

Core Challenges of Social Robot Navigation: A Survey

Paper:

Core Challenges of Social Robot Navigation: A Survey

CHRISTOFOROS MAVROGIANNIS, Paul G. Allen School of Computer Science & Engineering,
University of

Washington, USA

FRANCESCA BALDINI, Honda Research Institute and California Institute of Technology, USA

ALLAN WANG, Robotics Institute, Carnegie Mellon University, USA

DAPENG ZHAO, Robotics Institute, Carnegie Mellon University, USA

PETE TRAUTMAN, Honda Research Institute, USA

AARON STEINFELD, Robotics Institute, Carnegie Mellon University, USA

JEAN OH, Robotics Institute, Carnegie Mellon University, USA

Summary:

Introduction

- The vision of mobile robots navigating seamlessly through public spaces has inspired a significant amount of research over the past three decades.
- While early efforts were driven by engineering principles, it quickly became evident that the problem is inherently interdisciplinary, requiring insights from fields such as human-robot interaction, psychology, design research, and sociology.
- This understanding resulted in the emergence of a dedicated field of research, which is commonly referred to as social robot navigation, broadly addressing all algorithmic, behavioral, evaluation, and design aspects of the problem.

Decoupled Prediction and Planning:

- Decoupled approaches can be further classified into a set of dominant trends:
 - a) planning while *treating humans as dynamic, but non-responsive obstacles*
 - While these works achieved important milestones in terms of autonomous navigation capabilities and serving guidance requests, they totally ignored the human component.
 - Early works' underlying navigation frameworks treated humans as non-reactive obstacles without explicitly modeling human navigation, social engagement or interaction with and between humans.
 - The lack of models for human behavior tends to result in subtle robot behaviors that negatively impact user experience. For instance, robots that treat humans as non-reactive obstacles (and thus do not make predictions about human motion) tend to hinder or block human paths with their motion.
 - b) planning using interaction-agnostic models of uncertainty
 - c) planning by following hand-designed social objectives.

Coupled Prediction and Planning:

- Decoupled approaches tend to follow the conventional paradigm of decoupling motion prediction from motion planning, often resulting in undesired phenomena such as the freezing robot problem or the reciprocal dance problem.
- These observations motivated a new research trend focusing on modeling multi-agent interaction in crowd domains as cooperative collision avoidance (CCA).
- The authors observe two main trends depending on the type of cooperation models they employ:
 - explicit approaches
 - Explicit approaches leverage insights about the structure of multiagent collision avoidance to couple prediction and planning.
 - implicit approaches
 - Implicit approaches leverage insights about the principles underlying cooperative collision avoidance but do not impose explicit constraints on its structure.

Open Problems and Directions for Future Work

- Despite substantial and sustained interest in planning framework design for social robot navigation, several fundamental problems remain:
 - Modeling Interaction
 - Much of the motivation for coupled prediction and planning stems from the inability of a robot to move through a dense crowd unless the robot accounts for human cooperation.
 - Unfortunately, coupled models—where each agent's movement is dependent on and influences every other agent's movement—are, in general, computationally intractable.
 - Behavioral Assumptions
 - A significant design decision for any planning framework involves the assignment of a behavioral model for co-navigating humans.
 - Early works in the field made the assumption that other agents move as dynamic obstacles with primitive or nonexistent models of intelligence. This turns out to be problematic in practice because humans tend to attribute intentions to observed actions of others.
 - Much like humans, a competent robot must be able to adjust its behavior based on various types of human pedestrian behavior.
 - Existing works often make the assumption of homogeneity, i.e., all agents are assumed to be making navigation decisions by using the same exact planning mechanism.
 - Further, several works still often make the assumption that others move by propagating their current velocities forward at every timestep. Moving forward in terms of algorithm design would inevitably require leveraging more high-fidelity, high-granularity models of high-level behavior, capturing richer aspects of human decision making.
 - The role of context
 - The majority of prediction and planning approaches do not explicitly model the context in which the robot is deployed.
 - However, we are aware of the importance that the shape of a space (e.g., open space or narrow hallway), the timing (e.g., rush hour in a metro station or a museum) or the semantic map (e.g., at airports there are often queues next to the gates) plays on the type of crowd interactions.
 - Contexts have not been fully explored in the social robot navigation research.

Proxemics

- Proxemics describes a space around a pedestrian that, if intruded, might cause the pedestrian to feel discomfort.

Intentions

- Intentions often represent rich signals that offer hints to the future behavior of an agent.

- Literature from psychology and cognitive science has shown that humans tend to interpret observed actions as goal-directed. Therefore, by designing robot motion that is indicative of the robot's goal could enable an observer to infer the robot's future behavior.
- Legible robot motion may yield higher comfort for the co-navigating human.
- Modalities such as body posture, eye gaze, arm/hand gestures and facial expressions could also provide rich intention signals.
- Lichtenthäler et al. showed that legible motions increase the perceived safety by pedestrians.
 - C. Lichtenthäler, T. Lorenzy, and A. Kirsch. 2012. *Influence of legibility on perceived safety in a virtual human-robot path crossing task. In Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 676–681.

