

# 선박 초기 설계 단계에서의 강화학습을 이용한 배관경로 설계

신동선·박병철·임채욱·오상진·김기용·신성철†  
부산대학교 조선해양공학과

## Pipe Routing using Reinforcement Learning on Initial Design Stage

Dong-seon Shin·Byeong-cheol Park·Chae-og Lim·Sang-jin Oh·Gi-yong Kim·Sung-chul Shin†  
Department of Naval Architecture and Ocean Engineering, Pusan National University

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pipe routing is a important part of the whole design process in the shipbuilding industry. It has a lot of constraints and many tasks that should be considered together. Also, the result of this stage affects follow-up works in a wide scope. Therefore, this part requires skilled designers and a lot of time. This study aims to reduce the workload and time during the design process by automating the pipe route design on initial stage. In this study, the reinforcement learning was used for pipe auto-routing. Reinforcement learning has the advantage of dynamically selecting routes, unlike existing algorithms. Therefore, it is suitable for the pipe routing design in ship design process which is frequently modified. At last, the effectiveness of this study was verified by comparing pipelines which were designed by piping designer and reinforcement learning results.

**Keywords** : Pipe routing(배관경로 설계), Reinforcement learning(강화학습), Machine learning(머신러닝)

## 1. 서론

조선해양산업에서 선박, 해양플랜트 등에 설치되는 배관의 경로를 선정하는 것은 배관업무 뿐만 아니라 전체 설계업무 중에서도 큰 비중을 차지한다. 배관경로 설계는 함께 고려해야 하는 연관 업무가 많고 경로선정의 결과가 후속되는 작업에 많은 영향을 끼치며 설계 조건이 복잡해 업무 부담이 크다. 이는 전체 설계업무 시간에 최대 50%까지 차지할 수 있는 등 많은 시간과 인력을 필요로 한다 (Park & Storch, 2002).

최근 선박의 시스템이 복잡해짐에 따라 배관의 종류와 양이 많아져 배관경로 설계업무는 더욱 복잡해졌다. 이러한 배관경로 설계는 아직까지 자동화되지 못하여 설계자가 직접 수행하기 때문에 설계자마다 배치 결과가 다를 수 있다. 또한 설계자의 경험에 의존해 작업이 이루어지기 때문에 설계자의 역량이 설계품질에 큰 영향을 미친다. 그리고 설계과정에서 참조호선이 있는 경우에는 참고해 배관경로 설계를 진행하지만, 참조호선이 없는 경우에는 새로운 환경에서 설계를 수행해야하기 때문에 업무 부담이 더욱 높아진다. 따라서 초기 설계 단계에서 배관경로 설계를

자동화하여 설계자에게 최적경로를 제시함으로써 전체 설계공정에서 업무 효율을 높일 수 있는 방안이 필요하다.

배관경로 설계 자동화는 많은 연구자에 의해 수행되어 왔다 (Kim et al., 2013; Nguyen et al., 2016). 하지만 대부분의 알고리즘들은 배관경로 설계에 필요한 설계 조건을 제한적으로 반영하였으며, 현업에서 사용하기에는 부족한 면이 있었다 (Asmara, 2013). 현재까지 수행된 연구는 배관이 설치될 환경이 확정된 상태에서만 적용될 수 있지만 실제 현업에서는 배관경로 설계가 완료된 후에도 장비위치 및 방향 등의 수정작업이 빈번하게 발생한다. 따라서 배관경로 설계 자동화 알고리즘은 이러한 환경의 변화를 고려하는 것이 필요하다.

강화학습은 경로선정문제에 있어 환경(environment)을 탐험(exploration)하고 활용(exploitation)하여 최적경로를 선정하는 알고리즘으로서 사용되어 왔다 (Millan & Torras, 1992). 에이전트(agent)를 학습시켜 행동(action)을 결정하고 학습된 에이전트는 정책(policy)을 기반으로 변화된 환경 내에서 최적경로를 선정할 수 있으며 드론이나 청소용 로봇 등 경로선정이 필수적인 기계에 적용된 사례가 많다. 강화학습은 에이전트의 행동과 보상(reward)을 이용해 설계 조건에 대한 설계자의 경험지식을 보다

효율적으로 반영할 수 있다. 그리고 환경에 대해 동적으로 반응하면서 문제를 해결하기 때문에 환경변화에 대응할 수 있는 특징이 있다.

본 연구는 강화학습을 이용해 초기 설계 단계에서 선박 엔진룸의 배관경로 설계를 자동화하는 것에 목적을 두었다. 이를 위해 실무자에 의해 설계된 총 7개의 배관을 포함하고 있는 선박 엔진룸 CAD환경을 Unity를 이용해 모델링했으며 강화학습을 이용하여 배관경로 설계에서 필요한 설계 조건들을 고려해 배관의 경로를 선정했다. 또한 본 연구의 유효성을 검증하기 위해 동일한 환경에서 실무자가 배치한 배관의 경로와 연구의 결과를 비교하였으며 끝으로 강화학습을 이용한 배관경로 설계가 목적지의 변경에 대응할 수 있음을 보였다.

