JPNT 10(2), 113-120 (2021)
https://doi.org/10.11003/JPNT.2021.10.2.113

JPNT Journal of Positioning, Navigation, and Timing

A Study on Deep Reinforcement Learning Framework for DME Pulse Design

Jungyeon Lee, Euiho Kim†
Mechanical System and Design Engineering, Hongik University, 04066, Korea

ABSTRACT

The Distance Measuring Equipment (DME) is a ground-based aircraft navigation system and is considered as an infrastructure that ensures resilient aircraft navigation capability during the event of a Global Navigation Satellite System (GNSS) outage. The main problem of DME as a GNSS back up is a poor positioning accuracy that often reaches over 100 m. In this paper, a novel approach of applying deep reinforcement learning to a DME pulse design is introduced to improve the DME distance measuring accuracy. This method is designed to develop multipath-resistant DME pulses that comply with current DME specifications. In the research, a Markov Decision Process (MDP) for DME pulse design is set using pulse shape requirements and a timing error. Based on the designed MDP, we created an Environment called PulseEnv, which allows the agent representing a DME pulse shape to explore continuous space using the Soft Actor Critical (SAC) reinforcement learning algorithm.

Keywords: distance measuring equipment (DME), alternative position, navigation and timing (APNT), reinforcement learning, deep learning

1. 서론

Global Navigation Satellite System (GNSS) 가동 중단 시 안전한 항공 교통 관제 운영을 유지하기 위한 Alternative Position, Navigation, and Timing (APNT) 시스템에는 단기 백업 솔루션으로 Distance Measuring Equipment (DME) (DME/N 또는 DME Normal)를 사용한다 (FAA 2016). DME란 전파가 항공기로부터 보내어진 질문신호가 지상의 응답기에 의해 응답신호의 형태로 되돌아올 때까지의 걸리는 시간을 측정하여 지상의 특정 점까지의 거리를 측정하는 방법을 말한다. 지상의 응답기와 항공기의 송신기 사이에 교환된 한 쌍의 펄스의 경사 범위를 측정하게 되고, 둘 이상의 지상 응답기로부터 항공기 수평 위치를 측정된 경사 범위와 DME 지상 응답기의 알려진 좌표를 사용하여 계산할 수 있다. 1950년부터 사용되었고 기술적 완성도가 다른 APNT 대안 방법론들보다 우위에 있지만, 대부분의 DME 펄스에 적용되는 Gaussian pulse waveform은 범위 정확도가 낮기 때문에 Federal Aviation Administration (FAA)에서 요구하는 최대 0.3 nm (nautical miles) 정확도를 요구하는 FAA의 APNT 측위 성능을 제공하지 못한다. (Kim 2012) 따라서 오차에 가장 큰 영향을 주는 DME 펄스 형태를 새롭게 디자인하는 시도가 있었다 (Lilley & Erikson 2012). Smoothed Concave Polygon (SCP) 펄스 (Kim 2013, 2017)는 DME 다중 경로 효과를 상당히 완화시켜 다중 경로 유도 범위 오차를 Gaussian 펄스에 비해 약 50% 감소시켰다. 하지만 SCP 펄스의 설계 프로세스는 가능한 DME 펄스 중에서 좁은 탐색공간만 고려하기 때문에 고급 최적화 기법을 사용할 경우 SCP 펄스보다 우수한 다른 대체 DME 펄스가 존재할 수 있다. 이를 보완한 후속연구로는 Genetic Algorithms (GA)을 활용한 Stretched-FrOnt-Leg (SFOL) 펄스가 있다 (Kim & Seo 2017). GA를 이용하여 더 넓은 탐색공간에서 전역 최적 값을 찾아 Gaussian 및 SCP 펄스보다 다중경로 효과에 의한 timing error를 줄일 수 있었으며 전체적인 DME 펄스의 범위 정확도를 크게 향상시켰다. 그러나 노이즈를 고려한 민감도 분석에서 SFOL 펄스의 성능이 저하되고 Signal to Noise Ratio이 감소함에 따라 Gaussian 및 SCP 펄스에 대한 이점이 감소했다. 따라서 여전히 DME 펄스 형태 개발의 필요성이 있다.

따라서 본 연구에서는 딥러닝과 강화학습을 결합한 심층 강화학습 Deep Reinforcement Learning (DRL)을이용한 새로운 DME

Received Jan 28, 2021 **Revised** May 17, 2021 **Accepted** May 25, 2021
†**Corresponding Author**
E-mail: euihokim@hongik.ac.kr
Tel: +82-2-320-1636 Fax: +82-2-320-1636
Jungyeon Lee https://orcid.org/0000-0002-0802-4141
Euiho Kim https://orcid.org/0000-0002-6501-9330

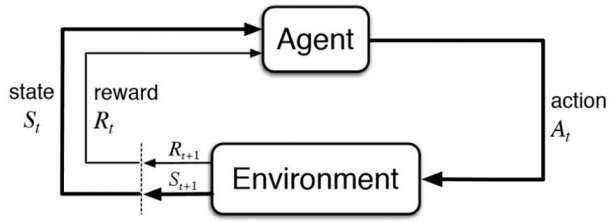

Fig. 1. Deep reinforcement learning.

