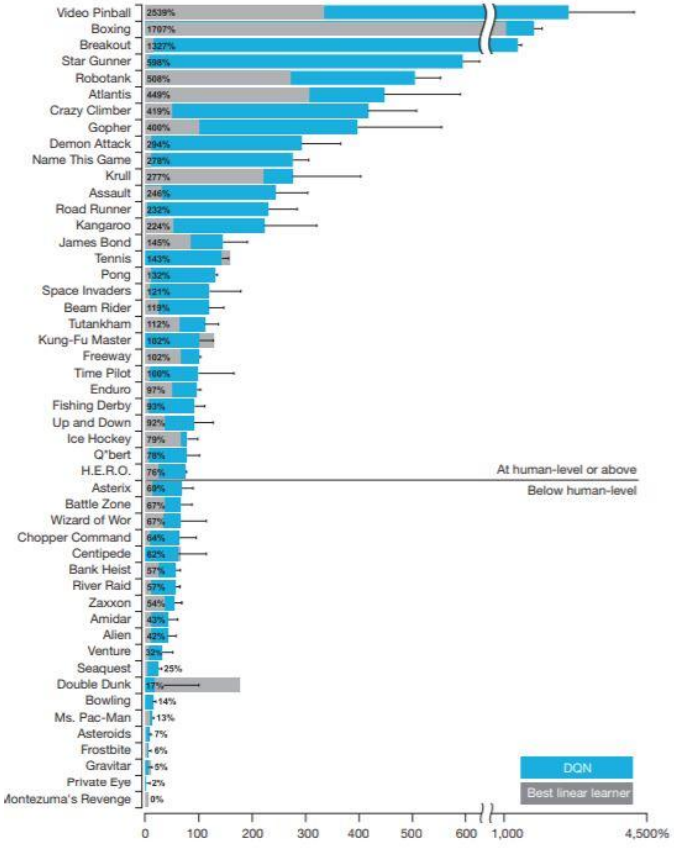
Sparse Reward and Temporal Abstraction: Hierarchical DQN

