실내 자율주행 자동차를 위한 심층 강화학습 알고리즘 비교 및 분석

김우현, 왕수현, 전두선, 엄두섭

고려대학교

koniwin06@korea.ac.kr, 08shwang@korea.ac.kr, jeonds1127@korea.ac.kr, eomds@korea.ac.kr

Comparison and Analysis of Deep Reinforcement Learning Algorithms for Indoor Autonomous Vehicles

Woo-Hyon Kim, Soo-Hyun Wang, Du-Seon Jeon, Doo-Seop Eom

Korea Univ.

요 약

본 논문에서는 실내 환경에서 영상정보를 이용한 End-to-End 방식의 자율주행을 위하여 심층 강화학습(Deep Reinforcement Learning) 기반의 알고리즘을 학습하고, 각 알고리즘 간의 성능을 비교 및 분석해본다. Deep-SARSA, Deep Q Network(DQN), Double DQN(DDQN)의 3가지 심층 강화학습 알고리즘을 사용하여 학습하였다. 학습 모델은 전방 카메라로부터 받아오는 깊이(depth) 이미지와 현재 위치에서 목적지까지의 방향각을 입력으로 받아서 학습하였다. 시뮬레이션을 통하여 학습 완료된 모델을 적용한 자동차는 안정적인 주행을 하며 목적지에 도착했음을 볼 수 있다.

Ⅰ. 서 론

최근 들어 공항 안내 로봇, 물류센터의 자동화 로봇, 가정의 로봇 청소기 등 실내 자율주행에 대한 수요가 늘어남에 따라 로봇 스스로가 주변 상황을 인지하여 자율적으로 길을 찾아 주행하는 연구가 활발하게 진행되고 있다. 기존 자율주행 연구에서 사용되던 상황인지, 판단, 제어 알고리즘은 인공지능의 발달로 신경망으로 대체되며 End-to-End 제어의 실현 가능성을 보인다[1].

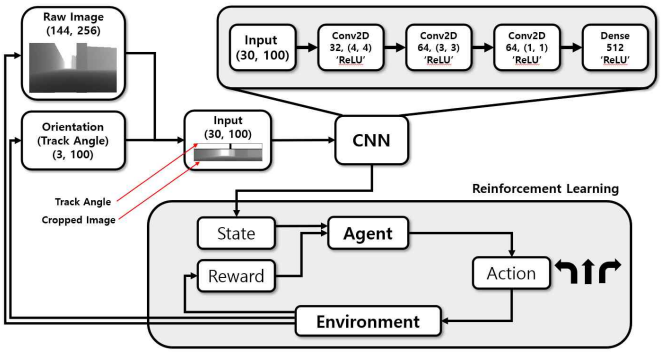
심층 강화학습은 알파고 등을 통하여 엄청난 잠재력이 입증됨에 따라 현재 많은 분야에 적용이 확대되고 있으며, 자율주행도 그중 하나다. 카메라 이미지를 입력으로 하는 심층 강화학습 기반의 자율주행 연구가 꾸준히 이루어지고 있으며, 많은 센서를 이용해서 환경에 대한 정보를 얻지 않아도 주행 가능하다는 이점이 있다.

본 논문에서는 깊이 이미지와 방향각만을 이용하여 목적지까지 최적의 경로를 탐색하는 실내 자율주행 자동차를 제안한다. 자동차는 심층 강화학습 기반의 알고리즘을 이용해 학습하고, Deep-SARSA, DQN, DDQN의 3가지 알고리즘에 대해 학습을 진행하고 결과를 비교, 분석한다.

Ⅱ. 본론

2.1 실내 자율주행을 위한 심층 강화학습 알고리즘

본 논문에서 제안하는 자율주행 방법은 [그림 1]과 같다. 전방 카메라를 통해 받아오는 깊이 이미지와 현재 위치에서 목적지까지의 방향각만을 입력으로 한다[2]. 깊이 이미지와 방향각은 전처리를 거쳐서 합성 곱 신경망(CNN)으로 들어가게 되는데 전처리 방법은 다음과 같다. [그림 1]과 같이 세로, 가로 144x256 화소의 원시 깊이 이미지 가운데를 기준으로 자른 20x100 화소의 이미지와 현재 위치에서 목적지까지의 방향값인 트랙 각을 표시한 10x100 화소 이미지를 합친 30x100 화소 이미지를 입력으로 사용한다. 원시 깊이 이미지를 입력으로 그대로 사용하지 않고, 자른 이미지를 사용한 이유는 자동차 주행에는 불필요한 천장과 같은 정보를 줄이고, 트랙 각을 이미지에 같이 표시함으로써 CNN에서 연산량을 줄일 수 있다



[그림 1] 깊이 이미지와 방향각을 입력으로 하는 강화학습 모델

는 이점이 있기 때문이다.

