Professorship of Business Analytics and Intelligent Systems
TUM School of Management
Technical University of Munich

Selecting Operators in LNS with Reinforcement Learning

Problem definition

The Adaptive Large Neighborhood Search (ALNS) (Pisinger & Ropke 2007) improves on Large Neighborhood Searches (Shaw 1998) by adding an adaptive layer, which selects from multiple predefined destroy- and repair-heuristics. The adaptive layer commonly picks heuristics via a roulette wheel selection. The probability of a heuristic being selected is adapted during an optimization run based on whenever an operator was used, depending on the success or failure of the last iteration. Turkes et al. (2021) performed an in depth meta-analysis of ALNS algorithms and found that, on average, meta-heuristics perform only 0.14% better when they adapt operator selection percentages during an optimization run. They argue that a robust pre-selection of operators is more important than having an adaptive operator selection layer.

Instead of simply removing the adaptive layer, an improvement to the operator selection process may be possible. Recently, Machine Learning-based approaches have been proposed to select operators. This thesis' objective is to investigate potential improvements to the standard Roulette Wheel Operator Selection (RWOS) via an agent trained with Reinforcement Learning (RL), which we will refer to as Reinforcement Learning Operator Selection (RLOS). This objective is motivated by shortcomings of RWOS and we propose RLOS to improve upon these, as detailed in the following:

- RWOS chooses operators only depending on their historic performance within one optimization run. The RLOS agent can be supplied with additional information, such as current iteration count and domain specific variables. The input could also include the historical performance of the operators within the current optimization run, meaning RLOS loses no information in comparison to RWOS.
- RWOS weights are reset whenever a new optimization run is started, leading to a loss of learnings from previous optimization runs. Depending on the initial probabilities, a long warm-up time might be needed before the RWOS can make "good" operator selections. While an ALNS meta-heuristic seldom needs to solve the same problem twice, problems solved by the same algorithm are usually close enough to each other that an RLOS actor trained on similar problems might reduce warm-up times compared to a traditional adaptive layer.
- The success of RWOS is highly dependent on the provided operators. Since the selection is random, a large pool of operators can disturb the selection and make it hard to consistently choose good operators. Therefore, a domain specific pool of (handcrafted) destroy- and repair-operators is needed to ensure good results (Johnn et al. 2023). In contrast, an RLOS agent could potentially be supplied with a high number of unfiltered operators and still make consistent good selections. While this could potentially slow down the training of the agent, time could be saved when creating an algorithm, as the operator selection could be skipped.

Aims and scope of the thesis

It is subject of this thesis to analyze whether an RLOS can significantly outperform an RWOS. This comprises the following research tasks:

- Creating an ALNS meta-heuristic to solve Capacitated Vehicle Routing Problems (CVRP).
- Comparing the standard RWOS with non-adaptive stochastic operator selection.
- Creating and training an RLOS actor.
- Comparing the RLOS actor against RWOS and non-adaptive stochastic operator selection.
- Performing a case study on the impact of different factors on the performance of RWOS and RLOS. These factors could for example be the amount of operators, the input information and the action space of the actor, the reward function (ex- or including rewards for faster operators), as well as the difference between training domain and other CVRPs. Since retraining a large amount of actors is infeasible, a preliminary test of the factors could be done on small instances and with a low amount of settings.

Related Research

Authors who replace RWOS with RLOS:

Bongiovanni, C., Kaspi, M., Cordeau, J.F., & Geroliminis, N. (2022). A machine learning-driven two-phase metaheuristic for autonomous ridesharing operations. Transportation Research Part E: Logistics and Transportation Review, 165.

Johnn, S.N., Darvariu, V.A., Handl, J., & Kalcsics, J. (2023). Graph Reinforcement Learning for Operator Selection in the ALNS Metaheuristic. https://doi.org/10.48550/arXiv.2302.14678

Lu, H., Zhang, X., & Yang, S. (2020). A Learning-based Iterative Method for Solving Vehicle Routing Problems. In: International Conference on Learning Representation 2020. https://openreview.net/forum?id=BJe1334YDH

Reijnen, R., Zhang, Y., Lau, H. C., & Bukhsh, Z. (2022). Operator Selection in Adaptive Large Neighborhood Search using Deep Reinforcement Learning. https://doi.org/10.48550/arXiv.2211.00759

Authors who replace handcrafted heuristics with learned heuristics:

Gao, L., Chen, M., Chen, Q., Luo, G., Zhu, N., & Liu, Z. (2020). Learn to Design the Heuristics for Vehicle Routing Problem. https://doi.org/10.48550/arXiv.2002.08539

Hottung, A., & Tierney, K. (2020). Neural Large Neighborhood Search for the Capacitated Vehicle Routing Problem. In G. de Giacomo, A. Catala, B. Dilkina, M. Milano, S. Barro, A. Bugarín, & J. Lang (Eds.), Frontiers in Artificial Intelligence and Applications: Volume 325. ECAI 2020: 24th European Conference on Artificial Intelligence 2020, Santiago de Compostela, Spain. IOS Press.

Wu, Y., Song, W., Cao, Z., Zhang, J., & Lim, A. (2022). Learning Improvement Heuristics for Solving Routing Problems. IEEE Transactions on Neural Networks and Learning Systems, 33(9), 5057–5069.