ISSN 2383-6318(Print) / ISSN 2383-6326(Online)
KIISE Transactions on Computing Practices, Vol. 23, No. 6, pp. 343-349, 2017. 6
https://doi.org/10.5626/KTCP.2017.23.6.343

대화력전 및 기계화 보병 시나리오를 통한 대규모 가상군의 POMDP 행동계획 및 학습 사례연구

(Case Studies on Planning and Learning for Large-Scale CGFs with POMDPs through Counterfire and Mechanized Infantry Scenarios)

이 종 민 [†] 홍 정 표 [††] 박 재 영 [††] 이 강 훈 [††]
(Jongmin Lee) (Jungpyo Hong) (Jaeyoung Park) (Kanghoon Lee)

김 기 응 [†††] 문 일 철 [††††] 박 재 현 [†††††]
(Kee-Eung Kim) (Il-Chul Moon) (Jae-Hyun Park)

요 약 대규모 가상군의 전투 모델링 및 시뮬레이션에서 자율적으로 행동하는 이성적 전투 개체의 행동 묘사는 향후 발생할 전투의 작전을 고도화하고 효율적인 모의 훈련을 가능하게 하는 핵심 요소이다. DEVS-POMDP 계층적 프레임워크는 전투 행동 교범에 따른 상위 단계 의사결정 및 구체적 서술이 어려운 하위 단계 자율 행동계획을 각각 DEVS 및 POMDP로 모델링함으로써 대규모 가상군을 모의하였으나, POMDP 최적 행동정책 계산에 있어서 많은 컴퓨팅 자원를 필요로 하는 단점이 있었다. 본 논문에서는 DEVS-POMDP로 모델링된 대화력전 모의 시나리오 및 기계화 보병여단 공격작전 모의 시나리오의 사례연구를 통해 효율적인 POMDP 트리 탐색 알고리즘을 제안하고 적군 행동 양상 모델의 학습을 통한 가상군 전투 개체의 성능 향상을 확인한다.

키워드: 전투 개체 모델링, 모델링 시뮬레이션, POMDP, DEVS

Abstract Combat modeling and simulation (M&S) of large-scale computer generated forces (CGFs) enables the development of even the most sophisticated strategy of combat warfare and the efficient facilitation of a comprehensive simulation of the upcoming battle. The DEVS-POMDP framework is

· 본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다(UD140022PD)
· 본 연구는 방위사업청과 국방과학연구소의 지원으로 한국과학기술원 초고속비행체특화센터에서 수행되었습니다.
· 이 논문은 2016 한국컴퓨터종합학술대회에서 '대규모 가상군의 POMDP 행동계획 및 학습 사례연구'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 한국과학기술원 전산학부
jmlee@ai.kaist.ac.kr
†† 비 회 원 : 한국과학기술원 전산학부
jphong@ai.kaist.ac.kr
jypark@ai.kaist.ac.kr
khlee@ai.kaist.ac.kr
††† 종신회원 : 한국과학기술원 전산학부 교수(KAIST)
kekim@cs.kaist.ac.kr
(Corresponding author임)
†††† 비 회 원 : 한국과학기술원 산업 및 시스템 공학과 교수
icmoon@kaist.ac.kr
††††† 비 회 원 : 국방과학연구소 국방과학연구소
forehand@add.re.kr

논문접수 : 2016년 11월 9일
(Received 9 November 2016)
논문수정 : 2017년 2월 25일
(Revised 25 February 2017)
심사완료 : 2017년 3월 9일
(Accepted 9 March 2017)

Copyright©2017 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회 컴퓨팅의 실제 논문지 제23권 제6호(2017. 6)

proposed where the DEVS framework describing the explicit behavior rules in military doctrines, and POMDP model describing the autonomous behavior of the CGFs are hierarchically combined to capture the complexity of realistic world combat modeling and simulation. However, it has previously been well documented that computing the optimal policy of a POMDP model is computationally demanding. In this paper, we show that not only can the performance of CGFs be improved by an efficient POMDP tree search algorithm but CGFs are also able to conveniently learn the behavior model of the enemy through case studies in the scenario of counterfire warfare and the scenario of a mechanized infantry brigade's offensive operations.

