Robot Skill Learning from Demonstrations in Cluttered Environments

M. Asif Rana, Mustafa Mukadam, S. Reza Ahmadzadeh, Sonia Chernova, and Byron Boots

***Abstract*—In this paper, we present *importance weighted skill learning* from demonstrations. Unlike prior learning from demonstration approaches, which assume obstacle-free demonstration environments, our approach is capable of learning generalizable robot skills from demonstrations provided in a cluttered environment. To further allow skill refinement as more demonstrations are provided, we present an incremental weighted skill learning approach. Experimental validation on a robot is provided.**

I. Introduction

Learning from demonstration (LfD) [1] provides an effective way to enable robots to continuously expand their skillsets and hence function in unstructured and dynamic environments. In LfD, observing a set of human-provided demonstrations, the goal for the robot is to learn and reproduce a skill in novel scenarios. The reproduction scenarios may involve handling new start/goal state constraints, while avoiding any arbitrarily placed obstacles.

With the exception of a few [2, 3], all prior work in trajectory-based LfD [4, 5, 6, 7] has been based on the assumption that demonstrations can be performed in uncluttered, minimally constrained environments. This assumption, however, is unrealistic in real-world environments where restructuring the world to remove all clutter is often an impractical solution. In this work, we tackle the problem of learning skills from a set of demonstrations which can be partially or fully influenced by the presence of obstacles. Since obstacles in the demonstration environments can introduce additional constraints in human demonstrations that are unrelated to the target skill, this can often lead to suboptimal skill models.

To tackle the problem of learning from demonstration in cluttered environments, we present *importance weighted skill learning* that is able to extract the true skill constraints from influenced demonstrations. Alongside a batch learning method for importance weighted skill learning, we also provide an incremental learning method. The incremental learning approach allows to continuously improving the learned skill as demonstrations are accumulated from different environments. The aforementioned methods are further incorporated into our previous work [8] to enable generalizable skill reproduction. As a proof of concept, we have verified our approach on a robotic *reaching* skill.

II. Combined Learning from Demonstration and Motion Planning

We adopt the probabilistic inference view on generalizable skill reproduction called CLAMP [8]. In CLAMP, skill reproduction involves carrying out *maximum a posteriori* (MAP) inference to find the desired trajectory.

Trajectory Prior: A trajectory is defined as a finite collection of the robot states $\boldsymbol{x}(t_i) \in \mathbb{R}^d$ at time instances $\{t_0, t_1, \ldots, t_N\}$. The prior trajectory distribution is given by,

$$p(\boldsymbol{x}) \propto \exp\{-\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{\mu}\|^2_{\boldsymbol{\mathcal{K}}}\}. \tag{1}$$

where,

$$\boldsymbol{x} \doteq [\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^T$$

$$\boldsymbol{\mu} \doteq [\boldsymbol{\mu}(t_0), \boldsymbol{\mu}(t_1), \ldots, \boldsymbol{\mu}(t_N)]^T,\ \boldsymbol{\mathcal{K}} \doteq [\boldsymbol{\mathcal{K}}(t_i, t_j)]\big|_{ij,0\leq i,j\leq N}$$

The prior enforces *optimality* and is learned from demonstrations.

Event Likelihood: The likelihood function encodes the constraints, associated with the occurrence of random events $\mathbf{e}$, in the skill reproduction scenario. These random events may include obstacle avoidance, a new start/goal state or via-point. The likelihood function [9] is defined as,

$$p(\mathbf{e}|\boldsymbol{x}) \propto \exp\{-\frac{1}{2}\|\boldsymbol{h}(\boldsymbol{x};\mathbf{e})\|^2_{\boldsymbol{\Sigma}}\}, \tag{2}$$

where $\boldsymbol{h}(\boldsymbol{x};\mathbf{e})$ is a vector-valued cost function with covariance matrix $\boldsymbol{\Sigma}$. The likelihood enforces *feasibility* on a given trajectory during skill reproduction.

