비마르코프 노이즈를 활용한 PPO 기반 양자 배터리 충전 프로토콜 최적화

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

{seokchu123, js03093351, pbeomdo}@gmail.com, {hjeong, htw398}@etri.re.kr,

sangkeum@hanbat.ac.kr

Optimization of a PPO-Based Quantum Battery Charging Protocol Using Non-Markovian Noise

Hyeonseok Jang[1], Junseong Park[1], Beomdo Park[1], Hoon Jeong[2], Taewook Heo[2], and *Sangkeum Lee[1]

*Hanbat National University[1], Electronics and Telecommunications Research Institute[2]

요 약

양자 배터리는 양자 역학적 특성을 활용한 차세대 에너지 저장 기술로 주목받고 있다. 하지만 물리적인 환경에서 양자 배터리는 노이즈로 인해 배터리 내부의 결맞음 상태가 붕괴되어 성능이 저하된다. 본 논문에서는 노이즈를 역으로 활용하여 충전 효율을 극대화하는 강화학습 모델을 제안한다. 제안 모델은 Proximal Policy Optimization (PPO) 알고리즘을 기반으로, 비마르코프 노이즈의 정보 역류 현상을 보상 함수에 반영하여 최대 가용 에너지인 에르고트로피를 높인다. Tavis-Cummings (TC) 시스템에서 시뮬레이션한 결과, 정보 역류 없는 PPO 모델과 노이즈 정보 역류 기반 PPO 모델은 최대 에르고트로피의 약 97.5%를 달성하였다. 이는 노이즈를 방해 요소가 아닌 유용한 정보로 활용하는 접근법의 유효성을 입증하며, 향후 양자 배터리 연구에 실질적인 기여를 할 것으로 기대된다.

Ⅰ. 서론

양자 기술의 발전에 따라, 양자 역학적 특성을 이용한 에너지 저장 장치인 양자 배터리에 대한 관심도가 증가하고 있다. 양자 역학적 특성은 배터리의 구성 요소인 큐비트 간 상대적인 위상 관계가 유지되는 결맞음 상태에서 나타난다. 그러나 외부 환경의 노이즈로 인해 결맞음 상태가 붕괴되면 충전 성능이 저하된다. 이를 해결하고자 선행 연구들은 양자 역학적 특성을 이용하여 배터리 충전 성능을 높이고 노이즈의 영향을 줄이는 데 집중한다[1]. 일반적으로 노이즈는 비가역적인 정보 손실을 유발하지만, 비마르코프 환경에서는 시스템의 과거 정보를 기억하며, 이는 충전에 유리한 자원으로 사용될 수 있다. 본 논문에서는 복잡한 환경에서도 효과적인 탐색을 수행하는 Proximal Policy Optimization (PPO) 강화학습 모델을 기반으로, 에이전트가 노이즈의 동역학을 학습하도록 프레임워크를 재구성한다. 제안된 모델의 성능 평가를 위해, 표준 실험 모델인 Tavis-Cummings (TC) 환경에서 배터리의 최대 사용 가능한 에너지인 에르고트로피로 모델 성능을 분석한다. 시뮬레이션 결과, 정보 역류 없는 PPO 모델은 거의 에르고트로피를 높이지 못한 반면, 제안된 모델은 최대값의 약 97.5%를 달성하였다. 이는 제안된 모델이 현실적인 환경에서 충전 과정에 효과적으로 기여했다는 것을 나타내며, 효율적인 양자 배터리 충전 프로토콜을 개발하는 데 이론적 토대를 제공할 것으로 기대한다.

