

강화 및 진화 학습 기능을 갖는 에이전트 기반 합정 교전 시뮬레이션

정찬호^{1*} · 박철영¹ · 지승도¹ · 김재익²

The Battle Warship Simulation of Agent-based with Reinforcement and Evolutionary Learning

Chan-Ho Jung · Cheol-Young Park · Sung-Do Chi · Jae-ick Kim

ABSTRACT

Due to the development of technology related to a weapon system and the info-communication, the battle system of a warship has to manage many kinds of human intervention tactics according to the complicated battlefield environment. Therefore, many kinds of studies about M&S(Modeling & Simulation) have been carried out recently. The previous M&S system based on an agent, however, has simply used non-flexible(or fixed) tactics. In this paper, we propose an agent modeling methodology which has reinforcement learning function for spontaneous(active) reaction and generation evolution learning Function using Genetic Algorithm for more proper reaction for warship battle. We experiment with virtual 1:1 warship combat simulation on the west sea so as to test validity of our proposed methodology. We consequently show the possibility of both reinforcement and evolution learning in a warship battle.

Key words : Reinforcement learning, Evolutionary Simulation, Battle Warship

요 약

합정 전투체계는 무기체계, 정보통신 등의 기술 발전으로 인한 복잡한 전장 환경에 따라 인간이 개입하여 다양한 전술을 운용해야 한다. 따라서 에이전트 기반의 국방 M&S 시스템의 연구가 최근 들어 활발히 진행되고 있다. 그러나 현존하는 에이전트 기반 M&S 시스템은 고정된 전술을 적용하여 분석하는데 그치고 있다. 본 논문에서는 합정 교전에서 보다 적합한 대응을 찾기 위해 환경변화에 능동적으로 대처할 수 있도록 강화 학습 기능을 갖으며, 또한 유전 알고리즘을 이용하여 세대별 진화 학습 기능을 갖는 에이전트 모델링 방법론을 제안하였다. 타당성 검증에 위해 서해상에서 벌어지는 가상의 1:1 합정교전 시뮬레이션을 수행하였고, 이를 통해 합정 교전에 있어 강화 및 진화 학습이 가능함을 검증하였다.

주요어 : 강화 학습, 진화 시뮬레이션, 합정 교전

1. 서 론

다대다 플랫폼간의 교전을 포함하는 최근의 전투실험은 전투개체의 독립적/협력적인 의사결정과 개체간의 상호작용에 따른 고도의 복잡성을 보이고 있다. 이러한 다

양하고 복잡한 전투상황의 묘사를 위해 높은 자율성을 갖는 에이전트를 탑재한 국방 M&S(Modeling and Simulation) 시스템에 관한 연구가 활발히 진행되고 있다. 에이전트 기반의 시스템들은 인간의 의사 결정 요소를 모델링하여 전장의 개체들 간의 복잡한 상호 작용을 표현한다. 현재 에이전트 기술을 도입한 국방 M&S 시스템들로는 미국에서 인공생명개념이 도입된 ISSAC/EINSTEIN^[1]을 시작으로 뉴질랜드에서 개발된 MANA^[2], 호주에서 CROCADILE^[3], WISDOM-II^[4]가 개발되었고, 최근의 미국군에서 개발된 OneSAF^[5] 등으로 발전되었다. 하지만 기존 시스템들을 살펴보면 에이전트에 탑재된 고정된 전술을 적용하여 분

접수일(2012년 10월 5일), 심사일(1차 : 2012년 11월 21일),
게재 확정일(2012년 12월 24일)
¹⁾ 한국항공대학교 컴퓨터공학과
²⁾ 국방과학연구소
주 저 자 : 정찬호
교신저자 : 정찬호
E-mail; chanho66@kau.ac.kr

석하는데 그치고 있으며, 고도의 전문운용이나 다양한 전장 환경과 같은 실질적인 시뮬레이션 효과를 얻기에는 미흡한 실정이다. 이를 극복하기 위해 고도의 전문적 운용이나 다양한 전장 환경과 같은 복잡도가 높은 문제 해결을 위해 운용자 에이전트를 DEVS기반으로 수준을 향상시키고, 단일 함정 간의 함정교전에서 유전 알고리즘을 탑재한 통제관 에이전트를 사용하여 행동 중심적이고, 창발)적 전술 생성 가능성을 연구하였고⁶⁾, 다대다 함정 교전 시뮬레이션을 통해 창발적인 협업 전술 생성의 가능성을 검증한 바 있다⁷⁾. 본 연구에서는 전술 생성의 보완을 위해 강화 학습을 적용하여 학습과 진화라는 두 가지의 적응 과정을 보이고자 한다. 강화 학습은 개체의 일생동안에 이루어지는 반면 진화는 세대교체에 의하여 전반적으로 이루어진다. 학습과 진화와의 상호작용에 관하여 많은 연구가 이루어지고 있는데, 학습이 개체의 생존에 영향을 미친다면 학습 능력이 뛰어난 개체는 많은 자손을 갖게 되며 학습에 관련된 유전인자는 자손에게 전달되게 된다⁸⁾. 이는 학습뿐만 아니라 진화가 지능 향상을 위하여 중요한 역할을 담당함을 의미한다. 기존에 카드게임을 이용하여 강화 및 진화 학습 기능을 갖는 시뮬레이션 연구를 진행하여 그 타당성을 검증한 바 있다⁹⁾. 그러나 이 연구는 단일 에이전트 중심으로 이루어져서 복잡한 상황 묘사가 어렵다는 단점이 있다. 본 연구에서는 이를 확장하여 다중 에이전트 기반의 강화 및 진화 학습 기능을 갖는 시뮬레이션 연구를 진행하고자 한다. 이를 검증하기 위해 함정 운용에 필요한 함장, 정보장교, 항해장교, 포술장교 등의 에이전트로 계층 구조적으로 구성하였다. 그리고 환경변화에 능동적으로 대처하여 효율적인 규칙을 도출할 수 있는 강화 학습 기능을 갖는 학습 모듈을 적용하였고, 유전 알고리즘을 이용하여 각 세대 안에서의 학습과 세대교체에 따른 진화 학습을 통하여 기존의 단일 세대에서의 학습에 비해 전반적인 지능의 향상과 강건한(Robust) 지능을 갖는 에이전트 기반 함정 교전 시뮬레이션을 진행하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 강화 학습 및 유전 알고리즘의 관련연구를 소개하고, 3장에서는 강화 및 진화 학습 기능을 갖는 에이전트 기반 함정 교전 모델링 방법론에 대해 설명한다. 4장에서는 사례 연구를 통해 타당성을 검토하고, 5장에서 결론과 향후 연구방향을 기술한다.

