강화학습 기반 무인 구급 로봇의 다중 부상자 구조 의사결정 최적화 연구

안희수, 이준형, 권현, 이준[1]
육군사관학교

heesuan10@gmail.com, junhyeong77428@gmail.com, hkwon.cs@gmail.com,
jun.lee.mistra@gmail.com

A Reinforcement Learning- Based Study on Optimizing Multi-Casualty Rescue Decision-Making for an Unmanned Medical Robot

Huisu An, Junhyeong Lee, Hyun Kown, Jun Lee
Korea Military Academy

요 약

본 논문은 다수 사상자 발생 상황에서 자율 의료구조 로봇이 제한된 임무 시간 내 구조 및 후송 우선순위를 최적화할 수 있는 강화학습 기반 의사결정 최적화 방법을 제안한다. 이를 위해서 제안한 시스템에서는 구조 임무를 마르코프 의사결정과정(MDP) 으로 모델링하고, 상태는 임무 시간과 함께 사상자별 거리, 지형 계수, 손상 유형(사지/접합부/흉부), 초기 생존확률, 처치 여부로 구성하였다. 사상자의 생존확률을 손상유형별 골든타임 및 시간에 따른 감소 모델로 정의하였다. 에이전트는 Maskable PPO 를 기반으로 학습하여 구조대상자가 남은 시간 내 후송이 불가능한 행동을 액션 마스킹으로 배제하여 현실적 제약을 반영하고 학습 효율을 향상시켰다. 또한 생존확률, 손상 가중치, 소요시간을 결합한 점수 기반 동적 Top-K 후보 제한을 적용해 탐색 공간을 줄이고 안정적인 정책 학습을 유도하였다. 시뮬레이션 기반 비교 실험에서 본 논문에서 제안한 강화학습 정책은 최근접 우선, 최대 생존확률 우선, 최소 생존확률 우선 등의 휴리스틱 대비 평균 생존자 수 및 보상에서 일관되게 우수한 성능을 보였고, 다양한 지형 분포에서도 강건함을 확인하였다.

Ⅰ. 서 론

본 현대 사회에서 과학기술의 발전과 인구 감소는 미래 전장을 변화시키는 두 개의 큰 축으로 작용하고 있다. 이에 따라 군에서는 전투력과 생존성을 보장하기 위해 워리어 플랫폼 기반의 차세대 전투체계를 구축하고 있으며, 디지털화된 전장 정보(생체 정보, 위치, 전투상황 정보 등)를 실시간으로 획득 및 공유하며 유기적인 상호작용으로 네트워크 중심전 환경을 발전시키고 있다. 더불어 유무인 체계를 통해 위험 지역에 대한 인명 손실을 최소화하고, 로봇과 인공지능의 능력을 전장에 적극적으로 도입하고 있다 [1].

이러한 변화 속에서 전장에서의 부상자 응급처치 및 후송 임무는 무인 자율체계가 담당하게 될 가능성이 높다. 전투 지역에서 부상자를 응급처치 및 후송하는 데에 사람의 역량만으로는 한계가 존재한다. 무인 체계는 물리적인 측면에서 인간의 한계를 보완할 수 있다. 또한, 판단의 측면에서 현실 전장에서는 부상자의 부상의 정도, 후송 지점까지의 거리, 지형 등 복합적인 요소들이 얽혀있기 때문에 사람이 즉각적으로 최적의 우선순위를 판단하기 어렵다. 또 제한된 시간 속에서 특정 지역 내 많은 부상자들에 대해 사람이 직접 조치를 취하기란 쉽지 않다. 전투원의 생존에 영향을 미치는 요소를 모두 고려한 의사결정 모델, 즉 강화학습 기반의 후송 정책 학습이 필요하게 될 것이다. 이는 전투력 유지와 직결되는 핵심 문제이다.

그러나 전장에서 부상자 후송 전략을 실제 환경에서 실험적으로 검증하는 것은 안전성·윤리적 측면에서 불가능하다. 따라서 전장 상황을 모델링하여 시뮬레이션 기반으로 정책을 검증하는 M&S(Modelling and Simulation)접근이 필요하다. 현대 군사 분야에서 M&S 는 실제 전투 상황을 모의 환경으로 재현하여 전술·작전 개념의 효과를 사전 검증하는 핵심 분석 도구로 활용되고 있다 [2]. 부상자 후송과 같은 실제 실험이 어려운 영역에서는 M&S 기반 시뮬레이션을 통해 후송 전략 및 자율체계 정책의 실효성과 적용 가능성을 검증할 수 있다. 본 연구는 전장 환경을 모델링하고

[1] 교신저자:이준, 안희수, 이준형은 동일한 비율로 논문 작성에 기여하였음.

구급후송 로봇을 에이전트로 설정하여 강화학습 시뮬레이션을 수행하였다.

