Knowledge Transfer in Markov Decision Processes

Caitlin Phillips
Supervised by: Prakash Panangaden, Joelle Pineau, and Doina Precup

August 18, 2006

Abstract

Markov Decision Processes (MDPs) are an effective way to formulate many problems in Machine Learning. However, learning the optimal policy for an MDP can be a time-consuming process, especially when nothing is known about the policy to begin with. An alternative approach is to find similar MDP, for which an optimal policy is known, and modify this policy as needed. We present a framework for measuring the quality of knowledge transfer when transferring policies from one MDP to another. Our formulation is based upon the use of MDP bisimulation metrics, which provide a stable quantitative notion of state similarity for MDPs. Given two MDPs and a state mapping from the first to the second, a policy defined on the latter naturally induces a policy on the former. We provide a bound on the value function of the induced policy, showing that if the two MDPs are behaviourally close in terms of bisimulation distance and the original policy is close to optimal then the induced policy is guaranteed to be close to optimal as well. We also present some experiments in which simple MDPs are used to test the tightness of the bound provided by the bisimulation distance. In light of the results of these experiments, we suggest a new similarity measure.