MAP Inference: The desired optimal and feasible trajectory that reproduces the skill is given by,

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmax}} \left\{p(\boldsymbol{x}|\mathbf{e})\right\} = \underset{\boldsymbol{x}}{\operatorname{argmax}} \left\{p(\boldsymbol{x})p(\mathbf{e}|\boldsymbol{x})\right\} \tag{3}$$

. Furthermore, we impose structure on the trajectory prior in (1). We assume that samples from this prior are governed by an underlying stochastic dynamical system. This assumption yields an exactly sparse precision (inverse covariance) matrix, suitable for efficient learning and inference. The problem of learning the trajectory prior is equivalent to estimating the underlying stochastic dynamics. For more details on the prior formulation, the reader is referred to Section 4.1 in [8].

III. Importance weighted skill learning

The aforementioned skill dynamics is governed by,

$$\boldsymbol{x}_{t+1} = \tilde{\boldsymbol{\Phi}}_{t+1}\tilde{\boldsymbol{x}}_t + \mathbf{w}_{t+1}, \qquad \mathbf{w}_{t+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{t+1}) \tag{4}$$

where,

$$\tilde{\boldsymbol{x}}_t = \begin{bmatrix} \mathbf{1} \\ \hline \boldsymbol{x}_t \end{bmatrix}, \qquad \tilde{\boldsymbol{\Phi}}_{t+1} = \left[\mathbf{u}_{t+1} \mid \boldsymbol{\Phi}_{t+1}\right]$$

$\boldsymbol{\Phi}_t$ and $\boldsymbol{u}_t$ are a time-varying transition matrix and a bias term, respectively, and $\mathbf{w}_t$ is additive white noise with time-varying covariance $\mathbf{Q}_t$.

Lets assume the availability of K trajectory demonstrations $\mathcal{T}^{\{1:K\}}$, with the k^{th} demonstration defined as $\boldsymbol{x}^k = [\boldsymbol{x}_0^k, \boldsymbol{x}_1^k, \dots, \boldsymbol{x}_N^k]^T$. Hence for each discrete time interval $(t, t+1]$, we have a set composed of K current states $\mathcal{X}_t = \{\boldsymbol{x}_t^1, \boldsymbol{x}_t^2, \dots, \boldsymbol{x}_t^K\}$ with a corresponding set of K next states $\mathcal{X}_{t+1} = \{\boldsymbol{x}_{t+1}^1, \boldsymbol{x}_{t+1}^2, \dots, \boldsymbol{x}_{t+1}^K\}$. Moreover, we also assume the availability of an importance weighting function $w : \mathbb{R}^d \mapsto \mathbb{R}$. The importance weighting function would give higher weights to the parts of demonstrations which are more likely to demonstrate the skill or the true intent of the human. We present two approaches to estimate the unknown parameters of the skill dynamics model in (4).

A. Batch Skill Learning

In the batch skill learning formulation, we seek to find an estimate of the unknown parameters $\tilde{\mathbf{\Phi}}_{t+1}$ and $\mathbf{Q}_{t+1}$ as follows,

$$\begin{aligned}\tilde{\mathbf{\Phi}}^*_{t+1}, &\boldsymbol{Q}^*_{t+1} \\ &= \underset{\tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1}}{\operatorname{argmin}} \left\{ \mathcal{L}(\tilde{\mathbf{\Phi}}_{t+1}, \mathbf{Q}_{t+1}) \right\} \\ &= \underset{\tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1}}{\operatorname{argmin}} \left\{ \operatorname{tr}\big(\boldsymbol{Q}_{t+1}^{-1} \mathbf{E}_{t+1} \mathbf{W}_t \mathbf{E}_{t+1}^T\big) + \lambda \|\tilde{\mathbf{\Phi}}_{t+1}\|_F^2 \right\}\end{aligned} \tag{5}$$

