

## 강화학습을 활용한 기만행위 모의방법 연구 : 해병대 상륙양동 사례를 중심으로

박대국<sup>\*,1)</sup> · 조남석<sup>1)</sup>

<sup>1)</sup> 국방대학교 국방과학학과

### A Study on Reinforcement Learning Method for the Deception Behavior : Focusing on Marine Corps Amphibious Demonstrations

Daekuk Park<sup>\*,1)</sup> · Namsuk Cho<sup>1)</sup>

<sup>1)</sup> Department of Defense Science, Korea National Defense University, Korea

(Received 9 March 2022 / Revised 1 July 2022 / Accepted 29 July 2022)

#### Abstract

Military deception is an action executed to deliberately mislead enemy's decision by deceiving friendly forces intention. In the lessons learned from war history, deception appears to be a critical factor in the battlefield for successful operations. As training using war-game simulation is growing more important, it is become necessary to implement military deception in war-game model. However, there is no logics or rules proven to be effective for CGF(Computer Generated Forces) to conduct deception behavior automatically. In this study, we investigate methodologies for CGF to learn and conduct military deception using Reinforcement Learning. The key idea of the research is to define a new criterion called a "deception index" which defines how agent learn the action of deception considering both their own combat objectives and deception objectives. We choose Korea Marine Corps Amphibious Demonstrations to show applicability of our methods. The study has an unique contribution as the first research that describes method of implementing deception behavior.

Key Words : Amphibious Demonstrations(상륙양동), Computer Generated Forces(가상군), Unity 3D(유니티 3D), Reinforcement Learning(강화학습), Deception Behavior Model(기만행위모델)

#### 1. 서론

기만작전이란 "아군의 작전 의도, 능력, 배치 등을 적이 오관하도록 유도하여 적을 아군의 의도대로 유

\* Corresponding author, E-mail: ncho64@gmail.com  
Copyright © The Korea Institute of Military Science and Technology

인하거나 적의 기도를 사전에 포기하게 하는 계획적인 작전 활동”으로 전장기만 방법에는 양공, 양동, 계략, 허식이 있다<sup>[1]</sup>. 기만작전은 전장에서 발휘되는 하나의 기습으로 적을 심리적으로 마비시킬 수 있으며, 최소한의 피해로 최대 효과를 얻을 수 있으므로 전투에서 승패를 좌우하는 핵심요소로 작용한다<sup>[2]</sup>.

기만작전의 방법 중 양동은 “적을 기만할 목적으로 야군이 결정적인 작전을 기도하고 있지 않은 지역에서 실시하는 무력시위”이며<sup>[3]</sup>, 해병대에서는 상륙작전 형태 중 하나인 상륙양동을 중요한 전투행위로 간주한다. 상륙양동은 “상륙작전부대의 무력시위를 통해 적이 불리한 방책을 채택하도록 상륙 시간·장소·규모에 대해 적을 기만하는 상륙작전”으로 정의되며, 적을 속이고 혼란에 빠뜨리기 위하여 다른 기만작전과 연계하여 실시하기도 한다<sup>[4]</sup>.

상륙양동과 관련된 전사를 보면 인천상륙작전과 노르망디 상륙작전이 대표적이다. 인천상륙작전에서는 인천 및 군산, 삼척 지역 일대에서의 기만작전과 군산 지역에서의 상륙양동으로 주요시설을 파괴하고 적의 주의를 전환시켜 적으로 하여금 상륙지역 및 시기를 오판케 하였으며<sup>[5]</sup>, 노르망디 상륙작전에서는 스칸디나비아 해안에서의 상륙양동으로 독일군을 스칸디나비아 반도에 고착시키고, 파 드 칼레 지역에서의 상륙양동으로 상륙지역 및 시기를 오판케 했다<sup>[6]</sup>. 이처럼 성공적인 기만작전 및 상륙양동은 전쟁 수행의 중요한 요소임에 틀림없다. 이외에도 중동전쟁과 걸프전쟁, 이라크전쟁 등에서도 기만작전은 광범위하게 실시되었으며, 합동작전을 중요시하는 현대전에서도 주도권 확보를 위한 기만작전은 전승을 위한 필수요소로 인식되고 있다<sup>[7]</sup>.

본 연구는 기만행위를 자동화 모의하기 위한 방법론을 다루고 있다. 특히, 위게임 모델 안에서 자동으로 모의되는 객체인 CGF(Computer Generated Forces, 가상군)가 자율적으로 기만행위를 할 수 있는 방법론에 대해서 연구한다<sup>[8]</sup>. 이를 위해 실험의 대상을 해병대 양동작전으로 제한하고 간략화된 상륙양동 모델을 통해 CGF의 자동화 모의를 구현하겠으며, 기만작전의 연구를 위한 모의에서는 활용된 바 없는 인공지능의 강화학습(Reinforcement Learning)을 다룬다.

본 연구의 주요 구성은 다음과 같다. 2장에서는 CGF 자동화 모의의 중요성에 대해 설명하고 관련 연구를 분석하고, 3장에서는 강화학습에 대해 간략히 소개하고, 기만행위 묘사를 위한 상륙양동 모델 행동규

칙을 설명한 후, 강화학습을 구현하기 위해 보상규칙과 실험방법에 대해 논의한다. 4장에서는 Unity 3D를 사용해서 구현한 상륙양동 모델의 환경 구성 및 설정에 대해 설명하고, 강화학습 결과에 대해 분석한다. 5장은 인천상륙작전으로 Case Study를 통해 학습된 모델을 적용해본다. 마지막으로 6장에서는 결론 및 향후 연구에 대해 설명한다.

## 2. 이론적 배경

본 장에서는 먼저 CGF의 정의 및 자동화 모의의 중요성에 대해 설명하고 관련 연구를 분석한다.

### 2.1 CGF 자동화 모의 중요성

CGF는 위게임 모델 등 국방 M&S(Modeling & Simulation) 분야에서 사람의 통제 없이 자율적으로 판단하고 자동으로 모의되는 객체를 의미한다. 예를 들어, 위게임 모델에서 게이머가 조작하는 청군은 상륙작전을 실시하며, 해안을 방어하는 홍군은 자동모의된다고 했을 때 홍군을 CGF라고 할 수 있다.

