Personalized Meta-Learning for Domain Agnostic Learning from Demonstration

Mariah L. Schrum
Georgia Institute of Technology
Atlanta, Georgia
mschrum3@gatech.edu

Erin Hedlund-Botti
Georgia Institute of Technology
Atlanta, Georgia
ehedlund6@gatech.edu

Matthew C. Gombolay
Georgia Institute of Technology
Atlanta, Georgia
matthew.gombolay@cc.gatech.edu

2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI) | 978-1-6654-0731-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/HRI53351.2022.9889377

***Abstract*—For robots to perform novel tasks in the real-world, they must be capable of learning from heterogeneous, non-expert human teachers across various domains. Yet, novice human teachers often provide suboptimal demonstrations, making it difficult for robots to successfully learn. Therefore, to effectively learn from humans, we must develop learning methods that can account for teacher suboptimality and can do so across various robotic platforms. To this end, we introduce Mutual Information Driven Meta-Learning from Demonstration (MIND MELD) [12, 13], a personalized meta-learning framework which meta-learns a mapping from suboptimal human feedback to feedback closer to optimal, conditioned on a learned personalized embedding. In a human subjects study, we demonstrate MIND MELD's ability to improve upon suboptimal demonstrations and learn meaningful, personalized embeddings. We then propose Domain Agnostic MIND MELD, which learns to transfer the personalized embedding learned in one domain to a novel domain, thereby allowing robots to learn from suboptimal humans across disparate platforms (e.g., self-driving car or in-home robot).**

***Index Terms*—meta-learning, personalized learning**

I. Introduction

Imagine a world in which robots are part of our everyday lives. Self-driving cars take us to and from work. A 7-DOF mobile arm empties the dishwasher and later cleans the home. Because many real-world tasks are novel and humans may differ in their preferences for how to accomplish a task, robots cannot be easily pre-programmed. Therefore, robots must learn such tasks from human demonstrators [3, 11]. Yet, because humans are unlikely to have expert knowledge about how best to demonstrate a new route to a self-driving car or show a robotic arm how to unload a new dishwasher, humans are likely to provide suboptimal demonstrations [5]. Consequently, there is a need for learning from demonstration (LfD) algorithms that can learn from suboptimal human demonstrators across robotic platforms and task domains [14].

To effectively learn from humans, prior work in LfD has investigated both human-centric (HC) and robot-centric (RC) approaches [5]. In HC LfD, the human performs a desired task and the robot learns from the demonstrated trajectory [2]. While intuitive to the demonstrator, HC LfD suffers from performance degradation due to covariate shift [6, 9]. Conversely, RC LfD requires the human to observe the robot's behavior and provide corrective feedback. While RC LfD solves the covariate shift problem and outperforms HC LfD when the feedback is high quality, prior work has shown that humans tend to provide low quality feedback to robots, making it difficult for robots to learn from humans in an RC paradigm [1, 5, 15]. Additionally, prior work has shown inter-demonstrator differences, meaning that human feedback is not only suboptimal but is also heterogeneous [7, 10, 16].

To overcome these limitations of RC learning, we introduce Mutual Information Driven Meta-Learning from Demonstration (MIND MELD) [12, 13], which meta-learns a personalized embedding describing an individual's feedback style and maps suboptimal feedback to better feedback, conditioned on this learned style. In a driving simulator domain, we demonstrate MIND MELD's ability to outperform prior work and show that MIND MELD has superior performance, likeability, intelligence, trust, and workload.

Teaching a car to drive to a goal is just one example of a domain in which a robot may learn from a human. When robots become ubiquitous in the real-world, humans will likely have to teach robots across different domains (e.g., in-home pick-and-place robot, self-driving car, etc.). Therefore, we must develop LfD techniques which are capable of effectively learning from humans across diverse platforms. To this end, we propose Domain Agnostic MIND MELD which learns a mapping of an individual's personalized embedding from one robotic domain to another. In future work, we propose to demonstrate Domain Agnostic MIND MELD's ability to learn an individual's personalized embedding in a pick-and-place task with a 7-DOF robotic arm and learn to map this embedding to a driving simulator domain, thereby improving upon RC LfD in both domains.

II. Approach

A. Calibration Tasks

In this section, we ground our approach in a driving simulator example in which the objective is to teach a car to drive to a goal location. To learn the personalized embedding describing an individual's suboptimal tendencies, or "style" of providing corrective feedback, we create a set of calibration tasks. These tasks are Wizard-of-Oz [8] rollouts representative of a policy learned via LfD and are drawn from the distribution of possible tasks that the demonstrator may encounter [4] within that

This work was supported by Georgia Tech State Funding, a NASA Early Career Fellowship (80HQTR19NOA01-19ECF-B1), MIT Lincoln Laboratory, Konica Minolta, and the National Science Foundation (1545287 and 20-604).

