오프라인 강화학습에서 데이터 품질에 따른 신경망 모델 크기 선택에 관한 연구

유승찬, 이정우
서울대학교

scyu@cml.snu.ac.kr, jungle@snu.ac.kr

A Study on Neural Network Model Size Selection Based on Data Quality in Offline Reinforcement Learning

Seung Chan Yu, Jung Woo Lee
Seoul National Univ.

요 약

본 논문에서는 오프라인 강화학습에서 데이터 품질에 따라 요구되는 신경망 모델 크기의 차이를 실험적으로 분석한다. TD3+BC 알고리즘을 기반으로 D4RL 벤치마크 환경에서 모델 크기를 단계적으로 축소한 결과, expert 데이터는 medium 데이터에 비해 모델 크기 감소에 더 민감한 성능 저하를 보였다. 이는 고품질 데이터가 더 높은 정책 표현 복잡도를 요구함을 시사한다.

Ⅰ. 서 론

오프라인 강화학습은 사전에 수집된 데이터만을 이용해 정책을 학습함으로써, 온라인 상호작용이 어려운 환경에서도 강화학습 적용을 가능하게 한다. 오프라인 데이터는 수집 정책의 숙련도에 따라 품질이 달라지며, 이는 학습 성능에 직접적인 영향을 미친다. 기존 연구들은 주로 데이터 품질에 따른 알고리즘 성능 비교에 초점을 두어 왔으나, 데이터 품질에 따라 요구되는 신경망 모델 크기가 어떻게 달라지는지에 대한 분석은 충분히 이루어지지 않았다. 본 논문에서는 오프라인 데이터 품질에 따른 신경망 모델 크기 감소 시 성능 민감도를 분석하여, 오프라인 강화학습에서의 모델 크기 선택에 대한 시사점을 제시한다.

Ⅱ. 본론

실험 방법

본 논문에서는 오프라인 강화학습에서 오프라인 데이터의 품질에 따라 요구되는 신경망 모델 크기가 어떻게 달라지는지를 분석한다. 이를 위해 동일한 강화학습 알고리즘과 학습 설정을 유지한 상태에서, 오프라인 데이터 품질과 신경망 모델 크기만을 변화시키는 방식으로 실험을 설계하였다.

모든 실험에는 오프라인 강화학습 알고리즘으로 TD3+BC[1]를 사용하였다. TD3+BC 는 TD3 구조에 행동 모방 항을 결합한 알고리즘으로, 오프라인 데이터 분포로부터 벗어난 행동을 억제하여 안정적인 학습이 가능하다. 본 논문에서는 알고리즘에 따른 영향을 배제하고 데이터 품질과 신경망 크기의 관계만을 분석하기 위해, TD3+BC 를 모든 실험에 공통적으로 적용하였다.

오프라인 데이터 품질은 D4RL[2] 벤치마크의 expert 및 medium 데이터셋을 사용하여 조절하였다. expert 데이터셋은 전문가 정책으로 수집된 고품질 데이터를 포함하며, medium 데이터셋은 Soft Actor-Critic(SAC)[2] 기반 정책을 온라인으로 학습한 후 조기에 학습을 중단하여 수집된 부분 학습 정책 데이터로 구성되어 있다.

신경망 모델 크기의 영향만을 비교하기 위해, 정책 신경망은 동일한 다층 퍼셉트론 구조를 유지하되 각 hidden layer 의 차원만을 변화시켰다. hidden dimension 을 128, 64, 32, 16, 8 로 설정한 여러 모델을 구성하였으며, layer 수와 기타 학습 설정은 모든 실험에서 동일하게 유지하였다. 이를 통해 성능 변화가 순수한 모델 용량 차이에서 기인함을 보장하였다.

각 환경(HalfCheetah, Hopper, Walker2D)과 데이터셋 조합에 대해 5 개의 random seed(0~4)를 사용하여 반복 실험을 수행하였으며, 최종 성능은 누적 보상의 평균을 기준으로 평가하였다. 이러한 실험 설계를 통해 오프라인 데이터 품질에 따라 요구되는 신경망 모델 용량의 차이를 정량적으로 분석하였다.