Keywords: combat entity modeling, modeling and simulation, POMDP, DEVSverification

1. 서 론

컴퓨터 모의 가상군(Computer Generated Forces; CGF)을 이용한 전투모의 모델링 및 시뮬레이션은 국방 분야의 무기체계와 작전 등의 설계, 분석 및 평가에 있어 중요한 기술이다[1,2]. 특히, 이성적이고 지능적인 가상군 전투 개체의 행동 묘사는 지휘관들의 효율적인 전투작전 설정 및 모의훈련을 가능하게 함으로써, 향후 발생할 전투의 작전을 고도화할 수 있다. 기존의 군 전문가에 의해 작성된 전투 행동 교범(Field Manual)은 전투 개체의 행동을 명시적으로 서술하고 있지만, 전장에서 일어나는 모든 상황을 명시하는 것이 불가능하고, 전투 개체 단위 행동에 대한 구체적인 서술이 불가능한 단점이 존재했다.

명시적으로 서술될 수 없거나, 서술된 행동 이상의 지능적인 행동을 보여야 할 경우, 전투 개체는 인공지능 의사결정 이론에 기반하여 자율적으로 최적의 행동 정책을 계산하여야 한다. 부분관찰 마코프 의사결정과정(Partially Observable Markov Decision Process; POMDP)[3]은 개체가 행동하는 주변환경의 불확실성을 확률적으로 모델링하여 순차적인 행동 정책을 계산하는 기계학습 방법론으로, 산업공학, 인지과학 및 인공지능 분야에서 사용되는 자율의사결정 프레임워크이다. 가상군을 POMDP로 모델링함으로써 명시적으로 서술된 교범 없이, 주변환경에 대한 확률 모델만으로 지능적인 행동 정책의 계산이 가능하지만[1], POMDP 행동 정책 최적화 알고리즘(Planning)의 높은 계산 복잡도로 인해 행동 정책 계산이 어렵고, 전투시 지켜야 할 교범을 고려하지 못하는 단점이 있었다. 이로 인해 최근 대규모 가상군 모의 시 교범에 따른 상위 단계 의사결정을 이산 사건 시스템 명세(Discrete Event System Specification; DEVS)[4]를 이용하고, 교범에 구체적인 서술이 힘든 하위 단계 자율 의사결정은 POMDP를 이용하는 계층적 모델링 기법인 DEVS-POMDP 프레임워크가 제안되었다[2].

본 논문에서는 대화력전 모의 시나리오 및 기계화 보병여단 공격작전 시나리오 사례연구를 통해 DEVS-POMDP 기반 대규모 가상군의 실시간 전투를 모의한다. 특히, 복잡한 환경 상태를 가지는 POMDP 문제에 대해 효율적인 실시간 탐색 알고리즘을 제안하고, 적군 행동 모델에 대한 불확실성을 경험을 통해 스스로 학습(Learning)함으로써, 가상군 전투 개체의 성능을 향상시킬 수 있음을 보인다.

2. 대화력전 모의 시나리오

그림 1은 적군의 포격 도발 상황을 모의한 전투 시나리오이다. 적군은 지도의 북쪽 개머리 진지에서 6개 포대규모의 갱도포를 통해 포격 도발을 하고, 아군은 지도 남쪽에서 2개 포대 규모의 자주포로 대응 포격을 하게 된다. 각각의 포대는 6문의 갱도포 및 자주포로 구성되어 있으며, 전투 행동 교범에 따라 같은 포대에 속한 포들은 같은 행동을 수행하게 된다. 이 때, 아군 포대는 적군 포대에 비해 수적 열세인 상황에서 최대한 적은 피해로 많은 수의 적군을 포격해 공격 불능 상태로 만드는 것이 목표가 된다.

