Autonomous Robotics: Defining Instincts and Learning Systems of Values

Michèle Sebag

University Paris-Sud, CNRS, France sebag@lri.fr

Abstract. Reinforcement learning aims at finding a good action policy, interacting with the environment in such a way that the agent (the robot) optimizes its cumulative reward along time. Where does the reward come from? In the robotics simulation context, the ground truth is available and the designer can use it to steer the robot learning toward the desired goals, through an appropriate reward function.

When reinforcement learning takes place on the robot, the ground truth is no longer available. A first issue then becomes to design an intrinsic reward function, or "instinct", providing the robot with internal incentives to act and explore its environment, visiting all states reachable on a given time budget. A second issue is to provide the robot with "values", indicating that not all reachable states are equal, and gradually steering the exploration toward the most promising behaviors.

Regarding the former issue, an intrinsic reward function will be discussed: viewing the robot as an information machine, a natural motivation thus is to maximize the quantity of information in its sensory datastream.

Regarding the latter issue, preference-based approaches can be used. A first possibility is to ask the designer to rank the behaviors demonstrated by the robot, thus enabling the robot to learn a policy return estimate. Iteratively, the robot builds a new and expectedly better controller during an active reinforcement learning phase. It thereafter demonstrates this controller to the designer, and uses the designer's feedback (it's better / it's worse) to update the policy return estimate.

Another possibility is to design ad hoc experiments to learn a preference-based value function directly on the state space. For instance, the designer can set the robot in a target position, and exploit the fact that the robot situation almost surely deteriorates along time when using naive controllers.

Keywords

reinforcement learning, preference learning, active learning, robotics