Inferring Navigation Policies for Mobile Robots from Demonstrations

Henrik Kretzschmar Markus Kuderer Wolfram Burgard

***Abstract*— We present an approach that allows a mobile robot to learn the behavior of pedestrians from observed trajectories. It maintains probability distributions over trajectories and represents these distributions as a mixture model. The upper level of this model represents a discrete distribution over classes of trajectories that are equivalent according to a set of features, such as passing on the left or passing on the right side. The lower level comprises continuous probability distributions over trajectories for each of these classes and captures physical features of the trajectories, such as velocities and accelerations. For each level, our method learns maximum entropy distributions that match the feature values of the observations. To estimate the feature expectations in the high-dimensional probability distributions over the trajectories, our approach applies Hamiltonian Markov Chain Monte Carlo sampling. The extensive experimental evaluation suggests that our method models human navigation behavior more accurately than state-of-the-art techniques.**

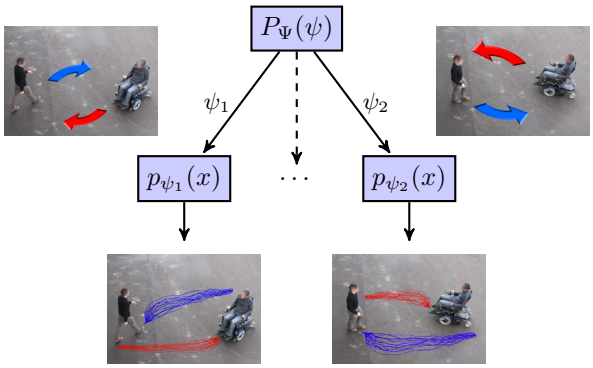

Fig. 1. We use a mixture distribution to model cooperative navigation behavior. The discrete probability distribution P_Ψ describes the choice of homotopy classes, whereas the continuous distributions p_ψ describe which trajectories of the class the agents are likely to choose.

I. Introduction

In the near future, more and more mobile robots are expected to populate our environments and coexist with us. Obviously, these robots will perform better if they understand and comply with mutually accepted social rules and behave accordingly when interacting with humans. In the past, there has been extensive research about modeling social rules and how mobile robots can utilize these models to predict and mimic the navigation behavior of humans. One popular approach to achieve socially compliant robot navigation, for example, is to implement predefined social rules the robot acts on [14, 10].

However, the cooperative behavior of pedestrians and also the desired behavior of the robot typically depends on the corresponding application. For example, a cleaning robot should be unobtrusive and should not unnecessarily hinder people, whereas a transportation robot that supplies an emergency room in a hospital must not delay its task by being overly cautious. Therefore, one needs flexible means for achieving the appropriate navigation behaviors for the applications at hand.

In this paper, we present an approach that utilizes supervised learning techniques to recover the underlying navigation policy from a set of observed trajectories that exhibit the desired cooperative behavior. Thereby we allow that the observation set for learning human navigation policies consists of both observations of pedestrians and also trajectories obtained by tele-operating the robot. While the first approach can directly learn from observing people, the second method requires active human control but allows the robot to learn the desired interaction behavior for a specific task.

Instead of fitting deterministic policies to observed behavior or assuming utility optimizing agents, our approach learns the probability distribution that underlies the demonstrated behavior, which typically is by no means optimal and often shows stochastic properties. We assume that this distribution depends on features that describe relevant properties of the trajectories. We use features that represent individual preferences of individuals, such as accelerations and velocities. Other features describe the interactive navigation behavior, such as distances between the agents, the preferred side of passing, or the acceptability to move through groups of people.

The contribution of this paper is twofold. Firstly, we propose a technique to infer the continuous probability distribution over trajectories using feature-based maximum entropy learning. To compute the feature expectations over the high-dimensional distributions, we use a Markov Chain Monte Carlo method that utilizes the gradient of the features to improve mixing. Secondly, we propose to represent a probabilistic navigation policy as a mixture distribution. The top level of this distribution contains discrete decisions during the navigation process, such as evading on the left or on the right. The bottom level represents continuous distributions over trajectories that depend on physical properties of the trajectories. We show how to infer the probability distributions on both levels using maximum entropy learning.

