그래프 신경망 기반 강화학습을 통한 양자 배터리 초흡수 이득 활용

박범도 [1], 장현석 [1], 박준성 [1], 정훈 [2], 허태욱 [2], *이상금 [1]
*국립한밭대학교 [1], 한국전자통신연구원 [2]

{pbeomdo, seokchu123, js03093351}@gmail.com, {leech, htw398}@etri.re.kr,
sangkeum@hanbat.ac.kr

Leveraging Superabsorption Gain in Quantum Batteries via Graph Neural Network-Based Reinforcement Learning

Beomdo Park[1], Hyeonseok Jang[1], Junseong Park[1], Hoon Jeong[2], Taewook Heo[2],
*Sangkeum Lee[1]
*Hanbat National University[1], Electronics and Telecommunications Research Institute[2]

요 약

다입자 양자 배터리의 최적 제어는 큐비트 수 N 에 따라 상태 공간이 지수적으로 확장되는 문제로 인해 강화학습 적용에 본질적인 확장성 한계를 가진다. 본 연구는 Tavis-Cummings 모델에서는 큐비트의 순서가 결과에 영향을 주지 않는다는 물리적 특성에 착안하여, 출력이 입력 순서에 무관(Permutation Invariant)한 그래프 신경망 정책을 강화학습에 도입함으로써 초흡수(superabsorption) 이득을 안정적으로 활용하는 방안을 제시한다. Graph Convolutional Network(GCN), Graph Attention Network(GAT), MLP(Set)을 순서 무관 정책으로 활용하고, 순서 정보를 그대로 사용하는 Multi-Layer Perceptron(MLP)을 비교군으로 설정하여 N=2~8 구간에서 PPO(Proximal Policy Optimization) 학습을 수행한다. 순서 무관 정책은 N=8 에서 99% 이상의 평균 최종 에너지를 달성한 것에 반해, MLP 는 동일 조건에서 73% 수준에 그쳤다. 특히 GCN 은 GAT, MLP(Set)와 동등한 충전 효율을 달성하면서도 파라미터 수와 학습 시간을 최소화하여, 대규모 양자 배터리 환경으로의 확장 가능성을 확인하였다.

Ⅰ. 서 론

양자 배터리는 양자 시스템의 집단 상호작용을 제어하여 기존 배터리를 능가하는 성능을 구현하려는 기술로써 주목받고 있다[1]. 특히 다수의 큐비트가 집단적으로 상호작용할 때 충전 효율이 N^2 에 비례하는 초흡수 현상이 나타나며, 이론적 충전 시간을 $O(1/N)$ 수준으로 단축할 수 있는 잠재력을 가진다. 이러한 양자 이득을 최대로 활용하기 위해서는 외부 제어 신호를 정밀하게 조정하는 것이 필수적이며, 복잡한 제어 전략을 데이터 기반으로 최적화할 수 있는 강화학습이 실질적인 수단으로 제시된다[2]. 그러나 다중 큐비트 환경에서의 강화학습은 큐비트 수가 증가함에 따라 상태 공간 차원이 2^N 으로 증가하면서, 정책망이 탐색해야 할 공간이 과도하게 넓어지는 학습 불안정과 과도한 계산 자원을 필요로 한다.

본 연구는 이러한 확장성 한계를 극복하기 위해, 제어 대상인 Tavis-Cummings(TC)모델의 물리적 특성을 정책 신경망 구조에 반영하는 전략을 제안한다. TC 모델은 모든 큐비트가 공진기와 균일하게 상호작용하므로, 큐비트의 순서를 어떻게 바꾸더라도 관측되는 물리량은 동일하다. 이러한 순서 무관성을 학습 모델에 반영하면, 본질적으로 동일한 상태의 중복 탐색을 막아 고차원 상태 공간에서도 안정적이고 효율적인 학습이 가능하도록 한다.

Ⅱ. 방법론

2-1. 양자 배터리 환경: Tavis-Cummings 모델

물리적 환경은 N 개의 동일한 2 수준 큐비트가 단일 모드 공진기와 상호작용하는 TC 모델을 기반으로 구성한다. 시스템의 해밀토니안은 다음과 같다.

$$\mathrm{H(t)} = \mathrm{H_{drift}} + \Omega(\mathrm{t})\mathrm{H_{drive}}$$

여기서 $\mathrm{H_{drift}}$ 는 시스템 자체의 운동을 기술한다. 큐비트의 고유 진동수 $\omega_0 = 1.0$, 결합 강도 $g = 0.1$ 으로 설정한다. $\Omega(\mathrm{t})$ 는 에이전트가 제어하는 제어 신호의 크기이며, $\mathrm{H_{drive}}$ 는 이 신호와의 상호작용을 나타낸다. 물리적 제약을 반영하여 신호 크기는 $|\Omega(t)| \leq 2.0$ 으로 제한한다. 보상은 각 시간 단위에서 저장 에너지 증가분에 제어 비용 $0.01|\Omega(t)|$을 뺀 값으로 정의한다.

