RIDM: Reinforced Inverse Dynamics Modeling for Learning from a Single Observed Demonstration in RoboCup 3D Simulation

Brahma Pavse, Faraz Torabi, Patrick MacAlpine, and Peter Stone

Department of Computer Science, The University of Texas at Austin

Imitation learning has long been an approach to alleviate the tractability issues that arise in reinforcement learning. However, most literature makes several assumptions such as access to the expert's actions, availability of many expert demonstrations, and injection of task-specific domain knowledge into the learning process. We propose reinforced inverse dynamics modeling (RIDM) [1], a method of combining reinforcement learning and imitation from observation (IfO) to perform imitation using a single expert demonstration, with no access to the expert's actions, and with little task-specific domain knowledge. Given only a single set of the expert's raw states, such as joint angles in a robot control task, at each time-step, we learn an inverse dynamics model to produce the necessary low-level actions, such as torques, to transition from one state to the next such that the reward from the environment is maximized. We have demonstrated that RIDM outperforms other techniques when we apply the same constraints on the other methods on six domains of the MuJoCo simulator and for two different robot soccer tasks for two experts from the RoboCup 3D simulation league on the SimSpark simulator.

RIDM works by collecting a single state-only demonstration from an expert, learning an inverse dynamics model to infer the actions needed to recreate the expert's demonstration state transitions, and then optimizing the parameters of the inverse dynamics model to maximize an environmental reward.

Algorithm 1 RIDM

Let $D^e = \{s_t^e\}$ be a single state-only demonstration of raw states per time-step
Let θ be the parameters of the inverse dynamics model
Randomly initialize θ
while not converged **do**
    Infer actions, $\{\widetilde{a_t^e}\}$, for $\{(s_t, s_{t+1}^e)\}$ using θ
    Execute $\{\widetilde{a_t^e}\}$
    Collect observed states $\{s_t\}$
    Collect cumulative episode reward R_{env}
    Update θ by optimizing θ for reward R_{env}
end while
return θ

The presentation will conclude with video demonstrations of fast walking and long kicking behaviors learned via RIDM from the demonstrations of other RoboCup 3D simulation teams.

1. B. S. Pavse, F. Torabi, J. Hanna, G. Warnell, and P. Stone. Ridm: Reinforced inverse dynamics modeling for learning from a single observed demonstration. In *Imitation, Intent, and Interaction (I3) Workshop at ICML 2019*, June 2019.