## 2. 강화학습 및 학습 조건

### 2.1 강화학습

강화학습은 머신러닝의 한 분야이다. 머신러닝은 지도학습, 비지도학습, 강화학습으로 구성된다. 지도학습은 데이터와 데이터에 대한 정답(label)을 이용해 학습을 수행하며 주로 회귀 및 분류 등의 문제에 적용된다. 비지도학습은 데이터를 이용해 학습을 진행하고 군집화 등에 사용된다. 마지막으로 강화학습은 Fig. 1과 같은 구조를 가지고 환경 내에서 행동의 주체가 되는 에이전트가 보상을 최대 받는 행동을 찾는 것을 목표로 하며 주로 최적화, 로봇 제어, 게임 등에 적용된다.

에이전트는 학습의 주체로서 행동을 통해 자신의 상태를 변경한다. 상태는 에이전트에 대한 정보로 관찰 가능한 모든 집합이며 에이전트는 행동으로 인한 상태의 변경을 통해 환경으로부터 보상을 받고 학습이 진행됨에 따라 보상을 가장 많이 받는 행동을 찾는다.

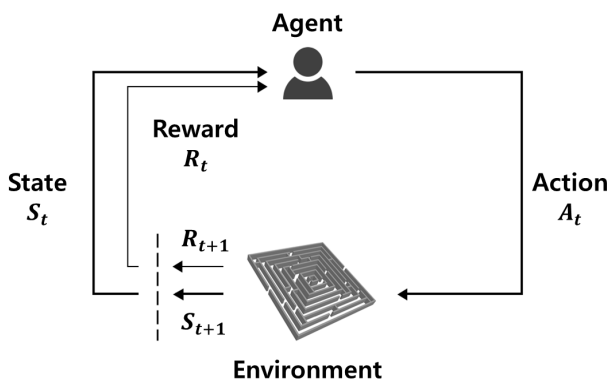


Fig. 1 Reinforcement learning process

강화학습은 예로부터 다양한 문제 해결을 위한 알고리즘으로 사용되어 왔다. 최근 DeepMind의 Deep Q-Network (Mnih et al., 2013)와 인공지능 바둑 프로그램 알파고(AlphaGo)의 개발 이후로 강화학습에 대한 연구가 머신러닝 연구에서 큰 부분을 차지하는 등 관심이 높아졌다. 본 연구에서는 강화학습 알고리즘으로

로 Proximal Policy Optimization(PPO)을 사용했다 (Schulman, et al., 2017).

강화학습에서 상태에 따라 선택하는 행동에 대한 확률을 정책이라고 하며 식 (1)과 같이 나타낸다.

$$\pi(s, a) = P[A_t = a | S_t = s] \quad (1)$$

식 (1)에서 우변은 시간이 t일 때, 상태 s에 있는 에이전트가 행동 a를 취할 확률이며 에이전트가 받은 보상을 최대로 하는 최적 정책을 찾는 것이 강화학습의 목표이다. 신경망을 이용한 강화학습에서는 신경망에 의해 정책이 결정되기 때문에 정책은 신경망의 노드를 연결하는 가중치  $\theta$ 를 이용해  $\pi_\theta(a | s)$ 로 나타낼 수 있다.

PPO는 Trust Region Policy Optimization(TRPO) (Schulman et al., 2015)의 단점을 해결한 알고리즘이다. 식 (2)와 같이 정책을 업그레이드하는 TRPO는 순차적 행동(continuous action)을 학습할 수 있으며, 학습의 진행에 따라 정책이 업데이트 될 때 불안정성을 줄이기 위해 정책변화의 정도를 제한했다.

$$Maximize_\theta E_t \left[ \frac{\pi_\theta(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)} \hat{A}_t \right] \quad (2)$$

$$s.t. \quad E_t [KL[\pi_{\theta_{old}}(a_t | s_t), \pi_\theta(a_t | s_t)]] \leq \delta$$

식 (2)에서  $A_t$ 는 advantage function으로 행동 가치함수에서 가치함수를 뺀 값이며 현재 행동이 얼마나 가치 있는 행동인지를 판단하는 기준이 된다. TRPO는 제한조건인 trust region에서 업데이트 된  $\pi_\theta$ 가  $\pi_{\theta_{old}}$ 로부터 크게 바뀌지 않게 되면서 학습에 있어 안정성을 가지게 되지만 업데이트의 범위를 제한하면서 과도한 계산량이 발생하며 구현이 복잡하고 어렵다는 단점이 있다. 이를 개선하기 위해 식 (3)과 같은 clip개념을 도입한 PPO가 개발되었다 (Schulman et al., 2017).

$$L^{CLIP}(\theta) = E_t \left[ \min \left( \frac{\pi_\theta(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)} \hat{A}_t, \text{clip} \left( \frac{\pi_\theta(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)}, 1 - \epsilon, 1 + \epsilon \right) \hat{A}_t \right) \right] \quad (3)$$

식 (3)에서 advantage function이 양수일 경우 정책 업데이트의 범위를  $1+\epsilon$ 을 선택함으로써 현재 업데이트 방향을 유지하며 advantage function이 음수일 경우  $1-\epsilon$ 을 경계로 선택함으로써 업데이트 방향을 반대로 취하게 된다. 이렇게 clip을 이용해 업데이트 크기의 범위를 줄이면서 TRPO에서 정책을 업데이트시키는데 필요한 계산량을 줄여 구현을 간편화하는 동시에 업데이트 범위를 효과적으로 제한해 신뢰할 수 있는 학습효과를 가진다는 장점이 있다.

