시나리오 균등화를 통한 강화학습 기반 표적기동분석의 안정성 및 정확도 향상

신준하, 고현석*
한양대학교

ipip0114@hanyang.ac.kr, *hyunsuk@hanyang.ac.kr

Improving DRL-based Target Motion Analysis through Stratified Scenario Equalization

Shin Junha, Ko Hyunsuk*
Hanyang Univ.

요 약

심층 강화학습(DRL) 기반의 표적기동분석(TMA)은 방위각 관측 데이터 기반 추정의 노이즈 강인성을 확보하는 효과적인 접근법이다. 그러나 시뮬레이션 환경에서 요구되는 가관측성 기준 필터링은 특정 표적 속력에 대한 시나리오의 분포를 편중시켜 시나리오 불균형 문제를 야기한다. 이 불균형은 DRL 에이전트를 특정 시나리오에 과적합시키며, 다양한 운용 환경에서의 일반화 성능을 저해하는 주요 요인으로 작용한다. 본 연구에서는 이 문제를 해결하고자 학습 시 생성 시나리오 균등화 기법을 도입하고 그 효과를 검증했다. 표적 속력을 기준으로 시나리오 그룹을 계층화하고, 각 그룹별로 목표 학습 샘플 수를 강제적으로 할당하는 계층적 샘플링 전략을 적용하여 훈련 데이터의 분포를 균등하게 맞춘다. 실험은 Residual-GRU 기반 DRL-TMA 프레임워크에 통합하여 Proximal Policy Optimization(PPO) 에이전트 학습을 진행하고, 실험 결과, 시나리오 균등화 기법을 적용했을 때 불균형 학습 환경 대비 초기 학습 단계에서의 안정성과 속력 추정 정확도가 개선되는 결과를 보였다.

Ⅰ. 서 론

수동적인 방위각 관측만을 이용하는 표적 운동 분석(Target Motion Analysis, TMA)은 수중 감시 및 추적 환경에서 핵심적인 기술이다. 전통적인 TMA 기법들은 높은 계산 효율성을 제공하지만, 추정 성능이 초기값 설정에 민감하고 낮은 신호 대 잡음비(SNR) 환경에서 취약하다[1]. 특히, 방위각 변화량이 적은 경우 발생하는 가관측성(Observability) 부족 문제로 인해 추정 오차가 누적되거나 발산하는 한계를 가진다[2]. 이러한 제약 사항들은 TMA 시스템의 강인성과 자동화에 대한 요구를 증대시킨다.

최근 순차적인 의사 결정 및 시계열 데이터 처리에 강점을 보이는 심층 강화학습(DRL)이 TMA 문제의 대안으로 부상했다. DRL 기반 TMA 는 관측된 방위각-시간 기록의 시계열 특징을 학습하기 위해 Residual-GRU 와 같은 순환 신경망 아키텍처를 활용하며, 기존 기법 대비 노이즈 강인성 및 비선형성 추정 능력을 향상시키는 잠재력을 보여준다.

그러나 DRL 에이전트를 훈련시키기 위해 구축되는 시뮬레이션 환경은 본질적인 문제에 직면한다. TMA 의 성공을 보장하기 위해 에피소드 시작 시 최소 방위각 변화량을 충족하는 시나리오만을 선별적으로 채택하는 필터링 과정이 필수적이다. 이 필터링 과정은 결과적으로 특정 표적 속력(V_T) 범위에 대한 시나리오의 분포를 편중시키는 시나리오 불균형을 초래한다. 이처럼 불균형한 데이터셋으로 학습된 DRL 에이전트는 특정 시나리오에 과적합되어, 실제 환경에서 마주하게 될 다양한 운용 시나리오에 대한 일반화 성능이 크게 저하된다.

본 논문은 DRL 기반 TMA 시스템의 일반화 성능을 저해하는 이 시나리오 불균형 문제를 해결하기 위한 시도로 시나리오 균등화 기법을 적용한다. 표적 속력 그룹에 기반한 계층적 샘플링을 통해, 훈련 데이터가 모든 속력 범위에 걸쳐 균등하게 분포되도록 강제한다. 이러한 시나리오 균등화(Scenario Equalization, SE)는 Residual-GRU 기반의 DRL-TMA 프레임워크에 통합하여 학습 안정성과 추정 정확도를 향상시킨다.

Ⅱ. 본론

Ⅱ-1. 시나리오 균등화 및 DRL 프레임워크

DRL 에이전트는 시뮬레이션 환경에서 생성된 N 개의 시나리오 $S = \{S_1, S_2, \cdots, S_N\}$ 를 통해 학습된다. TMA 시나리오 생성 시, 추정의 성공을 위해 필수적인 방위각 변화량 기준을 만족하지 못하는 시나리오는 사용되지 않게 된다. 이 과정에서 특정 표적 속력 그룹의 시나리오가 과도하게 제외되거나 잔존하여 시나리오 분포의 불균형이 발생한다.

