한국시뮬레이션학회 논문지
Vol. 29, No. 1, pp. 39-46 (2020. 3)
http://doi.org/10.9709/JKSS.2020.29.1.039
ISSN 1225-5904

깊은강화학습 기반 1-vs-1 공중전 모델링 및 시뮬레이션

문일철† · 정민재 · 김동준

Modeling and Simulation on One-vs-One Air Combat with Deep Reinforcement Learning

Il-Chul Moon† · Minjae Jung · Dongjun Kim†

ABSTRACT

The utilization of artificial intelligence (AI) in the engagement has been a key research topic in the defense field during the last decade. To pursue this utilization, it is imperative to acquire a realistic simulation to train an AI engagement agent with a synthetic, but realistic field. This paper is a case study of training an AI agent to operate with a hardware realism in the air-warfare dog-fighting. Particularly, this paper models the pursuit of an opponent in the dog-fighting setting with a gun-only engagement. In this context, the AI agent requires to make a decision on the pursuit style and intensity. We developed a realistic hardware simulator and trained the agent with a reinforcement learning. Our training shows a success resulting in a lead pursuit with a decreased engagement time and a high reward.

Key words : Artificial Intelligence, Dog fighting, Air Warfare, Reinforcement Learning

요 약

인공지능(AI)를 교전상황에 활용하는 것은 최근 10년간 국방 분야의 주요 관심사였다. 이러한 응용을 위해서, AI 교전 에이전트를 훈련해야 하며, 이를 위해 현실적인 시뮬레이션이 반드시 필요하다. 하드웨어 차원의 현실성을 가진 공중 무기체계 공중전 모델에서 AI 에이전트를 학습한 사례에 대해서 본 논문은 서술하고 있다. 특히, 본 논문은 기총만을 활용하는 공중전 상황에서 적을 어떻게 추적해야하는지 AI를 학습하였다. 본 논문은 현실적인 공중전 시뮬레이터를 작성하여, 에이전트의 행동을 강화학습으로 수행한 결과를 제시한다. 훈련 결과로는 Lead 추적을 활용하여 단축된 교전시간과 높은 보상을 갖는 에이전트의 학습에 성공하였다.

주요어 : 인공지능, 공중전, 공중 무기체계, 강화학습

1. 서론

인공지능(AI)을 교전상황에 활용하는 것은 최근 10년간 국방 분야의 주요 관심사였다. 인공지능(AI)의 활용을 위해서, AI 교전 에이전트를 훈련해야 하며, 이를 위해 현실적인 시뮬레이션이 반드시 필요하다. 현실적인 시뮬레이션은 다양한 방면에서 구성될 수 있는데, 작전 수립 및 교전 과정의 현실성, 장비체계의 현실성, 사기 및 심리 묘사의 현실성 등 다양한 차원의 현실성이 존재한다.

본 논문은 위의 현실성 문제를 장비체계의 현실성의 문제로 한정하여 연구를 진행하였다. 장비체계의 현실성을 다루기 위해서는 교전체계에서 활용되는 장비의 범위를 정하고, 정해진 장비의 모델링 해상도(Resolution)을 결정해야한다. 여기서 모델링 범위 및 해상도의 결정은 공중전 교전의 충실성을 담보할 수 있는 차원인지가 주요 의사 결정 인자가 되며, 이러한 의사 결정은 분야 전

* 본 연구는 국방과학연구소의 지원(계약번호: UD160081BD)으로 수행된 위탁과제 결과의 일부이며, 지원에 깊이 감사드립니다.

Received: 12 December 2018, **Revised**: 27 December 2019, **Accepted**: 27 December 2019

† **Corresponding Author**: Il-Chul Moon
E-mail: icmoon@kaist.ac.kr
KAIST
Industrial and Systems Engineering

문가 및 시뮬레이션 모델러가 협의하여 결정하게 된다.

본 논문은 하드웨어 차원의 현실성을 가진 공중 무기체계 공중전 모델을 수행하고 있다. 무기체계 간의 교전을 모델링할 때, 모델링 요소를 산출하기 위한 다양한 방식이 있을 수 있으나, 교전 행동 과정을 정의하고 그 교전 행동에서 활용되는 개체를 인지하여 이를 모델링 하는 과정도 논리적인 접근이다. 우리는 Observe, Orient, Decide, Act로 정의되는 OODA루프를 기반으로 교전 행동 과정을 가정하였다(Breton 등, 2005). OODA루프는 1) Observe 과정에서 탐지 모델, 2) Orient 과정에서 기동 모델, 3) Decide 과정에서 의사 결정 모델, 4) Act 과정에서 교전 행동 및 피해 평가 모델을 필요로 한다. 본 논문에서 활용된 시뮬레이터는 위에서 언급된 모델들을 분야 전문가와 협의하여 필요한 수준으로 모델링 해상도를 정하여 시뮬레이션을 구축하였다.

