Deep Reinforcement Learning for Truck Dispatching at Marine Container Terminal

Introduction
Maritime transportation has been expanding dramatically in the last few decades. As an essential operation in a container terminal, truck dispatching greatly affect the port throughput. Truck dispatching happens at vessel’s arrival. A truck fleet take charge of load and unload containers between quay cranes (QC) and yard cranes (YC). When a truck finish a task, the scheduler need to select a new task to it based on the states of all quay cranes. The objective is the total waiting time of each task because a vessel would stay short time in a berth if all the quay crane keep busy. Such problem could be solved by some methods such as integer programming or meta-heuristic. However, in non-deterministic formulated environment, these methods may lose their competitiveness. In this project, we proposed a deep reinforcement learning based dispatching method that trained on the environment with different type of uncertainties. The experimental results demonstrated its superiority and the potential for practical applications.

Container Terminal Simulation
A port terminal simulation based on the Anylogic platform is developed as the training environment of the reinforcement learning agent. The simulator is also support for other port-related problem such as yard crane scheduling, container relocation or container space allocation problem.

The uncertainty in the environment come from three part:
1. The truck travel speed at different area of the container terminal.
2. The service time of quay crane or yard crane operation.
3. Degree of yard congestion which depends on the number of trucks at a yard and their relative location.

Methodology
Reinforcement learning solves the sequential decision-making problems that formalized as the Markov Decision Process (MDP), where the action affects the current reward and subsequent state and rewards in the future. In last decade, reinforcement learning has continuously attracted enormous attention from the operation research communities.

The following features are considered as the state for RL agent at each decision point:
1. Amount of remain task of each quay crane.
2. Distance between the current truck and each quay crane.
3. Total number of trucks working for each quay crane.
4. Queue length of each quay crane at present.
5. Quay crane type: LOAD or DISPATCH.

We used the well-known REINFORCE algorithm to update the gradient

$$\nabla \mu J(\theta) = E_{\pi(\sigma|s)} \left( L(\pi|s) - b(s) \right) \nabla \log \pi(\sigma|s)
$$

Model
A similar pointer network framework is used as the actor in the reinforcement learning training. Since this is an on-line problem with uncertainties, only one step decision need to be made at each time step.

Experimental Result
A manually designed dispatching heuristic is deployed and considered as the benchmark of the reinforcement learning method. The heuristic is based on supply-demand difference in a pre-defined time window of each QC. The experiment result demonstrate the proposed RL-based method outperformed manually designed heuristic in different scenarios.

<table>
<thead>
<tr>
<th>Truck Fleet</th>
<th>Heuristic Method</th>
<th>RL Method</th>
<th>Improve</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 trucks</td>
<td>24390</td>
<td>21153</td>
<td>13.3%</td>
</tr>
<tr>
<td>80 trucks</td>
<td>16716</td>
<td>14958</td>
<td>10.5%</td>
</tr>
<tr>
<td>100 trucks</td>
<td>13658</td>
<td>12730</td>
<td>6.8%</td>
</tr>
</tbody>
</table>

Table 1: Performance of RL Method Under Different Size Truck Fleet