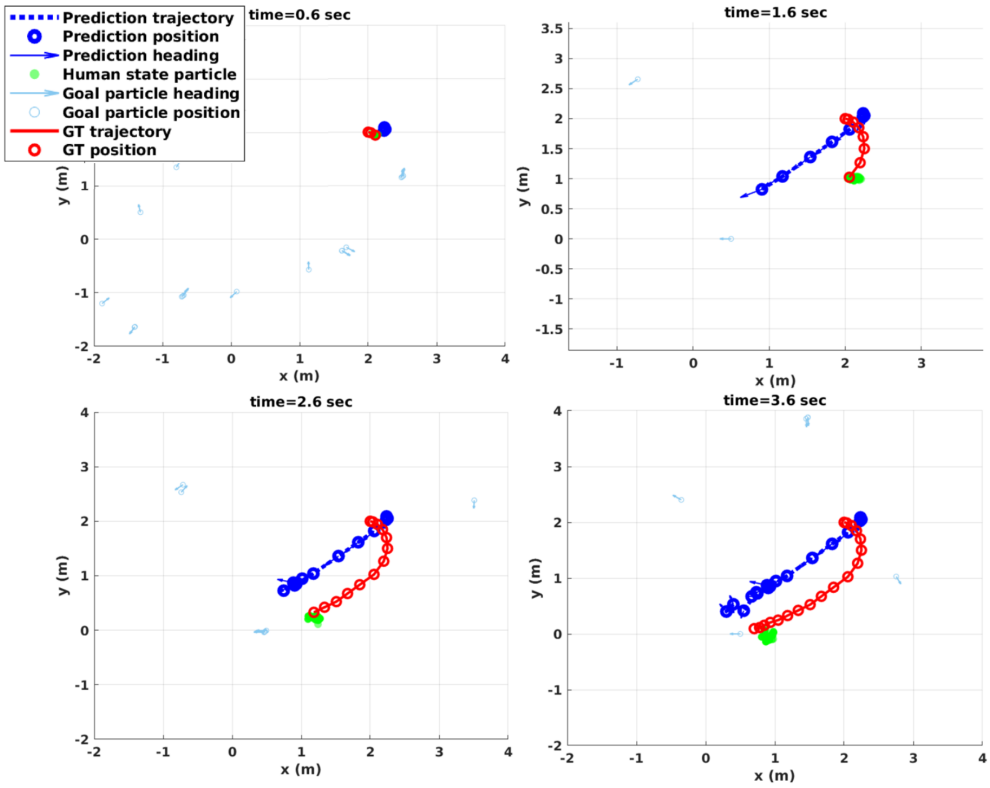


Figure 5.3: Prediction results for three steps ahead (0.6 seconds). The particle representation of $P(z_H^\tau)$ is shown as green dots, that of $P(g^\tau)$ as light-blue arrows, that of $P(z_H^{\tau+1}|z_H^\tau)$ as blue dotted arrows, and the ground truth path of the human as solid red line.

our assumptions about human behaviour model based on which our synthetic data were generated and, as the quantitative analysis for longer horizons suggests in tables 5.2 and 5.1, our approach outperforms the baseline.

