DQN 기반 커버리지 경로 계획에서의 하이퍼파라미터 분석

남승우, 유경민, 박재원, 김성현, 백의준, *김명섭

고려대학교

{nam131119, rudals2710, 2018270614, honeybeelawn, pb1069, tmskim}@korea.ac.kr

Hyperparameter Analysis for DQN-based Coverage Path Planning

Seung-Woo Nam, Gyeong-Min Yu, Jae-Won Park, Seong-Hyeon Kim, Ui-Jun Baek, Myung-Sup Kim*

Korea Univ.

요 약

커버리지 경로 계획(Coverage Path Planning, CPP)은 주어진 영역을 누락 없이 효율적으로 탐색하는 문제로, 자율 주행 로봇, 무인 항공기, 청소 로봇 등 다양한 응용 분야에서 핵심적인 과제로 여겨지고 있다. 최근 강화 학습 기반 접근이 주목받고 있으며, 그 중 Deep Q Network(DQN)[1]는 고차원 상태 공간에서도 효과적인 경로 탐색을 할 수 있다. 본 연구에서는 DQN 기반 CPP의 성능에 영향을 주는 주요 파라미터에 대한 분석을 수행하였다. 2차원 격자 환경에서 에이전트가 Deep Q Network를 기반으로 탐색을 수행한다. 이를 평가하기 위해 커버리지 비율, 누적 보상 등의 지표를 기반으로 성능을 비교하였다. 실험 결과, 각 파라미터의 선택이 CPP 성능에 영향을 끼치며 특정 파라미터 조합이 일반화된 성능 향상에 기여할 수 있음을 확인하였다.

Ⅰ. 서 론

Coverage Path Planning(CPP)은 주어진 공간을 중복 없이 완전하게 탐색하는 경로를 생성하는 문제이다. 로봇 공학, 자율주행, 정밀 농업, 청소 로봇 등 다양한 문제에 적용되고 있다. 전통적인 CPP 기법[2]은 휴리스틱이나 탐색 기반 알고리즘을 기반으로 하지만, 이는 복잡한 환경이나 동적인 상황에 대응하는 데에 한계를 가진다.

이를 해결하기 위해 인공 신경망을 사용한 Deep Q Network(DQN)이 제안되었다. DQN은 현재 상태와 행동에 대한 미래 가치를 예측하여 최적의 행동을 할 수 있게 하는 방법으로, 인공 신경망을 통해 복잡하고 상태의 가지 수가 많은 환경에서 효과적으로 사용할 수 있다. 그러나 DQN의 학습 성능은 다수의 하이퍼파라미터 설정에 크게 의존한다. 학습률(learning rate), 리플레이 메모리 크기, 배치 크기, 타겟 네트워크 업데이트 주기 등의 설정은 학습 수렴 속도와 정책의 질에 직접적인 영향을 미치며, 적절한 파라미터 조합을 찾는 것이 필요하다.

본 논문에서는 DQN 기반 CPP 알고리즘의 성능을 영향을 주는 주요 파라미터들을 체계적으로 분석하고, 파라미터 조합이 커버리지 비율, 누적 보상 등에 미치는 영향을 실험적으로 평가하여 CPP에 특화된 DQN의 하이퍼파라미터 튜닝 전략을 제시한다.

Ⅱ. Coverage Path Planning

CPP는 주어진 공간 내의 모든 영역을 누락 없이 완전하게 커버하도록 경로를 계획하는 문제로, 모바일 로봇, 산업 자동화, 청소 로봇, 농업 등 다양한 분야에서 중요한 과제로 여겨진다. CPP의 핵심 목표는 커버리지의 완전성을 유지하면서도 경로의 길이, 중복 탐색, 에너지 소비 등을 최소화하는 최적의 경로를 생성하는 것이다.

강화 학습 기반 CPP는 기존의 전통적인 알고리즘의 문제를 해결하기 위한 방법론이다. RL 기반 접근은 사전에 명시적인 경로를 정의하지 않고, 상태-행동-보상 구조를 통해 최적의 행동 정책을 학습한다는 점에서 기존 방식과 본질적인 차이를 가진다. 특히 Deep Q Network는 심층 신경망을 사용하여 복잡한 상태 공간에서도 효과적인 경로 생성을 가능하게 한다.

본 논문에서는 2차원 격자 환경에서의 CPP 문제를 다룬다. 2차원 격자 환경은 이동 방향은 4방향으로 위, 아래, 왼쪽, 오른쪽이다.

Ⅲ. Deep Q Network

DQN는 강화학습에서 Q-learning 알고리즘과 딥러닝을 결합한 방식으로, 에이전트가 고차원 상태 공간에서도 효율적으로 최적 정책을 학습할 수 있도록 설계된 기법이다. 이는 특히 환경 모델이 없거나 매우 복잡한 상황에서 강력한 탐색 능력을 발휘한다.