모의 시나리오는 적군의 최초 포격을 시작으로, 아군 연대 작전통제소의 지시 하에, 포격한 적군 갱도포대의 표적 획득 및 자주포를 통한 대응 포격을 진행한다. 포격 후, 아군은 주진지와 보조진지 사이를 이동하며 적군이 획득한 표적의 위치를 벗어나야 하고, 적군은 갱도 밖에서 포격 후 갱도 안으로 이동하여 아군의 포격을 방어함과 동시에 다음 사격에 대한 준비를 한다. 이 때, 갱도 밖에서 포격한 적군은 75%의 확률로 갱도 안으로 이동 및 방어 후 다시 갱도 밖으로 나가게 된다.

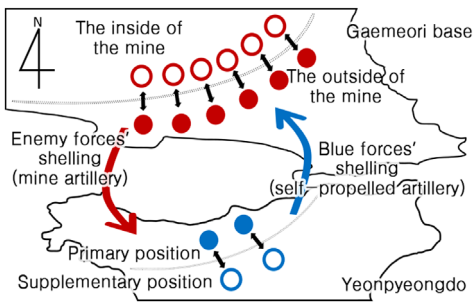


그림 1 대화력전 모의 시나리오

Fig. 1 Counterfire warfare simulation scenario

아군은 적군 행동 양상(즉, 갱도 안팎 이동 확률)을 고려하여 대응 포격 정책을 계획(Planning)해야 한다. 또한, 적군의 행동 양상에 대한 정보가 불확실할 경우, 적군의 행동 모델에 대한 학습을 동시에 고려할 수 있는 강화학습(Reinforcement Learning) 알고리즘이 필요하다. 본 논문에서는 적군 행동 양상을 알고 있다고 가정할 때 최적의 포격 정책을 계획하는 방법론과 함께, 적군 행동 양상에 대한 불확실성이 존재할 경우 필요한 강화학습 방법론을 제시한다.

2.1 POMDP 모델링을 통한 행동계획 및 학습 알고리즘

대화력전 모의 시나리오의 전투 시뮬레이션은 DEVS를 통해 모의된다. 이 때, 아군 자주포대의 자율적 포격 행동 전략은 POMDP로 계산되고, 이는 DEVS-POMDP 계층적 모델링[2]을 통해 구현가능하다. 다음의 내용에서는 아군 자율 포격 행동의 성능 향상을 위한 전투 개체의 POMDP 모델링, 최적 행동계획 및 학습 알고리즘을 제안한다.

2.1.1 POMDP 모델링 및 행동계획 알고리즘

POMDP는 <환경 상태, 행동, 관찰값, 상태 전이 확률, 관찰 확률, 보상 함수>로 구성된다. 대화력전 모의 시나리오의 아군 자주포대 POMDP 모델링은 표 1과 같다. POMDP는 환경 상태를 직접적으로 알 수 없고, 대신 관찰값으로부터 환경 상태들에 대한 확률 분포인 Belief-State를 계산하여 표현한다. 따라서, POMDP의 행동계획 문제란, 현재 Belief-State로부터 미래에 얻게 될 보상값의 합을 최대화하는 순차적 행동들을 효율적으로 계산하는 것이다.

POMDP 최적 행동계획을 위해서 몬테-카를로(Monte-Carlo) 트리 탐색을 이용한 POMDP 근사 알고리즘인 POMCP(Partially Observable Monte-Carlo Planning) [5] 알고리즘을 적용하였다. POMCP는 복잡한 환경 상태를 가진 문제에서 근사화된 Belief-State를 빠르게 계산할 수 있는 Particle Filtering 기법과 Monte-Carlo 시뮬레이션을 이용한 효율적인 트리 탐색을 적용해 최적의 행동 정책을 실시간으로 빠르게 찾을 수 있다.

2.1.2 POMDP 모델의 베이지안(Bayesian) 학습

앞서 POMCP 알고리즘을 이용한 대화력전 모의 시나리오의 최적 행동계획은 적군의 갱도 안 이동 확률 모델을 아군이 알고 있을 경우를 가정하였다. 하지만, 실제 전투시 적군의 정확한 확률 모델에는 불확실성이 존재하므로, 경험을 통한 강화학습으로 확률 모델의 불확실성을 줄여야 한다.