실험 결과

그림 1-3 은 HalfCheetah, Hopper, Walker2D 환경에서 오프라인 데이터 품질(expert, medium)에 따라 신경망 모델 크기(hidden dimension)를 단계적으로 감소시켰을 때의 성능 변화를 normalized score 기준으로 나타낸 결과이다. hidden dimension 128 은 expert 및 medium 데이터셋 모두에서 충분한 성능을 안정적으로 달성할 수 있는 기준 모델로 설정하였으며, 이후 모델 크기를 축소하면서 데이터 품질에 따른 성능 저하 양상의 차이를 분석하였다.

expert 데이터셋의 경우, 세 환경 모두에서 모델 크기 감소에 따라 성능 저하가 비교적 빠르게 발생하였다. HalfCheetah 환경에서는 hidden dimension 이 64 에서 32 로 감소하는 구간부터 성능이 급격히 하락하였으며, Hopper 및 Walker2D 환경에서도 32 또는 16 이하의 모델에서 뚜렷한 성능 저하가 관찰되었다. 이는 전문가

정책으로부터 수집된 고품질 데이터가 보다 정교하고 복잡한 상태-행동 매핑을 포함하고 있어, 이를 안정적으로 근사하기 위해서는 충분한 모델 용량이 필요함을 시사한다.

반면 medium 데이터셋의 경우, 모델 크기를 감소시키더라도 성능 저하가 상대적으로 완만하게 나타났다. 세 환경 모두에서 hidden dimension 을 64 또는 32 까지 축소하더라도 normalized score 의 감소 폭은 제한적이었으며, 성능 저하는 주로 hidden dimension 이 16 이하로 감소하는 구간에서 관찰되었다. 이는 medium 데이터가 상대적으로 단순한 행동 분포를 포함하고 있어, 작은 신경망 모델로도 일정 수준의 성능을 유지할 수 있음을 의미한다.

이러한 결과는 고품질 데이터일수록 모델 용량 감소에 더 민감한 성능 저하를 보이며, 이는 전문가 수준의 정책이 갖는 높은 표현 복잡도가 충분한 신경망 용량을 요구하기 때문임을 보여준다. 즉, 데이터 품질이 높을수록 작은 모델에서도 학습이 쉬워지는 것이 아니라, 정교한 행동 구조를 유지하기 위해 더 큰 모델이 필요함을 실험적으로 확인하였다.

