모션계획을 위한 강화학습 기반 트리 편향 확장 기술

Bias tree expansion using reinforcement learning for efficient motion planning

윤민성[1] · 박대형[2] · 윤성의[†]

Minsung Yoon[1], Daehyung Park[2], Sung-Eui Yoon[†]

Abstract: Motion Planning is a computational problem to find a valid and optimal path from the given start to the goal configuration. During the last few decades, sampling-based motion planning methods, such as Rapidly-exploring Random Tree* (RRT*), have been shown to work well even in a high-dimensional and continuous state space. Recently, with the advances of deep learning, sample efficiency of sampling-based motion planning has been improved by learning the bias (heuristic) to the near-optimal region considering the surrounding obstacles and goal position. However, the performance of a neural network trained using supervised learning is highly dependent on a set of demos previously collected for training. Therefore, this leads to problems such as distribution mismatch and performance bounding and overfitting to the demos. In this regard, we propose RL-RRT* to train the network using reinforcement learning and use it as a bias network. We validate our method in a 2-D environment showing improved anytime performance, including initial solution quality and time and reasonably fast cost convergence rate.

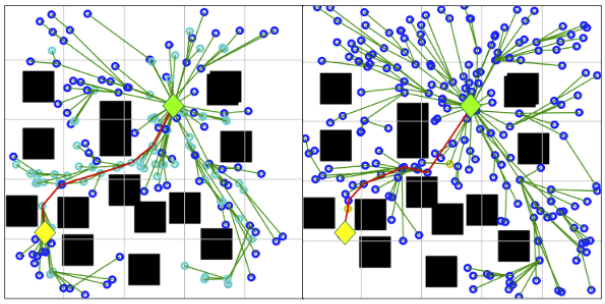
Keywords: Motion and Path Planning, Reinforcement learning.

1. 서 론

모션 계획(Motion Planning)은 로봇 공학과 로봇 지능에 있어서 핵심적인 연구 분야이다. 모션 계획 알고리즘의 목표는 시작 지점과 목표 지점을 연결하며, 충돌이 없고 비용(e.g., path length)이 적은 경로를 찾는 것이다.

샘플링-기반에 모션 계획 방식은 configuration space (C-space) 상에서 샘플링을 통해 트리 혹은 그래프 구조를 구축하여 충돌이 없는 자유 공간을 추정하고, 이를 기반으로 A*와 같은 그래프 서치 알고리즘을 사용하여 그래프상에 최적에 경로를 도출한다 [1]. 하지만, 기존 샘플링-기반 방식은 트리 구조를 확장하는 데 있어 오로지 랜덤 샘플링을 기반으로 동작을 하기에 목표 지점과는 상관없이 모든 방향으로 트리 구조의 확장이 이루어지고, 이에 따라 문제 해결에 있어 샘플의 효율성이 적었다.

최근에는 딥러닝의 발전에 따라 사전에 시간을 충분히 주어 수집한 optimal paths을 기반으로 지도학습(supervised learning)을 수행하고, 트리구조의 확장에 주변 환경과 목표 지점을 고려한 near-optimal 영역으로의 편향(bias)을 주는 연구가 진행되고 있다 [2]. 하지만 근본적으로 지도학습으로 학습된 네트워크의 성능은 사전에 준비된 데이터셋(demonstration set)에 상한선이 존재하고, 그에 쉽게 오버 피팅(overfitting)이 되는 경향이 있다. 또한 데이터셋에 존재하지 않는 상황에 있어서는 유의미한 행동을 학습할 수 없다는 단점이 존재한다.



[Fig. 1] The tree structure expanded in 1 second. (green diamond: start position, yellow diamond: goal position, blue circle: vertex generated with random sampling, cyan circle: vertex generated using bias network, green line: edge, red line: found path)

※ This project was funded by National Research Foundation (NRF) grant funded by Korea government (MSIT). No. 2019R1A2C3002833
1. MS candidate, Korea Advanced Institute of Science and Technology (KAIST), School of Computing, Daejeon, Korea (minsung.yoon2@gmail.com)
2. Assistant Professor, KAIST, School of Computing (daehyung@kaist.ac.kr)
†. Professor, Corresponding author: KAIST, School of Computing (sungeui@kaist.edu)

본 논문에서는 사전에 준비된 데이터 셋에 영향을 받지 않고, 학습 agent가 수많은 행동의 시행착오를 통해 최적에 행동을 학습하게 되는 강화학습을 적용하였다. 이를 통해 샘플링-기반 방식에서는 anytime performance가 향상되고, 강화학습에 측면에서는 safety와 completeness가 만족되는 것을 실험적으로 보인다.

2. 관련 연구

샘플링-기반 모션 계획에 딥러닝을 활용하여 트리 구조의 편향 확장에 대한 연구는 MPNET (motion planning network)에서부터 시작되었다 [2]. 해당 논문은 지도학습 기반으로 환경 맵, 목표 지점과 현재 위치를 입력으로 받고 어느 위치로 확장이 이루어져야 하는지를 내보낸다.

