Reinforcement Learning for Solving the Vehicle Routing Problem

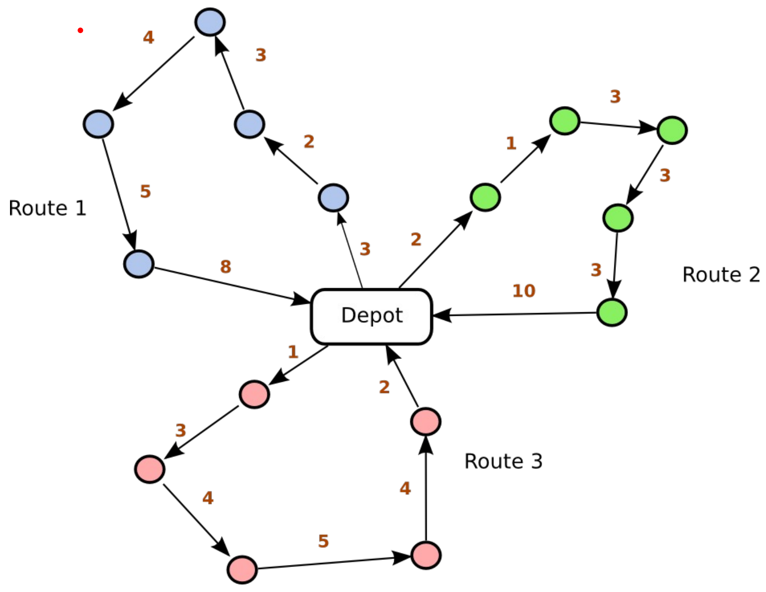
Authors: Mohammadreza Nazari, Afshin Oroojlooy, Martin Takac, Lawrence V. Snyder

Presenter: Abdullah Mamun

Date: Nov 23, 2022

A reinforcement learning approach to solve the vehicle routing problem.

- ❑ One depot
- ❑ One vehicle
- ❑ 10 to 100 customer nodes
- ❑ Dynamic demands



The Problem

- The depot has some goods
- The customers are located at different coordinates
- Customers demand different amount of goods
- The vehicle has to deliver to those customers and return to the depot
- The vehicle has a limited capacity, so occasionally it may require multiple trips to the depot to refill.

- The problem is NP Hard

Background

- This is a combinatorial optimization problem
- Somewhat similar to TSP = Traveling Salesman Problem
- Except the vehicle may have to refill
- Previously used methods on VRP and similar problems:
    - Classical and Modern Heuristics [Laporte et al. 2020]
    - Exact and Approximation Algorithms [Luong et al. 2020]
    - PDDL [Cheng et al. 2014]
    - Pointer Networks [Vinyals et al. 2015, Bello et al. 2016]
    - Attention mechanism [Bahdanau et al. 2015]

The model

At the start, Depot and the Vehicle are at the same point

- Set of inputs, $X = \{x^i, i = 1, \ldots, M\}$
- $x^i = \{x^i_t \doteq (s^i, d^i_t)\}$
- s^i = static features (**coordinates** of customer i)
- d^i_t = dynamic features (customer i’s **demands** at time *t*)

- Note that there is no *t* in the static features.

- X_t = *set of all input states at a fixed time t*

I need 7 gallons
I need 5 gallons

The model

At the start, Depot and the Vehicle are at the same point

- At every decoding time *t (t = 0, 1, …),*

  y_{t+1} *points to one available input from* X_t

Meaning, selects a customer to serve next.

And it continues until the terminating condition

: no more demand to satisfy

Then the vehicle returns to the depot.