2. 관련 연구

2.1 강화 학습

강화 학습(Reinforcement Learning)은 동적 환경 하에 있는 에이전트(Agent)의 행동에 대한 보상(Reward)을 최대화하는 상태-행동 규칙이나 행동 발생 전략을 찾는 학습법이다. 즉 보상의 최대화라고 하는 목적만을 가지고 있기 때문에 본질적으로 불확실성과 지연 보상이 존재하지만 환경으로부터의 보상과 벌칙만으로 우수한 정보를 획득할 수 있다는 장점 때문에 애초에 정답이 없는 문제의 접근의 길을 열어주고 있다. 강화 학습은 일반적으로 제어기 또는 에이전트의 행동에 대한 보상을 최대화하는 상태-행동 규칙이나 행동 발생 전략을 찾는 것이다. 그러나 많은 실제계의 경우에 있어서 목표 상태에 도달할 때까지는 중간 단계의 행동에 대한 즉각적인 보상이 주어지지 않는다. 이러한 경우 외부로부터의 강화 신호가 없기 때문에 학습이 일어나지 않게 된다. 그러한 경우에도 목표 상태에 도달하기 위해서는 지속적인 학습이 이루어져야 하므로 일시적인 신뢰 할당이 이루어져야 한다. 이것을 신뢰 할당 문제(Credit-Assignment Problem)라고 하며 강화 학습에 있어서 가장 중요한 문제라고 할 수 있다. 이 문제에 대한 가장 일반적인 접근 방법은 강화 신호를 생성하는 외부 평가 함수보다 더 자세한 정보를 얻을 수 있는 내부 평가 함수를 구현하는 것이다. 강화 학습법은 현재까지 Sutton의 TD Method에 의한 Actor-Critic 구조와 Watkins의 Q-learning, Modified Dempster-Shafer Theory 등이 활용되고 있다⁹⁻¹¹⁾.

국내외적으로 강화 학습을 사용한 국방 M&S 시스템에 적용된 사례를 살펴보면 국외에서는 군 부대의 계획 수립과 임무 수행을 찾기 위해 계층적 강화 학습¹²⁾을 적용하였고, 동적으로 변하는 환경에서 차량의 이동경로 최적화 문제를 자동화하는데 강화 학습¹³⁾을 사용하였다.

국내에서는 무인항공기의 미사일 회피 전략에 강화 학습을 사용하였고¹⁴⁾, 무인헬리콥터의 자동비행의 성능 개선을 위해 강화 학습이 적용된바 있다¹⁵⁾.

본 연구에서는 강화 학습을 위한 신뢰 할당 문제를 해결하기 위해 믿음 값과 가능성 값을 통한 증거기간 설정과 같은 효과적인 대안을 제공하는 Modified Demster-Shafer 이론을 적용하고자 한다.

2.2 유전 알고리즘

유전 알고리즘은 자연 선택의 원리와 자연계의 생물 유전학에 기본 이론을 두며 병렬적이고 전역적(Global)인

1) 창발이란 인간의 고정관념에 의해 부여된 것이 아닌, 개체간의 상호작용과 진화의 과정을 통해서 발현된 새로운 패턴(전술)을 말한다.

탐색 알고리즘으로서, 최근에 생물의 진화과정, 즉 자연선별과 유전법칙 등을 모방한 진화전략(Evolution Strategies), 유전 프로그래밍(Genetic Programming)등 여러 형태의 이론과 기법들이 최적해 탐색 알고리즘이 필요한 산업, 공학 등의 다양한 분야에서 활용되고 있다¹⁶⁾.

국내외적으로 국방 M&S 시스템에 유전 알고리즘을 적용한 사례를 살펴보면 국외에서는 적군 레이더들을 탐지하는 센서들의 최적의 탐 설계를 목적으로 유전 알고리즘¹⁷⁾을 적용하였고, UAV를 이용해 감시 임무를 수행하는 적합한 경로를 결정하기 위해 유전 알고리즘¹⁸⁾을 사용하였다.

국내에서는 소대이동시 적군의 포병공격으로부터 아군의 손실을 최소화하고, 목표지점까지 신속한 이동으로 적군보다 유리한 지점에 위치할 수 있도록 최적의 이동경로를 찾기 위해 유전 알고리즘¹⁹⁾을 적용하여 수행된 바 있으며, 해양환경에서 소나시스템의 효율적인 운영을 위해

유전 알고리즘을 이용하여 최적 탐색 경로를 설정하는 알고리즘²⁰⁾을 제시하였다. 따라서 본 연구에서는 복잡한 환경 묘사와 다양한 전술을 수행하고 분석해야 하는 국방 M&S 시스템에 적합한 기법이라 할 수 있는 유전 알고리즘을 적용하고자 한다.

