근접 정책 최적화를 통한 드론 자율운항 모델예측제어 알고리즘의 고도화

강남욱[1,2], 윤우승[1,3], 한경림[1,4]*

[1]한국과학기술연구원 뇌과학연구소 뇌융합기술연구단, [2]건국대학교 스마트운행체공학과, [3]서울대학교 물리천문학부, [4]국가연구소대학교 KIST스쿨 바이오-메디컬 융합전공

namwook99@kist.re.kr, riowoosung@kist.re.kr, *khan@kist.re.kr

Advancement of Model Predictive Control Algorithms for Drone Autopilot Using Proximal Policy Optimization

Namwook Kang[1,2], Wooseung Yoon[1,3], Kyungreem Han[1,4]*

[1]Center for Brain Technology, Brain Science Institute, Korea Institute of Science and Technology, Seoul 02792, Korea, [2]Department of Smart Vehicle Engineering, Konkuk University, Seoul 05029, Korea, [3]Department of Physics and Astronomy, Seoul National University, Seoul 08826, Korea, [4]Division of Bio-Medical Science & Technology, KIST School, Korea National University of Science and Technology, Seoul 02792, Korea

요 약

최근 드론 제작 기술의 빠른 발전과 이에 따른 응용 분야의 확장은 기존에 사용되어 온 모델 예측 제어(MPC)나 비례-적분-미분 제어(PID) 기반 운항 알고리즘의 고도화를 요구한다. 본 연구에서는 드론 운항 제어에 널리 사용되어 온 MPC 알고리즘에 근접 정책 최적화(PPO)를 적용한 새로운 개념의 PPO 강화학습-MPC 알고리즘을 제안한다. 제안하는 강화학습기반 MPC 알고리즘의 원리와 주요 특징을 기존의 MPC 알고리즘과 비교하였고, 움직이는 게이트를 통과하는 시뮬레이션 시스템을 통하여 제어 성능의 향상을 검증하였다. PPO 강화학습-MPC 알고리즘은 기존 MPC 알고리즘에 비해 움직이는 게이트 통과 확률이 의미 있게 높았으며, 운항 궤적이 목표 달성에 더욱 효율적으로 조절됨을 확인하였다. 이 연구는 앞으로 두뇌 기저핵의 운동 선택 강화학습 기전을 모사한 뉴로모픽 PPO 강화학습-MPC 개발의 핵심 플랫폼을 제공한다.

Ⅰ. 서 론

최근 드론의 활용이 급증하면서 영상 촬영, 농업, 군사 기술 등 다양한 분야에서 드론 제어 기술의 연구가 활발하게 이루어지고 있다. 특히 드론의 궤적 생성 알고리즘은 드론 자율운항 제어의 핵심 요소로, 드론의 비행 성능과 안정성에 직접적인 영향을 미친다. 기존 MPC (Model Predictive Control)나 PID (Proportional-Integral-Derivative) 제어법을 사용한 연구들에서는 일반적으로 MPC가 PID에 비해 더 우수한 성능을 보였으나, MPC 기반 제어법이 여전히 모든 환경에서 최적의 성능을 보장하지는 않을 것이다. 본 연구에서는 강화학습법 중 하나인 근접 정책 최적화를 적용하여 기존 MPC 알고리즘을 개선하고, 이를 통한 드론 궤적 생성 최적화 방법을 제안한다.

Ⅱ. 본론

본 연구에서는 기존의 MPC 알고리즘에 강화학습법 중 하나인 근접 정책 최적화를 적용하여 향상된 PPO (Proximal Policy Optimization)-MPC 알고리즘을 개발한다. 제안된 MPC 알고리즘은 강화학습을 통해 드론 제어 파라미터를 최적화하여, PPO 미적용에 비하여 움직이는 게이트 통과 성능 향상을 보였으며 운항 궤적이 목표 달성에 더욱 효율적으로 조절됨을 확인하였다.

1. 드론 제어 MPC 알고리즘 분석

MPC는 현재 상태를 기반으로 미래 상태를 예측하고, 이를 바탕으로 최적의 제어 입력을 정하는 제어 방법이다.[1] 드론에서 MPC의 주요 목표는 정확한 경로 추적과 실시간 제어를 달성하는 것이다. MPC 알고리즘은 시스템의 동적 모델을 사용하여 여러 시간 단계에 걸쳐 예측을 수행하고, 각 단계에서 비용 함수를 최소화하는 제어 입력을 결정한다. 기존의 MPC는 고정된 제어 파라미터를 사용하여 드론의 궤적을 생성하지만, 이는 고속 비행이나 복잡한 기동을 요구하는 환경에서 한계가 드러난다. 이를 개선하기 위해, 기존의 MPC 알고리즘에 강화학습을 적용하여 적응형 제어 파라미터를 사용하는 향상된 MPC 알고리즘을 개발할 수 있다.[2] 이러한 접근 방식은 다양한 환경에서 안정적인 성능을 보장한다.