All authors are with the Department of Computer Science, University of Freiburg, Germany.

This work has partly been supported by the EC under FP7-ICT-248873-RADHAR, by the German Research Foundation (DFG) under contract number SFB/TR-8, and by the Hans L. Merkle-Stiftung.

II. Related Work

Supervised learning methods that aim to recover the underlying policy from observations have been investigated extensively. If the policy can be modeled in terms of a cost function, the problem is known as inverse reinforcement learning. Many authors proposed techniques to learn the underlying cost function of a Markov decision process from observed sequences. Similar to our work, Abbeel and Ng [1] introduced the idea of matching feature values that describe the relevant behavior. However, feature matching is ambiguous since in general there is no unique cost function that explains the behavior of the experts. To resolve this ambiguity, Ziebart et al. [20] proposed maximum entropy inverse reinforcement learning. Inverse reinforcement learning methods have been used in a variety of applications, for example robot control [11] and route planning for outdoor mobile robots [16], and was also used to learn pedestrian navigation behavior [21].

These techniques, however, are limited to discrete Markov decision processes. An alternative way of dealing with the continuous trajectories of pedestrians is to model physical aspects of human navigation behavior and to use features over the continuous space of trajectories. For example, several authors proposed to model pedestrians as utility-optimizing agents that try to minimize a parametrized cost function comprising relevant properties of human motion behavior in continuous spaces. The proposed cost functions represent drifting from the planned trajectory, accelerations, and closeness to other pedestrians [8], smoothness [15] or the derivative of the curvature [3]. Typically, these approaches aim to find parameters that lead to trajectories that are similar to the observations. Mombaur et al. [13] use inverse optimal control to learn human-like navigation policies for humanoid robots. In contrast to these approaches, we do not assume utility optimizing agents. Since the observed trajectories are not optimal, we model the navigation behavior as a probability distribution that depends on a set of features. Furthermore, we do not model individual agents but assume cooperative agents and model the joint interactive behavior, which was also proposed by Trautman and Krause [18].

Cooperative behavior is also modeled by Kuderer et al. [12] who use a spline-based representation and learn a maximum entropy distribution over joint trajectories in a continuous state space. Our approach described in this paper is an extension of the work of Kuderer et al. [12] with respect to two relevant aspects. Firstly, we use a mixture distribution that allows us to learn the distribution over discrete decisions such as the preference to evade others on the left or on the right side. Secondly, Kuderer et al. [12] approximate the expectation of the continuous distribution over trajectories with Dirac delta functions at the modes of the distribution. This approximation underestimates the expected "cost" of the policy compared to the demonstrations that are not optimal with respect to a cost function. Vernaza and Bagnell [19] deal with the inference of continuous trajectories in high-dimensional state spaces by constraining the features to have a certain low-dimensional structure. We allow arbitrary features and compute the feature expectations using a Markov chain Monte Carlo technique that utilizes gradient information of the features. Thus, the method proposed in the following is able to capture important properties of human navigation behavior and can predict their behavior more accurately than state-of-the-art approaches.

III. Learning Navigation Policies for Mobile Robots from Human Demonstrations

The objective of this work is to learn the navigation policies of a set A of cooperatively interacting pedestrians by observing their trajectories

$$x \in \mathcal{X} = \prod_{a \in A} x^a(t), \tag{1}$$

where $x^a(t)$ continuously maps each point in time to a configuration space of agent $a \in A$. We assume that the behavior of the pedestrians can be described by a probability distribution $p(x)$ that depends on features that capture relevant properties of the behavior. We assume that the trajectories of the pedestrians are samples drawn from this distribution. Hence, our goal is to find the distribution for which the expected behavior matches the observed behavior in terms of the features.

