Towards Socially Competent Navigation of Pedestrian Environments

Christoforos I. Mavrogiannis and Ross A. Knepper

***Abstract*—We present a planning framework for producing socially competent robot behaviors in pedestrian environments. The framework is designed according to conclusions of recent psychology studies on action interpretation and sociology studies on human pedestrian behavior. The core of the approach is a novel topological representation of the pedestrian scene, based on braid groups. Thanks to this representation, our online algorithm is able to reason about several topologically distinct scene evolutions, simultaneously predicting future behaviors of other agents and planning the robot?s role in the scene. This is especially important in crowded pedestrian scenes of high uncertainty. Preliminary simulation results demonstrate the potential of our approach for application in real world scenarios.**

I. INTRODUCTION

Navigating a pedestrian workspace is a hard problem in robotics, due to the lack of formal rules guiding traffic and the lack of explicit communication among pedestrians. Humans relax those complications through implicit communication, implemented by encoding their intentions into their motion. This allows them to negotiate and agree on a joint motion plan. This form of negotiation, described by sociologists [4] as *pedestrian bargain*, is guided by two foundational principles: (1) *people must behave like competent pedestrians* and (2) *people must trust copresent others to behave like competent pedestrians*. The enforcement of the pedestrian bargain imposes a form of *trust* among pedestrians, which is described as the foundation of socially competent behavior. Therefore, imbuing robots with a humanlike ability to seamlessly navigate pedestrian environments, requires a planning framework that reproduces a human-robot pedestrian bargain.

The first step towards this goal is enabling robots to reproduce the implicit communication that the pedestrian bargain relies on. To do so, we need to understand its underlying *language*, i.e., understand how humans act in front of others and interpret observed actions. Deriving a model of this mechanism is important not only for understanding human actions but also for planning socially acceptable robot behaviors. According to Csibra and Gergely [2], humans tend to attribute context-specific intentions to observed behaviors. Following this reasoning, Dragan and Srinivasa [3] formalized criteria for generating intent-expressive robot motion, based on the interplay between *action* and *goal*.

Following the aforementioned principles, we develop a framework for generating intent-expressive robot motion in the context of a crowded pedestrian scene. To this end, we model a collection of observed trajectories to be the *action* and a topological pattern of agents' future trajectories to be the *goal*. We employ braid groups [1] as a tool to symbolically characterize a pattern of trajectories.

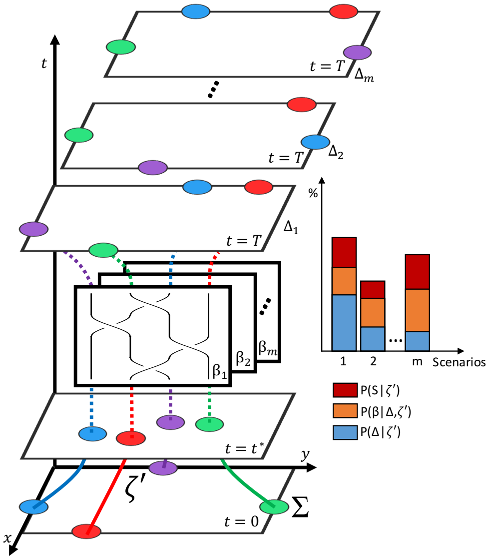


Fig. 1: Conceptual visualization of our model and planning architecture. Four agents are moving from their starting locations, Σ, to their intended destinations. At time $t = t^*$, the red agent has formed a belief about the emerging *pedestrian scenario*, $P(S|\zeta')$, based on a distribution over agents' destinations, $P(\Delta|\zeta')$, and a distribution over the emerging topological pattern of their future trajectories, $P(\beta|\Delta, \zeta')$, given observations of all agents' trajectories so far ζ' (depicted in solid lines). The planner will then generate a trajectory that belongs to the most likely class S and best communicates the robot's role in that class.

Our planner considers several different, topologically distinct potential scenarios, scores them according to their likelihood of occurrence in the observed pedestrian context and generates motion that best communicates the robot's role in the most likely emerging scenario. The main goal of our planning framework is to generate motion that manipulates the observers' belief towards the scenario that appears to be the most socially acceptable and likely at a given time, while ensuring smooth overall behavior.