Q-learning은 상태(state)- 행동(action) 쌍에 대한 가치 함수인 Q-value를 학습함으로써 최적 정책을 추론하는 방식이다. 이 때 Q-value는 (1) 식으로 정의한다.

$$Q(s,a) = E(R_t | s_t = s, a_t = a) \qquad (1)$$

전통적인 Q-learning은 Q 테이블을 활용하지만, 상태 공간이 커질수록 계산량과 메모리 요구가 폭증한다. 이를 해결하기 위해 DQN은 Q 함수를 딥러닝 모델로 근사한다. 핵심 요소는 아래와 같다.

- Experience Replay : 경험 데이터를 버퍼에 저장한 후, 무작위 샘플을 통해 학습
- Target Network : 타겟 Q 값 계산 시 별도 네트워크를 사용
- e-greedy Exploration : 탐험과 활용을 균형 있게 조절하는 정책

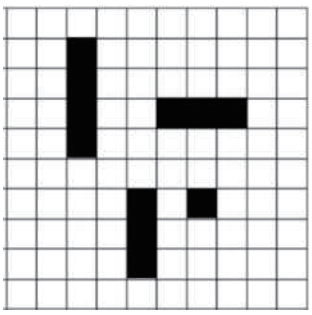

그림 1 10 x 10 크기의 맵 정보

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(00230661, 하이브리드 양자키분배 방법 및 망 관리 기술 표준개발) 및 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(00235509, ICT융합 공공 서비스• 인프라의 암호화 사이버위협에 대한 네트워크 행위기반 보안관제 기술 개발)을 받아 수행된 연구임.

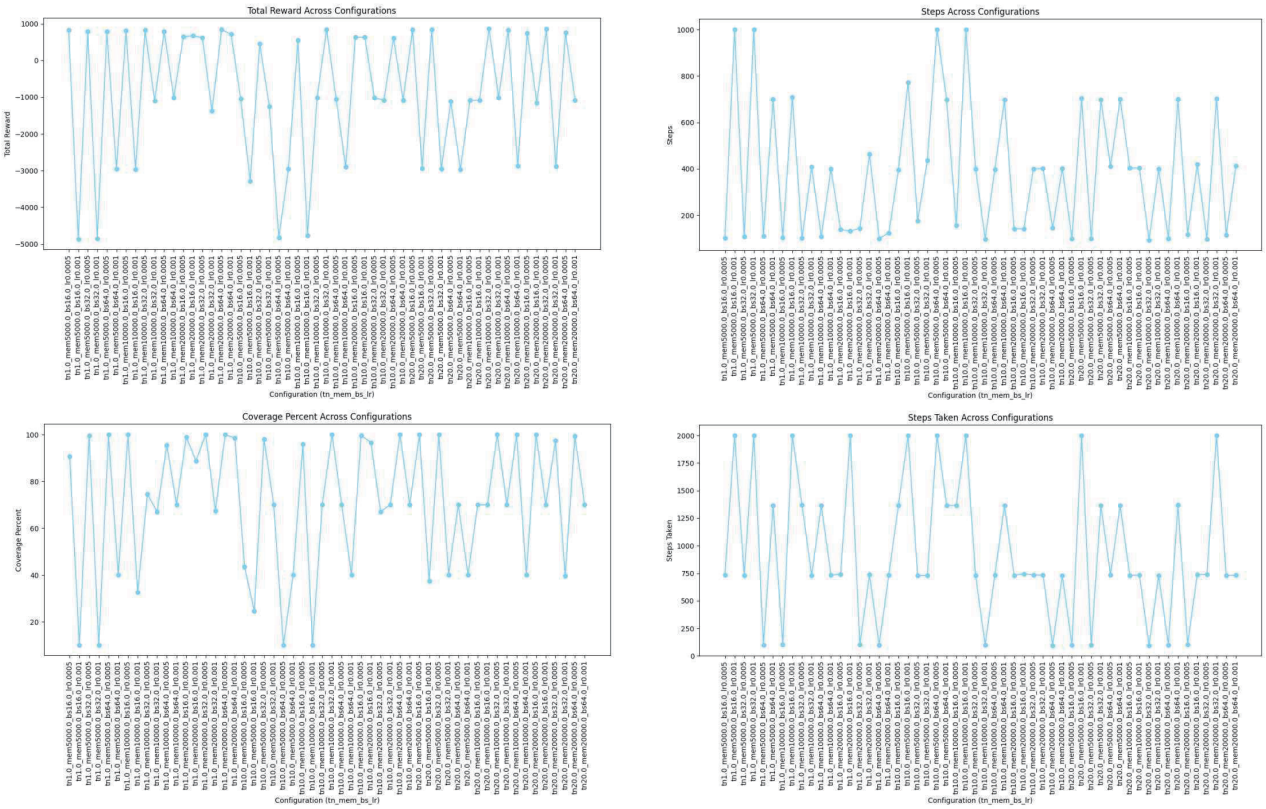

그림 2 실험 결과

위와 같은 기법을 사용하여 DQN은 다양한 분야에 적용되며 많은 연구가 진행되어왔다.

