Scaling Reinforcement Learning Paradigms for Motor Control

Jan Peters, Sethu Vijayakumar, Stefan Schaal
University of Southern California,
Computational Learning and Motor Control Laboratory
3461 Watt Way, Los Angeles, CA 90089

May 1, 2003

Abstract

Reinforcement learning offers a general framework to explain reward related learning in artificial and biological motor control. However, current reinforcement learning methods rarely scale to high dimensional movement systems and mainly operate in discrete, low dimensional domains like game-playing, artificial toy problems, etc. This drawback makes them unsuitable for application to human or bio-mimetic motor control. In this poster, we look at promising approaches that can potentially scale and suggest a novel formulation of the actor-critic algorithm which takes steps towards alleviating the current shortcomings. We argue that methods based on greedy policies are not likely to scale into high-dimensional domains as they are problematic when used with function approximation – a must when dealing with continuous domains. We adopt the path of direct policy gradient based policy improvements since they avoid the problems of unstabilizing dynamics encountered in traditional value iteration based updates. While regular policy gradient methods have demonstrated promising results in the domain of humanoid notor control, we demonstrate that these methods can be significantly improved using the natural policy gradient instead of the regular policy gradient. Based on this, it is proved that Kakade's 'average natural policy gradient' is indeed the true natural gradient. A general algorithm for estimating the natural gradient, the Natural Actor-Critic algorithm, is introduced. This algorithm converges with probability one to the nearest local minimum in Riemannian space of the cost function. The algorithm outperforms non-natural policy gradients by far in a cart-pole balancing evaluation, and offers a promising route for the development of reinforcement learning for truly high-dimensionally continuous state-action systems.

Keywords: Reinforcement learning, neurodynamic programming, actor-critic methods, policy gradient methods, natural policy gradient

References

[1] Kakade, S. A Natural Policy Gradient. Advances in Neural Information Processing Systems 14, 2002

[2] Amari, S. Natural Gradient Works Efficiently in Learning. Neural Computation 10, 1998

[3] Nedic, A. and Bertsekas, D. P., Least-Squares Policy Evaluation Algorithms with Linear Function Approximation, LIDS Report LIDS-P-2537, Dec. 2001; to appear in J. of Discrete Event Systems, 2002.

[4] Peters, J., Vijayakumar, S., Schaal, S. Reinforcement Learning for Humanoid Robotics, submitted to Humanoids, 2003.

[5] Sutton, R.S., McAllester, D., Singh, S., and Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. In Advances in Neural Information Processing Systems 12. MIT Press, 2000.

[6] Konda, V., and Tsitsiklis, J.N. Actor-Critic Algorithms. In Advances in Neural Information Processing Systems 12. MIT Press, 2000.