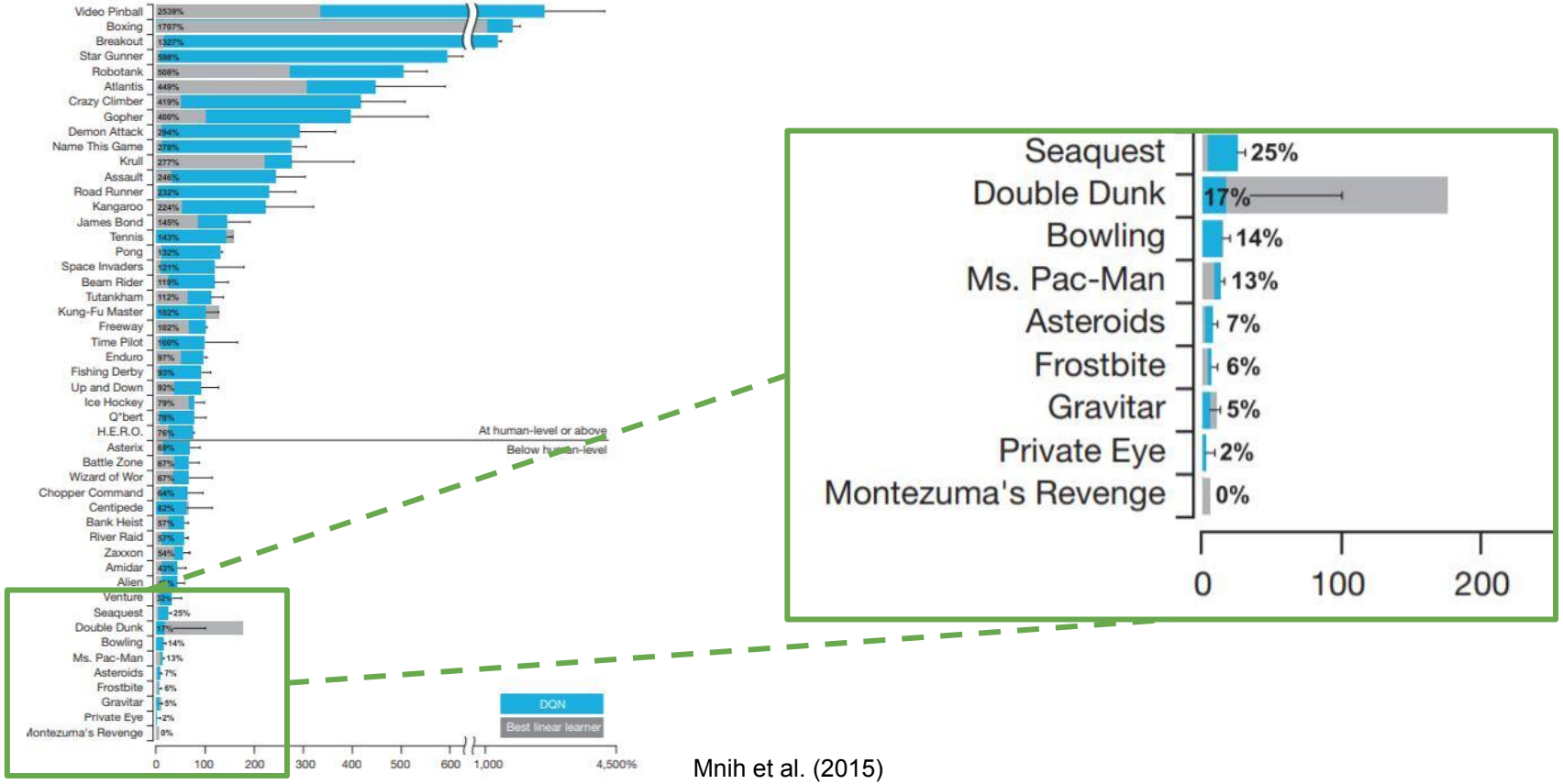
Gabe Margolis

Motivation: DQN and Atari



Mnih et al. (2015)