선행연구<sup>[8-10]</sup> 등에서 논의된 CGF의 자동화 모의의 필요성은 다음과 같이 요약할 수 있다. 첫째, 병력자원의 부족으로 비전투 분야의 병력감축이 요구되며, 위게임을 조작하고 관리하는 인원의 감축도 요구된다. 구체적으로 보면, 인구절벽으로 인해 병역자원은 '22년에는 25만명 수준으로 떨어지고, '40년 이후에는 16만명으로 급감할 것으로 예상된다<sup>[10]</sup>. 반면, 위게임 모델은 장차전에서 그 중요성이 더욱 커져 위게임 운용에 필요한 과업의 양은 오히려 더 증가하고 있다. 이러한 상황에서 CGF의 자동화 모의는 위게임에 필요한 병력의 수는 줄이면서, 확대되는 위게임 훈련 및 분석을 지원할 수 있는 효율적인 대안이 될 수 있다.

둘째, 4차 산업혁명 시대에 많은 기술발전과 함께 국방 M&S와 연관된 많은 기술들이 진보하고 있다. 국방 M&S의 기술 중 많은 것들은 단순히 위게임 안에서만 적용되는 것이 아니라, 실 전장에서 활약할 무인무기체계 플랫폼에서 활용할 수 있는 기술로 확대 발전할 수 있다. 예를 들어, CGF의 자동화 표적할당 기술이 개발되면, 해당 기술이 곧 무인무기체계의 표적할당 기술로 활용이 가능하다는 것이다. 이러한 의미에서, CGF의 자동화 모의를 위한 다양한 기술의 개발은 중요하다.

셋째, CGF는 실 전장을 구성하는 무기체계, 부대 등을 표현하여 구성이 가능할 뿐 아니라, 아직 도입되지 않은 무기체계나 전술 등을 구현하는 것도 가능하기 때문에 CGF의 자동화모의는 다양한 전장을 묘사하는데 그 역할을 할 수 있다. 즉, 시나리오에 의한 형식적인 임무 수행 모의뿐만 아니라 예측하기 어려운 전술 상황을 묘사할 수 있고, 게이머의 개입 없이 지속적인 운용이 가능하기 때문에 시간, 비용 등 제반사항이 해소되면서 다양하고 창의적인 연구 및 분석이 가능하게 된다.

### 2.2 CGF 자동화 모의를 위한 연구

CGF 자동화 모의를 다룬 연구로 Cho, et. al<sup>[9]</sup>은 국방 M&S의 CGF 행위 모델링 방법론들에 대해 논의하고, 이어서 이러한 방법론을 다양한 분야별 어떻게 적용할 수 있는지에 고찰하여 향후 자동화 모의 기술 개발 및 사업을 추진하는 군에게 가이드 라인을 제시하였으며, Jung<sup>[11]</sup>은 CGF의 목표를 우선적으로 탐색하는 목표지향행위계획과 우발상황 시 새로운 계획을 시행하는 재계획 방법에 대하여 제안했으며, 공군 전투비행기 모델에 적용하여 검증하였다. 또한 Lee, et. al<sup>[8]</sup>은 CGF 기술의 중요성과 기술 개발 및 적용 사례를 설명하고, 군 전투훈련, 전력분석, 체계획득 등 국방분야 적용방안에 대해 연구 하였으며, Han, et. al<sup>[12]</sup>은 METT+TC에 기초한 과업 목록과 퍼지 추론의 적용 과정에 대해 설명하고, 이를 소부대 전투 시뮬레이션에 적용하여 CGF의 자율능화 방안에 대해 고찰하였다. 이외에도 Hong, et. al<sup>[13]</sup>은 연평도 일대 적 포격 도발 상황을 모의한 전투 시나리오를 구성하고, 적 행동 양상 모델의 학습을 통해 CGF 전투 개체의 성능 향상을 확인했으며, Lee, et. al<sup>[14]</sup>은 적 예상 기동로에 따른 아군 전차소대의 대응에 대한 CGF 자율행위 모델링을 구현하고, 최적 행동정책이 실제 교리와 일치하는 것을 확인하였다.

이처럼 CGF의 자동화 모의를 위한 연구는 다수 있었으나, 1절에서 설명한 바와 같이 전투에서 매우 중요한 행위인 기만행위를 구현하기 위한 연구는 이뤄진 바 없었다. 이를 통해 도출한, 본 연구의 차별성은, 그 동안 이루어진바 없는 기만행위를 자동화 모의하기 위한 방법론을 다루고, 그것을 실제로 구현해보았다는 데 있다.

### 3. 방법론

본 장에서는 강화학습에 대해 간략히 소개하고, 기만행위 묘사를 위한 상륙양동 모델 행동규칙을 설명한 후, 강화학습을 구현하기 위해 보상규칙과 실험방법에 대해 논의한다.

#### 3.1 강화학습(Reinforcement Learning)

강화학습은 머신러닝(Machine Learning, ML)의 한 영역으로써, 에이전트가 어떠한 환경 안에서 반복적인 시행착오 상호작용을 통해 목표를 달성하는 학습 과정이다. 또한 마코프 결정 과정(Markov Decision Process)을 이용해서 수학적으로 모델링 되며, 의사결정 문제를 순차적으로 풀기 위한 최적화 방법이다<sup>[15]</sup>.

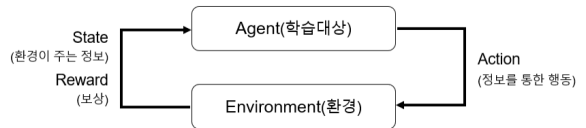


Fig. 1. Mechanism of reinforcement learning

Fig. 1과 같이 학습대상인 에이전트(Agent)는 환경(Environment)과 상태(State)를 통해 관측된 정보를 토대로 행동(Action)을 선택하고 보상(Reward)을 받으며, 반복 학습을 통해 보상을 최대화한다. 에이전트가 행동을 선택할 때마다 환경의 상태와 보상은 계속해서 바뀌며, 특정 상태에서 어떠한 행동을 선택하는 규칙이 정책(Policy)이다<sup>[16]</sup>.

예를 들어, 강화학습 알고리즘 중 가장 대표적인 Q-Learning은 특정 상태에서 행동을 선택했을 때, 행동의 가치를 나타내는 Q 함수를 이용하여 각각의 상태에서 최적의 정책을 학습한다<sup>[17]</sup>.

