![Stanford University]

Robot Navigation in Crowds via Meta-Learning

Bilan Yang, Michelle Zhang

WN19 CS234 Final Project

{bilany, zhangmx@stanford.edu}

Highlight

- Study robot 2D navigation policies in pedestrian-rich environments towards a randomly assigned goal
- Propose the robot navigation policy network and optimize it via metal-learning framework
- Enable the robot with human-behavior capturing mechanism, and facilitate the awareness of unobservable intentions
- Demonstrate that the robot pursues efficient path while avoiding pedestrians with fast adaptation to a new environment

Related Work

Model-Agnostic Meta-Learning[1] (MAML)

- MAML optimizes model parameters to minimize the surrogate loss among all tasks
- It enables quickly adaptation to any random task in execution.

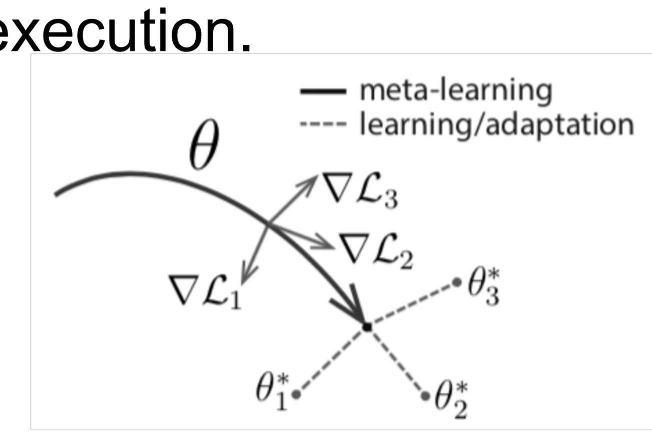

Fig. 1: MAML optimizes params. over all tasks

Social-Attention Reinforcement Learning[2] (SARL)

- SARL captures interactions among agents (pedestrians) occurring in dense crowds
- It can effectively learn to avoid pedestrians but is subject to a specific environment
- The robot learns through a value based network given state transition model

[1] Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks
[2] Chen, C., Liu, Y., Kreiss, S., and Alahi, A. Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning

Approaches

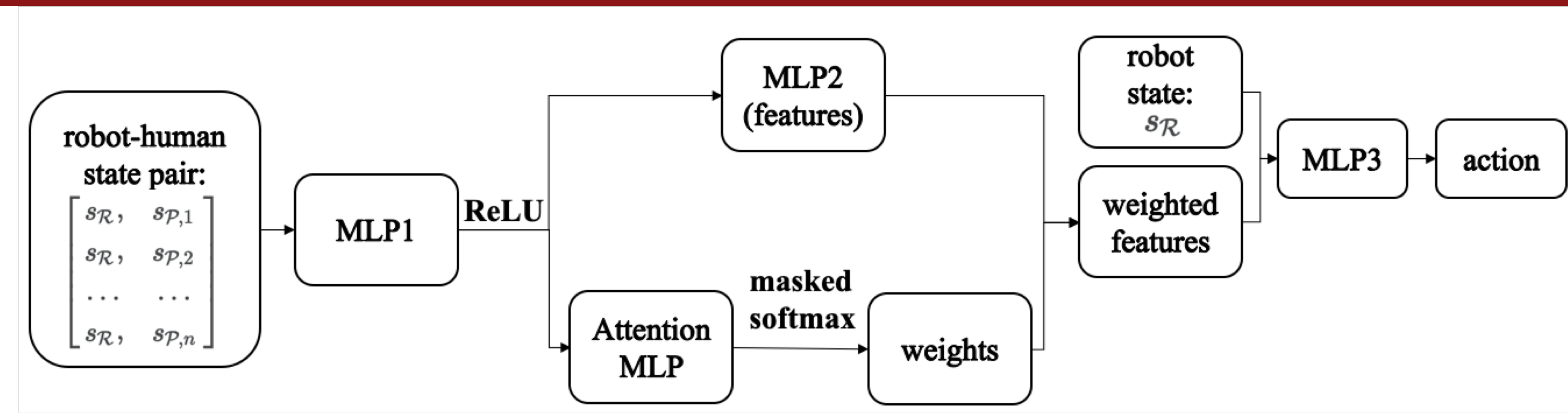

Fig. 2: Robot navigation policy network with robot-human interaction features

- **Tasks** are defined by uniformly random-sampled goal position, pedestrian speeds and moving directions
- **Observations** consist of the robot states and all pedestrian states:

$$s_{\mathcal{R}} = [p_x, p_y, v_x, v_y, g_x, g_y],$$
$$s_{\mathcal{P},i} = [p_{x,i}, p_{y,i}, v_{x,i}, v_{y,i}],\ i = 1, ..., n.$$