2. 근접 정책 최적화(PPO) 알고리즘 분석

근접 정책 최적화(PPO) 알고리즘은 강화학습에서 사용되는 정책 기반 최적화 알고리즘으로, 정책의 안정성과 표본추출 효율성을 높인다.[3] PPO는 정책 업데이트 과정에서 큰 변화를 방지하기 위해 확률 비율을 고정(clipping)하는 기법을 사용하여, 학습 과정의 불안정성을 줄이고 샘플 효율성을 향상한다. PPO는 구현이 비교적 간단하며, 다양한 환경에서 안정적으로 적용되어 드론 자율운항과 같은 실시간 제어 문제에 효과적이다. 드론의 고속 비행과 다양한 환경에서의 기동성을 고려할 때,

PPO의 안정적인 정책 갱신 방식은 큰 이점을 제공한다. 이러한 특성 덕분에 PPO는 드론과 같은 복잡한 시스템에서 강화학습을 효과적으로 수행한다.

3. PPO-MPC 알고리즘

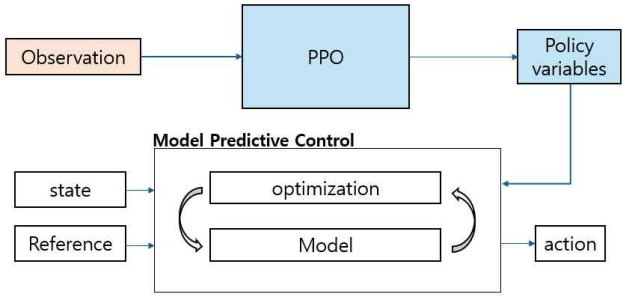

그림 1. PPO를 적용한 모델 예측 제어 구조의 개요

기존의 MPC 알고리즘은 고정된 제어 파라미터를 사용하여 드론의 궤적을 생성한다. 이는 드론이 다양한 환경에서 최적의 성능을 발휘하지 못하게 한다. 본 연구에서는 PPO를 사용하여 MPC의 제어 파라미터를 강화학습으로 최적화하는 PPO-MPC를 제안한다. PPO-MPC는 드론의 현재 상태와 주변 환경에 적응하는 제어 파라미터를 학습함으로써, 기존의 MPC보다 우수한 성능을 발휘한다. PPO-MPC는 PPO의 정책 최적화와 MPC의 제어 방식을 결합하여, 드론의 실시간 제어와 경로 최적화 문제에 대한 접근 방식을 제공한다.

4. 성능 검증 시뮬레이션

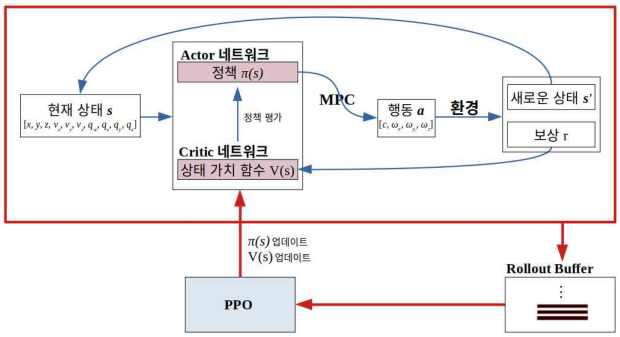

그림 2. PPO를 통한 드론제어 학습 과정

본 연구에서는 PPO-MPC 제어법의 성능을 시험하기 위한 방편으로 드론의 움직이는 게이트 통과 시뮬레이션을 수행하였다.[4] 드론은 현재 상태를 기반으로 배우(actor)-비평가(critic) 네트워크를 통해 행동을 결정하고, 실험 환경에서 드론은 여러 에피소드를 반복하면서 최적의 행동을 선택하였다. 배우 네트워크는 정책을 생성하고, 비평가 네트워크는 상태의 가치를 평가하여 생성된 정책을 기반으로 MPC를 통해 최종 행동을 결정하였다. 이 과정에서 수집된 데이터는 Rollout Buffer에 저장되며, Rollout Buffer의 데이터는 PPO를 통해 정책을 샘플링하고 이를 업데이트하는 데 사용되었다. 정책 업데이트 과정에서는 비평가 네트워크 평가를 통해 이점을 계산하고, 누적된 보상을 정규화하여 손실을 최소화하는 방향으로 모델의 파라미터를 조정하였다. 또한, 일정 간격으로 정책의 파라미터를 저장하여 이후 학습에 활용하였다. 정책 업데이트 과정에서는 정책 클리핑을 통해 안정적인 업데이트를 수행하였다. 게이트는 진자 모형으로 구현되었으며 드론의 제어 입력은 질량 기준 추력 벡터 c와 몸체 프레임에서의 각속도 ω_x, ω_y, ω_z로 구성되었다. 그 결과 PPO-MPC가 기존 MPC 알고리즘에 비해 게이트를 더 잘 통과하는 성능을 보였다. 특히 PPO-MPC는 급격한 게이트 움직임에도 드론의 경로 안정성을 유지하며 게이트를 통과하는 능력이 뛰어났다. 또한 실험은 다양한 궤적을 수행함으로써 신뢰도를 높였다.