입력 이미지의 특징을 추출한 CNN의 결과는 강화학습의 상태로 들어가게 된다. 에이전트는 상태와 보상을 받아서 그에 알맞은 행동을 취하면서 학습을 한다. 행동 공간은 불연속적인 3가지 행동으로 1.5m 직진 후 정지, 30도 좌회전 후 정지, 30도 우회전 후 정지이다. 방향전환은 복도의 너비와 자동차의 회전 반경을 고려하였다. 자동차가 좁은 폭의 복도에서 한 번에 90도를 회전할 경우, 다시 이전의 방향으로 복귀하기에 어려움이 있다. 따라서 회전을 3번에 나누어서 90도를 회전하게 하여 목적지를 향한 방향으로의 복귀도 가능하게 했다.

보상은 환경과 충돌할 경우 -100을 주며, 한 에피소드가 끝나고 다음 에피소드를 시작한다. 무제한적인 탐색을 막고, 추후에 실제 환경에서 진행할 자동차의 배터리 자원을 고려하여 충돌하지 않더라도 보상이 -100에 도달하면 에피소드가 끝나게 된다. 자동차가 목적지 1m 이내에 위치하면 도착으로 간주하고 100의 보상을 주며 에피소드가 끝난다. 목적지까지 이동하는 동안에는 보상 감소로 -2를 추가하여 매 step에 대한 보상은 "-2 + (이전 위치에서 목적지까지의 거리 - 현재 위치에서 목적지까지의 거리)" 이다. 이 보상 식을 통해 목적지 방향으로 이동하면 양수 보상을 받고, 잘못된 방향으로 이동하면 음수 보상을 받게 하였다. 학습을 진행하

는 총 step 수는 200,000번이며, 학습률은 0.00025이다.

환경과 입력이 동일한 상황에서 심층 강화학습 기반의 알고리즘 Deep-SARSA, DQN, DDQN의 3가지에 대해 각각 실험을 진행하였다.

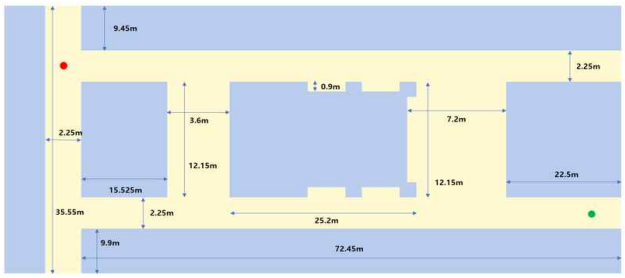
Deep-SARSA 모델은 학습 정책과 행동 정책이 같아야만 학습이 가능한 on-policy 알고리즘이며, Q-러닝과 다르게 상태와 행동에 대한 최대값을 구해서 Q값을 업데이트 하지 않고, 다음 행동을 확률적으로 선택하여 Q값을 업데이트 한다.

반면에 DQN[3]은 학습 정책과 행동 정책이 달라도 학습이 가능한 off-policy 알고리즘이며, SARSA와 다르게 정책이 업데이트 되더라도 이전의 경험들을 학습에 사용할 수 있다는 것이 특징이다.

DDQN[6] 모델은 DQN 모델에서 신경망의 사용으로 발생하는 과적합에 의한 학습 저하 문제를 개선한 알고리즘이다. DQN 모델을 2개 사용해서 하나는 계속 업데이트가 되도록 하고, 다른 하나는 가끔씩만 새로 학습된 네트워크로 업데이트한다. 학습과 최적화를 따로 진행하기 때문에 다양한 시도를 하면서도 학습 저하 문제를 해결할 수 있다.

2.2 실험 환경

강화학습을 진행하기 위해서 Epic Games사의 Unreal Engine과 Microsoft사의 AirSim을 이용해 실제 환경과 동일한 크기의 환경을 시뮬레이션에 모델링하였다[3]. 실제 환경은 고려대학교 신공학관 5층 복도이며 환경에 대한 정보는 [그림 2]와 같다. 전체 크기는 가로 74.7m, 세로 35.55m로 가로로 긴 복도 2개가 평행하게 있고, 세로로 된 복도 3개가 두 가로 복도를 잇는다. 학습 에이전트의 시작점은 초록점이고, 목적지는 빨간점이다.



