

Learning from Failure

[Extended Abstract]

Daniel H Grollman

Learning Algorithms and Systems Laboratory
Ecole Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
daniel.grollman@epfl.ch

Aude G Billard

Learning Algorithms and Systems Laboratory
Ecole Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
aude.billard@epfl.ch

ABSTRACT

In the canonical Robot Learning from Demonstration scenario a robot observes performances of a task and then develops an autonomous controller. Current work acknowledges that humans may be suboptimal demonstrators and refines the controller for improved performance. However, there is still an assumption that the demonstrations are successful examples of the task. We here consider the possibility that the human has failed, and propose a model to minimize the possibility of the robot making the same mistakes.

Categories and Subject Descriptors: I.2.6 [Artificial Intelligence]: Learning; I.2.9 [Artificial Intelligence]: Robotics

General Terms: Theory

Keywords: Learning from Demonstration

1. INTRODUCTION

Robot Learning from Demonstration (RLfD) is potentially a means by which humans can instantiate robot controllers without performing analytical decomposition of the task itself or explicit coding. Ideally, an RLfD-enabled robot would be able to learn any performable task it observes demonstrated. Such robots would find a myriad of uses in both industrial and domestic settings.

Currently, RLfD typically proceeds by first collecting a set of examples, wherein a human user demonstrates acceptable task execution. From these examples a generalized controller is extracted by one of various methods [1]. Early work typically took the demonstrations as indicative of correct or *optimal* behavior [3]. More recently, research has focused on dealing with suboptimalities in the humans' demonstrations by, for instance, allowing for additional corrective demonstrations [2] or reinforcement learning [5].

Drawing inspiration from work showing that infants are able to successfully perform tasks that they have only seen failed examples of [6], we propose that the next logical step is to consider the possibility that the observed demonstrations are in fact failures. Rather than discarding such data as is now commonly done, we develop a model which uses variance in the demonstrations to guide exploration while avoiding replication of the demonstrations themselves. Particularly, where the demonstrations are consistent (low variance), we infer that the human is confident that this part of the task is being performed well and proceed as normal in

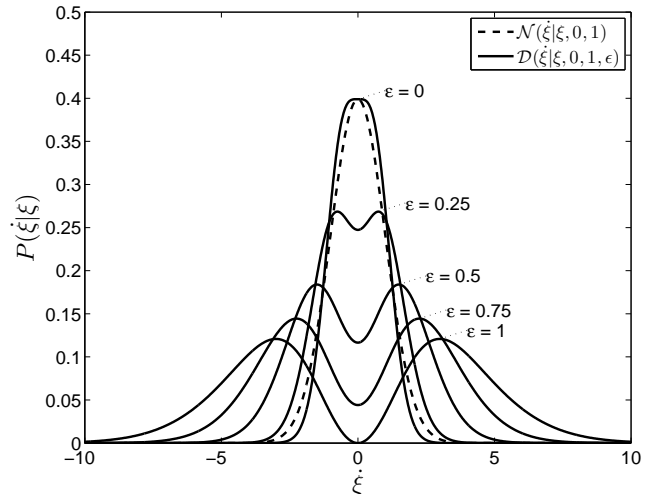


Figure 1: The donut pseudo-inverse has an exploration parameter to control the distance between the peaks.

RLfD. Inconsistent demonstrations (high variance) are then indicative of error (or non-importance for the task), and we actively explore possibilities other than those observed.

2. METHODOLOGY

Our method is an adaptation of Gaussian Mixture Model (GMM) based Dynamical Systems [4]. To N demonstrations (position and velocity over time: $\{\xi_t, \dot{\xi}_t\}^n$, $t \in [1, T_n]$, $n \in [1, N]$) we fit a GMM using the Bayesian Information Criterion to find the number of components (K) and Expectation Maximization to fit the priors (ρ^k), means (μ^k) and covariances (Σ^k), collectively termed θ . Standardly, the velocity at a position ξ is computed as the expected value of the conditional distribution over velocities at that position:

$$\dot{\xi}_{\text{std}} = E_{P(\dot{\xi}|\xi, \theta)}[\dot{\xi}] = \sum_{k=1}^K \rho^k \mu^k \quad (1)$$

$$P(\dot{\xi}|\xi, \theta) = \sum_{k=1}^K \rho^k(\xi, \theta) \mathcal{N}(\dot{\xi}; \tilde{\mu}^k(\xi, \theta), \tilde{\Sigma}^k(\theta)) \quad (2)$$

where the conditional's parameters are given by:

$$\tilde{\mu}^k(\xi, \theta) = \mu_{\xi}^k + \Sigma_{\xi\xi}^k \Sigma_{\xi\xi}^{k-1} (\xi - \mu_{\xi}^k) \quad (3)$$

$$\tilde{\Sigma}^k(\theta) = \Sigma_{\xi\xi}^k - \Sigma_{\xi\xi}^k \Sigma_{\xi\xi}^{k-1} \Sigma_{\xi\xi}^k \quad (4)$$

$$\tilde{\rho}^k(\xi, \theta) = \frac{\rho^k \mathcal{N}(\xi; \mu_{\xi}^k, \Sigma_{\xi\xi}^k)}{\sum_{k=1}^K \rho^k \mathcal{N}(\xi; \mu_{\xi}^k, \Sigma_{\xi\xi}^k)} \quad (5)$$

We instead replace each Gaussian in the GMM ($\mathcal{N}(\xi; \tilde{\mu}^k, \tilde{\Sigma}^k)$) with a variable-width pseudo inverse termed the Donut distribution which is itself a difference of two Gaussians:

$$\mathcal{D}(\xi; \tilde{\mu}, \tilde{\Sigma}, \epsilon) = \gamma \mathcal{N}(\xi; \tilde{\mu}, \frac{\tilde{\Sigma}}{r_{\alpha}^2}) - (\gamma - 1) \mathcal{N}(\xi; \tilde{\mu}, \frac{\tilde{\Sigma}}{r_{\beta}^2}) \quad (6)$$

whose means are the same as that of the base distribution, and whose covariances are defined by scalar ratios r_{α} and r_{β} . $\gamma > 1$ is an arbitrary constant.