where $\mathbf{E}_{t+1} = \boldsymbol{X}_{t+1} - \tilde{\mathbf{\Phi}}_{t+1} \tilde{\boldsymbol{X}}_t$ defines the error matrix. The inputs are collected into a matrix $\tilde{\boldsymbol{X}}_t = [\tilde{\boldsymbol{x}}_t^1, \tilde{\boldsymbol{x}}_t^2, \dots, \tilde{\boldsymbol{x}}_t^K]$ while the corresponding targets into a matrix $\boldsymbol{X}_{t+1} = [\boldsymbol{x}_{t+1}^1, \boldsymbol{x}_{t+1}^2, \dots, \boldsymbol{x}_{t+1}^K]$. The matrix $\mathbf{W}_t = \operatorname{diag}\big(w(\boldsymbol{x}_t^1), w(\boldsymbol{x}_t^2), \dots, w(\boldsymbol{x}_t^k)\big)$ defines a state-dependent importance weight matrix. Furthermore, λ is the regularization coefficient.

The solution to the batch skill learning problem in (5) is found using linear ridge regression,

$$\tilde{\mathbf{\Phi}}^*_{t+1} = \boldsymbol{X}_{t+1}^T \mathbf{W}_t \tilde{\boldsymbol{X}}_t \big(\tilde{\boldsymbol{X}}_t \mathbf{W}_t \tilde{\boldsymbol{X}}_t^T + \lambda \mathbf{I}\big)^{-1} \tag{6}$$

$$\begin{aligned}\mathbf{Q}^*_{t+1} &= \frac{1}{z} \mathbf{E}_{t+1} \mathbf{W}_t \mathbf{E}_{t+1}^T \\ z &= \frac{\operatorname{tr}(\mathbf{W}_t)^2 - \operatorname{tr}(\mathbf{W}_t^T \mathbf{W}_t)}{\operatorname{tr}(\mathbf{W}_t)},\end{aligned} \tag{7}$$

B. Incremental Skill Learning

The batch skill learning procedure assumes that there are enough demonstrations available sufficient to learn an optimal skill model. However, as demonstration environment evolves and new demonstrations are aggregated, it is desirable to have the ability to update the skill model. To achieve this, we propose to carry out incremental weighted skill learning.

Similar to the skill reproduction method, our incremental skill learning method is also based on MAP inference. In this formulation, we associate an uninformed prior probability distribution over the unknown parameters in the skill dynamics equation (4). In the presence of k new demonstrations $\mathcal{T}^{1:k}$, the distribution is updated by conditioning the prior on the likelihood of observing these demonstrations. At any instance, the mode of this distribution provides an estimate of the unknown parameters.

Skill Dynamics Prior: The joint prior distribution over the unknown parameters $\tilde{\mathbf{\Phi}}_{t+1}$ and $\boldsymbol{Q}_{t+1}$ is given by

$$p(\tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1}) = p(\tilde{\mathbf{\Phi}}_{t+1} | \boldsymbol{Q}_{t+1}) p(\boldsymbol{Q}_{t+1}), \tag{8}$$

where,

$$p(\tilde{\mathbf{\Phi}}_{t+1} | \boldsymbol{Q}_{t+1}) = \mathcal{MN}(\boldsymbol{M}^0_{t+1}, \boldsymbol{Q}^0_{t+1}, \boldsymbol{R}^0_{t+1}), \tag{9}$$

$$p(\boldsymbol{Q}_{t+1}) = \mathcal{W}^{-1}(\boldsymbol{V}^0_{t+1}, \nu^0_{t+1}), \tag{10}$$

$\mathcal{MN}$ refers to a matrix normal distribution with matrix-valued mean $\boldsymbol{M}_{t+1}$ and covariances $\boldsymbol{Q}_{t+1}$ and $\boldsymbol{R}_{t+1}$ for the rows and columns respectively. Furthermore, $\mathcal{W}^{-1}$ refers to an inverse Wishart distribution with the parameter $\boldsymbol{V}_{t+1}$ being a positive definite scale matrix, and the parameter ν_{t+1} denoting the degrees of freedom.

