Rollout Sampling Approximate Policy Iteration*

Christos Dimitrakakis[1] and Michail G. Lagoudakis[2]

[1] Informatics Institute
University of Amsterdam
Amsterdam, The Netherlands
dimitrak@science.uva.nl
[2] Department of Electronic and Computer Engineering
Technical University of Crete
Chania 73100, Crete, Greece
lagoudakis@intelligence.tuc.gr

Several researchers [2,3] have recently investigated the connection between reinforcement learning and classification. Our work builds on [2], which suggests an approximate policy iteration algorithm for learning a good policy represented as a classifier, without explicit value function representation. At each iteration, a new policy is produced using training data obtained through rollouts of the previous policy on a simulator. These rollouts aim at identifying better action choices over a subset of states in order to form a set of data for training the classifier representing the improved policy. Even though [2,3] examine how to distribute training states over the state space, their major limitation remains the large amount of sampling employed at each training state.

We suggest methods to reduce the number of samples needed to obtain a high-quality training set. This is done by viewing the setting as akin to a bandit problem over the states from which rollouts are performed. Our contribution is two-fold: (a) we suitably adapt existing bandit techniques for rollout management, and (b) we suggest a more appropriate statistical test for identifying states with dominating actions early and with high confidence. Experiments on two classical domains (inverted pendulum, mountain car) demonstrate an improvement in sample complexity that substantially increases the applicability of rollout-based algorithms. In future work, we aim to obtain algorithms specifically tuned to this task with even lower sample complexity and to address the question of the choice of sampling distribution.

References

1. Dimitrakakis, C., Lagoudakis, M.: Rollout sampling approximate policy iteration. Machine Learning 72(3), 157–171 (September 2008)
2. Lagoudakis, M.G., Parr, R.: Reinforcement learning as classification: Leveraging modern classifiers. In: Proceedings of the 20th International Conference on Machine Learning (ICML), Washington, DC, USA, August 2003, pp. 424–431 (2003)
3. Fern, A., Yoon, S., Givan, R.: Approximate policy iteration with a policy language bias. Advances in Neural Information Processing Systems 16(3) (2004)

* This is an extended abstract of an article published in the Machine Learning journal [1]. This project was partially supported by grant MCIRG-CT-2006-044980.

W. Daelemans et al. (Eds.): ECML PKDD 2008, Part I, LNAI 5211, p. 7, 2008.