The Robustness of the GNN approach in application of different data-generation methods

Jaša Lemut

June 21, 2026

Abstract

Sea cargo transportation relies on the efficiency of ports and how effectively ships are allocated to berths. This study evaluates the robustness of the Graph Neural Network proposed by Moya et al. [7] for the Berth Allocation Problem. It uses real world scenarios and known cases to scrutinize the agent in comparison to common heuristics and a baseline optimal solution. While the agent created schedules that were in general on par with the baseline heuristics, the agent performed particularly poorly in changes to the handling times variable.

1 Introduction

Ships all around the world use ports to load and unload cargo. At any given moment a port is not going to have enough berths for all the ships to be accommodated at any given moment. The current system works on a scheduling system, such that a ship can request a berth for a specific arrival time, then based on the congestion and priority at the time, the ship can be placed into a berth after some waiting time. While there exist many solutions, the focus of this research is the Graph Neural Network approach mentioned in Moya et al. [7] that uses reinforcement learning to find a near optimal arrangement for this Dynamic problem. As of yet, this method has only been tested on environments modeled for training, which is not thorough enough for usage in the logistics industry.

As such this research aims to expand on this testing phase of the aforementioned algorithm, aiming to use rare unlikely scenarios and more accurate distributions that model closer to reality. Much like how the original agent was tested, here the agent will be tested against an optimal solution (within a timeout period) to check for optimality and also tested against 4 known heuristics for the Berth Allocation Problem: FCFS, SPT, Priority, and WTSP.

The paper will be structured in the following. The next section provides the related work and background information about the task. In section 3, we see a breakdown of the methodology of the paper and what testing precautions were taken. In section 4, we evaluate the agent and how it has performed compared to the aforementioned heuristics and optimal solutions.

2 Related Work

This work is based on the Graph Neural Network approach mentioned in Moya et al. [7] which is an agent developed to solve the berth allocation problem. The GNN uses random variable generation for the evaluation process. The three random variables it introduces are the priority variable, the Handling times variable, and the Arrival times variable.

The priority variable is a fictitious variable and does not have a real world equivalent unit. Its affect on the evaluation function is also trivial, and hence not the focus of this paper.

The two main variables to be scrutinized are the (inter-)arrival times variable and the handling times variable. The arrival times variable is generated using an exponential distribution to calculate the difference between two consecutive vessels. As such no order for the vessels is required as they are pre-generated in an arbitrary order as they have no chronological dependencies.

The handling times variable is calculated individually for each berth-vessel pair as a uniform distribution [10,210]. Figure 1 is an example of a generated instance with 10 vessels and five berths.

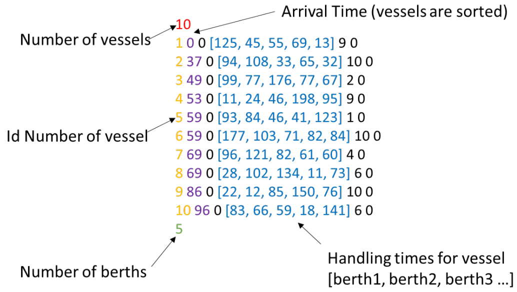

Figure 1: Example instance of the berth allocation problem

3 Methodology

3.1 Berth Specific Handling Times

Uniform handling times across different berths are not the best representation of how the real world harbours work. Faster machinery, more organised personnel, berth infrastructure, among others factors, could all affect the average handling of a specific berth. Therefore it would also be unrealistic to assume all berths have the same average handling times or could be represented by the same distribution.

Rather than designating one berth as the most efficient, and using the same distribution for all other berths, which would once again be unrealistic as all other berths are once again identical. Berths were assigned increasingly higher handling times such that a range of berth efficiencies was generated.

There was two ways this distribution was approached in this paper (as shown in Figure 1). Method one splits the total range into n equally sized subranges, where n is the number of berths. This ensures that berths cannot be interchangeable in order of efficiency, and each berth has a strictly higher efficiency than the successive one.