강화학습은 가치함수 기반 기법과 정책함수 기반 기법으로 나뉜다. 이 중 연속적인 행동 공간(continuous action space)을 학습하는 데에 유리한 정책함수 기반 기법을 선택했다. 정책함수 기반 기법 중 정책을 결정적 함수로 만들고 이의 기울기를 구하는 기법을 결정론적 정책 기울기(Deterministic Policy Gradient)라고 한다(Silver 등, 2014). 여기서 정책함수를 인공신경망으로 근사하는 방법을 깊은 결정론적 정책 기울기(Deep Deterministic Policy Gradient)라고 한다(Lillicrap 등, 2016). 본 논문은 결정론적 정책 기울기, 줄여서 DDPG를 이용해 기체의 행동 정책을 학습하였다.

본 논문은 위와 같이 구성된 공중전 시뮬레이션에서 AI 에이전트를 학습하여 기총 기반 공중전 교전을 수행하였다. 이 상황에서 AI 에이전트는 추적 위치를 결정해야 한다. 현실에서는 위와 같은 추적을 피아간의 상호 위치를 인지하여 조종사가 적절하게 기체를 조정하여 적을 추적한다. 이 실험의 목적은 위와 같은 사람의 의사 결정을 AI 에이전트가 효과적으로 재현해 낼 수 있는지 점검한다.

상태는 아군기와 적군기의 개별 개체의 기동 정보와 상호 기동 정보로 이루어져 있다. 여기서 개별 개체의 기동 정보는 지면좌표계에서의 값뿐만 아니라 아군기좌표계, 적군기좌표계에서의 값도 상태로 넣었다. 그리고 적군기좌표계에서 추적할 위치를 행동으로 넣었다. 마지막으로 적군기를 격추시켰을 때의 큰 보상뿐만 아니라 위험 영역 반경과 상관있는 변수들로 작은 보상을 정의했다. 본 논문의 실험 결과로 단축된 교전시간과 높은 보상을 갖는 에이전트의 학습에 성공하였다.

2. 기존 연구

공중전 교전을 위한 다양한 시뮬레이션 연구가 진행되었다. 예를 들어, 박현주 등(2015)에 따르면, Score 기반의 교전 방법론이 연구된 사례가 있는데, 이는 피아간의 위치 관계를 Pay-off 행렬로 모델링을 한 사례로 게임이론에 기반을 두었다고 볼 수 있다.

이동진 등(2009)에 따르면, 공중전 교전을 강화학습을 이용하여 모델링을 한 사례도 있으며, 이는 정밀한 기체 기동 정보 모델링을 하였지만, 강화학습의 방식이 기존에 흔히 활용되는 Markov Decision Process기반의 강화학습이어서, 인공신경망 기반의 교전 강화학습 모델링의 사례는 찾을 수 없었다.

해외의 경우 공중전에 인공지능을 부여한 사례들이 더욱 많이 존재한다. 예를 들어, Ernest 등(2016)에 따르면 단순 Fuzzy 로직 기반의 인공지능만으로도 사람과의 대결에서 승리한 공중전 인공지능 모델의 사례도 존재한다.

이렇게 인공지능을 활용하여 최적 교전을 수행하는 사례는 점차 증가하고 있으며, 최근에는 깊은 강화 학습(Deep reinforcement learning) 기반의 교전 사례도 발표되고 있다. 예를 들어, Toghiani-Rizi 등(2017)에 따르면 컴퓨터 생성군(Computer Generated Force)이 인공지능으로 통제되어 지상 작전을 수행하는 사례가 발표되었다.

3. 시뮬레이션 모델

위에서 논의한 바와 같이 본 논문의 시뮬레이션은 OODA 루프를 기반으로 하여, 탐지, 기동, 의사 결정, 교전, 피해평가로 구현 모델을 식별하였다. 본 절은 각 모델의 구현에 대한 내용을 소개한다.

3.1 기체 모델

그림 1은 시뮬레이션으로 구현된 한 기의 공중 교전 기체의 구조를 나타내고 있다. 이 구조는 아래와 같은 하위 컴포넌트 모델을 가지고 있다.