3. 강화 및 진화 학습 기능을 갖는 에이전트 기반 함정 교전 모델링 방법론

본 연구에서는 강화 및 진화 학습 기능을 갖는 에이전트 기반 함정 교전 모델링을 위해 아래와 같은 모델 구조도에 따른 접근을 제안한다(Fig. 1 참조). 먼저 함정 교전 모델링을 위해 운용자 에이전트 모델, 함정 모델, 전장 환경 모델, 통제관 에이전트 모델들로 구성하였으며, 이를 살펴보면 다음과 같다.

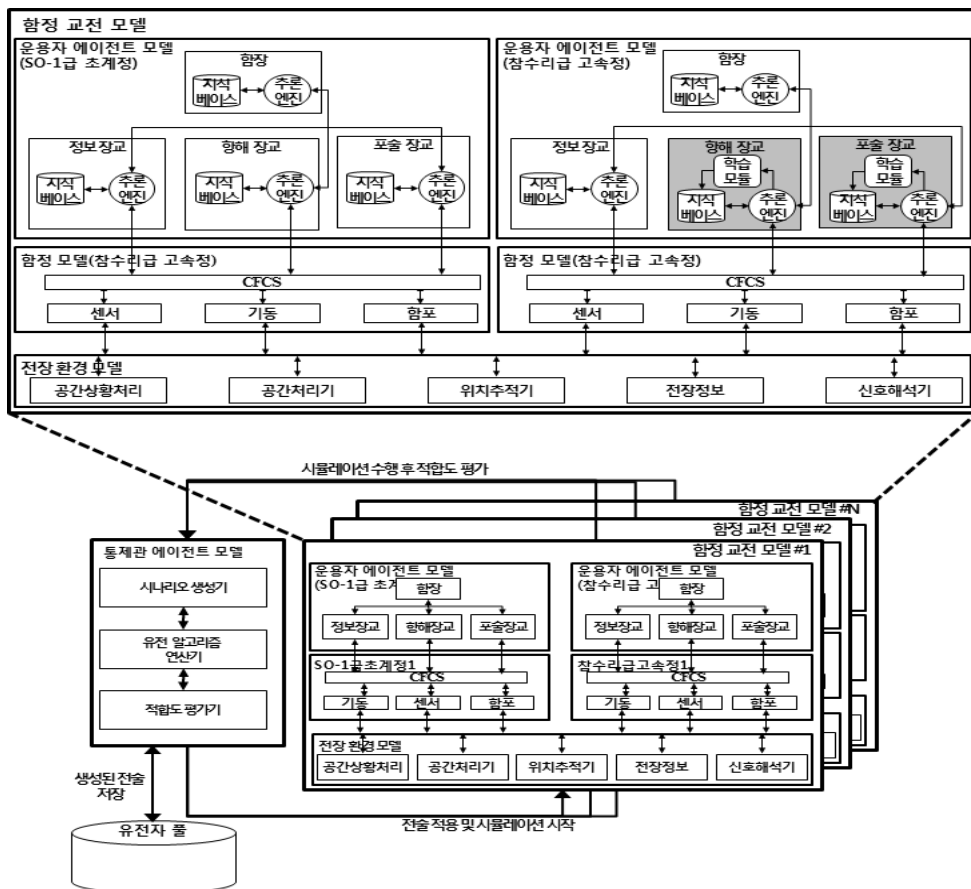


Fig. 1. The Structure of Agent-based Battle Warship Model with Reinforcement and Evolutionary Learning

- **운용자 에이전트 모델** : 아군, 적군 함정 모델들은 운용자 에이전트 모델의 의사결정에 따라 운용된다. 운용자 에이전트는 각 함정의 함장, 정보장교, 포술장교, 항해장교 등의 에이전트 모델 등으로 계층 구조적으로 구성하였다.
- **함정 모델** : 각 함정은 운용에 필요한 CFCS, 기동, 센서, 함포 모델 등으로 구성하였다.
- **전장환경 모델** : 함정 간 또는 함정과 환경 간의 상호작용에 대한 처리를 담당 한다. 공간상황처리, 공간처리기, 위치추적기, 전장정보, 신호해석기 등의 모델로 구성하였다.
- **통제관 에이전트 모델** : 시나리오를 생성하고, 제어하며, 교전결과 분석을 담당한다. 구성 요소로는 시나리오 생성기, 유전 알고리즘 연산기, 그리고 적합도 평가기로 구성하였다. 먼저 시나리오 생성기는 시나리오의 생성과 유전 알고리즘의 진화 세대(Generation) 수 및 개체(Population) 수를 결정하는 일을 담당한다. 유전 알고리즘 연산기는 시뮬레이션 제어 및 적합도 평가기로부터 얻어진 적합도를 통해 선택, 교차, 변이 등의 유전 알고리즘 연산 수행을 담당한다. 다음으로 적합도 평가기는 시뮬레이션의 평가와 최종 시뮬레이션 결과의 분석을 담당한다.

앞서 설명한 운용자 에이전트 모델 안에 탑재한 항해장교, 포술 장교 에이전트는 추론 엔진, 지식 베이스, 학습 모듈로 구성되어 있으며, 이를 살펴보면 다음과 같다(Fig. 1 참조).