Table 1. Table 1. Markov decision process table

Notation	Description
$\mathcal{S}$	A set of states
$\mathcal{A}$	A set of actions
$\mathcal{P}$	A state transition probability matrix $\mathcal{P}_{ss'}^{a} = \mathbb{P}[S_{t+1} = s' \mid S_t = s, A_t = a]$
$\mathcal{R}$	A reward function $\mathcal{R}_s^a = \mathbb{E}[R_{t+1} \mid S_t = s, A_t = a]$
γ	A discount factor

펄스 형태 개발 프레임 워크를 제안하고 새로운 펄스 형태 개발 가능성을 보여주고자 한다. DRL은 다양한 분야에서 그 실효성이 입증되었다. 57개 Atari 게임에서 탐색공간이 넓은 raw pixel input을 가지고 사람보다 더 좋은 성능을 냈으며 (Badia et al. 2020), 게임 조작키와 같은 Discrete 공간에서의 제어뿐만 아니라 연속적이고 차원의 크기가 큰 Continuous 공간에서의 제어가 필요한 4족 보행로봇과 정교한 로봇 손 동작에서도 뛰어난 성능을 보여주었다 (Haarnoja et al. 2019). 이러한 연구 결과에서 볼 수 있듯이 DRL은 탐색공간이 크고 제어가 복잡한 task들에서도 문제를 잘 해결하는 솔루션이 될 수 있음을 보여준다. 이러한 DRL의 학습구조는 Fig. 1에서 보이는 것과 같이 Agent와 Environment라는 2개의 구조로 이루어져 있으며, 다른 인공지능 방법론들과 다르게 정답(label)이 주어지지 않고 Agent와 Environment 사이의 상호작용으로부터 학습하게 된다. Agent는 환경으로부터 받는 누적 리워드를 가장 크게 하는 최적 정책을 학습하게 된다. 이때 신경망 기술과 결합된 심층 강화학습은 정책을 신경망으로 근사하여 함수 최적화로 다양한 문제들을 풀 수 있다. 그래서 다중 경로 효과에 의한 DME 펄스의 timing error를 줄이기 위해 DME 펄스 형태를 디자인하는 문제는 넓은 탐색공간을 가지고 있기 때문에 DRL을 적용하여 문제를 풀어보고자 한다.

강화학습의 학습과정은 다음과 같다. Agent가 현재 state에서 action을 취하게 되면, Environment는 다음 state와 reward를 Agent에게 주게 된다. 이렇게 timestep에 따라 Environment에게 받는 reward를 기준으로 Agent가 학습하게 된다. 이때 Agent는 어떻게 action을 선택해야 할지 알려주는 정책 신경망을 통해 최종적으로 Environment로부터 받는 timestep에 따라 누적된 reward, 반환값 (G, return)을 가장 크게 하는 최적 정책을 학습하게 된다. 강화학습으로 문제를 풀기 위해서는 Table 1에서 볼 수 있는 Markov Decision Process (MDP)로 정의할 수 있어야 하며, 수식으로는 MDP를 기반으로 시간 t에 대해서 할인율을 고려한 누적 reward, 반환 값을 $G_t = \sum_{i=t}^{T} \gamma^{(i-t)} r(s_i, a_i)$로 나타낼 수 있다.

따라서 본 연구에서는 새로운 DME 펄스 형태를 제안하기 위해, DME 펄스 디자인 문제를 MDP로 설계하고 이를 기반으로 Pulse Environment의 줄임말로 PulseEnv라는 Environment를 만들었다. Agent는 DRL 알고리즘 중 하나인 Soft Actor Critic (SAC)을 이용하였다 (Haarnoja et al. 2019). 구체적인 MDP에 대해서는 2.1절에서 설명하였으며, PulseEnv와 SAC Agent에 대해서는 각각 2.2절과 2.3절에서 자세히 설명하였다.

2. 연구 방법

2.1 DME Pulse Design MDP

DRL은 Environment와 Agent의 상호작용으로 학습하는 인공지능 방법이다. DRL은 MDP로 문제가 정의된다. DME 펄스 디자인의 MDP는 Table 1에서와 같이 $\mathcal{S}$, $\mathcal{A}$, $\mathcal{P}$, $\mathcal{R}$ 4개의 공간으로 정의되며, 먼저 $\mathcal{S}$는 state space를 나타내고 각 timestep t에 대한 샘플 state $\mathbf{s}_t \in \mathcal{S}$는 펄스 형태을 나타내는 seed point들의 위치가 된다. DME 펄스는 −6 ~ 6 μs time domain에서 총 61개 seed point들을 가지게 되며, 각 seed point들은 time domain에서 일정한 간격 0.4 μs를 유지한다. 총 61개의 seed point 개수를 사용하였지만 이는 임의의 개수로 추후 Agent의 action에 따라 변경될 수 있으며 각 seed point의 amplitude 값은 0 ~ 1 사이의 실수 값을 가진다. 따라서 pulse 형태의 state space는 61개의 0 ~ 1 사이의 연속적인 값을 가지는 벡터가 된다. Discrete한 61개의 seed point들을 continuous한 space에서 완성시키기 위해 seed point들의 간극은 smoothing factor를 이용하여 cubic spline interpolation 기법을 사용하여 형태를 완성시킨다. Fig. 2에서 이렇게 만들어진 펄스의 형태의 예시를 볼 수 있다. 이때 smoothing factor는 0에서 1사이의 값으로, 1에 가까울수록 seed point들에 fit한 곡선의 형태를 완성한다. Fig. 3에서 보았을 때 주황색 곡선의 smoothing factors는 0.5이고 초록색 곡선의 smoothing factor는 0.99로, 초록색 곡선이 seed point들에 대해 더 fit하며 이때의 smoothing factor가 1에 더 가까운 것을 확인할 수 있다.