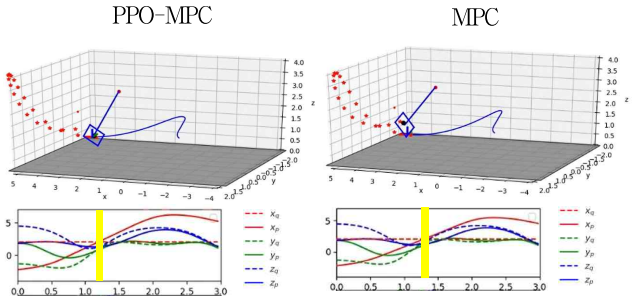

그림 3. PPO-MPC(왼쪽)와 MPC(오른쪽)의 시뮬레이션 결과 비교 PPO-MPC는 게이트가 급격히 변할 때도 드론이 통과하는 확률이 높지만 MPC는 통과하기 어렵다. 상단은 여러 시도 중의 하나를 보여준다. 하단에는 드론(q)과 게이트 중심(p)의 x, y, z 좌표의 시간에 따른 변화를 제시하였다. 형광색 세로선은 드론이 게이트를 통과하는 시점을 나타낸다.

Ⅲ. 결론 및 전망

본 연구에서는 최적 참조 궤적 선택 방법을 제안하고, 이를 MPC 알고리즘과 강화학습을 결합하여 제어하는 방식을 탐구하였다. 제안된 방법은 기존 MPC 알고리즘의 한계를 극복하고, 실시간으로 최적 경로를 생성하여 드론의 비행 성능을 향상할 수 있음을 시뮬레이션을 통해 보였다. 특히, PPO를 활용하여 정책이 급격히 변하는 것을 방지하여 탐험(exploration)과 활용(exploitation)의 균형을 유지하면서도 학습의 안정성을 확보하였다.

뉴로모픽 알고리즘과 하드웨어 개발에는 에너지 고효율성을 위해 신경세포를 모방한 스파이킹 뉴럴 네트워크 (SNN, Spiking Neural Network)의 원리를 널리 활용한다. 그러나 SNN과 강화학습의 결합은 아직 초기 단계에 있으며, 두 전략이 드론 제어에 상승작용을 일으키기 위해서는 본 연구에서 제안하는 PPO-MPC 알고리즘과 상보적인 작용을 할 수 있는 신경과학적 메커니즘의 발굴이 필요하다. 이에 행동 선택에 특화된 도파민 강화 시스템인 두뇌 기저핵의 기능[5]이 중요한 출발점 구실을 할 수 있다. 그림 4에서 제안하듯이 기저핵 원리기반 PPO-MPC 알고리즘의 개발로, 드론이 고효율로 복잡한 환경에서 더 안정적이고 적응력 있는 성능을 발휘할 것으로 기대된다.

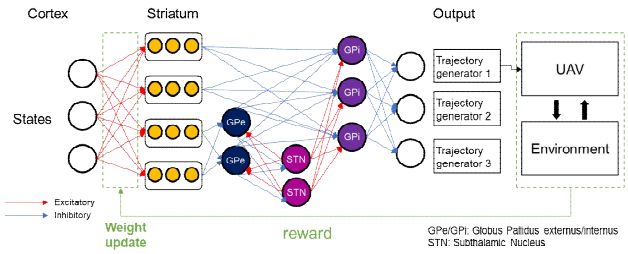

그림 4. 기저핵 모사 강화학습 기반 드론 경로 선택 모델 모식도

ACKNOWLEDGMENT

이 논문은 한국과학기술연구원 미래원천 뇌과학 기술개발사업 (고효율 예측 뇌 기능 모사 알고리즘 개발: 2E32921, 기여율 70%)과 정부(과학기술정보통신부)의 재원으로 한국연구재단 차세대 지능형 반도체 기술개발(소자)사업 (소뇌 모델 기반 자율이동로봇 보정학습 알고리즘 및 비폰노이만 구조 CMOS칩 개발: 2021M3F3A2A01037808, 기여율 30%)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Mayne, D. Q. "Model predictive control: Recent developments and future promise," Automatica, 50(12) pp. 2967-2986, 2014.

[2] Sawant, S., Anand, A. S., Reinhardt, D., and Gros, S. "Learning-based MPC from big data using reinforcement learning," arXiv preprint arXiv:2301.01667, 2023.

[3] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.

[4] Song, Y., and Scaramuzza, D. "Learning high-level policies for model predictive control," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 7629-7636, 2020.

[5] Doya, K. "What are the computations of the cerebellum, the basal ganglia and the cerebral cortex?" Neural Networks, 12(7-8), pp. 960-974, 1999.