Formations and Social Spaces

- Spatial behavior of the pedestrians and their encompassed social spaces contain rich social signals. These signals guide a robot by informing the robot which social areas are inappropriate to trespass and by offering the robot additional insights to help with pedestrian motion tracking and predictions.
- Most popular research on pedestrian spatial behavior is on grouping.
 - Grouping is thought by some as a synchronization process where a few individuals attune their behavior to that of the people around them.
 - Some have investigated how robots can adjust their behavior to fit into human groups.
- Current works on groups can be divided into static and dynamic settings:
 - Static groups:
 - In static groups, people do not move and form interactions, such as conversation groups. Social space formed within static group formations should be respected by mobile robots.
 - Although sometimes not explicitly related to navigation, these static groups should be counted as impassable spaces and should be avoided during navigation.
 - Dynamic groups:
 - Dynamic groups are groups implicitly formed by moving pedestrians. Similar to static groups, social spaces are present within dynamic groups and are best not to be interrupted.

Open Problems and Directions for Future Work

- Despite the large body of work on pedestrian behavior, the biggest behavioral challenge of social navigation is still on understanding pedestrian behavior.
- At the current stage, research in pedestrian behavior understanding takes a trial-and-error based approach, where researchers design a model for a specific aspect of pedestrian behavior and often show in a small-scale lab study that the model shows promise.

- While the lab studies and validation methods used to establish scientific significance of these models are sufficient, lab studies are often heavily controlled and lack authenticity when compared to real world scenarios.
- Pedestrian behavior analysis is a conglomeration of the various behavioral aspects that potentially benefit social navigation. Researchers often focus on one of these aspects to explore.
 - However, it is often unclear how much the inclusion of such particular behavioral aspects benefit social navigation.

Core Evaluation Challenges of Social Navigation

- An agent's navigation performance can be evaluated based on how successful in arriving at a destination.
 - It is generally subjective to rank alternatives according to multiple preferences or soft constraints.
 - Social compliance, for example, is a soft constraint that is comprised of many aspects including safety, comfort, naturalness, and other societal preference

Metrics

- Navigation success rate:
 - The arrival rate measures how often an agent reaches its goal.
 - ex) A time constraint can be added to enforce an agent to reach its destination within a certain deadline.
 - The arrival rate depends not only on an agent's navigation algorithm but also on its environmental context such as the crowd density.
 - Combined with the average speed of agents, the arrival rate can be used to measure the level of congestion of an environment.
- Path Efficiency and Optimality
 - Path-efficiency metrics measure the quality of the trajectories.
- Safety / Collision Avoidance
 - The safety performance can be measured in terms of the number of constraint violations, for instance, the number of collisions.
- Behavioral Naturalness
 - Naturalness is not directly connected to a planning performance, but the concept represents a fitness score for a robot to blend in a human environment.
 - It is a common method to evaluate the naturalness of a robot navigation in terms of the distance between an agent's plans and the actual trajectories taken in the recordings.
 - Two distance metrics that are prominent in existing works are the average displacement error(ADE) and the final displacement error (FDE).
 - ADE is the average of Euclidean differences between temporally aligned points in the two paths given by a navigation algorithm and the ground truth path from the dataset, respectively.
 - FDE is the displacement error only at the final time step.

Evaluation Methodologies

- Datasets
 - Using pedestrian datasets for evaluation is favored due to two reasons:
 - First, although the recorded behaviors are not fully representative of human pedestrians, the datasets still contain the trajectories of humans navigating in a real environment.
 - Second, it is practically much more convenient to evaluate on a dataset when compared to testing in a real human environment.
- Simulation
 - Real-life experiments are usually expensive in both time and resources. Crowd simulators can be easily customized to support new tasks and environments in a way that would not be otherwise possible in the real world. They are essential tools to evaluate social navigation algorithms and train data-driven or reinforcement learning models.
 - Simulator software:
 - PedSim, OpenSteer, Menge, Continuum, both written in C++, are the most common tools used to simulate crowd scenarios.

Open Problems and Directions for Future Work

- Limitations of Existing Simulation Practices
 - it is crucial to carefully construct an experimental setup that can provide meaningful insights.
 - This requires determining the models governing the behavior of simulated humans, the specific scenarios in which we evaluate an algorithm, the metrics with respect to which we evaluate them, and more.
- Behavioral Model Assumptions
 - Current crowd models fail to describe the relationships between human behaviors systematically and to enable heterogeneity.
 - One way to improve existing simulator suites would be to consider combining different modeling approaches that increase the crowd heterogeneity while modeling and representing complex individual behavior in different scenarios.

Conclusion

- The authors shared thoughts on fundamental challenges underlying the problem of planning robot motion in crowded human environments. Specifically, planning in such spaces is NP-hard, and the non-convexity of balancing standard efficiency/safety tradeoffs prohibit guaranteeing practical performance in real-world applications.
- Additional complications arise from the difficulty of modeling human agents, and from the sensitivity to different types of contexts.