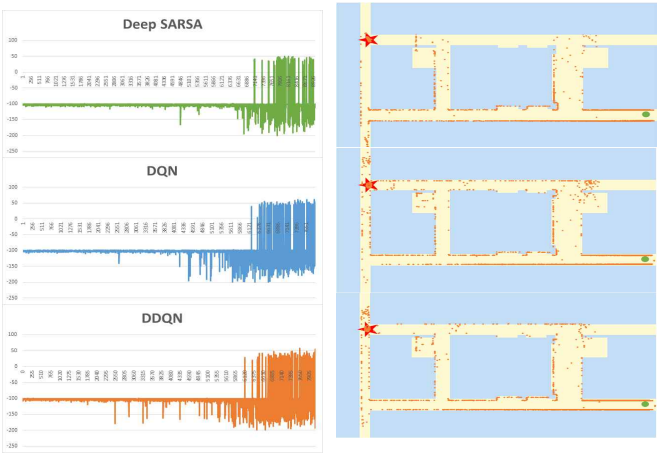
〔그림 2〕 a. 강화학습을 적용한 실제 환경 (고려대학교 신공학관 5층 복도) 시작점(초록점), 목적지(빨간점)

2.3 실험 결과 및 분석

각 알고리즘을 학습한 결과는 [표 1]과 같다. 3가지 알고리즘 중 DDQN이 DQN, Deep-SARSA에 비해 각각 9688 steps, 14785 steps 더 빠른 학습 속도와 25회, 40회 높은 성공 횟수를 보인다. [그림 3]의 b를 보면 Deep-SARSA에 비해 off-policy 정책을 가지는 DQN과 DDQN의 에이전트 최종 위치가 더 넓게 퍼져있다. off-policy 알고리즘은 이전의 경험들을 학습에 사용할 수 있고, 하나의 정책을 따르면서 여러 개의 정책을 학습할 수 있기 때문에 탐색을 계속하면서도 최적의 정책을 학습할 수 있다[4]. 따라서 Q 값을 업데이트할 때, 확률적인 방법보다 max를 이용한

〔표 1〕 각 알고리즘 학습 결과

알고리즘	첫 성공 step	성공 횟수	성공한 에피소드의 평균 보상값	성공한 에피소드의 평균 step
Deep-SARSA	88491	84	31.54	181.69
DQN	83394	99	39.40	181.58
DDQN	73706	124	45.77	175.86



〔그림 3〕 시뮬레이션을 이용한 각 알고리즘의 학습 결과 a. 에피소드에 따른 보상값(좌), b. 에피소드에 따른 에이전트의 최종 위치(우)

greedy 한 방법이 더 좋은 성능을 보임을 알 수 있다. [그림 3]의 그래프를 통해 off-policy 정책을 가지는 DDQN이 다른 모델보다 더 많은 지역을 탐색하고, 목적지에 빠르게 도달했음을 알 수 있다. 그러면서도 성공한 에피소드의 평균 보상값이 높고, 평균 스텝 수가 낮다. 많은 탐색을 통해 얻은 환경에 대한 정보를 이용해 최적의 정책을 찾는다고 볼 수 있다.

Ⅲ. 결론 및 향후 연구과제

본 논문에서는 환경에 대한 정보가 없는 상황에서 깊이 이미지와 방향각만을 이용하여 목적지까지 최적의 경로를 탐색하는 실내 자율주행 자동차를 제안하고, Deep-SARSA, DQN, DDQN의 알고리즘 각각에 대한 결과를 비교, 분석하였다. 3가지 알고리즘 중 DDQN이 가장 빠른 학습 속도와 많은 성공 횟수로 우수한 성능을 보이고, 더 많은 지역을 탐색하며 최적의 경로를 찾았음을 알 수 있다. 동일한 환경에서 장애물이 추가되거나 다른 목적지를 설정하여 주행하더라도 높은 성공률을 보일 것으로 기대된다. 향후 계획으로는 더 복잡한 상태 공간과 연속적인 행동 공간에 적합한 알고리즘인 결정론적 정책경사 (DDPG) 모델과 장-단기 기억을 고려한 LSTM 이용하여 학습해보고 결과를 비교해보고자 한다. 또한, 모형 자동차를 이용하여 학습이 완료된 모델을 실제 환경에서 테스트할 계획이다.

참 고 문 헌

[1] S. Levine, C. Finn, T. Darrell, and P. Abbeel. End-to-end trainingof deep visuomotor policies.Journal of Machine Learning Research,17(39):1 - 40, 2016.

[2] Kersandt, Kjell. "Deep reinforcement learning as control method for autonomous UAVs" MS thesis. Universitat Politècnica de Catalunya, 2018.

[3] S.Shah, D. Dey, C. Lovett, and A. Kapoor "AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles", Field and Service Robotics conference 2017 (FSR 2017), 2017

[4] Demin, Vladimir. "Cliff walking problem."

[5] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." arXiv preprint arXiv:1312.5602 (2013).

[6] Van Hasselt, Hado, Arthur Guez, and David Silver. "Deep reinforcement learning with double q-learning." Thirtieth AAAI conference on artificial intelligence. 2016.