PPO를 활용한 APF 알고리즘 인력/척력 계수의 환경 적응형 자동 최적화 모델

이종훈, 이세비, 유현석*
*㈜빅앤딥

jhlee@bigndeep.co.kr, leesebi@bigndeep.co.kr, *hsyoo@bigndeep.co.kr

Environmentally Adaptive Automatic Optimization Model of the Attractive/Repulsive Force Coefficient of the APF algorithm using PPO

Lee Jong Hun, Lee Se Bi, Yoo Hyun Suk*
*BIG & DEEP Co., Ltd.

요 약

최근 방위 산업 업계에서는 무인 이동 객체를 활용한 군사 작전 수립 전략에 딥러닝(Deep Learning)을 결합한 방산 인공지능이 주목받고 있다. 방산 인공지능 중에서도 무인 로봇의 경로 계획(Path Planning) 알고리즘은 아군 객체의 적진 우회, 군수물자 운반, 적군 추적 등 다양한 전시 상황에 응용될 수 있다. 본 연구에서는 지역 경로 계획 알고리즘 중 하나인 인공 퍼텐셜 필드(Artificial Potential Field, APF)의 고도화를 위해 강화학습 방법론을 활용한 APF 인력/척력 계수 결정 모델을 제시한다. 제안 방안은 국지 최소값 진입 현상을 인지하고 이탈하는 알고리즘을 포함하고 있으며, 시뮬레이터 기반 강화학습을 통해 제안된 모델이 APF 평가 지표에서 종합적으로 안정적인 성능을 보임을 확인하였다.

Ⅰ. 서론

로봇의 경로 계획은 전역 경로 계획(Global Path Planning)과 지역 경로 계획(Local Path Planning)으로 분류된다. 전역 경로 계획은 전체 지도를 기반으로 장거리 단위의 중간 지점을 생성함으로써, 최단 경로를 따르는 동시에 목표 지점까지의 안정적인 이동을 보장한다. 지역 경로 계획은 전역 경로 계획에서 생성한 중간 지점 사이의 이동 경로를 결정한다. 지역 경로 계획은 장애물 회피 능력과 실상황에서의 실시간 대응 계획 등 여러 경로 계획 요소의 통합적 고려가 요구된다.

지역 경로 계획 알고리즘인 APF(Artificial Potential Field)는 가상의 인력과 척력 포텐셜 필드를 사용하여 경로를 생성하는 알고리즘이다. APF 는 타 알고리즘 대비 낮은 계산 복잡도로 부드러운 경로 생성을 가능케 하나, 국지 최소값 (Local Minima)진입 현상을 극복하지 못하여 경로 생성에 실패할 가능성이 있다[1]. 이러한 국지 최소값 진입 현상은 알고리즘 수정[2] 또는 강화학습 적용 방법론[3]을 활용하여 개선될 수 있다. 본 연구에서는 강화학습 알고리즘 중 PPO(Proximal Policy Optimization)[4]를 접목한 환경 적응형 최적화 모델의 학습 방안과 이를 학습시키기 위한 시뮬레이터 구조를 제안한다.

Ⅱ. 환경 적응형 APF 인력/척력 계수 최적화 모델

경로 계획 에이전트는 이동 객체가 출발지에서 목적지로 향하게 하는 인력 계수 K_{att} 와 장애물로부터 이동 객체를 밀어내게 하는 척력 계수 K_{rep}를 결정하여 2 차원 평면에서의 APF 이동 경로를 생성한다. 에이전트는 높은 경로 생성 성공률, 최대 장애물 근접 거리, 최단 경로, 최소 Local Minima 진입 횟수를 도출하는 APF 인력/척력 계수 선택을 목적으로 학습한다.