- **Action** is the robot's velocity: $a = [v_x, v_y] \in \mathcal{R}^2$
- **Reward** is the weighted sum of robot distance to the goal position r_d and robot-pedestrian collision indicator r_c:

$$r(s) = r_d(s) + r_c(s),$$
$$r_d(s) = -w_d \cdot \sqrt{(g_x - p_x)^2 + (g_y - p_y)^2},$$
$$r_c(s) = \sum_{i=1}^{n} r_{c,i}(d_i).$$

$$r_{c,i}(d_i) = \begin{cases} -1.5 & \text{if } d_i \leq r_{\text{crit}}, \\ -0.2(r_{\text{safe}} - d_i) & \text{if } r_{\text{crit}} < d_i \leq r_{\text{safe}}, \\ 0 & \text{if } d_i > r_{\text{safe}}. \end{cases}$$

Experiments and Training Results

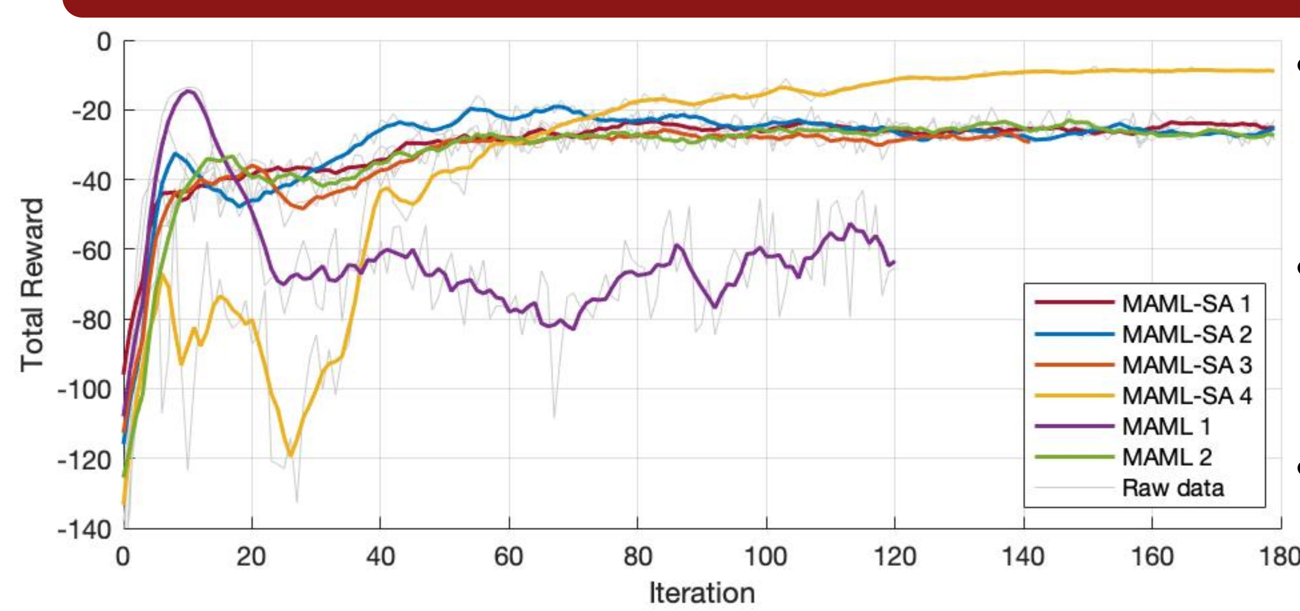

Fig. 3: Training results

- "MAML-SA 4" has the highest distance reward while having very low variance
- "MAML 1" has the worst distance reward and high variance
- Robot-centric frame boosts the efficiency of learning the joint behaviors in robot-human interactions using our proposed network

Model	Pedestrian number	Interaction feature number	Observable states	Coordinate system	Distance rewards	Collision rewards	Total rewards
MAML-SA 1	4	4	robot + pedestrians	global	-21.3	-1.30	-22.6
MAML-SA 2	8	4	robot + pedestrians	global	-21.4	-2.65	-24.0
MAML-SA 3	8	50	robot + pedestrians	global	-23.9	-2.81	-26.7
MAML-SA 4	8	50	robot + pedestrians	robot-centric	-7.41	-0.55	-7.95
MAML 1	8	N/A	robot	global	-53.9	-1.87	-55.3
MAML 2	8	N/A	robot + pedestrians	global	-22.6	-2.44	-25.1

Simulation Results

- MAML-SA policies enable fast adaption to a new environment usually within 3-step gradient updates
- We show policy improvement with one-step task-specific policy with 8 pedestrians ("MAML-SA 4")
- Robot also demonstrates pedestrian-avoiding capabilities while not deviate too much from its original path

