UCMERCED

Computational Cognitive Neuroscience Laboratory
University of California, Merced

Lateral Inhibition Overcomes Limits of Temporal Difference Learning

Jacob Rafati & David C. Noelle
Computational Cognitive Neuroscience Laboratory
Electrical Engineering and Computer Science
University of California, Merced

Abstract

There is growing support for Temporal Difference (TD) Learning as a formal account of the role of the midbrain dopamine system and the basal ganglia in learning from reinforcement. This account is challenged, however, by the fact that realistic implementations of TD Learning have been shown to fail on some fairly simple learning tasks — tasks well within the capabilities of humans and non-human animals. We hypothesize that such failures do not arise from natural learning systems because of the ubiquitous appearance of lateral inhibition in the cortex, producing sparse conjunctive internal representations that support the learning of predictions of future reward. We provide support for this conjecture through computational simulations that compare TD Learning systems with and without lateral inhibition, demonstrating the benefits of sparse conjunctive codes for reinforcement learning.

Keywords:

- *reinforcement learning*
- *temporal difference*
- *lateral inhibition*
- *sparse conjunctive codes*
- *computational cognitive neuroscience*

Scan this QR Code to download poster, paper and source code or contact us.

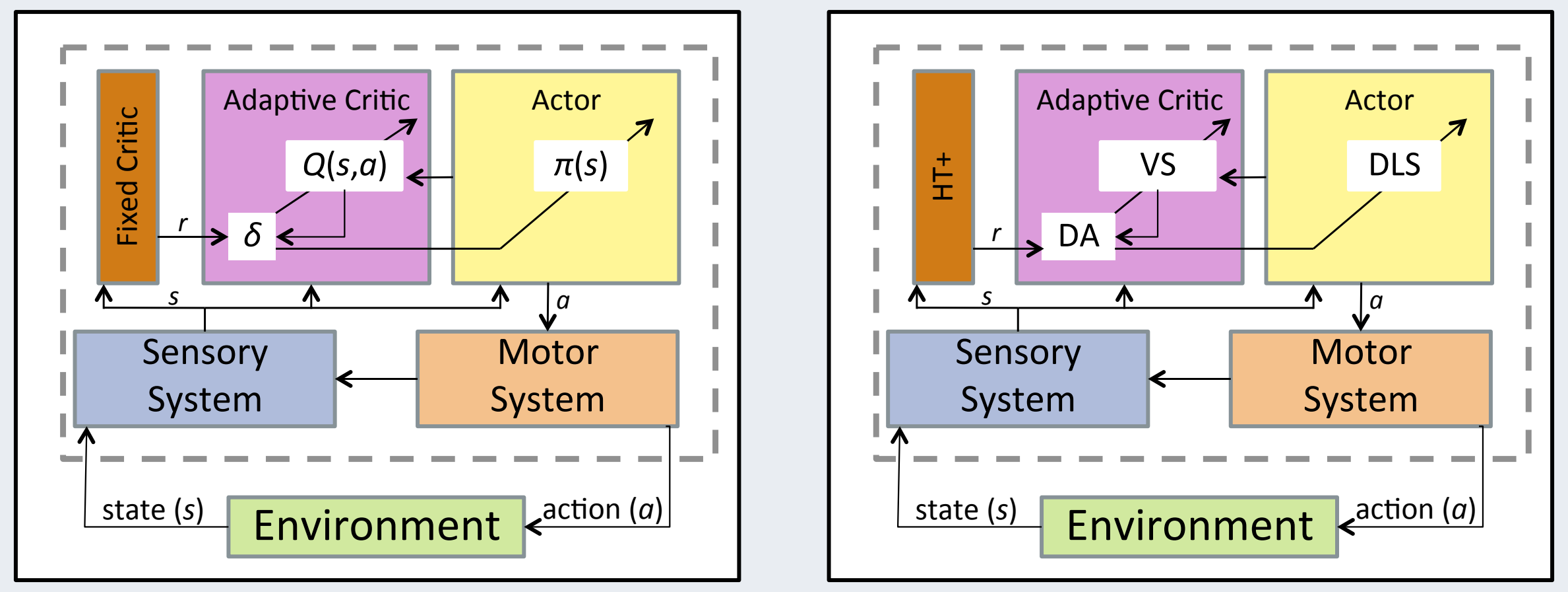
CONTACT

Jacob Rafati
jrafatiheravi@ucmerced.edu
David C. Noelle
dnoelle@ucmerced.edu

Introduction

- The midbrain dopamine (DA) system is essential for reward-based learning and adaptation to the environment.
- Temporal Difference (TD) Learning is a powerful class of reinforcement learning algorithms which successfully describes the information processing role of DA in learning.

Temporal Difference Learning



TD SARSA Algorithm

```
initialize Q(s,a) arbitrarily;
for each episode
      initialize s;
      choose a from s using policy derived from Q (ε-greedy);
      while(s != goal or steps# < allowed#)
            take action a, get reward r, update state to s′;
            choose a′ from s′ using policy derived from Q;
            δ = r + γQ(s′,a′) − Q(s,a);  % compute TD error
            Q(s,a) = Q(s,a) + αδ;  % update value function
            s = s′, a = a′;
      end
end
```

The Problem

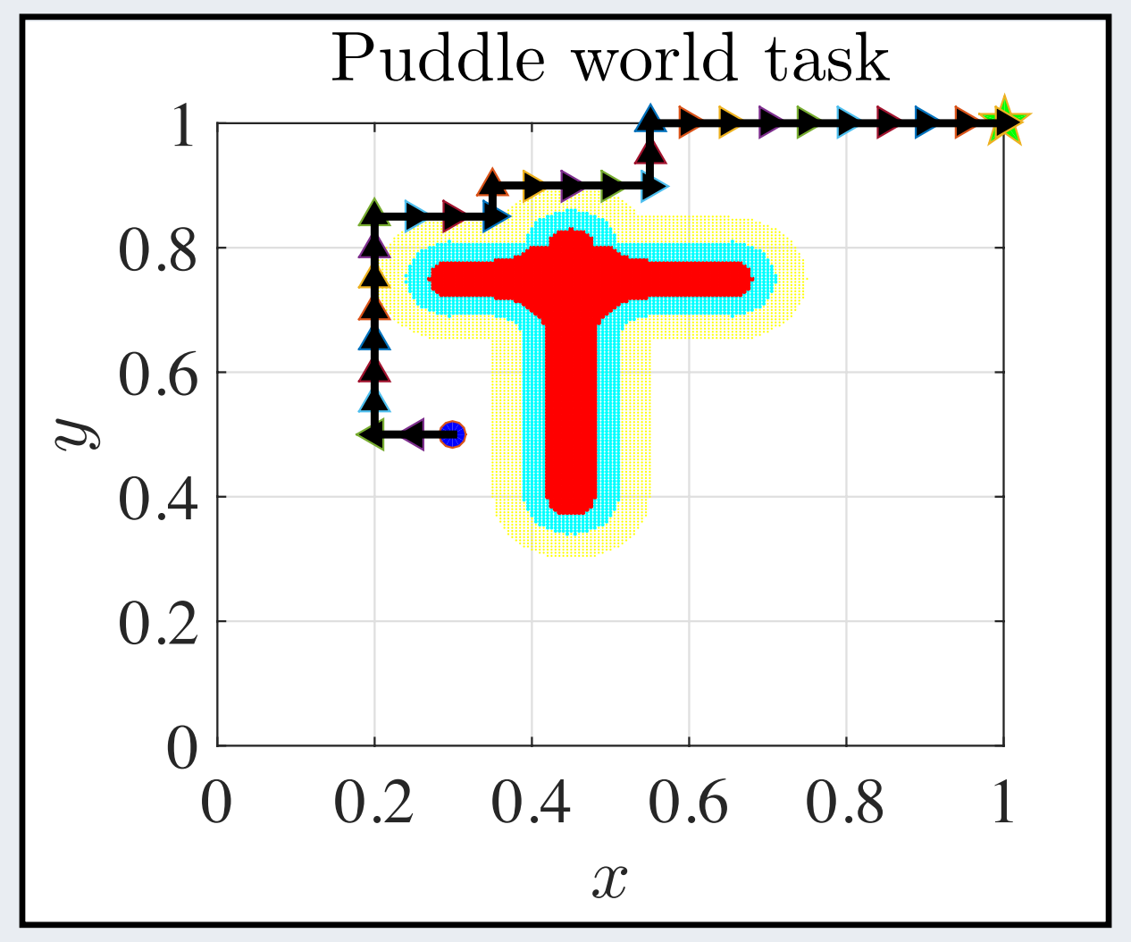
Problem: Sometimes the space of sensory states of the reinforcement learning agent is so large that it is intractable to store the agent's learned value (i.e. expectation of future reward) for each state in a look-up table.