Demonstration Likelihood: Given the availability of k new demonstrations, we group the input states, target states and weights in matrices form $\tilde{\boldsymbol{X}}_t$, $\boldsymbol{X}_{t+1}$, and $\mathbf{W}_t$ respectively as before. The likelihood of observing the data governed by the stochastic dynamics (4) is then given by

$$p(\boldsymbol{X}_{t+1} | \tilde{\boldsymbol{X}}_t, \tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1}) \propto \exp\left\{ -\frac{1}{2} \operatorname{tr}(\boldsymbol{Q}_{t+1}^{-1} \mathbf{E}_{t+1} \mathbf{W}_t \mathbf{E}_{t+1}^T) \right\} \tag{11}$$

Note that the likelihood is scaled by the weight matrix in order to incorporate the importance weighting.

Skill Dynamics MAP Inference: The optimal skill dynamics parameters after assimilation of the new demonstrations is given by the *maximum a posteriori* (MAP) estimate of the parameters

$$\tilde{\mathbf{\Phi}}^*_{t+1}, \boldsymbol{Q}^*_{t+1} = \underset{\tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1}}{\operatorname{argmax}} \left\{ p(\tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1} | \boldsymbol{X}_{t+1}, \tilde{\boldsymbol{X}}_t) \right\}, \tag{12}$$

where the posterior distribution is found by conditioning the prior distribution on the likelihood of demonstrations

$$\begin{aligned}p(\tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1} | \boldsymbol{X}_{t+1}, \tilde{\boldsymbol{X}}_t) \propto \\ p(\boldsymbol{X}_{t+1} | \tilde{\boldsymbol{X}}_t, \tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1}) p(\tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1})\end{aligned} \tag{13}$$

Due to the choice of the prior in (8) for the likelihood function in (11), it is possible to obtain an analytical solution for the MAP inference problem in (12). To achieve this, we decompose the posterior into the product of two distributions, similar to the prior distribution

$$\begin{aligned}p(\tilde{\mathbf{\Phi}}_{t+1}, \boldsymbol{Q}_{t+1} | \boldsymbol{X}_{t+1}, \tilde{\boldsymbol{X}}_t) = \\ p(\tilde{\mathbf{\Phi}}_{t+1} | \boldsymbol{X}_{t+1}, \tilde{\boldsymbol{X}}_t, \boldsymbol{Q}_{t+1}) p(\boldsymbol{Q}_{t+1} | \boldsymbol{X}_{t+1}, \tilde{\boldsymbol{X}}_t)\end{aligned} \tag{14}$$

where,

$$p(\tilde{\mathbf{\Phi}}_{t+1} | \boldsymbol{X}_{t+1}, \tilde{\boldsymbol{X}}_t, \boldsymbol{Q}_{t+1}) = \mathcal{MN}(\boldsymbol{M}^k_{t+1}, \boldsymbol{Q}_{t+1}, \boldsymbol{R}^k_{t+1}), \tag{15}$$

$$p(\boldsymbol{Q}_{t+1} | \boldsymbol{X}_{t+1}, \tilde{\boldsymbol{X}}_t) = \mathcal{W}^{-1}(\boldsymbol{V}^k_{t+1}, \nu^k_{t+1}), \tag{16}$$

The solution to the MAP estimation problem in (12) is equivalent to finding the mode of the two contributing distributions. According to the properties of matrix normal and inverse Wishart distributions, this results in

$$\tilde{\boldsymbol{\Phi}}^{*}_{t+1} = \boldsymbol{M}^{k}_{t+1} \tag{17}$$

$$\boldsymbol{Q}^{*}_{t+1} = \frac{1}{\nu^{k}_{t+1} + d + 1}\boldsymbol{V}^{k}_{t+1} \tag{18}$$

