An Active Learning Interface for Bootstrapping Robot's Generalization Abilities in Learning from Demonstration

E. Gribovskaya
elena.gribovskaya@epfl.ch

Florent d'Halluin
florent.dhalluin@epfl.ch

Aude Billard
aude.billard@epfl.ch

Learning Algorithms and Systems Laboratory
Ecole Polythechnique Federale de Lausanne, Switzerland

I. Introduction

We consider combining the *non-linear dynamical* systems' motion representation with the *active learning* paradigm as an approach to enhance motion learning and generalization abilities in robots. We suggest to design the learning process so that a robot can *gradually expand* its *generalization* capacities by requesting new demonstrations that extend a region of the applicability of the learned dynamics of a motion; see Figure 1. To allow for such a learning process, we extend our previous work on learning motion dynamics [1] with the iterative algorithm for estimating the region of applicability and the algorithm for estimating the consistency of new demonstrations with respect to the learned model.

In the previous work of ours [1], we propose an algorithm for estimating the *non-linear dynamics* of motion in the *state-space*. Our system encapsulates local correlation patterns between the motion's variables and provides the actual temporal robustness. The dynamics learned this way has a local character due to its non-linear form (local stability) and the limited generalization power of the statistical inference (farther from the demonstrated data the reliability of the inferences degrades.)[1] To be efficiently applied in practice, the learned dynamical model requires an estimate of its *region of applicability*, i.e. the estimate of the boundaries of the invariant sub-space where all trajectories converge to the target and where the confidence of the statistical inference allows generating the relevant trajectories. The size of this region determines the generalization abilities of the learned representation. Here we discuss how this region can be systematically expanded through additional demonstrations obtained with support of the suggested active learning interface.

Our approach to motion learning combines several demonstrations and, therefore, introduces a view on motion learning which is essentially different than the one, adopted in the other approaches to learning motion dynamics, where the authors consider learning from a single demonstration [2]. It has been acknowledged that combining several demonstrations is advantageous for motion learning [3]. Indeed, in our case, this allows a robot to generalize its knowledge from sub-optimal demonstrations and accurately reproduce the task's trajectories, starting from any point in the region of applicability.

However, while introducing the advantages in terms of improved generalization, learning from multiple demonstrations requires more efforts from a human teacher. In addition to repeating a motion several times, he/she should control the variability and consistency of the demonstrations. As an attempt to support the human during the teaching process, we present a generic interface for robot active learning that allows the robot learner to become cognitively involved in the process and support the teacher through querying for new demonstrations and by providing the feedback on those that have been acquired. We report on some preliminary experiments conducted with the i-Cub robot.

[1]This, however, should not be considered as a principal drawback as the non-linear motions demonstrated in the vicinity of manipulated objects, are not equally relevant far from these objects.

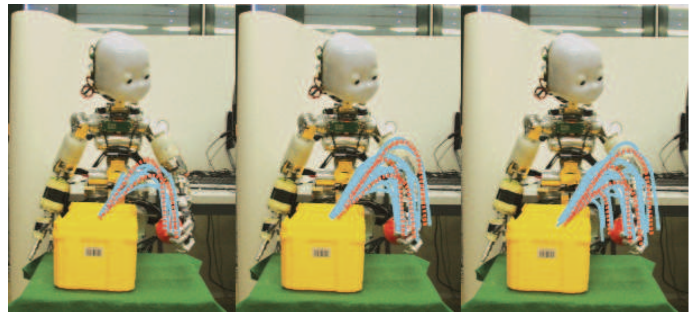

Fig. 1. Improvements in the generalization abilities of the i-Cub robot as a human teacher provides additional demonstrations (in dotted red) requested by the active learning system. The growth of the sub-space where the i-Cub can accurately accomplish the task and put the ball into the box while avoiding the box' boundaries, is represented by the increase in the volume of the trajectories' flow which the learned dynamical model can generate (reproduced trajectories are in blue). Note the expansion of the region of applicability after providing two, three, and four demonstrations.