- **학습 모듈** : 강화 학습을 위해 학습 모듈은 상황에 맞는 규칙을 적용하였는지에 대한 평가를 하여 신뢰도(Reliability) 변경을 통해 학습을 하고 지식 베이스에 업데이트 하는 역할을 한다.
- **추론 엔진** : 에이전트의 의사결정을 담당하는 기관으로 지식 베이스와 분리하여 구현함으로써 규칙이 변경되더라도 시스템의 변경 없이 동작이 가능하도록 한다. 추론 엔진은 순방향 추론(Forward Chaining Inference) 방법을 이용하여 문제를 해결하거나 의사결정을 하게 된다.
- **지식 베이스** : 에이전트의 의사결정에 필요한 전술을 저장한 지식 저장소로서 규칙(rule)을 탑재하고 있다.

학습 모듈의 규칙 신뢰도 할당 문제를 해결하기 위해

Modified Dempster-Shafer Theory를 사용한다. Modified Dempster-Shafer Theory에서는 (ef, ea, n, x) 4개의 요소로 신뢰도를 표현한다.

각 요소가 의미하는 바는 다음과 같다.

- ef : evidence for, 어느 정도 근거가 있다.
- ea : evidence against, 어느 정도 근거가 없다.
- n : natural evidence, 모르겠다.
- x : contradiction, 모순이 있다.

여기서 $0.0 \leq ef, ea, n, x \leq 1.0$ 이고, $ea+ea+n+x = 1.0$ 이다. 따라서 (ef, ea, n, x)가 (1, 0, 0, 0)이면 긍정(True), (0, 1, 0, 0)이면 부정(False), (0, 0, 1, 0)이면 모름(Unknown), (0, 0, 0, 1)이면 모순(Contradiction)을 의미한다.

학습 모듈의 강화 학습 과정은 Fig. 2와 같으며, 이를 포술 장교 에이전트로 예를 들어 설명하면 다음과 같다. 포술 장교 에이전트의 학습 모듈은 추론엔진으로부터 점화된 규칙을 전달 받는다. 아군이 i번째 공격을 수행하기 전 적군의 누적 피해량인 D_i 를 이용하여 단일 공격에 의한 피해량(단일 공격의 피해량은 0~9로 설정)의 값(VA: Actual Value)을 수식 (1)의 방법으로 구한다.

$$V_A = D_{i+1} - D_i \tag{1}$$

단일 공격에 의한 피해량의 기대값(VE: Expected Value)과 VA를 이용하여 계산된 결과를 Fuzzy화 한 값(Δe)을 신뢰도(R_{tran})의 형태로 변환한다(수식 (2) 참조). R_{tran} 과 점화된 규칙의 기존 신뢰도(R_n)를 Modified Dempster-Shafer Combine Algorithm을 사용하여 합성하고, 합성된 신뢰도(R'_n)를 점화된 규칙의 신뢰도에 반영한다.

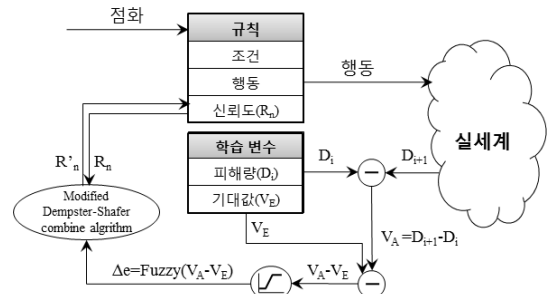


Fig. 2. Example of The Reinforcement Learning Process^[9]

$$\begin{aligned} \Delta e &= Fuzzy(VA - VE) \\ R_{tran} &= \begin{cases} (\Delta e, 0, 1 - \Delta e, 0) & \text{if } (\Delta e > 0) \\ (0, -\Delta e, 1 + \Delta e, 0) & \text{if } (\Delta e < 0) \end{cases} \\ R_n &= (ef_a', ea_a', n_a', x_a'), R_{tran} = (ef_b', ea_b', n_b', x_b') \\ R_n' &= \begin{cases} ef_a' = ef_a' * n_b' + ef_b' * n_a' + ef_a' * ef_b' \\ ea_a' = ea_a' * n_b' + ea_b' * n_a' + ea_a' * ea_b' \\ n_a' = n_a' * n_b' \\ x_a' = 1 - (ef_a' + ea_a' + n_a') \end{cases} \quad (2) \\ R_n' &= (ef_a', ea_a', n_a', x_a') \end{aligned}$$

위와 같은 과정으로 점화된 규칙에 의한 실제값을 이용하여 규칙의 신뢰도를 학습함으로써 다양한 규칙들 사이에서 우수한 규칙을 찾을 수 있다.

함장, 정보 장교, 항해 장교, 포술장교 에이전트의 지식 베이스 중 실험을 단순화하기 위해 휴리스틱과 랜덤으로 설정 해놓은 규칙 부분이 있으며(Table 1 참조), 휴리스틱의 구조는 조건(Condition)과 행동(Action)으로 구성되어 있다. 예를 들어 설명하면 정보장교의 보고가 ‘이상무’일 경우 함장 에이전트의 명령은 ‘순찰’을 내리게 된다. 랜덤 구조는 조건과 행동, 신뢰도로 구성되어 있다. 예를 들어 설명하면 함장의 명령 중 ‘공격’을 하달 받는 항해 장교 에이전트는 함정의 이동 속도를 초기에는 랜덤으로 선택하게 되고, 포술 장교 에이전트 또한 초기에는 함포의 종류, 함포의 사거리를 랜덤으로 선택하여 진행을 하게 된다. 당대의 환경반응에 적응하는 학습을 통해 항해 장교 에이전트는 규칙 신뢰도가 높은 함정의 이동 속도를 선택

하고, 포술 장교 에이전트 또한 강화 학습을 통해 규칙 신뢰도가 높은 함포 종류, 함포의 사거리를 선택하게 된다.