POMDP 모델의 효과적 학습과 전투 목표의 달성을 동시에 고려하는 방법론으로 베이지안 강화학습이 있다. 베이지안 강화학습은 모델의 불확실성을 환경 상태에 포함시켜, 동일한 문제를 증강 환경 상태 (Augmented State)를 가지는 행동계획 문제로 변환시켜 풀 수 있다. 이러한 베이지안 모델은 모델 학습과 전투 목표 달성의 두가지 행동 균형에 대해 베이지안 입장에서 최적 해를 줄 수 있다. 이에 따라, 대화력전 시나리오 POMDP 모델의 강화학습 문제는 Bayes-Adaptive POMDP 모델[6]의 최적 행동계획 문제로 새롭게 모델링하여 앞서 사용한 POMCP 알고리즘을 적용하였다. 베이지안 강화학습 모델링 및 트리 탐색을 이용한 최적 강화학습 행동계획 문제의 적용은 POMCP 알고리즘의 베이지안 강화학습 확장 알고리즘인 BAMCP(Bayes-Adaptive Monte-Carlo Planning)[7]를 참고하였다.

2.2 대화력전 모의 시나리오 실험 결과

그림 2는 대화력전 모의 시나리오에서 적군 행동에 대한 모델을 알고 있을 경우, 아군의 행동계획에 대한 실험 결과이다. POMDP 모델링을 통한 행동계획과 비교를 위해, 포격 후 진지 이동을 반복하는 규칙을 (Rule-Based) 가지는 모델과의 비교 실험을 수행하였다. 그림에서 공격 가능한 아군 자주포 수는 파란색으로, 공격 가능한 적군 갱도포 수는 빨간색으로 표현하였고, POMDP로 모델링한 아군의 행동계획 결과는 실선으로, 규칙을 기반으로

표 1 대화력전 모의 사나리오의 POMDP 모델링

Table 1 POMDP modeling of counterfire warfare simulation scenario

State	Position, availability to fire and move, and the number of available shells of each blue and red agents (The number of states ≈ 1034)
Action	Targetable enemy position per each artillery battalion with the number of shells to fire (0~3) / Target position to move (The number of actions = 2401)
Observation	States of blue agents and positions of the detected red agents (The number of observations ≈ 1011)
Transition Probability	Position is changed with respect to moving actions of blue and red agents. Firing ability or movability is lost with probability 0.7 when fired. Red agents move into the mine with probability 0.75.
Observation Probability	Blue agents are fully observable. Detection of shooting of red agents fails with probability 0.3.
Reward Function	(the number of red agents unable to fire) + (the number of blue agents able to fire) - (the number of blue agents' firing shells)

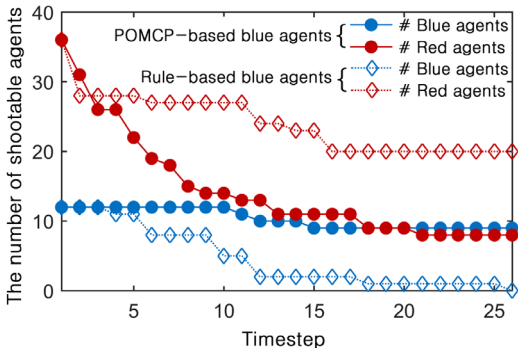


그림 2 규칙 기반 및 POMDP 행동계획 성능 비교 (대화력전 시나리오)

Fig. 2 Performance comparison between rule-based and POMDP-based agents

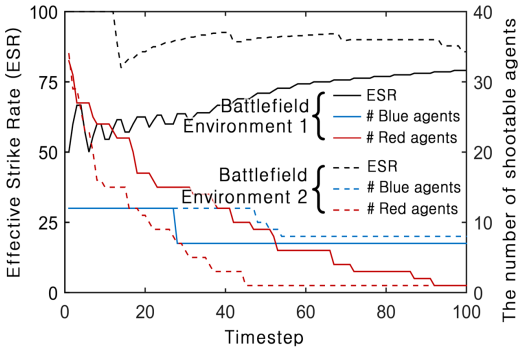


그림 3 서로 다른 환경 모델의 베이지안 강화학습(대화력전 시나리오)

Fig. 3 Bayesian reinforcement learning results for two different environment models

행동한 결과는 점선으로 표현하였다.

