Multicopy Reinforcement Learning Agents

Alicia P. Wolfe. Oliver Diamond, Brigitte Goeler-Slough,
Remi Feuerman, Magdalena Kisielinska, Victoria Manfredi
Wesleyan University
Middletown, CT, United States
pwolfe,odiamond,bgoelersloug,rfeuerman,mkisielinska,vumanfredi@wesleyan.edu

ABSTRACT

Inspired by the problem of optimal packet duplication in Mobile Wireless Networks, this paper examines a specific type of cooperative multi-agent problem in which an agent makes multiple identical copies of itself in order to achieve a noisy and difficult single agent task more reliably or efficiently. The agent must balance the cost of sending more copies with the improvement in speed or success rate generated by the extra copies. Due to the space of possible joint reward functions, this specific case has some differences from the more general cooperative multi-agent problem. We propose a learning algorithm for this multicopy problem which takes advantage of the structure of the value function to efficiently learn how to balance the costs and benefits of adding additional copies.

KEYWORDS

Multiagent Reinforcement Learning, Cooperative Agents, Wireless Networks

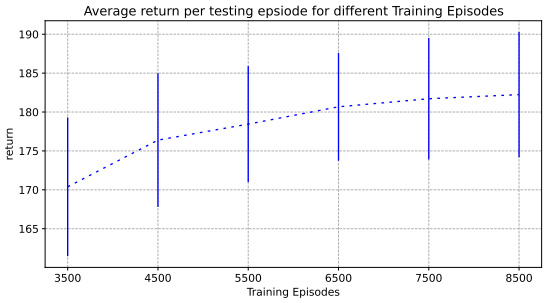
1 APPENDIX

This appendix details some of the experiments we ran to determine reasonable parameters for various parts of the main algorithm.

Unless otherwise noted, all experiments in this section were run using the Three Bridges Gridworld from the main paper, with noise 0.2, cost -3, Monte Carlo learning rate of 0.05 and Q-learning learning rate of 0.2. All experiments were repeated 100 times and results during the 500 step testing phase were averaged.

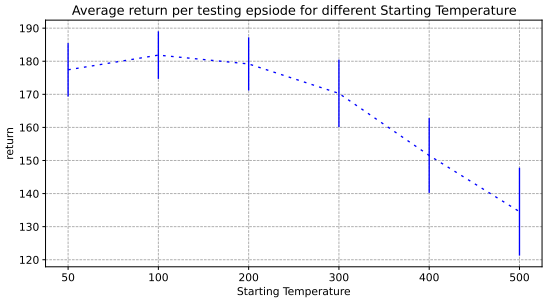
1.1 Training Episodes (Training Data)

The earliest point at which learning seems to have converged is around 7500 episodes.



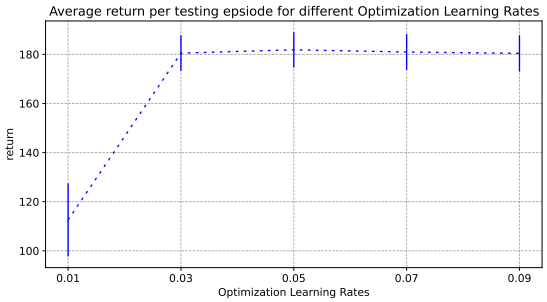
1.2 Starting Temperature (Exploration)

The performance of different starting temperatures for this particular gridworld has a fairly clear peak around 100.



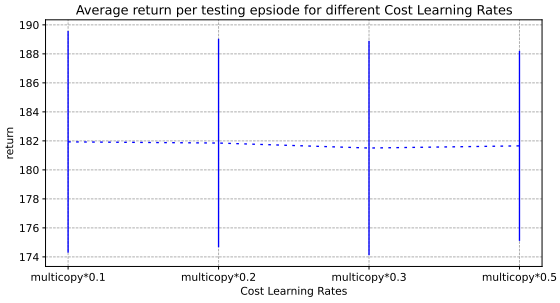
1.3 Optimization Learning Rate

The optimization learning rate results are fairly steady within the range 0.03-0.07. We chose the middle of this range.



1.4 Cost Learning Rate

The cost learning rate results are fairly steady within the range 0.1-0.5. We chose 0.2 due to its frequency in the literature on gridworlds.



Proc. of the Adaptive and Learning Agents Workshop (ALA 2025), Avalos, Aydeniz, Müller, Mohammedalamen (eds.), May 19 – 20, 2025, Detroit, Michigan, USA, ala-workshop.github.io. 2025.