진화 학습 기능을 갖는 통제관 에이전트는 유전 알고리즘의 생성, 평가, 선택, 그리고 연산 등의 4단계 과정을 거친다. 이를 살펴보면 다음과 같다.

- **생성 단계** : 진화에 사용될 염색체는 에이전트 모델 내의 규칙들로 구성된다(Table 1 참조). 함장의 규칙은 해당 함정의 상태를 고려하여 명령을 결정한다. 이렇게 정의된 하나의 함장 규칙이 결합하여 함대의 전술이 생성된다.
- **평가 단계** : 교전 후 우수한 개체를 식별하기 위하여 적합도(F: Fitness)를 계산한다. 각 염색체의 적합도는 한 개체의 교전 시뮬레이션 종료 후에 마지막 피해량(D_i)으로 정의하였으며(마지막 피해량은 100을 넘을 수 없도록 설정), 수식 (3)과 같이 표현된다.

$$F = D_i \quad (0 < D_i \leq 100) \quad (3)$$

- **선택 단계** : 평가 단계에서 계산된 염색체의 적합도 평가를 기준으로 교차 및 변이 연산에 사용될 염색체를 선택한다. 룰렛 휠 선택 기법을 적용하여 염색체가 가지는 적합도 비율에 따라 선택될 확률이 높아지도록 설정하였다.

Table 1. Knowledge Base of Agent

(a) Captain Knowledge Base				(b) Intelligence Officer Knowledge Base			
규칙명	R1	R2	R3	규칙명	R1	R2	R3
구분				구분			
조 건	정보장교의 보고 = ‘이상무’	정보 장교의 보고 = ‘적군의 출현’	정보 장교의 보고 = ‘적군의 공격’	조 건	센싱 정보 = ‘이상무’	센싱 정보 = ‘적군 발견’	센싱 정보 = ‘적군의 공격’
행 동	명령 = ‘순찰’	명령 = ‘추적’	명령 = ‘공격’	행 동	함장의 보고 = ‘이상무’	함장의 보고 = ‘적군 출현’	함장의 보고 = ‘적군의 공격’

(c) Navigate Officer Knowledge Base				(d) Fire Control Officer Knowledge Base		
규칙명	R1	R2	R3	규칙명	R1	
구분				구분		
조 건	함장의 명령 = ‘순찰’	함장의 명령 = ‘추적’	함장의 명령 = ‘공격’	조 건	함장의 명령 = ‘공격’	
행 동	이동 속도 = ‘상’	이동 속도 = ‘상’	이동 속도 = ‘상’, ‘중’, ‘하’ 중 임의 선택	행 동	함포 종류(함포 성능, 장전속도) = ‘매우강력, 매우느림’, ‘강력, 느림’, ‘중간, 중간’, ‘약함, 빠름’, ‘매우약함, 매우빠름’ 중 임의 선택	함포 사거리 = ‘500’, ‘450’, ‘400’, ‘350’, ‘300’ 중 임의 선택
신뢰도	-	-	(0,0,1,0)	신뢰도	(0,0,1,0)	(0,0,1,0)

- 연산 단계 : 선택된 염색체를 짝을 지어 교차 및 변이 연산을 적용한다. 이 때 교차점의 위치는 임의로 선택 가능하고, 복수 선택도 가능하다. 염색체 구조와 교차 연산의 예를 Fig. 3(a)에서 보여주고 있다. 포술장교 에이전트의 부모 1과 부모 2간의 3번째 교차점(임의설정)을 기준으로 유전자를 교환함을 의미한다. 염색체들 사이에서 적합도에 영향을 미칠 수 있게 돌연변이는 0.03%의 확률로 발생하고, 돌연변이가 일어나는 포인트, 돌연변이 값 역시 랜덤하게 정하도록 하였다. 특정 유전자에서 변이가 발생할 경우 기존 행동을 새로운 행동으로 변경하게 된다. Fig. 3(b)는 염색체의 세 번째 유전자(R3)에 변이 연산이 적용된 예를 보인다. 그림에서와 같이 기존의 함포 종류 및 사거리가 변이 발생 후 함포 종류 및 사거리가 변경된다.

앞서 설명한 모델링 방법론을 적용하여 강화 학습 및 진화 학습을 통해 항해 장교 에이전트는 신뢰도가 높은 함정의 이동 속도를 선택, 포술 장교 에이전트 또한 신뢰도가 높은 함포 종류, 함포의 사거리를 선택하고 진화하였는지에 대한 여부를 사례 연구를 통해 검증하고자 한다.

교차점(임의설정)

구분	R1	R2	R3	R4	R5
조건	공격	공격	공격	공격	공격
행동	함포 종류 = '강력', '노림' 함포 사거리 = '350'	함포 종류 = '중간', '중간' 함포 사거리 = '500'	함포 종류 = '약함', '빠른' 함포 사거리 = '350'	함포 종류 = '중간', '중간' 함포 사거리 = '400'	함포 종류 = '약함', '빠른' 함포 사거리 = '350'
신뢰도	(0.5, 0, 0, 0)	(0.3, 0, 0.7, 0)	(0.3, 0.4, 0.1, 0)	(0, 0.4, 0, 0)	(0.7, 0, 0.2, 0)