2.1 인력 계수 기반 척력 계수 정의

본 연구에서는 인력 계수와 척력 계수가 특정 범위 내에서 반비례 관계를 갖는다는 특성에 주목하여 두 계수를 선형식에 근사하여 정의하였다. 인력 계수를 $K_{att} \in [K_{att_{min}}, K_{att_{max}}]$, 척력 계수를 $K_{rep} \in [K_{rep_{min}}, K_{rep_{max}}]$ 로 나타낼 때, K_{rep} 은 다음과 같이 정의할 수 있다.

$$K_{rep} = K_{rep_{max}} - \frac{K_{rep_{max}} - K_{rep_{min}}}{K_{att_{max}} - K_{att_{min}}} \times \left(K_{att} - K_{att_{min}}\right)$$

2.2 강화학습 기반 인력/척력 계수 자동 최적화 모델

지역 경로 계획을 수행하는 에이전트가 특정 시점에서 APF 경로 계획을 생성하기 위해서는 과거 상태가 아닌 현재 상태, 즉 현재 시뮬레이터의 결과와 장애물의 위치만을 필요로 한다. 따라서 해당 과정은 마르코프 특성을 만족하므로 강화학습 문제로 모델링할 수 있다.

(x^*, y^*) 좌표축으로 이루어진 평면에서 장애물 $O_{i\in[1,N]}$ 가 $\left(O_{i_{x^*}}, O_{i_{y^*}}\right)$ 에 위치해 있을 때, 에이전트가 관측하는 상태(State) S는 다음과 같다.

$$S = [O_{1_{x^*}}, O_{1_{y^*}}, \cdots O_{N_{x^*}}, O_{N_{y^*}}, K_{att}]$$

APF 에서는 장애물의 척력 세기에 따라 에이전트의 이동 경로가 달라지므로, 여러 장애물 간의 밀도와 분포는 에이전트의 학습에 있어 유의미한 입력으로 작용할 수 있다. 본 논문에서는 다수의 장애물과 에이전트의 상대적인 좌표를 일관된 상태로 표현하기 위해 기존 평면좌표계를 대체한 (x^*, y^*) 좌표계를 활용하였다. x^* 축은 기존 좌표계에서의 시작점과 도착점을 지나는 직선이며, 기존 시작점과 장애물들 간

거리를 비교하였을 때 가장 작은 Δx값을 갖는 장애물이 포함된 x^*축의 수직선을 y^*축으로 정의한다.

에이전트는 인식된 상태에 대해 실수 공간 $[K_{att_{min}}, K_{att_{max}}]$ 내에서 인력 계수 K_{att}를 추론하고, K_{att}와 K_{rep}에 따라 생성된 APF 에서 이동한 결과를 보상으로 얻는다. 경로 생성 여부 $P_{success}$ 에 따른 보상함수는 다음과 같다.

$$R = \begin{cases} r_{success} - \alpha \times \dfrac{D_{path} - D'}{D'} - \beta \times \dfrac{1}{{D_{obs}}^2} - \gamma \times N_{LM}, & if\ P_{success} \\ r_{fail}, & otherwise \end{cases}$$

경로 생성이 가능한 인력계수가 선택되었다면 에이전트는 기본 보상 $r_{success}$를 받게 되며, 해당 경로의 길이 D_{path}가 시작점과 도착점을 잇는 최소 거리 D'에 대해 차이가 적을수록 높은 보상을 얻는다. 에이전트는 가장 가까운 장애물과의 근접 거리인 D_{obs}가 작을수록 높은 페널티를 받으며, 국지 최소값 진입 횟수 N_{LM}이 적을수록 높은 보상을 얻을 수 있다.

2.3 국지 최소값(LM) 진입 감지 및 이탈 알고리즘

본 연구에서는 지역 경로 계획 에이전트의 국지 최소값 진입 현상을 완화하기 위해 LM 진입 감지 및 이탈 알고리즘[1]을 추가한 APF 시뮬레이터를 사용하였다. 에이전트가 최근 m 포인트 이내에 방문한 기록이 있는 위치에 다시 방문한다면, APF 시뮬레이터는 에이전트가 LM 상태에 진입한 것으로 판단한다. 이 때 시뮬레이터는 LM 발생 지점 주변의 포텐셜 값을 기존 최대값보다 높은 임의의 값으로 수정하여 에이전트로 하여금 해당 위치를 벗어나게 한다.

Ⅲ. 모의실험 실행 및 결과

강화학습 에이전트는 무작위로 분포된 N 개의 장애물 환경에서 학습한다. 본 모의실험에서는 $N \in [2,5]$ 개의 장애물이 무작위로 혼합되어 있는 2×10^4 개의 환경에서 학습을 진행하였다.

시뮬레이션에 사용된 인력 및 척력 계수는 이동 경로 길이가 약 $1\,km$, 장애물의 척력 적용 반경이 $100\,m$인 경우를 가정하여 각각 $K_{att} \in [0.01,\ 0.05]$, $K_{rep} \in [10^4,\ 5 \times 10^5]$ 로 설정하였다.