Node visiting sequence: $Y = \{y_t, t = 0, \ldots T\}$

The model

Input length was M (number of customers)

Sequence length T

M may not be equal to T

As the vehicle may need to refill

The model’s policy

Suppose,

$Y_t = \{y_0, \ldots, y_t\}$

That is the sequence decoded so far (until time t)

$$P(Y \mid X_0) = \prod_{t=0}^{T} \pi\,(y_{t+1}, \mid X_t, Y_t) \quad (1)$$

$$X_{t+1} = f\,(y_{t+1}, X_t) \quad (2)$$

$$\pi\,(. \mid Y_t,\ X_t) = \mathit{softmax}\,(\, g(h_t\,, X_t)) \quad (3)$$

Attention mechanism is used to calculate the components of the right side of (1).

h_t = the hidden state representation RNN decoder summarizing steps $y_0, \ldots, y_t$.

g = affine function that outputs an input-sized vector.

f = state transition function. (will be explained soon)

Now let's explain the transition function

- $X_{t+1} = f\ (y_{t+1}, X_t)\ \ (2)$
- *We have the current states (coordinates and demands)*
- *We have the next hop (where to go)*

- *Now, we need the next states (coordinates will be the same, demands may update)*

$$d^i_{t+1} = \max(0, d^i_t - l_t), \quad d^k_{t+1} = d^k_t \text{ for } k \neq i, \text{and} \quad l_{t+1} = \max(0, l_t - d^i_t) \qquad (7)$$

- $l = load$
- *So, remaining demand = demand – load served*
- *and remaining load = load – demand served*

Masking/pruning for faster training

Nodes not to visit:

1. nodes with zero demand
2. No customer visited, if the vehicle’s remaining load is exactly 0
3. the customers whose demands are greater than the current vehicle load

Exception of 3: if split delivery is allowed, those customers can be served in more than one visits.

In the experimental setup of this paper, split delivery was not allowed

Experimental methods

Problems:

VRP10, VRP20, VRP50, VRP100

i.e. Number of customers varied from 10 to 100

Vehicle capacity varied 20, 30, 40, 50

Customer demand is discrete from 1 to 9

***Proposed algorithms**:*

Greedy RL (at any step, the node with highest probability is selected)

Beam search RL (keeps track of the most probable paths and chooses the one with minimum tour length) (varying beam width = 5 and 10)

