Abstract

One of the key problems in reinforcement learning is balancing exploration and exploitation. Another is learning and acting in large or even continuous Markov decision processes (MDPs), where compact function approximation has to be used. In this paper, we provide a practical solution to exploring large MDPs by integrating a powerful exploration technique, RMAX, into a state-of-the-art learning algorithm, least-squares policy iteration (LSPI). This approach combines the strengths of both methods, and has shown its effectiveness and superiority over LSPI with two other popular exploration rules in several benchmark problems.