시나리오에서 아군은 수적 열세인 상황이므로 규칙 기반 행동은 25번의 교전 후 아군이 전멸하는 결과를 보이지만, 적군의 행동 양상을 고려한 POMDP 모델의 최적 행동계획은 아군 피해를 최소화하고 적군 포대를 공격 불능 상태로 만든다. 특히, POMDP 모델링을 통한 아군 행동은 협업을 통해, 한 포대가 적군 공격을 유인하는 동안 다른 포대가 적군을 포격하는 전략을 계산함으로써 아군의 피해를 최소화하였다.

그림 3은 적군 행동 양상의 불확실성이 있을 때, 베이지안 강화학습을 통해 최적 행동 정책을 계산한 결과이다. 실선으로 표시된 [전장상황1]에선 적군 행동 양상으로 90%의 확률로 갱도 안에서 두 차례 휴식기를, 10%의 확률로 즉시 갱도 밖으로 나온다. 점선으로 표시된 [전장상황2]는 반대로 10%의 확률로 갱도 안에 있고, 90%의 확률로 갱도 밖으로 나오게 된다. 아군은 적군의 두 행동 양상에 대해 사전 정보 없이 시작하여 (즉, [전장상황1]일 확률 0.5, [전장상황2]일 확률 0.5), 실제 적군의 행동 양상이 어떤 것인지에 대한 학습과 동시에 적군 포대에 대한 효과적인 포격을 수행한다. 그림에서 검은색으로 표현된 왼쪽 y축은 포탄의 활용 척도를 나타내기 위한 유효 타격률로 정의되고, 파란색 및 빨간색으로 표현된 오른쪽 y축은 아군 및 적군의 공격 가능한 포의 수를 나타낸다.

$$(\text{유효타격률}) = \frac{(\text{적군이 갱도 밖에 나와 있을 때 포격한 횟수})}{(\text{전체 포격 횟수})}$$

검은 실선으로 표현된 [전장상황1]의 유효 타격률은 적군이 높은 확률로 갱도 안에 있기 때문에 초반 유효 타격률이 낮지만, 적군이 갱도 밖으로 나올 확률을 학습함에 따라, 유효 타격률이 점점 높아짐을 알 수 있다.

3. 기계화 보병여단 공격작전 모의 시나리오

여단급 대규모 가상군의 자율 행동계획으로의 확장을 위해, 본 장에서는 포대 단위 전투인 대화력전 모의 시나리오를 확장하여 기계화 보병여단 공격작전 모의 시나리오의 DEVS-POMDP 계층적 모델링 및 POMDP 행동계획 알고리즘을 제안한다.

그림 4는 기계화 보병여단 공격작전 모의 시나리오이다. 아군은 북쪽에 위치한 적군이 점령한 진지를 향해 진격 후 적군 진지 탈취를 목표로 한다. 아군 및 적군은 전차, 장갑차, 자주포 및 박격포 등 서로 다른 11개 무기 체계를 보유하고 있다. 아군은 포병 3대대(18개체), 전차 및 장갑차 3대대(105개체)를 포함해 총 135개체를, 적군은 포병 3대대(18개체), 장갑차 3중대(36개체) 및 보병 2소대를 포함해 총 93개체로 구성되어 있으며, 같은 대대에 속해있는 각각의 전투 개체들은 전투 행동 교범에 따라 동일한 종류의 행동을 수행하게 된다.