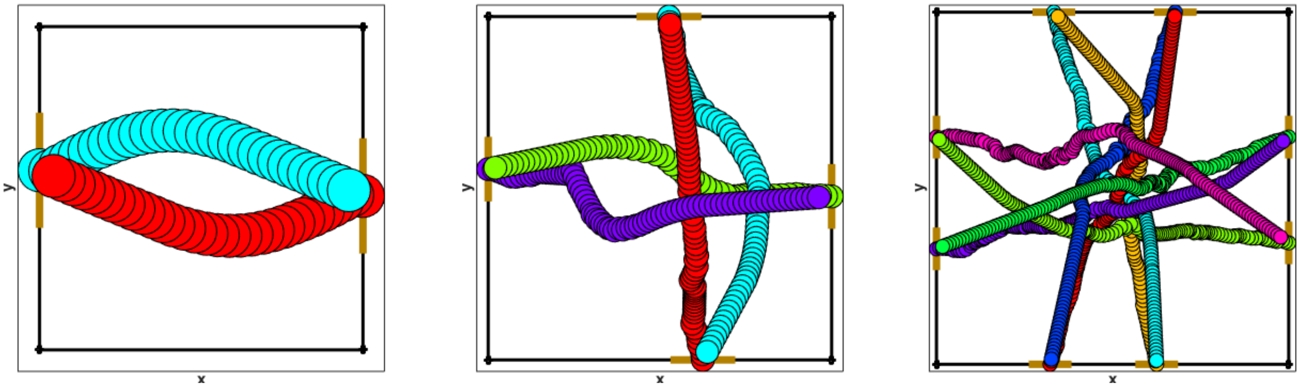
II. APPROACH

A. Pedestrian Scene Model

Consider n agents moving in a workspace $\mathcal{W} \subset \mathbb{R}^2$. Each agent a follows a time-parametrized trajectory ξ_a from an initial position $s_a \in \mathcal{W}$ towards an intended destination $d_a \in \mathcal{D}$, where $\mathcal{D} \subset \mathcal{W}$ is a finite set of possible destinations. We collect the starting locations and the intended destinations of all agents in the tuples $\Sigma = \langle s_1, ..., s_n \rangle$ and $\Delta = \langle d_1, ..., d_n \rangle$ respectively, and their trajectories in the tuple $\zeta = \langle \xi_1, ..., \xi_n \rangle$. The agents' trajectories form a topological pattern τ in space-time, as they connect their initial locations to their intended

---

C. I. Mavrogiannis is with with the Sibley School of Mechanical & Aerospace Engineering, Cornell University and R. A. Knepper is with the Department of Computer Science, Cornell University. Email: cm694@cornell.edu, rak@cs.cornell.edu.



(a) Two agents. (b) Four agents. (c) Eight agents.

Fig. 2: Online Execution: Swept volumes of agents navigating a common workspace in three different scenarios.

destinations. We formally model this pattern using braid groups [1]. Using the aforementioned definitions, we abstract a pedestrian scene into a *Pedestrian Scenario* $S = \langle \Sigma, \Delta, \beta \rangle$. Fig. 1 illustrates our pedestrian scene abstraction.

B. Pedestrians' Inference Mechanism Model

As pedestrians move in the workspace, they make inferences regarding the future behaviors of other agents and their own role in the emerging pattern of trajectories. We model their inference mechanism as a distribution over pedestrian scenarios (see Fig. 1), given a set of observed, partial trajectories ζ', i.e., $P(S|\zeta')$, which we decompose into 1) a prediction of destinations $P(\Delta|\zeta')$ and 2) a prediction of the entanglement pattern $P(\beta|\Delta, \zeta')$:

$$P(S|\zeta') = P(\Delta|\zeta')P(\beta|\Delta, \zeta'). \tag{1}$$

C. Socially Competent Motion Generation

Our motion planning scheme generates information-rich robot motion as a reaction to a weighed consideration of several topologically distinct pedestrian scenarios (see Fig. 1). It simultaneously predicts and plans collective motion by explicitly considering the effect of each agent's behavior on other agents' decisions, following the conclusions of Csibra and Gergely [2] regarding action interpretation. This is implemented by employing the *predictability* and *legibility* metrics, formalized by Dragan and Srinivasa [3]. Intuitively, *predictable* is the motion that an observer expects from an agent whose goal is known, whereas *legible* is the motion that allows an observer to quickly, confidently and accurately predict the agent's goal.

At every replanning cylce, our planner determines a set of topologically distinct pedestrian scenarios $\mathcal{S}$, and represents them as collections of *predictable* trajectories. The robot's trajectory is synthesized from *legible* motion reactions to all scenarios in $\mathcal{S}$, weighed according to $P(\mathcal{S}|\zeta')$. Legible motion serves as a manipulation of pedestrians' belief towards the emerging scenario. By considering a set of scenarios, instead of committing to a single prediction, our algorithm is robust to the uncertainties of pedestrian environments.

III. SIMULATION RESULTS

Fig. 2a, Fig. 2b and Fig. 2c depict the swept volumes of 2, 4 and 8 agents respectively in different scenarios. Each agent is running a separate online instance of our algorithm and plans its motion without communicating at all with the other agents. Our algorithm appears to result in a faster consensus regarding the joint motion plan that the agents will be following to adjust their initially conflicting strategies.

For the generation of the presented simulations, the distributions $P(\Delta|\zeta')$ and $P(\beta|\Delta, \zeta')$ were analytically approximated, both favoring low energy scenarios and penalizing costly ones, under the assumption of rational action. Predictability was defined using a joint efficiency cost, while legibility was modeled as a clearance functional, favoring trajectory modifications that occur early. The trajectory optimization computations are implemented using CHOMP [5].

Future work involves learning a model of $P(\beta|\Delta, \zeta')$ from human data and conducting real world experiments with our robot platform to assess the performance of our approach.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1526035. We are grateful for this support.

REFERENCES

[1] J. S. Birman. *Braids Links And Mapping Class Groups*. Princeton University Press, 1975.

[2] G. Csibra and G. Gergely. 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, 124(1):60–78, Jan. 2007.

[3] A. D. Dragan and S. Srinivasa. Integrating human observer inferences into robot motion planning. *Auton. Robots*, 37 (4):351–368, 2014.

[4] N. H. Wolfinger. Passing Moments: Some Social Dynamics of Pedestrian Interaction. *Journal of Contemporary Ethnography*, 24(3):323–340, 1995.

[5] M. Zucker et al. Chomp: Covariant hamiltonian optimization for motion planning. *The International Journal of Robotics Research*, 32(9-10):1164–1193, 2013.