#### 3.2 상륙양동 모델

강화학습은 에이전트가 행동하고 보상을 획득하는 모델이 필요한데, 본 연구에서는 간략화된 상륙양동 모델을 Fig. 2와 같이 구성하여 연구진의 방법론을 구현한다. 상륙양동은 1절에서 설명한 바와 같이 가장 대표적인 기만행동으로 공자는 상륙지점을 속이기 위한 노력을, 방자는 상륙지점을 올바르게 판단하기 위한 노력을 지속한다.

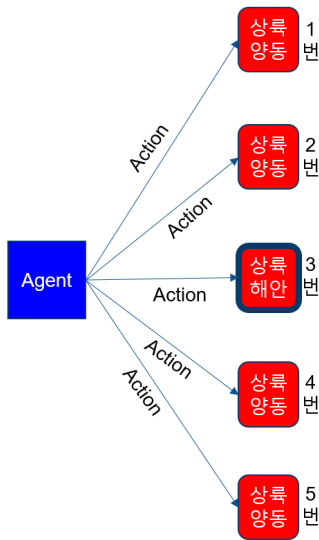


Fig. 2. Model of amphibious demonstrations

Fig. 2와 같이 아군은 싱글 에이전트로 상륙작전을 하는 입장이며, 적은 1~5번 해안 지역에서 방어하는 입장이다. 아군은 3번을 상륙해안으로 선정하였으며, 나머지 지역에서는 양동작전을 할 예정이다. 아군은 상륙해안을 알고 있지만 적은 모른다.

아군은 상륙해안을 포함한 5개소에 대해 적을 오판케 하기 위하여 총 10회의 타격을 실시하며, 1회 타격 시 적 전투력은 5% 감소한다. 에피소드는 10회차 타격 후 종료되며, 매 타격 시마다 전투력 충원이 이루어진다.

전투력 충원은 인접 지역 1개소에서 2%의 충원을 원칙으로 시행하며, 적 전투력이 0%일 시에는 인접 지역 1개소에서 5% 전투력을 충원 받는다. 만약, 전투력 충원이 불가할 시에는 또 다른 지역에서 충원한다.

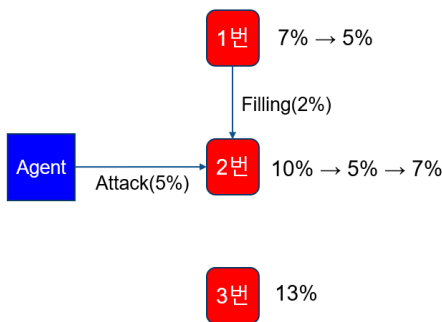


Fig. 3. Rule for combat power covering

예를 들어, 위에서 설명한 규칙의 예는 Fig. 3과 같다. 현재 전투력은 1번 7%, 2번 10%, 3번 13%이며, 에이전트가 2번 지역을 타격함으로써 2번 지역 전투력은 5%로 감소 되었다. 전투력 충원은 인접 지역인 1, 3번 지역 중에서 랜덤으로 선택되며, 1번 지역에서 2%를 충원해서 2번 지역의 전투력은 최종 7%가 되었다.

기본적으로 전투력 충원은 교리에 의거 예비대에서의 충원이 기본이지만, 본 모델에서는 에이전트의 기만행위를 통해 적 부대 주의를 전환되고 타격 받은 지역에 추가 전투력이 투사되는 것을 묘사하기 위해 인접 지역에서 전투력을 충원한다고 가정한다. 이러한 가정은 모델의 validity를 저하시키는 것이 분명하지만, 기만작전에 대한 적의 ‘속는 행위’를 묘사하는데 필요한 가정이기 때문에 포함하였다.

### 3.3 강화학습 구현

본 절에서는 강화학습 구현에서 핵심이 되는 보상 규칙에 대해 설명한다.

보상은 한 번의 에피소드 종료 후, 지역별 잔여 전투력으로 판단하고, 상륙해안의 전투력 순위에 따라 차등적으로 보상(1~5점)을 부여한다. 이는 상륙양동의 목적은 적을 오판케 하는 것이지만, 성공적인 상륙작전을 위해서는 상륙해안의 전투력도 최소화 시켜야 하기 때문이다. 예를 들어, 10회차 타격 후, 잔여 전투력은 1번 15%, 2번 10%, 3번 12%, 4번 20%, 5번 17%라고 가정한다. 상륙해안 3번의 전투력이 2번째로 낮으므로 4점을 부여한다. 두 번째 에피소드가 시작되면 다시 전투력은 초기화된다. 두 번째 에피소드 종료 후, 잔여 전투력이 1번 13%, 2번 10%, 3번 8%, 4번 15%, 5번 10%라고 가정하면 상륙해안의 전투력이 가장 낮으므로 가장 높은 5점의 보상을 부여한다.

현재까지 구현된 행동규칙으로 강화학습을 실행하면, 에피소드가 진행될수록 에이전트는 높은 보상을 받기 위해 상륙해안으로만 타격을 하면서 상륙해안의 전투력을 최소화 시킨다. 이는 적으로 하여금 상륙해안을 정확하게 판단하게 할 수 있기 때문에, 기만행위로 볼 수 없다. 따라서, 강화학습의 에이전트가 ‘속이는 행위’에 대해 보상을 받고 스스로 학습을 할 수 있는 보상체계가 필요하다.

본 연구에서는 적을 속이는 행위를 수치화(정량화)하기 위해, 새로운 개념인 ‘기만지수(Deception Index)’를 정의하였다. 기만지수를 일반적으로 정의하면, 적의 입장에서 아군의 작전 의도를 판단하는데 사용되

는 수치이다. 예를 들어, 육군의 여단급 이상 부대의 공격작전 시 주공 방향을 결정하게 되는데 적은 아군의 지휘소 위치나 또는 여건조성을 위한 상급부대 화력 등으로 아군의 주공 방향을 판단할 수 있다. 이 경우, 기만지수는 아군의 화력 빈도, 지휘소 관측 여부 등으로 계산할 수 있을 것이다.

이러한 일반적 정의를 바탕으로, 본 연구에서 활용한 해병대 상륙양동 모델에서의 기만지수는 적이 아군의 상륙해안을 예측하는 일종의 판단기준이며, 에피소드 종료 후 적군이 맞게 판단할 시에는 아군에게 감점을 주고, 오판할 시에는 아군에게 추가 보상을 주는 체계로 작용한다.

기만지수 판단기준은 다양하게 구성이 가능하지만, 본 연구에서의 판단기준은 다음과 같다.