$\mathcal{A}$는 action space를 나타내고, action은 time domain에서 연속된 3개의 seed point들을 움직이는 것으로 정의했다. DME 펄스 Agent가 1번의 action을 수행하는 것은 총 2단계로 나누어져 있다. 첫번째 단계로는 action으로 amplitude 값을 변화시킬 연속된 3개의 seed point들을 선택하는 것이다. Time domain에서 연속된 3개의 점이므로, 3개의 seed point들 중 time이 빠른 시작점을 고르는 것이 amplitude를 변화시킬 3개의 점을 선택하는 것과 같다. 시작점이 될 수 있는 seed point들은 총 39개이다. 펄스의 안정적인 형태 구현을 위해서 총 61개의 seed point들 중 펄스의 앞부분의 10개의 점과 뒷부분의 10개의 점들은 amplitude 값을 0으로 고정시켜 두었다. 그렇기 때문에 연속되는 3개의 점들의 시작점이 될 수 있는 seed point들의 수는 61−(20×2)−3+1=39임을 알 수 있다. 이때 시작점이 될 수 있는 seed point들은 0 ~ 38 정수의 인덱스를 가지고 있다. 두번째 단계로는 선택된 3개의 seed point들의 amplitude 변화 정도를 −0.05 ~ 0.05 (전체 amplitude의 스케일은 0 ~ 1) 사이의 연속된 실수 값들 중에서 선택하는 과정이다. 3개의 seed point들은 amplitude가 변화되는 정도가 서

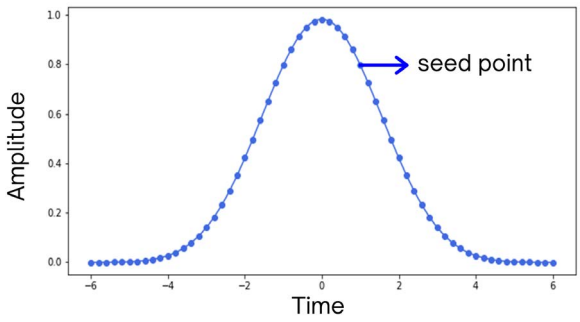

Fig. 2. Example of state space.

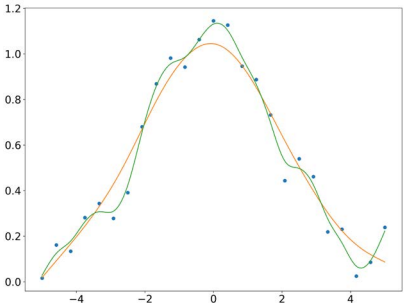

Fig. 3. Comparison of smoothing factors (smoothing factor value of green is closer to 1 than that of orange).

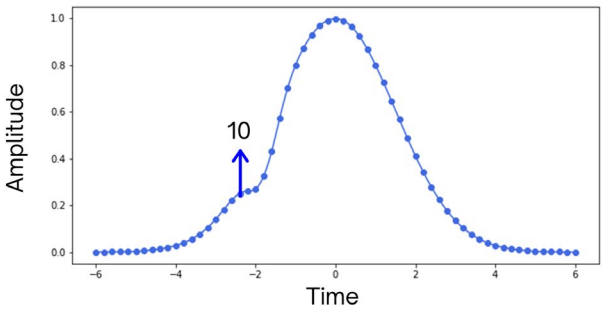

Fig. 4. Example of the action.

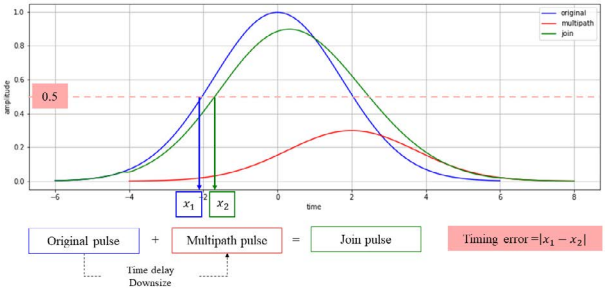

Fig. 5. The process of calculating timing error.

로 다르다. Fig. 4에서 한 가지 예시를 보면 action이 "10 [−0.03598 −0.0202 −0.0327]"으로 선택되었을 때, 연속하는 seed point들의 시작점은 인덱스 10에 해당하는 seed point이며, 인덱스 10, 11, 12에 해당하는 seed point들이 각각 −0.03598, −0.0202, −0.0327 만큼 amplitude 값이 변화하게 된다.

$\mathcal{R}$은 현재 state와 action을 가지고 Environment로부터 받는 reward로, 심층강화학습 방법에서 Agent의 목표를 나타내는 중요한 지표이다. 본 연구의 목적은 DME 펄스가 다중 경로효과에 의한 형태 변형이 일어나 original 펄스 형태와 변형된 펄스 형태 차이에서 나타나는 timing error를 줄이는 것이 목표이다. 이를 고려하여 reward는 다중경로 효과를 고려한 변형 펄스 (Join pulse)와 원래 펄스 (Original pulse)로부터 timing error를 구하고 이 값을 이용하여 설계하였다. Timing error를 구하는 과정은 먼저 원래 펄스에서 일어나는 time delay와 amplitude의 downsize 효과를 고려하여 다중경로 펄스 (Multipath pulse)를 구한다. 그 다음 원래 펄스와 다중경로 펄스를 합친 변형 펄스를 구한다. 펄스의 timing은 원래 펄스의 amplitude 50% 지점을 기준으로 측정되므로, Fig. 5에서 보듯이 0.5 값을 가지는 time domain의 x값을 구하게 된다. 원래 펄스와 변형 펄스 각각에서 amplitude 0.5 값에 해당하는 time x를 x_1, x_2로 구하게 되고, 이 x_1과 x_2의 차이가 timing error가 됨을 알 수 있다. 또한 DME 펄스는 4개의 요구조건, rise time, fall time, width, power pass가 있다. 이는 Table 2에 자세히 나와있으며 이를 penalty로 정해서 reward에 반영하였다. 앞서 설명한 것을 기반으로 Agent가 Environment으로부터 받게 되는 reward는 Eq. (1)과 같다.

Table 2. DME ground transponder pulse shape requirements.

