

하이브리드 선박 에너지 최적화를 위한 오프라인 강화학습 연구

김서영, 김웅섭

동국대학교 정보통신공학과

ksy5098@dongguk.edu, woongsup@dongguk.edu

A Study of Offline Reinforcement Learning for Optimal Hybrid Ship Energy Management

Seo-Young Kim, Woongsup Kim

Department of Information Communication Engineering, Dongguk University

요약

하이브리드 선박의 에너지 관리 시스템 최적화는 연료 효율성과 배출가스 저감을 위한 핵심 과제이나, 실측 데이터 부족으로 인해 온라인 강화학습 알고리즘을 적용하기에는 한계를 보인다. 본 연구는 이러한 배경에서 Deep Double Q-Network(DDQN)과 Conservative Q-Learning(CQL) 알고리즘을 시뮬레이션 환경과 오프라인 환경에 적용하여 그 효율성을 비교·분석하였다. 결과적으로 시뮬레이션 환경에서 CQL 이 DDQN 보다 우수한 성능을 보였으나, 오프라인 환경에서의 CQL 은 학습 안정성 문제를 나타냈다. 이는 오프라인 강화학습의 한계와 실제 환경과의 갭을 보여준다. 본 연구는 이를 통해 하이브리드 선박의 강화학습 기반 에너지 관리 시스템의 실용성을 평가하고, 최적화된 선박 에너지 관리 방법론 개발에 기여할 수 있는 발전방향을 제시한다.

I. 서론

하이브리드 선박은 해양 환경 보호와 에너지 효율성 향상을 위해 점점 더 중요해지고 있다. 이러한 선박에서는 배터리와 연료전지 등 다양한 에너지원의 효율적 운용을 통해 연료 소비를 줄이고 배출가스를 최소화하는 것이 핵심 과제이다. 따라서 현재 강화학습을 활용한 하이브리드 선박 에너지 관리 최적화에 대한 다양한 연구가 수행되고 있다[1].

그러나 고품질의 하이브리드 선박 운항 실측 데이터를 충분히 확보하기 어렵고 강화학습의 특징인 trial-and-error 를 통한 학습이 시뮬레이터를 통해서만 가능하다는 점에서 강화학습을 적용하는데 있어 한계를 보이고 있다. 이에 따라, 그 대안으로 오프라인 강화학습이 주목받고 있다. 오프라인 강화학습에서는 이미 수집된 데이터셋을 이용하여 학습을 수행하며, 대표적인 오프라인 강화학습 알고리즘으로는 CQL 과 Batch Constrained Q-Learning(BCQ) 등이 있다[2,3].

본 연구에서는 해양 선박 데이터를 기반으로 온라인 환경에서 DDQN 과 CQL 알고리즘을 적용한 결과와 오프라인 환경에서 CQL 알고리즘을 적용한 결과를 비교하고 그 결과를 통해 오프라인 알고리즘의 효율성을 분석하였다.

II. 본론

2.1 실험 환경

본 연구에서 하이브리드 선박의 에너지 관리 시스템하의 강화학습 실험 환경을 구축하였다. 이 환경은 실제 선박 운용 조건을 모사하여, 에너지 효율을 최대화하는 정책 학습을 목표로 한다. 연구에서 적용한 하이브리드 선박은

추진에도 배터리가 사용되는 폴 하이브리드 선박[4]으로 메인 엔진의 최대 출력과 배터리의 용량은 24,000kWh 로 설정하였다.

강화학습 에이전트의 상태 공간 S 는 다음과 같이 정의하였다.

$$S = \{spA, p_{dem}, x_t, SoC, v\} \quad (1)$$

$spA \in \{0, 1\}$: 선박 운용 모드 (0: 운항, 1: 정박)

$p_{dem} \in [0, p_{max}]$: 현재 전력 수요량 (kW)

$x_t \in [0, x_{max}]$: 이전 단계의 엔진 출력 (kW)

$SoC \in [0, 1]$: 배터리 충전률

$v \in [0, v_{max}]$: 선박 속도 (km/h)

여기서 p_{max} , x_{max} , v_{max} 는 각각 최대 전력 수요, 최대 엔진 출력, 최대 선박 속도를 나타낸다.