Ⅱ. 본론

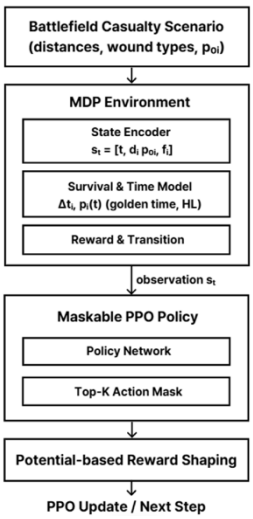

그림 1 제안 방법론 아키텍처

본 연구는 다수 사상자가 동시에 발생한 상황에서 자율 의료구조 로봇이 '누구를 먼저 구조하고 후송할지'를 최적화 할 수 있는 강화학습 방법을 제안한다. 제안한 시스템은 전장 구조 임무를 시뮬레이션 환경에서 의사결정 문제로 정식화하고, 생존에 영향을 주는 요인(시간, 거리, 지형, 손상 심각도)을 한꺼번에 고려하는 정책을 학습을 시키며 전체 시스템 구조는 그림 1 과 같다. 먼저 MDP 환경은 임무 시간 제한을 가지며, 매 의사결정 시점에서 구조로봇은 현재 시간과 각 사상자의 정보(거리, 지형 난이도, 손상 유형, 초기 생존확률, 이미 후송되었는지 여부)를 관측한다. 지형이 험할수록 이동이 느려지는 효과를 반영하기 위해, 단순 거리만이 아니라 구조자를 이송시에 몸무게에 대한 부담이 커지도록 설계한다. 로봇이 특정 사상자를 선택하면 해당 위치로 이동해 처치하고 후송하며, 이 과정에서 임무 시간이 소모된다. 또한 현실성을 위해, 남은 시간 안에 구조가 불가능한 사상자는 선택할 수 없도록 제한한다.

학습 알고리즘은 Maskable PPO 를 사용한다 [3]. 즉, 구조 과정에서 이미 후송한 사상자나 시간 제약상 실행 불가능한 선택지를 사전에 제거하는 행동 마스킹 기법을 적용한다. 추가로, 후보가 너무 많을 때는 생존확률과 손상 심각도, 이동 소요시간을 함께 고려한 점수로 상위 후보만 남기는 방식(동적 Top-K 후보 제한)을 적용해 탐색 효율을 높인다. 전이 과정에서는 후송 완료 시점의 생존확률에 따라 실제 생존 여부가 확률적으로 결정되도록 하여, 불확실성이 존재하는 전장 환경을 모사한다. 본 논문에서 보상 설계는 단순히 사람만 많이 구하는 것이 아니라, 구조 대상자의 부상 중요도(예: 흉부 손상 > 접합부 > 사지)를 가중치로 반영하여, 생존확률이 높기만 한 경상자를 무조건 우선하지 않도록 만든다. 또한, 구조대상자가 사망했더라도 도착 시점 생존확률을 일부 보상으로 주어 학습 신호가 끊기지 않게 한다. 특히 사상자 생존확률은 시간에 따라 감소하도록 모델링하며, 손상 유형별로 골든타임 및 half-life 개념을 반영해 심각한 부상일수록 더 빠르게 생존가능성이 떨어지도록 구성한다.

표 2 평지지형에서 생종자수 비교

Agent	평균 보상	생존자 수
RL	8.20	4.84
Nearest-Distance Greedy	8.04	4.68
Highest-Survival Greedy	6.16	4.82
Lowest-Survival Baseline	5.10	1.48

시뮬레이션 실험에서는 에피소드마다 사상자 배치와 조건을 무작위로 생성해 일반화 성능을 점검했고, 최근접 우선, 최대 및 최소 생존확률 우선 등의 휴리스틱 방법들과 비교했다. 그 결과 강화학습 정책이 표 1 과 같이, 평균 생존자 수와 평균 보상에서 전반적으로 더 우수했음을 알 수 있다.

Ⅲ. 결론

본 연구는 전장 환경에서의 구조 의사결정 문제를 강화학습 관점에서 모델링하고, 구급 로봇이 불확실한 상황 속에서도 최적의 후송 순서를 스스로 학습할 수 있도록 설계하였다. 부상자의 위치, 부상 부위, 생존율이 시간에 따라 비선형적으로 변화하는 복잡한 환경을 구축하였으며, 이를 Maskable PPO 기반의 강화학습 알고리즘으로 해결하였다. 학습 환경은 현실적인 제약을 반영했지만 다양한 기법으로 학습 안정성을 확보하였다. 또한, 액션 마스킹 기법을 통해 불가능하거나 비효율적인 행동을 사전에 차단함으로써, 탐색 효율성과 정책의 현실 적합성을 동시에 향상시켰다. 본 논문에서 제안한 시스템을 사용한다면 불확실성과 시간 압박이 공존하는 실전 상황에서도 강화학습 기반 로봇이 스스로 합리적 구조 결정을 내릴 수 있음을 알 수 있었다. 향후에는 다중 로봇 협업 구조, UAV·UGV 실시간 센서 연동 등을 통한 구조 시스템 개발을 수행할 예정이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2024-00455788).

참 고 문 헌

[1] Kwon, Y., Kim, T., Chae, J. W., & Kim, J., "Study on survival effectiveness of intelligent system for warrior platform by using AWAM", Journal of the Korea Institute of Military Science and Technology, 23(3), 277-285, 2020.

[2] Maria, A., "Introduction to modeling and simulation", In Proceedings of the 29th conference on Winter simulation (pp. 7-13), 1997.

[3] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O., "Proximal policy optimization algorithms", arXiv preprint, arXiv:1707.06347, 2017.