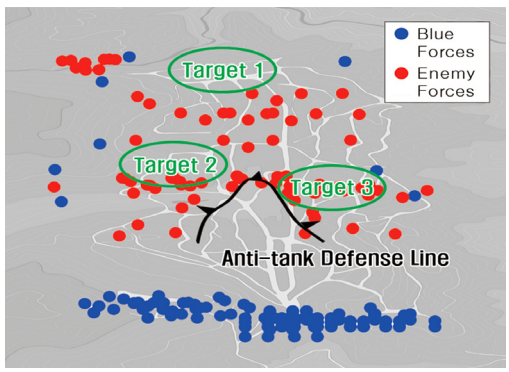


그림 4 기계화 보병여단 공격작전 모의 시나리오

Fig. 4 Mechanized infantry brigade's offensive operations scenario

시나리오는 아군 정찰팀이 적군의 위치를 파악한 뒤 적군 주요 화기에 포병 화력을 요청하면서 시작된다. 식별된 적군 주요 표적에 포병/박격포 화력이 집중된 후, 아군 1, 2 대대가 전진 공격하게 된다. 최초 적군 경보병 소대와 접촉하면 1 대대 2 중대와 2 대대 4 중대는 적 경계 부대를 격멸하고, 나머지는 주요 화기만 격멸하며 기동한다. 적군은 반땅크지탱점에 대전차화기 및 전차/장갑차를 집중 배치하여 아군의 점령으로부터 저항한다. 1대대가 목표 3을 점령하는 동안 2대대는 계속해서 공격하며 목표 2를 향해 전진하게 된다. 이후 1대대는 3대대가 목표 1을 향해 진격할 수 있도록 엄호하는 동안 2대대는 목표 2를 점령한다. 목표 2가 점령되면 마지막으로 통합 화력(포병+헬기)를 운용하여 3대대와 1,2대대 일부 잔여 병력이 목표 1을 향해 기동 및 점령하게 된다. 위 시나리오에서 아군은 최대한 적은 피해로 많은 수의 적군을 제압하며 목표 1을 점령하는 것이 목표가 된다.

본 기계화 보병여단 공격작전 모의 시나리오는 대화력전 모의 시나리오와 마찬가지로, DEVS를 통해 전투를 모의하고 아군 전투개체의 자율적 포격 행동 전략은 POMDP를 통해 계획하는 DEVS-POMDP 계층적 프레임워크로 모델링하였다.

3.1 POMDP 모델링 및 행동계획 알고리즘

기계화 보병여단 공격작전 시나리오는 아군 135개체와 적군 93개체로 구성된 여단급으로 대규모 확장된 모의 시나리오로서, 아군 모든 대대의 행동을 POMDP 행동 계획 문제로 푸는 것은 현실적인 시간 내에 어렵다. 따라서 전체 시나리오 중 포병대대의 행동 결정 부분에 대해서는 POMDP 행동계획 문제로 모델링 하고 나머지 다른 대대의 전투 개체에 대해서는 규칙 기반의 행동 정책을 정의하여 모의하였다. 본 시나리오의 아군 포병대대 POMDP 모델링은 표 2에 나타나있다. 여기서 아군 3개 포대만을 POMDP로 모델링했음에도, 대화력전 모의 시나리오와 비교해 매우 큰 POMDP로 모델링이 되었음을 알 수 있다.

이 때 POMCP와 같은 몬테-카를로 트리 탐색 알고리즘은 모든 행동을 최소 한 번 이상의 수행이 필요하고, 알고리즘의 성능은 수행 가능한 행동의 수가 많아짐에 따라 현격히 저하된다. 기계화 보병여단 공격작전 시나리오의 80여만 개 행동 공간은 각 행동 당 단 한 번의 액션을 시뮬레이션 하기 어려울 정도로 크다. 따라서 몬테-카를로 시뮬레이션 과정에서 최초 모든 행동을 한 번씩 수행하는 대신, 수행 허용 시간 동안 균일 샘플링(Uniform Sampling)으로 행동을 선택 및 시뮬레이션하고 허용 시간이 끝난 시점 최종적으로 가장 가치가 높은 행동을 선택하는 샘플 기반 탐색 알고리즘을 제안하였다.

