비마르코프 환경의 OQBs 충전 최적화를 위한 LSTM-SAC 프레임워크

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

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

LSTM-SAC Framework for Optimizing OQBs Charging in Non-Markovian Environments

Hyeonseok Jang[1], Beomdo Park[1], Junseong 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 역학을 활용하여 기존 배터리의 물리적 한계를 보완하는 에너지 저장 체계로 주목받고 있다. 그러나 외부 환경과 상호작용하는 Open Quantum Batteries (OQBs) 에서는 에너지 손실로 인해 충전 효율이 저하된다. 특히, 비마르코프 환경의 OQBs 는 소산된 정보를 기억하는 특성이 있으며, 배터리와 환경 간 결합 강도 $g(t)$의 동역학에 따라 에너지 상호작용이 변화한다. 표준 강화학습은 마르코프 의사 결정을 전제하므로, 과거 정보를 고려하는 비마르코프 특성을 학습하기 어렵다. 본 연구는 Soft Actor-Critic (SAC) 강화학습 모델에 Long Short-Term Memory (LSTM) 을 결합하여 비마르코프 환경의 OQBs 충전 프로토콜을 최적화하는 프레임워크를 제안한다. LSTM-SAC 는 LSTM 으로 환경의 맥락을 파악하여, 충전 과정에서의 환경으로 인한 에너지 손실을 최소화한다. 실험 결과, LSTM-SAC 는 최대 가용 에너지인 에르고트로피를 최대 충전량의 약 93%를 충전하였으며, LSTM-SAC 의 각 제어 변수는 $g(t)$와 평균 -0.6 의 음의 상관관계를 보였다. 이는 LSTM-SAC 이 비마르코프 환경에서 강건한 충전 정책을 도출했음을 입증한다.

Ⅰ. 서 론

최근 센서, 통신 등 여러 분야에서 미세한 규모의 디바이스가 활용됨에 따라, 원자 단위에서도 안정적으로 에너지를 저장하는 양자 배터리가 주목받고 있다. 양자 배터리는 양자 역학적 특성을 이용하여 빠르고 효율적으로 충전되기 때문에, 기존 배터리의 물리적 한계를 극복한다. 그러나 양자 배터리는 외부 환경과 상호작용하는 Open Quantum Batteries (OQBs) 를 시뮬레이션 하는 과정에서 에너지 손실이 발생한다[1]. 이전 OQBs 충전 시뮬레이션은 주로 마르코프 의사 결정을 전제로 한 표준 강화학습 기반의 최적화를 진행하였다. 하지만 비마르코프 환경은 과거 정보에 따라 다음 상태가 달라지는 시간적 상관관계가 존재하므로, 표준 강화학습만으로는 충전 성능을 확보하기 어렵다. 본 연구는 비마르코프 환경의 시계열 특성을 처리하기 위해, Long Short-Term Memory (LSTM) 네트워크를 Soft Actor-Critic (SAC)의 Actor 와 Critic 구조에 결합한 LSTM-SAC 프레임워크를 제안한다. MLP 기반의 표준 SAC 는 버퍼에 데이터를 저장하고, 학습 시 무작위 샘플을 선정하여 학습한다. 그러나 이러한 학습 방식은 데이터 간 상관관계를 고려하지 않아, 비마르코프 환경에서는 부적절하다. 반면, LSTM-SAC 는 저장된 데이터의 특정 패턴을 활용하여 학습하기 때문에, 비마르코프 환경에서도 안정적인 충전 성능을 유지한다. 시뮬레이션 결과, LSTM-SAC 는 배터리의 최대 가용 에너지인 에르고트로피를 이론적 최대값인 1.00 $\hbar\omega_0$의 약 93%인 $0.93\hbar\omega_0$를 달성하였고, MLP 기반의 표준 SAC 대비 약 8 배의 성능 향상을 보였다. LSTM-SAC 의 제어 변수를 분석한 결과, $g(t)$와 최대 -0.7 의 강한 음의 상관관계를 가진 역펄스를 생성하여 충전 프로토콜을 최적화하였으며, 이는 LSTM-SAC 가 환경에 대한 강건한 충전 정책을 학습하였음을 시사한다.