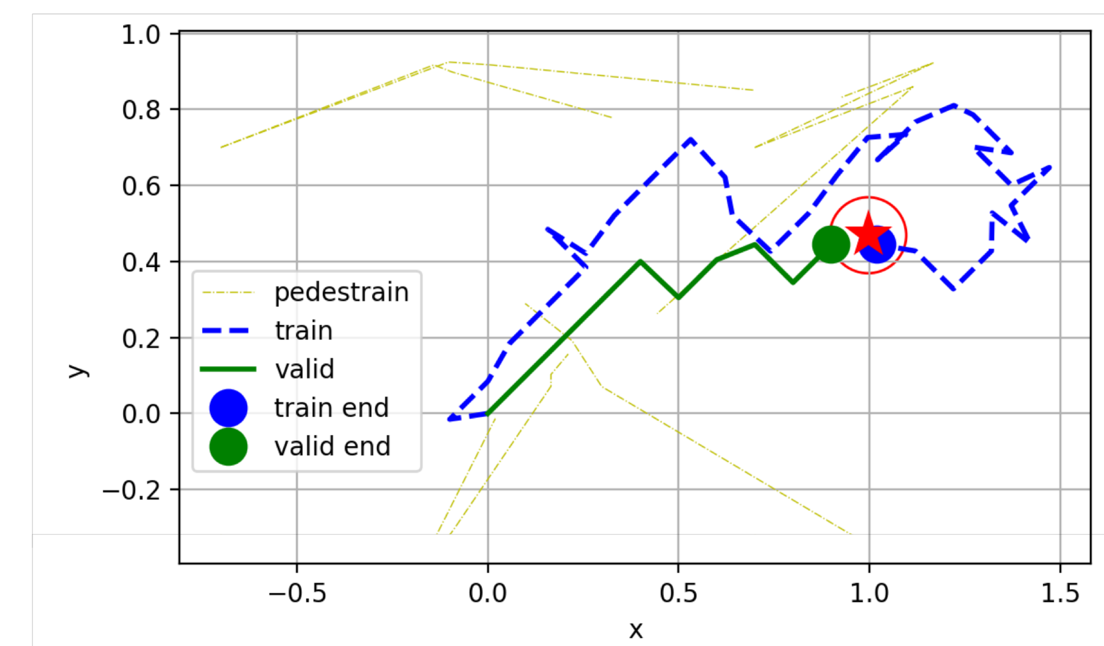

Fig. 4: Trajectory comparison between pre-update policy and one-step task-specific policy

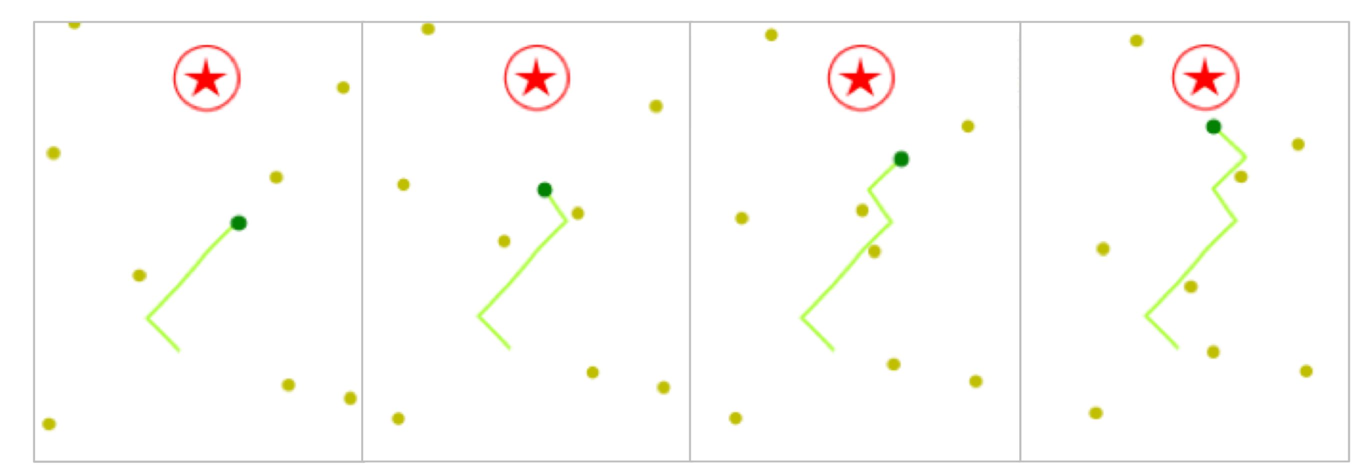

Fig. 5: Robot avoids pedestrians when navigating

Future Work

- Improve robustness by only considering neighboring pedestrians when we evaluate pedestrians' attention levels during training
- Incorporate human-human interactions into the policy network