3.2 기계화 보병여단 공격작전 시나리오 실험 결과

그림 5는 기계화 보병여단 공격작전 모의 시나리오에서 포격 행동에 따른 전투 개체의 피해 평가 모델을 알고 있을 때, 아군 포병대대의 행동 계획에 대한 실험 결과이다. BASELINE과 RANDOM 두 가지의 행동 결정 방식과 POMDP 행동 계획의 성능과 비교하였다. RANDOM은 각 포병 대대가 관측한 적군들 중 임의의 개체를 선택해 타격하며, BASELINE은 관측된 적군 개체 각각을 중심으로 하는 직사각형을 그려 범위 내에 가장 많은 적군이 포함된 개체를 타격하는 규칙으로 행동을 결정한다.

그림 5에서 볼 수 있듯이 POMDP 모델링을 통한 아군 행동은 RANDOM 및 BASELINE에 비해 더 빠른 속도로 적군을 제거해 나간다. 최종적으로 아군에게 피해를 입히지 않는 범위 중, 포격 위치의 오차까지 고려했을 때 확률적으로 가장 많은 적군을 동시에 타격할 수 있는 위치에 포격하는 전략을 계산함으로써 아군의 임무 수행을 성공적으로 돕는다. 반면 RANDOM은 때때로 아군까지 피해를 입는 범위에 타격할 수 있으며 BASELINE은 계산된 포대의 타격 범위가 겹치고 포격

표 2 기계화 보병여단 공격작전 사나리오의 POMDP 모델링

Table 2 POMDP modeling of mechanized infantry brigade's offensive operations scenario

State	Position (real numbers), and availability to fire and move of each blue and red agent (The number of states = 1044 × [0,1]2)
Action	Targetable enemy position per each artillery battalion (The number of actions = 804,357)
Observation	States of blue agents and states of the observed red agents (The number of total observations ≈ 1044)
Transition Probability	Each blue and red agent becomes immovable or destroyed according to the result of engagement
Observation Probability	Red agents being detected and blue agents are fully observable. No information of the red agents not being observed is given.
Reward Function	(the number of destroyed red agents) + (the number of immovable red agents) - (the number of destroyed blue agents) - (the number of immovable blue agents)

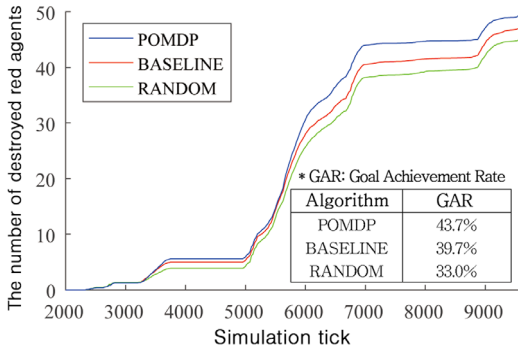


그림 5 규칙 기반 및 POMDP 행동계획 성능 비교(기계화 보병여단 공격작전 시나리오)

Fig. 5 Performance comparison between rule-based and POMDP-based agents

위치 오차가 고려되지 않아 POMDP에 비해 성능이 떨어진다. 결과적으로 POMDP 행동계획 알고리즘의 목표 도달률이 다른 비교 대상 알고리즘의 도달률을 상회하고 있다.

4. 결론 및 향후 연구계획

본 논문은 인공지능 의사결정 프레임워크 POMDP 모델링의 행동계획 및 베이지안 강화학습의 효과를 대화력전 모의 시나리오 사례연구를 통해 보여 주었다. 또한 포대급 모의 시나리오를 확장하여 여단급 모의 시나리오인 기계화 보병여단 공격작전 모의 시나리오에서도 POMDP 행동계획 알고리즘 적용이 가능함을 확인하였다.

POMDP 행동계획 알고리즘을 이용한 전투 개체는 적군의 행동 양상을 효과적으로 고려하여 전투 시뮬레이션의 성능을 향상시켰고, 또한 적군 행동 양상에 대한 불확실성이 존재하는 전장 상황에서는 적군의 행동 양상을 학습함으로써 전투의 불확실성에 강인하게 행동할 수 있음을 알 수 있었다.

