Adaptive Human-Aware Local Planner with Reinforcement Learning

Mauro Martini[1], Noé Pérez-Higueras[2], Andrea Ostuni[1], Marcello Chiaberge[1], Fernando Caballero[2], Luis Merino[2]

***Abstract*— Human-aware navigation is a complex task for mobile robots. Social planners usually add additional costs or constraints to the objective function, leading to intricate tuning process. Machine Learning can provide planners with enhanced versatility and learning complex social behaviours from data. In this work, an adaptive social planner is proposed, using a Deep Reinforcement Learning agent to dynamically adjust the weighting parameters of the cost function used to evaluate trajectories. The resulting planner combines the robustness of a social version of DWA based on Social Force Model, with the flexibility of learning methods to boost the overall performance on social navigation tasks.**

I. INTRODUCTION

The recent advancements in service robotics have opened up new avenues for robotics research, particularly in the domain of human-aware navigation. Different simple social navigation scenarios have been partially categorized in literature to build benchmarks [1], [2], highlighting the set of unique challenges of each situation. However, standard social planners struggle in performing properly in all of them, considering the population of the space and the environmental geometry in cluttered narrow passages or wide open spaces. Therefore, the diversity and uncertainty of social scenarios necessitate a more flexible and adaptive approach.

Machine Learning (ML) techniques represent a potential solution to tackle this problem. Recent studies tried to address the challenges posed by human-aware navigation with Deep Reinforcement Learning (DRL) [3].

However, end-to-end learning approaches may often present less robust performance and poor generalization to unseen conditions than classical planners. On contrary, hybrid approaches, that take advantage of the benefits of classical planners and learn planner parameter policies, seems to be a promising approach for human-aware navigation.

This work proposes an adaptive hybrid approach inspired by the adaptive planner parameter-learning approach [4]. In contrast to it, we address the complex problem of human-aware navigation by teaching a DRL agent to dynamically adapt the cost weights of a human-aware local planner. A human-aware version of the popular Dynamic Window Approach (DWA) is used as social controller, adding a social cost term based on the robot-pedestrians interaction according to the Social Force Model (SFM) [5].

[1]Department of Electronics and Telecommunications, Politecnico di Torino, 10129, Torino, Italy. mauro.martini@polito.it, andrea.ostuni@polito.it, marcello.chiaberge@polito.it

[2]School of Engineering, Pablo de Olavide University, Crta. Utrera km 1, Seville, Spain noeperez@upo.es, fcaballero@upo.es, lmercab@upo.es

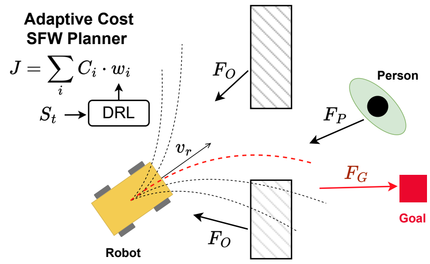

Fig. 1: The SFWP combines DWA and SFM. A DRL agent optimizes the weights of the trajectory scoring function.

II. METHODOLOGY

A. Social Force Window Planner (SFWP)

A human-aware local planner has been developed by adding a social cost to the trajectory scoring function of the classic Dynamic Window Approach (DWA) [6]. A "social work" quantity is adopted as social cost by using the well-known Social Force Model (SFM) [5], [7]. This social cost is computed as: $C_s = W_r + \sum_i W_{p,i}$, where W_r is the sum of the modulus of the robot social force (F_p) and the robot obstacle force (F_O) along the trajectory. W_p is the modulus of the social forces generated by the robot for each pedestrian i along the trajectory (see Fig. 1).

The overall trajectory scoring function is formulated as:

$$J = C_s \cdot w_s + C_o \cdot w_o + C_v \cdot w_v + C_d \cdot w_d + C_h \cdot w_h \quad (1)$$

where we have a single cost term for social navigation C_s, obstacles in the costmap C_o, robot velocity C_v, distance C_d and heading C_h from a local waypoint on a given global path. The costs are combined using weights w that regulate the impact of each term in the velocity command selection. The SFWP is publicly available in Github[1].

B. SFM Adaptive Cost Approach

In this work, a DRL agent learns a policy to dynamically set the weights of the cost function (1) that govern the control algorithm of the robot. The agent receives the local features of the surrounding environment, and induce the local planner to choose optimal velocity commands. The overall methodology represents a robust hybrid solution which efficiently integrate the flexibility of the DRL agent policy with the reliability of a classical navigation algorithm.

[1]https://github.com/robotics-upo/social_force_window_planner

A dedicated reward function has been designed to let the agent learn an optimal cost weights regulation policy. It is composed of the following terms:

- Goal alignment r_w. A path alignment reward term is defined to encourage the approach of the next waypoint on the global path.
- Velocity r_v. A velocity reward to promote faster motion
- Obstacle r_o. An obstacle reward is included to safely avoid collisions.
- Social penalty. Two different social rewards have been considered:
  - A proxemics-based reward r_p penalizes the robot when intruding in the intimate or personal space of a person.
  - The social work r_s as done in the cost of the SFWP.

