Learning multi-objective robot control policies from demonstration

Daniel H Grollman and Odest Chadwicke Jenkins
Department of Computer Science, Brown University
Providence, RI 02912-1910
{dang,cjenkins}@cs.brown.edu

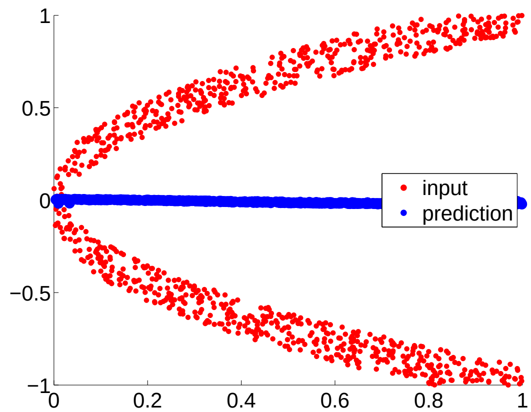

Figure 1: $\sqrt{x}$ learned with a functional regression algorithm (LWPR). The multiple outputs are averaged.

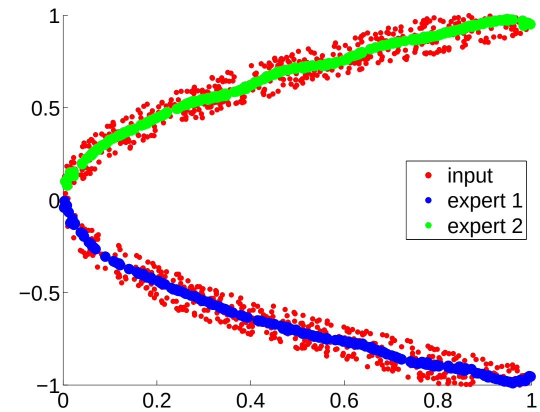

Figure 2: $\sqrt{x}$ learned with ROGER, a multimap learning algorithm. Learned experts shown by color.

When demonstrating unknown robot tasks via teleoperation, a human user may leverage information, latent in their mind, that is not observable to the robot. Such information may include user preferences as to how a task should be performed, state information observable to the human but not the robot, or task structure information such as subtask objectives. Multiple, different actions may thus occur in what the robot perceives to be the same state. The observed mapping from perceived states to actions, $\pi : \hat{s} \rightarrow a$, may then be a one-to-many *multimap*, instead of a one-to-one or many-to-one function.

Our interactive learning from demonstration architecture [1] enables robot learning from teleoperative demonstration via direct policy approximation (regression). Functional regression algorithms, however, are not appropriate for learning multimap policies (Fig. 1). Instead, we have developed ROGER (Realtime Overlapping Gaussian Expert Regression), a multimap regression algorithm for interactive learning from demonstration (Fig 2).

ROGER is based on the Infinite Mixture of Gaussian Processes model [2]. We achieve interactivity with a human user by reformulating it incrementally as a particle filter and using the Sparse Online Gaussian Process formulation [3]. Current work focuses on improving the algorithm's sparse and realtime properties and applying it to real robot tasks.

Whilst learning a multimap in this manner, the overall task is decomposed into a collection of overlapping, *functional*, experts, each corresponding to a subtask. Properly selecting an expert or subtask at run time is analogous to learning the topology of a finite-state machine describing the transitions between the subtasks.

We can annotate each datapoint in the human demonstration with the expert (or possible experts) that generated it. From the resulting expert trace, we can learn pre- and post-conditions for each expert, as in [4]. These conditions will be used when the robot is behaving autonomously to switch experts appropriately, and perform the task correctly.

Each expert is a single-objective functional mapping from states to actions. Thus, techniques such as inverse reinforcement learning [5] can be used to improve them. By estimating the underlying objective for each expert, the function for that expert can then be optimized, resulting in improved performance.

References

[1] Daniel H Grollman and Odest Chadwicke Jenkins. Dogged learning for robots. In *ICRA 2007*.

[2] Edward Meeds and Simon Osindero. An alternative infinite mixture of gaussian process experts. In *NIPS*, pages 883–890, 2006.

[3] Lehel Csató and Manfred Opper. Sparse online gaussian processes. *Neural Computation*, 14(3):641–669, 2002.

[4] Monica Nicolescu and Maja J Matarić. Natural methods for robot task learning: Instructive demonstration, generalization and practice. In *AAMAS*, pages 241–248, 2003.

[5] Pieter Abbeel and Andrew Y. Ng. Apprenticeship learning via inverse reinforcement learning. In *ICML 2004*.