향후 연구로 기계화 보병여단 공격작전 모의 시나리오와 같이 POMDP 모델의 행동 공간이 매우 큰 경우, 행동들을 군집화 하거나 계층적 구조를 활용하는 등 균일 샘플링보다 진보된 방식의 알고리즘을 이용하면 행동 계획 알고리즘의 효율을 높일 수 있을 것으로 기대된다.

References

[1] K. Lee, H. Lim, and K. Kim, "A Case Study on Modeling Computer Generated Forces based on Factored POMDPs," *Proc. of Korea Computer Congress 2012*, Vol. 39, No. 1(B), 2012. (in Korean)

[2] J. Bae, K. Lee, H. Kim, J. Lee, B. Goh, B. Nam, I, Moon, K. Kim, and J. Park, "Modeling Combat Entity with POMDP and DEVS," *Journal of the Korean Institute of Industrial Engineers*, Vol. 39, No. 6, pp. 498-516, Dec. 2013. (in Korean)

[3] E. J. Sondik, "The optimal control of partially observable Markov processes," Ph.D. thesis, Stanford University, 1971.

[4] B. P. Zeigler, H. Praehofer, and T. Kim, *"Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems,"* Academic press, 2000.

[5] D. Silver, and J. Veness, "Monte-Carlo planning in large POMDPs," *Proc. of Advances in Neural Information Processing Systems*, 2010.

[6] S. Ross, B. Chaib-draa, and J. Pineau, "Bayes-Adaptive POMDPs," *Proc. of Advances in Neural Information Processing Systems*, 2007.

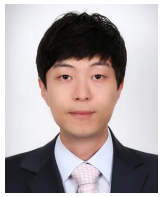
[7] A. Guez, D. Silver, and P. Dayan, "Efficient Bayes-Adaptive Reinforcement Learning using Sample-Based Search," *Proc. of Advances in Neural Information Processing Systems*, 2012.

이 종 민

2014년 서울대학교 컴퓨터공학부 졸업(학사). 2017년 한국과학기술원 전산학부 졸업(석사). 2017년~현재 한국과학기술원 전산학부 박사과정. 관심분야는 인공지능, 기계학습, 강화학습

홍 정 표

2014년 한국과학기술원 전산학과 졸업(학사). 2016년 한국과학기술원 전산학부 졸업(석사). 2016년~현재 한국과학기술원 전산학부 박사과정. 관심분야는 인공지능, 기계학습, 역강화학습



박 재 영

2008년 KAIST 전산학과, 수리과학과(학사). 2010년 KAIST 전산학과(석사). 2011년~2013년 서치솔루션. 2013년~2014년 LG전자. 2016년~현재 KAIST 박사과정 관심분야는 인공지능, 기계학습, 강화학습

이 강 훈

2008년 KAIST 전산학과, 수학과 졸업 (학사). 2008년~현재 KAIST 전산학부 석박사통합과정. 관심분야는 베이지안 강화학습, 마코프 의사결정과정

김 기 응

1995년 KAIST 전산학과 학사. 1998년 브라운대학교 전산학과 석사. 2001년 브라운대학교 전산학과 박사. 2001년~2003년 삼성 SDS 책임연구원. 2004년~2005년 삼성종합기술원 책임연구원. 2005년~현재 KAIST 전산학과 부교수. 관심분야는 인공지능, 기계학습, 강화학습

문 일 철

문일철 교수는 카네기멜론대학교에서 전산학 박사학위를 취득하고(2008), 현재 KAIST 산업 및 시스템공학과에 재직중 관심분야는 행위자 기반 시뮬레이션, 인공지능, 군사학, 정보학

박 재 현

1999년 1월~현재 국방과학연구소 책임연구원. 1992년 2월~1998년 12월 국방정보체계연구소. 관심분야는 SOA, IoT, M&S 등