where the parameters of the posterior distribution are given by

$$\begin{aligned}
\boldsymbol{R}^{k}_{t+1} &= \tilde{\boldsymbol{X}}_t \boldsymbol{W}_t \tilde{\boldsymbol{X}}^{T}_{t} + \boldsymbol{R}^{0}_{t+1} \\
\boldsymbol{M}^{k}_{t+1} &= (\boldsymbol{X}_{t+1}\boldsymbol{W}_t\tilde{\boldsymbol{X}}^{T}_{t} + \boldsymbol{M}^{0}_{t+1}\boldsymbol{R}^{0}_{t+1})(\boldsymbol{R}^{k}_{t+1})^{-1} \\
\boldsymbol{V}^{k}_{t+1} &= (\boldsymbol{X}_{t+1} - \boldsymbol{M}^{k}_{t+1}\tilde{\boldsymbol{X}}_t)\boldsymbol{W}_t(\boldsymbol{X}_{t+1} - \boldsymbol{M}^{k}_{t+1}\tilde{\boldsymbol{X}}_t)^{T} \\
&\quad + (\boldsymbol{M}^{k}_{t+1} - \boldsymbol{M}^{0}_{t+1})\boldsymbol{R}^{0}_{t+1}(\boldsymbol{M}^{k}_{t+1} - \boldsymbol{M}^{0}_{t+1})^{T} + \boldsymbol{V}^{0}_{t+1} \\
\nu^{k}_{t+1} &= k + \nu^{0}_{t+1}
\end{aligned}$$

IV. Obstacle Based Importance Weighting Function

We seek to define the importance weighting function such that it gives lower importance to the parts of a demonstration which are more likely to be influenced by the presence of an obstacle and hence exhibit deviation from the true intent of the human.

We conjecture that the parts of demonstrations which are closer to obstacles are influenced by the obstacles and hence fail to satisfy the skill constraints well, assigning them low importance. For a given state $\boldsymbol{x}_t$, we define the importance weight as equivalent to the likelihood of staying collision-free [9]. For this likelihood function, we first define a hinge loss function

$$c(\boldsymbol{x}_t) = \begin{cases} -d(\boldsymbol{x}_t) + \epsilon & d(\boldsymbol{x}_t) \leq \epsilon \\ 0 & d(\boldsymbol{x}_t) > \epsilon \end{cases},$$

where $d(\cdot)$ is the signed distance from the closest obstacle in an environment and ϵ specifies the 'danger area' around the obstacle. With the mentioned hinge loss, we assume that an obstacle affects a state only when it is within the danger area around the obstacle. Outside this danger area, the obstacle has no influence on the state. The importance weight is given by a function in the exponential family,

$$w(\boldsymbol{x}_t) = \exp\left\{-\frac{c(\boldsymbol{x}_t)^2}{2\sigma^2_{obs}}\right\}, \tag{19}$$

where the parameter σ_{obs} dictates the rate of decay of the importance weight for states within the danger area. The smaller the value of σ_{obs}, the faster the importance weight will decay down to zero.

V. Experiments

We evaluate the performance of our method on a *reaching* skill. Multiple demonstrations via kinesthetic teaching on a JACO^2 7-DOF manipulator are provided. The recorded states $\boldsymbol{x}_t$ are composed of vector concatenation of the instantaneous end-effector positions and velocities. The demonstrations are also time-aligned using dynamic time warping (DTW) [10].

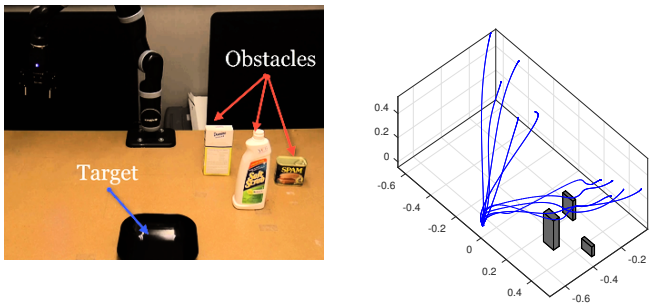

Fig. 1: Human demonstrations for a *reaching* skill. The right half of the demonstration environment cluttered with obstacles which influences some of the demonstrations.