구분	R1	R2	R3	R4	R5
조건	공격	공격	공격	공격	공격
행동	함포 종류 = '약함', '빠른' 함포 사거리 = '350'	함포 종류 = '중간', '중간' 함포 사거리 = '400'	함포 종류 = '약함', '빠른' 함포 사거리 = '350'	함포 종류 = '중간', '중간' 함포 사거리 = '400'	함포 종류 = '중간', '중간' 함포 사거리 = '500'
신뢰도	(0.5, 0, 0, 0)	(0, 0.4, 0, 0)	(0.7, 0, 0.2, 0)	(0.3, 0, 0.7, 0)	(0.3, 0.4, 0.1, 0)

↓ 교차점(임의설정)

구분	R1	R2	R3	R4	R5
조건	공격	공격	공격	공격	공격
행동	함포 종류 = '강력', '노림' 함포 사거리 = '400'	함포 종류 = '중간', '중간' 함포 사거리 = '500'	함포 종류 = '약함', '빠른' 함포 사거리 = '350'	함포 종류 = '강력', '노림' 함포 사거리 = '400'	함포 종류 = '중간', '중간' 함포 사거리 = '500'
신뢰도	(0.5, 0, 0, 0)	(0.3, 0, 0.7, 0)	(0.3, 0.4, 0.1, 0)	(0.3, 0, 0.7, 0)	(0.3, 0.4, 0.1, 0)

구분	R1	R2	R3	R4	R5
조건	공격	공격	공격	공격	공격
행동	함포 종류 = '약함', '빠른' 함포 사거리 = '350'	함포 종류 = '중간', '노림' 함포 사거리 = '400'	함포 종류 = '약함', '빠른' 함포 사거리 = '300'	함포 종류 = '중간', '중간' 함포 사거리 = '400'	함포 종류 = '약함', '빠른' 함포 사거리 = '350'
신뢰도	(0.5, 0, 0, 0)	(0, 0.4, 0, 0)	(0.7, 0, 0.2, 0)	(0, 0.4, 0, 0)	(0.7, 0, 0.2, 0)

(a) Crossover Operation of Fire Control Officer

규칙명	R1	R2	R3	R4	R5
구분	R3				
조건	공격				
행동	함포 종류 = '강력', '노림' 함포 사거리 = '400'				
신뢰도	(0.5, 0, 0, 0)				

↓ 돌연변이 연산수행

구분	R3'
조건	공격
행동	함포 종류 = '중간', '중간' 함포 사거리 = '400'
신뢰도	(0, 0, 1, 0)

(b) Mutation Operation of Fire Control Officer

Fig. 3. Example of Fire Control Officer Evolution Operation

4. 사례연구 : 1:1 함정 교전 시뮬레이션

4.1 시뮬레이션 시나리오 및 초기 조건

제안한 강화 및 진화 학습 기능을 갖는 에이전트 기반 함정 교전 모델링 방법론의 타당성 검증을 위해 1:1 함정 교전 시뮬레이션을 진행하였다. 시나리오는 서해 교전을 참조하여 적군 함정이 NLL을 침범하고, 아군 함정은 이를 방어하는 시나리오를 설정하였다. 전장 환경은 연평도 근해의 지형을 모의하였으며, Fig. 4와 같이 적군(SO-1급 경비정)과 아군(참수리급 고속정)의 함정은 각각 1척으로 설정하였다. 적군의 전술은 아군 함정을 공격하도록 설정하였고, 아군은 각 에이전트마다 특정 상황에서 취할 수 있는 행동을 정의하고(Table 참조), 학습 및 진화를 위해 임의의 행동들을 결합하여 유전 알고리즘의 염색체를 구성하였다. 실험을 단순화 하기 위해 진화 세대 수는 20세대, 개체 수는 40개로 설정하였다. 시뮬레이션 종료 조건은 한 쪽 진영의 함정이 침몰되거나, 포탄을 다 소진하였으나 침몰 시키지 못했을 경우 무승부로 설정하였다.

4.2 시뮬레이션 결과 분석

시뮬레이션 결과로 아군 함정의 세대별 전체 염색체의 평균 적군 피해량의 변화를 Fig. 5와 같이 그래프로 나타내었다. 1세대에서 약 57이던 평균 적군 피해량이 4세대 이후 83로 증가하였고, 그림에서 보는 바와 같이 14세대 이후 83에 수렴하는 모습을 확인할 수 있다.

세대 수 증가에 따른 교전에 대한 승리 횟수를 분석한 그래프는 Fig. 6과 같다. 여기서 '승리'란 적군의 피해량이 100이 되는 것을 말한다. 14세대 이후 아군의 승리 횟수가 적군보다 더 많아진 것을 확인할 수 있다.

강화학습의 검증을 위하여 휴리스틱한 규칙을 탑재한 적군과 학습 전/후의 규칙을 탑재한 아군과의 교전 실험을 진행하였다. 실험 결과 Fig. 7에서와 같이 1세대에서

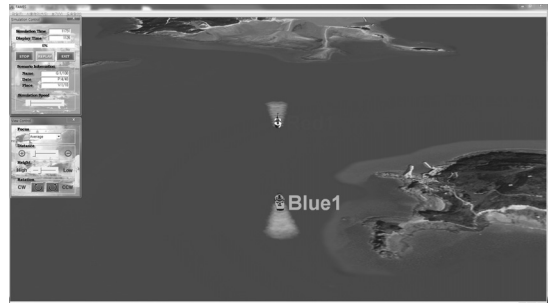


Fig. 4. GUI for Battle Warship Simulation

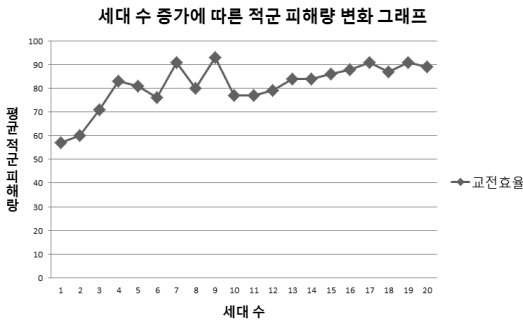


Fig. 5. Graph of relationship between increasing number of generations and average damage of the opponent (the enemy)

