Learning to Navigate in Human Crowds

Abdulrahman Alabdulkareem
CSAIL MIT

Meshal Alharbi
LIDS MIT

Mohamad Alrished
CSE MIT

Abstract

Social robot navigation requires that the robot follows social norms while navigating towards its goal. Current algorithms model pedestrians as independent agents, making the problem computationally intractable and degrading overall performance in dense human crowds. In this work, we explore various ways to enhance the robot's performance in environments with a higher number of pedestrians and aim to achieve well-behaved scaling. Specifically, we compare different approaches common in the literature such as state reduction, reward shaping, and curriculum learning. We find that the use of curriculum learning closely approximates optimal (human-like) behavior. This report serves as a supplemental information to the presentation. Please refer to the presentation for the results of the experiments including animated visualizations.

1 Introduction

In the context of social robot navigation, the primary goal is for the robot to reach its target while sharing space with humans and other entities, ensuring that it avoids any collisions in the process. A particularly complex issue in the development of social robot motion plans is the management of increasing computational complexity when avoiding dynamic obstacles. Current methods largely depend on simulating interactions on a one-to-one basis. However, early findings indicate that the computational load escalates significantly with the addition of more individuals, presenting scalability challenges at larger scales. Furthermore, several techniques that were introduced that intended to enhance scalability—like attention mechanisms (1; 2)—have shown limited efficacy in improving performance.

When the obstacles have structured dynamics (e.g., grouped crowds (3)), we hypothesize that Reinforcement Learning (RL) agents can be trained to navigate dynamic obstacles effectively, even in scenarios with a large number of individual obstacles using appropriate engineering of the state representation as well as a properly engineered reward function which we expect to be the main difficulties in achieving a strong policy.

Solving this challenge will help in ensuring safety, as robots must navigate without posing risks to humans. Additionally, it can significantly boost operational efficiency in various settings, such as healthcare and logistics, by improving the robot's speed. Achieving natural and unobtrusive robot navigation is crucial for public acceptance, as it requires robots to understand and predict human behaviors to coexist comfortably.

2 Related Work

Navigating in human-populated environments poses some significant challenges that have been addressed in literature. One such method, namely (4), is the integration of social norms and behaviors into robot path planning. The work investigates the incorporation of social behavior models into the decision-making processes of robots. Similar approach (1), where attention mechanisms are utilized to enhance the robot's awareness of the crowd's dynamics. The work emphasizes the importance of

6.8200 Computational Sensorimotor Learning Final Project (Spring 2024).

"pooling" the different elements of the crowd to improve navigation strategies in complex, densely populated environments. Other work (2) further builds on these ideas by proposing a model that recognizes and respects the cohesive nature and not to intrude into pedestrian groups. (3) extends these ideas by specifically focusing on predicting the motion of groups in crowded settings.

Another line of work focuses on avoiding freezing behavior. (5) introduces a novel concept of creating temporary *freezing zones* around pedestrians to ensure safety without halting the robot's movement completely.

3 Problem Formulation

In social robot navigation, the goal is to learn a control policy that allows a robot to navigate to a target location while sharing space with pedestrians (6). An optimal policy should minimize the number of steps required to reach the target location while avoiding collisions and adhering to social norms. This task can be formulated as a Markov decision process. In particular, if K is the total number of pedestrians, the state and action spaces can be defined as follows:

State space $\mathcal{S}$: For each pedestrian $i \in [N]$[1] in the scene, their observable state information is captured by the following vector:

$$\text{ped}_i = (p_x^i, p_y^i, v_x^i, v_y^i, r^i) \tag{1}$$

where p_x^i and p_y^i are the pedestrian's x and y coordinate, v_x^i and v_y^i are the pedestrian's x and y velocity, and r is the radius of the pedestrian. For the robot, the state information vector is defined as:

$$\text{rob} = (p_x, p_y, v_x, v_y, r, g_x, g_y) \tag{2}$$

where the first four quantities are as defined for the pedestrians, g_x and g_y are the x and y coordinates of the goal. Therefore, a state $s \in \mathcal{S}$ that describes the full state of the environment at a particular point in time will be given by:

$$s = (\text{rob}, \text{ped}_1, \text{ped}_2, \cdots, \text{ped}_N) \tag{3}$$

We note that such state representation is common in the literature (4).

Action space $\mathcal{A}$: We assume holonomic kinematics for the robot. That is, the robot can move in any direction independent of its current orientation (i.e. it's last step direction). Thus, the actions $a \in \mathcal{A}$ can be described by a tuple (v_x^a, v_y^a) where v_x^a and v_y^a are the one step x and y velocity vector (i.e., $(p_x)_{t+1} = (p_x)_t + \Delta t(v_x + v_x^a)$ where Δt is the discretization of time).