- ① 특정 지역의 적 전투력이 35 % 이하일 경우 해당 지역을 상륙해안으로 판단한다.
- ② 아군이 특정 지역을 연속적으로 타격할 때 해당 지역을 상륙해안으로 판단한다.
- ①과 ② 중 한 가지 조건을 만족할 경우 상륙지역으로 판단한다.

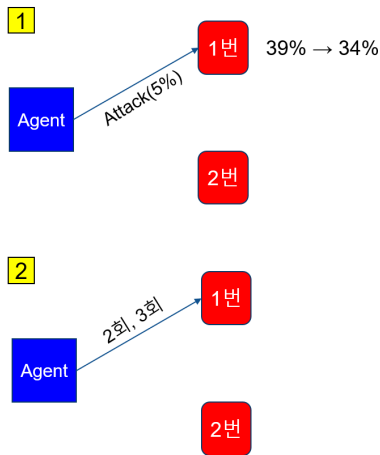


Fig. 4. Deception index criteria

예를 들어, Fig. 4의 ①은 전투력에 의한 기만지수 판단조건을 보여주며, 에이전트가 1번 지역을 타격해서 전투력이 34 %가 되면서 적은 1번 지역을 상륙해안으로 판단한다. ②는 연속 타격에 의한 기만지수 판단조건을 보여주며, 에이전트가 2회차와 3회차에 연속으로 1번 지역을 타격하면서 적은 1번 지역을 상륙해안으로 판단하는 모습을 보여준다. 기만지수에 의한

적의 판단은 타격이 진행되면서 만족하는 조건이 바뀌면 판단하는 대상도 계속 바뀌게 된다.

### 3.4 실험방법

앞 절에서 설명한 상륙양동 모델과 강화학습 논리를 구현하기 위한 Tool로는 R<sup>[18]</sup>의 Reinforcement Learning<sup>[19]</sup> 패키지, NetLogo<sup>[20]</sup>의 Reinforcement Learning Maze<sup>[21]</sup>, Unity 3D<sup>[22]</sup>의 ML\_Agent Toolkit<sup>[23]</sup>, MATLAB<sup>[24]</sup>의 Reinforcement Learning Toolbox<sup>[25]</sup> 등이 있다. 하지만, 연구진이 제시한 상륙양동 모델은 1회차 타격부터 10회차 타격까지의 공격이 모두 독립적으로 평가받지 않고, 일관된 목표를 가진 일련의 연속된 행동이기 때문에 각각의 상태들을 복원추출(Sampling with Replacement) 형태로써 독립적인 형태로 행동을 결정하고 최적의 정책을 구현하는 여타의 툴로는 구현이 어렵다. 본 연구에서는 에피소드 내의 상태들이 독립적인 형태가 아닌 연속적인 형태로 이어지며, 에피소드 종료 후 보상을 받는데 활용 가능한 Unity 3D를 활용한다.

Unity 3D는 게임을 만드는데 최적화 되어 있는 엔진이며, 가상현실(VR), 3D 애니메이션 등 인터랙티브 콘텐츠 제작이 가능한 통합 도구이다. 또한 3D 모델링 객체의 제작 및 삽입, 인공지능, 특수효과, 지형, 등을 사실적으로 표현이 가능하다<sup>[26]</sup>. ML\_Agent는 Unity 3D 환경에서 강화학습 연구가 가능한 머신러닝 플랫폼이다.

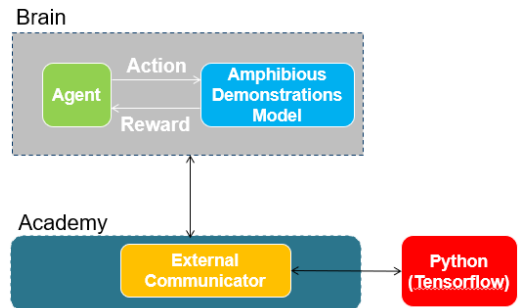


Fig. 5. ML\_Agent structure

Fig. 5는 ML\_Agent의 구조를 도식화하여 나타내고 있다. 에이전트는 상륙양동 모델에서 상태 값을 가지고 행동을 수행하고, 상륙양동 모델은 에이전트에게 행동에 대한 보상을 주면서 상호작용을 한다. 이를 통해 브레인인 정책을 결정하여 학습의 에피소드를 관리하는 아카데미 External Communicator를 통해 외부

Python 프로그램에게 모델을 제공하여 학습하고 최적의 정책을 결정한다<sup>27)</sup>.

#### 4. 연구 결과 및 분석

본 장에서는 Unity 3D를 사용해서 구현한 상륙양동 모델의 환경 구성 및 설정에 대해 설명하고, 강화학습 결과에 대해 분석 후 학습된 모델의 효과성에 대해 논의한다.

##### 4.1 상륙양동 모델 환경 구성

상륙양동 모델의 환경 구성은 Fig. 6과 같으며, 에이전트 오브젝트는 항공기로 생성해서 어떤 지역을 타격하는지 확인 가능하도록 레이캐스트(Raycast)를 사용한다. 그리고 상륙해안 및 상륙양동 지역은 사각형 오브젝트로써, 좌측 1번 지역부터 5번까지이다. 또한 타격 이후 각 지역 전투력을 확인할 수 있는 Bar를 화면 상단에 표시하고, 잔여 타격 횟수도 포함한다. 마지막으로 타격 이후 지역별 전투력이 감소하고 충전되는 모습을 표현한다.

모델은 Unity 3D(2020.3.12.fl)로 구현하였으며, 모든 실험은 CPU 1.60GHz, RAM 8GB의 컴퓨터에서 실시한다.



(a) Strike



(b) Filling

Fig. 6. Snapshot of amphibious demonstrations model in Unity 3D

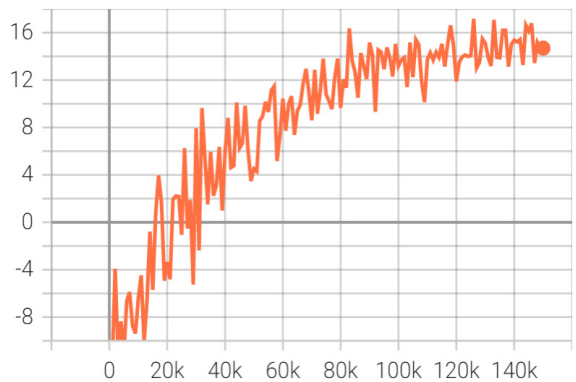
##### 4.2 환경 설정

강화학습을 위한 환경 설정은 다음과 같다.

