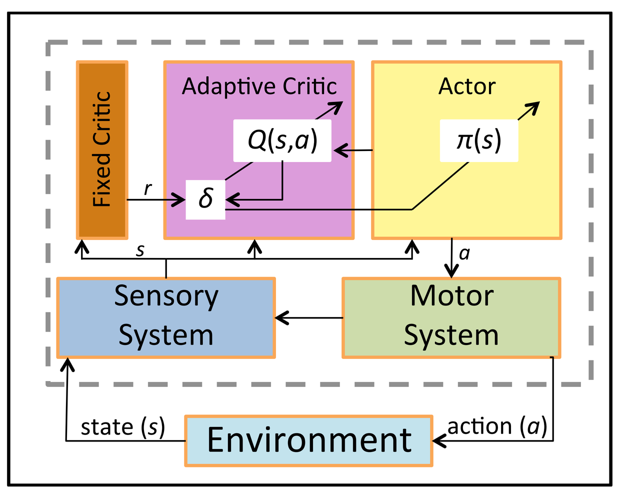
Sparse Coding of Learned State Representations in Reinforcement Learning

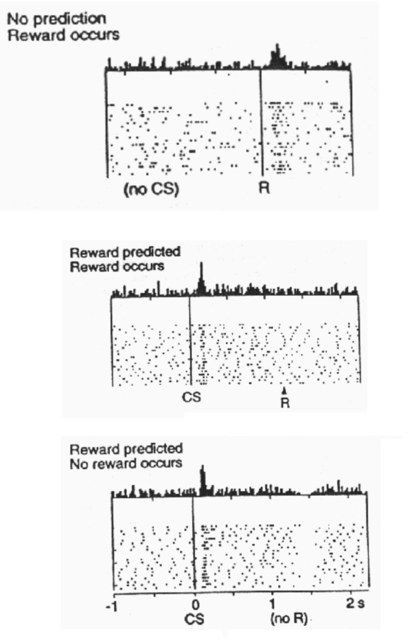
Jacob Rafati & David C. Noelle
Computational Cognitive Neuroscience Laboratory
University of California, Merced
{jrafatiheravi, dnoelle}@ucmerced.edu

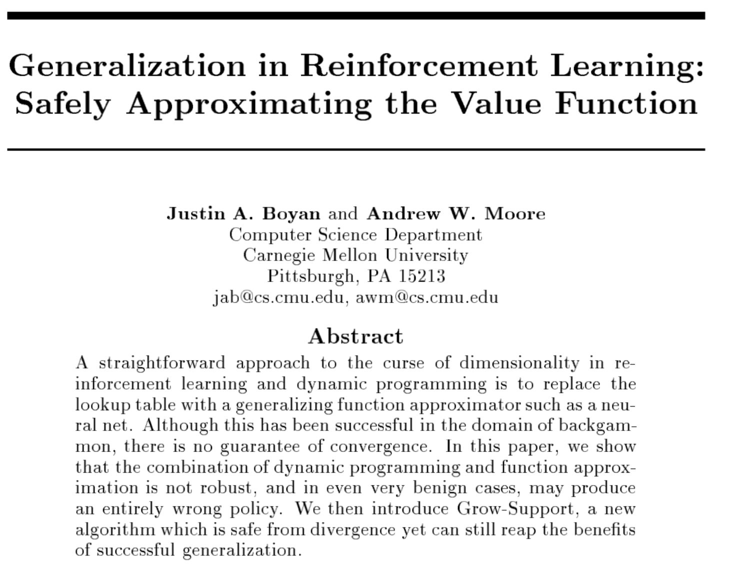
Convergence of Reinforcement Learning

Temporal Difference (TD) Learning is a leading account of the role of the dopamine system in reinforcement learning. TD Learning has been shown to fail to learn some fairly simple control tasks, however, challenging this explanation of reward-based learning. We conjecture that such failures do not arise in the brain because of the ubiquitous presence of lateral inhibition in the cortex, producing sparse distributed internal representations that support the learning of expected future reward. We provide support for this position by demonstrating the benefits of learned sparse representations for two problematic control tasks: mountain car and acrobat.



