Robust Learning of Sequential Motion from Human Demonstrations to Enable Robot-guided Exercise Training

PI: Momotaz Begum, Co-PI: Dain LaRoche and Sajay Arthanat
PhD Students: Paul Gesel and Mostafa Hussein

Introduction

What does a robot need to learn from, and eventually instruct, humans how to perform structured exercises (e.g. therapeutic and recreational)? Four things: 1) a low-level trajectory learner, 2) a high-level policy learner, 3) a set of good metrics to evaluate human movements and 4) human-robot interaction capabilities. We are currently working on the first two.

Low-Level Trajectory Learning from Demonstrations

State-of-the-Art: Three methods are prominent in learning trajectory.

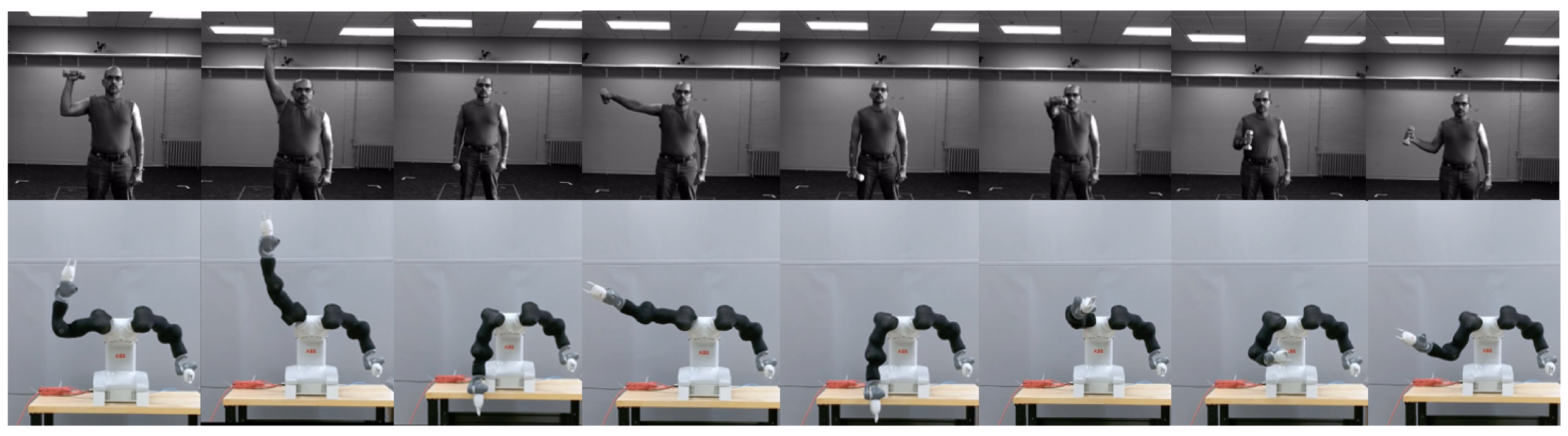
- DMP: generates motion by modulating stable second order dynamics with a non-linear forcing function. In case of large perturbations a phase re-indexing method is required.
- SEDS: dynamic system based mapping of states to action. It requires demonstrations throughout the state space to generate meaningful trajectories.
- GMM: creates a statistical representation of the motion from several demonstrations of the task. It is time dependant, can not avoid obstacles or change the goal configuration online.

Phase Space Model (PSM): A New Way of Trajectory Learning

- The PSM uses piece-wise Phase Space Transition Functions (PSTF)(equation 1) that drive an initial state to a desired phase space states (position, velocity) [1].

$$\ddot{x} = k\left(x + \frac{x_n^2 - x_c^2}{2\,(x_n - x_c)}\right) + \frac{\dot{x}_n^2 - \dot{x}_c^2}{2\,(x_n - x_c)} \qquad (1)$$

- Each PSTF contains a free parameter, which can be fit to the demonstration with regression.
- We applied the PSM to learn a variety of motions, including: feeding a human, rolling a cylinder, and several therapeutic exercises.
- We have formulated a quadratic program that optimizes velocity profile over a set of PSTFs according to an objective function [2].



A PSM-powered Yumi demonstrating therapeutic exercises learned from a therapist

- The PSM approach includes several advantages over existing low-level LfD approaches, namely time invariant dynamics, online obstacle avoidance and goal adaptation, and the ability to optimize the velocity profile as a quadratic program.

Broader Impacts

- LfD-powered robots with human-like anthropomorphism can be a potential solution to the lack of ubiquitous availability of motor rehabilitation services.
- Capstone project: Three undergraduate students are working on the vision-based control of YuMi for playing ping-pong as their Capstone project.

High-Level Policy Learning from Demonstrations

State-of-the-art: Reinforcement learning (RL), inverse-RL (IRL), and deep RL (DRL) are among the most successful contemporary methods for learning task-policies from human demonstrations.

- An ample amount of demonstrations, often hundreds, is needed to build the dynamics of the model. Which is unrealistic in robotics applications.
- RL-based methods require hand crafting the reward function.
- IRL-based methods are computationally expensive since they run an RL algorithm in an inner loop.