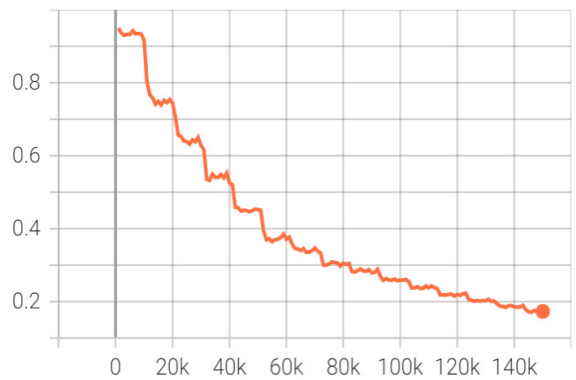
첫째, 적 전투력은 5개 지역 50%로 동일하게 하였으며, 둘째, 보상은 3.2절에서 설명한 바와 같이 상륙해안 지역 전투력이 가장 낮을 시 5점의 보상을 부여한다. 셋째, 기만지수에 의한 보상은 적이 상륙해안을 맞게 판단했을 때 아군 패널티 -15점, 적이 상륙해안을 오판했을 때 아군 추가 보상 15점으로 설정한다. 넷째, 학습 효과를 가장 잘 구현할 수 있도록 150,000회를 학습한다.

##### 4.3 강화학습 결과

강화학습 한 결과는 누적값(Cumulative Reward)과 엔트로피(Entropy)를 시각화하여 학습이 잘 되었는지 분석한다.



(a) Cumulative reward graph



(b) Entropy graph

Fig. 7. Reinforcement learning results graph

Fig. 7 (a)는 한 에피소드가 진행되는 동안 에이전트가 받은 보상과 감점의 누적값이다. 부분적인 감소 구간이 있긴 하지만 학습이 진행됨에 따라 지속적으로 증가하는 것을 볼 수 있으며, 이는 정상적인 학습을 수행하고 있음을 보여주고 있다.

(b) 엔트로피는 모델이 학습하는 동안 행동을 얼마나 랜덤하게 선택하는지를 나타내는 값이며, 이 값이 클수록 Exploration의 비중이 크다고 할 수 있다. 그래프에서 엔트로피 값이 점차 감소하는 것을 볼 수 있는데, 이는 정책 결정의 랜덤성이 줄어드는 것으로 볼 수 있고 정책이 최적값에 수렴하고 있다고 볼 수 있다.

4.4 학습 모델 분석

학습된 모델의 결과는 상륙해안인 3번을 타격 및 충원하는 횟수에 따라 크게 세 가지 경우로 분류되는데 이는 다음과 같다.

학습된 에이전트의 기만행위 결과 중 다수로 도출된 패턴은 Table 1, Table 2와 같다. 적색은 에이전트

가 타격한 지역이며, 청색은 충원한 지역이다. 두 Table의 결과는 모두 상륙해안인 3번을 4회 타격하고, 상륙해안에서 다른 지역으로 1회 충원한 결과를 보여주고 있는데 타격하는 횟수나 순서는 매번 달라진다. 표 안의 숫자는 타격과 충원이 일어난 후 변동되는 각 지역의 전투력을 의미한다. 표의 제일 하단에 기록한, 적 예측지역은 매 타격 시마다 판단을 하며, 2회 차까지는 충족되는 조건이 없어서 예측을 못 하다가 3회차에서 특정지역을 연속적으로 타격하는 조건을 충족하여 상륙해안을 3번으로 판단한다. 이후 8회차에서 특정지역을 연속적으로 타격하는 조건을 충족하여 5번을 상륙해안으로 판단하면서 에피소드 종료 후 최종 5번을 상륙해안으로 판단한다. 전투력은 상륙해안이 가장 낮게 나오며, 최종 상륙해안의 전투력을 최소화 하면서 적이 상륙해안을 오판케 하는 목적과 부합한 결과가 나왔다. 100회의 에피소드 실험 결과 이와 같은 에피소드가 70 %로 가장 많으며, 적이 상륙해안을 맞게 예측하는 경우는 없었다.

Table 1. Reinforcement learning results 1

		타격 충원									
	1회	2회	3회	4회	5회	6회	7회	8회	9회	10회	
1번	50	-	-	48	-	45	-	-	-	45	
2번	50	48	-	45	43	41	-	-	-	39	
3번	48	45	42	-	39	-	-	-	-	36	
4번	47	-	45	-	-	-	43	41	39	39	
5번	50	-	-	-	-	-	47	44	41	41	
예측	-	-	3번	3번	3번	3번	3번	5번	5번	5번	

Table 3. Reinforcement learning results 3

		타격 충원									
	1회	2회	3회	4회	5회	6회	7회	8회	9회	10회	
1번	50	47	-	-	-	-	-	-	-	47	
2번	50	48	-	46	43	-	41	-	-	41	
3번	47	-	44	41	39	-	36	34	-	34	
4번	48	-	46	-	-	43	-	40	37	37	
5번	50	-	-	-	-	48	-	-	46	43	
예측	-	-	-	3번	3번	3번	3번	3번	4번	4번	

Table 2. Reinforcement learning results 2

		타격 충원									
	1회	2회	3회	4회	5회	6회	7회	8회	9회	10회	
1번	50	-	-	-	-	-	-	47	44	44	
2번	50	47	-	-	45	-	-	43	41	39	
3번	47	45	-	-	42	-	39	-	-	36	
4번	48	-	46	44	-	41	39	-	-	39	
5번	50	-	47	44	-	42	-	-	-	42	
예측	-	-	-	5번	5번	5번	5번	5번	1번	1번	

Table 4. Reinforcement learning results 4

		타격 충원									
	1회	2회	3회	4회	5회	6회	7회	8회	9회	10회	
1번	50	-	-	-	47	-	-	45	43	43	
2번	48	45	-	-	43	-	-	40	37	37	
3번	47	45	42	-	-	40	37	-	-	34	
4번	50	-	48	46	-	43	41	-	-	39	
5번	50	-	-	47	-	-	-	-	-	47	
예측	-	-	-	-	-	-	-	-	2번	3번	

Table 3, Table 4는 상륙해안을 4회 타격하고, 상륙해안에서 다른 지역으로 2회 충원한 결과이며, 타격하는 횟수나 순서는 매번 달라진다. 전투력은 상륙해안이 가장 낮지만, 적이 상륙해안을 맞게 예측하는 경우도 생긴다. 100회 에피소드 실험 결과 이와 같은 에피소드는 20 % 정도이며, 그중 대략 10 % 정도는 적이 상륙해안을 맞게 예측할 수 있다.

