Optimal Evacuation Route Prediction in Fpso Based on Deep Q-Network

Seokyoung Hong, a Kyojin Jang, a Jiheon Lee, a Hyungjoon Yoon, a Hyungtae Cho, b Il Moon a*

a Department of Chemical and Biomolecular Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea
b Green Materials & Processes Group, Korea Institute of Industrial Technology, 55 Jongga-ro, Jung-gu, Ulsan, 44413, Republic of Korea
pselab@yonsei.ac.kr

Abstract

As a major production facility for oil and gas exploration, a dangerous working environment can cause serious accidents in FPSOs. The hull of the FPSO can be severely damaged or flooded, and in such cases, it is very difficult for workers to escape due to the large and complex structure of the FPSO. Even in an emergency situation, rational reasoning is blurred, and objective and prompt judgment cannot be made. It is necessary to provide quick and efficient evacuation guidance in case of emergency in order to prevent casualties. In this study, deep Q-network, a method of reinforcement learning, was applied to the optimal path prediction model to calculate the optimal evacuation route for workers in case of an accident. Deep Q-network can be applied to models with large and complex structure, and the environment of the model consists of 4 parts: deck, accommodation, backward, and frontward. The agent receives penalties for every move and is rewarded when it arrives at one of the four exits on the deck. As a result, the average number of movements is less than 30 when escaping from all locations in the environment consisting of 621 grids. This study contributes to provide an optimal escape route that is fast and safe in a changing environment. Through the analysis of the results of this study, it is expected that the safety inspection on the FPSO can be carried out and it will help the safe FPSO design.

Keywords: Deep Q-Network, Optimal Route Prediction, 3D Environment.

1. Introduction

Floating Production, Storage and Off-loading (FPSO) is a primary method of oil and gas processing and storage for deep-water fields. The number of FPSOs in operation or available for deployment has grown by 33% over the past 10 years. However, the FPSOs are still prone to accidents with a high probability of harsh working. Once serious accident occurs, FPSO may be seriously destroyed and even flooded. In these instances, it is very difficult for workers to evacuate out of the FPSO due to the large scale and complex structure of the FPSO. To avoid causalities, offering workers the effective escape route is essential. Evacuation route analysis methods can be roughly divided into simulation and optimization. Optimization approaches in evacuation planning are adopted by numerous
previous researches. In mathematical modeling, linear programming and dynamic programming models are widely used. Network flow algorithm is utilized to model crowd movement with blocking effect (Luh et al., 2012). As a meta-heuristic approach, ant colony optimization algorithm is implemented to solve multi-objective evacuation routing problem (Fang et al., 2011) and has evolved into quantum ant colony algorithm (Liu et al., 2016). Genetic algorithm (GA) is also a popular heuristic solution method. Multi-criteria problem (Goerigk et al., 2014) and level of service design (Li et al., 2019) for efficient evacuation planning are solved by GA. Compared to mathematical modeling, using heuristic based algorithm reduces the computational cost but does not guarantee exact solution.

To solve this, Q-learning algorithm has been attempted to find a safe and short path (Su et al., 2011). Q-learning is very suitable for solving discrete motion sequence decision problem like path planning, but it is no longer able to approximate the value function as the number of states and actions increases. Most of maritime evacuation researches have been focused on passenger ship and little attention has been paid to the FPSO evacuation. In case of FPSO, different from ships or buildings, the structure of workspace is complex, passages are narrow, and the operating equipment are concentrated. There is a still lack of model to solve such a large-scale problem.

Thus, in this paper, we introduced deep Q-network (DQN) (Mnih et al., 2015) algorithm that combines deep learning with reinforcement learning to optimize the evacuation route in FPSO. We modeled the environment referring to the actual FPSO design (Figure 1). The agent is trained to find the shortest path without stepping back or passing the dangerous area. We compare the results of 2 scenarios and have a discussion about the evacuation system.

2. Problem statement and objective

Spill/release of chemicals occurs most frequently in the types of FPSO accidents. Since the accident may lead to the massive fire, operators who are working on the FPSO have to evacuate safely and quickly. When an accident happens, the operators are evacuated to the lifeboats which are located on both sides of the ship following the exit signs. Therefore, objective is to find an evacuation plan to get to the nearest destination of the four points from current position.

Figure 1. Structure of FPSO: ‘A’ for accommodation, ‘B’ for backward, ‘F’ for forward, ‘D’ for deck
However, as the accident progressed, the optimal route to the exits can be changed. Thus, we considered 2 possible evacuation scenarios under this situation. In the first scenario, an agent finds the nearest exit and chooses the shortest route in no accident, which is called ‘normal situation’. The second scenario considered the case that one of the lifeboats reached the maximum capacity, which is called ‘accident situation’. In this case, the exit is not available and a sign has to guide the route to another lifeboats before the agent goes in vain. The agent learns the optimal evacuation route and the evacuation time is calculated for each case.