Dynamic model: Several models are considered for our choice of the dynamic model. The Social Force Model is a dynamic model that conceptualizes human behavior within social contexts as a sum of forces on a particle. Essentially, the model treats individuals as particles subject to social forces—such as attraction to a destination or repulsion from other individuals—which influence their motion. These forces represent the "internal impulses" that drive individuals towards their goals while avoiding obstacles (7). Another notable dynamic model is the Optimal Reciprocal Collision Avoidance (ORCA) (8). ORCA functions as a dynamic model by continuously adapting to the changing positions and velocities of multiple agents within a shared environment under reciprocal assumptions. Note that both models are reactive methods in multi-agent navigation setting (1).

Objective function: Let $d_{\text{goal}}(t)$ be the distance of the agent to goal at time t. $d_{\text{coll.}}$ is the fixed collision distance. $d_{\text{disc.}}$ is the fixed discomfort distance.

The objective is to find a policy which makes the agent we want to control reaches the goal $(d_{\text{goal}}(T) < d_{\text{coll.}})$ for some value T while ensuring not to collide with pedestrians $\sum_i \mathbb{1}(d_i(t) < d_{\text{coll.}}) = 0, \forall t : 0 \leq t \leq T$ while also minimizing the discomfort criteria $\sum_i (d_{\text{disc.}} - d_i(t)) \mathbb{1}(d_{\text{coll.}} \leq d_i(t) \leq d_{\text{disc.}})$.

Success Criterion: We define the success criterion for a single episode of the environment if the agent manages to reach the goal $(d_{\text{goal}}(T) < d_{\text{coll.}})$ within the time limit without colliding with any pedestrians. If the time limit is reached or the agent collides with a pedestrian then the episodes terminates with a failure. Additionally, if the agent intrudes into a group, the episode terminates with a failure. Note that the aforementioned condition will be relaxed in the curriculum learning setting.

[1]For any integer $n \in \mathbb{Z}^+$, we define $[n] \coloneqq \{1, 2, \ldots, n\}$.

Reward function $R(t)$: Several reward function designs exist for social robot navigation. In this project, we aim to identify a suitable reward function that allows for group abstraction. Initially, we adjust the reward function defined in (2) to account for group abstraction. For the updated state vector, we drop the group intrusion cost in (2) and their reward function is reduced to:

$$\begin{aligned} R(t) =& C_{\text{prog.}} \left(d_{\text{goal}}\left(t-1\right) - d_{\text{goal}}\left(t\right)\right) \\ &+ C_{\text{goal}}\, \mathbb{1}\left(d_{\text{goal}}\left(t\right) < d_{\text{coll.}}\right) \\ &- C_{\text{disc.}} \sum_i \left(d_{\text{disc.}} - d_i(t)\right) \mathbb{1}\left(d_{\text{coll.}} \leq d_i(t) \leq d_{\text{disc.}}\right) \\ &- C_{\text{coll.}} \sum_i \mathbb{1}\left(d_i(t) < d_{\text{coll.}}\right) \end{aligned} \tag{4}$$

where $C_{\text{prog.}}$ is progression reward, C_{goal} is getting to goal reward, $C_{\text{disc.}}$ is the discomfort cost and $C_{\text{coll.}}$ is the collision cost. Where the objective is to maximize the cumulative discounted rewards.

Base Learning Algorithm: We intend to use Proximal Policy Optimization (PPO) as our base learning algorithm, as it is a natural fit for our problem, and GroupNav (2) which is a social robot navigation algorithm based on PPO. In the following section, we describe the specific changes we are proposing to our algorithms of choice.

4 Experiments Details

In this section, we provide further details for the results shown in the presentation.

4.1 Reduced state representation

In our first experiment, we modify GroupNav to reduce the dimensionality of the state space $\mathcal{S}$. Initially, we assume the knowledge of group membership. This assumption is reasonable since there exist a number of group identification algorithms, such as (9), which can be used as an input to our work. Let $\{G_1, G_2, \cdot, G_k\}$ be a grouping of the pedestrian, such that $G_i \cap G_j = \phi$ if $i \neq j$ and $G_1 \cup G_2 \cup \cdots \cup G_k = [N]$. We define the group centroid's position and velocity as:

$$(p_x^{G_j}, p_y^{G_j}, v_x^{G_j}, v_y^{G_j}) = \frac{1}{|G_j|} \sum_{i \in G_j} (p_x^i, p_y^i, v_x^i, v_y^i) \tag{5}$$

For r^{G_j}, we define it as the minimum radius (from the group centroid $(p_x^{G_j}, p_y^{G_j})$) that covers all the penetrations in a group G_j. Thus, the reduced state representation will be:

$$\text{group}_j = (p_x^{G_j}, p_y^{G_j}, v_x^{G_j}, v_y^{G_j}, r^{G_j}) \tag{6}$$

$$s_{\text{reduced}} = (\text{rob}, \text{group}_1, \text{group}_2, \cdots, \text{group}_K) \tag{7}$$

Hence, using the reduced state representation reduces the group to a circle defined by the group centroid $(p_x^{G_j}, p_y^{G_j})$ and radius r^{G_j}. Note that the robot information vector rob remains unchanged.

