Inverse Reinforcement Learning of a Dictionary of Primitive Tasks

Olivier Mangin and Pierre-Yves Oudeyer

Human behaviors feature a huge variety both in their nature and in their complexity, yet understanding and imitating them is a key requirement for robots to collaborate with humans or even to be accepted in their living environments. Inverse reinforcement learning (IRL) [6] consist in learning, from observation of a behavior, a reward function that models the agent's intention, that we hence denote as *task*. Previous work in IRL [5] have shown that instead of directly mimicking the observed behavior, learning a model of the task can provide better generalization to new contexts.

Often, one can think of common human behaviors as having a combinatorial structure in the sense that these complex behaviors can be understood as combinations of several simpler behaviors. For example, a child learning to play a new ball game won't learn from scratch how to grasp a ball, how to throw it, or how to run. While observing a demonstration from the new game, the child would not focus on every position of each body joint during the demonstration, but rather recognize a set of basic behaviors, e.g. running to a position while keeping the ball in hands and then throwing the ball to an other position.

In similar situations, a learning system would be expected to leverage the combinatorial structure of the demonstrated behaviors to improve its learning capabilities and eventually learn tasks of increasing complexity. The typical decription of IRL focuses on learning a single task from an expert, which, for non-trivial models of states and actions, already requires a huge number of demonstrations. Therefore one often uses adequate features to keep the number of required demonstrations feasible, which, however, requires prior knowledge on the task. In this work, we propose to learn a dictionary of features from several related but different tasks in order to reduce the amount of demonstrations required for each task and achieve transfer of knowledge between tasks.

Some recent work addresses aspects of these questions. Levine et al. [4] introduce an algorithm to learn features to efficiently represent a task; Jetchev and Toussaint [3] present an inverse feedback learning approach to a grasping problem and demonstrated that enforcing sparsity of the feedback function is a way to discover the features relevant to the task. Babes-Vroman et al. [2] present an EM-based algorithm that performs clustering on unlabeled demonstrations of several tasks. Almingol et al. [1] also recently developed similar ideas. In this work we introduce an algorithm, based on the gradient IRL algorithm of Neu et al [5], to learn a dictionary of primitive reward functions that can be combined together to model the intention of the expert in each demonstration. Not only does it learn from demonstrations of several tasks, as in [2], but it also enables transfer of knowledge from one task to another.

We performed two experiments on a toy problem on which we compare our factorial algorithm to learning separately the parameters that represent each task, denoted as *flat learning*. More precisely, we generate a random dictionary of features and n distinct composite tasks that are linear combinations of the dictionary elements; n experts then show demonstrations for each tasks, that are observed by both flat and factorial learners. While the flat learners independently learn a model of each task, the factorial learner reconstructs a dictionary, shared amongst tasks together with mixing coefficients. We evaluate the learners on each learned task by measuring their average performance on the MDP corresponding to the demonstrated task. We also compare the results with the ground truth dictionary and more naive strategies to build features.

The results demonstrate that our approach enables the learning of the shared structure of the tasks; they furthermore demonstrate and illustrate the fact that this approach enables more accurate representation of new additional tasks from only one short demonstration whereas the classical inverse reinforcement learning approach fails to generalize to unobserved parts of the space due to the lack of adequate features. The comparison of our algorithm with naive approaches also shows that these approaches also fail to learn the relevant structure of the demonstrated tasks.

References

[1] Javier Almingol, Luis Montesano, and Manuel Lopes. Learning Multiple Behaviors from Unlabeled Demonstrations in a Latent Controller Space. In *International conference on Machine learning*, 2013.

[2] Monica Babes-Vroman, Vukosi Marivate, Kaushik Subramanian, and Michael Littman. Apprenticeship learning about multiple intentions. In *International conference on Machine learning*, 2011.

[3] Nikolay Jetchev and Marc Toussaint. Task Space Retrieval Using Inverse Feedback Control. In *International Conference on Machine Learning*, pages 449–456, 2011.

[4] Sergey Levine, Zoran Popovic, and Vladlen Koltun. Feature construction for inverse reinforcement learning. *Advances in Neural Information Processing Systems*, pages 1–9, 2010.

[5] Gergely Neu and Csaba Szepesvári. Apprenticeship Learning using Inverse Reinforcement Learning and Gradient Methods. In *Conference on Uncertainty in Artificial Intelligence*, pages 295–302, Vancouver, Canada, 2007.

[6] Andrew Y. Ng and Stuart Russell. Algorithms for inverse reinforcement learning. *International Conference on Machine Learning*, 2000.