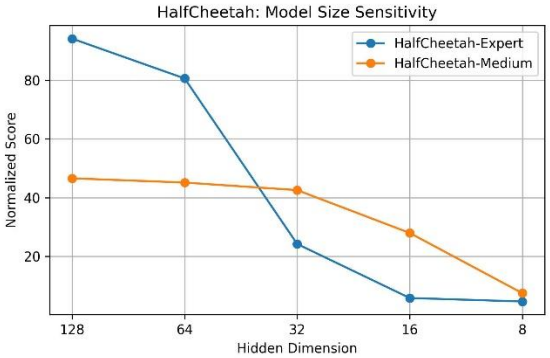

그림 1. HalfCheetah 환경에서 신경망 크기 감소에 따른 성능 변화

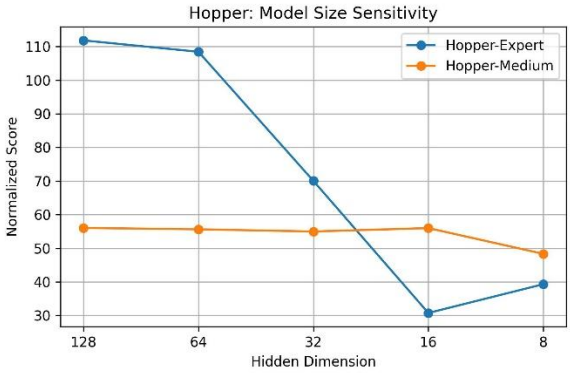

그림 2. Hopper 환경에서 신경망 크기 감소에 따른 성능 변화

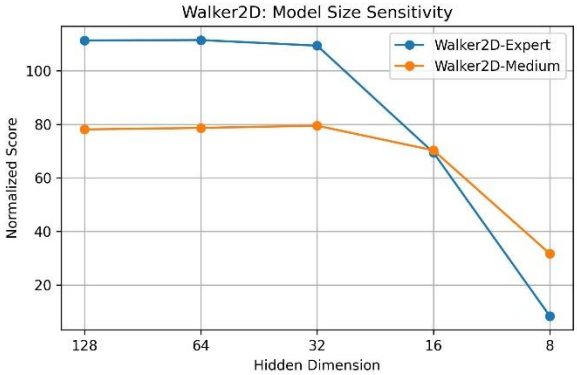

그림 3. Walker2D 환경에서 신경망 크기 감소에 따른 성능 변화

Ⅲ. 결론

본 논문에서는 오프라인 강화학습 환경에서 오프라인 데이터 품질에 따라 요구되는 신경망 모델 크기가 어떻게 달라지는지를 실험적으로 분석하였다. TD3+BC 알고리즘을 기반으로 HalfCheetah, Hopper, Walker2D 환경에서 D4RL 의 expert 및 medium 데이터셋을 사용하여, 신경망 모델 크기를 단계적으로 축소하는 실험을 수행하였다.

실험 결과, 모델 크기 감소 시 expert 데이터셋은 medium 데이터셋에 비해 성능 저하가 더 빠르게 발생하였으며, 이는 전문가 정책으로부터 수집된 고품질 데이터가 보다 정교하고 복잡한 상태-행동 매핑을 포함하고 있음을 시사한다. 반면, medium 데이터셋은 상대적으로 단순한 행동 분포를 포함하고 있어 작은 신경망 모델에서도 일정 수준의 성능을 유지할 수 있음을 확인하였다.

이러한 결과는 고품질 데이터일수록 정책의 표현 복잡도가 증가하며, 이를 안정적으로 근사하기 위해서는 충분한 신경망 용량이 필요함을 의미한다. 따라서 모델 크기 선택 시 데이터 품질을 고려하지 않은 일률적인 신경망 구조 적용은 비효율적일 수 있다. 본 연구는 오프라인 강화학습에서 데이터 품질에 따른 신경망 모델 크기 선택의 중요성을 실험적으로 규명하였으며, 향후 다양한 알고리즘과 데이터 품질 수준에 대한 분석을 통해 보다 일반적인 모델 설계 가이드라인으로 확장될 수 있을 것으로 기대된다

ACKNOWLEDGMENT

This work is in part supported by National Research Foundation of Korea (NRF, RS-2024-00451435(20%), RS-2024-00413957(20%), RS-2025-22442984(20%)), Institute of Information & communications Technology Planning & Evaluation (IITP, RS-2025-02305453(10%), RS-2025-02273157(10%), RS-2025-25442149(10%) RS-2021-II211343(10%)) grant funded by the Ministry of Science and ICT (MSIT), Institute of New Media and Communications(INMAC), and the BK21 FOUR program of the Education, Artificial Intelligence Graduate School Program (Seoul National University), and Research Program for Future ICT Pioneers, Seoul National University in 2026.

참 고 문 헌

[1] Fujimoto, S. *et al.* “A Minimalist Approach to Offline Reinforcement Learning.” Advances in Neural Information Processing Systems (NeurIPS), 2021.

[2] Fu, J. *et al.* “D4RL: Datasets for Deep Data-Driven Reinforcement Learning.” Advances in Neural Information Processing Systems (NeurIPS), 2020.

[3] Haarnoja, T. *et al.* “Soft Actor-Critic.” International Conference on Machine Learning (ICML), 2018.