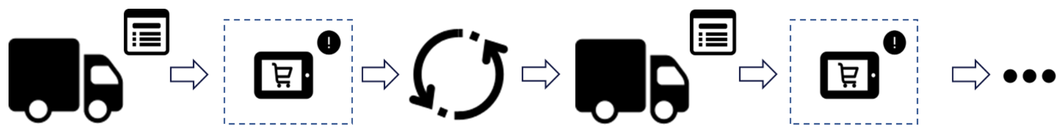
Fujitsu-SMU Urban Computing & Engineering Corp. Lab

Deep Reinforcement Learning Approach to Solve Dynamic Vehicle Routing Problem with Stochastic Customers

Waldy JOE, Hoong Chuin LAU
waldy.joe.2018@phdcs.smu.edu.sg, hclau@smu.edu.sg

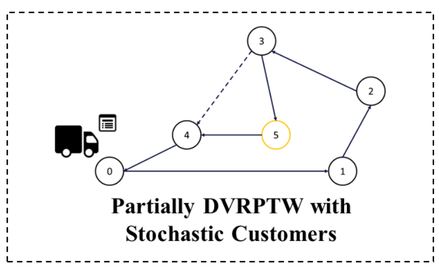
MOTIVATION

In real-world urban logistics operations, **changes to the routes and tasks** occur in response to dynamic events. To ensure customers' demands are met, logistics service providers need to **dynamically respond** to these changes quickly. Thus, **a quick decision support solution** is required to address this problem.

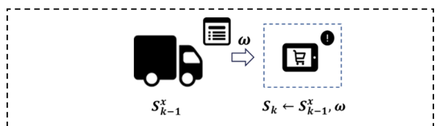


PROBLEM

We consider a **Dynamic Vehicle Routing Problem (DVRP)** with time windows and both known and stochastic customers. We formulate this problem as a **route-based MDP** where state, action space and reward computation take into account the set of planned routes.



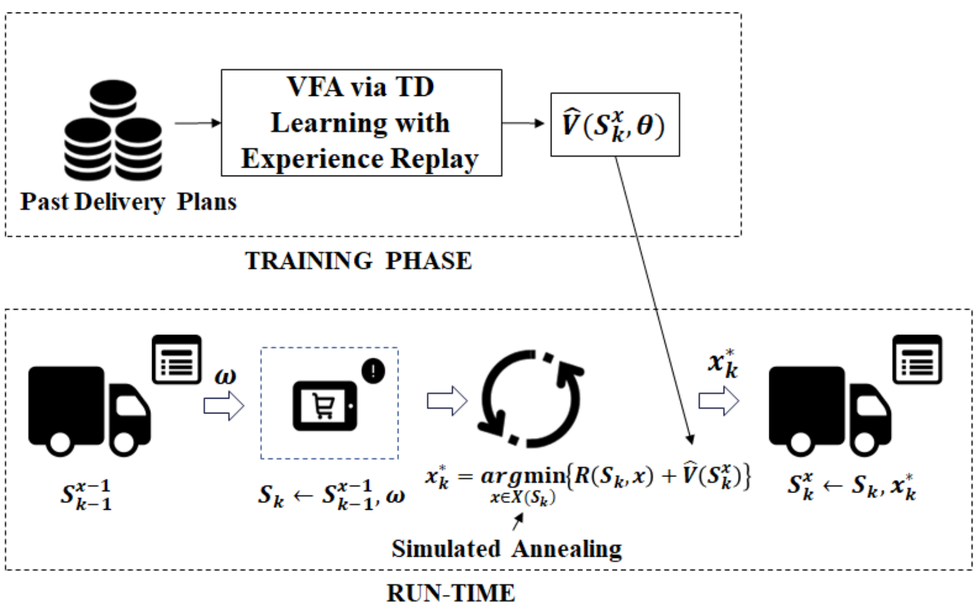
Partially DVRPTW with Stochastic Customers



Route-based MDP

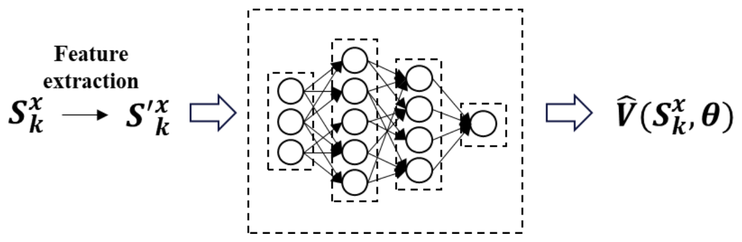
SOLUTION APPROACH

DEEP REINFORCEMENT LEARNING WITH SIMULATED ANNEALING (DRLSA)



Value Function Approximation Step

- Approximate value function using neural network.
- Use TD learning with experience replay algorithm to learn the parameter $\boldsymbol{\theta}$.



State Representation

- Use the following state representation to reduce state space, exploit problem structure and extract distinctive features.

$$S'(k) = \begin{bmatrix} \text{Time} \\ \text{Cost of remaining route for vehicle 1} \\ \text{Cost of remaining route for vehicle 2} \\ \text{Penalty cost for time violation for vehicle 1} \\ \text{Penalty cost for time violation for vehicle 2} \end{bmatrix}$$

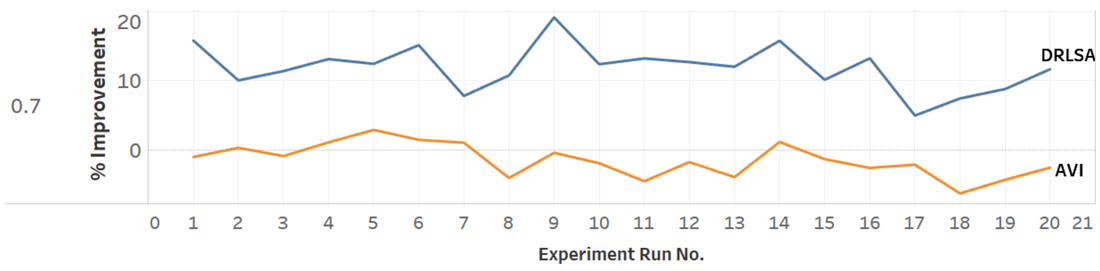
**Assume a problem setting with 2 vehicles*

Routing Optimization Step

- Use the trained value function to guide SA in finding optimal re-routing quickly and effectively.

EXPERIMENTAL RESULTS

- Evaluated DRLSA against **Approximate Value Iteration (AVI)** (offline method) and **Multiple Scenario Approach (MSA)** (online method) based on % improvement over myopic.



- DRLSA achieved on average **12% improvement,** outperforming AVI on problems with Degree of Dynamism (DoD) above 0.5.
- However, DRSLA performed **poorly** when DoD **drops below 0.5**.
- MSA only managed to achieve on **average 2% improvement** and it requires **larger sample size and longer computational time** to outperform DRLSA.
- DRLSA was also able to achieve **consistent improvement** with **increasing DoDs** and **larger problem size** (tested up to 4 vehicles and 40 orders).

CONCLUSION

Our approach is **fairly generic** and can be applied to tackle other dynamic planning and scheduling problems.
Our approach is oblivious to probability distributions of demand uncertainty and only require **a relatively small training set** based on historical data.

NATIONAL RESEARCH FOUNDATION
PRIME MINISTER'S OFFICE
SINGAPORE

This research project is partially supported by National Research Foundation Singapore under its Corp Lab @ University scheme and Fujitsu Limited as part of the A*STAR-Fujitsu-SMU Urban Computing and Engineering Centre of Excellence.