Pulse shape parameters	Range
Rise time	2.5 (+0.5, -1.0) μs
Pulse top	No instantaneous fall below a value which is 95% ofthe maximum voltage amplitude of the pulse.
Pulse duration (width)	3.5 (±0.5) μs
Fall time	2.5 (±0.5) μs

$$r = -(timing\ error) \times e^{penalty \times 0.5} \quad (1)$$

이를 기반으로 Agent는 시간 t에 대해서 할인율을 고려한 누적 reward, 반환 값 (Returns)을 $G_t = \sum_{i=t}^{T} \gamma^{(i-t)} r(s_i, a_i)$을 계산하게 되고 이는 학습되는 정도의 지표로 삼을 수 있다. 각 timestep의 reward값의 크기로도 Agent의 행동의 적절성을 판단할 수 있지만, episode내의 전체 timestep의 reward의 누적합이 클수록 더 좋은 정책(행동)이라고 볼 수 있으므로 episode의 반환 값이 Agent 정책의 최적성 판단에 사용된다.

Agent의 정책신경망은 Gradient Descent optimization 기법으로 최적화를 하기 때문에 이를 고려하여 음의 값으로 바꿔주었고 penalty는 비선형적으로 가중시키기 위해 exponential 함수를 이용하였다. 만약 펄스 요구조건들 중 2개를 만족하지 못한다면 reward는 $-(timing\ error) \times e^{2\times0.5} = -(timing\ error) \times e^{1}$이 된다.

$\mathcal{P}$는 현재 state와 선택한 action으로부터 다음 state가 무엇이 될 것인지에 대한 확률 공간이다. Transition probability라고도 하는 $p(s_{t+1} \mid s_t, a_t)$는 deterministic한 Environment에서는 1이다. 또한 max timestep을 300으로 설계하여 done signal은 max timestep이 되었을 때 True로 반환된다. Done signal이 1번 True

Table 3. MDP for DME pulse design.

	Gym space / Data type	Description
$\mathcal{S}$	Box (low = 0, high = 1, shape=(N_SEED_POINTS ,), dtype = np.float16)	N_SEED_POINTS = 61 Amplitude values for time domain (-6 ~ 6 μs)
$\mathcal{A}$	Box (low = 0, high = 1, shape = (1 + N_SELECT_POINTS,), dtype = np.float16)	N_SELECT_POINTS = 3 Select the start point and make 3 points group. Each point has different degree of change of amplitude.
$\mathcal{P}$	1 (Deterministic environment)	$p(s_{t+1} \mid s_t, a_t)$
$\mathcal{R}$	$-(timing\ error) \times e^{penalty \times 0.5}$	Measure timing error and add penalty term for pulse requirements. Make negative value to use gradient descent optimization method.
γ	0.99	Close to 1
Done	True/False	When max timestep 300 is reached.

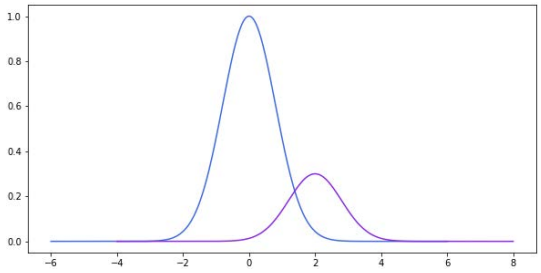

Fig. 6. In phase multipath effect (blue: original pulse, purple: multipath pulse).

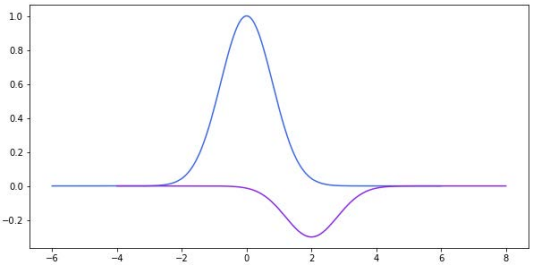

Fig. 7. Out phase multipath effect (blue: original pulse, purple: multipath pulse).

가 되었을 때 episode가 1개 완성된 것이며 timestep은 0으로 초기화 되어 다른 새로운 episode를 시작한다. 이와 같이 설계한 MDP를 정리하면 Table 3과 같다. 이때 유한한 timestep을 가지기 때문에 시간 할인율 γ을 1에 가까운 값인 0.99로 설정했다.

2.2 Pulse Environment

2.1절에서 정리한 MDP를 기반으로 Gym 형식으로 PulseEnv를 설계하였다. Environment 내의 다중경로 효과는 time delay를 2 μs를 주었고 펄스의 amplitude는 기존의 30%로 설정했다. 다중경로 효과는 다중경로 펄스의 위상과 원래 펄스의 위상차이로 in phase와 out phase로 나눌 수 있다. 원래 파장의 위상과 다중경로 펄스의 위상이 같으면 in phase이며 (Fig. 6), 위상이 180° 차이가 나게 되면 out phase이다 (Fig. 7). 이렇게 설정된 다중경로효과를 고려한 변형 펄스는 원래 펄스와 다중경로 펄스를 합친 다음 amplitude 값의 최소값을 0 최대값을 1로 표준화하여 형태가

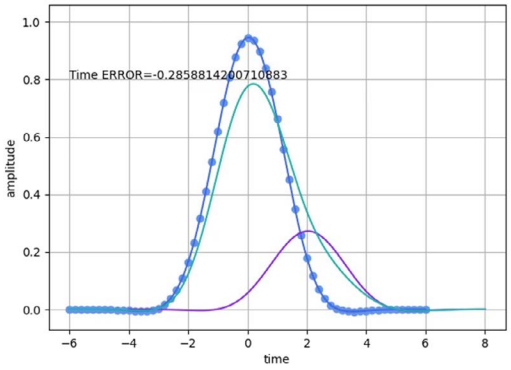

Fig. 8. Original pulse (blue), multipath pulse (purple), and join pulse (green).

완성되게 된다. Fig. 8에서 원래 펄스는 파란색, 다중경로 펄스는 보라색, 변형 펄스는 초록색으로 각각 나타냈다.