Table 5, Table 6은 상륙해안을 5회 이상 타격하는 결과이며, 타격하는 횟수나 순서는 매 실험마다 변경된다. 상륙해안에 최소 5회 이상의 타격을 하기 때문에 전투력은 가장 낮았지만, 적이 상륙해안을 맞게 판단하는 경우도 증가한다. 100회 에피소드 실험 결과 이와 같은 에피소드는 10 % 정도이며, 그중 대략 70 % 정도는 적이 상륙해안을 맞게 예측할 수 있었다. 그래서 이런 경우에는 실제로 상륙하고자 하는 상륙해안의 전투력이 최소화된 효과는 가지지만, 적의 입장에서는 계속해서 상륙해안을 맞게 판단하기 때문에 올바른 양동작전이 이뤄졌다고 보기는 어렵다.

Table 5. Reinforcement learning results 5

타격 충원										
	1회	2회	3회	4회	5회	6회	7회	8회	9회	10회
1번	50	47	-	-	-	-	-	45	43	43
2번	50	48	-	46	-	-	44	41	38	38
3번	47	-	45	42	39	-	36	-	-	33
4번	48	-	45	-	43	41	-	-	-	39
5번	50	-	-	-	-	47	-	-	-	47
예측	-	-	-	-	3번	3번	3번	2번	2번	3번

Table 6. Reinforcement learning results 6

타격 충원										
	1회	2회	3회	4회	5회	6회	7회	8회	9회	10회
1번	50	-	-	-	47	44	-	-	-	44
2번	47	-	45	-	43	41	-	-	39	39
3번	48	45	42	-	-	-	39	37	34	31
4번	50	48	-	46	-	-	44	41	-	39
5번	50	-	-	47	-	-	-	-	-	47
예측	-	-	3번	3번	3번	1번	1번	1번	3번	3번

이후, 학습된 모델이 랜덤 모델보다 얼마나 상륙양동 계획의 목적과 부합하는지 확인하기 위해 비교 실험을 한다. 랜덤모델에서 타격 지점의 선정은 Uniform 분포를 따른다. 랜덤모델은 방자가 공격을 예측할 수 없다는 측면에서 일종의 기만 성격을 가지고 있기 때문에, 만약 기만을 학습한 모델이 없다면 택할 수 있는 하나의 기만행동 모델이다. 따라서, 모형의 비교대상으로 적합하다고 할 수 있다.

Table 7. Performance comparison between Random Model and Reinforcement Model

구분		랜덤 모델	학습 모델 (150,000회)
실험횟수		100회	
상륙해안 전투력	평균	41.72	35.15
	분산	14.21	2.98
상륙해안 예측 횟수		27회	9회

Table 7은 랜덤 모델과 학습 모델을 비교 실험한 결과표이다. 랜덤 모델은 학습된 에이전트가 없기 때문에 타격은 랜덤하게 하되, 상륙해안을 예측하는 절차는 동일하게 적용하였다. 실험 횟수는 100회 에피소드이며, 각각의 에피소드 종료 후 상륙해안인 3번 전투력의 평균, 분산, 상륙해안 예측 횟수를 기술하였다. 평균과 분산은 학습된 모델이 랜덤 모델보다 낮았으며, 적이 상륙해안을 맞춘 예측 횟수는 학습 모델이 9회, 랜덤 모델은 27회로 도출되었다. 랜덤 모델에서 에이전트는 상륙해안의 전투력을 낮추지 못하고 적은 100회 중 27회를 상륙해안을 제대로 맞추면서 학습 모델과 차이가 있음을 보인다.

이와 같은 결과는 학습 모델이 상륙해안의 전투력을 최소화 시키면서 적으로부터 상륙해안을 오판케 하는 두 가지 목적을 효과적으로 달성하게끔 학습이 되었다는 것을 확인할 수 있다. 또 하나 학습된 모델의 중요한 특징은 상륙해안인 3번의 인접 지역인 2번과 4번 해안에 많은 타격을 하는 것이다. 이는, 인접 지역의 전투력이 낮아지면 이를 보강하기 위해 3번에서 전투력 차출이 이루어져 결과적으로 3번의 전투력을 낮추면서 동시에 적으로 하여금 3번이 상륙해안이라는 예측을 하지 못하게 하는 전술적 판단이 학습된 것이다.



다음은, 학습 모델의 현실적용 가능성을 확인하기 위해 상륙작전사에서 가장 대표적인 인천상륙작전을 대상으로 Case Study를 실시한다.

### 5. Case Study : 인천상륙작전

양동작전을 보여주는 전사로 인천상륙작전이 있다. 1장에서 설명한 바와 같이 인천상륙작전을 성공적이었다고 할 수 있었던 요인 중 하나로 치밀한 양동작전이 있었음은 틀림없다. 즉, 전세를 완전히 뒤바꿀 수 있었던 인천상륙작전의 성공은 다양한 형태의 양동작전이 없었다면 불가능했다고 할 수 있을 것이다.

본 장에서는 학습된 모델을 인천상륙작전에 적용해서 양동작전을 실시하였다. 목표 달성을 위해서는 인천지역의 전투력을 최소화하며, 적이 상륙해안을 오판케 하는 것이다.

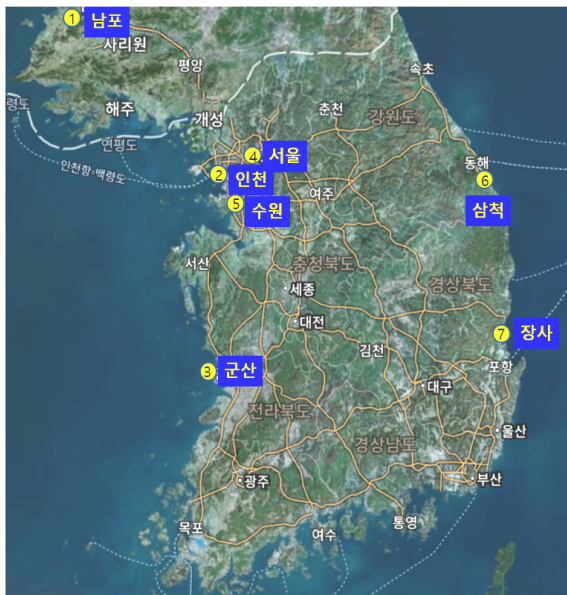


Fig. 8. Candidates of landing point

Fig. 8은 인천상륙작전에서 실제로 양동작전을 실시한 7개 지역으로 표시했다.

