Supplementary Material for: Reinforcement Learning for Solving Stochastic Vehicle Routing Problem

Zangir Iklassov ZANGIR.IKLASSOV@MBZUAI.AC.AE
Ikboljon Sobirov IKBOLJON.SOBIROV@MBZUAI.AC.AE
Ruben Solozabal RUBEN.SOLOZABAL@MBZUAI.AC.AE
Martin Takáč MARTIN.TAKAC@MBZUAI.AC.AE
MBZUAI, UAE, Abu-Dhabi

Editors: Berrin Yanıkoğlu and Wray Buntine

Appendix A. Classical Formulation

The objective of the problem is to

$$\begin{aligned}
\text{minimize} \quad & \textstyle\sum_{i,j\in C} c_{ij}x_{ij} + \mathscr{R}(x), && (1)\\
\text{subject to} \quad & \textstyle\sum_{j=1}^{n} x_{0j} = 2|K|, && (2)\\
& \textstyle\sum_{i<k} x_{ik} + \sum_{j>k} x_{kj} = 2 \quad \forall k \in N, && (3)\\
& \textstyle\sum_{i,j\in C} x_{ij} \leq |C| - [\sum_{i\in C} \mathbb{E}[\xi_i]/Q], && (4)\\
& 0 \leq x_{ij} \leq 1 \quad \forall i,j \in C, && (5)\\
& 0 \leq x_{0j} \leq 2 \quad \forall j \in N, && (6)\\
& x = (x_{ij}) \ \text{ integer} \quad \forall i,j \in N. && (7)
\end{aligned}$$

where,
N : set of customers and depot,
C : customers set,
c_{ij} : stochastic travel cost between nodes i and j,
c_{ijs} : s^{th} realization of $\{ij\}$ travel cost,
ξ_i : stochastic demand of customer i,
μ_{is} : s^{th} realization of the demand of customer i,
K : number of vehicles,
Q : maximum capacity of vehicle,
x_{ij} : binary variable that shows whether (i, j) is used in the route.

The goal of the formulation is to minimize the total traversing cost in Equation 1, which is the sum across the arc costs in the route. Several constraints are set to guarantee that (i) each vehicle starts and finishes at a depot (Equation 2), (ii) each customer is paid a visit only once (Equation 3), (iii) the maximum load each vehicle has covers the total expected demand (Equation 4), and (iv) the decision variable x_{ij} is an integer for every arc (Equations 5, 6, 7). The objective of setting such rules is to find a feasible solution that considers the underlying constraints of the problem.

In case of situations when a vehicle fails to fulfill the demand of a customer due to a shortage of load, the recourse cost $\mathscr{R}(x)$ provided in Equation 1 accounts for it. To be precise, the term is responsible for the cost incurred for traversing to the depot and back to refill in order to meet the customer demand. The incurred cost mathematically is represented as:

$$\mathscr{R}(x) = \sum_{k=1}^{K} \mathscr{R}^k(x),$$
$$\mathscr{R}^k(x) = 2\sum_{j=2}^{t}\sum_{l=1}^{j-1} P(\sum_{s=2}^{j-1}\xi_s \leq lQ < \sum_{s=2}^{j}\xi_s)c_{0j}.$$

where the total incurred cost (i.e. the total recourse cost) equals the sum of the additional cost each vehicle incurs (k), which is mathematically expressed as $\mathscr{R}^k(x)$. It computes the likelihood of the l^{th} failure case at the j^{th} customer along the route.

Appendix B. Implementation

Parameters. The resulting outputs, representing $P(p_{1:K}^{t+1}|\cdot)$, are passed through a Critic network, which consists of two fully-connected layers, with D and 1 neurons. The Rectified Linear Unit (ReLU) activation function is employed. The dimensionality of the embeddings D is set to 128.

Training. We employ the Xavier initialization method. During the training phase, we utilize the Adam optimizer with a learning rate of 10^{-4}. To prevent overfitting, we apply dropout with a probability of 0.1. The model is trained on a GPU system consisting of a NVIDIA A100 SXM 40GB GPU and 2x AMD EPYC 7742 CPUs (8 cores) with 256GB RAM, for a total of 10,000 iterations for each problem size. The implementation of the model is available online [1].