A. Modeling Human Navigation Using a Mixture of Maximum Entropy Distributions

In particular, we model the behavior using a mixture distribution, where each mixture component $\psi_j \in \Psi$ represents a class of trajectories that are homotopy equivalent. Two trajectories $x_1, x_2 \in \mathcal{X}$ are homotopy equivalent if each pair of agents passes each other on the same side. The probabilities $P_\Psi(\psi)$ of the mixture components Ψ reflect the probability of choosing a trajectory of the corresponding class. Each component distribution $p_\psi(x)$ describes the probability of the trajectories that belong to ψ. Fig. 1 illustrates this mixture distribution.

We assume that both the discrete mixture proportions $P_\Psi(\psi)$ and the component densities $p_\psi(x)$ depend on features that describe relevant properties of the navigation behavior. A high level feature

$$f_i^{\text{high}} : \Psi \mapsto \mathbb{R} \tag{2}$$

is a function that describes properties of classes $\psi \in \Psi$. Futhermore, a low level feature

$$f_i^{\text{low}} : \mathcal{X} \mapsto \mathbb{R} \tag{3}$$

is a function that captures physical properties of the trajectories $x \in \mathcal{X}$, such as velocities, accelerations, or the clearance between agents. To mimic these physical properties, we want each distribution $p_\psi(x)$ to yield expected feature values that match the empirical feature values of a set $\mathcal{D}$ of observed trajectories $\tilde{x}_j \in \mathcal{X}$:

$$E_{p_\psi}[\mathbf{f}^{\text{low}}] = \tilde{\mathbf{f}}^{\text{low}} = \sum_j \frac{\mathbf{f}^{\text{low}}(\tilde{x}_j)}{|\mathcal{D}|}. \tag{4}$$

Similarly, to capture the humans choice with respect to homotopy classes, we want the distribution $P_\Psi(\psi)$ to yield expected feature values that match the empirical feature values, which leads to

$$E_p[\mathbf{f}^{\text{high}}] = \tilde{\mathbf{f}}^{\text{high}} = \sum_j \frac{\mathbf{f}^{\text{high}}(\psi_{\tilde{x}_j})}{|\mathcal{D}|}, \tag{5}$$

where $\psi_{\tilde{x}_j}$ is the homotopy class of $\tilde{x}_j$.

B. Maximum Entropy and Feature Matching

Our goal is to match the expected feature values of the policy to the feature values of the observed behavior without imposing any further assumptions. In general, there is no unique distribution that matches the features. However, the principle of maximum entropy states that the distribution with the highest entropy represents the given information best [9]. This distribution and the gradient of its parameters are well known in information theory and have been used to infer trajectories in discrete [21] and continuous [12] state spaces. In the following, we will shortly outline the derivation of a continuous maximum entropy distribution under the condition that the feature expectations $E_p[\mathbf{f}]$ match empirical values $\tilde{\mathbf{f}}$.

Our goal is to find the distribution $p^\star$ that maximizes the differential entropy

$$H(p) = \int_x -p(x) \log p(x) dx \tag{6}$$

subject to these constraints. Applying constrained optimization we therefore need to maximize the Lagrangian

$$H(p) - \sum_i \theta_i (E_p[f_i] - f_{i,\mathcal{D}}) - \alpha \Big(\int_x p(x)\, dx - 1 \Big) \tag{7}$$

with respect to the distribution p and the Lagrangian multipliers $[\theta_1, \ldots, \theta_n] = \boldsymbol{\theta}$ and α, where the last term assures that the probability integrates to one. Taking derivatives with respect to p reveals that the desired probability distribution belongs to the exponential family distribution

$$p_{\boldsymbol{\theta}}(x) \sim e^{-\boldsymbol{\theta}^T \mathbf{f}(x)}. \tag{8}$$

The term $\boldsymbol{\theta}^T \mathbf{f}(x)$ can be interpreted as a cost function with $\boldsymbol{\theta}$ as weights. Therefore, states with higher cost are less likely. Unfortunately, it is not feasible to analytically compute the parameters $\boldsymbol{\theta}^\star$ that maximize the Lagrangian. However, we can use its gradient of the dual function with respect to θ, which is given by

$$\mathbf{f}_{\mathcal{D}} - E_p[\mathbf{f}(x)], \tag{9}$$

and apply an optimization technique to determine the maximum entropy probability distribution $p^\star$ that satisfies our constraints. Interestingly, Eq. (9) is equal to the gradient with respect to the likelihood of the observed data under the exponential family distributions. Therefore, the maximum entropy distribution equals the exponential family distribution that maximizes the likelihood of the training data. The derivation of the discrete maximum entropy distribution follows the same idea, where integrals are substituted by sums.

