Articles for Seminary Paper

General Guidelines

You are required to prepare a written summary of two related research articles (primary and secondary) in the area of approximate dynamic programming (ADP) and Reinforcement Learning.

The primary article may be chosen from the list below, or from any other source, but should get my approval. It is generally required to be a full journal paper (not a conference paper), and to have a significant theoretical part.

In any case send me an e-mail with your choice for approval.

The secondary article should be related to the primary one (but not a conference version thereof), and chosen by you. For example, it may be chosen from papers that are cited in or cite the main paper.

Your seminary paper should contain two parts. Part one is a summary of the two articles, and should be up to 5 pages long. The summary should describe the basic problems, methods and ideas, focusing on the main issues. Part 2 should contain a critical assessment and comparison of the two articles. For example, you may point out the good and bad points, emphasize the main contribution and new ideas, put them in the context of what we learned in the course, and compare the two papers in terms of their inter-relation and contribution.

Paper List

(Shaded papers cannot be selected)

A. Antos, C. Szepesvari and R. Munos (2008). Learning near-optimal policies with Bellman-residual minimization based fitted policy iteration and a single sample path. *Machine Learning,* 71(1):89-129.

Bhatnagar, S., Sutton, R. S., Ghavamzadeh, M., and Lee, M. (2009). Natural actor-critic algorithms. *Automatica*, 45:2471-2482, 2009.

Borkar, V. and Meyn, S.P., The O.D.E. method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal of Control and Optimization*, 38(2):447-469.

Brafman, R. and Tennenholtz, M., R-Max – a general polynomial time algorithm for near-optimal reinforcement learning. Jounral of Machine Learning Research 3, 213-231, 2002.

T. Dietterich (2000). Hierarchical Rinforcement Learning with the MAXQ value function decomposition. *Journal of Machine Learning Research* 13, pp. 227-303. []

D. Ernst, P. Geurts and L. Wehenkel (2006). Tree-Based Batch Mode Reinforcement Learning". *Journal of Machine Learning Research,* Vol. 6, pp. 503-556.

D. P. de Farias, and B. Van Roy (2003). The linear programming approach to approximate dynamic programming. *Operations Research*, 51(6):850-865.

A. P. George and W.B. Powell, (2006). Adaptive stepsizes for recursive estimation with applications in approximate dynamic programming. *Machine Learning,* 65:167-198.

Guestrin, C. E., Koller, D., Parr, R., and Venkataraman, S. (2003). Efficient Solution Algorithms for Factored MDPs. *J. of Artificial Intelligence Research,* Vol. 19, pp. 399-468.

Kearns, M. and Singh, S. P. (1998). Near-optimal Reinforcement Learning in polynomial time. *Machine Learning*, 49, 209-232, 2002.

M. Kearns, Y Mansour and A. Ng (2002). A Sparse Sampling Algorithm for Near-Optimal Planning in Large Markov Decision Processes. *Machine Learning,* pp. 193-208 []

V. R. Konda, and J. N. Tsitsiklis (2003). On actor-critic algorithms. *SIAM J. Control and Optimization*, 42(4):1143-1166.

Lagoudakis, M. and Parr, R. (2003). Least-squares policy iteration. *Journal of Machine Learning Research*, 4:1107-1149.

M. G. Lagoudakis and R. Parr (2003). Least-squares policy iteration. *Journal of Machine Learning Research*, vol. 4.

Munos, R., and Szepesvari, C. (2008). Finite-Time Bounds for Fitted Value Iteration. *Journal of Machine Learning Research*, Vol. 1, pp. 815-857.

Rust, J. (1996). Using randomization to break the curse of dimensionality. Econometrica, 65:487{516.

Sutton, R. S., Precup, D., and Singh, S. P. (1999b). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. Artificial Intelligence, 112:181{211.

Tsitsiklis, J. N. and Van Roy, B. (1997). An analysis of temporal difference learning with function approximation. IEEE Transactions on Automatic Control, 42:674{690.

J. N. Tsitsiklis and S. Mannor (2004). The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research,* 5:623-648. []

B. Van Roy (2006). Performance loss bounds for approximate value iteration with state aggregation. Mathematics of Operations Research, 31(2):234{244.

Xu, X., He, H., and Hu, D. (2002). Efficient reinforcement learning using recursive least squares methods. Journal of Artificial Intelligence Research, 16:259-292.

Added 21/6/11

T. Jaksch, R. Ortner and P. Auer (2010). Near-optimal Regret Bounds for Reinforcement Learning, *Journal of Machine Learning Research* 11, pp. 1563-1600. []

L. Li, M. Littman, T. Walsch and A. Strehl (2011). Knows what it knows: a framework for self-aware learning, *Machine Learning*, 82(3):399-443. []