* Aircraft Structure
: 기체 구조를 모델링하여 Radar Cross Section (RCS)을 제시함
* Radar Warning Radar
: 적 레이더의 탐지 신호를 받아 조종사에게 경고함
* Aircraft Maneuver
: 기체를 원하는 위치 혹은 자세각으로 제어함

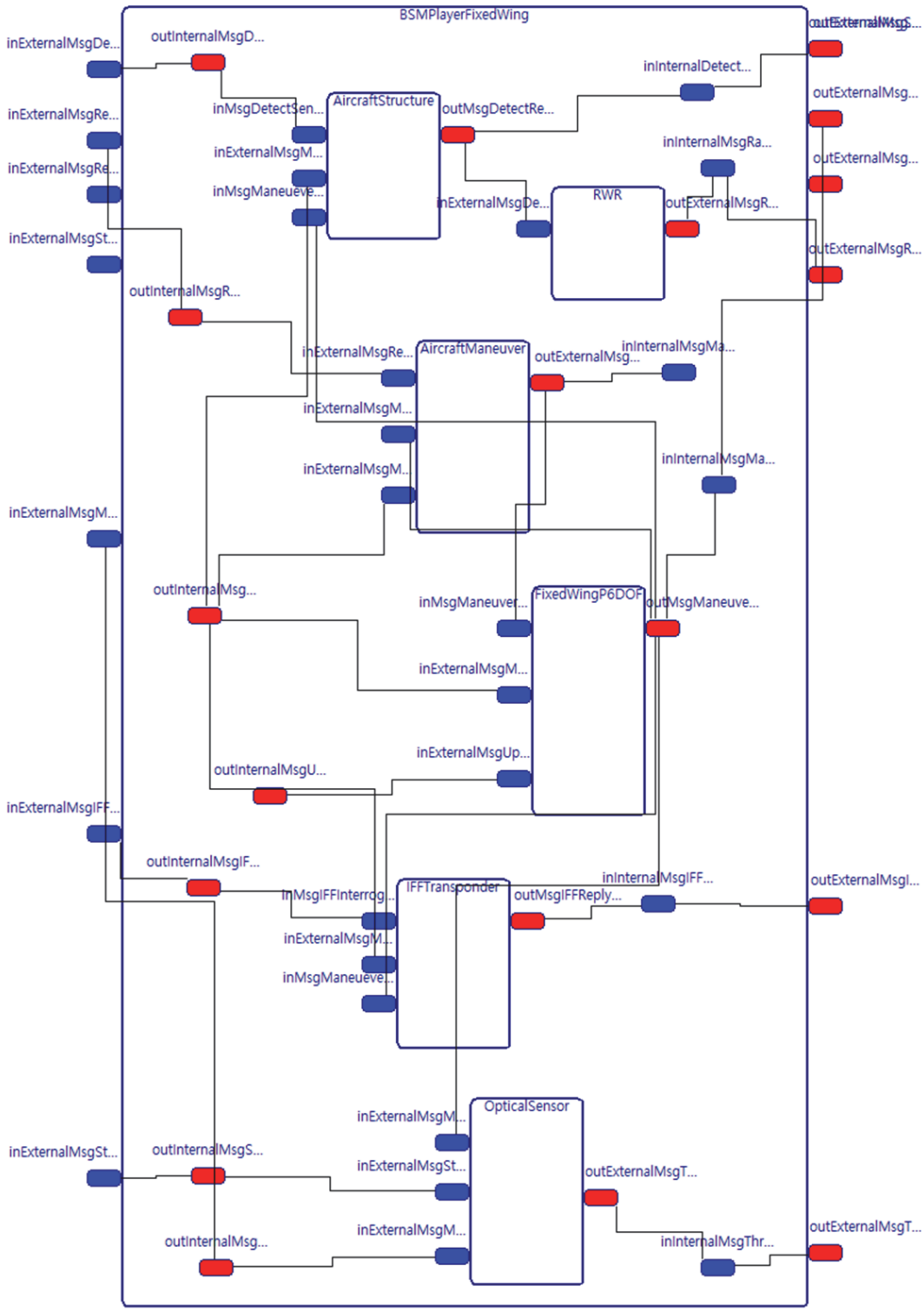

Fig. 1. Components and their message passing structure of a single fixed-wing combat aircraft

* FixedWingP6DOF
: 기동 방정식을 의사 6자유도 수준으로 모델링함
* IFF Transponder
: 레이더 탐지시의 피아식별 장치를 모델링함
* Optical Sensor
: 가시거리 내의 주변 개체에 대한 정보를 수집함

위와 같은 구조 모델에서, 본 논문은 문제를 기총 기반 교전으로 제한하여 강화학습을 진행하므로, Optical Sensor, Aircraft Maneuver, FixedWingP6DOF가 핵심 모델이 된다.