### 2.2 선박 배관경로 설계 제약 및 운용조건

일반적으로 배관의 경로를 선정하는데 고려해야하는 설계 조

건은 아래와 같다 (Park & Storch, 2002).

- (1) 배관은 장애물을 피해 설치되어야 한다.
- (2) 원활한 장비의 운용을 위해 장비와 배관 사이에 최소한의 거리가 유지되어야 한다.
- (3) 배관 밸브의 높이는 작업자의 편의를 고려해야 한다.
- (4) 지지대(Support)를 최대한 공유해야 한다.
- (5) 경로상의 길이와 굵힘을 최소화해야 한다.

위에 명시된 설계 조건 외에도 배관은 통로 상에 설치되지 않아야 하며 노즐의 시작점과 끝점이 확실하게 연결되어야 한다. 또한 안전상의 이유로 물 전용 배관의 경로는 전기 장비 위를 지나가지 않아야 하며 연소 엔진 위는 기름전용 배관의 경로로써 선택해서는 안 된다는 등의 배관을 통하는 물질과 배관 주위의 설치장비 사이의 관계에 대한 선급규칙도 반영해야 한다. 또한 상황에 따라 지름이 큰 배관의 경로를 우선적으로 선정해야 하며, 경우에 따라 굵힘에 대한 비용보다 지지대 설치에 대한 비용이 더 크기 때문에 지지대 설치를 우선적으로 고려한다.

### 2.3 배관경로 설계 문제 환경 및 학습 조건

본 연구에서는 경로선정시뮬레이션 및 학습을 위해 Unity ML-agents을 이용했다. Unity ML-agents는 3D 환경을 만들고 이를 학습에 이용할 수 있으며, 본 연구에서 사용된 에이전트는 Unity 환경 내에서 상, 하, 전, 후, 좌, 우로 총 6가지의 행동을 취할 수 있다( $A_t = \{up, down, forward, backward, left, right\}$ ). 행동을 취하면 선택한 방향을 기준으로 단위 거리만큼 이동하며 에이전트의 상태는 목적지까지의 x, y, z축 방향 거리이고 에이전트, 에이전트의 목적지, 장비 및 장애물로 이루어진 선박 엔진룸 환경에서 구 형태의 에이전트가 같은 색을 가진 정육면체 형태의 목적지에 도착하는 것이 강화학습 모델의 최종 목표이다. 그리고 Fig. 2는 배관경로 설계 알고리즘 검증을 위해 사용된 7개의 배관경로를 포함하고 있는 환경이다.

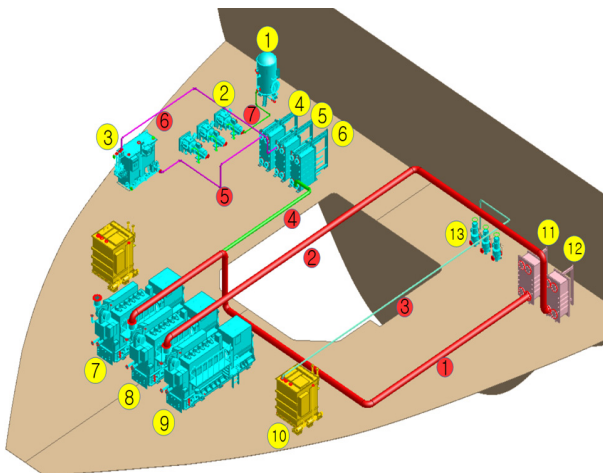


Fig. 2 Pipeline in simplified engine room model

위 환경은 단순화시킨 선박의 엔진룸으로 실무자에 의해 3D CAD 프로그램을 바탕으로 설계된 배관의 경로이다. Fig. 3과 Fig. 4는 이를 Unity 환경에 구현한 것으로 환경의 복잡성을 낮추고 에이전트의 불필요한 환경탐색을 방지하며 학습의 효율을 높이기 위해 장비 및 장애물은 직육면체로 단순화시켰다. 또한 모델 중앙에 있는 홀을 구현하기 위해 해당 위치에 투명한 박스를 두어 배관의 경로로써 선정되지 못하게 했다. 그리고 장비를 장비보다 큰 투명박스로 감싸 배관과 장비사이의 이격거리를 고려했다. Table 1은 각 장비를 연결하는 배관의 번호이다.