$$\delta \;=\; r \;+\; \gamma\,\hat{Q}(s', a') \;-\; \hat{Q}(s, a)$$





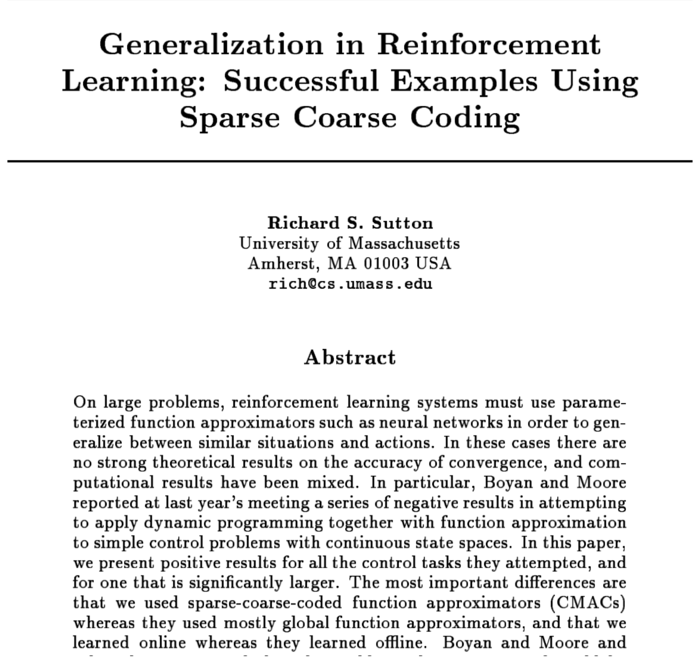
Generalization in Reinforcement Learning: Safely Approximating the Value Function

Justin A. Boyan and Andrew W. Moore
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213
jab@cs.cmu.edu, awm@cs.cmu.edu

Abstract

A straightforward approach to the curse of dimensionality in reinforcement learning and dynamic programming is to replace the lookup table with a generalizing function approximator such as a neural net. Although this has been successful in the domain of backgammon, there is no guarantee of convergence. In this paper, we show that the combination of dynamic programming and function approximation is not robust, and in even very benign cases, may produce an entirely wrong policy. We then introduce Grow-Support, a new algorithm which is safe from divergence yet can still reap the benefits of successful generalization.

Sparse Coded State Representations

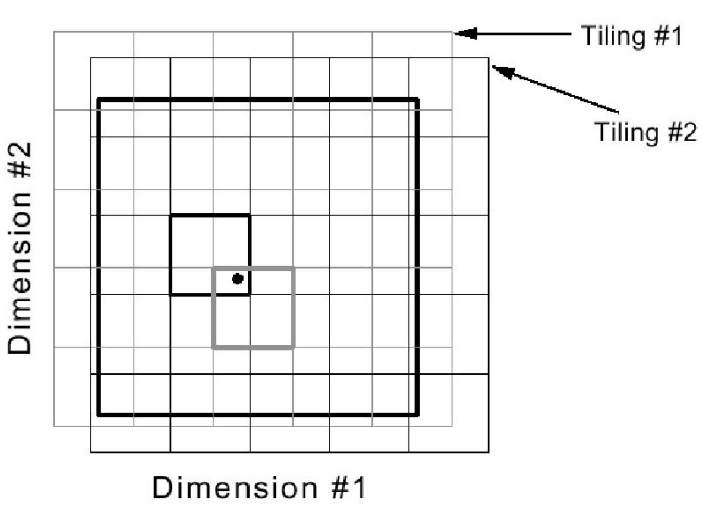


Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding

Richard S. Sutton
University of Massachusetts
Amherst, MA 01003 USA
rich@cs.umass.edu