Solution: Use a function approximator, such as an artificial neural network, to map sensory states and considered actions to values, encoding the value function *Q*(*s*,*a*).

Benefits: This approach supports generalization by including a bias toward mapping similar sensory states to similar predictions of future reward.

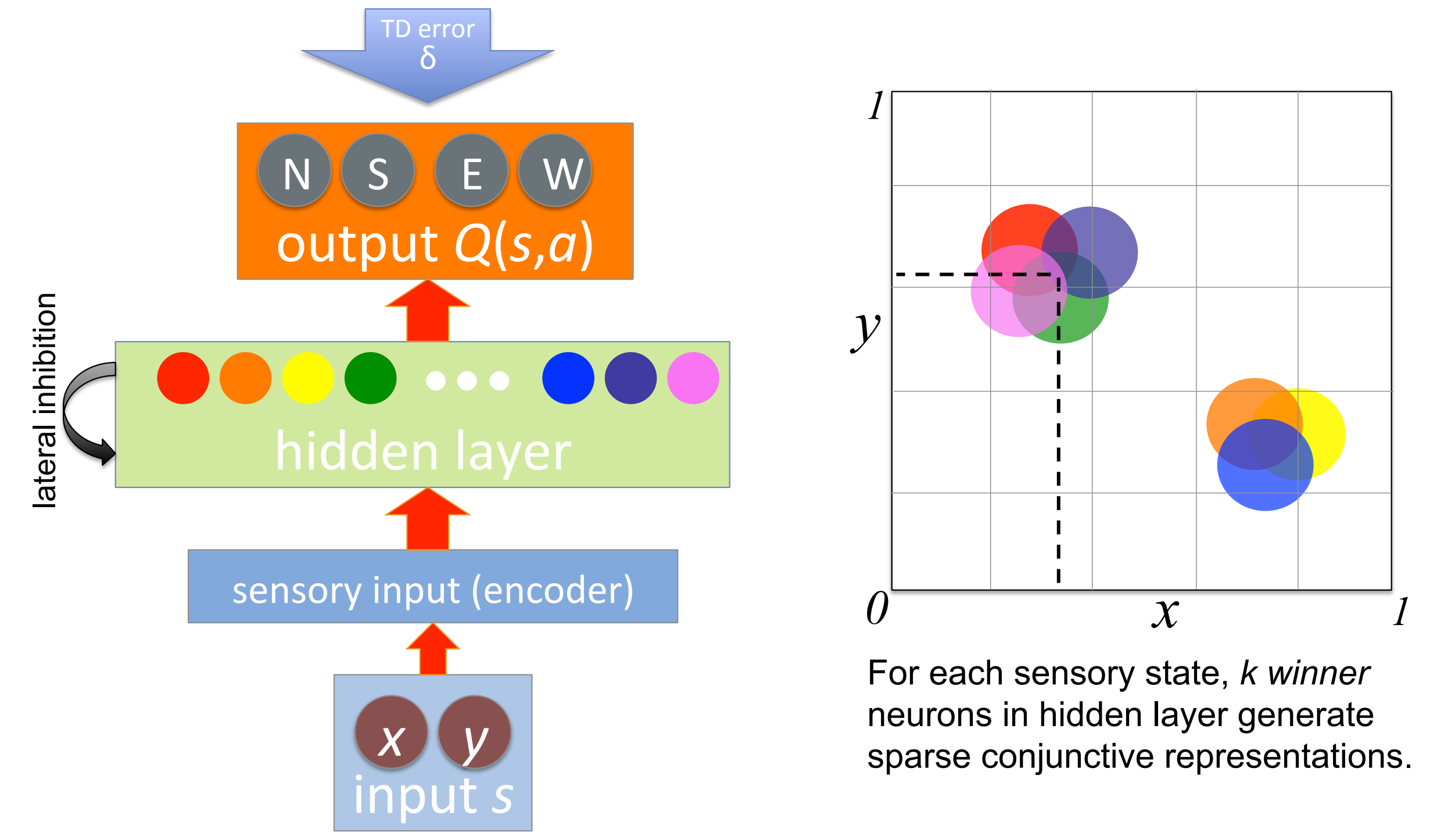
Problem: When a function approximator is used to encode the value function, TD learning can fail, even on some simple learning tasks.