When talking about the Donut distribution, we define the height (η) as the ratio between the Donut and base values at the mean and the width (λ) as the ratio between the donut’s peak-to-mean distance and the base’s standard deviation. To set r_{α} and r_{β} , we use ϵ to interpolate between a point where we most closely approximate the base ($\eta = 1, \lambda = 0$) and one where we are maximally different from it ($\eta = 0, \lambda = \lambda^*$) as seen in Figure 1. We use $\lambda^* = 6, \gamma = 2$.

We then predict a velocity at ξ by finding the most likely velocity in the resulting Donut Mixture Model (DMM):

$$\dot{\xi}_{\text{dnt}} = \operatorname{argmax}_{\dot{\xi}} \sum_{k=1}^K \tilde{\rho}^k \mathcal{D}(\xi; \tilde{\mu}^k, \tilde{\Sigma}^k, \epsilon) \quad (7)$$

$$\epsilon = 1 - \frac{1}{1 + \|V[\dot{\xi}|\xi, \theta]\|} \quad (8)$$

$$V[\dot{\xi}|\xi, \theta] = -\dot{\xi}_{\text{std}} \dot{\xi}_{\text{std}}^{\top} + \sum_{k=1}^K \tilde{\rho}^k (\tilde{\mu}^k \tilde{\mu}^{k\top} + \tilde{\Sigma}^k) \quad (9)$$

where ϵ is set by the overall variance of the conditional GMM, giving us the desired behavior described above. Since we learn from failure, generated trajectories that fail to perform the task are simply incorporated as more data.

3. EXPERIMENTS

We illustrate this approach on the task in Figure 2, which is to get a square foam block to stand on end by hitting a protruding edge from below. The setup is such that the block cannot be lifted to a standing position while in contact with the robot, instead the robot must impart momentum to the block. However, too much momentum and the block will topple over. This task is deceptively simple, in that human demonstrators often fail a few times before succeeding. Under a standard RLfD approach, these failures would be discarded, but we instead use them to learn the task.

We collected 2 failed demonstrations of this task, one where too little momentum is transferred, and another with too much. Using these demonstrations to initialize our model, we generate new trajectories and run them on the robot. If the task is still not performed, we incorporate the new trajectory into the model and repeat until it is consistently successful (5 consecutive trials). Over multiple restarts, this usually occurs in less than 10 iterations.

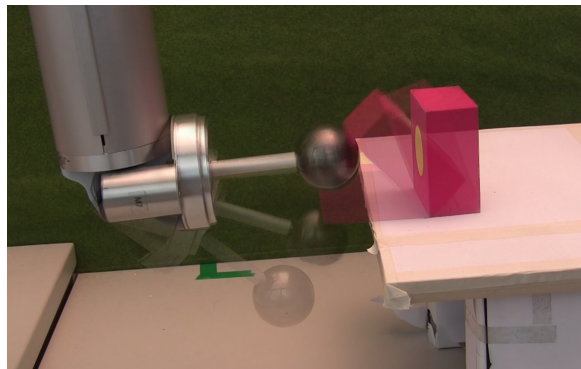


Figure 2: The FlipUp Task: get the foam block to stand on end. Shown is a successful performance learned from failure.

4. FUTURE WORK

Currently, there is no sense of direction guiding the search. In fact, we observe that the system will often alternate between trying a “too-fast” trajectory, a “too-slow” trajectory, and an “almost-there” trajectory several times before generating a successful trial. We believe that by incorporating a more informative reward signal (instead of a binary success/failure), we can guide the system better.

Overall, we hypothesize that a technique that uses failure cases such as this will lead to reduced overall training time in RLfD, as the system will be able to learn from data that is currently discarded. Also, it is possible that a system that learns alongside a human (i.e. the same task at the same time) may be able to highlight issues in the human’s learning process as well.

Acknowledgements

This work was supported in part by the European Commission under contract numbers FP7-248258 (First-MM) and FP7-ICT-248311 (Amarsi).

5. REFERENCES

- [1] Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469 – 483, May 2009.
- [2] Daniel H Grollman and Odest Chadwicke Jenkins. Dogged learning for robots. In *International Conference on Robotics and Automation*, pages 2483 – 2488, 2007.
- [3] Gillian Hayes and John Demiris. A robot controller using learning by imitation. In *International Symposium on Intelligent Robotic Systems*, 1994.
- [4] M. Hersch, F. Guenter, S. Calinon, and A. Billard. Dynamical system modulation for robot learning via kinesthetic demonstrations. *IEEE Transactions on Robotics*, pages 1463–1467, 2008.
- [5] Jens Kober and Jan Peters. Policy search for motor primitives in robotics. In *Neural Information Processing Systems*, 2008.
- [6] Andrew N. Meltzoff. Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31(5):838–850, 1995.