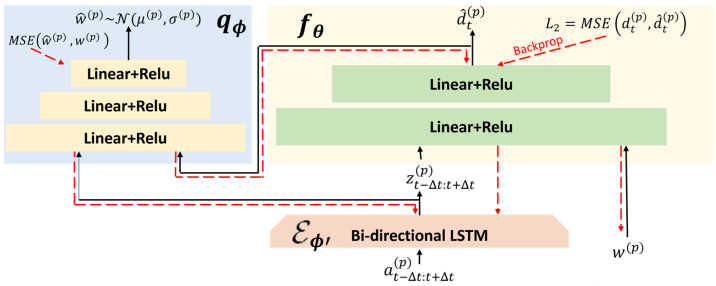

Figure 1: MIND MELD architecture

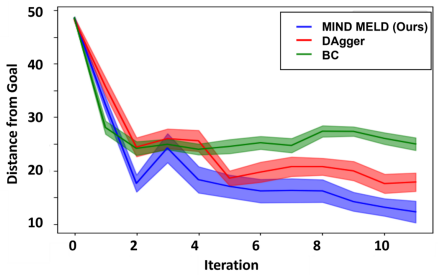

Figure 2: Average distance from goal

domain. In our case, the calibration tasks are representative of the car attempting to drive from point A to point B. We then calculate the optimal, corrective feedback (i.e., steering angle of the car) for each point in time along each calibration trajectory. Participants provide corrective feedback by turning the steering wheel in the direction they desire the car to go for each of the calibration tasks. The participant feedback and known ground truth actions are used to train our MIND MELD architecture and learn an individual's personalized embedding. For a similar, but different task, such as lane keeping, the calibration tasks and ground truths should be designed to capture this behavior. More information on the calibration tasks and architecture training can be found in [12, 13].

Fig. 1 shows our network architecture for learning the personalized embedding, $w^{(p)}$, and mapping from suboptimal corrective labels to improved labels. The personalized embedding, $w^{(p)}$, describes the way in which an individual is suboptimal (e.g., over- or under-correcting in a driving simulator). The subnetwork, $\mathcal{E}_{\phi'}$, maps a sequence of corrective feedback from the calibration tasks and outputs encoding, $z^{(p)}_{t-\Delta t:t+\Delta t}$. f_θ maps the encoding, conditioned on personalized embedding, $w^{(p)}$, to the difference between the ground truth label, o_t, and the demonstrators labels, $a_t^{(p)}$, at time, t, ($d_t^{(p)} = o_t - a_t^{(p)}$). q_ϕ learns a mapping from the difference, $d_t^{(p)}$, and encoding, $z^{(p)}_{t-\Delta t:t+\Delta t}$, to a posterior distribution over the embedding, $w^{(p)}$. Via variational inference, we maximize a lower bound on mutual information between $d_t^{(p)}$ and $w^{(p)}$, ensuring that $w^{(p)}$ can represent various and distinct feedback styles.

III. Human-Subjects Study

In prior work [12], we conduct a human-subjects study, demonstrating MIND MELD's ability to outperform RC and HC LfD. We compare MIND MELD to Dataset Aggregation (DAgger) [9] and Behavioral Cloning (BC) [2] in a within-subjects study in which participants were tasked with teaching a car to drive to a goal in a simulator. Fig. 2 shows that MIND MELD outperforms DAgger and BC in terms of average distance from goal. We find that MIND MELD is viewed more favorably in terms of likeability, intelligence, trust, and workload with $p < .01$ and that the learned embeddings significantly correlated with stylistic tendencies ($p < .001$) and video game experience ($p = .038$).

IV. Future Work

In the following section, we discuss our future work in which we propose to demonstrate MIND MELD's ability to generalize across platforms and tasks. We introduce Domain Agnostic MIND MELD which learns to transfer a personalized embedding learned in one domain to a novel domain.

We anticipate heterogeneous robotic platforms may produce different behavior and stylistic tendencies amongst participants. Therefore, in future work, we propose to investigate MIND MELD's ability to improve upon suboptimal demonstrations in a 7-DOF arm domain to show that MIND MELD can generalize to diverse robotic platforms. We hypothesize that humans will be less optimal in the arm domain compared to the car domain due to the complexity of control required.