Optical Sensor는 적기에 대한 피아 구분, 적기의 NED(North, East, Down) 좌표계에 대한 위치, Roll, Pitch, Yaw에 대한 자세각 정보를 제공한다. 이러한 정보 제공은 적기의 기동에 따른 정보를 메시지로 작성하여, 아군기의 Optical Sensor 컴포넌트에 전달하는 것으로 구현되었다. 이를 통하여, Optical Sensor를 통한 기총 교전 수준의 탐지를 모델링하였다.

3.2 고정의 의사6자유도 모델 및 기동 제어 모델

기체 모델의 FixedWingP6DOF는 고정익의 의사 6자유도 기동 모델을 컴포넌트로 구현한 것이다. 여기에서 쓰인 의사 6자유도는 아래와 같은 운동방정식을 가진다(Park 등, 2016). 위에 점이 찍혀 있는 변수는 순간변화율 즉, 시간에 대한 미분 값을 의미한다. 표 1은 각 변수를 정리한다.

Table 1. Variables of Pseudo 6DOF

No.	**Notation**	Variables
1	V_t	**동체 축에서의 속도 벡터 [m/s]**
2	T_t	**동체 축에서의 추력 벡터 [N]**
3	α	**받음각 [rad]**
4	γ_t	**상하각 [rad]**
5	ψ_t	**방위각 [rad]**
6	X	**관성 좌표계에 대한 X축 위치 [m]**
7	Y	**관성 좌표계에 대한 Y축 위치 [m]**
8	h	**관성 좌표계에 대한 고도 [m]**
9	L	**양력 벡터 [N]**
10	D	**항력 벡터 [N]**
11	W	**중력 벡터 [N]**
12	m_t	**동체 질량 [kg]**
13	g	**중력 가속도 [m/s^2]**
14	ϕ	**롤 자세각 [rad]**

$$
\begin{aligned}
\dot{V}_t &= (T_t - D)/m - g\sin\gamma_t \\
\dot{\gamma}_t &= \frac{(L + T_t\sin\alpha)\cos\phi - W\cos\gamma_t}{m_t V_t} \\
\dot{\psi}_t &= \frac{(L + T_t\sin\alpha)\sin\phi}{m_t V_t\cos\gamma} \\
\dot{X} &= V_t\cos\gamma_t\cos\psi_t \\
\dot{Y} &= V_t\cos\gamma_t\sin\psi_t \\
\dot{h} &= V_t\sin\gamma_t
\end{aligned}
$$

의사 6자유도를 통하여 기체를 제어한다고 하더라도, 기체가 원하는 지점까지 이동하게 하는 다양한 기동 방법론이 존재한다. 여기서 우리는 Line-of-Sight(LOS)을 활용하여 위에서 나열한 추력 T_t, 양력 L, 롤 각속도 $\dot{\phi}$의 제어를 수행하였다.

만약 x_a를 현재 기체의 NED 좌표계의 위치, x_t를 동일 좌표계의 목표점으로 가정한다면 $LOS = x_t - x_a$로 정의할 수 있다. 기체의 자세각(롤, 피치, 요)으로 계산되는 DCM(Direction Cosine Matrix)을 이용해 NED 좌표의 LOS를 기체 좌표계(Body Axis)의 LOS_{BA}로 변환시킬 수 있다.

$$LOS_{BA} = DCM \bullet LOS$$

PG 알고리즘으로 움직이기 위해서 가속도를 LOS_{BA}의 방향으로 주면 된다. 기체 좌표계에서의 이동 속도 벡터를 V라고 정의하였을 때 수식 1로 가속도 a를 구할 수 있다. V와 LOS_{BA}사이의 각도를 ϕ로 뒀을 때 a의 크기는 $\| V\| \sin\phi$로 속력과 ϕ 각도에 비례한다.

$$a = \frac{(V \times LOS_{BA}) \times V}{\| V\| \ \| LOS_{BA}\|} \tag{1}$$

여기서 제시된 가속도의 값은 비례항, 적분항, 미분항을 이용하는 PID(Proportional Integral Differential) 제어를 통해 추력, 양력, 롤 각속도 값을 조절할 수 있게 한다.