PulseEnv에서 SAC Agent가 학습하는 과정을 도식화하면 Fig. 9와 같다. 먼저 펄스의 초기 형태를 Gaussian pulse 형태와 SFOL pulse 형태, 2가지 방법으로 각각 초기화 시켜 실험을 진행하였다. 초기화된 펄스의 형태에서 시작하여 펄스 Agent는 action을 하게 되고 취해진 action을 기반으로 새로운 state, 즉 새로운 펄스 형태가 만들어지게 된다. 이 새로운 펄스 형태를 가지고 PulseEnv에서 timimg error와 pulse requirements를 각각 판단하여 reward를 계산하고, 이러한 과정으로 만들어진 reward와 새로운 state를 다시 펄스 Agent에게 주게 된다.

2.3 SAC Algorithm Agent

SAC 알고리즘 (Haarnoja et al. 2019)은 Actor-Critic 계열의 알고리즘이다. Fig. 10에 있는 SAC의 구조를 보면 정책 신경망 네트워크를 가진 Actor와 가치함수 신경망 네트워크를 가진 Critic로 Agent가 이루어져 있다. Critic은 현재 상태s_t와 선택한 action a_t를 input으로 하며 행동 가치인 Q_θ를 output으로 출력하는 신경망이다. Actor 신경망에서는 input으로 현재 상태s_t를 입력 받고 output으로 a_t를 반환하여 Agent가 어떤 action을 선택하게 되는지 결정하게 된다. 또한 SAC는 model-free 알고리즘으로 Environment의 구성요소 transition probability인 $\mathcal{P}^a_{ss'}$에 대한 별도의 모델링이 없이 적용가능 하다. 마지막으로 off-policy이기 때문에 on-policy 알고리즘에 비해 data efficiency가 높아 효과

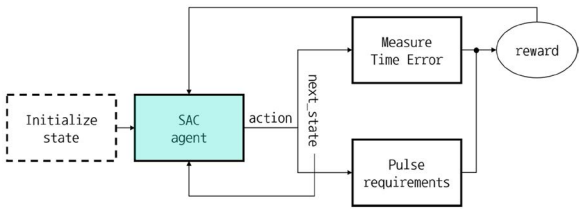

Fig. 9. DME pulse design system structure.

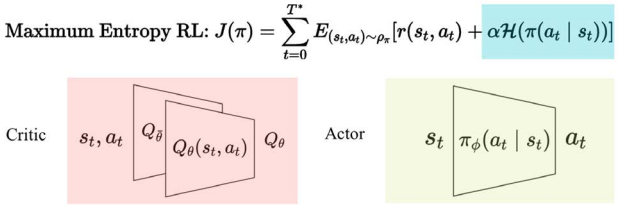

Fig. 10. Structure of SAC algorithm.

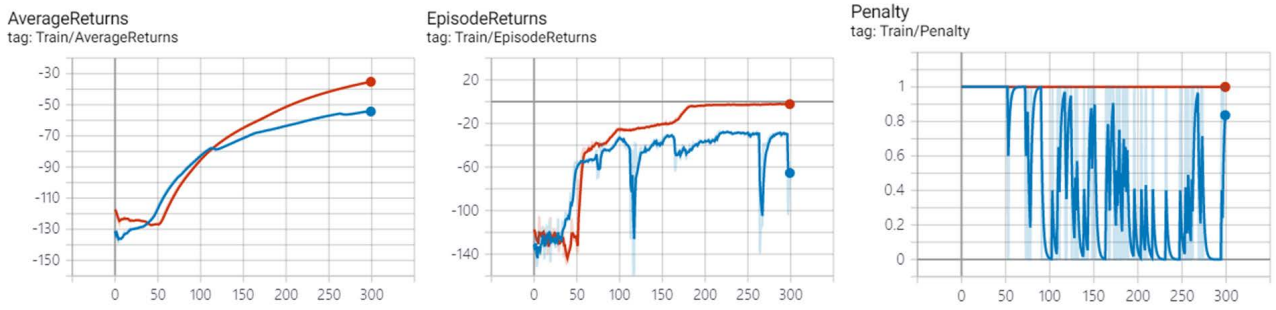

Fig. 11. Gaussian initialization and smooth factor 0.95 (red), SFOL initialization and smooth factor 0.95 (blue).

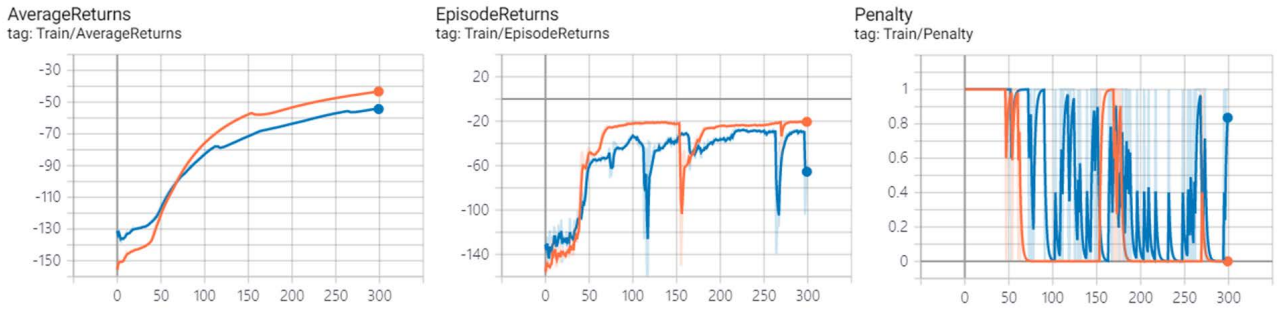

Fig. 12. SFOL initialization and smooth factor 0.85 (orange), SFOL initialization and smooth factor 0.95 (blue).

적인 학습을 가능하게 해주는 알고리즘이라고 할 수 있다.