이를 해결하기 위해 본 연구는 시나리오 균등화를 도입한다. 균등화는 추정 목표인 표적 속력을 기준으로 시나리오를 K 개의 구간으로 나누고, 에피소드 시작 시 각 속력 그룹에서 목표 학습 샘플 수가 채워질 때까지 우선적으로 시나리오를 선택한다.

$$P(\text{sampling } S \in \text{Group}_k) \propto \frac{\text{Target Sample Count} - \text{Current Count}_k}{\sum(\text{Target Sample Count} - \text{Current Count}_i)} \quad (1)$$

학습 환경은 속력 그룹(V_T)을 기반으로 계층적 샘플링을 수행하며, 에이전트는 모든 속력 범위에 대해 균등한 경험을 얻고 정책의 일반성을 확보한다.

II-2. Residual-GRU 기반 DRL 아키텍처

본 연구의 DRL 프레임워크는 Proximal Policy Optimization(PPO) 알고리즘을 사용하며, 상태 공간(S)의 표현을 위해 Residual-GRU 특징 추출기를 채택한다. 상태 s_t는 관측된 방위각-시간 기록을 Residual-GRU 를 통과시켜 추출된 정제된 특징 벡터이며, 행동 a_t 는 에이전트가 추정하는 표적 속력(V_T)이다. Residual-GRU 는 1D-Conv 레이어로 노이즈를 필터링하고 잔차 연결을 통해 정보 보존 및 GRU 학습 안정성을 달성한다.

$$r_t = -\left|\hat{V}_T - V_T^{true}\right| \quad (2)$$

보상은 (2)와 같이 추정 속력 ($\hat{V}_T$)과 실제 속력 (V_T^{true})간의 오차 기반 페널티로 정의된다.

II-3. 실험 결과 및 분석

실험은 모의 TMA 시뮬레이터 환경에서 수행되었으며, PPO 알고리즘을 사용했다. 관측 데이터에는 가우시안 노이즈($\sigma = 1.0$)를 주입했다. 성능 지표로는 평균 절대 오차(Mean Absolute Error, MAE), 평균 제곱 오차(Mean Squared Error, MSE), 그리고 특정 오차 기준내에서 추정에 성공한 비율인 성공률(Success Rate)을 사용했다. 시나리오 균등화의 효과를 검증하기 위해 균등화를 적용하지 않은 선행 연구를 Baseline 모델로 사용하여 비교했다.

Model	MAE	MSE	Success Rate
GRU-Only [2]	0.99	1.75	61.96%
Residual-GRU	0.99	1.84	61.75%
Residual-GRU (SE)	**0.88**	**1.57**	**71.23%**

표 1. 시나리오 균등화 적용 전후의 속력 추정 성능 비교 (300 timestep 기준)

표 1. 의 결과는 시나리오 균등화 기법이 학습 안정성에 미치는 긍정적인 영향을 명확히 보여준다. SE 적용 모델은 초기 학습 단계에서 MAE 를 0.99 에서 0.88 로 11.11% 감소시켰고, 특히 성공률을 61.75%에서 71.23%로 약 9.5% 향상시켰다. 이는 시나리오 균등화를 통해 학습 초기 단계에서 다양한 환경에 대한 경험을 얻어, DRL 에이전트가 초기 단계의 편향을 극복하고 더 빠르고 안정적인 정책을 수립했음을 의미한다.

Ⅲ. 결론

본 논문은 DRL 기반 TMA 시스템의 시뮬레이션 환경에서의 시나리오 불균형 문제를 해결하기 위한 시도로 시나리오 균등화 기법을 적용했다. 제안된 SE 를 Residual-GRU 기반 DRL 프레임워크에 통합한 결과, 특히 학습 초기 단계에서 표적 속력 추정의 정확도와 성공률을 획기적으로 개선하여 학습의 안정성을 향상시키는 결과를 확인했다. 이는 TMA 문제 해결에 있어 모델 구조뿐만 아니라 학습 데이터의 분포 관리가 핵심적인 요소임을 입증한다. 향후 연구에서는 본 시나리오 균등화 기법을 다중 표적 추정 문제로 확장할 계획이다.

참 고 문 헌

[1] J.-M. Passerieux, D. Pillon, P. Blanc-Benon, and C. Jauffret, “Target motion analysis with bearings and frequencies measurements via instrumental variable estimator (passive sonar),” Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 2645-2648, 1989.

[2] S. Jang, J. Shin, D. Kim, J. Lee, and H. Ko, “Reinforcement learning-based automated target motion analysis in underwater environments,” *Ocean Eng.*, vol. 342, p. 122946, 2025