3.3 기총 모델 및 피해 평가 모델

본 교전은 기총만을 활용한 교전을 수행하므로, 기총에 대한 모델만을 상세히 기술한다. 기총의 개별 총알의

탄도 궤적을 모델링 하는 것이 정밀한 물리 모델을 반영하는 방법이나, 이런 경우 수천 개의 탄환 궤적을 생성해야 한다는 문제점이 발생한다. 이를 극복하기 위해, 총알이 산포되는 영역을 모델링하기로 하였으며, 이를 위험 영역으로 모델링하였다. 즉, 적 기체가 기총의 위험 영역에 진입한 상태에서, 기총의 발사 메시지가 발생한다면, 적이 피해를 입는 것으로 모델링을 하였다.

기총의 위험 영역을 모델링하기 위해서는 두 가지 파라미터가 필요하며, 이는 위험 영역 반경과 위험 영역의 원뿔 꼭지각이다. 원뿔의 중심축은 기체의 진행방향이며, 위에서 정의된 영역 반경과 꼭지각을 활용하여 적 기체는 피해를 입게 된다. 본 시뮬레이션의 기총 기반 교전에서 개별 기체는 적 기체가 위험 영역에 진입하면, 즉각 기총 사격 메시지를 발생시키는 것으로 모델링을 수행하였다. 이를 통하여, 기총을 활용한 공중전 교전에서 인공지능이 학습하는 영역은 피아의 위치를 감안한 추적 방식만으로 제한되게 된다. 본 논문은 위험 영역 반경은 2000m, 위험 영역의 원뿔 꼭지각은 0.1rad으로 설정했다.

4. 인공지능 모델

위에서 소개한 바와 같이, 본 논문은 강화학습 기반으로 에이전트의 학습을 수행하였다. 본 절은 에이전트의 학습 모델 구성에 대해 소개한다.

4.1 강화학습 상태 정의

강화학습 방법론 을 활용하기 위해서, 강화학습이 일어나는 상태와 행동을 정의하고, 상태(State) 및 행동(Action) 쌍(Pair)에 대한 보상(Reward)을 정의해야 한다(Sutton 등, 1998). 이 중에서 상태의 정의는 시스템 전체에서 강화학습 과정에 필요한 정보량을 식별하여 입력하는 첫 번째 작업이다.

기본적으로 기총 기반 공중전의 핵심은 교전 기하(Engagement Geometry)로 연구되는 “꼬리잡기 기동”이 핵심 행동이 된다. 그에 따라 상태 정보의 모든 변수는 피아간 기동 정보를 나타내는데 집중하였다.

피아 기동 정보는 개별 개체의 기동 정보와 상호 기동 정보로 구분할 수 있다. 개별 개체의 기동 정보는 위치, 속도로 나타낼 수 있으며, 본 연구는 기체의 기동에 대한 의사 6자유도 모델을 수행하기 때문에 자세각(롤, 피치, 요)도 상태 정보로 포함할 수 있다.

아군기가 적군기를 쫓을 때 절대적인 지면 좌표계보다 아군기 기체 좌표계에서의 위치, 자세각, 속도를 더 고려하게 될 것이다. 이는 적군기가 아군기를 회피할 때도 마찬가지이다. 따라서 본 논문은 개별 개체의 기동 정보를 상대의 좌표계로 변환시킨 값 역시 강화학습의 상태로 넣었다.

Table 2. Reinforcement learning state definition

No.	**State Variables**	Variable Dimensions
1	Player Position on NED coordinate (North, East, Down)	**3**
2	Player Position on Opponent coordinate (North, East, Down)	**3**
3	Opponent Position on NED coordinate (North, East, Down)	**3**
4	Opponent Position on Player coordinate (North, East, Down)	**3**
5	Player Attitude on NED coordinate (Roll, Pitch, Yaw)	**3**
6	Player Attitude on Opponent coordinate (Roll, Pitch, Yaw)	**3**
7	Opponent Attitude on NED coordinate (Roll, Pitch, Yaw)	**3**
8	Opponent Attitude on Player coordinate (Roll, Pitch, Yaw)	**3**
9	Player Velocity on NED coordinate (Velocity of North, East, Down)	**3**
10	Player Velocity on Opponent coordinate (Velocity of North, East, Down)	**3**
11	Opponent Velocity on NED coordinate (Velocity of North, East, Down)	**3**
12	Opponent Velocity on Player coordinate (Velocity of North, East, Down)	**3**
13	Player-Opponent Distance	**1**
14	Line of Sight Angle from Player to Opponent (Yaw, Pitch)	**2**
15	Line of Sight Angle from Opponent to Player (Yaw, Pitch)	**2**
16	Bearing Angle from Player to Opponent	**1**
17	Bearing Angle from Opponent to Player	**1**
	Total	**43 Variables**