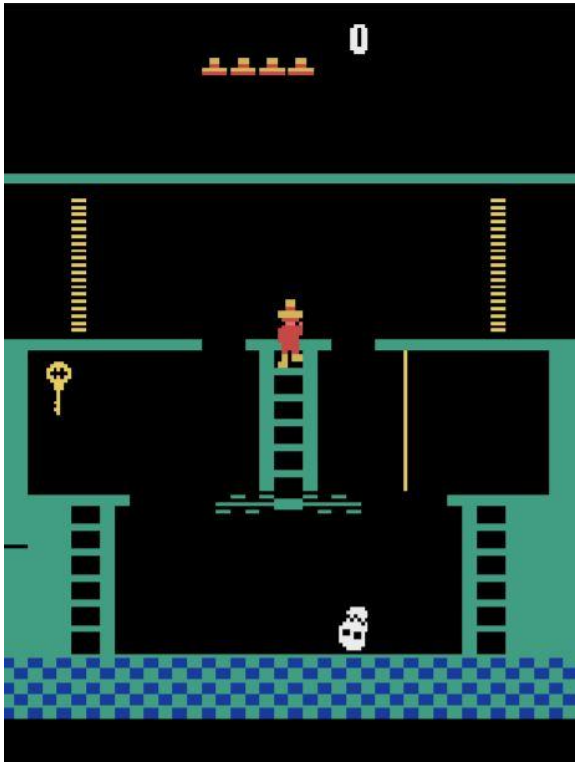
Motivation: DQN and Atari

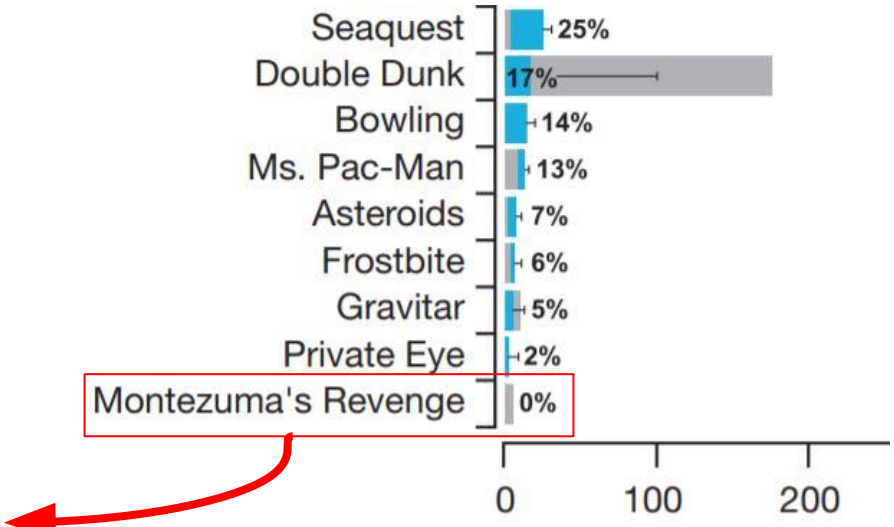


Mnih et al. (2015)

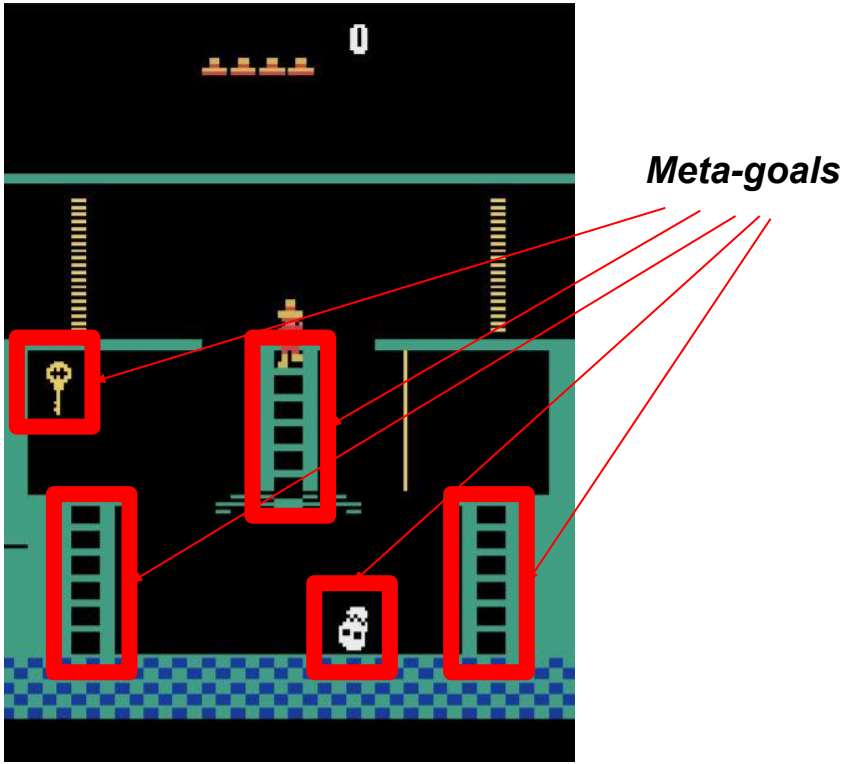
DQN Fails to Learn from Sparse, Delayed Reward