SAC 알고리즘은 기존의 심층 강화학습 알고리즘들과 다르게 Maximum Entropy 개념을 강화학습 알고리즘에 도입함으로써 object function J(π)에 entropy term인 αH(π(a_t | s_t))을 추가하여 정책끼리 비슷한 continuous space에서도 exploration을 잘 하게 만들 수 있었고 노이즈에 대해서도 강건한 최적 정책을 구할 수 있다. 이러한 SAC 알고리즘으로 Agent를 설계함으로써 state와 action space가 큰 차원인 DME 펄스 디자인 문제에도 적용할 수 있었다.

3. 연구 결과

In phase 다중경로 효과 실험은 초기화 상태를 SFOL 펄스로 할 것인지, Gaussian 펄스로 할 것인지에 대한 구분과 smooth factor의 값을 0.85으로 하였는지, 0.95로 하였는지에 대한 구분을 하여 진행하였다. 측정 지표로는 Average Returns, Episode Returns, Penalty이 있다. Average Return은 Agent가 PulseEnv에서 정의된 max timestep으로 설정한 episode 단위와 상관없이 학습하고 있는 모든 episode들을 합쳐서 받은 반환 값 G_t를 지금까지 학습한 episode 총 수로 나누어 평균을 낸 값이다. Episode return은 episode 단위로 Agent가 받은 반환 값을 평균을 낸 값이다. 마지막으로 Penalty는 펄스 요구조건 4개 중에 만족하지 못한 개수를 나타낸 값이다. 심층강화학습 알고리즘의 목표는 반환 값을 최대화하는 최적 정책을 찾는 것이 목표이므로, Average Returns, Episode Returns는 학습이 진행됨에 따라 상승하는 경향을 보이는 것이 학습이 잘 되고 있다는 지표로 볼 수 있으며, Penalty는 펄스 요구 조건이 반영된 값이므로 모든 요구 조건을 만족시켰을 때의 값인 0에 가까울수록 좋다.

Figs. 11과 12의 각각 y축에는 Average Returns, Episode Returns, Penalty를 x축에는 학습한 episode 수로 나타낸 것이다. Fig. 11에서는 smooth factor는 0.95로 고정시키고 빨간색 그래프는 Gaussian 펄스 형태, 파란색 그래프는 SFOL 펄스 형태로 초기화 하여 학습하였다. 실험 결과 2개의 그래프 모두 Average return과 Episode return에서 상승하는 경향을 대체적으로 잘 보여주었으나, SFOL 펄스의 초기화 경우에 좀 더 불안정하고 return 값이 더 낮은 값으로 수렴하는 결과를 보여주었다. Penalty는 2개의 경우 모두 rise time과 width 조건을 잘 만족하지 못하는

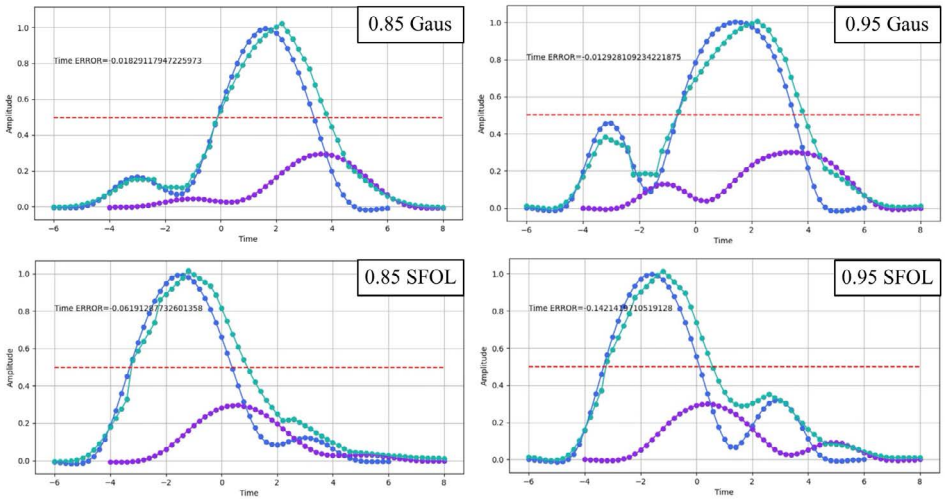

Fig. 13. In phase results, original pulse (blue), multipath pulse (purple), and join pulse (green).

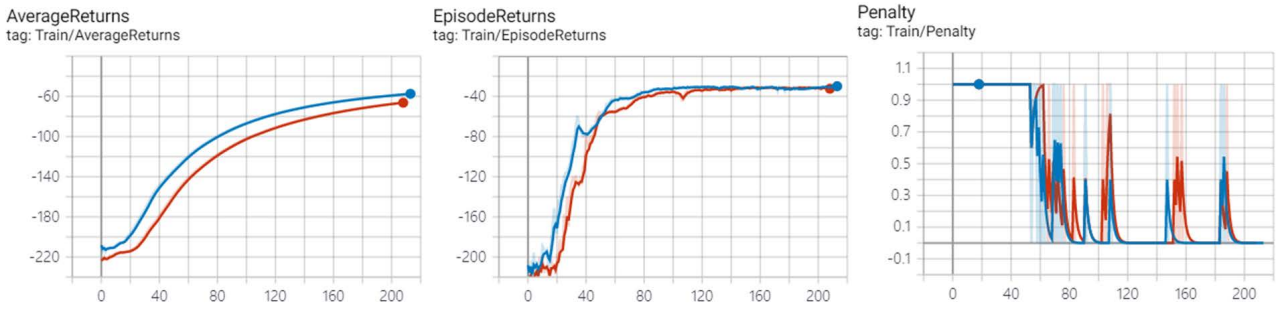

Fig. 14. 1.0 μs time delay, SFOL initialization (blue), Gaussian initialization (red).