Agent Policy We define the parameterized agent policy with a Deep Neural Network. To train we employ the Soft Actor-Critic (SAC) algorithm [8], which allows for a continuous action space and a fast convergence. In particular, we instantiate a stochastic Gaussian policy for the actor and two Q-networks for the critics.

State definition The state s_t has been designed as the ensemble of:

- Goal information with respect to the robot in polar coordinates.
- The set of cost weights of the SFWP predicted at the previous time step, $[w_d, w_h, w_v, w_o, w_s]_{t-1}$.
- The position and velocity of the four closest people in the robot frame. Position is expressed in polar coordinates whilst velocity with module and orientation.
- A reduced set of Lidar 2D points (closer than $3m$) for obstacle avoidance.

ANN and Output actions The neural network architecture of the agent is simply composed by two dense layers of 256 neurons each. The policy network predicts an action $a_t = [w_d, w_h, w_v, w_o, w_s]$ at each time step with a frequency of $2Hz$, which directly represents the new set of cost weights for the Social Force Window local planner.

III. EXPERIMENTS AND RESULTS

We employ Gazebo simulator to train and to test in diverse scenarios. The *HuNavSim* plugin [9] has been adopted to instantiate realistic moving people and to collect the evaluation metrics. *PIC4rl-gym* [10] is used as ROS 2 framework for DRL agent training. The training scenario contains a set of 30 diverse episodes, involving pedestrians passing, overtaking and crossing tasks in both narrow and open spaces. The agent is tested over another 10 different environments, always considering tasks of crossing, passing and overtaking. For the DWA baseline we use the same implementation of the SFWP, setting the social cost to zero value.

A. Preliminary Results

The adaptive social planner, SFW-SAC, is evaluated first with a single run experiment in each environment, comparing the results obtained with the baseline DWA and the SFWP with static costs. Standard navigation metrics such as clearance time $[s]$, path length (PL) $[m]$ and average linear velocity $v_{avg}[m/s]$ are employed to evaluate the effectiveness of the planner from a classic perspective. On the other hand, the social work (SW) metric is included in the quantitative results to show the social compliance of the navigation, measuring the social forces generated by the robot on the pedestrians and vice versa. The average metrics are presented in Table I. It can be noticed that the SFW-SAC method improves the overall trade-off between optimal navigation and human-awareness.

TABLE I: Average results over 10 scenarios obtained with DWA, SFWP, and SFW-SAC with adaptive cost.

Method	**Success** %	**Time** [s]	**PL** [m]	$\mathbf{v_{avg}}[\frac{m}{s}]$	**SW**
DWA	90.00	16.51	8.13	0.40	2.09
SFWP	100.00	19.70	8.38	0.37	3.14
SFW-SAC	100.00	16.80	8.38	0.40	1.77

IV. CONCLUSIONS AND FUTURE WORK

Promising initial results has been presented. We expect to further evaluate the approach with a richer set of metrics and the following future work:

- To enrich the experiments with different scenarios and test the system on the real robot, including a perception system to estimate the necessary state information of the closest people to the robot;
- To extend the method to other social controllers, including both controller parameters and costs.

REFERENCES

[1] Y. Gao and C.-M. Huang, "Evaluation of socially-aware robot navigation," *Frontiers in Robotics and AI*, vol. 8, p. 721317, 2022.

[2] C. Mavrogiannis, F. Baldini, A. Wang, D. Zhao, P. Trautman, A. Steinfeld, and J. Oh, "Core challenges of social robot navigation: A survey," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 3, pp. 1–39, 2023.

[3] Y. F. Chen, M. Everett, M. Liu, and J. P. How, "Socially aware motion planning with deep reinforcement learning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1343–1350.

[4] X. Xiao, Z. Wang, Z. Xu, B. Liu, G. Warnell, G. Dhamankar, A. Nair, and P. Stone, "Appl: Adaptive planner parameter learning," *Robotics and Autonomous Systems*, vol. 154, p. 104132, 2022.

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

[6] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.

[7] M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G. Theraulaz, "Experimental study of the behavioural mechanisms underlying self-organization in human crowds," *Proceedings of the Royal Society B: Biological Sciences*, vol. 276, no. 1668, pp. 2755–2762, 2009.

[8] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*. PMLR, 2018, pp. 1861–1870.

[9] N. Pérez-Higueras, R. Otero, F. Caballero, and L. Merino, "Hunavsim: A ros 2 human navigation simulator for benchmarking human-aware robot navigation," *IEEE Robotics and Automation Letters*, vol. 8, no. 11, pp. 7130–7137, 2023.

[10] M. Martini, A. Eirale, S. Cerrato, and M. Chiaberge, "Pic4rl-gym: a ros2 modular framework for robots autonomous navigation with deep reinforcement learning," *arXiv preprint arXiv:2211.10714*, 2022.