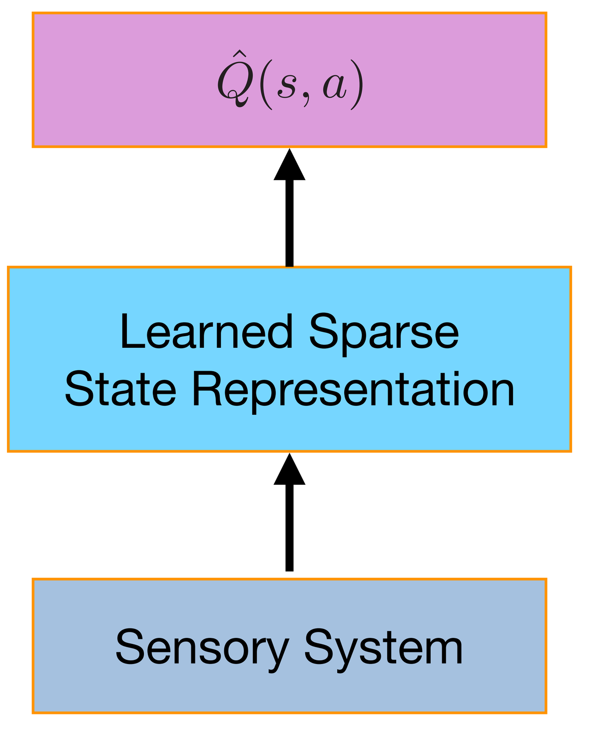
Abstract

On large problems, reinforcement learning systems must use parameterized function approximators such as neural networks in order to generalize between similar situations and actions. In these cases there are no strong theoretical results on the accuracy of convergence, and computational results have been mixed. In particular, Boyan and Moore reported at last year's meeting a series of negative results in attempting to apply dynamic programming together with function approximation to simple control problems with continuous state spaces. In this paper, we present positive results for all the control tasks they attempted, and for one that is significantly larger. The most important differences are that we used sparse-coarse-coded function approximators (CMACs) whereas they used mostly global function approximators, and that we learned online whereas they learned offline. Boyan and Moore and



CMAC

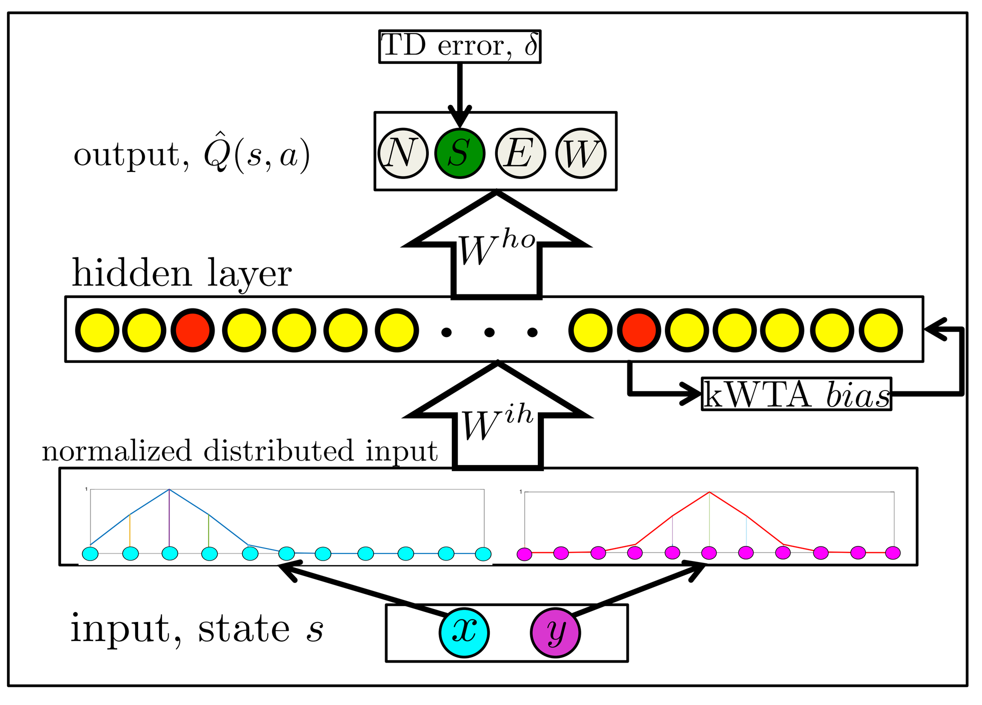
Sutton (1996) provided evidence that the sparse coarse coding of state representations can address the problem of failure to converge during reinforcement learning, when using a value function approximator, as reported by Boyan & Moore (1995). The proposed approach required the engineering of appropriate sparse coarse codes for each learning problem, however. A more general learning mechanism could be had if sparse coded internal state representations were learned, rather than engineered.



Lateral Inhibition

Learning in a value function approximator could be regularized to encourage the development of sparse coarse codes. An alternative is to introduce a homeostatic dynamic that rapidly adapts state representations to a fixed level of sparsity.

Fast pooled lateral inhibition is ubiquitous in mammalian cortex. There is evidence that lateral inhibition can introduce a soft k-winners-take-all (kWTA) attractor dynamic (O'Reilly, 2001) that produces sparsity.