상호 기동 정보는 피아간의 거리, LOS 각도, 베어링 각도(Bearing Angle)가 있다. LOS 각도는 LOS 벡터의 방향을 각도로 변환시킨 것이고 베어링 각도는 LOS 벡터와 기체의 정면 축이 이루는 각도를 말한다. 베어링 각도 θ의 계산은 수식2와 같다. LOS_{BA-N}은 LOS_{BA}의 N 방향 값이다.

$$\theta = \frac{LOS_{BA-N}}{\| LOS_{BA} \|} \quad (2)$$

위와 같은 기동정보들은 두 개체의 기동 정보 및 연관 관계에 대해 의사 6자유도 기반으로 생성되는 모든 데이터를 포괄하고 있다. 표 2는 상태로 사용한 변수들을 정리한 것이다

4.2 강화학습 행동 정의

교전시 적 추적의 방식은 Lead, Lag, Pure 추적의 형태로 이루어 질 수 있다(Shaw, 1985). 각 추적의 형태는 기동 상태에 따라서 선호가 달라질 수 있다. 만약 피아간의 베어링 각도가 0에 가깝다면, Pure 추적이 적절하다. 그러나 피아간의 베어링 각도가 0에서 차이가 난다면, 적을 앞서서 추적할 것인지(Lead Pursuit) 혹은, 적의 차후 선회가지 고려하여 추적할 것인지(Lag Pursuit)을 결정해야 한다.

본 논문은 더 나아가 적군기 좌표계에서의 NED 위치를 행동으로 정의했다. 그림 2처럼 행동이(N,E,D)로 정의 되었을 때 적군의 위치에서의 NED 위치를 Pursuit하게 된다. 이렇게 정의한다면 위의 추적 형태보다 더 풍부하게 행동을 사용할 수 있다. 각각의 좌표의 절댓값은 1000이하의 값을 가진다.

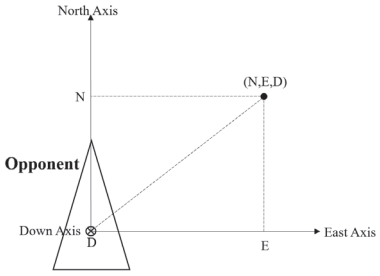

Fig. 2. Action definition on engagement model

4.3 강화학습 보상 정의

위험 영역 반경은 위험 영역의 원뿔 꼭지각은 각각 거리 D와 베어링 각도 θ에 대응시킬 수 있다. 따라서 보상(reward) r은 베어링 각도나 거리가 위험 영역 이하이면 보상에 양수의 영향을 주고 이상이면 보상에 음수의 영향을 주는 보상함수를 정의해야한다. 보상은 수식 3으로 정의했다.

$$r = \tanh(1-10*\theta) + \tanh(1-D/2000) \quad (3)$$

위의 보상은 상대편을 격추했을 때 제시하는 큰 보상(본 논문은 100000) 이외에 주어지는 상태에 따른 보상이다. 이와 같은 상태에 따라 연속적으로 주어지는 보상은 우리가 원하는 교전 기하의 모양으로 행동을 유도하게 되며, 이러한 기법을 보상 형성(Reward Shaping)이라 한다(Ng 등, 1999).

4.4 강화학습 모델 구성

$$\nabla_\theta J = E_{\pi_\theta}[\nabla_\theta \log \pi_\theta(a|s)\, Q_\phi^{\pi_\theta}(s,a)] \quad (4)$$

Policy Gradient Theorem에 의해 정책 기울기(Policy Gradient) $\nabla_\theta J$는 수식 4로 나타낼 수 있다. π는 정책이고 Q는 Q함수이다. θ와 ϕ는 각각 정책과 Q함수의 파라미터이다. DDPG(Deep Deterministic Policy Gradient)의 두 네트워크 Actor와 Critic은 각각 정책과 Q함수를 학습한다(Lillicrap 등, 2016). Actor는 정책 기울기로, Critic은 MSE(Mean Squared Error)로 학습을 진행한다. 자세한 학습 방법은 Lillicrap 등(2016)에 나와 있다.