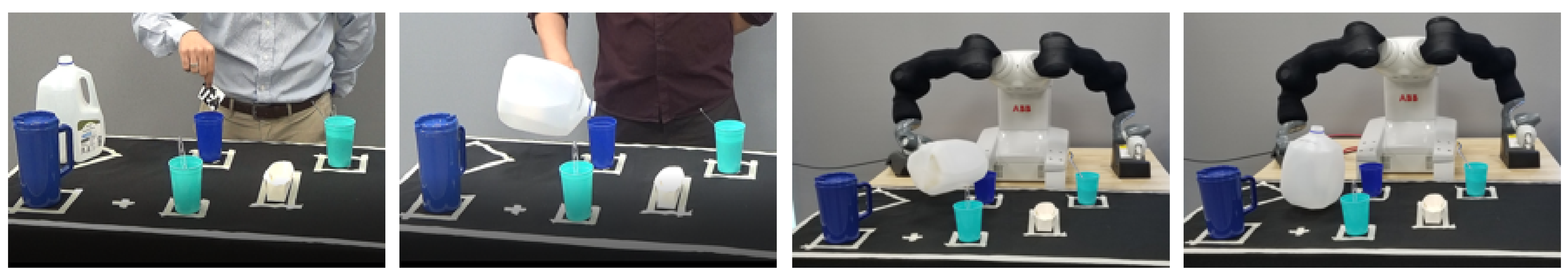
MAximum entropy Policy LEarning (MAPLE): A New Way of Learning High-level Policy MAPLE learns a policy through a deeper understanding of how different actions (during demonstrations) manipulate environmental features.

- MAPLE defines the task in terms of a set of easily identifiable feature constraints and does not require knowing the complete task dynamics from a large number of demonstrations.

$$\begin{aligned} \max_{p(a|s)} \quad & H(p) \equiv -\sum \tilde{p}(s)p(a \mid s)\log p(a \mid s) \\ \text{s.t.} \quad & \mathbb{E}_{\tilde{p}}[f_i] - \mathbb{E}_{p}[f_i] = 0 \quad 1 \le i \le n \\ & \sum_a p(a|s) - 1 = 0 \quad \forall\ s \end{aligned} \qquad (2)$$

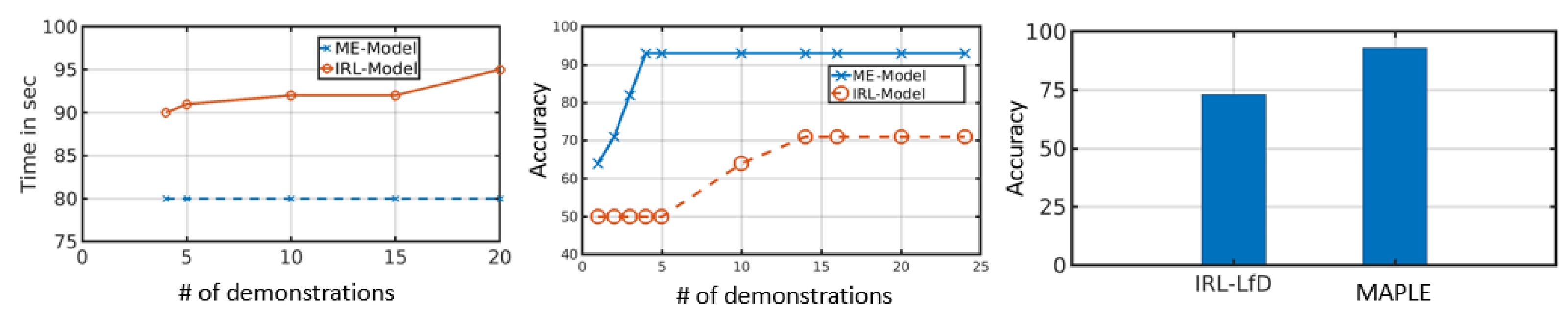
- It uses the principle of maximum entropy to derive the most 'relaxed' policy that only complies with the those feature constraints.
- Invoking features of various levels of specificity in policy learning makes MAPLE sample efficient.
- MAPLE does not require learning or hand-crafting a reward function and therefore is computationally efficient.

We employed MAPLE to learn a tea-making task from demonstrations [3]. MAPLE was able to produce a consistent policy after observing different ways of making tea by five demonstrators.



(a) (b) (c) (d)

(a-b) Demonstrations (c-d) Execution by Yumi



Performance of MAPLE

Reference

1) P. Gesel, M. Begum and D. LaRoche, "Learning Motion Trajectories from Phase Space Analysis of the Demonstration", IEEE ICRA 2019
2) P. Gesel et al., "Learning Optimized Human Motion via Phase Space Analysis", IEEE IROS 2020 (in preparation)
3) M. Hussein et al., "A Maximum Entropy Framework to Learn Policies of High-level Sequential Tasks", IEEE IROS 2020 (in preparation)

2020 National Robotics Initiative (NRI) Principal Investigators' Meeting **Award ID: 1830597**

FEBRUARY 27 - 28, 2020 | ARLINGTON, VIRGINIA