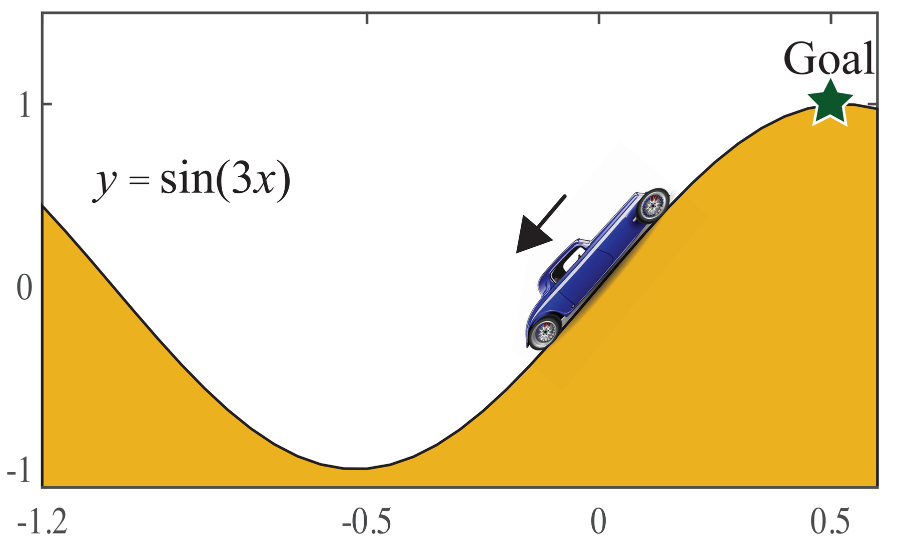


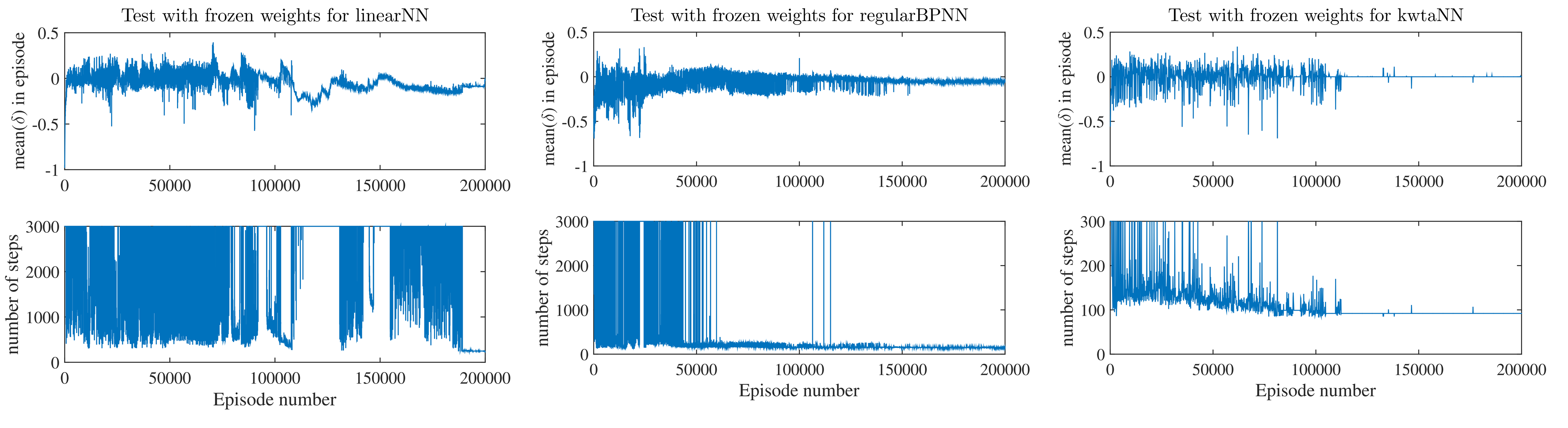
The simple control tasks investigated by Boyan & Moore (1995) and Sutton (1996) were examined using three different value function approximators:

- linear mapping from sensory inputs to value
- backpropagation network with a single hidden layer
- backpropagation network with a single hidden layer and kWTA enforced (k = 10% of hidden units)