Because humans must be able to effectively teach various robotic platforms to perform novel tasks, it is necessary to be able to transfer the learned knowledge about an individual's suboptimal tendencies from one domain to another. For example, an individual may need to teach a self-driving car a new route but also teach a dishwashing robot how to move dishes from cabinet to sink. In both domains, the suboptimality of the individual's demonstration must be taken into account to effectively learn from the human. We propose Domain Agnostic MIND MELD which learns to map the personalized embedding learned in one domain to an embedding describing the human's suboptimality in another domain.

To learn this mapping, we propose to collect human calibration data to learn an individual's personalized embedding, $w_d^{(p)}$, in the driving simulator domain. The same participants will also perform calibration tasks in the 7-DOF arm domain to learn their embedding, $w_a^{(p)}$. We will then learn a mapping of the personalized embeddings, $w_d^{(p)} = m_{a\to d}(w_a^{(p)}, \vec{c}^{(p)})$ from the arm domain to the car domain, conditioned on relevant covariates, $\vec{c}$ (e.g., experience with video games, experience teleoperating a robot). Next, we will conduct a human subjects study in which we demonstrate that the embedding learned in one domain can be transferred to a novel domain given demographic information and the mapping, m. This transferred embedding will be used to improve upon suboptimal labels in the new domain, thereby enhancing the robot's ability to learn from human feedback across disparate domains.

For robots to effectively learn from humans, there is a need for LfD algorithms which take into account human suboptimality and heterogeneity across disparate platforms and domains. Domain Agnostic MIND MELD will meet this need by learning transferable personalized embeddings to account for human suboptimality and heterogeneity in RC LfD.

REFERENCES

[1] Saleema Amershi, Maya Cakmak, W. Bradley Knox, and Todd Kulesza. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4):105–120, 2014.

[2] Brenna Argall, Sonia Chernova, Manuela M. Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, May 2009.

[3] Sonia Chernova and Manuela Veloso. Interactive policy learning through confidence-based autonomy. *Journal of Artificial Intelligence Research*, 34:1–25, 2009.

[4] Muhammad Abdullah Jamal and Guo-Jun Qi. Task agnostic meta-learning for few-shot learning.

[5] Michael Laskey, Caleb Chuck, Jonathan Lee, Jeffrey Mahler, Sanjay Krishnan, Kevin Jamieson, Anca Dragan, and Ken Goldberg. Comparing human-centric and robot-centric sampling for robot deep learning from demonstrations. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 358–365, 2017.

[6] Takayuki Osa, Gerhard Neumann, and Jan Peters. An Algorithmic Perspective on Imitation Learning. 7(1):1–179, 2018.

[7] Rohan Paleja and Matthew Gombolay. Inferring personalized bayesian embeddings for learning from heterogeneous demonstration. *arXiv*, 2019.

[8] Laurel D. Riek. Wizard of oz studies in hri: A systematic review and new reporting guidelines. *J. Hum.-Robot Interact.*, 1(1):119–136, July 2012.

[9] Stéphane Ross, Geoffrey J Gordon, and J. Andrew Bagnell. No-regret reductions for imitation learning and structured prediction. *Aistats*, 15:627–635, 2011.

[10] Claude Sammut. Automatically Constructing Control Systems by Observing Human Behaviour. *Second International Inductive Logic Programming Workshop*, (May), 1992.

[11] Stefan Schaal. Learning from demonstration. In M. C. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9. MIT Press, 1997.

[12] Mariah L. Schrum, Nina Moorman Erin Hedlund, and Matthew C. Gombolay. Mind meld: Personalized meta-learning for robot-centric imitation learning. *ACM/IEEE International Conference on Human-Robot Interaction*, 2022.

[13] Mariah L. Schrum, Erin Hedlund, and Matthew C. Gombolay. Improving robot-centric learning from demonstration via personalized embeddings. *CoRR*, abs/2110.03134, 2021.

[14] Felipe Leno Da Silva, Garrett Warnell, Anna Helena Reali Costa, and Peter Stone. Agents teaching agents: a survey on inter-agent transfer learning. *Autonomous Agents and Multi-Agent Systems*, 34, 4 2020.

[15] Jonathan Spencer, Sanjiban Choudhury, Matt Barnes, Matthew Schmittle, Mung Chiang, Peter Ramadge, and Siddhartha Srinivasa. Learning from Interventions: Human-robot interaction as both explicit and implicit feedback. 2020.

[16] Sebastian Weigelt, Vanessa Steurer, and Walter F. Tichy. At your command! an empirical study on how laypersons teach robots new functions. In *2020 IEEE 14th International Conference on Semantic Computing (ICSC)*, pages 468–470, 2020.