II. Model Overview

The overview of the learning process is presented in Figure 2. The details of the approach used to learn motion dynamics can be found in [1]. In essence, learning consists in estimating a dynamical function $\dot{\xi} = \hat{f}(\xi)$ ($\xi \in \mathbb{R}^N$ is the state of the robot's end-effector) through the encoding of the training data as a joint probability distribution $\mathcal{P}(\xi, \dot{\xi})$ represented with the Gaussian Mixture Models and ensuring the stability of the estimate. We further briefly review the major components of the proposed active learning interface.

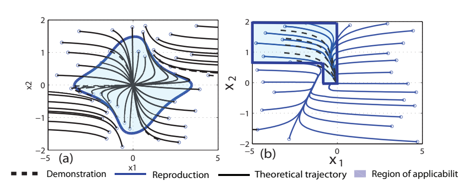

Fig. 3. The problem arising while applying a SoS-based approach to estimating the region of attraction (ROA) of a learned dynamics: the assumption of the polynomial form of the ROA's boundaries cannot be hold in general, as the learned dynamical vector flow is often asymmetric around the attractor. (a) An original *theoretical* dynamics is non-linear and locally stable. The ROA (thick blue) is symmetric and can be estimated as a level-set of the 7-th order polynomial. (b) The dynamics learned from the several samples (dashed lines), the reproduced vector flow (solid lines) is asymmetrical and the SoS estimate of the ROA is negligibly small. The estimate of the ROA generated with our incremental algorithm is highlighted in blue.

Incremental estimation of the region of applicability of a learned dynamical representation. We define the region of applicability

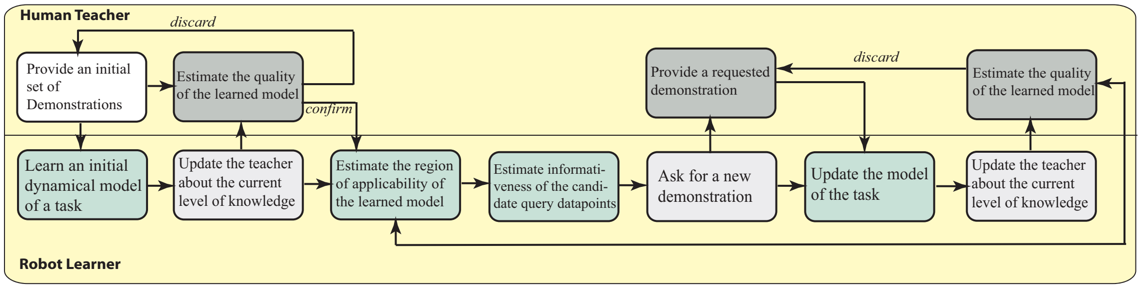

Fig. 2. Overview of the interactive learning process. In our framework learning is a bidirectional process, the direction of the information exchange between the human figure teacher and the robot learner is presented with arrows. Different colors of the blocks on the schema represent the different stages of active learning: the green blocks in the learner flow denote to the *acquisition* of new knowledge and *consolidation* of information; the light grey blocks in the same flow specify *behavioral responses* of the learner which aim at communicating a current level of understanding to the teacher. The dark grey blocks in the teacher flow relate to the *scaffolding process*, in which the teacher estimates the learned model and provides additional demonstrations to improve performance.

as the region of attraction (ROA) of a learned dynamics with an additional constraint that the likelihood of the points inside this region with respect to the learned model should exceed a given threshold.

However recently the Sum-of-Square optimization methods (SoS) have been intensively applied to estimating the ROA of non-linear systems [4], application of these methods to the motion dynamics learned from the human demonstrations is associated with some difficulties; see Figure 3. Here we develop an iterative computational approach to estimating the ROA; see Figure 4 for illustration and the brief explanation.

Assessing the consistency of a new demonstration with respect to the learned model. For efficient learning the demonstrated data should represent a consistent pattern, however, among the demonstrations provided by the human teacher outlier trajectories[2] are inevitable even if a task is demonstrated by an experienced user. To endow the robot with an ability to check the validity of a new demonstration and, if necessary, to discard it, we integrate an algorithm for estimating the *learner variance*[3] similar to [5]. The learner variance controls an error of the statistical inference, i.e. an average statistical error of an estimate $\hat{\dot{\xi}} = \hat{f}(\xi)$ in comparison with an actual unknown value $\dot{\xi} = f(\xi)$. Therefore, before adding a new demonstration into the training set, the robot verifies how this demonstration will affect the inference accuracy of the already existing model. If the estimated variance of the learned model increases above a given threshold (which means the decrease in the accuracy) the robot queries the teacher whether he/she prefers to keep the demonstration or it may be discarded.