Table 1 Pipe number

Pipe number	Start equipment	Destination equipment
1	8	11
2	9	12
3	10	13
4	7	6
5	3	4
6	3	5
7	1	2

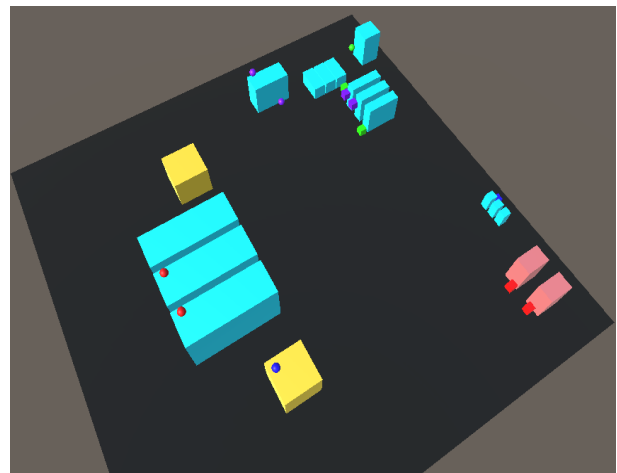


Fig. 3 Isometric view of learning environment

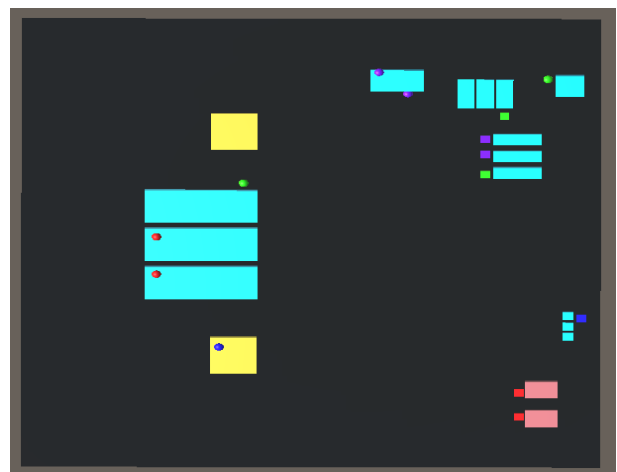


Fig. 4 Top view of learning environment

환경 안에서 에이전트가 목적지까지 도달하면서 생성한 경로는 배관이 설치될 경로로 선정되며 학습을 거쳐 보상을 최대로 받을 수 있는 경로를 최종 경로로 선정한다. 본 연구에서 배관경로 설계에 강화학습을 적용하기 위해 사용된 설계 조건 및 그에 따른 보상은 아래와 같다.

- (1) 길이 최소화를 위한 시간에 따른 보상 [-0.10]
- (2) 굽힘 최소화를 위한 경로변경에 따른 보상 [-0.10]
- (3) 목적지에 가까워지는 행동에 따른 보상 [+0.10]
- (4) 목적지에 멀어지는 행동에 따른 보상 [-0.10]
- (5) 목적지 노즐진입방향에 따른 보상 [+0.10]
- (6) 목적지 도착에 따른 보상(짧은 경로) [+5.00]
- (7) 목적지 도착에 따른 보상(긴 경로) [+10.0]
- (8) 경로로서 천장을 이용함에 따른 보상(천장용 배관) [+0.10]
- (9) 경로로서 바닥을 이용함에 따른 보상(바닥용 배관) [+0.10]

에이전트가 도착지까지 도달하는데 필요한 총 행동 횟수를 최소화함으로써 배관의 길이를 최소화하고 배관의 굽힘을 최소화하기 위해 학습에 소요되는 시간 및 경로변경에 음의 보상을 부여했고 학습의 효율을 위해 에이전트와 목적지와의 거리를 이용해 보상을 부여했다. 또한, 장비와 배관을 연결하는 노즐의 방향을 맞추기 위해 목적지에 도달하기 전 특정 위치를 지나치면 양의 보상을 부여했다. 그리고 Fig. 2에서 1번 및 2번 배관처럼 다른 배관보다 비교적 긴 경로의 경우 학습되지 않은 초기 에이전트가 목적지를 찾는데 오래 걸리고 목적지까지 가는데 배관 길이 및 굽힘에서 음의 보상이 비교적 잦게 발생하기 때문에 목적지 도착에 따른 보상을 더 크게 부여했다. 또한 바닥용 배관 에이전트는 경로선정에 있어 바닥을 이용하는 행동을 하면 +0.10점을 받으며 천장용 배관 에이전트는 천장을 이용하는 행동에 +0.10점을 받는다. 이렇게 강화학습은 배관경로 설계에 있어 특정조건에 보상을 부여함으로써 에이전트의 행동양상을 조절해 일반적인 경우가 아닌 특수한 경우의 배관을 고려하는 것이 가능하다. 아래 Table 2는 본 연구에서 고려한 설계 조건 및 보상의 종류와 이에 해당하는 배관의 번호이다.

Table 2 Agents design reward

Design reward	Pipe number	Value
Action for minimum length	1, 2, 3, 4, 5, 6, 7	-0.1
Action for minimum bending	1, 2, 3, 4, 5, 6, 7	-0.1
Action for close to the destination	1, 2, 3, 4, 5, 6, 7	+0.1
Action for far from the destination	1, 2, 3, 4, 5, 6, 7	-0.1
Action for the inlet nozzle direction	1, 2, 3, 4, 5, 6, 7	+0.1
Reach to the destination (short path)	3, 4, 5, 6, 7	+0.1
Reach to the destination (long path)	1, 2	+0.1
Reward for using the floor as the route	1	+0.1
Reward for using the ceiling as the route	2	+0.1

학습에 사용된 신경망과 강화학습의 Hyper-parameters와 하드웨어의 사양은 아래 Table 3 및 4와 같다.