The actual scheme is simplified and expressed as three-dimensional grids. It used to be multi-story structure, but each section is reduced into one or two floors. Therefore, the environment considered in this paper assumes 4 sections, total 621 states. The cells represent different area as the following:

1) Work space, room, or passage (white areas in Figure 2). Agent can move freely and safely.

2) Stair (pink, orange, yellow areas in Figure 2). Pink, orange, yellow respectively means that stairs link the deck to the accommodation, forward and backward section.

3) Blocked area (black areas in Figure 2). Agent cannot move to this kind of areas.

With this environment, the model should satisfy two requirements: 1) It should solve large size problem considering the scale and complexity of FPSO. 2) It should be fast in order to be used in real-time evacuation.

3. Reinforcement learning: Deep Q-Network

Reinforcement learning (RL) is one category of machine learning but it is different from supervised or unsupervised learning. It consists of an agent, an environment, states, actions, rewards and policies. The agent chooses a proper action at each state by trial-and-error and an optimal policy (or a series of actions) maximizing the reward back from the environment can be found. From the point of policy control method, the way to find the optimal policy, on-policy and off-policy learning are considered. A typical off-policy algorithm is Q-learning. In Q learning, a function $Q(s,a)$ is defined representing the expected discounted reward for action $a$ at state $s$. Q-learning algorithm produces and updates a Q table to find the best action at a given state. However, this method is
ineffective in large space environment since there are countless number of cases. To
overcome the limit of Q-learning, deep Q network (DQN) algorithm was proposed.
DQN algorithm uses neural network to approximate Q values instead of Q table. A
neural network with weights \( \theta \) is referred to as Q network. Q-network learns a state-
action value function Q by minimizing loss function

\[
L_i(\theta_i) = \frac{1}{2} \left[ r + \gamma \max_{a'} Q(s', a'; \theta_{i-1}) - Q(s, a; \theta_i) \right]^2
\]

Eq. (1):

Also, the network is trained with a mini-batch from a replay memory which stores
experience data collected during learning. This technique called experience replay
makes the network more stable. (Target network) Therefore, the algorithm achieves
stability by breaking temporal dependency and can obtain an optimal strategy.
In this paper the agent is assumed to be a worker on the FPSO. At each cell, according
to the result of the neural network the agent decides an action between 4 directions - up,
down, left and right. Also, agent can move up or down to the other floor by the stairs.
The problem is how the agent finds the safe and fast route to the lifeboats. The agent
gains the reward \( r_t \) for the selected action. The agent makes a decision sequentially to
maximize the reward. Therefore, the reward represents the learning tendency and it also
means objectives and constraints of the action. There are five different rewards
according to the area types.

\[
r_t = \begin{cases} 
  r^{ga}, & \text{if the agent reaches the goal} \\
  r^{ba}, & \text{if the agent moves to the blocked cell} \\
  r^{va}, & \text{if the agent moves to the visited cell} \\
  r^{sa}, & \text{if the agent moves to the stair cell} \\
  r^{na}, & \text{if the agent moves to the normal cell} 
\end{cases}
\]

and the reward value always satisfies \( r^{ga} \gg r^{na} > r^{sa} > r^{va} > r^{ba} \). \( r^{ga} \) has a positive
value to encourage the agent to find the goal and the others have negative values to find
the optimal routes. Additionally, we restricted the minimum value of reward to prevent
the agent wander endless.
At first, agent wanders the whole map without any information. Then sometimes the
agent may reach the minimum reward we constrained and sometimes arrive at the
destination by chance. Whether the agent succeed or not, the chosen action and reward
the agent got at every discrete state he visited are used to update the neural network.
4. Result and discussion

The proposed DQN is trained with 20000 epochs. The policies and networks have been mentioned in previous section. To evaluate the performance of the model, success rate is recorded. Success rate is defined as the ratio of the number of episodes that successfully find the goal to the recent 300 episodes. Also, average number of moves is calculated in both normal and accident situation. The results are compared to verify whether the agent can find an alternative route properly.

In normal situation, all exits are available then the agent finds the nearest exit from the current position. When starting at 2nd floor of accommodation, the agent finds the shortest route to the exit via the 1st floor and deck and Figure 5 shows the result. The average moves of all starting points in each section are shown in Figure 6 and verify that the agent can evacuate safely regardless of where it starts.

In accident situation, one of four exits is not available due to the capacity of boats and the agent must detour by alternative route. Figure 6 shows how the agent bypasses when exit 3, which is available under normal situation, is closed. Seeing Figure 5, the closure of exit 3 affects escape from all sections except the backward. The reason the backward is not affected is that all escape routes in the backward are not involving exit 3. Although the agent bypasses the route by the effect of exit 3 closure, it finds an alternative route quickly and accurately.

5. Conclusion

A success rate at evacuation from FPSO consistently rises to 100% as training as the training progresses. Although it takes a lot of time to train the model at first, the model makes a quick calculation for every starting point once the train is completed.
Deep Q-Network model can find the shortest evacuation route in complex structure of FPSO. In addition, the route decision is flexible according to accidents and environmental changes. Used in actual marine rescue, it would be helpful to find the optimal evacuation route quickly and rationally in an emergency.

References