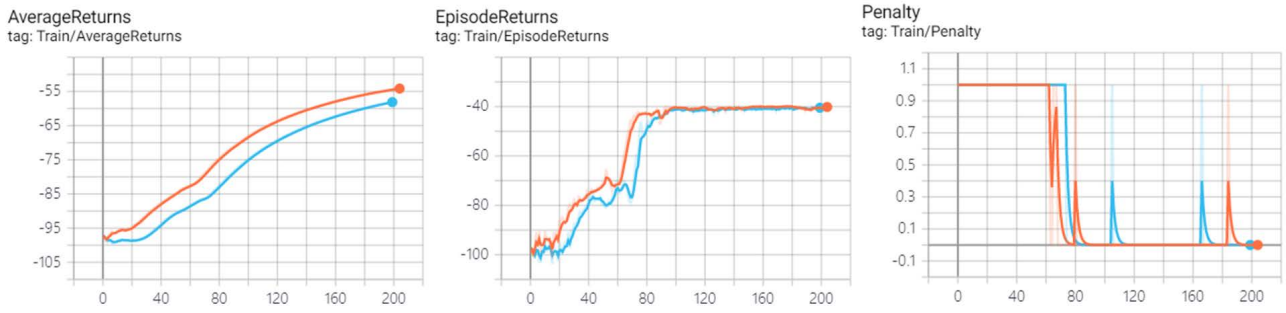

Fig. 15. 0.4 μs time delay, SFOL initialization (orange), Gaussian initialization (skyblue).

것으로 결과가 나왔다. 학습 후 완성된 DME 펄스 형태는 Fig. 13 오른쪽 2개의 그림에서 볼 수 있다. Gaussian 펄스 초기화의 경우 reward가 −0.0129, SFOL 펄스 초기화의 경우 −0.1421로 Gaussian 펄스 초기화의 경우가 더 reward 값이 높음을 알 수 있다.

다음으로 Fig. 12에서는 SFOL 펄스 초기화로 고정하고 smooth factor를 0.85, 0.95로 나누어 학습시켜보았다. Smooth factor가 0.85인 경우가 return 값도 더 높았고 penalty 값도 학습이 진행됨에 따라 줄어드는 것으로 보아 펄스 요구 조건도 잘 충족시키는 것으로 확인할 수 있었다. 학습 완료 후 DME 펄스의 형태는 Fig. 13의 하단 2개의 그림에서 확인할 수 있으며, smooth factor가 0.85일 때는 −0.0619, smooth factor가 0.95일 때는 −0.1421로 reward 값이 나왔다.

Out phase의 경우 smooth factor를 0.85로 고정하고, 다중경로 효과에서 고려되는 time delay 값을 1.0 μs와 0.4 μs, 초기화 셋팅을 Gaussian/SFOL 각각 나누어 학습시켰다. 변인으로 설정한 time delay의 값은 앞선 선행연구에서 in phase/out phase 각각의 경우에서 다중경로 효과에 취약하다고 판단되었던 값을 차용하여 설정하였다. Time delay를 Fig. 14에서는 1.0 μs로, Fig. 15에서는 0.4 μs로 각각 설정하고 실험하였다. 먼저 Fig. 14에서의 time delay가 1.0 μs인 학습 그래프를 보면, Return 그래프가 안정

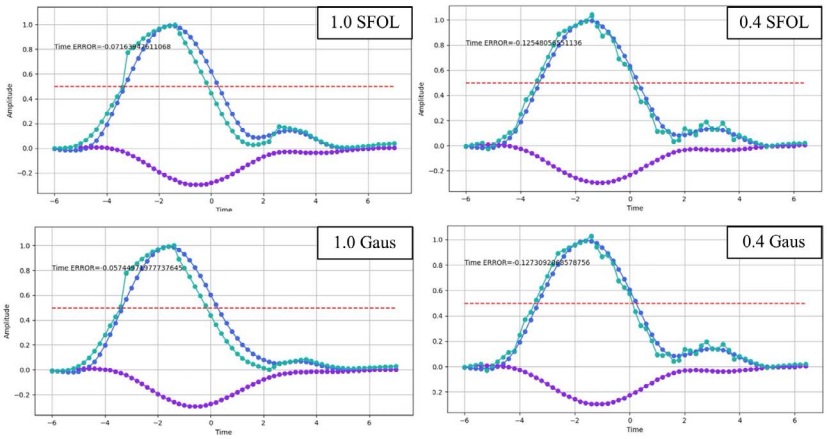

Fig. 16. Out phase results, original pulse (blue), multipath pulse (purple), and join pulse (green).

적으로 수렴하는 형태를 그리고 penalty도 학습이 진행됨에 따라 줄어들고 있는 것을 확인할 수 있다. 이때 대체적으로 만족하지 못하는 펄스 조건은 width와 fall time임을 알 수 있었다. 이 학습 셋팅에서 학습한 후 완성된 펄스의 형태는 Fig. 16의 왼쪽 2개의 그림에서 확인할 수 있으며, SFOL 초기화일 때는 −0.0716, Gaussian 초기화일 때는 −0.0574로 reward 값이 나왔다.

다음으로 Fig. 15에서 진행된 학습은 time delay가 0.4 μs인 경우, 각각 SFOL과 Gaussian 펄스로 초기화하여 진행하였다. 2가지 경우 모두 비슷하여 학습이 수렴하고 펄스 요구조건도 잘 충족하고 있음을 알 수 있었다. Fig. 16에서 오른쪽 2개의 그림에서 학습된 후 DME 펄스의 형태를 볼 수 있으며, SFOL 초기화일 때는 −0.1254, Gaussian 초기화일 때는 −0.1273으로 reward 값이 나왔다. in phase의 경우와는 다르게 out phase의 경우 4가지 서로 다른 PulseEnv 셋팅에 대해서 비슷한 펄스 형태를 나타내고 있음을 알 수 있었다.