Table 3 Hyper-parameters

Parameter	Value
Batch size	256
Buffer size	2560
Beta	0.01
Epsilon	0.2
Gamma	0.9
Lambda	0.9
Time horizon	64
Number of hidden layer	2
Number of epoch	3
Hidden units	128
Learning rate	0.001
Max step	500,000

Table 4 Experimental hardware specification

	Value
CPU	Intel(R) Core(TM) i7-8700K CPU @ 3.60GHz
GPU	NVIDIA GeForce GTX 1080 8.0GB
RAM	32.0GB

경로선정문제에 있어 최적경로를 완벽하게 도출해내기 위해서는 문제 환경에 대한 전역적 탐색과정을 거쳐야한다. 하지만 이는 시간적 제약이 존재하는 현실 문제에서는 적용하기 어려울 뿐만 아니라 효율성 또한 떨어진다. 최적경로를 찾을 때 음의 가치를 허용하지 못한다면 피해야 할 설계 조건에 대한 적용이 번거로워지기도 한다. 강화학습은 설계 조건에 대한 보상으로 음의 값을 부여함으로써 지양해야하는 조건을 직관적이고 쉽게 적용할 수 있다. 또한 잦은 수정 및 반복 작업을 거치는 선박 배관경로설계에서 재계산 없이 기존 학습결과를 재활용함으로써 일정 수준 이내의 환경 변경에 대한 대응이 가능하기 때문에 기존의 경로선정 알고리즘과 비교해 이점을 가지고 있으며, 이를 Table 5에 나타내었다.

Table 5 Path-finding algorithm comparison

Function \ Algorithm	Dijkstra	A*	Reinforcement learning
Global searching	O	X	X
Negative weight value	X	X	O
Response to destination change	X	X	O

### 3. 학습 결과

본 연구에서는 강화학습을 이용해 총 7개 배관의 경로를 선정했다. 에이전트들의 평균 학습시간은 약 14분 7초가 소요됐다. Fig. 5와 Fig. 6는 바닥용 배관과 천장용 배관의 학습을 진행할 때 step에 따른 보상으로 학습에 따른 보상의 경향성 및 계산의 효율성을 위해 1,000 step 별 평균 보상값을 나타냈으며, 바닥용 배관의 보상은 16.03, 천장용 배관의 보상은 15.01로 학습을 종료했다. 또한 나머지 5개 배관의 1,000 step 별 평균 보상값은 Fig. 5~11과 같다. 이들의 보상은 순서대로 4.60, 4.90, 5.00, 4.80, 5.00으로 학습을 종료했다. 그리고 최종적으로 선정된 배관의 경로는 Fig. 12 및 Fig. 13과 같다.