행동 공간 A 는 변화된 엔진 출력량을 나타낸다. A 에는 미세한 수준까지 표현되는 엔진 출력량과 엔진 가동이 완전히 중지하는 옵션을 포함한다.

보상 함수 R 은 세 가지 주요 요소를 고려하여 설계하였다. 세 가지 주요 요소는 1) 배터리 SOC 의 안정성 2) 연료 소비량 최소화 3) 전력 수요 충족 이다.

연료 소비량 계산에는 G90ME-C10.5 엔진의 Specific Fuel Oil Consumption(SFOC) 특성 모델[4]을 사용하였다. 이를 통해 엔진 출력량에 따른 연료 소비량을 정확히 추정할 수 있게 해준다.

시뮬레이션 환경과 오프라인 환경에서 사용한 데이터셋으로는 6 개월 동안 15 초 간격으로 측정된 선박의 실제 운항 데이터를 가공해서 사용하였다. 원본 데이터셋은 15 개의 열('MAIN Engine #1', 'MAIN Engine #1.1', 'Total Main Engine', 'DFDG #1', 'DFDG #2', 'DFDG #3', 'DFDG #4', 'Total DFDG', 'Shaft Generator #1',

'Shaft Generator #2', 'Total Shaft Generator', 'Cargo Pump Run Count', 'Vapor Return Pump Current', 'Draft', 'Main Engine #1 RPM')과 1,048,320 개의 행으로 구성되어 있다.

2.2 심층 강화학습 알고리즘 - DDQN, CQL

본 연구에서는 하이브리드 선박의 에너지 관리 시스템 최적화를 위해 두 가지 주요 강화학습 알고리즘인 DDQN 과 CQL 을 적용하였다.

DDQN 은 DQN 의 주요 약점인 Q-value 과대평가 문제를 해결하기 위해 개발된 강화학습 알고리즘으로, 두 개의 Q-network 를 사용하며, 하나의 네트워크는 최적 행동을 선택하고, 다른 네트워크는 선택된 행동의 가치를 평가하도록 하여 Q-value 추정의 편향을 줄이고, 결과적으로 학습의 안정성과 성능을 향상시킨다.

CQL 은 오프라인 강화학습의 핵심 과제인 분포 이동 문제를 해결하기 위해 고안된 알고리즘이다. CQL 은 CQL 은 데이터셋에서 관찰된 행동의 Q-value 는 보존하는 동시에 모든 가능한 행동에 대한 Q-value 의 기대값을 최소화하는 정규화 항을 도입함으로써 데이터 분포 밖 행동에 대한 Q-value 를 하향 조정하여 과대 평가를 방지한다. CQL 은 데이터에 내재된 정보를 효과적으로 활용하면서도 과도한 외삽(extrapolation)을 방지하여 오프라인 설정에서의 훈련 안정성을 크게 향상시킨다.

III. 결과

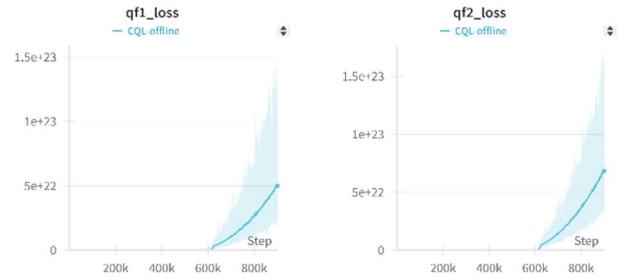
시뮬레이션 환경에서 DDQN 과 CQL 알고리즘의 성능을 비교하였다. 그 결과, CQL 알고리즘은 DDQN 대비 우수한 성능을 보였다. 특히, 시뮬레이션 환경에서 CQL 은 에피소드 진행에 따라 높은 보상 값을 유지하며 학습이 안정적으로 진행되었다(그림 1a). 또한, CQL 은 연료 소비량 측면에서도 DDQN 보다 일관된 감소 추세를 보이며, 에너지 소모량을 감소하는 데 기여하였다(그림 1b).