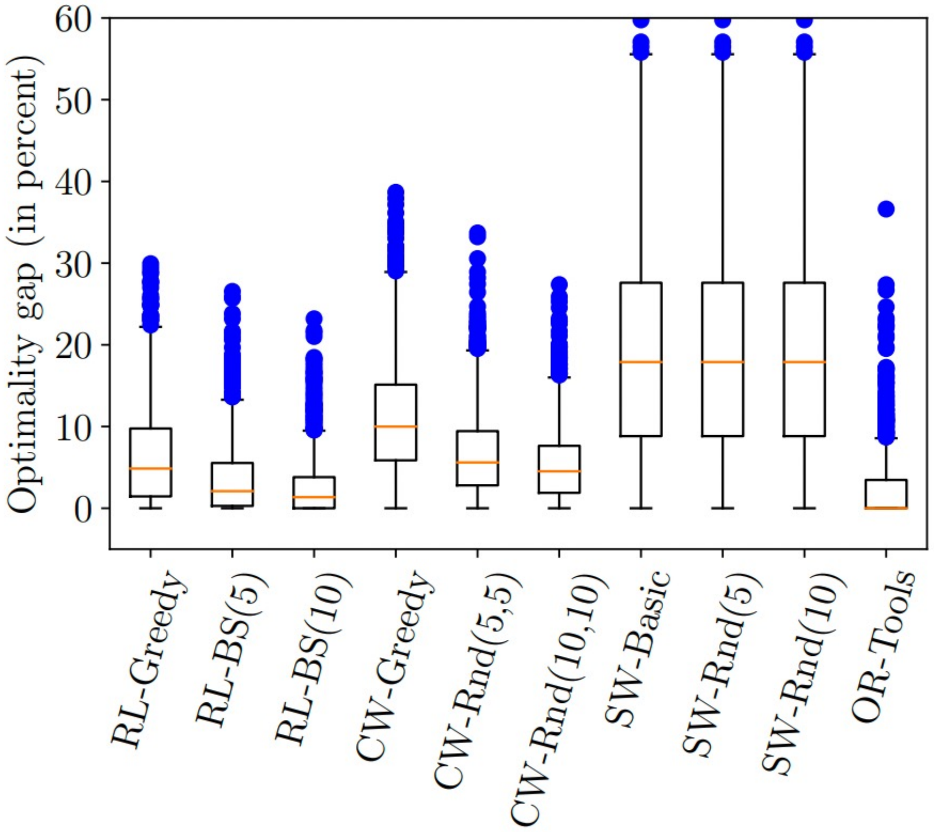
***Benchmark algorithms**:*

Clarke-Wright savings heuristics

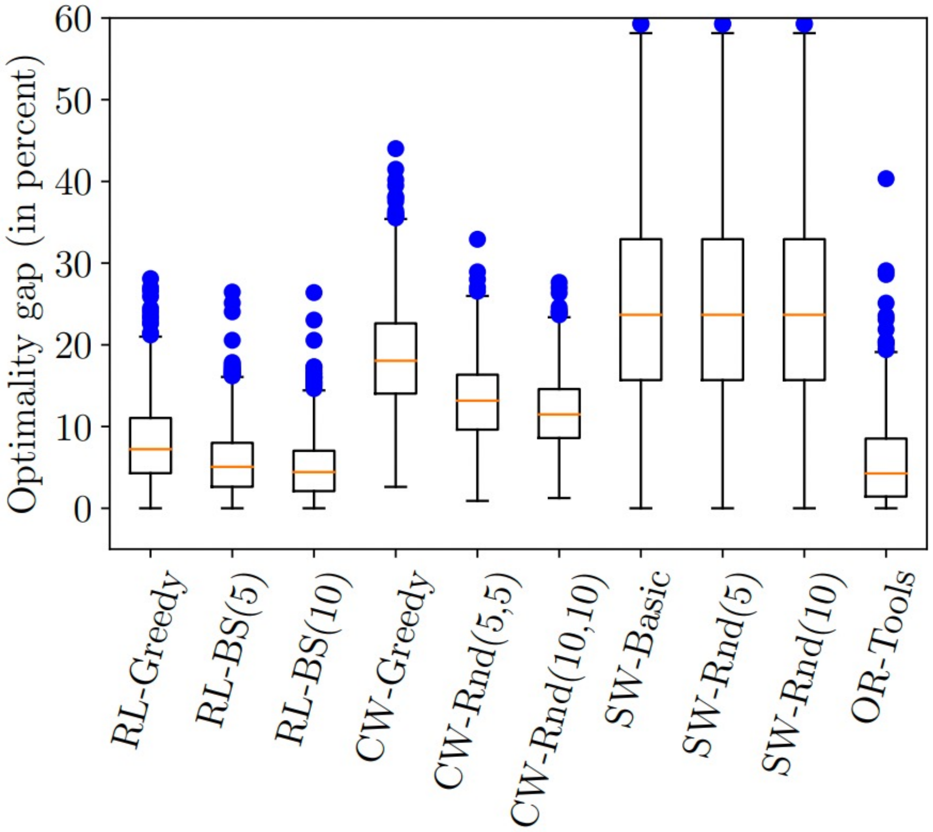
Sweep heuristic

OR-Tools (Google's optimization tools)