2-2. 상태 표현과 정책망

네 가지 유형의 정책망을 PPO(Proximal Policy Optimization) 알고리즘 상에서 학습시켜 그 성능을 비교 분석한다[3]. 각 모델은 동일한 물리 환경과 상호작용하지만, 상태 정보를 처리하는 방식에서 차이를 보인다. 실험은 N={2, 3, 4, 6, 8} 큐비트 수에서 진행한다.

- **MLP**(Multi-Layer Perceptron): 전체 양자 상태 벡터(2^N)를 실수 표현으로 변환해 입력하며, 이 구조는 큐비트의 순서 정보에 민감하게 반응하여, 동일한 물리적 상태라도 큐비트 순서가 다르면 다른 출력 값을 생성할 수 있다.

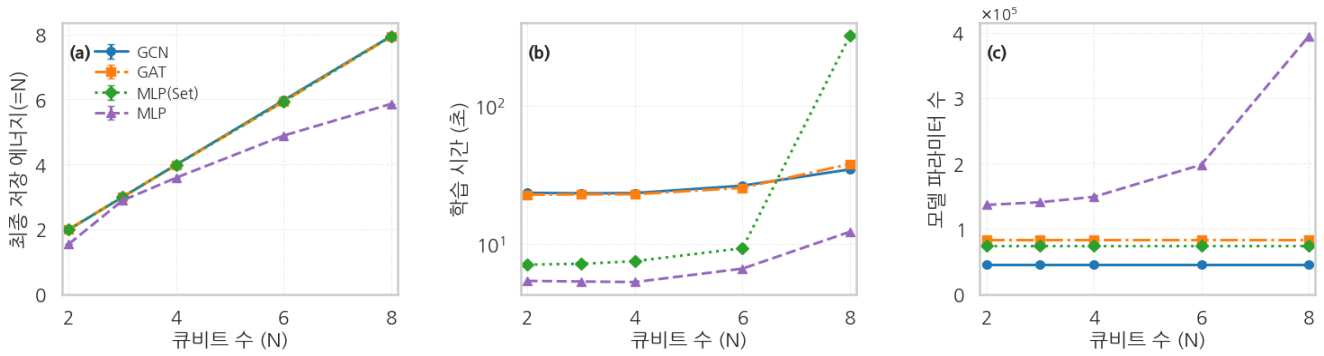

그림 1. 큐비트 수 N 에 따른 (a) 최종 에너지, (b) 학습 시간, (c) 파라미터 수 비교.

- **MLP(Set)**: 각 큐비트의 개별 관측값 $(\langle X_i\rangle, \langle Y_i\rangle, \langle Z_i\rangle)$ 을 동일한 가중치를 가진 MLP 로 처리한 후, 각 특징들의 평균을 구해 순서에 무관한 집합 표현을 만든다.
- **GCN**(Graph Convolutional Network): 큐비트를 노드로, 상호작용을 엣지로 하는 완전 연결 그래프를 구성한다. 그래프 합성곱 연산을 통해 각 큐비트의 정보를 이웃 큐비트들과 모아서 통합하고, 최종적으로 전체 노드 정보의 평균을 취해 순서에 무관한 그래프 수준 표현을 얻는다.
- **GAT**(Graph Attention Network): GCN 과 동일한 그래프 구조를 사용하지만, 어탠션 메커니즘을 통해 노드(큐비트) 간 정보 교환 가중치를 동적으로 학습한다[4]. 이후 시스템 상태에 따라 중요한 상호작용에 더 비중을 둘 수 있고, 전체 평균을 통해 순서 무관성을 확보한다.

Ⅲ. 실험 및 결과

3-1. 모델 구조별 성능

그림 1-(a)에서 순서 무관 정책망인 GCN, GAT, MLP(Set) 모델은 최종 저장 에너지는 N=8 기준에서 이론적 최대값(8.0)의 99% (7.95 ± 0.01)를 달성하였다. 반면, MLP 방식은 최대 에너지의 약 73% (5.86) 수준에 머물러, 고차원 공간에서 초흡수 현상 유도의 안정성이 떨어진다. 그림 1-(b), (c)는 모델의 효율성을 평가한다. GCN 은 GAT 와 학습 시간이 약 8% 소폭 빨랐으며, 파라미터 규모는 GAT, MLP(Set)보다 절반 이하의 현저히 적은 파라미터를 사용하여 계산 효율성과 추론 속도 측면에서 효과적이다.