C. Learning the Mixture Distribution

When we apply the principle of maximum entropy to the mixture distribution to model human navigation behavior, we have

$$p_\psi(x) \propto e^{-\boldsymbol{\theta}^{\text{low}} \mathbf{f}^{\text{low}}} \text{ and } P_\Psi(\psi) \propto e^{-\boldsymbol{\theta}^{\text{high}} \mathbf{f}^{\text{high}}}, \tag{10}$$

where we refer to $\boldsymbol{\theta}^{\text{low}}$ and $\boldsymbol{\theta}^{\text{high}}$ as feature weights. Consequently, to learn the mixture distribution we need to find the feature weights that give rise to a distribution whose feature expectations match the empirical feature values, as described in the previous section. Since the features $\mathbf{f}^{\text{low}}$ do not depend on the mixing proportions of the homotopy classes, we can first learn the probability distributions p_ψ and then the mixing proportions $P_\Psi(\psi)$. Note that our model allows a feature $\mathbf{f}^{\text{high}}$ to depend on the probability distribution p_ψ of the trajectories in the corresponding homotopy class. To learn both $\boldsymbol{\theta}^{\text{low}}$ and $\boldsymbol{\theta}^{\text{high}}$, we use the gradient-based optimization technique RPROP [17].

IV. Modeling Physical Properties of Human Navigation Behavior

To capture relevant physical properties of the navigation behavior, we use the features

$$f^a_{\text{acceleration}} = \int_t \|\ddot{x}^a(t)\|^2 \, dt, \tag{11}$$

$$f^a_{\text{time}} = \int_t 1 \, dt, \tag{12}$$

$$f^a_{\text{velocity}} = \int_t \|\dot{x}^a(t)\|^2 \, dt, \tag{13}$$

$$f^a_{\text{distance}} = f^a_{\text{time}} \sum_{b \neq a} \int_t \frac{1}{\|x^a(t) - x^b(t)\|^2} \, dt, \tag{14}$$

which were introduced by Kuderer et al. [12], and we represent the continuous state space over the trajectories of all using cubic $2d$-splines. Our model assumes that the agents know the target positions of all the other agents, whereas it does not assume that the agents know when they reach their targets.

A. Matching Feature Expectations over Trajectories in Continuous State Spaces

Similar to our work, Kuderer et al. [12] aim to learn a maximum entropy distribution that matches empirical feature values. However, they approximate the expected feature values using Dirac delta functions at the modes of the probability distribution. Accordingly, they do not account for the variation of the trajectories in each homotopy class and use only the most likely state of that homotopy class. As a consequence, they underestimate the expected feature values. Our experiments suggest that this approximation leads to suboptimal results, especially when the human behavior is not optimal with respect to the cost function. We will show a comparison using extensive training data in the experimental section.

In the following, we show how to improve this approximation and how to compute the feature expectations $E_{p_\psi}[\mathbf{f}]$ over

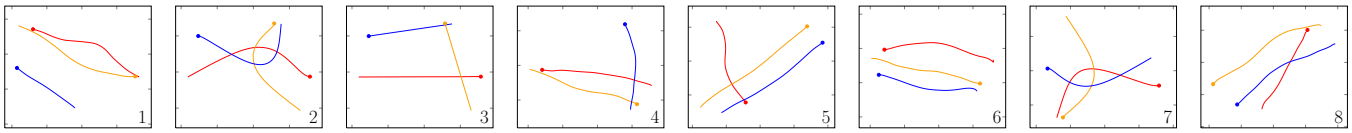

Fig. 2. Examples of trajectories that participants of a Turing test were asked to label as either human or machine. To help the participants understand the dynamics of the interactions of the pedestrians, we showed them animations of the trajectories.