Ⅳ. 실험 및 실험 결과

본 연구의 실험은 2차원 격자 기반의 시뮬레이션 환경에서 수행하였다. 실험에 사용된 맵은 10 x 10 크기의 정사각형 격자로 구성되었으며, 각 셀은 이동 가능한 자유 공간 혹은 장애물로 설정되어 있다. 맵 정보는 그림 1과 같으며, 해당 맵에서 경로를 계획하도록 설계하였다. 단일 에이전트로 진행하였으며 이동은 상, 하, 좌, 우의 네 방향이 가능하며, 각 방향으로 한 칸씩 이동할 수 있다. 로봇 에이전트는 왼쪽 위 모서리 지점에서 시작하여 탐색 가능한 모든 셀을 한 번씩 방문하는 커버리지 경로를 생성하는 것을 목표로 한다.

하이퍼파라미터 튜닝을 위한 실험에서는 다음 네 가지 주요 파라미터를 변수로 설정하였다.

- Target Update : Target Network의 파라미터 업데이트 주기
- Replay Memory 크기 : 샘플을 저장하는 메모리 크기
- Learning Rate : 학습률
- Batch Size : 학습할 때의 배치 사이즈

보상 함수는 (2) 식으로 설정하였다.

$$reward = \begin{cases} 10 & \text{if } (x,y) = \text{탐색하지 않은 셀} \\ -5 & \text{if } (x,y) = \text{장애물} \\ -5 & \text{if } (x,y) = \text{탐색한 셀} \end{cases} \quad (2)$$

모든 실험은 동일한 맵 조건에서 독립적으로 3회 반복하였으며 각 설정에 대해 누적 보상, 최종 에피소드의 스텝 수, 테스트 시 커버리지 비율과 스텝 수로 측정하여 비교 분석하였다.

실험 결과는 그림 2에 정리하였다. 각 기준에 대해 학습을 3회 진행하였으며, 3회에 대한 결과를 평균으로 계산하였다. 실험 결과 특정 하이퍼파라미터에서 성능이 좋고 안정적인 모습을 확인하였다. 타겟 네트워크 업데이트 주기는 20, 리플레이 메모리 크기는 10000, 배치 사이즈 크기는 32, 학습률은 0.0005일 때 커버리지 비율은 100%이며 중복 탐색률도 가장 적은 것을 확인하였다. 학습 안전성이 높은 기준을 도출했을 때, 학습에 영향을 미치는 하이퍼파라미터는 타겟 네트워크 업데이트 주기, 학습률, 배치 사이즈였다. 3회의 학습 동안 학습 안전성이 높은 기준은 각각 20, 32, 0.0005였다. 이러한 결과를 통해 DQN 기반 커버리지 경로 계획 성능은 하이퍼파라미터 설정에 따라 크게 달라질 수 있음을 확인하였다.

Ⅴ. 결론

본 논문에서는 DQN 기반 Coverage Path Planning의 성능에 영향을 미치는 주요 하이퍼파라미터에 대한 실험적 분석을 수행하였다. 2차원 격자 기반 시뮬레이션 환경에서 여러 기준을 변화시키며 학습 및 평가를 진행한 결과, 각 파라미터 조합이 커버리지 비율과 학습 성능에 유의미한 차이를 발생시킴을 확인하였다.

다만 본 연구는 고정된 10x10 정사각형 맵 환경과 제한된 실험 반복 횟수에 기반하고 있어, 일반화된 결론을 도출하기에는 한계가 존재한다. 따라서 향후 연구에서는 실험 반복 횟수를 확대하고, 다양한 크기와 구조를 갖는 맵 환경에서의 실험, 추가적인 하이퍼파라미터 및 복합적 파라미터 조합 탐색을 진행하고자 한다.

참 고 문 헌

[1] FAN, Jianqing, et al. A theoretical analysis of deep Q-learning. In: Learning for dynamics and control. PMLR, 2020. p. 486-489.
[2] GALCERAN, Enric; CARRERAS, Marc. A survey on coverage path planning for robotics. Robotics and Autonomous systems, 2013, 61.12: 1258-1276.