4. 결론

Table 2에서 정리했던 DME 펄스는 4개의 요구조건에 대하여 새로 개발된 펄스의 형태가 만족하는 지를 실험을 통해 확인해 보면 In phase와 Out phase 2가지 경우의 실험 모두에서 width 조건이 잘 만족되지 않았음을 확인할 수 있었으며, 각각의 경우에 대해 추가적으로 In phase 경우에서는 rise time 조건이, Out phase 경우에서는 fall time 조건이 충족되지 않는 것을 Figs. 11, 12, 14, 15의 penalty 그래프를 통해 확인할 수 있었다. 이번 연구에서는 모든 4가지 요구 조건들에 대하여 구분없이 부합하지 못하는 조건들의 정황만 파악하고 요구 조건 별 심층파악을 하지 않았기 때문에 이러한 충족하지 못한 조건들에 대해서는 추후 실험 조건 개선과 학습을 진행하며 표준에 부합하는 펄스를 개발하는 방향으로 연구를 진행할 예정이다.

딥러닝과 강화학습을 결합한 심층 강화학습, DRL을 이용한 새로운 DME 펄스 형태 개발 프레임 워크를 제안하고 새로운 펄스 형태들을 살펴보았다. DME 펄스 디자인을 다중경로 효과와 DME 펄스 요구조건을 고려한 PulseEnv를 설계하고, 심층강화학습 알고리즘인 SAC을 이용하여 연구하였다. In phase, out phase, 서로 다른 상태 초기화 조건 등 다양한 다중 경로 효과 조건에서 심층 강화학습 방법론이 DME 펄스 디자인의 솔루션이 될 수 있다는 가능성을 확인해 볼 수 있었다.

본 논문에서 도출한 펄스들은 2개의 피크를 가지고 있어 실제 DME 장비에서 사용될 수 있는 형태가 아닌 초기 연구결과이며 이는 향후 연구에서 개선될 것이다. 그 외 추후 연구 방향으로는 현재 penalty로만 고려한 펄스 요구 조건을 더 엄격하게 만족시키며 본래의 목적인 다중경로 효과에 의한 time error도 더 감소시킬 수 있는 방안을 고려하는 것이다. 또한 본 연구에서는 학습되는 시간과 신호 송수신에서 중요한 부분인 환경변화와 실시간성에 대한 고려가 빠져 있음을 인지하여 이 방향으로도 연구할 예정이다. 본 연구에서는 하나의 multipath 신호에 대하여 특정 시간차와 amplitude를 한정하여 연구가 수행되었으나 만약 다른 형태의 multipath 신호가 있거나, 혹은 복수의 multipath 신호와 결합한 경우에는 Agent가 학습하는 환경이 더 계층적이고 다양하므로 본 연구의 방법을 적용했을 때는 한계점이 있을 것이다. 따라서 이후 연구로는 다양한 다중경로 효과 조건들에 대해서 일반화된 Agent나 계층적 학습을 하는 Agent로 다양한 환경에서도 다중 경로효과 문제를 해결할 수 있는 솔루션을 찾는 것이 이후 연구방향이 될 것이다.

ACKNOWLEDGMENTS

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었습니다 (과제번호 21TBIP-C155921-02).

AUTHOR CONTRIBUTIONS

Evaluation of DME pulse shape excellence K.E.; validation of learning models K.E.; methodology, K.E. and L.J.; investigation, L.J.; writing—original draft preparation and editing, L.J.; visualization, L.J.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Badia, A., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., et al. 2020, Agent57: Outperforming the Atari Human Benchmark. https://arxiv.org/abs/2003.13350

Federal Aviation Administration (FAA) 2016, Performance-Based Navigation (PBN) National Airspace System (NAS) Navigation Strategy 2016; U.S. Department of Transportation: Washington, DC, USA

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., et al. 2019, Soft Actor-Critic Algorithms and Applications. https://arxiv.org/abs/1812.05905

Kim, E. 2012, Investigation of APNT Optimized DME/DME Network Using Current State-of-the-Art DMEs: Ground Station Network, Accuracy, and Capacity, In Proceedings of the IEEE/ION PLANS 2012, Myrtle Beach, SC, USA, 24–26 April 2012. https://doi.org/10.1109/PLANS.2012.6236876

Kim, E. 2013, Alternative DME/N pulse shape for APNT, In Proceedings of the 2013 IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC), East Syracuse, NY, USA, 5–10 Oct. 2013. https://doi.org/10.1109/DASC.2013.6712591

Kim, E. 2017, Improving DME Performance for APNT Using Alternative Pulse and Multipath Mitigation, IEEE Trans. Aerosp. Electron. Syst., 53, 877-887. https://doi.org/10.1109/TAES.2017.2667058

Kim, E. & Seo, J. 2017, SFOL Pulse: A High Accuracy DME Pulse for Alternative Aircraft Position and Navigation, Sensors, 17, 2183. https://doi.org/10.3390/s17102183

Lilley, R. W. & Erikson, R. 2012, DME/DME for Alternative Position, Navigation, and Timing (APNT), APNT White Paper, Department of Transportation: Washington, DC, USA, 2012.

Jungyeon Lee is currently an undergraduate student in the Department of Mechanical & System Design Engineering, Hongik University, South Korea. Her current research interests include reinforcement learning and robotics.

Euiho Kim received the Bachelor's degree from the Department of Aerospace Engineer- ing, Iowa State University, Ames, IA, USA, and the Ph.D. and Master's degree from the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, USA. He is currently an Assistant Professor in the Department of Mechanical & System Design Engineering, Hongik University, South Korea. Prior to this, he was a Research Associate in the Department of Aerospace Engineering, University of Kansas; and the Technical Lead of the ground-based augmentation system of GPS and FAA's alterna- tive position, navigation, and timing programs. His current research interests include satellite-based navigation, aircraft navigation using ground nav-aids, indoor navigation, and robotics.