the trajectories. This high-dimensional non-linear distribution impedes evaluating these feature expectations analytically. Monte Carlo sampling methods yet allow us to approximate the expectations using a set of samples independently drawn from the target distribution. However, drawing samples directly from the distribution p_ψ over trajectories is also not feasible. Fortunately, Markov chain Monte Carlo (MCMC) methods can be used to obtain samples from such high-dimensional distributions [2, 4]. These methods aim to explore the state space by constructing a Markov chain whose equilibrium distribution is the target distribution. In particular, the widely-used Metropolis-Hastings algorithm [6] generates a Markov chain in the state space using a proposal distribution and a criterion to accept or reject the proposed steps. This proposal distribution and the resulting acceptance rate, however, have a dramatic effect on the mixing time of the algorithm, e. g., the number of steps after which the distribution of the samples can be considered to be close to the target distribution. In practice, it is often difficult to design a proposal distribution that leads to satisfactory mixing. Simply increasing the step size often leads to a higher rejection rate.

Fortunately, using the Hybrid Monte Carlo algorithm [5], we can improve mixing when sampling from the distribution p_ψ by incorporating information about the gradient with respect to the spline parameters of the trajectories. The algorithm considers an extended target density

$$p(x, \mathbf{u}) = p(x)\mathcal{N}(u; 0, I_n), \tag{15}$$

where $\mathbf{u} \in \mathbb{R}^n$ are auxiliary momentum variables that simulate a fictitious physical system. After performing a number of "frog leaps" in $\mathbf{u}$ and x, Hybrid Monte Carlo relies on the Metropolis-Hastings algorithm to accept or reject the resulting candidate sample drawn from the density $p(x, \mathbf{u})$.

V. Modeling Topological Properties of Human Navigation Behavior

We assume that the decisions of which homotopy class the agents choose can be described by a probability distribution $P_\Psi(\psi)$ that depends on features $\mathbf{f}^{\text{high}}$.

To compute the homotopy class of a trajectory x, we compute on which side each pair of agents passes each other. For all agents a and b, we integrate the derivative of the angle $\alpha_a^b(t)$ of the vector $x^b(t) - x^a(t)$ and the vector $(1, 0)^T$ over time, which leads to

$$\kappa_a^b = \int_t \dot{\alpha}_a^b(t)\, dt. \tag{16}$$

For instance, two agents a and b that pass each other on the right-hand side yield a positive κ_a^b.

A. Physical Properties of Homotopy Classes

We assume that the decision of which homotopy class the agents choose depends on the cost of the trajectories that belong to that homotopy class. We use a feature

$$f^{\text{high}}_{\text{ml\_cost}}(\psi) = \min_{x \in \psi} \boldsymbol{\theta}^T \mathbf{f}^{\text{low}}(x) \tag{17}$$

that corresponds to the cost of the most likely trajectory x of homotopy class ψ. As a result, our model assumes that the agents decide which homotopy class they choose depending on the cost of the most likely trajectory of that homotopy class with respect to the learned distribution $p_\psi(x)$.

B. Tendencies to Pass Left or Right

When avoiding others, most humans prefer either the right- or the left-hand side to pass each other. To capture these tendencies, we use a feature

$$f^a_{\text{angle}} = \sum_{b \neq a} \kappa_a^b. \tag{18}$$

C. Group Behavior

Moving through a group of people might be considered inappropriate and is therefore typically avoided if possible. Our approach aims at learning to which extent the agents avoid to move through groups. To check whether an agent moves through a group of agents, we rely on κ, as defined in Eq. (16), which indicates the side on which two agents pass each other. An agent that passes two members of a group on different sides moves through the corresponding group. Therefore, we have

$$f^a_{\text{group}} = |\{G \in \mathcal{G} \mid \exists b, c \in G : b, c \neq a \wedge \kappa_a^b \kappa_a^c < 0\}|, \tag{19}$$

where $\mathcal{G}$ is the set of groups of agents.