III. Results and Conclusion

To validate the performance of our interface, we conducted two experiments: (1) the i-Cub should have learned to put the ball into the box; see Figure 1. (2) The ping-pong task, where the robot had to learn how to approach the ping-pong ball with the rocket in its hand. This experiment has been conducted with four users to verify the consistency of improvements in the generalization abilities provided by the proposed method; as a *measure* of the generalization abilities we chose the *volume of the region of applicability*. The preliminary results confirmed that the use of our interactive interface allowed to expand the robot's generalization abilities; and that the expansion has been consistent across the users.

In our work we address the problem of active involvement of a robot learner into the teacher-learner communication; which allows

[2]The trajectories exhibiting the shape different than other demonstrations in the dataset or following a distorted velocity profile.

[3]The expected average output variance of the learned model if the new demonstration would be added into the training set

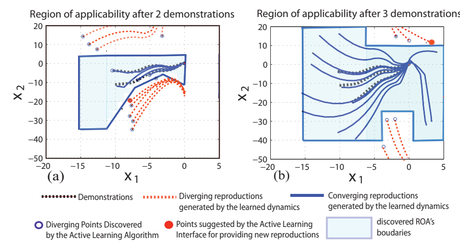

Fig. 4. An example of expanding the region of applicability of the learned figure dynamics with the support of the suggested active learning algorithm (the demonstrations are generated from a given theoretical dynamical system.) The initial model is learned from two demonstrations. The active learning system then starts suggesting the human teacher the new starting positions for demonstrations. The candidate points are those that lie outside of the current region of attraction and have the maximum likelihood with respect to the current learned model. The region of applicability gradually expands once a user adds demonstrations. Here we develop an iterative computational approach to estimating the ROA. The general idea consists in covering the volume containing demonstrated trajectories with a mesh and gradually expanding this mesh depending on whether a trajectory starting from a particular node converges to the origin. The convergence is verified by forward integration of the trajectories starting at the nodes of the mesh. The verification of stability on the boundaries of the region of attraction is sufficient to guarantee the stability inside. This property also allows to decrease the computational load: the complete forward integration of trajectories is only required to get the first estimate of the region of applicability; further, during the expansion, the algorithm only verifies the convergence of the trajectories towards the boundaries of the previous estimate of the ROA. This allows to apply the algorithm to real-time interactive learning.

improving the generalization abilities of the robot, that are of the important concern in the state-of-the-art Learning from Demonstration. The results obtained so far, though being promising, are still preliminary and require additional experiments with unexperienced users. We also planning to consider another alternatives of the constructive active learning to choose the regions to be queried for demonstrations.

Acknowledgment

This work was supported by the European Commission through the EU Projects FEELIX-GROWING (FP6-IST-045169) and ROBOT@CWE (FP6-034002).

References

[1] E. Gribovskaya and A. Billard, "Learning Nonlinear Dynamics of Motion for Position and Orientation Control of Robotic Manipulators," in *9th IEEE-RAS International Conference on Humanoid Robots*, 2009.

[2] P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal, "Learning and generalization of motor skills by learning from demonstration," in *Proceedings of International conference on Robotics and Automation*, 2009.

[3] A. Coates, P. Abbeel, and A. Ng, "Learning for control from multiple demonstrations," in *Proceedings of the International Conference on Machine learning*, 2008.

[4] R. Tedrake, I. Manchester, M. Tobenkin, and J. Roberts, "Lqr-trees: Feedback motion planning via sums of squares verification," *International Journal of Robotics Research*, 2010.

[5] G. Z. J. M. Cohn, D., "Active learning with statistical models," *Journal of Artificial Intelligence Research*, vol. 4, pp. 129–145, 1996.