The other method devised for this was to have increasingly higher averages for handling times for each berth, and then having a normal distribution centred at the that average. This does allow for some overlap given that normal distributions have a non-zero possibility for every real number. The variances are set

Unlike the definition of robustness in Kolley et al. [5], where the robustness of the algorithm is more to do with uncertainties of arrival times, this method tests the robustness of the agent by requiring it to make decisions beyond simply matching vessels to the most suitable berth. the robustness is tested after optimization. In the case that a berth is sufficiently slow, it may be preferable to skip it entirely rather than attempting to distribute all vessels across all available berths.

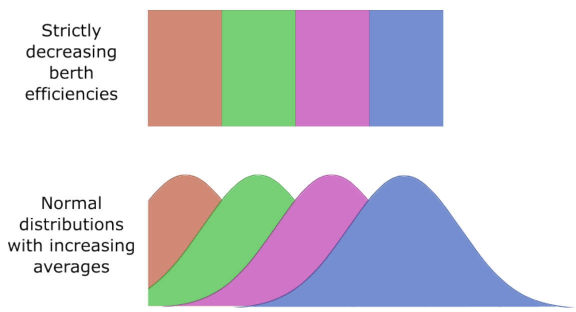

Figure 2: Two different approaches to Berth Specific Handling Times

3.2 Berth-Vessel Pair Handling Times

A berth being assigned the same handling time distribution across all vessels is also not the best representation of reality. Vessels vary in size, cargo, and docking requirements, all affect the handling time of a vessel for a berth. So, to assume that there is no relation between vessel and berth in handling times would be unrealistic.

For each vessel we arbitrarily pick a berth that is preferred. All preferred berths have a uniform distribution between [10,105] and non-preferred berths have a uniform distribution from [106,210]. To maintain modularity in the distribution and allow for independent changing of the underlying distribution, this setup allows for modification to either of the two distributions. This means that this can be combined with either the Erlang distribution tests and/or the Berth Specific handling times tests.

In the case that a berth were chosen as the preferred berth for multiple vessels, results would overlap with the Berth Specific Handling Times results. In prevention of this, vessels are evenly distributed across the berths in modulo of the vessel numbers.

This distribution is meant to test the robustness of the GNN agent by generating instances that prefer a very compact solution. These instances require the agent to identify the preferred berth for the vessel and also to preserve these preferred berth-vessel pairs as it generates the schedule.

3.3 Erlang-2 Distribution

The third method of instance generation is meant to represent the average scenario. For an individual berth, the handling time taken for each berth is unrealistic to be uniform, as that would mean that taking little to no time is as common as taking an average amount of time. Following empirical data of the Port of Kobe in Japan, the Erlang-2 distribution was found to be the most accurate to the real handling times[4].

The Erlang-2 distribution is a variation of the gamma distribution where the shape parameter is specifically two. This distribution appears to be a positively skewed normal distribution.

This method will test the robustness of the agent by providing instances that are closest to reality if all berths and vessels are taken equally. While it does not generate the edge cases that might normally be used to strain a scheduling algorithm, it establishes a realistic baseline.

3.4 Higher Congestion

For the first method of testing the robustness of the agent through changing the Arrival Times Variable is to increase vessel congestion. While this has been shown to have no significant change in the operational costs of the scheduling[3], the aim of increasing the congestion is to increase the number of possible collision points between two vessels for a berth.

It should be noted that congestion here is meant to define the average amount of time between two vessels' arrivals. While this does justify its places in the tests that work with the Arrival times variable, it does not aim to change the shape of the underlying distribution, but rather only compress the distribution.

A vessel could have multiple "reasonable" berths to be allocated to, so its scheduling will depend on the handling times of all the other berths in addition to its own. This congestion test will also reduce the percentage of the search space that might offer a good schedule, as there could be many ways to make a bad schedule, but far less to make a good one.[2]

3.5 Poisson Distribution

The arrival times variable is most commonly near zero and hence is very small compared to the handling times variable. This test aims to change that by working the variable into a Poisson distribution. This distribution would offer instances differing substantially from the baseline training data. This new instance generation environment offers a way to test if the agent performs well in instances where the non-zero arrival times are far more likely.

Heuristics like FCFS often have a fault in not accounting for future arrivals, in that a preferred berth might not be available, but in best case the vessel waits for it to be available. This distribution would increase situations where this fault could be more pronouced since it decreases situations where two vessels that are competing for a berth arrive simultaneously, as that sub-problem is trivial.