VI. Experimental Evaluation

The goal of this section is to show that our approach is able to learn cooperative behaviors in navigation tasks from human demonstrations. To collect training data, we recorded one hour of encounters of two pedestrians and a wheelchair user in a variety of different situations and tracked their positions using a motion capture system. The participants thereby repeatedly chose an arbitrary target position in an area of 7 m $\times$ 7 m and then approached that position while naturally evading each other. These recordings resulted in 230 demonstrations that exhibit many evasive maneuvers. To compare the performance of our method, we implemented the approach presented by Kuderer et al. [12] and the social forces algorithm proposed by Helbing and Molnar [7]. In the following, we will demonstrate that the probabilistic policies

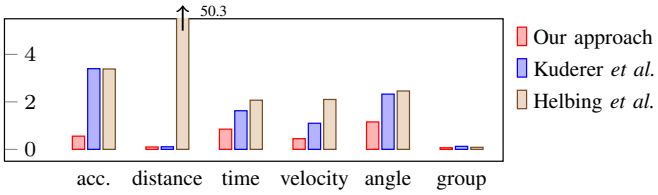

Fig. 3. Cross validations. The results suggest that our approach is able to better capture human navigation behavior in terms of each of the features than the other two methods.

computed by our approach model the observed behavior more accurately with respect to features of the trajectories than these state-of-the-art techniques. Furthermore, we will show that our approach is able to apply learned behavior to different scenarios.

A. Cross Validation

We evaluated how the behavior learned by our approach and by the methods introduced in Kuderer et al. [12] and Helbing and Molnar [7] generalize to different data sets. To allow for a fair comparison, we used the same features to train and evaluate each of the approaches, i.e., the aforementioned features that capture accelerations, velocities, the time to reach the target, distances between the agents, the sides pairs of agents pass, and a feature that indicates whether the agents move through groups of people that belong together. To train the social forces method, we applied stochastic gradient descend to minimize the norm of the error of the feature values.

We performed a 10-fold cross validation on the training set of 230 runs. For each fold, we evaluated the policies that were learned using the remaining folds. Fig. 3 depicts the results. The bars show the mean error of the empirical value and the expected value of each feature. Our approach was able to learn policies that better capture the observed behavior than the other methods with respect to each of the features. Fig. 4 shows how the gradient norm approaches zero while first learning p_ψ and then P_Ψ. Learning p_ψ relies on MCMC sampling and takes about one hour per iteration on a standard desktop computer. Once p_ψ has been computed, learning P_Ψ is carried out within a couple of milliseconds.

The social forces method is a reactive method that is not able to capture the predictive navigation behavior of pedestrians. In contrast to that, our approach reasons about entire trajectories to the target positions.

Kuderer et al. [12] also consider entire trajectories to the target positions of the agents. However, they approximate the feature expectations using Dirac delta functions at the modes of the distribution. As a consequence, they underestimate the feature expectations, and their learning method therefore tends to distribute the probability mass among unlikely homotopy classes to compensate for this approximation error. In the Turing test that we will describe in the next section the trajectories induced by their approach were not regarded as being human-like.