Case study task: Navigation in 2D puddle world



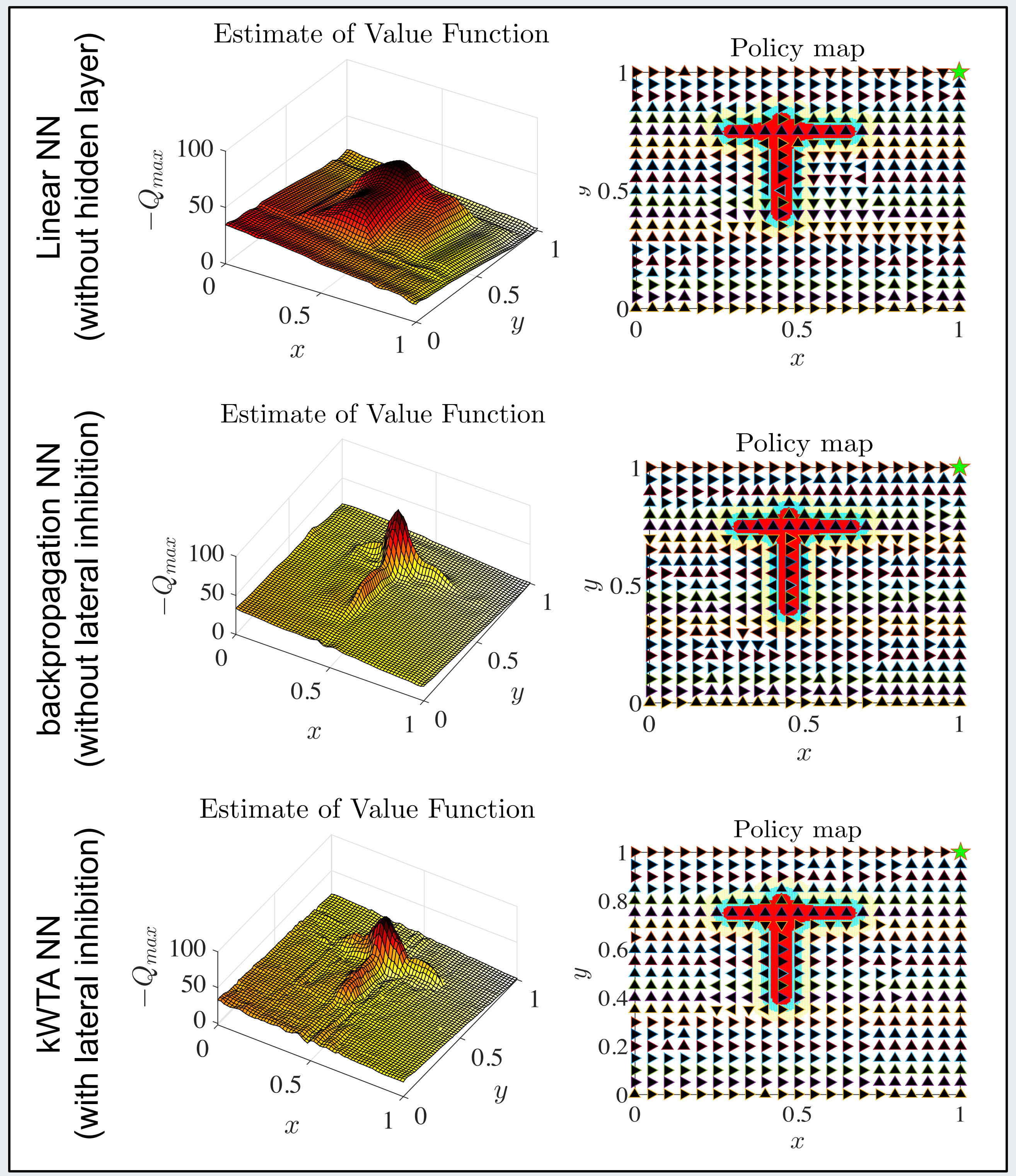
Our Approach

- ❖ Encode the state of the agent in a more continuous manner, modeling the mapping from sensory information to state representation.
- ❖ Allow the value function approximator to learn an appropriate sparse conjunctive representation of the agent state (without hand-wiring that encoding). Such a representation can arise by employing a k-Winners-Take-All (kWTA) mechanism, akin to the result of fast pooled lateral inhibition in biological neural networks.