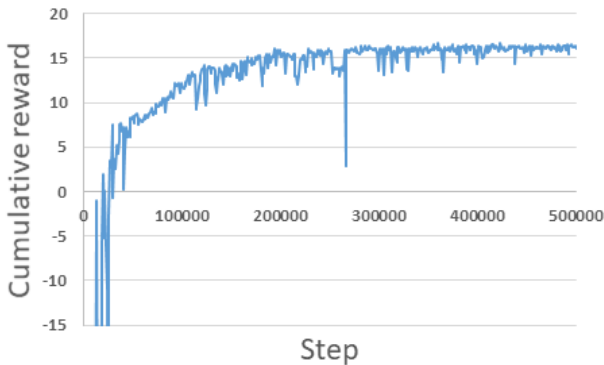


Fig. 5 Cumulative reward per step of Pipe No.1

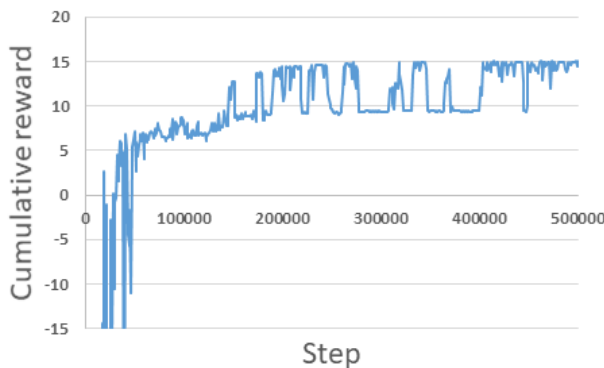


Fig. 6 Cumulative reward per step of Pipe No.2

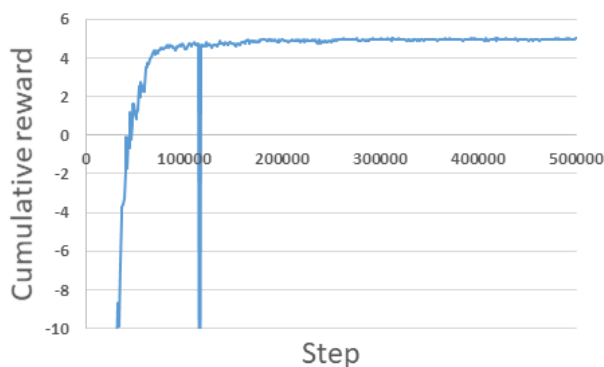


Fig. 7 Cumulative reward per step of Pipe No.3

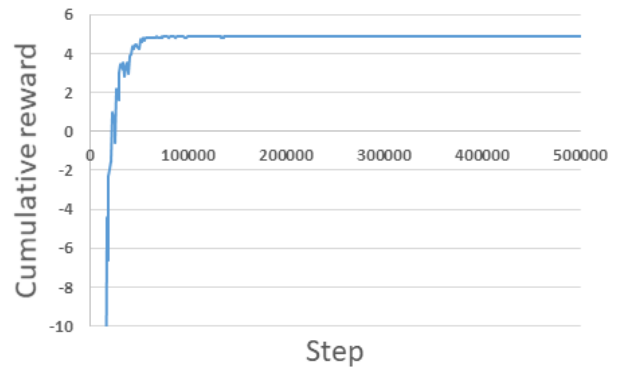


Fig. 8 Cumulative reward per step of Pipe No.4

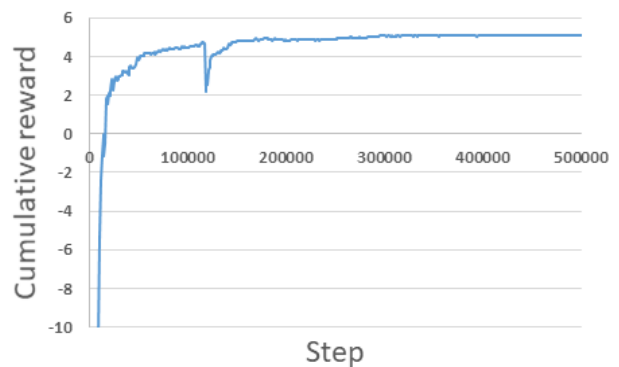


Fig. 9 Cumulative reward per step of Pipe No.5

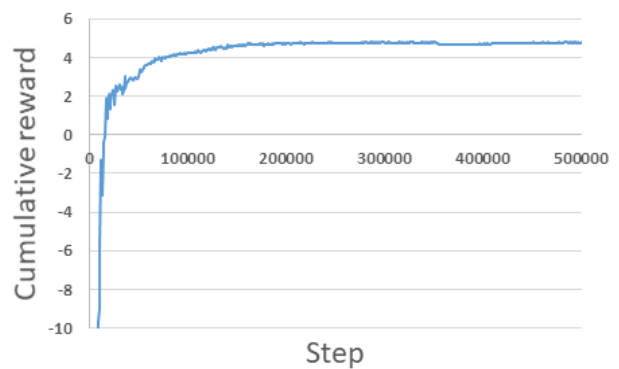


Fig. 10 Cumulative reward per step of Pipe No.6

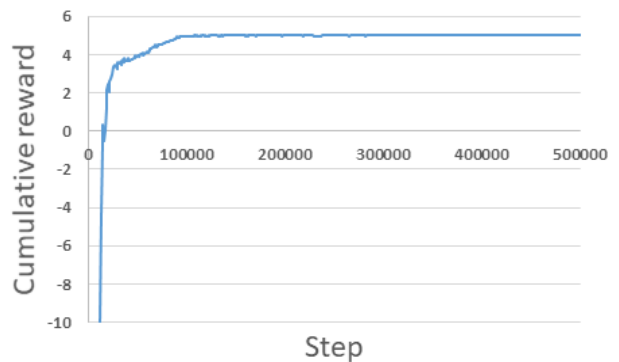


Fig. 11 Cumulative reward per step of Pipe No.7

Fig. 5 및 7에서 보상값이 급격하게 감소하는 것을 볼 수 있는데 이는 에이전트가 학습 도중 장비 혹은 장애물에 부딪히거나 특정 구간에서 목적지까지 가까워지는 행동을 하지 않으면서 시간에 따른 음의 보상만을 받는 상태이다. 하지만 에이전트는 학습이