Ⅱ. 본론

2.1 비마르코프 환경의 OQBs

마르코프 환경은 주파수 영역이 평탄하기 때문에, 에너지는 비가역적으로 소실된다. 반면, 비마르코프 환경은 구조화된 주파수 영역으로부터 발생하는 메모리 효과로 인해 환경과 배터리 간 에너지는 양방향으로 상호작용한다[2]. 비마르코프 환경의 OQBs 는 충전기, 배터리, 구조화된 저장소로 구성된다. 배터리는 외부 펄스가 직접 가해질 경우 내부 상태가 붕괴되므로, 간접적인 제어인 충전기로부터 결합 강도 $\kappa(t)$에 따라 에너지를 얻는다. 또한, 배터리에 저장된 에너지는 배터리와 환경의 결합 강도 $g(t)$에 비례하여 소산한다. 해당 시스템의 전체 해밀토니안은 다음과 같다:

$$H(t) = H_B + H_C + H_e + H_I(t)$$

H_B, H_C, H_e는 각각 배터리, 충전기, 환경에 대한 고유한 상수값이다. $H_I(t)$는 배터리와 환경 간 상호작용을 의미하며, 수식은 다음과 같다:

$$H_I(t) = g(t)\,(\sigma_+^B\sigma_-^e + \sigma_-^B\sigma_+^e)$$

$H_I(t)$는 충전에 해당하는 $\sigma_+^B\sigma_-^e$ 연산자와 에너지 소산에 해당하는 $\sigma_-^B\sigma_+^e$ 연산자가 동시에 동작하는 양방향 상호작용을 의미하며, $g(t)$에 따라 상호작용의 강도와 방향성이 동적으로 변화한다.

2.2 LSTM-SAC 프레임워크

비마르코프 환경의 복잡한 동역학을 처리하기 위해 LSTM-SAC 는 은닉 층을 LSTM 으로 대체하여 $g(t)$의 시계열 특성을 파악하고 정책에 반영한다[3]. Actor 는 시점 t의 현재 상태 s_t와 과거 정보가 압축된 은닉 상태 h_{t-1}를 입력받아, 정책 $\pi(a_t|s_t,h_{t-1})$을 산출하고, 이를 통해 제어 변수인 $\kappa(t)$와 구동장 펄스 진폭 $\eta(t)$를 결정한다. Critic 은 s_t와 a_t 뿐만 아니라, LSTM 네트워크로부터 출력된 h_t를 반영하여 누적 보상의 기대값인 Soft Q-value $Q(s_t,a_t,h_t)$를 계산한다. Actor 네트워크는 계산된 Soft Q-value 로부터 피드백을 받아, 정책을 업데이트한다. 보상 함수 R_t는 비마르코프 환경에서 에이전트가 안정적으로 충전하도록 설계하였으며 수식은 다음과 같다:

$$R_t = W_1\Delta E(t) + R_{gating}(t)$$

$\Delta E(t)$ 는 에르고트로피 증가량으로써 다음과 같이 정의된다:

$$\Delta E(t) = E(t) - E(t-\Delta t)$$

이는 에르고트로피 증가량을 의미하며, 값이 양수인 경우 보상을 부여한다. $R_{gating}(t)$ 항은 $g(t)$ 로부터 강건한 충전 전략을 학습하도록 유도되며, 다음과 같이 정의된다:

$$R_{gating}(t) = -W_2 g(t)\kappa(t) - W_3 g(t)\eta(t)$$

이는 $g(t)$가 양수일 때, 제어 변수 값을 줄여 에너지 손실을 방지하고, 음수일 때 값을 증폭시켜 에너지를 확보한다.

Ⅲ. 실험 및 결과

3.1 실험 설정

ω_0와 $\hbar$는 양자 시스템의 고유 단위로, 1.0 으로 설정하여 물리량을 정규화한다. $g(t)$ 는 주기적으로 결합 강도가 변화하도록 $g(t) = 0.5 \cdot cos\,(2\pi t)$로 근사한다.