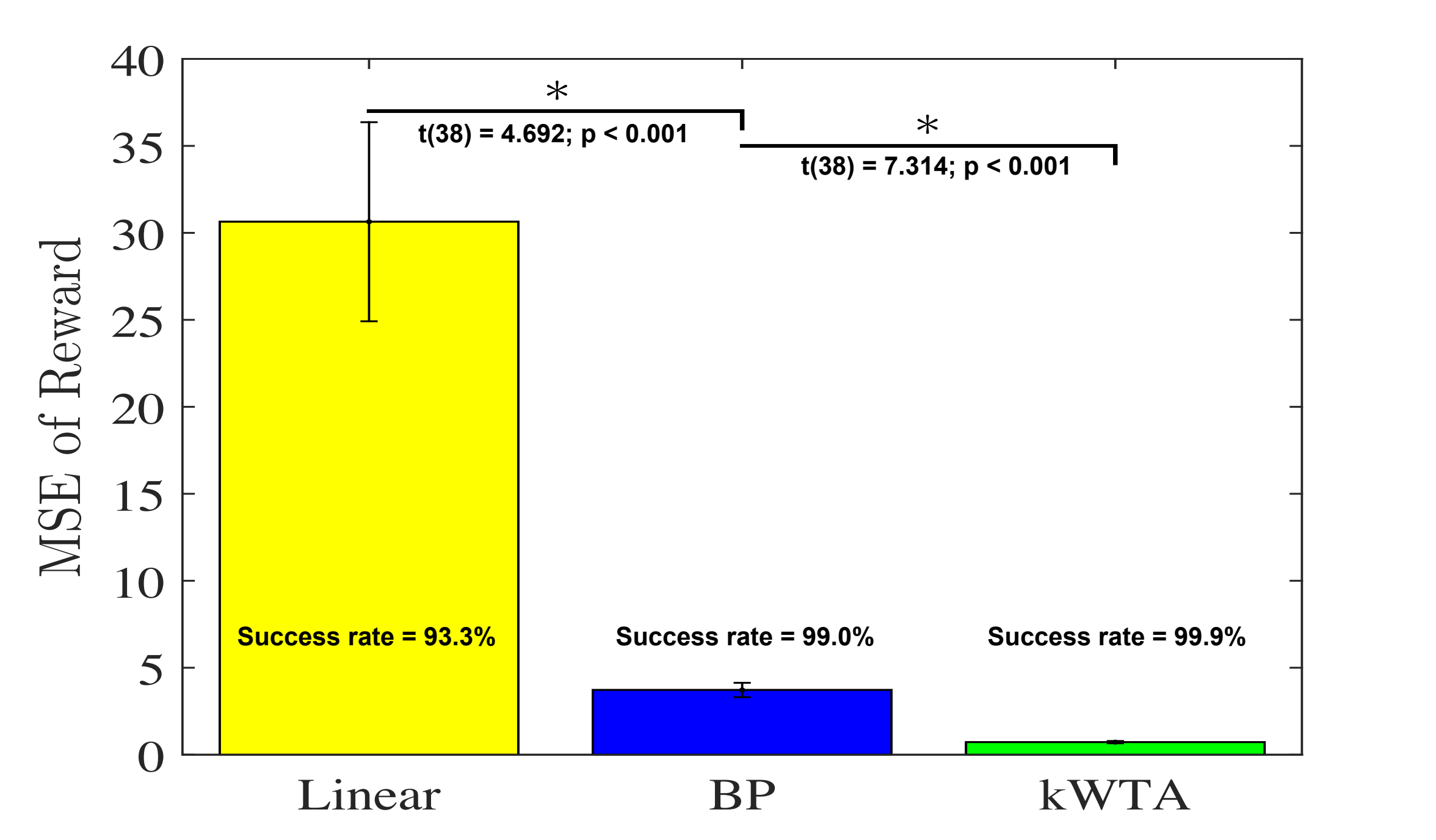


For each sensory state, *k winner* neurons in hidden layer generate sparse conjunctive representations.

Results: Optimum Values & Policy Maps



Results: Statistical Comparisons



Conclusions, Discussion & Future Work

- ❑ A mechanism for learning sparse conjunctive codes for the agent's sensory state can help overcome learning problems observed when using TD Learning with a value function approximator.
- ❑ Artificial neural networks can be biased toward producing sparse codes over their hidden units by including a process akin to the sort of pooled lateral inhibition that is ubiquitous in the cerebral cortex.
- ❑ These results support the hypothesis that the midbrain dopamine system implements a form of TD learning, and observed problems with TD Learning do not arise in the brain because sensory state information is encoded using circuits that make use of lateral inhibition.

Future directions:

- ❖ We are extending this work by applying our kWTA value function approximator to other reinforcement learning problems that have posed difficulties for TD Learning ("mountain car" and "acrobat" control problems).
- ❖ The brain's hippocampus can be seen as generating sparse conjunctive representations. We are exploring the utility of mechanisms like those theorized to arise in the hippocampus to support hierarchically organized learning tasks.

Acknowledgments

Helpful comments were provided by five anonymous reviewers and by members of the University of California, Merced (UCM) Computational Cognitive Neuroscience Laboratory (CCNL).

References

Boyan, J. A., & Moore, A. W. (1995). Generalization in reinforcement learning: Safely approximating the value function. *Advances in neural information processing systems 7* (pp. 369–376). Cambridge, MA: MIT Press.

Dayan, P. (1992). The convergence of TD(λ) for general λ. *Machine Learning*, *8*, 341–362.

O'Reilly, R. C., & Munakata, Y. (2001). *Computational explorations in cognitive neuroscience*. Cambridge, Massachusetts: MIT Press.

Sutton, R. S. (1996). Generalization in reinforcement learning: Successful examples using sparse coarse coding. *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.

Tesauro, G. (1995). Temporal difference learning and TD-Gammon. *Communications of the ACM*, *38*(3).