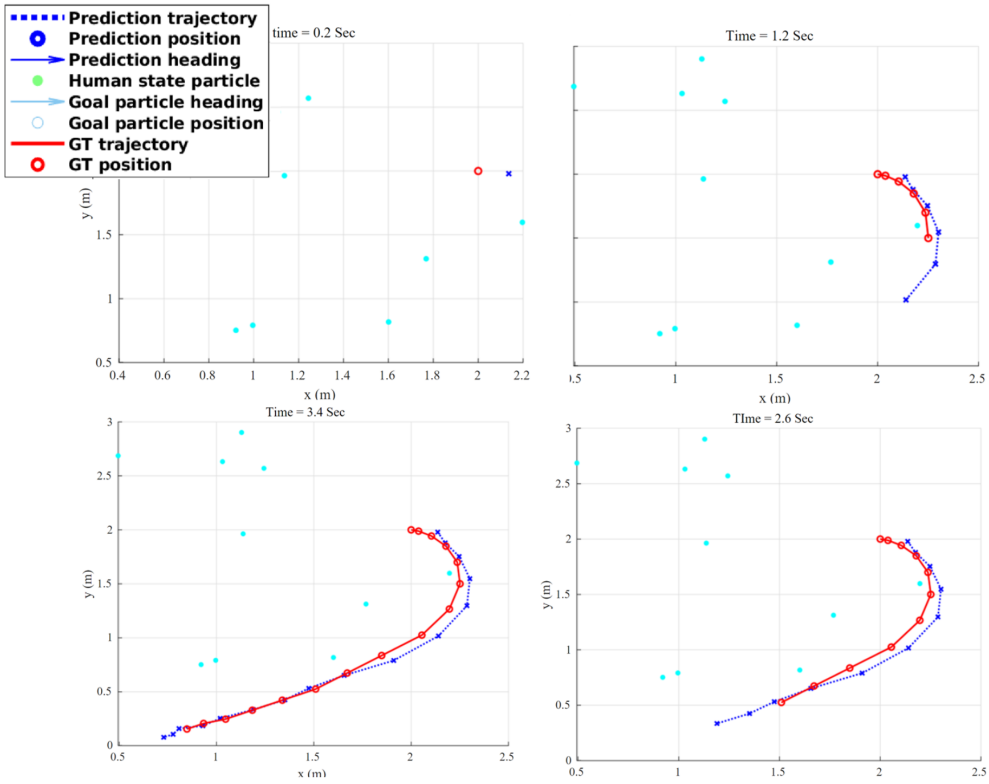


Figure 5.4: Prediction results for one step ahead (0.2 seconds) on real-world data. The particle representation of $P(g^\tau)$ is shown in light blue dots and that of $P(z_H^{\tau+1}|z_H^\tau)$ as blue dotted arrows, and the ground truth path of the human is shown as a solid red line.

Table 5.1: Comparing the displacement error of our approaches (BF) and (PF) on real data with the baseline (WA CA) [11], $k = 1$ step ahead

Errors	x (m)	y (m)	θ (rad)	v (m/s)
PF (ours)	0.0231	0.0229	0.0384	0.0312
BF (ours)	0.0083	0.0079	0.0104	0.0087
WA CA	0.0323	0.0367	NA	NA

Table 5.2: Comparing the displacement error of our approaches (BF) and (PF) on synthetic data with the baseline (WA CA) [11], k is the number of steps ahead

Errors	x (m)	y (m)	θ (rad)	v (m/sec)
PF (ours) ($k = 1$)	0.3796	0.3738	0.7980	0.7193
BF (ours) ($k = 1$)	0.1462	0.1261	0.7043	0.5903
WA CA ($k = 1$)	0.2813	0.2408	NA	NA
PF (ours) ($k = 3$)	0.8491	0.8552	1.9848	1.6163
BF (ours) ($k = 3$)	NA	NA	NA	NA
WA CA ($k = 3$)	1.5234	1.4256	NA	NA

Chapter 6

Discussion

In this thesis we tackled the problem of human navigational intent prediction from probabilistic and optimal approaches. Building upon recent work in this topic, we proposed a probabilistic framework that covers some of the missing aspects of the previous work, using a more complex and complete model of human decision-making and makes real-time predictions, and also outperforms the baseline. To this end, assuming a noisily-rational model for the human, we consider a Time-To-Reach value function and develop the state-action value function or Q function of the human based on that. Furthermore, our model introduced three internal parameters for the human model, the optimality factor β that controls the degree of optimality of human actions with respect to the predictions of our model, the γ or farsightedness parameter to account for immediate or long-term rewards in terms of closer or further goals. And navigational intent parameter g or goal. Our framework maintains a Bayesian belief over model parameters at the same time as inference. We used particle filter representation to overcome the computational limitations imposed by the usage of 4D state space and also for introducing arbitrary goal locations.

To account for the sudden changes in human navigational intent we evolved the goal particles using random walk, a more sophisticated method can be used to better propagate the human goals. Other future directions are to use more information from environment to extract context and find a sense of attractions for human to evolve the hypothetical goal locations around those regions of interests. For simplicity, we did not include obstacles in our model of the environment. This is one of the shortcomings of our work that could be addressed in the future. These obstacles could be moving objects or humans, which requires modeling of multiple dynamic agents at once and brings additional challenges.

Bibliography

[1] Pedram Agand, Hamid D Taghirad, and Ali Khaki-Sedigh. Particle filters for non-gaussian hunt-crossley model of environment in bilateral teleoperation. *ICROM*, 2016.

[2] Pedram Agand, Mahdi Taherahmadi, Angelica Lim, and Mo Chen. Human navigational intent inference with probabilistic and optimal approaches. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 8562–8568. IEEE, 2022.

[3] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. *CVPR*, 2016.

[4] Chris L Baker, Joshua B Tenenbaum, and Rebecca R Saxe. Goal inference as inverse planning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 29, 2007.

[5] RICHARD Bellman. Dynamic programming, princeton univ. *Press Princeton, New Jersey*, 1957.

[6] Andreea Bobu, Dexter RR Scobee, Jaime F Fisac, S Shankar Sastry, and Anca D Dragan. Less is more: Rethinking probabilistic models of human behavior. In *Proceedings of the 2020 acm/ieee international conference on human-robot interaction*, pages 429–437, 2020.