3.2 실험 결과

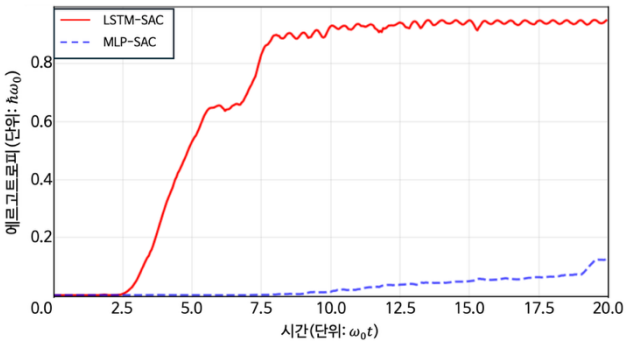

그림 1. 에르고트로피 충전량

그림 1 은 LSTM-SAC 와 MLP 기반의 표준 SAC 의 시간에 따른 에르고트로피 충전량을 보여준다. LSTM-SAC 는 약 $2.5\omega_0 t$ 동안 $g(t)$의 동역학 패턴을 학습한 후, 약 $8\,\omega_0 t$ 까지 에르고트로피를 충전하였다. 또한 충전 이후에도 안정적으로 에르고트로피를 유지하였다. 반면, 표준 SAC 는 최종 에르고트로피가 약 $0.12\ \hbar\omega_0$로, 이는 $g(t)$에 대한 강건한 충전 정책을 학습하지 못하였음을 의미한다.

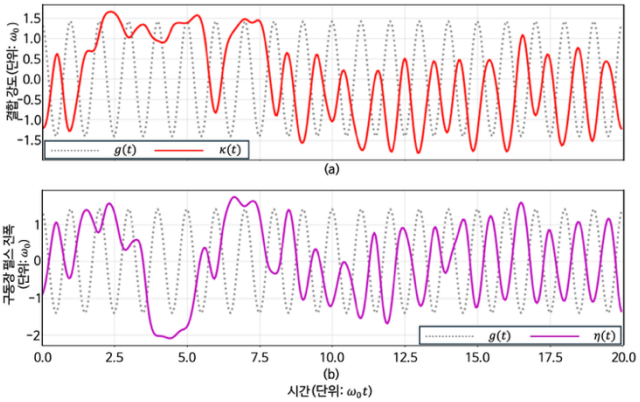

그림 2. LSTM-SAC 제어 변수 펄스

그림 2 는 LSTM-SAC 가 조절한 각 제어 변수 $\kappa(t)$, $\eta(t)$의 $g(t)$에 대한 제어 펄스이다. 그림 2-(a)의 $\kappa(t)$는 약 $8\,\omega_0 t$ 까지의 충전 구간에서 값을 높여 에너지 흡수를 최대화한다. 그러나, 그림 2-(b)의 $\eta(t)$는 동일 구간에서 값이 감소하는데, $\eta(t)$의 값이 높을 경우 환경과 배터리 간 상호작용 빈도가 커져 충전이 저하되기 때문이다. 반면, 충전 구간 이후 $\kappa(t)$ 와 $\eta(t)$ 는 모두 $g(t)$ 의 역펄스로 나타난다. 이는 저장된 에너지가 손실되지 않도록 에이전트가 제어 펄스를 동적으로 조절한 것을 의미한다. 해당 구간에서 $g(t)$가 양수일 경우, $\kappa(t)$ 와 $\eta(t)$ 값이 커도 흡수 에너지보다 소산되는 에너지가 많아, 두 제어 변수를 억제한다. $g(t)$가 음수일 때는, 두 값을 증가시켜 에너지 교환을 조절한다. 따라서, LSTM-SAC 에이전트는 시간에 따라 펄스 제어 전략을 변화시켜 강건한 충전 프로토콜을 학습하였음을 알 수 있다.

Ⅳ. 결론

비마르코프 환경의 OQBs 시스템에서 발생하는 충전 효율 저하 문제를 해결하고자, 환경의 시계열 특성을 학습에 활용하는 LSTM-SAC 프레임워크를 제안하였다. LSTM-SAC 는 MLP 기반의 표준 SAC 과 달리, 환경의 동역학을 추론하여 최적화된 충전 전략을 학습하였다. 향후 연구는 LSTM-SAC 성능의 일반화 검증을 위해 다양한 실험 환경에서 시뮬레이션을 수행할 예정이며, 이를 통해 양자 배터리 관리 시스템 및 에너지 저장 장치 연구에 이론적 토대를 제공하고자 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] S. Zakavati et al., “Optimizing the Charging of Open Quantum Batteries using Long Short-Term Memory-Driven Reinforcement Learning,” arXiv preprint arXiv:2504.19840, 2025.

[2] Breuer H. P. et al. " Colloquium: Non-Markovian dynamics in open quantum systems", Reviews of Modern Physics, vol. 88, no. 2, pp. 021002, 2016

[3] I. Kim et al., “Frequency Hopping Synchronization by Reinforcement Learning for Satellite Communication System,” arXiv preprint arXiv:2503.04266, 2025.