Results



(a) Comparison for VRP10

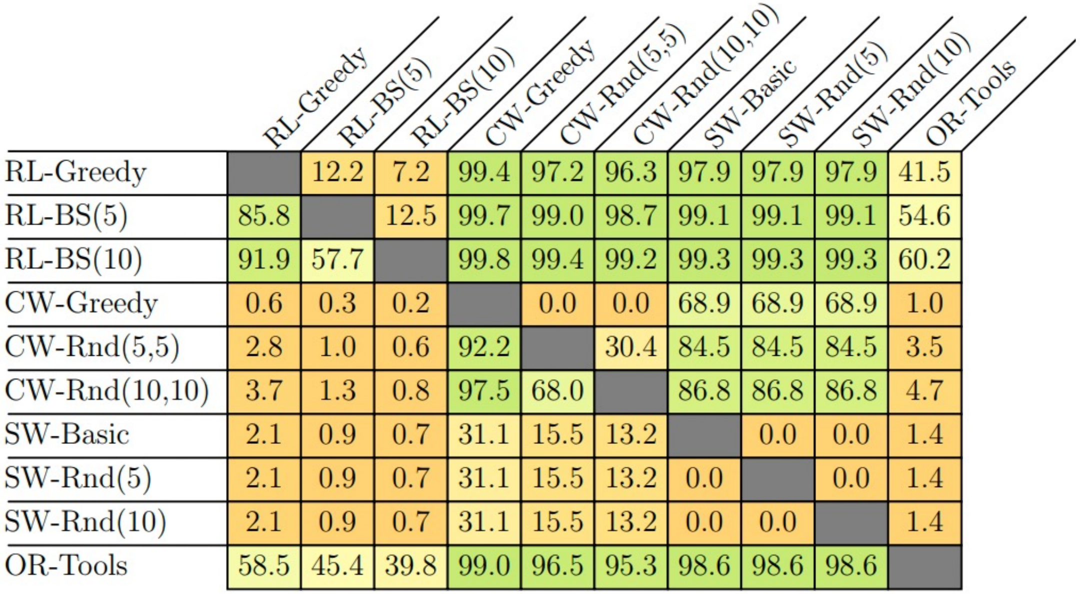


(b) Comparison for VRP20

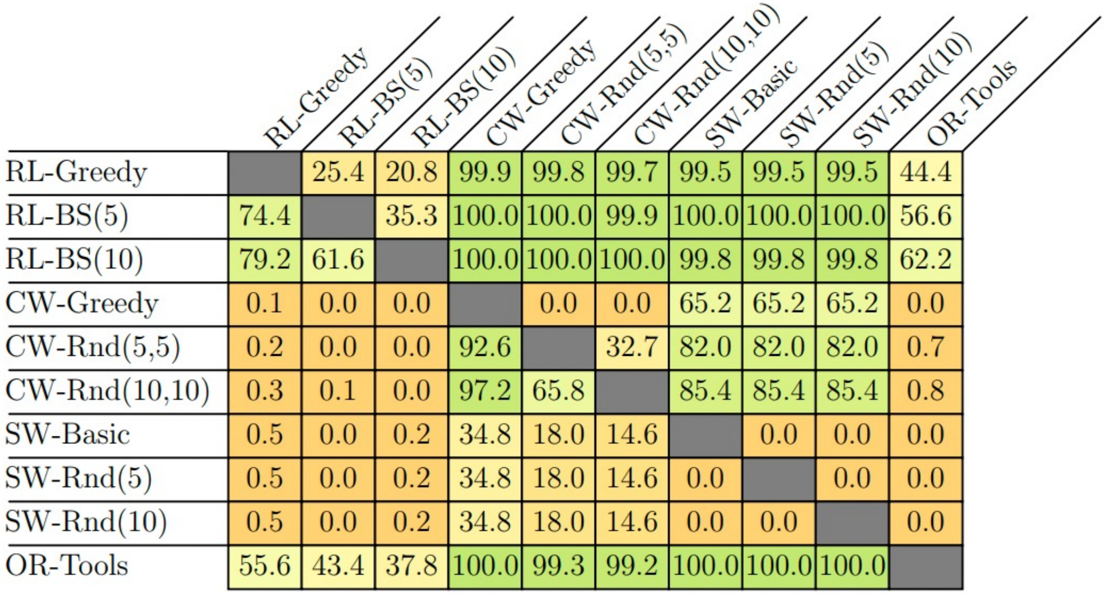
Results

(a) Comparison for VRP10

(b) Comparison for VRP20



	RL-Greedy	RL-BS(5)	RL-BS(10)	CW-Greedy	CW-Rnd(5,5)	CW-Rnd(10,10)	SW-Basic	SW-Rnd(5)	SW-Rnd(10)	OR-Tools
RL-Greedy		12.2	7.2	99.4	97.2	96.3	97.9	97.9	97.9	41.5
RL-BS(5)	85.8		12.5	99.7	99.0	98.7	99.1	99.1	99.1	54.6
RL-BS(10)	91.9	57.7		99.8	99.4	99.2	99.3	99.3	99.3	60.2
CW-Greedy	0.6	0.3	0.2		0.0	0.0	68.9	68.9	68.9	1.0
CW-Rnd(5,5)	2.8	1.0	0.6	92.2		30.4	84.5	84.5	84.5	3.5
CW-Rnd(10,10)	3.7	1.3	0.8	97.5	68.0		86.8	86.8	86.8	4.7
SW-Basic	2.1	0.9	0.7	31.1	15.5	13.2		0.0	0.0	1.4
SW-Rnd(5)	2.1	0.9	0.7	31.1	15.5	13.2	0.0		0.0	1.4