당시 전선 상황으로 북한군은 대부분의 모든 전투역량을 낙동강 전선에 집중하고 있었으며, 후방지역에는 지역 경비부대, 병참선 경비부대, 훈련이 미숙한 신편부대를 배치하고 있었다<sup>[28]</sup>. 또한 북한군은 인천

상륙작전 D-Day 2주 전부터 상륙작전을 예상하고 있었던 것으로 보인다. 이유는 서해안 방어사령부를 조직하여 인천-서울 방어부대와 군산-인천에 배치된 경비대, 보안대들을 여기에 소속시켜 전투역량과 기재를 인천-서울에 집중하도록 하였으나, 정확한 상륙시간 / 장소 / 규모를 모르기 때문에 적극적인 대책이 미흡했던 것으로 판단된다<sup>[28]</sup>. 북한군 병력 규모는 서울에 약 5,000명, 인천지역은 경비부대를 포함하여 약 2,000명으로 추정되고, 추가 증원병력 약 2,500여 명을 포함하면 서울-인천지역에서 활동하고 있는 북한군 총병력은 약 10,000여 명이었던 것으로 알려져 있다<sup>[28]</sup>. 그 외 남포지역에는 지역 경비부대와 1개 해안 육전여단, 1개 전차연대가 배치되어 있었다<sup>[28]</sup>. 이처럼 각 지역에 배치된 북한군 병력의 수를 합산하여 전투력을 산출하였다. 각 지역 전투력으로 남포 30 %, 인천 40 %, 군산 25 %, 서울 50 %, 수원 20 %이며, 삼척과 장사는 경비부대와 보안대만 방어하고 있는 것으로 가정하여 각각 15 %로 산출하였다.

상륙작전을 위한 실제 양동작전으로 항공 폭격은 남포:인천:군산을 30:40:30으로 배분함으로써, 비록 인천에 가장 많은 폭격을 실시했지만, 인천지역이 크게 두드러지지 않도록 했으며, 무력시위(상륙양동) 등을 통해 군산지역으로 북한군의 관심이 특별히 많이 쏠리도록 했다고 알려져 있다<sup>[29]</sup>. 이외에도 서울, 수원 등 인천 주변 지역을 폭격하였으며, 함선은 삼척지역에 함포사격을 실시했다. 또한 군산 및 장사지역에서는 상륙양동을 실시했다고 알려져 있다<sup>[30]</sup>.

본 연구에서 개발한 학습된 모델로 D-Day 2주를 고려 20회의 양동작전을 실시한 결과, 인천지역을 30 %로 가장 많이 양동작전을 실시했고, 다음으로는 남포와 군산지역에 각각 20 %, 15 %의 양동작전을 실시했다. 그 외 지역에는 10 % 이하로 양동작전을 실시했다. 또한 4.4절의 학습된 모델과 같이 인천지역에서의 증원과 상륙해안을 오판케 하기 위해 인접지역인 남포와 군산지역으로 타격을 실시했다. 전투력 부분에서는 인천지역 전투력이 40 %에서 12 %로 가장 크게 감소했으며, 북한군은 처음에는 인천지역으로 상륙해안을 판단하다가 최종적으로 전투력에 의한 기만지수 판단 기준에 따라 남포지역 전투력이 10 %가 되어 상륙해안으로 판단했으며, 최종적으로 북한군은 상륙지역을 오판했다. 실제 양동작전에서는 상륙해안을 군산지역으로 오판케 하기 위함이었으나, 학습된 모델은 남포지역을 상륙지역으로 오판하게끔 하는 행동을 하였다.

## 6. 결론 및 향후 연구

군은 4차 산업혁명의 첨단기술을 기반으로 한 ‘AI 기술 강군 육성’ 구현을 비전으로 한 국방혁신 4.0을 추진하고 있다. 이러한 상황 속에서 CGF 자동화 모의 기술 개발은 미래 전장상황에서 지휘관의 전략·전술적 의사결정지원을 위한 인공지능 전투참모단으로 활용이 가능하며<sup>[31]</sup>, 나아가 CGF 지능을 고도화하여 능력 및 역할을 부여하게 된다면 다양한 분야에서 활용이 가능할 것이다.

본 연구는 CGF의 기만행위를 자동화 모의하기 위한 방법론을 설명하고, Unity 3D를 사용해 간략한 상륙양동 모델을 구현하였으며, 인공지능의 한 분야인 강화학습을 활용하여 기만행위를 학습할 수 있는 에이전트를 구현하였다. 결론적으로, 에이전트가 자신의 임무 목적과 기만 목적의 동시 달성을 위한 최선의 행동을 학습할 수 있음을 실험 결과로 제시하였다. 마지막으로 인천상륙작전으로 Case Study를 통해 학습된 모델을 적용하여 활용성을 점검하였다.

본 연구의 제한점과 향후 연구방향은 다음과 같다.

첫째, 연구의 가장 중요한 가정인 기만지수의 조건 설정이 술의 영역에 있기 때문에, 다른 가정의 정의에 따라 다른 연구 결과가 얼마든지 도출될 수 있다는 점이다.

둘째, 상륙양동 모델 구현 간 변수는 전투력만 포함했지만 다양한 변수를 활용해서 연구가 진행되면 더욱 성과 있는 연구가 될 것이다. 예를 들어, 적 증원 세력 및 예비대 능력, 적 접근로 위협순위 등이 있다.

셋째, 본 연구는 싱글 에이전트로서 아군 에이전트만 학습을 하면서 최적의 정책을 결정했다. 그러나 두 개 이상의 에이전트를 학습시키는 멀티 에이전트 강화학습(MARL : Multi-Agent Reinforcement Learning)을 구현하면 아군 에이전트의 종류를 다양화해서 협업을 통하여 보상을 얻을 수 있고, 적 에이전트도 학습을 하고 아군 에이전트와의 경쟁을 통해 높은 보상을 받게 할 수도 있다.