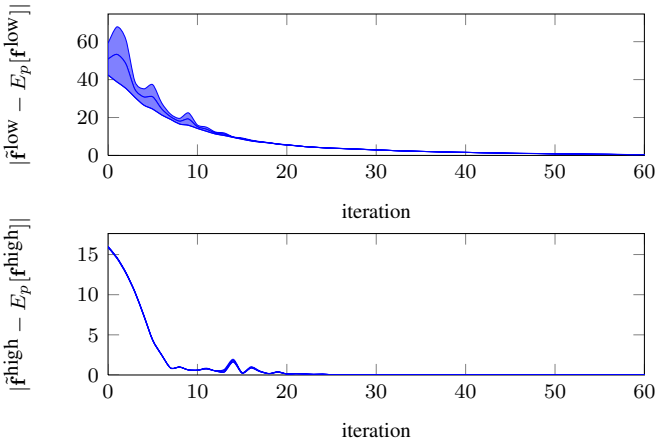

Fig. 4. The evolution of the norm of the discrepancy of the feature expectations and the empirical feature values while learning the feature weights. The figure furthermore depicts the variance of this norm over the training sets of the 10-fold cross validation. Top: Learning the physical properties of human navigation behavior. Bottom: Learning the topological properties of human navigation behavior.

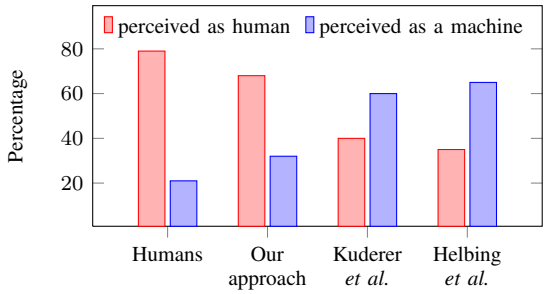

Fig. 5. Results of a Turing test that evaluates whether the behaviors induced by our approach, the approach proposed by Kuderer et al. [12], and the social forces method proposed by Helbing and Molnar [7] qualify as human. The results of the Turing test suggest that the behavior induced by our approach most resembles human behavior.

B. Turing Test

We carried out a Turing test to evaluate how human-like the behavior generated by our approach is. We asked 10 human subjects to distinguish recorded human behavior from behavior generated by an algorithm. We evaluated how well the subjects performed on a test set of runs that were randomly drawn from the recorded human demonstrations. For each of the runs, we showed them animations of trajectories that were either recorded from the human demonstrations or the prediction of an algorithm. In total, we presented 40 runs to each of the human subjects, where the trajectories were equally drawn from the human demonstrations, from the prediction of our approach, from the prediction of the approach presented by Kuderer et al. [12], and from the prediction of the social forces method. Fig. 2 depicts some sample trajectories that were shown to the human subjects, and Fig. 5 summarizes the results. The human subjects correctly identified 79 % of all the human demonstrations, but they mistook 68 % of the predictions of our approach, 40 % of the predictions of the approach by Kuderer et al. [12], and 35 % of the predictions of the social forces method for human

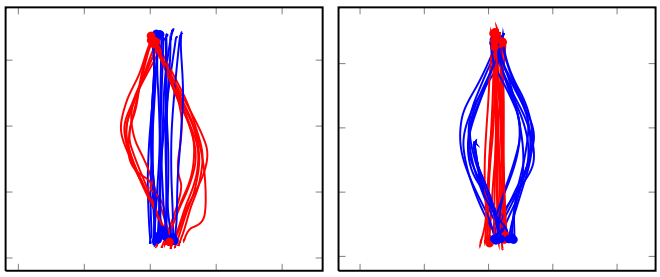

Fig. 6. Trajectories driven by an autonomous robotic wheelchair (red) while interacting with a pedestrian (blue). The robotic wheelchair thereby behaved according to a policy that had been learned by our approach from conservative behavior (left) and from aggressive behavior (right).

behavior. In summary, the results of this Turing test suggest that the behavior induced by our approach can be regarded as significantly more human-like than the behavior induced by the other two methods.

C. Application to Robot Navigation

We applied the presented model of cooperative navigation behavior to mobile robot navigation. Fig. 6 (b) depicts the trajectories of a set of experiments where a robotic wheelchair was controlled by a policy learned by our approach. To interact with the pedestrians, the robot constantly maintains the probability distribution over the trajectories in the current situation and acts according to the predicted cooperative behavior. Whenever the pedestrians do not behave according to its predictions, the robot recomputes the probability distribution and adapts its plan to the trajectory that is predicted to be most likely in the new situation. Our system computes the most likely trajectories in less than 100 milliseconds, which allows the robot to quickly react to changes in its environment.

