Kernel-Based Reinforcement Learning in Average-Cost Problems

D. Ormoneit and P. W. Glynn

- OrmoneitG02.pdf

Reinforcement learning (RL) is concerned with the identification of optimal controls in Markov decision processes (MDPs) where no explicit model of the transition probabilities is available. Many existing approaches to RL, including "temporal-difference learning", employ simulation-based approximations of the value function for this purpose. This procedure frequently leads to numerical instabilities of the resulting learning algorithm, especially if the function approximators used are parametric, such as linear combinations of basis functions or neural networks. In this paper, we propose an alternative class of RL algorithms which always produces stable estimates of the value function. In detail, we use "local averaging" methods to construct an approximate dynamic programming (ADP) algorithm. Nearest-neighbor regression, grid-based approximations, and trees can all be used as the basis of this approximation. We provide a thorough theoretical analysis of this approach and we demonstrate that ADP converges to a unique approximation in continuous-state average–cost MDPs. In addition, we prove that our method is consistent in the sense that an optimal approximate strategy is identified asymptotically. With regard to a practical implementation, we suggest a reduction of ADP to standard dynamic programming in an artificial finite-state MDP.