Actor 네트워크는 입력을 상태로 가지고 출력을 행동 선택지로 가진다. Critic 네트워크는 입력을 상태와 행동의 조합으로 가지며 출력을 그에 따른 Q함수 값을 가진다. Actor는 결정론적인 정책을 사용하기 때문에 학습 과정에서는 소음을 주어 다양한 행동을 취할 수 있도록 한다. 각 교전 에피소드에서 얻어지는 행동과 상태를 조합하여 Actor와 Critic을 번갈아가며 학습하게 된다.

5. 실험 결과

본 논문에서 소개된 시뮬레이터와 강화학습 방법론을 활용하여 학습 및 교전을 실행하였다. 교전은 random한 위치에서 일정한 거리 20000m를 두고 서로 마주보는 상태로 시작한다. 학습을 위해 총 5000회의 시뮬레이션이 실행되었고 각각은 50000번의 timestep을 최대 시뮬레이션 시간으로 가진다. 100회의 에피소드마다 50번의 시뮬레이션을 통해 테스트를 거쳤다.

또한 교전 환경으로 아군기는 기총만으로 무장하였으며, 기총은 충분히 많은 탄환을 가지고 있어서 탄소모에 따른 행동 학습은 수행하지 않았다. 적기는 비무장인 상태에서 회피기동만 수행한다고 가정하였다.

그림 3과 그림 4는 100 교전 에피소드 동안 발생하는 교전 시간 및 보상의 변화이다. 각각의 그래프에서 선과

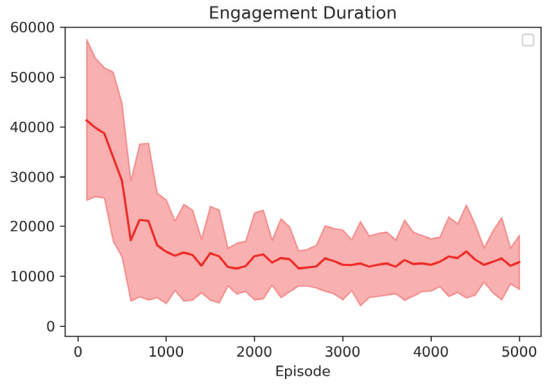

Fig. 3. Engagement duration over the training episodes

Fig. 4. Engagement rewards over the training episodes

이를 둘러싸는 범위는 평균과 표준편차를 의미한다. 시뮬레이션의 평균 교전 시간이 학습의 진행에 따라 줄어드는 것을 볼 수 있다. 이에 맞추어 평균 보상도 빠르게 높아지는 모습을 볼 수 있다.

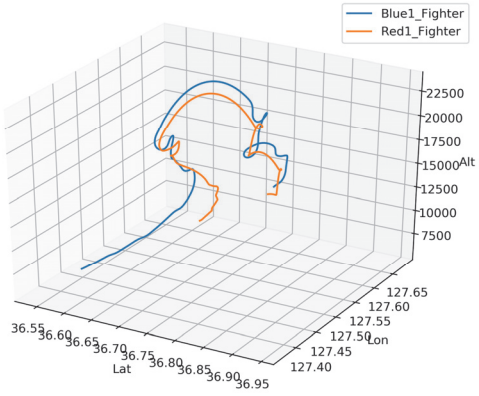

Fig. 5. Trajectory of two aircrafts before learning

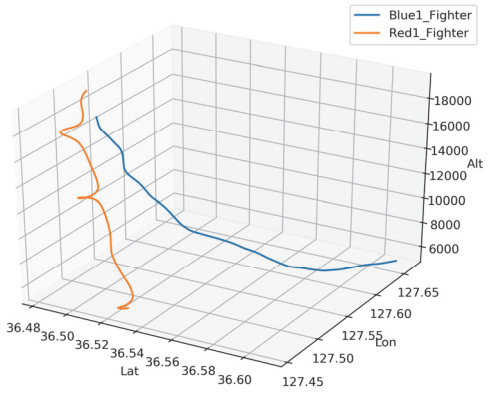

Fig. 6. Trajectory of two aircrafts after learning

이와 같은 학습 결과를 바탕으로, 학습이 완료된 결과물을 3D Plot을 이용해서 육안 관찰하였다(그림 5, 그림 6). 초기에는 적기의 격추가 관찰되지 않았는데, 학습 에피소드의 진행에 따라 적기의 격추가 가능해졌다.