Appendix C. Results

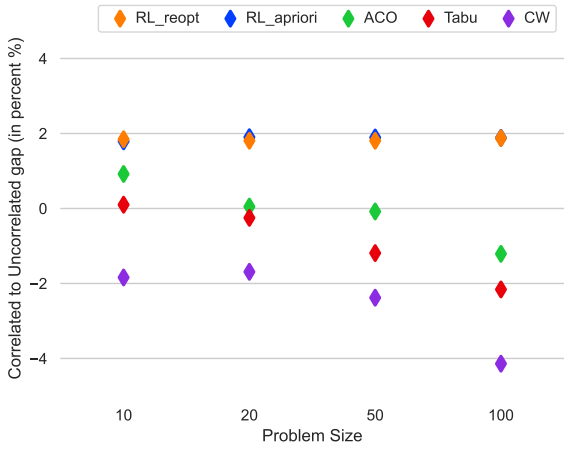


Figure 1: Percentage gap in travel costs of correlated variables setting to $\mathrm{A}, \mathrm{B}, \Gamma = 0.8, 0.2, 0.0$ to uncorrelated variables with $\mathrm{A}, \mathrm{B}, \Gamma = 0.8, 0.0, 0.2$.

1. https://github.com/Zangir/SVRP

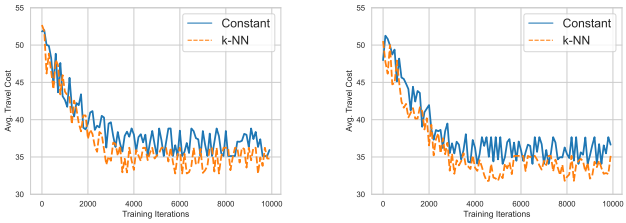
Table 1: The results obtained from various delivery types on the SVRP test dataset.

	10		20		50		100	
Delivery Type	Apr.	Reopt.	Apr.	Reopt.	Apr.	Reopt.	Apr.	Reopt.
Full delivery only	3.85	3.73	7.86	7.63	16.40	15.90	36.00	34.92
Partial delivery	**3.51**	**3.40**	**7.16**	**6.95**	**14.93**	**14.48**	**32.78**	**31.80**

Appendix D. Training Curves

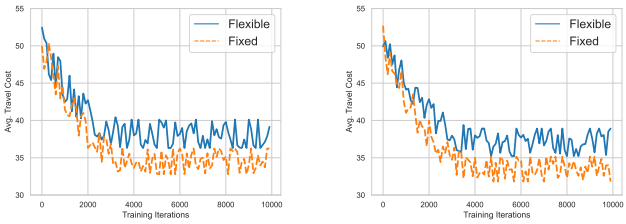
The training curves depicting the performance of different variable estimates are presented in Figure 2. The k-NN estimate outperforms the constant method, indicating its ability to learn and adapt to the problem dynamics more effectively.

Figure 3 displays the training curves comparing the performance of different customer positioning approaches. Figure 4 presents the training curves illustrating the performance of different delivery types.



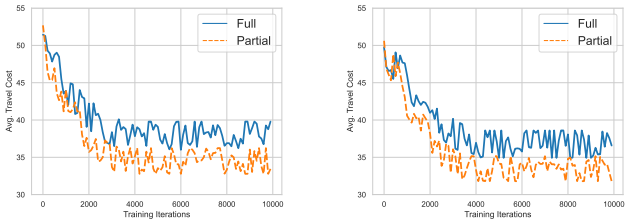
A priori Reoptimization

Figure 2: The travel cost during the training phase, using two types of variable estimates: Constant and k-NN. The results indicate robust learning in both cases.



A priori Reoptimization

Figure 3: The travel cost during the training phase, using two types of customer positions: fixed and flexible.



A priori Reoptimization

Figure 4: The travel cost during the training phase, using two types of delivery: full and partial.