넷째, 본 연구는 기만행위 학습 방법에 대한 연구로 간략한 상륙양동 모델을 구현했지만, 해병대 상륙작전을 위한 CGF 자동화 모의 기술 연구가 계속해서 이루어진다면 상륙작전 전 단계에 걸쳐 활용이 가능할 것이며, 나아가 전방위 위협에 신속하게 대응이 가능한 스마트 국가전략기동군으로 해병대가 지향해야 할 목표와 방향에 한 단계 더 다가갈 수 있을 것이다.

## References

- [1] R.O.K Joint Chiefs of Staff Homepage, “Deception Operations,” <https://www.jcs.mil.kr/user/indexSub.action?codyMenuSeq=71157&siteId=jcs>, accessed 1 Nov. 2021.
- [2] S. B. Park, “The Importance of Military Deception and the Direction of Development of Our Military through the Transcription,” Study Thesis, Korea National Defense University, pp. 1-2, 2019.
- [3] R.O.K Joint Chiefs of Staff Homepage, “Demonstrations,” <https://www.jcs.mil.kr/user/indexSub.action?codyMenuSeq=71157&siteId=jcs>, accessed 1 Nov. 2021.
- [4] R.O.K Joint Chiefs of Staff Homepage, “Amphibious Demonstrations,” <https://www.jcs.mil.kr/user/indexSub.action?codyMenuSeq=71157&siteId=jcs>, accessed 1 Nov. 2021.
- [5] Institute for Military History, MND, “Inchon Landing Operation and Counterattack Operations,” pp. 122-124, 2009.
- [6] WIKIPEDIA Homepage, “Normandy Landings,” [https://en.wikipedia.org/wiki/Normandy\\_landings](https://en.wikipedia.org/wiki/Normandy_landings), accessed 1 Nov. 2021.
- [7] K. S. Yoon, “The Direction of Joint Military Deception for Securing Initiative,” Study Thesis, Korea National Defense University, pp. 1-10, 2014.
- [8] S. K. Lee, et. al., “A Study on the Application of Computer Generated Forces Technology to Defense,” Review of Korean Society for Internet Information, Vol. 13, No. 4, pp. 17-22, 2012.
- [9] N. S. Cho, et. al., “The Study on CGF Behavior Modeling Methodologies for Defense M&S: Focusing on Survey and Future Direction,” The Korea Society for Simulation, Vol. 29, No. 2, pp. 35-44, 2020.
- [10] R.O.K Marine Corps, “Marine Corps Vision 2049,” pp. 62-84, 2020.
- [11] S. H. Jung, “A Goal Oriented Action Planning and Replanning method of Computer Generated Forces in Wargame,” Journal of Korean Institute of Intelligent Systems, Vol. 23, No. 2, pp. 120-125, 2013.
- [12] C. H. Han, et. al., “A Study for Autonomous

- Intelligence of Computer-Generated Forces,” The Korea Society for Simulation, Vol. 21, No. 7, pp. 69-77, 2011.
- [13] J. P. Hong, et. al., “A Case Study on Planning and Learning for Large-Scale CGFs with POMDPs,” The Korean Institute of Information Scientists and Engineers Conference, pp. 983-985, 2016.
- [14] K. H. Lee, et. al., “Modeling Combat Entity with POMDP and DEVS,” Korean Institute Of Industrial Engineers, Vol. 39, No. 6, pp. 498-516, 2013.
- [15] S. Y. Jang, et. al., “Research Trends on Deep Reinforcement Learning,” Electronics and Telecommunications Trends, Vol. 34, No. 4, pp. 1-2, 2019.
- [16] W. J. Jung, et. al., “Build Reinforcement Learning AI Process for Cooperative Play with Users,” Journal of Korea Game Society, Vol. 20, No. 1, pp. 57-66, 2020.
- [17] H. V. Kim, et. al., “Trajectory Optimization of UAV Using Q Learning,” The Korean Society for Aeronautical & Space Sciences Conference, pp. 1108-1109, 2019.
- [18] R Core Team, “R: A Language and Environment for Statistical Computing,” <https://www.r-project.org/>, 2020.
- [19] R and RSudio Homepage, “Reinforcement Learning in R,” <https://cran.r-project.org/web/packages/ReinforcementLearning/vignettes/ReinforcementLearning.html>, accessed 10 Nov. 2021.
- [20] Wilensky, U, “Netlogo,” <http://ccl.northwestern.edu/netlogo/>, 1999.
- [21] Netlogo Homepage, “Reinforcement Learning Maze,” <http://ccl.northwestern.edu/netlogo/models/community/Reinforcement%20Learning%20Maze>, accessed 10 Nov. 2021.
- [22] Unity3D Homepage, “Unity 3D,” <https://unity.com/solutions/gaming-services>, accessed 10 Nov. 2021.
- [23] Unity3D Homepage, “ML-Agent Toolkit,” <https://unity.com/products/machine-learning-agents>, accessed 10 Nov. 2021.
- [24] MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc, “MATLAB & Simulink - MathWorks,” <https://kr.mathworks.com/products/matlab.html>
- [25] MATLAB Homepage, “Reinforcement Learning Toolbox,” <https://kr.mathworks.com/products/reinforcement-learning.html>
- [26] B. G. Park, et. al., “Implementation of Horror Game Using Unity Game Engine,” Proceedings of KIIT Conference, pp. 483-484, 2017.
- [27] H. B. Choi, et. al., “Design and Implementation of Reinforcement Learning Environment Using Unity 3D-based ML-Agents Toolkit,” KIPS Conference, Vol. 26, No. 1, pp. 548-551, 2019.
- [28] Institute for Military History, MND, “Inchon Landing Operation and Counterattack Operations,” pp. 87-143, 2009.
- [29] K. S. Kim, “The Inchoen Landing Operation and Response of North Korean People’s Army,” Journal of Military History Studies, Vol. 130, pp. 69-120, 2010.
- [30] J. S. Park, “Review of the Changsa Landing Operation during the Korean War,” Northeast Asian History Foundation, Vol. 69, pp. 243-282, 2020.
- [31] C. E. Lee, et. al., “Technical Trends of AI Military Staff to Support Decision-Making of Commanders,” Electronics and Telecommunications Research Institute, Vol. 36, No. 1, pp. 89-98, 2021.