Ⅱ. 본론

2.1. 비마르코프 노이즈

현재 정보에만 의존하는 마르코프 환경에서 노이즈의 감쇠율은 시간에 독립적인 상수로 표현된다. 또한, 노이즈로 인해 산일된 양자 배터리 시스템의 결맞음 정보는 비가역적으로 소실되어 배터리의 성능이 저하된다. 그러나 이전 시점의 정보를 반영하는 비마르코프 환경에서 노이즈의 감쇠율은 시간에 따라 바뀌며, 과거 정보는 환경에 유지된다[2]. 감쇠율이 양수일 때는 결맞음 정보가 시스템에서 환경으로 유출되는 소실 과정이 일어나지만, 음수일 때는 소실되었던 정보가 양자 시스템으로 역류하는 정보 역류 현상이 발생한다. 강화학습 에이전트가 정보 역류 구간에서 복구된 결맞음 정보를 바탕으로 충전 펄스를 제어하면, 양자 배터리의 결맞음 상태가 회복되어 충전 성능이 향상될 수 있다. 따라서, 본 연구는 에이전트가 시간에 따른 감쇠율을 관찰하여 정보 역류 현상을 인지하고, 이를 활용하도록 학습 프레임워크를 설계한다.

2.2. 제안된 PPO 프레임워크

PPO 는 정책 경사 계열의 알고리즘으로 이전 정책과 새 정책의 비율에 클리핑 기법을 적용하여 양자 배터리의 복잡한 환경에서도 안정적인 학습이 가능하다[3]. 정보 역류 현상을 활용하기 위해 PPO 의 환경은 비마르코프 노이즈를 포함한 양자 배터리의 모델링된 동역학으로

설정한다. 상태는 양자 배터리의 물리량과 노이즈의 감쇠율을 의미한다. 이때, 에이전트는 환경과 배터리 간 상호작용 정도를 조절하는 펄스의 진폭을 제어한다. 에이전트가 정보 역류 현상을 활용하도록 하는 보상 함수는 다음과 같다:

$$R_{total}(t) = R_{battery}(t) + R_{backflow}(t)$$

$R_{battery}(t)$ 는 에이전트가 양자 배터리를 충전하기 위한 기본적인 목표를 학습하도록 유도하며, 수식은 다음과 같다:

$$R_{battery}(t) = w_e \cdot \Delta E_t + w_p \cdot R_p(t) + w_{loss} \cdot R_{loss}(t)$$

ΔE_t 는 이전 타임스텝에 대한 에르고트로피 변화량으로 양수일 때 보상을 부여한다. $R_p(t)$ 는 펄스 변화에 대한 보상 항으로 급격한 펄스 변화에 대해 페널티를 부여해 안정적인 학습을 유도한다. $R_{loss}(t)$ 는 충전 중 손실되는 에너지에 대한 페널티를 부여해 불필요한 에너지 낭비를 방지한다. 다음으로 정보 역류 현상에 대한 항인 $R_{backflow}(t)$의 수식은 다음과 같다:

$$R_{backflow}(t) = w_{backflow} \cdot \Delta E_t \ (if, \gamma(t) < 0 \ and \ \Delta E_t > 0)$$

$R_{backflow}(t)$ 는 감쇠율 $\gamma(t)$ 와 에르고트로피 변화량에 따라 보상을 부여한다. 시점 t에서 감쇠율이 음수가 될 때, 정보 역류 현상이 발생하며 동시에 에르고트로피의 변화량이 양수일 경우 에이전트는 보상을 받게 된다.

Ⅲ. 시뮬레이션

3.1. 시뮬레이션 설정

제안된 모델의 성능 검증을 위해 TC 배터리 모델에서 시뮬레이션을 한 뒤, 정보 역류 없는 PPO 모델과 에르고트로피를 비교하여 결과를 분석하였다. 주요 파라미터는 표 1 과 같다.