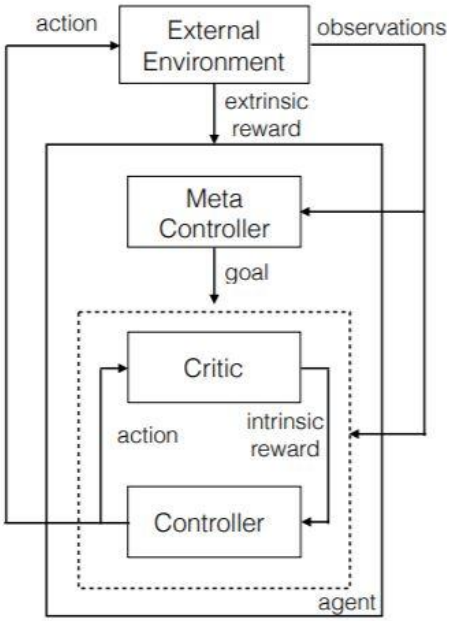
Montezuma’s Revenge (Atari)





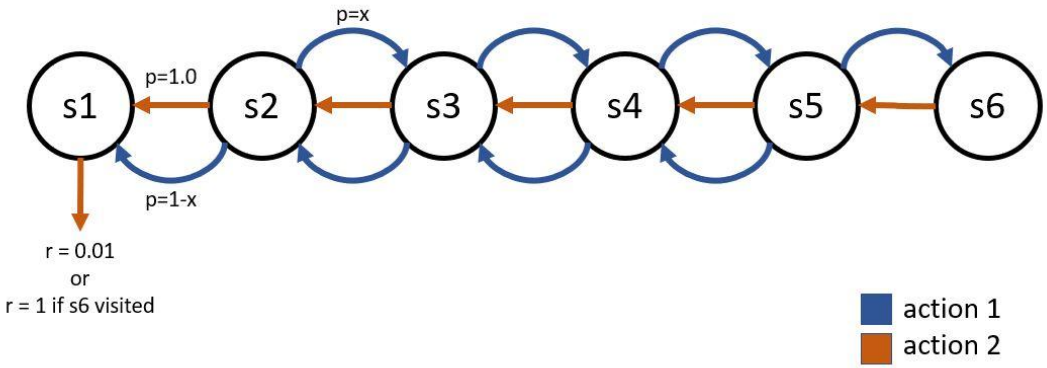
Hierarchical Deep Reinforcement Learning (hDQN)





hDQN Architecture (Kulkarni et al. 2016)

Sparsity Simplified: Discrete Stochastic Decision Process



Replication of Results

Kulkarni et al.

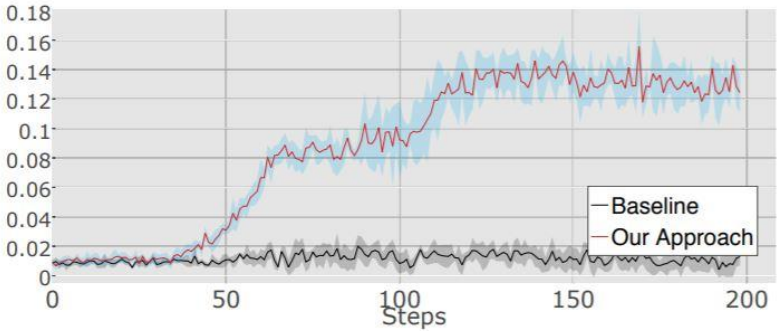
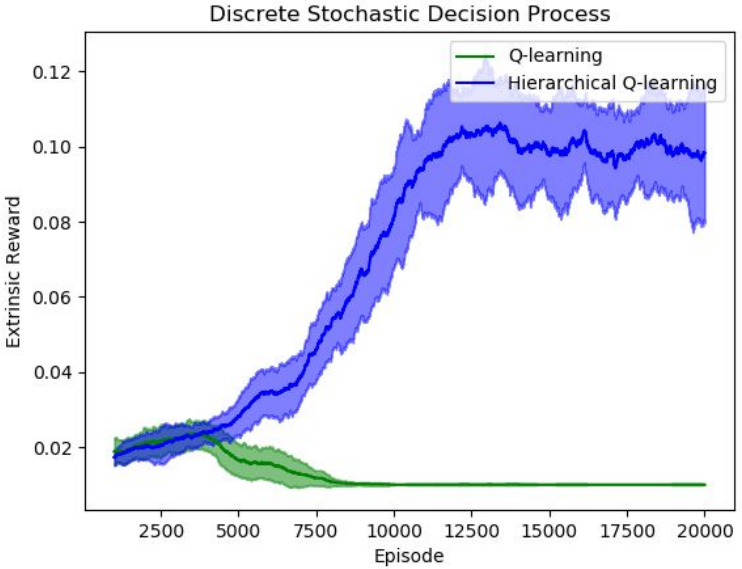