3. RL-RRT*: A NEURAL MOTION PLANNING

본 연구팀이 제안하는 네트워크의 구조는 [Fig. 2]와 같다. 특정 문제에 global 정보인 맵 이미지에 추가적으로 local lidar scan을 시뮬레이션하여 국소적인 정보를 줌으로써, agent의 학습 효율성 및 편향 성능이 증가되었다.

문제 정의를 위한 Markov Decision Process (MDP)는 아래와 같이 구성하여 학습을 진행하였고, 강화학습 알고리즘에 있어서는 soft actor-critic (SAC)이 사용되었다.

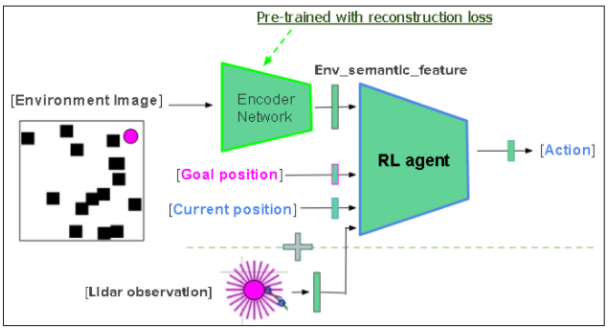
- Observation: $[\mathrm{Enc(map)}|\mathrm{Pos_{goal}}|\mathrm{Pos_{cur}}|\text{Lidar scan}]$
- Action: $[-1, 1] \in \mathrm{R^{dimension\ of\ problem}}$
- Reward: $+5$ if reach a goal, -0.5 if collision occurs

최종적으로 학습된 agent를 샘플링-기반 트리 확장에 사용하게 된다. 한 가지 주목할 점은 네트워크는 다양한 환경, 환경상에 가능한 모든 목표 위치와 현재 위치에 대한 최적에 확장 방향을 학습해야 한다. 하지만, 뉴럴 네트워크 함수에 복잡도에는 한계가 존재하고, 잘못된 방향으로 편향을 줄 가능성이 있다. 따라서 실질적 사용에는 optimality와 completeness를 보장하기 위해 편향 네트워크 사용에 확장과 랜덤 샘플링을 50 %의 비율로 사용하였다.

4. 실험 구성 및 결과

RL-RRT*의 성능평가를 위해 RRT* [1]와 MPNET [2]을 비교군으로 선정하였고, 랜덤으로 16개의 박스를 배치하여 400개의 2-D 맵을 구축하였다. 이 중 350개는 학습용으로, 나머지 50개는 테스트용으로 구성하였다.

MPNET은 지도학습 기반으로 5일에 걸쳐 24.7 M의 데이터를 수집하였고, 이를 기반으로 학습에 7시간이 소요되었다. RL-RRT*는 강화학습 기반으로, 학습에 1 M번의 행동을 취하였고 총 7.5시간이 걸렸다.



[Fig. 2] Architecture of RL agent
(input: map, current & goal position, and lidar scan, output: action)

50개의 테스트 맵 각각에서 무작위로 100개의 시작과 목표 지점을 준비하였다(총 5,000개의 문제). 각 문제에 해결 시간은 3초가 주어졌고, 이에 대한 정량적 평가는 [Table 1], 정성적 평가는 [Fig. 1]에서 확인할 수 있다.

실험 결과, 두 기존 방법들에 비해 5000개의 모든 문제를 해결하는데 걸리는 시간과 찾은 솔루션의 길이 측면에서 성능 향상을 보였다. 한 가지 흥미로운 점으로는 오로지 RL만을 사용하여 문제를 해결하였을 경우에는 최종 성공률이 86.5 %로 이외에 문제들에 있어서는 충돌이 나거나 문제를 해결하지 못하였다.

[Table 1] Experiment results (Only RL's success rate: 86.5 %)

	RRT* [1]	MPNET [2]	RL-RRT* (Ours)
Time to 100 % success rate (unit: sec)	1.6	2.75	**1.4**
Initial solution path quality (unit: m)	8.469	8.311	**8.213**

5. 결 론

본 논문은 강화학습 기법을 사용하여 뉴럴 네트워크를 학습하고, 이를 샘플링-기반 모션 계획에 편향 네트워크로써 사용하는 방식을 제시하였다. 이에 따른 샘플에 효율성 향상을 여러 메트릭을 통해 보였다. 결론적으로, 오직 강화학습만을 사용하거나 Classical 한 방식만을 사용할 것이 아닌 여러 방식을 적절히 융합함으로써 효율적인 방식을 도출하는 것이 가능할 수 있음을 시사한다.

6. 참고 문헌

[1] Karaman, Sertac, and Emilio Frazzoli. "Sampling-based algorithms for optimal motion planning." *The international journal of robotics research*, vol.30.7, pp.846-894, June.2011.

[2] A. H. Qureshi, Y. Miao, A. Simeonov and M. C. Yip. "Motion planning networks: Bridging the gap between learning-based and classical motion planners." *IEEE Transactions on Robotics*, vol. 37, no. 1, pp. 48-66, Feb.2021