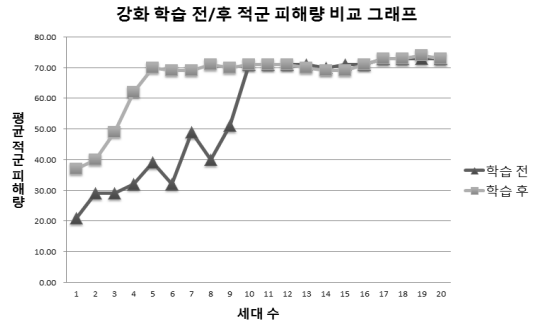


Fig. 7. Average damage comparison for each generation's learning effect

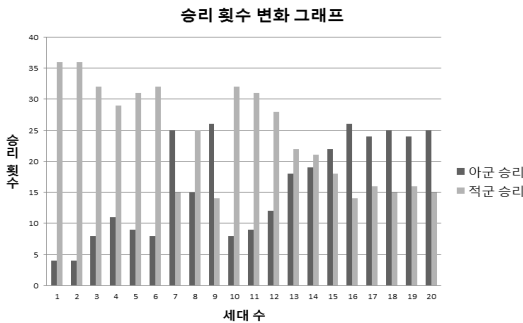


Fig. 6. Graph of count changed by increasing number of generations

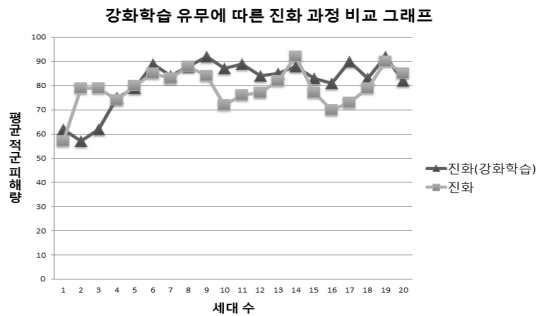


Fig. 8. Comparison of the opponent's average damage for absence and presence of Reinforcement learning of each generation

학습 전의 평균 적군 피해량이 21인 반면 학습 후 37의 학습 효과를 보였고, 2세대 또한 29에서 40의 학습 효과를 보였다. 전체적으로 1~9세대 까지의 학습 전/후를 살펴보면 10~35의 차이를 보이는 학습 효과를 보임으로서 학습 후의 규칙을 탑재하여 교전을 수행한 결과 평균 적군 피해량이 높게 나와 우수한 규칙을 찾았다는 것을 알 수 있었다.

평균 적군 피해량과 승리 횟수, 강화 학습 전/후 분석을 통해 공격 위주의 휴리스틱 전술로 설정된 적군과의 교전에서 강화 학습 및 진화 과정을 통해 적군의 피해량이 증가하는 결과를 보임으로써 적합한 이동 속도, 함포 종류, 사거리가 선택되었다고 할 수 있다.

본 연구에서는 진화 및 강화 학습을 이용한 실험이 지능 향상에 도움을 주는 지를 비교하기 위해 진화 학습을 이용한 실험을 진행하였다. 실험 결과 Fig. 8과 같으며 진화 학습만을 이용한 실험의 경우 50세대까지 실험을 진행한 결과 25세대 이후 평균 적군 피해량 70 ~ 90 사이의

결과를 반복적으로 보였다. 그에 반해 진화 및 강화 학습을 이용한 실험에서는 13세대 이후 평균 적군 피해량 80 이상을 나타냄으로써 더 효율적인 학습이 이루어짐을 검증 하였다.

5. 결 론

본 논문에서는 강화 및 진화 학습 기능을 갖는 에이전트 기반 1:1 함정 교전 시뮬레이션을 진행하였다. 이를 위해 학습 모듈을 추가하여 단일 세대에서의 학습 능력을 부여하였고, 또한 단일 세대에서의 학습 능력을 갖는 지능 개체 시스템을 다중 개체 시스템으로 확장하고 유전 알고리즘과 진화 모델링을 적용하였다. 제안된 시뮬레이션의 특징은 1) 교전중 적군의 행동 등과 같은 환경변화에 적응 하는 강화 학습과 세대별 적장 생존을 통한 진화 학습과정을 통합하여 학습기능을 극대화시킬 수 있는 에이전트를 성공적으로 제안했다. 2) 함정 교전 시뮬레이션

환경에 탑재하기 위한 에이전트 모델 구조도를 제안하고, 시뮬레이션을 통해 강화 및 진화 학습 기능의 효과를 검증하였다. 현재 1:1 함정교전에 적용된 강화 및 진화 학습 기능을 갖는 에이전트 기반 함정 교전 모델링 방법론을 향후 다대다 함정 교전에 적용하여 다양하고 실질적인 전술 생성에 관한 연구가 진행되어야 할 것이다.

사 사

본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다(계약번호 UD10004KD).