Figure 3: Average reward for 10 runs of our approach compared to Q-learning.

Ours



Replication of Results

Kulkarni et al.

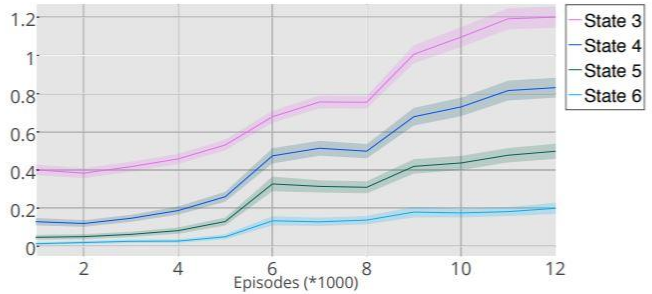
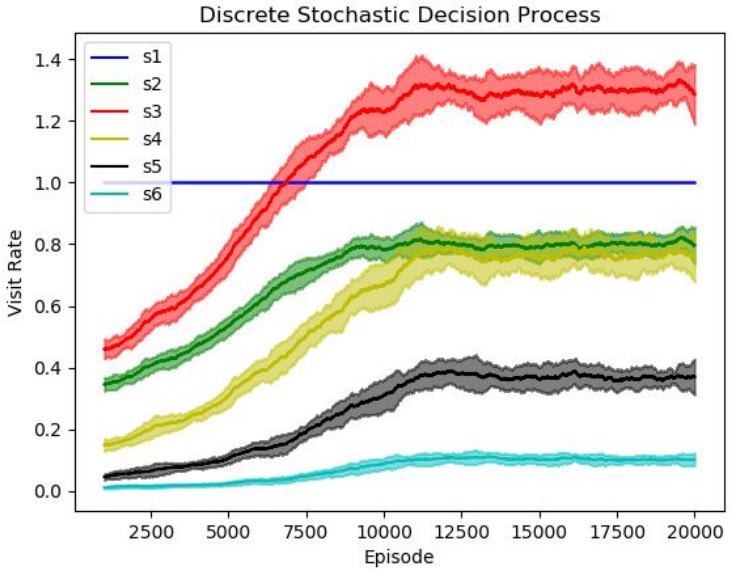
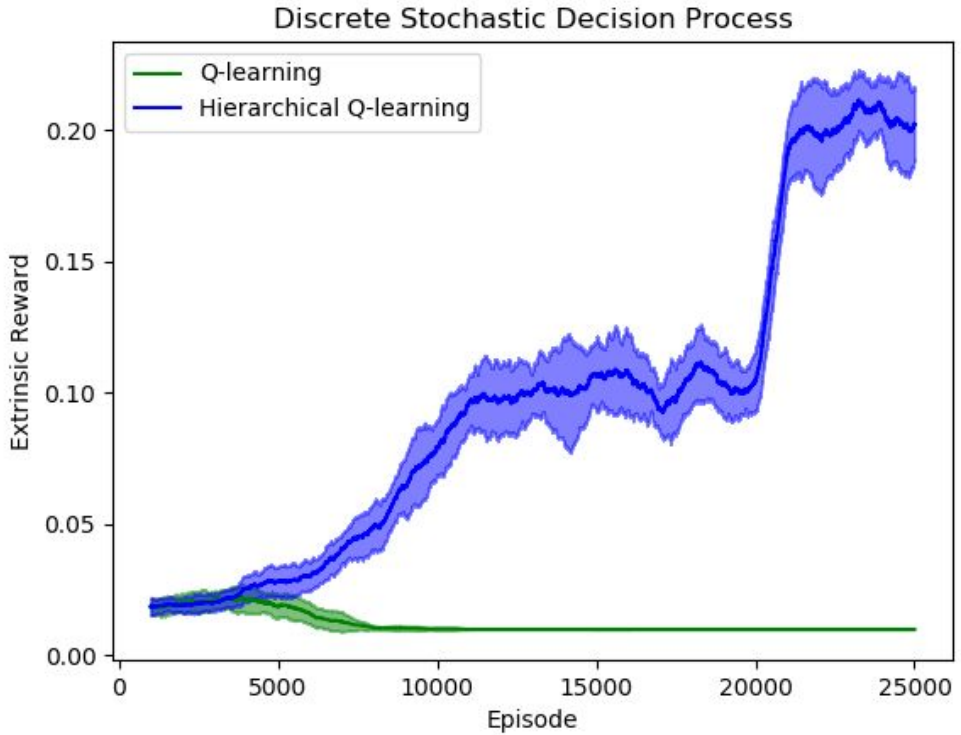