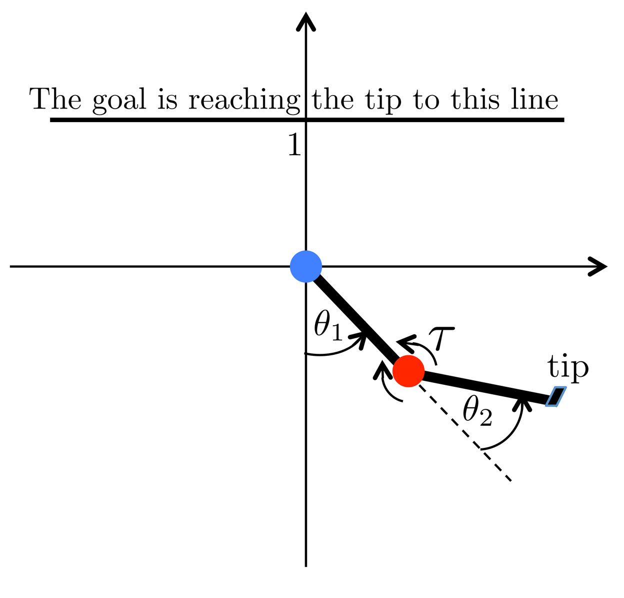
Sensory input representations were identical across the three conditions, and hidden layers were matched in size. Results on the "puddle world" task have already been reported (Rafati & Noelle, 2015).

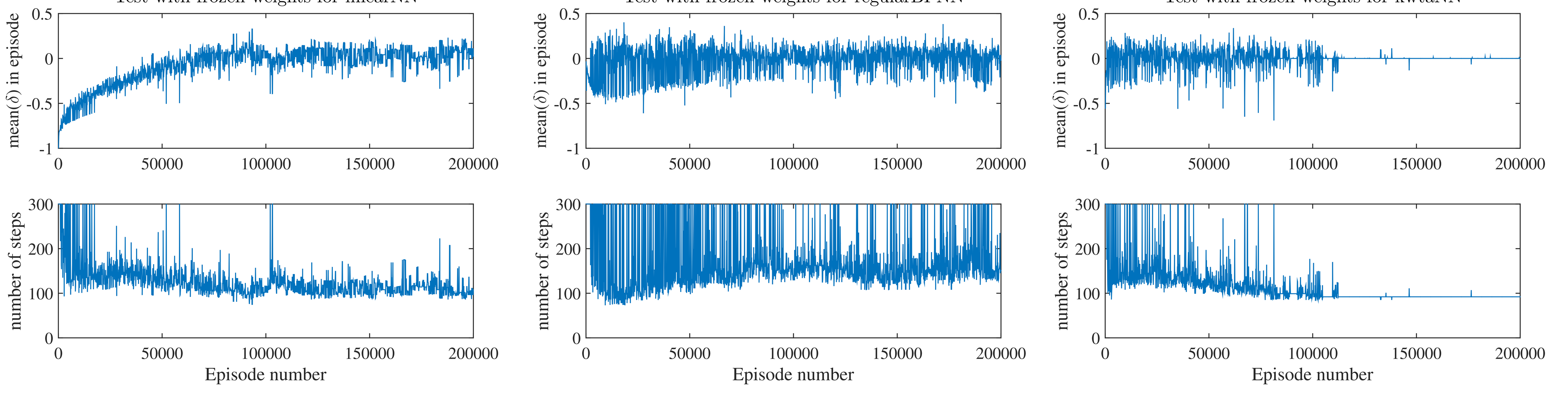
Mountain Car Task





Acrobat Task





References

Boyan, J. A., & Moore, A. W. (1995). Generalization in reinforcement learning: Safely approximating the value function. In G. Tesauro, D. S. Touretzky, & T. K. Leen (Eds.), *Advances in Neural Information Processing Systems 7* (pp. 369–376). Cambridge, MA: MIT Press.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16, 1936-1947.

O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation*, 13, 1199–1242.

Rafati, J., & Noelle, D. C. (2015). Lateral inhibition overcomes limits of temporal difference learning. In D. C. Noelle et al. (Eds.), Proceedings of the 37th Annual Meeting of the Cognitive Science Society. Pasadena, CA: Cognitive Science Society.

Sutton, R. S. (1996). Generalization in reinforcement learning: Successful examples using sparse coarse coding. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in Neural Information Processing Systems 8* (pp. 1038–1044). Cambridge, MA: MIT Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.