Heuristics for the Traveling Salesperson Problem based on Reinforcement Learning

January 8, 2022

Corrado Coppola, Giorgio Grani, Marta Monaci, Laura Palagi
Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy
{corrado.coppola, g.grani, marta.monaci, laura.palagi}@uniroma1.it

Abstract. We use Reinforcement Learning, in particular a deep Q-Learning algorithm and an adaptation of two actor-critic algorithms (Proximal Policy Optimization and Phasic Policy Gradient), originally proposed for robotic control, to solve the metric Traveling Salesperson Problem (TSP) on Hamiltonian graphs with a given costs distribution. We embed the TSP into a Markov Decision Process (MDP) environment. We use Convolutional Neural Networks (CNNs) to approximate action-value and state-value functions and we develop a neural architecture centered on the idea of considering the weighted incidence matrix of a graph with respect to arcs costs as the agent's environment representation at a given instant. Our computational experience shows that during the training process the CNNs-based architecture is both efficient in terms of CPU time and effective in terms of solutions found. Using the state-of-the-art solver for routing problems ORTOOLS during the test process, we find that, differently from deep Q-Learning, both PPO and PPG can achieve the same benchmark solution with a smaller computational effort; we also find that our trained models are able to orient themselves through new unseen graphs and with different costs distributions. Eventually, the test results suggest that the computational efficiency grows more than proportionally with the instants size.

Keywords: traveling salesperson problem; reinforcement learning; heuristic

References

[1] Cobbe, Karl W., et al. "Phasic policy gradient." International Conference on Machine Learning. PMLR, 2021.

[2] Dmitry Krass Jerzy A. Filar. The embedding of the traveling salesman problem in a markov decision process. Proceedings 01 the 26th Conference on Declslon and Control, Los Angeles, 1987.

[3] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." nature 518.7540 (2015): 529-533.

[4] Schulman, John, et al. "Trust region policy optimization." International conference on machine learning. PMLR, 2015.

[5] Schulman, John, et al. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).