Figure 4: Number of visits (for states s_3 to s_6) averaged over 1000 episodes. The initial state is s_2 and the terminal state is s_1.

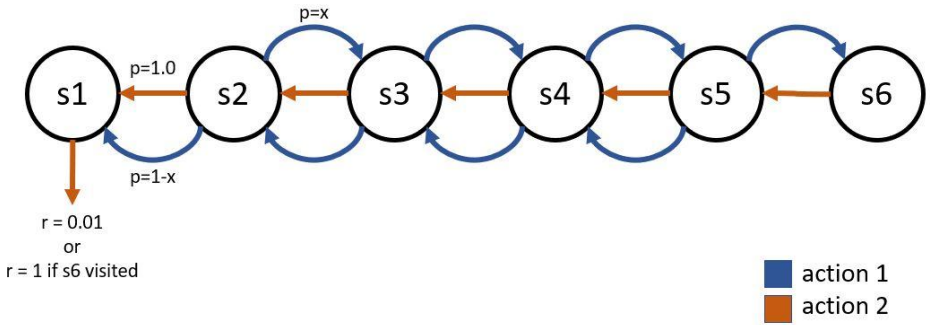
Ours



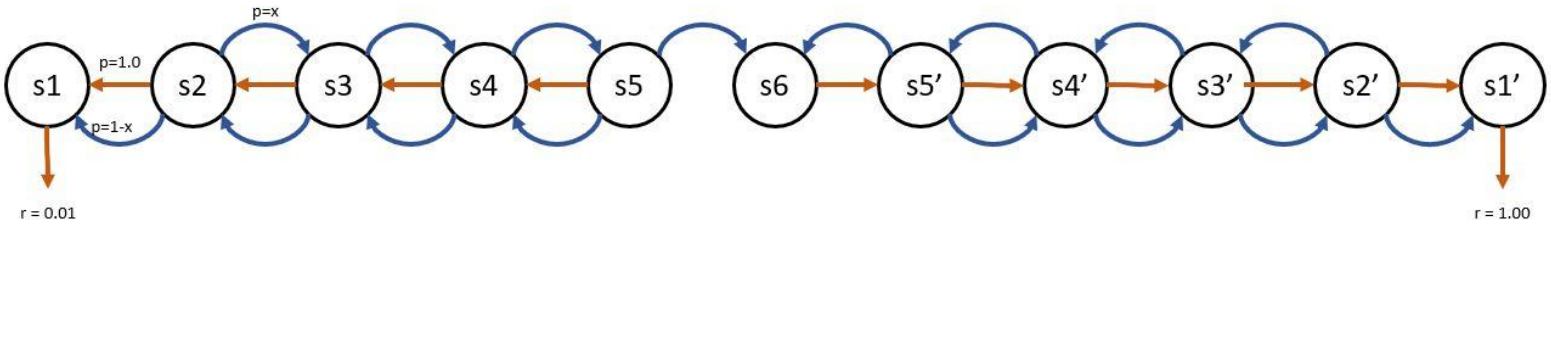
Optimal Reward Achieved under Pure Exploitation