파라미터	값
큐비트 개수	2
공진 주파수(w_0)	1.0
노이즈 진폭	$0.2 \times w_0$
노이즈 주파수	$0.5 \times w_0$
엔트로피 계수	0.02
전체 타임스텝	1×10^5
학습률	1×10^{-4}

표 1. 주요 파라미터 설정

3.2. 시뮬레이션 결과

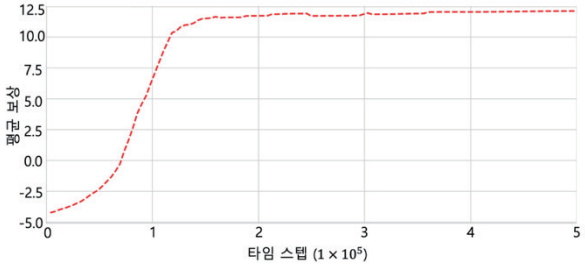

그림 1. 제안된 PPO 모델의 학습 곡선

그림 1 은 제안된 모델의 학습 곡선으로 타임스텝에 따른 평균 보상을 의미한다. 초기에는 무작위 탐색으로 인한 페널티의 영향이 커 보상이 음수에 있지만, 점진적으로 보상이 증가하여 약 120,000 번의 타임스텝부터 보상이 수렴한다. 이는 제안된 모델이 비마르코프 노이즈가 포함된 TC 배터리 시스템에 대한 최적의 제어 정책을 효과적으로 학습했음을 입증한다.

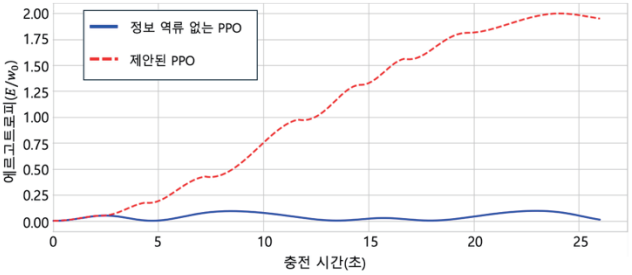

그림 2. 모델 별 에르고트로피 변화

그림 2 는 정보 역류 없는 PPO 모델과 제안된 PPO 모델의 충전 시간에 대한 에르고트로피이다. 충전 결과, 파란 선에 해당하는 정보 역류 없는 PPO 모델은 최종 에르고트로피를 약 0.01 E/w_0로 유지하였다. 반면, 빨간 선에 해당하는 제안된 PPO 모델은 에르고트로피를 약 1.95 E/w_0까지 달성하여, 최대값인 2.00 E/w_0의 약 97.5%에 달하는 성능을 보였다. 정보 역류 없는 PPO 모델은 노이즈의 동역학을 학습하지 못해 좁은 탐색 범위에서 지역 최적해로 수렴하였지만, 제안된 PPO 모델은 정보 역류 현상을 활용하여 넓은 탐색 범위에서 최적의 제어 정책을 구했다.

Ⅳ. 결론

본 연구는 비마르코프 환경에서 노이즈의 정보 역류 현상을 양자 배터리 충전 과정에 활용하는 PPO 알고리즘 기반의 강화학습 모델을 제안한다. 시뮬레이션 결과, 제안된 모델은 정보 역류 없는 PPO 모델과 달리 에르고트로피를 최대값에 근사하게 달성하였고, 이를 통해 에이전트가 정보 역류 현상을 학습하여 노이즈 환경에서도 최적의 제어 경로를 탐색할 수 있음을 보여주었다. 향후 연구 방향은 노이즈 환경을 확장하여 강건한 충전 성능을 보여주는 프레임워크를 개발하고, 궁극적으로 양자 배터리의 상용화에 기여하고자 한다.

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] Erdman, P. A., Andolina, G. M., Giovannetti, V., & Noé, F. (2024). Reinforcement learning optimization of the charging of a Dicke quantum battery. *Physical Review Letters*, *133*(24), 243602.

[2] Ghosh, S., Chanda, T., Mal, S., & Sen, A. (2021). Fast charging of a quantum battery assisted by noise. *Physical Review A*, *104*(3), 032207.

[3] Nengroo, S. H., Har, D., Jeong, H., Heo, T., & Lee, S. (2025). Continuous variable quantum reinforcement learning for HVAC control and power management in residential building. *Energy and AI*, *21*, 100541.