SW-Rnd(10)	2.1	0.9	0.7	31.1	15.5	13.2	0.0	0.0		1.4
OR-Tools	58.5	45.4	39.8	99.0	96.5	95.3	98.6	98.6	98.6	



	RL-Greedy	RL-BS(5)	RL-BS(10)	CW-Greedy	CW-Rnd(5,5)	CW-Rnd(10,10)	SW-Basic	SW-Rnd(5)	SW-Rnd(10)	OR-Tools
RL-Greedy		25.4	20.8	99.9	99.8	99.7	99.5	99.5	99.5	44.4
RL-BS(5)	74.4		35.3	100.0	100.0	99.9	100.0	100.0	100.0	56.6
RL-BS(10)	79.2	61.6		100.0	100.0	100.0	99.8	99.8	99.8	62.2
CW-Greedy	0.1	0.0	0.0		0.0	0.0	65.2	65.2	65.2	0.0
CW-Rnd(5,5)	0.2	0.0	0.0	92.6		32.7	82.0	82.0	82.0	0.7
CW-Rnd(10,10)	0.3	0.1	0.0	97.2	65.8		85.4	85.4	85.4	0.8
SW-Basic	0.5	0.0	0.2	34.8	18.0	14.6		0.0	0.0	0.0
SW-Rnd(5)	0.5	0.0	0.2	34.8	18.0	14.6	0.0		0.0	0.0
SW-Rnd(10)	0.5	0.0	0.2	34.8	18.0	14.6	0.0	0.0		0.0
OR-Tools	55.6	43.4	37.8	100.0	99.3	99.2	100.0	100.0	100.0	