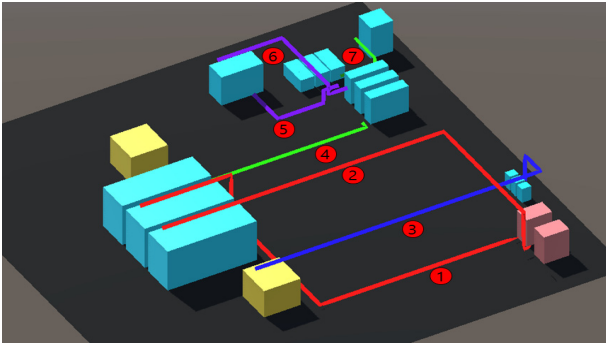


Fig. 12 Learning result

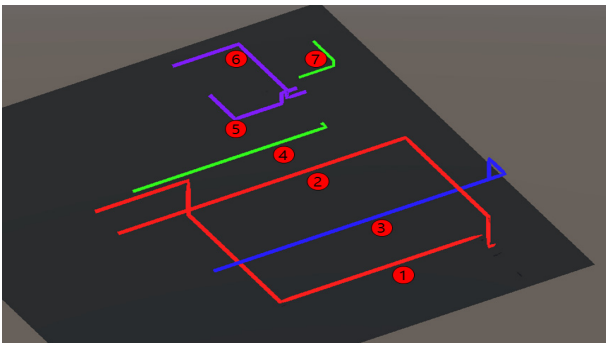


Fig. 13 Learning result without structures

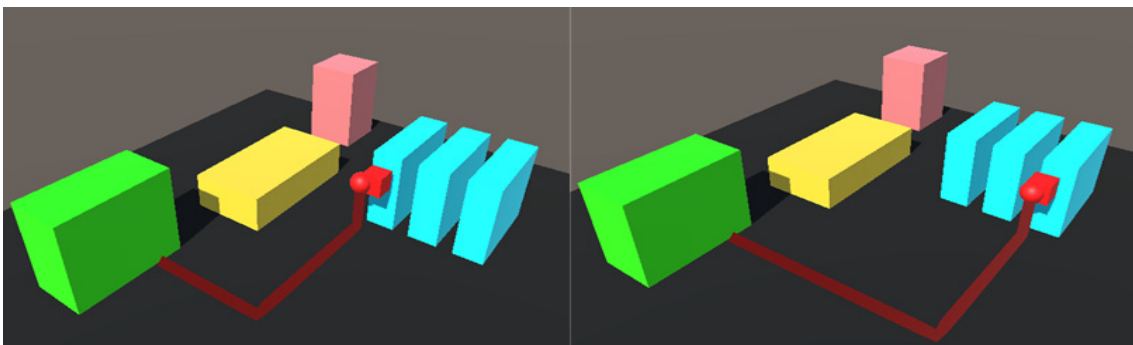
진행됨에 따라 해당 행동이 좋지 않은 행동임을 알고 본래 보상 값으로 수렴하는 것을 볼 수 있다.

단순화시킨 선박 엔진룸 모델인 Fig. 2와 강화학습의 배관경로 설계 결과 Fig. 12 및 13을 비교하면 7개 배관의 경로 상 굽힘 횟수가 동일하고 도착노즐의 진입방향이 동일함을 확인할 수 있으며 불필요한 경로를 만들어내지 않고 최단 거리로 경로를 선정했다. Table 6는 각 결과의 배관 굽힘 횟수를 비교한 표이다.

Fig. 14은 강화학습을 이용한 배관경로 설계가 학습이 종료된 후에 목적지가 변화더라도 이에 대응할 수 있음을 보여준다. 그림(a)는 본래 환경에서의 도착점으로 학습이 완료된 후 경로선정을 끝낸 상태이다. 그리고 그림(b), (c), (d)는 도착점을 변경한 뒤 이미 학습이 완료된 에이전트를 이용해 선정된 배관의 최적 경로이다. 강화학습을 이용한 경로선정이 도착점이 변경된 뒤에도 다시 경로선정 작업을 거칠 필요 없이 이미 학습된 데이터를 바탕으로 목적지의 변화에 대응하여 합리적인 경로를 선정할 수 있는 것을 확인할 수 있다.

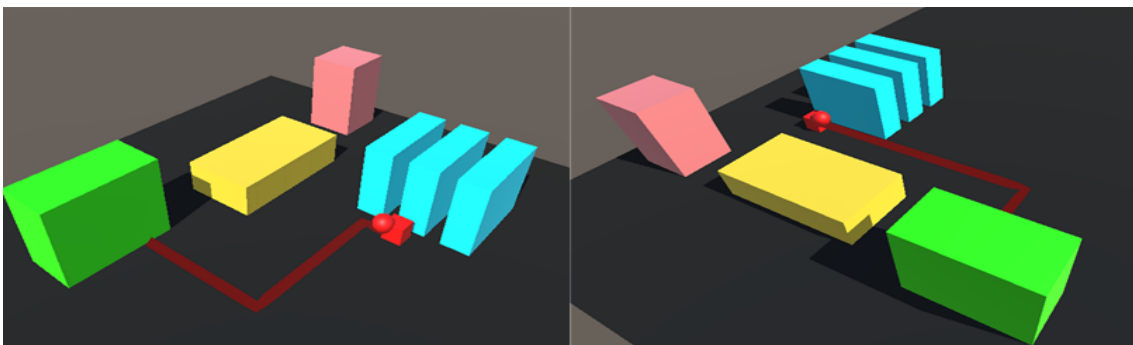
Table 6 Pipe bending number comparison

Pipe number	Ship engine room model	Learning result
1	3	3
2	3	3
3	2	2
4	1	1
5	3	3
6	3	3
7	2	2



(a)

(b)



(c)

(d)

Fig. 14 Original pipeline(a) & modified pipeline after changes its destination(b, c, d)

## 4. 결론

본 연구를 통한 결론은 아래와 같다.