6. 결론

본 논문은 강화학습 기반의 인공지능 공중전 사례 연구를 소개한다. 인공지능 공중전의 현실성을 높이기 위하여, 의사 6자유도 기반의 기동 모델을 구현하였다. 이와 같은 기동 모델을 바탕으로 기총 기반의 공중전 교전을 시뮬레이션 한다.

이후, 시뮬레이션과 통신하는 강화학습 모델을 구현하였다. 구체적으로 기총기반 공중전의 적기 추적 행동의 최적 의사 결정을 수행하는 DDPG 기반 강화학습을 수행하였다. 학습을 통해 좋은 성능의 AI 에이전트를 학습할 수 있음을 볼 수 있었다.

본 논문은 현실적인 공학 기반의 시뮬레이션에 어떻게 인공지능 기반 교전 의사 결정을 부가할 수 있는지에 대한 사례를 제시하며, 이러한 사례에 쓰인 방법론은 다양한 국방 시뮬레이션 및 인공지능 교전 모델에 쓰일 수 있다.

References

이동진, & 방효충. (2009). 강화학습을 이용한 무인전투기 (UCAV) 근접 공중전. 한국항공우주학회 학술발표회 초록집, 249-252.

박현주, 이병윤, 유동완, & 탁민제. (2015). Scoring Function Matrix를 활용한 전투기 3 차원 공중전 기동

생성. 한국항공우주학회 학술발표회 초록집, 442-445.
Breton, R., & Rousseau, R. (2005, June). The C-OODA: A cognitive version of the OODA loop to represent C2 activities. In Proceedings of the 10th International Command and Control Research Technology Symposium.
Ernest, N., Carroll, D., Schumacher, C., Clark, M., Cohen, K., & Lee, G. (2016). Genetic fuzzy based artificial intelligence for unmanned combat aerial vehicle control in simulated air combat missions. J Def Manag, 6(144), 2167-0374.
Toghiani-Rizi, B., Kamrani, F., Luotsinen, L. J., & Gisslén, L. (2017, October). Evaluating deep reinforcement learning for computer generated forces in ground combat simulation. In Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on (pp. 3433-3438). IEEE.
Park, H., Lee, B. Y., Tahk, M. J., & Yoo, D. W. (2016). Differential game based air combat maneuver generation using scoring function matrix. International Journal of Aeronautical and Space Sciences, 17(2), 204-213.
Sutton, R. S., & Barto, A. G. (1998). Introduction to reinforcement learning (Vol. 135). Cambridge: MIT press.
Shaw, R. L. (1985). Fighter Combat. Naval Institute Press.
Ng, A. Y., Harada, D., & Russell, S. (1999, June). Policy invariance under reward transformations: Theory and application to reward shaping. In ICML (Vol. 99, pp. 278-287).
Sutton R. S., McAllester D., Singh S., Mansour Y. (2000) Policy Gradient Methods for Reinforcement Learning with Function. In NIPS.
Silver, D., Lever G., Heess N., Degris T., Wierstra D., Riedmiller M. (2014). Deterministic Policy Gradient Algorithms. In ICML (JMLR: W&CP volume 32.).
Lillicrap T. P., Hunt J. J., Pritzel A., Heess N., Erez T., Tassa Y., Silver D. & Wierstra D. (2016) Continuous Control with Deep Reinforcement Learning. In ICLR.

문 일 철 (ORCID : https://orcid.org/0000-0002-1798-1306 / icmoon@kaist.ac.kr)

2004 서울대학교 컴퓨터공학부 공학사
2008 카네기멜론 대학교 공학박사
2008 KAIST 전기및전자공학과 연수연구원
2011~ KAIST 산업및시스템공학과 조교수, 부교수

관심분야 : 인공지능, 모델링 및 시뮬레이션

정 민 재 (ORCID : https://orcid.org/0000-0002-2405-8464 / extraord96@kaist.ac.kr)

2018 KAIST 산업및시스템공학과 공학사
2020 KAIST 산업및시스템공학과 공학석사
2020~ 티맥스 소프트

관심분야 : 인공지능

김 동 준 (ORCID : https://orcid.org/0000-0002-1117-5848 / dongjoun57@kaist.ac.kr)

2016 KAIST 수리과학과 이학사
2016~ KAIST 산업및시스템공학과 석박사통합과정

관심분야 : 인공지능