[7] Zhe Cao, Hang Gao, Karttikeya Mangalam, Qi-Zhi Cai, Minh Vo, and Jitendra Malik. Long-term human motion prediction with scene context. In *ECCV*, pages 387–404. Springer, 2020.

[8] Patrick Dendorfer, Aljosa Osep, and Laura Leal-Taixé. Goal-gan: Multimodal trajectory prediction based on goal position estimation. *ACCV*, 2020.

[9] Jianguo Ding and A Rebai. Probabilistic inferences in bayesian networks. *Bayesian Network*, pages 39–53, 2010.

[10] Jaime F Fisac, Andrea Bajcsy, Sylvia L Herbert, David Fridovich-Keil, Steven Wang, Claire J Tomlin, and Anca D Dragan. Probabilistically safe robot planning with confidence-based human predictions. In *Robotics: Science and Systems XIV*, 2018.

[11] David Fridovich-Keil, Andrea Bajcsy, Jaime F Fisac, Sylvia L Herbert, Steven Wang, Anca D Dragan, and Claire J Tomlin. Confidence-aware motion prediction for real-time collision avoidance1. *The International Journal of Robotics Research*, 39(2-3):250–265, 2020.

[12] Charles J Geyer and Elizabeth A Thompson. Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):657–683, 1992.

[13] Tarasha Khurana, Achal Dave, and Deva Ramanan. Detecting invisible people. *ICCV*, 2021.

[14] Bahare Kiumarsi, Kyriakos G Vamvoudakis, Hamidreza Modares, and Frank L Lewis. Optimal and autonomous control using reinforcement learning: A survey. *IEEE transactions on neural networks and learning systems*, 29(6):2042–2062, 2017.

[15] Henrik Kretzschmar, Markus Spies, Christoph Sprunk, and Wolfram Burgard. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research*, 35(11):1289–1307, 2016.

[16] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. In *Computer graphics forum*. Wiley Online Library, 2007.

[17] R Duncan Luce. The choice axiom after twenty years. *Journal of mathematical psychology*, 15(3):215–233, 1977.

[18] R Duncan Luce. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.

[19] Xubo Lyu and Mo Chen. Ttr-based reward for reinforcement learning with implicit model priors. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5484–5489. IEEE, 2020.

[20] Karttikeya Mangalam, Yang An, Harshayu Girase, and Jitendra Malik. From goals, waypoints & paths to long term human trajectory forecasting. *ICCV*, 2021.

[21] Karttikeya Mangalam, Harshayu Girase, Shreyas Agarwal, Kuan-Hui Lee, Ehsan Adeli, Jitendra Malik, and Adrien Gaidon. It is not the journey but the destination: Endpoint conditioned trajectory prediction. *ECCV*, 2020.

[22] Kenneth O May et al. R. duncan luce, individual choice behavior, a theoretical analysis. *Bulletin of the American Mathematical Society*, 66(4):259–260, 1960.

[23] Katja Mombaur, Anh Truong, and Jean-Paul Laumond. From human to humanoid locomotion—an inverse optimal control approach. *Autonomous robots*, 28(3):369–383, 2010.

[24] Alessandro Vittorio Papadopoulos, Luca Bascetta, and Gianni Ferretti. Generation of human walking paths. *Autonomous Robots*, 40(1):59–75, 2016.

[25] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th international conference on computer vision*, pages 261–268. IEEE, 2009.

[26] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Multi-agent generative trajectory forecasting with heterogeneous data for control. *ECCV*, 2020.

[27] Roger N Shepard. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22(4):325–345, 1957.

[28] Dizan Vasquez, Billy Okal, and Kai O Arras. Inverse reinforcement learning algorithms and features for robot navigation in crowds: an experimental comparison. *IROS*, 2014.

[29] John Von Neumann and Oskar Morgenstern. Theory of games and economic behavior. In *Theory of games and economic behavior*. Princeton university press, 2007.

[30] Insoon Yang, Sabine Becker-Weimann, Mina J Bissell, and Claire J Tomlin. One-shot computation of reachable sets for differential games. In *HSCC*, pages 183–192, 2013.

[31] Zhitian Zhang, Jimin Rhim, Angelica Lim, and Mo Chen. A multimodal and hybrid framework for human navigational intent inference. *IROS*, 2021.

[32] Zhitian Zhang, Jimin Rhim, Mahdi TaherAhmadi, Kefan Yang, Angelica Lim, and Mo Chen. Sfu-store-nav: A multimodal dataset for indoor human navigation. *Data in Brief*, 33:106539, 2020.

[33] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al. Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008.

[34] Brian D Ziebart, Nathan Ratliff, Garratt Gallagher, Christoph Mertz, Kevin Peterson, J Andrew Bagnell, Martial Hebert, Anind K Dey, and Siddhartha Srinivasa. Planning-based prediction for pedestrians. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3931–3936. IEEE, 2009.