그러나, 오프라인 환경에서의 CQL 학습은 불안정한 학습 결과를 보였다. CQL 알고리즘의 Q-function 손실이 에피소드가 진행됨에 따라 기하급수적으로 증가하였다(그림 2). 이러한 결과는 오프라인 강화학습이 적은 실제 데이터만을 사용할 때, 학습 안정성이 저하될 수 있음을 보여준다. 특히, CQL 알고리즘은 Q-value 의 과대평가를 억제하는 특성이 있음에도 불구하고, 데이터의 수가 부족할 경우 성능 저하가 발생하는 것을 확인할 수 있다.



[그림 1] DDQN 과 CQL 알고리즘의 학습 결과

a. 에피소드에 따른 보상값(상). b. 에피소드에 따른 연료 소모량(하)



[그림 2] 오프라인 환경에서의 CQL 알고리즘 학습 시 Q-function loss 값 결과

IV. 고찰 및 결론

본 연구는 하이브리드 선박의 에너지 관리 시스템 최적화를 위한 강화학습 알고리즘의 적용 가능성을 탐구하였다. 시뮬레이션 환경에서 CQL 이 DDQN 보다 우수한 성능을 보인 점은 주목할 만하다. CQL 의 보수적 학습 방식이 에너지 관리 시스템의 변동성을 효과적으로 제어할 수 있음을 시사하며, 이는 과대평가된 행동 가치를 체계적으로 규제하는 CQL 의 특성에 의한 것으로 보인다.

그러나 오프라인 데이터셋만을 사용한 CQL 의 성능 저하는 오프라인 강화학습의 한계를 드러낸다고 할 수 있다. 이는 실제 환경과의 상호작용 부재로 인한 것으로 추정되며, 선박 운용의 오프라인 데이터만으로 효과적인 정책을 학습하는 데 있어 여전히 한계가 있음을 보여준다.

이에 따라 오프라인 학습의 한계를 극복하기 위해서는 추가적인 연구가 필요함을 확인할 수 있으며, 향후 연구 방향으로 다음을 제시한다.

1. 다양한 운영 시나리오를 고려한 양질의 오프라인 데이터셋 구축
2. 실제 환경과 시뮬레이션 환경 간의 갭을 줄이기 위한 실제 데이터를 기반으로 한 시뮬레이션 모델의 지속적인 개선
3. 오프라인 환경에서의 학습 안정성을 높이기 위한 알고리즘 개선

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-학석사연계 ICT 핵심인재양성 지원을 받아 수행된 연구임 (IITP-2024-00436744)

참고 문헌

- [1] Wang, Xiaoyu, et al. "Energy Management Strategy for Diesel-Electric Hybrid Ship Considering Sailing Route Division Based on DDPG." IEEE Transactions on Transportation Electrification 10.1 (2023): 187-202.
- [2] Fujimoto, Scott, David Meger, and Doina Precup. "Off-policy deep reinforcement learning without exploration." International conference on machine learning. PMLR, 2019.
- [3] Kumar, Aviral, et al. "Conservative q-learning for offline reinforcement learning." Advances in Neural Information Processing Systems 33 (2020): 1179-1191.
- [4] MAN Energy Solutions. "Batteries on board ocean-going vessels" (2019).
- [5] Van Hasselt, Hado, Arthur Guez, and David Silver. "Deep reinforcement learning with double q-learning." Proceedings of the AAAI conference on artificial intelligence. Vol. 30, No. 1. 2016.