- (1) 선박의 엔진룸 환경에서 강화학습을 이용해 총 7개 배관의 경로를 선정했다.
- (2) 설계자에 의해 배치된 경로와 강화학습의 결과를 비교해 본 연구의 유효성을 검증하였다.
- (3) 에이전트의 최종 step에서의 보상은 배관의 번호 순서대로 16.03, 15.01, 4.60, 4.90, 5.00, 4.80, 5.00로 바닥 및 천장용 배관은 선호되는 공간을 경로로 사용함으로써 설계 조건으로부터 추가적으로 보상을 받아 경로선정완료 시 받는 +10.0보다 큰 보상을 받았으며 나머지 배관들은 양의 보상과 음의 보상을 받는 설계 조건들이 적절하게 적용되어 도착완료 시 받는 +5.0과 크게 차이나지 않는 수치로 학습을 종료했다.
- (4) 보상을 이용해 특수한 경로를 고려할 수 있음을 보였다.
- (5) 강화학습을 이용한 배관경로 설계는 경로선정이 완료된 후 목적지가 변하더라도 경로선정을 위한 재계산 없이 학습된 데이터를 바탕으로 변화에 대응하여 경로를 선정할 수 있음을 보였다.

실무자가 배치한 배관의 경로와 연구의 결과는 대부분이 유사했지만 약간의 차이를 보였다. 이는 배관경로 설계가 설계자의 경험을 기반으로 이루어지는 업무로서 설계자마다 배치결과가 달라질 수 있기 때문에 연구에 사용된 모델이 완벽한 경로는 아니며 모델이 설계될 때 고려되어야 할 설계 조건이 학습에 충분히 반영되지 못했기 때문이다. 실제 선박의 엔진룸은 복잡하지만 연구에 적용된 모델은 비교적 간단하기 때문에 배관경로 설계의 주요한 설계 조건만을 적용해 유효한 결과를 얻었다. 하지만 본 연구가 실제 산업에 직접적으로 사용되기 위해선 배관 서포터 공유, 선호되는 밸브 높이, 배관 열변형 등의 설계 조건까지 고려되어야 한다. 향후 연구를 통해 정밀한 환경을 기반으로 다양한 설계 조건을 적절하게 반영하고 이 설계 조건들의 상대적인 우선순위를 보상으로써 적절하게 나타낼 수 있도록 알고리즘을 구성한다면 더욱 정교한 배관경로 설계 자동화를 이룰 수 있을 것이다.

본 연구에서는 강화학습을 이용한 경로선정이 목적지 변화에 대응할 수 있음을 보였지만 변경된 목적지까지 도달하는 경로 상에 장애물이 생기거나 변경된 목적지의 위치가 기존 목적지와 크게 다르다면 합리적인 경로를 선정하지 못했다. 이는 강화학습 에이전트가 학습된 정보를 바탕으로 경로를 선정하기 때문에 학습되지 못한 행동범위는 행할 수 없기 때문이다. 목적지 변경뿐만 아니라 장비위치의 변경 및 새로운 장비 출현 등의 다양한 환경변화에 완벽하게 대응하기 위해서는 추후 다양한 행동을 학습할 수 있는 환경과 에이전트를 구축하는 것이 필요하다.

마지막으로 본 연구가 위 내용을 바탕으로 발전되어 초기 설계 단계에 적용된다면 설계자에게 배관의 경로를 제시해줌으로써 설계자가 겪는 업무부하를 줄여 전체 배관경로 설계에 소요되는 시간을 줄이고 전체 설계 단계의 효율을 높일 수 있을 것이다.

## 후기

본 연구는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단(No. 2011-0030013)과 2020년도 국토교통부의 재원으로 국토교통과학기술진흥원의 지원을 받아 수행된 연구임 (20FIP-B133625-04, 이수 혼합 및 처리시스템 구조 안정성 평가)

## References

- Asmara, A., 2013. *Pipe routing framework for detailed ship design*. Ph.D. Delft: Delft University of Technology.
- Kim, S.H., Ruy, W.S. & Jang, B.S., 2013. The development of a practical pipe auto-routing system in a shipbuilding CAD environment using network optimization. *International Journal of Naval Architecture and Ocean Engineering*, 5(3), pp.468-477.
- Milan, J., D., R., & Torras, C., 1992. A reinforcement connectionist approach to robot path finding in non-maze-like environments. *Machine Learning*, 8(3-4), pp.363-395.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D. & Riedmiller, M., 2013. Playing Atari with Deep Reinforcement Learning. *arXiv:1312.5602*.
- Nguyen, H., Kim, D.J. & Gao, J., 2016. 3D Piping route design including branch and elbow using improvements for dijkstra's algorithm. *International conference on artificial intelligence: Technologies and Applications*, pp.309-312.
- Park, J.H. & Storch, L., R., 2002. Pipe-routing algorithm development: case study of a ship engine room design. *Journal of the Expert Systems with Applications*, 23(3), pp.299-309.
- Schulman, J., Levine, S., Moritz, P., Jordan, M. & Abbeel, P., 2015. Trust region policy optimization. *arXiv:1502.05477v1*.
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O., 2017. Proximal policy optimization Algorithms. *arXiv:1707.06347v2*.



신 동 선



박 병 철



임 채 욱



오 상 진



김 기 웅



신 성 철