VII. Conclusion

In this paper we presented a novel supervised approach to learning navigation policies from human demonstrations. Our approach models the observed behavior using a mixture distribution that captures physical features of the trajectories as well as of the people's decisions according to homotopy classes of trajectories. To compute the feature expectations with respect to the high-dimensional continuous probability distributions over trajectories, we apply Hamiltonian Markov chain Monte Carlo sampling. Our method has been implemented and extensively evaluated using real-world data acquired from pedestrians and a wheelchair user. The evaluation of a Turing test indicates that our approach is able to generate highly human-like navigation behavior that can hardly be distinguished by people from trajectories really taken by humans. Furthermore, cross validation experiments demonstrate that our method outperforms two state-of-the-art techniques.

References

[1] P. Abbeel and A.Y. Ng. Apprenticeship learning via inverse reinforcement learning. In *ICML*, 2004.

[2] C. Andrieu, N. De Freitas, A. Doucet, and M.I. Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50(1):5–43, 2003.

[3] G. Arechavaleta, J.-P. Laumond, H. Hicheur, and A. Berthoz. An optimality principle governing human walking. *IEEE Transactions on Robotics (T-RO)*, 24(1):5–14, 2008.

[4] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

[5] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216–222, 1987.

[6] W.K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[7] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical Review E (PRE)*, 51:4282–4286, 1995.

[8] S. Hoogendoorn and P.H.L. Bovy. Simulation of pedestrian flows by optimal control and differential games. *Optimal Control Applications and Methods*, 24(3):153–172, 2003.

[9] E. T. Jaynes. Where do we stand on maximum entropy. *Maximum Entropy Formalism*, pages 15–118, 1978.

[10] R. Kirby, R. Simmons, and J. Forlizzi. Companion: A constraint optimizing method for person-acceptable navigation. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 607–612, 2009.

[11] J.Z. Kolter, P. Abbeel, and A.Y. Ng. Hierarchical apprenticeship learning with application to quadruped locomotion. *Advances in Neural Information Processing Systems*, 20:769–776, 2008.

[12] Markus Kuderer, Henrik Kretzschmar, Christoph Sprunk, and Wolfram Burgard. Feature-based prediction of trajectories for socially compliant navigation. In *Proc. of Robotics: Science and Systems (RSS)*, Sydney, Australia, 2012.

[13] K. Mombaur, A. Truong, and J.-P. Laumond. From human to humanoid locomotion – an inverse optimal control approach. *Autonomous Robots*, 28:369–383, 2010.

[14] A.K. Pandey and R. Alami. A framework for adapting social conventions in a mobile robot motion in human-centered environment. In *International Conference on Advanced Robotics (ICAR)*, pages 1–8, 2009.

[15] Q.-C. Pham, H. Hicheur, G. Arechavaleta, J.-P. Laumond, and A. Berthoz. The formation of trajectories during goal-oriented locomotion in humans. II. a maximum smoothness model. *European Journal of Neuroscience*, 26:2391–2403, 2007.

[16] N.D. Ratliff, J.A. Bagnell, and M.A. Zinkevich. Maximum margin planning. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 729–736. ACM, 2006.

[17] M. Riedmiller and H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *Proceedings of the IEEE International Conference on Neural Networks (ICNN)*, pages 586–591, 1993.

[18] P. Trautman and A. Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 797–803, 2010.

[19] P. Vernaza and D. Bagnell. Efficient high dimensional maximum entropy modeling via symmetric partition functions. In *Advances in Neural Information Processing Systems 25*, pages 584–592, 2012.

[20] B.D. Ziebart, A. Maas, J.A. Bagnell, and A.K. Dey. Maximum entropy inverse reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pages 1433–1438, 2008.

[21] B.D. Ziebart, N. Ratliff, G. Gallagher, C Mertz, K. Peterson, J.A. Bagnell, M. Hebert, A.K. Dey, and S. Srinivasa. Planning-based prediction for pedestrians. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3931–3936, 2009.