참 고 문 헌

1. A. Ilachinski, "Towards a Science of Experimental Complexity: An Artificial Life. Approach to Modeling Warfare", Proceedings of 5th Experimental Chaos Conference. Orlando. FL. 2000.
2. Michael Babilot, "COMPARISON OF A DISTRIBUTED OPERATIONS FORCE TO A TRADITIONAL FORCE IN URBAN COMBAT", The Master's Thesis of Naval Postgraduate School, September, 2005.
3. Easton A., and Barlow M., "CROCADILE: An Agent Based Distillation System Incorporating Aspects of Constructive Simulation", Proceedings of SimTect 2002, Melbourne Australia, 13-16 May, 2002.
4. Yang A., H. A. Abbass and R. Sarker, "Evolving Agents for Network Centric Warfare", Proceedings of the 2005 GECCO Conference on Genetic and Evolutionary, pp. 25-29, June 2005.
5. Lovgsdon J., D. Nash and M. Barnes, "OneSAF Tutorial", 2008 Defense Modeling and Simulation Conference (DMSC), Orlando, FL, USA, March 2008.
6. Yong-Jun You, Sung-Do Chi, Chan-Ho Jung et al, "A Study of Agent-based No-Human-in-the-Loop Battleship Warfare M&S System", The 7th Conference of marine weapon, pp. 29, 2008.
7. Chan-Ho Jung, Yong-Jun You, Sung-Do Chi et al, "Many-to-Many Warship Combat Tactics Generation Methodology Using the Evolutionary Simulation", Journal of The Korea Society for Simulation, Vol. 19, No. 3, pp. 79-88, 2011.
8. M. Mitchell, and S. Forrest, "Genetic Algorithms and Artificial Life", ArtificialLife, MIT Press, Cambridge, 1995.
9. Young-Kwang Kim, Jang-Se Lee, Sung-Do Chi, "Endomorphic Modeling of Intelligent Systems : Intelligent Card Game Players", Journal of KIISE, Vol. 26, No. 12, pp. 1507-1518, 1999.
10. L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning : a survey", Journal of Artificial Intelligence Research 4, 237-285, 1996.
11. B. P. Zeigler, "Some Properties of Modified Demster-Shafer Operators in Rule-Based Inference Systems", Int. J. General Systems, 1987.
12. Amandeep Singh Sidhu, Narendra S. Chaudhari, and Ghee Ming Goh, "Hierarchical Reinforcement Learning Model for Military Simulations", International Joint Conference on Neural Networks, Canada, July 2006.
13. NGAI Chi Kit, "Reinforcement-Learning-Based Autonomous Vehicle Navigation in a Dynamically Changing Environment", for the Degree of Doctor of Philosophy at The University of Hong Kong, November 2007.
14. Dong-Jin Lee, Hyo-Choong Bang, "Missile Evasive Strategies for Unmanned Aircrafts using Reinforcement Learning", The Conference of The Korean Society for Aeronautical & Space Sciences, pp 470-475, 2012.
15. Dong-Jin Lee, Hyo-Choong Bang, "Adaptive Linear Quadratic Regulator for Unmanned Helicopters via Reinforcement Learning", The Conference of The Korean Society for Aeronautical & Space Sciences, pp 1343-1346, 2011.
16. D.E. Goldberg, "Genetic Algorithm in Search, Optimization & Machine Learning", Addison-Wesley, 1989.
17. A.S. Yilmaz, B.N. McQuay, H. Yu, A.S. Wu, and J.C. Sciortino. "Evolving Sensor Suites for Enemy Radar Detection", In Genetic and Evolutionary Computation Conference Proceedings, Part II, pp. 2384-2395. July 2003.
18. S.W. Soliday. "A Genetic Algorithm model for mission planning and dynamic resource allocation of airborne sensors", In Proceedings, 1999 IRIS National Symposium on Sensor and Data Fusion, Laurel, MD, May 1999.
19. Sung-Young Lee, Sung-Ho Jang, Jong-Sik Lee, "Modeling and Simulation of Optimal Path Considering Battlefield-situation in the War-game Simulation", Journal of The Korea Society for Simulation, Vol. 19, No. 3, pp. 27-35, 2010.
20. Jung-Hong Cho, Jung-Hae Kim, Jea-Soo Kim et al, "Optimal Acoustic Search Path Planning Based on Genetic Algorithm in Discrete Path System", Journal of The Korea Society of Ocean Engineers, Vol. 20, No. 1, pp. 69-76, 2006.



정 찬 호 (chanho66@kau.ac.kr)

2006 청운대학교 컴퓨터공학과 학사
 2008 한국항공대학교 컴퓨터공학과 석사
 2008~현재 한국항공대학교 컴퓨터공학과 박사 과정

관심분야 : 모델링 및 시뮬레이션, 위게임



박 철 영 (momano@kau.ac.kr)

2012 한국항공대학교 컴퓨터공학과 학사
 2012~현재 한국항공대학교 컴퓨터공학과 석사 과정

관심분야 : 모델링 및 시뮬레이션, 위게임, 유전 알고리즘, 강화 학습



지 승 도 (sdchi@kau.ac.kr)

1982 연세대학교 전기공학과 학사
 1984 연세대학교 전기공학과 석사
 1985~1986 두산 컴퓨터 (현 한국 디지털) 근무
 1991 미국 아리조나대학교 전기전산공학과 박사
 1991~1992 미국 SIMEX Systems and S/W 회사 S/W 담당자로 근무
 1992~현재 한국항공대학교 항공전자 및 정보통신공학부 교수

관심분야 : 이산사건 시스템 모델링 및 시뮬레이션, 컴퓨터 보안, 지능시스템 디자인 방법론, 시뮬레이션 기반 인공생명, 교통 모델링



김 재 익 (jaeick@add.re.kr)

1990 경북대학교 전자공학과 학사
 1992 경북대학교 전자공학과 석사
 1992~현재 국방과학연구소 선임연구원

관심분야 : 함정 전투체계 무장통제장치 개발, 전투체계 모델링 및 시뮬레이션, 전투체계 교전 효과도 분석