Learning 2-opt Local Search for the Traveling Salesman Problem

Paulo R. de O. da Costa, Jason Rhuggenaath, Yingqian Zhang, and Alp Akcay

Eindhoven University of Technology, 5612 AZ Eindhoven, Netherlands
{p.r.d.oliveira.da.costa, j.s.rhuggenaath, yqzhang, a.e.akcay}@tue.nl

Abstract. Recent works using deep learning to solve the Traveling Salesman Problem (TSP) have focused on learning construction heuristics. Such approaches require additional procedures such as beam search and sampling to improve solutions and achieve state-of-the-art performance. However, few studies have focused on improvement heuristics, where a given solution is improved until reaching a near-optimal one. In this work, we propose to learn a local search heuristic based on 2-opt operators via deep reinforcement learning. We propose a policy gradient algorithm to learn a stochastic policy that selects 2-opt operations given a current solution. Moreover, we introduce a policy neural network that leverages a pointing attention mechanism, which unlike previous works, can be easily extended to more general *k*-opt moves. Our results show that the learned policies can improve even over random initial solutions and approach near-optimal solutions at a faster rate than previous state-of-the-art deep learning methods.

Keywords: Deep Reinforcement Learning · Combinatorial Optimization · Traveling Salesman Problem.
Acknowledgments: This research is funded by NWO Big data: Real Time ICT for Logistics, project number 628.009.012
Publication: The full paper of this abstract has been accepted at the 12th Asian Conference on Machine Learning (ACML), 2020. Available at http://proceedings.mlr.press/v129/costa20a/costa20a.pdf

1 Introduction

The Traveling Salesman Problem (TSP) is a well-known NP-hard combinatorial optimization problem. Exact methods for the TSP such as linear programming [1] are guaranteed to find an optimal solution but are often too expensive computationally. On the other hand, designed heuristics require specialized knowledge and their performances are often limited by algorithmic design decisions.

Thus, a machine learning method could potentially learn better heuristics by extracting useful information directly from data. We focus on methods in which a given solution is improved sequentially until reaching an (local) optimum. Thus, we propose a deep reinforcement learning algorithm to learn improvement heuristics based on 2-opt moves. Our approach can achieve near-optimal results that are better than previous deep learning methods based on construction and improvement heuristics.

2 Methods

Our neural network follows the general encoder-decoder architecture. The encoder embeds both graph topology and the positions of each node in a solution. Given node and sequence embeddings the *policy* decoder is autoregressive and samples output actions one element at a time. The *value* decoder operates on the same representations but generates real-valued outputs to estimate state values.

In our formulation, we resort to the Policy Gradient learning rule, to optimize our policy. Our model is close to REINFORCE [3] but with periodic episode length updates. Thus, at the start the agent learns how to behave over small episodes for easier credit assignment, later tweaking its policy over larger horizons.

3 Results

We learn policies for TSP instances with 20, 50 and 100 nodes, and depict the optimality gap for 10,000 test instances in Table 1. The results show that we can learn effective policies that decrease the optimality gap over the training epochs and can outperform the effective Graph Attention (GAT) [2] and are close to the optimal solutions.

Table 1: Performance of TSP methods w.r.t. Concorde. *Type*: **RL**: Reinforcement Learning, **S**: Sampling, *Time*: Time to solve 10,000 instances.

Method	Type	TSP20			TSP50			TSP100		
		Cost	Gap	Time	Cost	Gap	Time	Cost	Gap	Time
Concorde [1]	Solver	3.84	0.00%	(1m)	5.70	0.00%	(2m)	7.76	0.00%	(3m)
GAT [2]	RL,S	3.84	0.08%	(5m)	5.73	0.52%	(24m)	7.94	2.26%	(1h)
Ours	RL	**3.84**	**0.00**%	(15m)	5.70	0.12%	(29m)	**7.83**	**0.87**%	(41m)

References

1. Applegate, D.L., Bixby, R.E., Chvatal, V., Cook, W.J.: The traveling salesman problem: a computational study. Princeton university press (2006)
2. Kool, W., van Hoof, H., Welling, M.: Attention, learn to solve routing problems! In: Proceedings of the 7th International Conference on Learning Representations (ICLR) (2019)
3. Williams, R.J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning **8**(3-4), 229–256 (1992)