3.6 Clustered Distribution

As with the third variation of the handling times variable, this distribution aims to increase realism. In practise, evenly spread over time. External factors such as poor weather conditions or operational disruptions can affect the arrival times of multiple vessels, creating delaying in the constant line of vessels. This creates clusters of shortly spaced arrivals once the conditions improve or the problems are averted.

There burst arrivals increase the average inter-arrival time since vessels were disrupted from their normal inter-arrival times. This also means that the congestion is on average lower than the training data. This distribution helps capture a more realistic scenario with external factors.

4 Results

The primary objective of the experiments conducted was to look for scenarios in which the agent might have trouble with the allocations. They were designed to be realistic instances that a port might see in regular work-days or to stress test the agent in scenarios that might require different heuristics. Figure 3 shows an overview of the results.

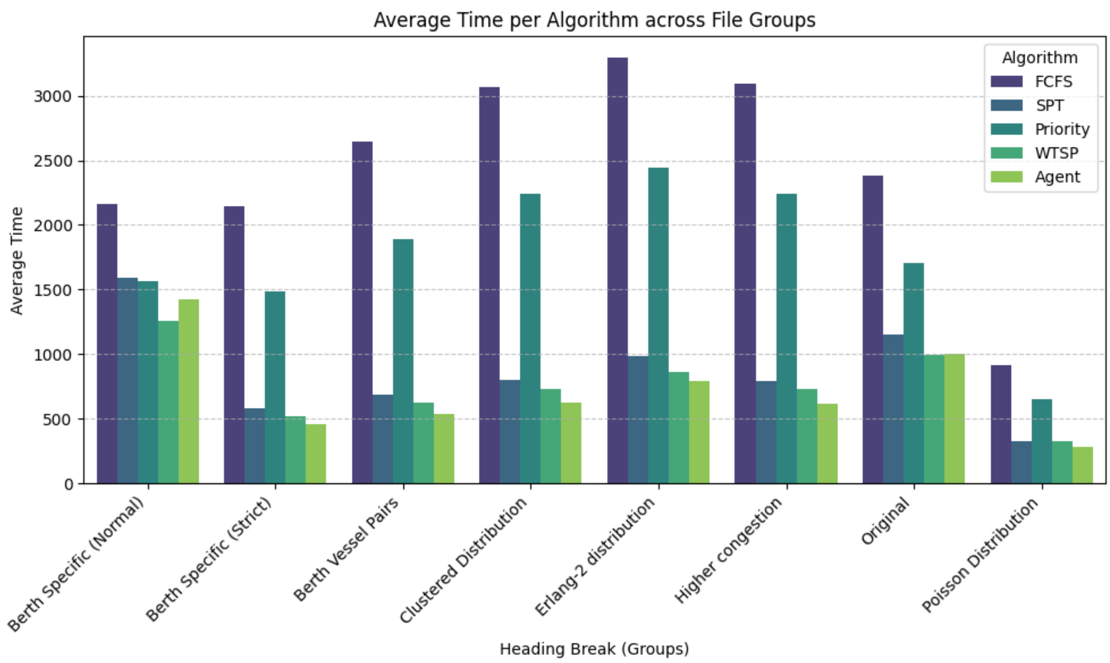

Figure 3: Average total schedule cost for each test

The distribution was designed to force the agent to understand that some berths are more efficient than others. This test was especially insightful into the shortcomings of the agent. The agent performed poorly, only managing to create a better schedule than all in less than a half the test cases in Normally distributed and none of the cases in Strictly greater.

The table below shows a breakdown of how well each heuristic did compared to the agent for both methods of generating instances for this test.

While in general it did poor, earlier instances of Berth Specific handling time show that it understood the importance of waiting with ships rather than trying to utilize the free berths. Figure 4 shows a very good example of this on a smaller instance.