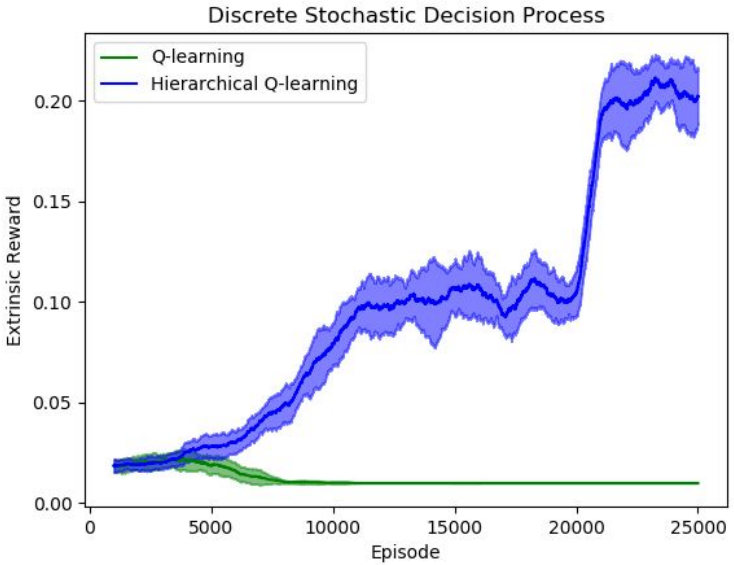
Analysis: Sparsity vs Observability

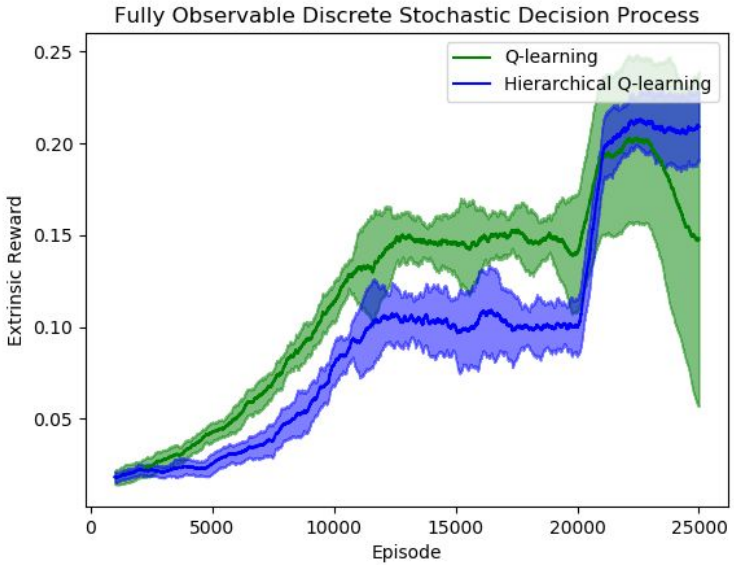


Analysis: Sparsity vs Observability



Experiments in Fully Observable Environment





Summary

- Successfully replicated h-DQN paper results
- Separated impacts of sparsity and partial observability/temporal abstraction
- Beat Montezuma’s Revenge! (If I were Google)