3-2. 초흡수 제어 전략 분석

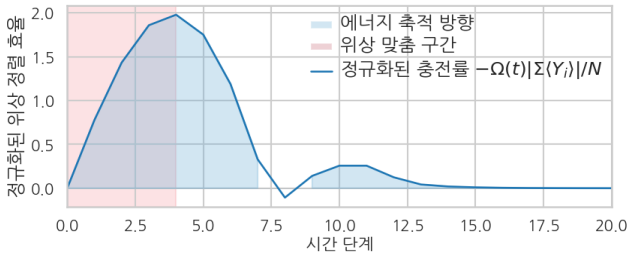

그림 2. 시간에 따른 정규화된 충전률 (GCN, N=8)

그림 2 는 N=8 환경에서의 GCN 정책망의 정규화된 위상 정렬 효율($-\Omega(t)|\sum_i\langle Y_i\rangle|/N$)의 시간적 변화를 보여준다. 이는 외부 신호와 큐비트 집단의 위상이 얼마나 일치하는지를 나타내는 척도이다. 초기 구간(붉은 음영)에서 강한 제어 신호 세기를 인가하여 큐비트들의 위상을 빠르게 맞춘다. 그와 동시에 효율이 양수인 구간(푸른 음영)에서는 제어 신호와 위상이 동기화되어 에너지가 지속적으로 축적된다. 충분한 에너지 충전 이후 신호 강도를 점차 줄여가는 전략을 학습했다. 에너지와 충전시간 모두에서 효율적인 제어 달성 방법을 찾아낸 것으로, 하드웨어 적용 시 제어 에너지를 줄일 수 있다.

추가적으로, N=8 에서 기본, 에피소드별, 스텝별 무작위 순서 입력을 각각 100 회 평가한 결과, 순서 무관 정책(GAT/GCN/MLP(Set))은 세 조건 모두에서 최종 에너지가 7.66±0.25 수준으로 거의 움직이지 않았고, 순열에 민감한 MLP 는 5.66±1.42 까지 흔들려 입력 순서에 따라 초흡수 성능이 크게 저하됐다. 이는 정책이 순서 무관성을 내재화할수록 결정적 평가에서 보이지 않던 변동성까지 억제할 수 있음을 보였다.

Ⅳ. 결론

본 연구는 다중 큐비트 양자 배터리의 TC 모델이 갖는 순서 무관성을 활용하여, 강화학습의 확장성 문제를 해결하는 그래프 신경망 기반 제어 프레임워크를 제안했다. 입력 순서에 무관하도록 설계된 GCN, GAT, MLP(Set)과 순서 의존적 MLP 를 비교한 결과, GCN 이 초흡수 현상을 가장 안정적으로 구현하면서 학습 시간과 파라미터 효율성에서 가장 우수했다. 이는 물리 시스템의 특성을 모델 구조에 반영하는 것이 고차원 제어 문제에 효과적이다. 향후에는 노이즈 특성을 반영한 실제 양자 기기의 강건한 제어 정책 학습, 대규모 다입자 양자 시스템으로 확장하는 연구를 진행할 계획이다.

ACKNOWLEDGMENT

This work was partly supported by Korea Evaluation Institute of Industrial Technology(KEIT) grant funded by the Korea government(MOTIE) (No.RS-2025-04752989, Quantum battery core technology for ultra-fast charging 100x faster than traditional lithium-ion batteries)

참 고 문 헌

[1] J. Q. Quach and D. V. Vasilyev, “Quantum batteries: The future of energy storage?” Joule, vol. 7, no. 9, pp. 1962–1986, 2023.

[2] S. H. Nengroo, D. Har, J. Hoon, T. Heo, and S. Lee, "Continuous-Variable Quantum Reinforcement Learning for HVAC Control and Power Management in Residential Building," Energy and AI, 2025.

[3] A. Raffin, A. Hill, A. Gleave, O. Kanervisto, M. Ernestus, and N. Dormann, "Stable-Baselines3: Reliable Reinforcement Learning Implementations," Journal of Machine Learning Research, vol. 22, no. 268, pp. 1-8, 2021.

[4] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph Attention Networks," in Proc. International Conference on Learning Representations (ICLR), 2018.