표 1 은 $N \in [2,\ 5]$ 혼합 환경에서의 비교군 대비 제안 방안의 성능 비교를 나타낸다. 성능 평가는 10^4 번의 테스트 환경에서 진행하였으며, 평균 경로 길이 D^*_{path}, 평균 장애물 근접거리 D^*_{obs}, 평균 LM 진입 횟수 N^*_{LM}, 평균 보상 R^*, 성공률 q 를 기준으로 학습 성과를 판단한다. $b_{i\in[1,B]}$ 는 인력 계수 범위를 B등분하여 만든 인력 및 척력 계수 조합 B개 중 i번째 조합을 나타내며, 본 실험에서는 $B = 5$로 설정하여 5개의 비교군에 대한

	b_1	b_2	b_3	b_4	b_5	제안방안
$D^*_{path}(m)$	1173	1125	1093	1057	1009	**1086**
$D^*_{obs}(m)$	97.40	85.24	75.21	63.65	41.02	**82.09**
N^*_{LM}	2.937	1.598	0.971	0.401	0.048	**0.358**
R^*	4.354	6.400	7.075	7.303	3.829	**8.004**
$q(\%)$	79.36	90.62	95.10	98.52	100.0	**99.55**

표 1. 혼합 환경에서의 비교군 대비 제안 방안 성능

제안방안의 성능을 평가하였다. 제안 방안으로 학습된 에이전트는 높은 경로 생성 성공률, 먼 장애물 근접 거리, 짧은 경로 길이, 최소 LM 진입 횟수를 동시에 갖는 인력 및 척력 계수를 선택함을 알 수 있다.

Ⅳ. 결론

본 연구에서는 5 개 이하의 장애물 배치 환경에서 APF 알고리즘의 인력 및 척력 계수를 적응적으로 결정하는 환경 적응형 APF 강화학습 모델을 제시하였다. 시뮬레이션 기반 강화학습 접근법을 통해 APF 시뮬레이터를 개발하였으며, 연속 공간에서의 PPO 알고리즘을 적용하였다. 행동 공간을 1 차원으로 축소하여 학습 난이도를 완화하고, 장애물 위치 정보를 전처리하여 분포 패턴을 정규화함으로써 유사한 구조적 특성을 갖는 환경을 동일하게 처리할 수 있도록 하였다. 결과적으로 다양한 장애물 분포 환경에서 장애물 배치에 강건하게 대응하며 효과적인 경로 계획을 생성할 수 있는 APF 에이전트를 강화학습을 통해 구현하였다.

본 연구의 결과를 바탕으로, 적군과 같은 위험요소가 존재하는 전장 환경, 고밀도 도시 환경, 안티드론 분야의 레이더 탐지 영역 분포 환경 등의 다양한 3 차원 실제 환경 하에서 실용성 있는 기능을 제공하는 강건한 강화학습 모델 연구를 할 계획이다. 또한 객관적인 성능 분석을 위하여 복잡한 실제 환경에서 D-APF(Dynamic Artificial Potential Field) 등 최신 알고리즘과 강화학습 에이전트 간의 상세한 비교 검증을 할 계획이다.

ACKNOWLEDGMENT

본 결과물은 방위사업청의 재원으로 국방기술진흥연구소의 방산혁신기업 100 프로젝트 기술개발 전용지원사업의 지원을 받아 연구되었음 (과제관리번호: R230106).

참 고 문 헌

[1] Park, Min Gyu, and Min Cheol Lee. "A new technique to escape local minimum in artificial potential field based path planning." KSME international journal 17 (2003): 1876-1885.

[2] Jayaweera, Herath Mpc, and Samer Hanoun. "A dynamic artificial potential field (D-APF) UAV path planning technique for following ground moving targets." IEEE access 8 (2020): 192760-192776.

[3] Shao, Mingzhi, et al. "Research on UAV Trajectory Planning Algorithm Based on Adaptive Potential Field." Drones 9.2 (2025): 79.

[4] Sakai, Atsushi, et al. "Pythonrobotics: a python code collection of robotics algorithms." arXiv preprint arXiv:1808.10703 (2018).

[5] Schulman, John, et al. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).