4.2 Reward shaping

For reward shaping, our goal is to incentivize the agent to anticipate the pedestrian path and avoid it. First, we try modifying the value of $C_{\text{disc.}}$. Moreover, we consider the alternative reward function:

$$\begin{aligned} R(t) =& C_{\text{prog.}} \left(d_{\text{goal}}\left(t-1\right) - d_{\text{goal}}\left(t\right)\right) \\ &+ C_{\text{goal}}\, \mathbb{1}\left(d_{\text{goal}}\left(t\right) < d_{\text{coll.}}\right) \\ &- C_{\text{avoid.}} \sum_i \left(d_{\text{disc.}} - d_i(t)\right) \\ &- C_{\text{coll.}} \sum_i \mathbb{1}\left(d_i(t) < d_{\text{coll.}}\right) \end{aligned} \tag{8}$$

Notice in this modified function we remove the binary indicator from the third term, leading to a smoother function. In the presentation, we show the training performance versus different values of $C_{\text{avoid.}}$.

4.3 Curriculum Learning

In our previous experiments, we notice that the models seem to always prefer walking in a straight line towards the goal without avoiding pedestrians which causes a collision and a failed termination. Thus, we hypothesize that the original environment is too difficult for models to learn from and models were stuck in local optimums where they exhibit the behaviour of walking in a straight path towards the goal without attempting to avoid obstacles or learning a more complicated path-finding behavior.

Thus, we develop a modified environment that is considerably easier and less punishing to the agent by relaxing the group intrusion termination condition. Training on this simpler environment made the agents quickly learn some basic obstacle-avoiding behavior due to the environment being simpler and less punishing, then we transfer this agent to the more difficult environment (the original one) which causes it to have a strong head-start in performance. At the end of training, the agent achieves an increased success rate in the original environment (70% success rate) compared to other models that were not given a curriculum to learn from (60% success rate).

5 Conclusions

In this project, we investigate the role of different methodologies attempting to improve the scaling behavior of the pedestrian navigation environment as we increase the number of human pedestrians. We show that state reduction for the problem of social robot navigation offers slight improvements in learning efficacy for models. Moreover, models trained in stricter environments tended to get stuck in local optimums. As the number of humans in the simulation increased, the models struggle to escape these local optimums. To address this, the implementation of curriculum learning proved beneficial. Thus, the use of curriculum learning shows a promising direction to find policies that mimic human behavior to navigate densely populated environments more effectively.

References

[1] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, "Crowd-Robot Interaction: Crowd-aware Robot Navigation with Attention-based Deep Reinforcement Learning," Feb. 2019, arXiv:1809.08835 [cs]. [Online]. Available: http://arxiv.org/abs/1809.08835

[2] K. Katyal, Y. Gao, J. Markowitz, S. Pohland, C. Rivera, I.-J. Wang, and C.-M. Huang, "Learning a Group-Aware Policy for Robot Navigation," Jul. 2022, arXiv:2012.12291 [cs]. [Online]. Available: http://arxiv.org/abs/2012.12291

[3] A. Wang, C. Mavrogiannis, and A. Steinfeld, "Group-based Motion Prediction for Navigation in Crowded Environments."

[4] Y. F. Chen, M. Everett, M. Liu, and J. P. How, "Socially Aware Motion Planning with Deep Reinforcement Learning," May 2018, arXiv:1703.08862 [cs]. [Online]. Available: http://arxiv.org/abs/1703.08862

[5] A. J. Sathyamoorthy, U. Patel, T. Guan, and D. Manocha, "Frozone: Freezing-Free, Pedestrian-Friendly Navigation in Human Crowds," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4352–4359, Jul. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9099106/

[6] P. Trautman and A. Krause, "Unfreezing the robot: Navigation in dense, interacting crowds," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Taipei: IEEE, Oct. 2010, pp. 797–803. [Online]. Available: http://ieeexplore.ieee.org/document/5654369/

[7] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, p. 4282, 1995.

[8] J. van den Berg, J. Snape, S. J. Guy, and D. Manocha, "Reciprocal n-body collision avoidance," in *Robotics research*. Springer, 2011, pp. 3–19.

[9] F. Solera, S. Calderara, and R. Cucchiara, "Socially constrained structural learning for groups detection in crowd," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 5, pp. 995–1008, 2015.