(c) Comparison for VRP50

(d) Comparison for VRP100

Computation time and Generalization

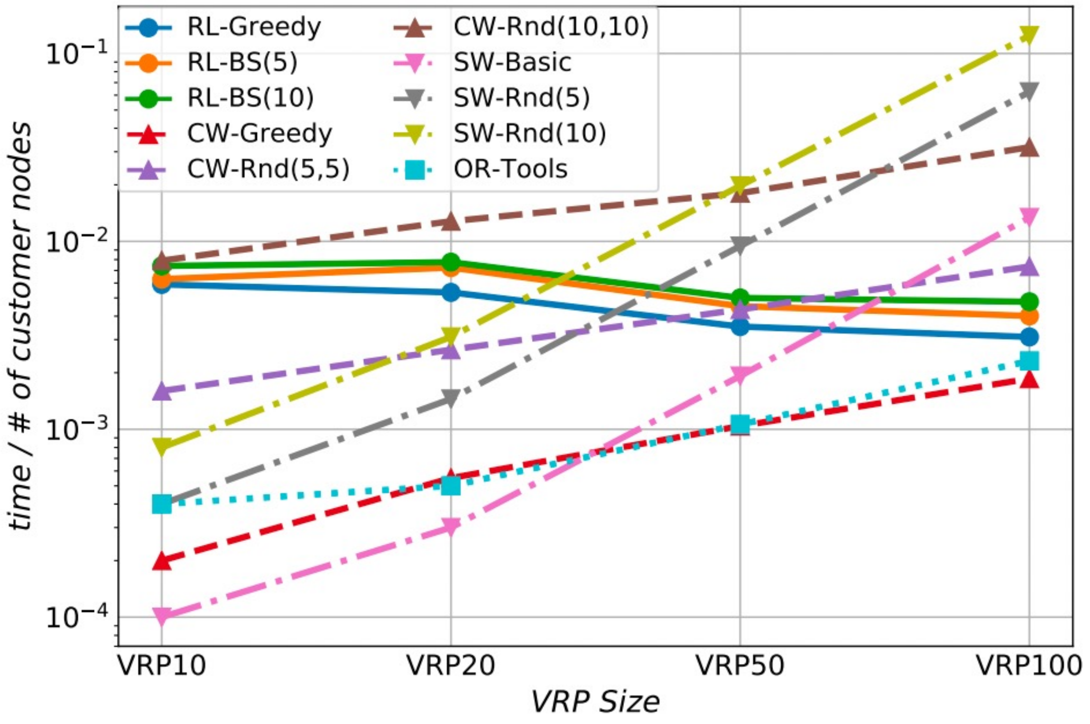


Figure 4: Ratio of solution time to the number of customer nodes using different algorithms.

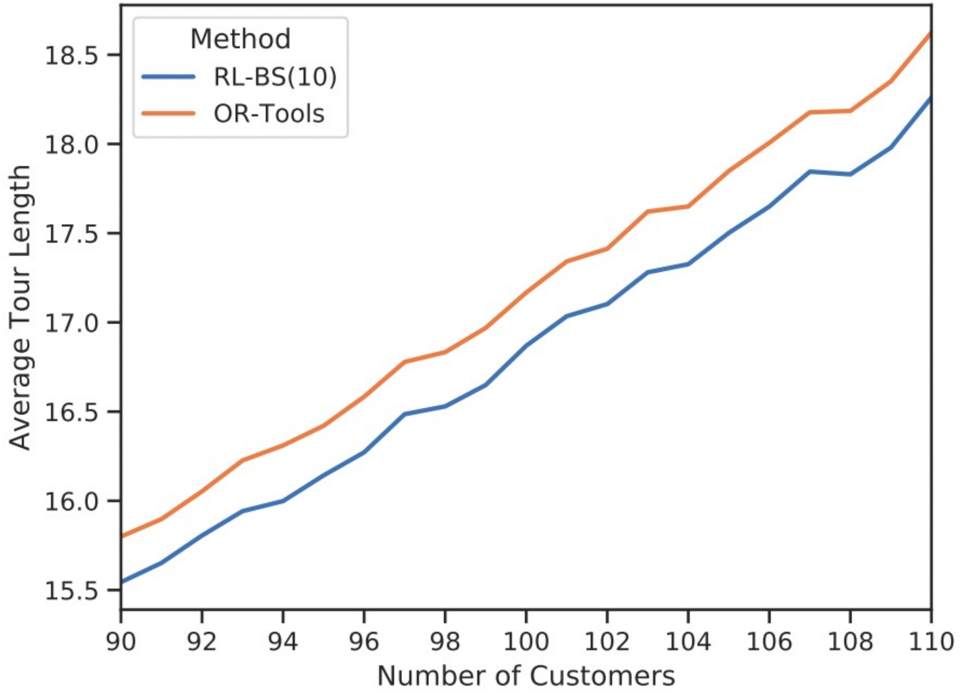


Figure 5: Trained for VRP100 and tested for VRP90-VRP110.

Limitations and Future works

- ❑ *Only one vehicle was used. In future, multiple vehicle system may be explored with these algorithms*
- ❑ *This current work covers only one-depot system.*
- ❑ *More constraints may be applied, e.g. multiple visits to the same customer is not allowed. (will have to send multiple vehicles for very large orders,but need to deliver at the same time for convenience).*
- ❑ *Split deliveries can be allowed.*

https://abdullah-mamun.com
a.mamun@asu.edu