	Cost Ratio	Win Ratio	Cost Ratio	Win Ratio
	Normally Distributed		Strictly Greater	
FCFS	2.240	100%	1.450	100%
SPT	1.141	92.6%	1.105	88.9%
Priority	1.614	100%	1.064	70.4%
WSTP	0.989	51.9%	0.877	0.0%

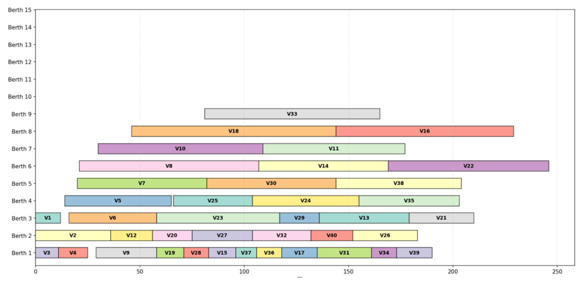

Figure 4: Instance of Berth Specific Handling Time (with normal distribution) with the schedule made by the agent

The agent also struggled with other variations on the handling times variable, making a usual fastest schedule around 80% of the time. This does show how the over reliance on uniform distribution in the agent affects its results poorly.

On the other hand, the agent did quite well in instances where the arrival times variable was altered. In the Poisson distribution, there was only one instance where the agent was outperformed by a heuristic. While in the Clustered distribution, the agent did have a lower score against the poorer heuristics like FCFS (getting a better scehdule 88.9% of the time), it still had an average schedule cost 0.87 times that of WSTP.

5 Conclusion and Further Work

Since this is a scheduling problem, it relies very heavily on the fact that small changes in the local scheduling affects the entire global schedule.[9]. The agent has a heavy reliance on the handling times, as can be seen by the poor results in all the handling times variables. It most likely over fits on the uniform distribution, accepting values when they are low, when comparatively they could very well be high if a berth is preferred.

For future works I suggest that dimensionality reduction be used on the agents final layer, to see what heuristics, if any, have been found and used in the instance. An example of this would be the learned heuristic that was found in the Berth Specific Handling Times instances, as this heuristic could be fully verified through the going into the last layer of the agents calculations.

The agent should also be trained on more rigorous data, much like the ones found here, as the overfitting on the uniform data shows poorly. This would also open the door to more specialized research, as when the berths and vessels do not follow uniform and exponential distributions, relationships between specific berths and vessels can be identified more easily.

References

[1] Peter Cheeseman. “Where the really hard problems are”. In: *International Joint Conference on Artificial Intelligence*. 1991.

[2] Jam Dai et al. “Berth allocation planning optimization in container terminals”. In: *Supply chain analysis: a handbook on the interaction of information, system and optimization*. Springer, 2008, pp. 69–104.

[3] Mihalis Golias et al. "The Berth-Scheduling Problem: Maximizing Berth Productivity and Minimizing Fuel Consumption and Emissions Production". In: *Transportation Research Record* 2166.1 (2010), pp. 20–27. DOI: 10.3141/2166-03.

[4] Akio Imai and Etsuko Nishimura. "A Berth Allocation Problem in the Public Berth System with Consideration of the Starting Time of Planning Horizon". In: *Tsuchi ki-kei-ga-gaku Ken Kiwamu-ron bun-shū [Civil engineering planning research and paper collection]* 15 (1998).

[5] Lorenz Kolley et al. "Robust berth scheduling using machine learning for vessel arrival time prediction". In: *Flexible Services and Manufacturing Journal* 15 (2022), pp. 29–69.

[6] Lorenz Kolley et al. "Robust berth scheduling using machine learning for vessel arrival time prediction". In: *Flexible services and manufacturing journal* 35.1 (2023), pp. 29–69.

[7] Carlos March Moya et al. "A Reinforcement Learning Approach for the Dynamic Berth Allocation Problem". In: *LOGMS26* (2026).

[8] Takahiro Ogura, Teppei Inoue, and Naoshi Uchihira. "Prediction of arrival time of vessels considering future weather conditions". In: *Applied Sciences* 11.10 (2021), p. 4410.

[9] Toby Walsh et al. "Search in a small world". In: *Ijcai*. Vol. 99. 1999, pp. 1172–1177.

[10] Ya Xu, Qiushuang Chen, and Xiongwen Quan. "Robust berth scheduling with uncertain vessel delay and handling time". In: *Annals of operations research* 192.1 (2012), pp. 123–140.