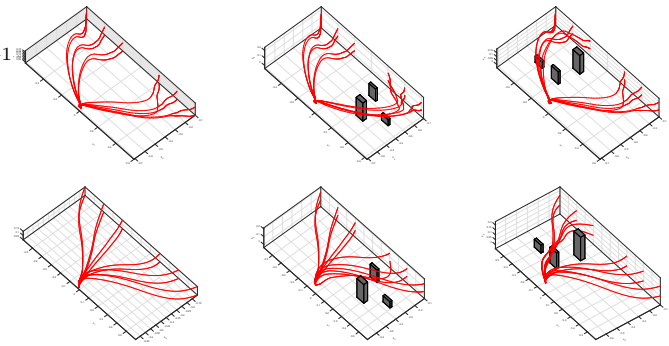

Fig. 2: Reproduced trajectories from different initial states for the *reaching* skill with prior learned **without** importance weighting (top row) and **with** importance weighting (bottom row). *Left:* Reproduction in an obstacle-free environment. *Middle:* Reproduction in the same environment as demonstration. *Right:* Reproduction in a different environment.

Since the goal is to reach an object from different locations, all the demonstrations share the same goal state while the initial state varies. As shown in Figure 1, the demonstrations follow a smoother path to the goal in the absence of obstacles on left-hand side of the state space. The demonstrations on the right however deviate from this behavior. We carry out the importance weighted skill learning procedure to remove these obstacle influence. For the importance weight function in (19), we empirically selected the parameters that worked in most of our experiments. A value of $\epsilon = 0.3m$ provided sufficient bounding region around the obstacles where the obstacle influence takes effect. Moreover, a value $\sigma_{obs} = 0.01m$ enabled nullifying the obstacle influence in most cases.

Figure 2 shows reproduction from a trajectory prior with and without considering the mentioned importance weighting procedure. We evaluated the skill reproduction in three different environments. When importance weighing is not considered, the reproduced trajectories exhibit deviation from the desired smooth path to the target even in the absence of obstacles in the vicinity. However, when importance weighting is considered, the obstacle influenced demonstrations reaching the target from right-hand side are given low importance. This enables extracting the true skill prior. The reproduced trajectories follow the desired smooth path, deviating only when obstacles are re-introduced in the vicinity.

REFERENCES

[1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.

[2] A. Rai, F. Meier, A. Ijspeert, and S. Schaal, "Learning coupling terms for obstacle avoidance," in *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on*. IEEE, 2014, pp. 512–518.

[3] A. M. G. E., C. Paxton, G. D. Hager, and L. Bascetta, "An incremental approach to learning generalizable robot tasks from human demonstration," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 5616–5621.

[4] S. Schaal, J. Peters, J. Nakanishi, and A. Ijspeert, "Control, planning, learning, and imitation with dynamic movement primitives," in *Workshop on bilateral paradigms on humans and humanoids, IEEE International Conference on Intelligent Robots and Systems*, 2003, pp. 1–21.

[5] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, no. 2, pp. 286–298, 2007.

[6] S. M. Khansari-Zadeh and A. Billard, "Learning stable nonlinear dynamical systems with Gaussian mixture models," *IEEE Transactions on Robotics*, vol. 27, no. 5, pp. 943–957, 2011.

[7] A. Paraschos, C. Daniel, J. R. Peters, and G. Neumann, "Probabilistic movement primitives," in *Advances in neural information processing systems*, 2013, pp. 2616–2624.

[8] M. A. Rana, M. Mukadam, S. R. Ahmadzadeh, S. Chernova, and B. Boots, "Towards robust skill generalization: Unifying learning from demonstration and motion planning," in *Conference on Robot Learning*, 2017, pp. 109–118.

[9] M. Mukadam, J. Dong, X. Yan, F. Dellaert, and B. Boots, "Continuous-time Gaussian process motion planning via probabilistic inference," *arXiv preprint arXiv:1707.07383*, 2017.

[10] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007.