TRPO 알고리즘 기반 무인항공기 비행경로계획에 관한 연구

안명기, 이석재, 성길영

LIG넥스원

myeonggi.ahn@lignex1.com, seokjae.lee@lignex1.com, kilyoung.seong@lignex1.com

A Study on Flight Path Planning of UAV based on TRPO Algorithm

Myeong Gi Ahn, Seok Jae Lee, Kil Young Seong

LIG Nex1

요 약

본 논문은 강화학습 기법 중 하나인 TRPO(Trust Region Policy Optimization) 알고리즘을 활용하여 무인항공기의 비행경로계획 수립 시 최적의 비행경로계획을 수립할 수 있는 알고리즘을 제안한다. 무인항공기는 임무 수행 시 사전에 예측하지 못한 장애물들이 나타났을 때 실시간으로 충돌을 회피할 수 있는 경로로 비행을 해야 한다. 이를 위해 TRPO 알고리즘을 적용하여 최적의 비행경로계획 알고리즘을 제안하였다.

Ⅰ. 서 론

무인항공기(UAV : Unmanned Aerial Vehicle)는 우리가 일반적으로 알고 있는 유인항공기와 달리 조종사가 탑승하지 않으며, 지상에서 지상통제시스템(GCS : Ground Control System)을 활용하여 비행경로 계획 등 통제명령에 따라 비행을 수행하는 비행체이다.

주로 감시정찰, 표적탐지 등 군에서 사용해오던 무인항공기는 최근 들어서는 무인항공기의 한 종류인 드론을 활용하여 레저, 방송촬영, 물품 수송 등 우리 생활과 밀접한 관계를 가지고 있다.

하지만 무인항공기는 제한된 연료로 인해 임무를 수행하는 데에 한계가 있다. 이를 해결하기 위해서는 최적의 비행경로계획을 수립하여 효율적인 연료 사용과 임무를 수행해야 한다.[1]

비행경로계획을 수립할 때에는 사전에 식별된 안전지역을 기반으로 비행경로계획을 수립한다. 하지만 무인항공기가 낮은 고도에서의 비행임무 수행 시에는 건물, 산악지형 등과 같은 장애물들을 식별하여 최적의 비행경로를 계산하여 비행해야 한다.

본 논문에서는 무인항공기가 임무수행을 하면서 건물, 산악지형 등과 같은 장애물들과의 충돌을 회피하기 위해 강화학습 알고리즘 중 하나인 TRPO 알고리즘 기반 비행경로계획 기법을 제안한다.

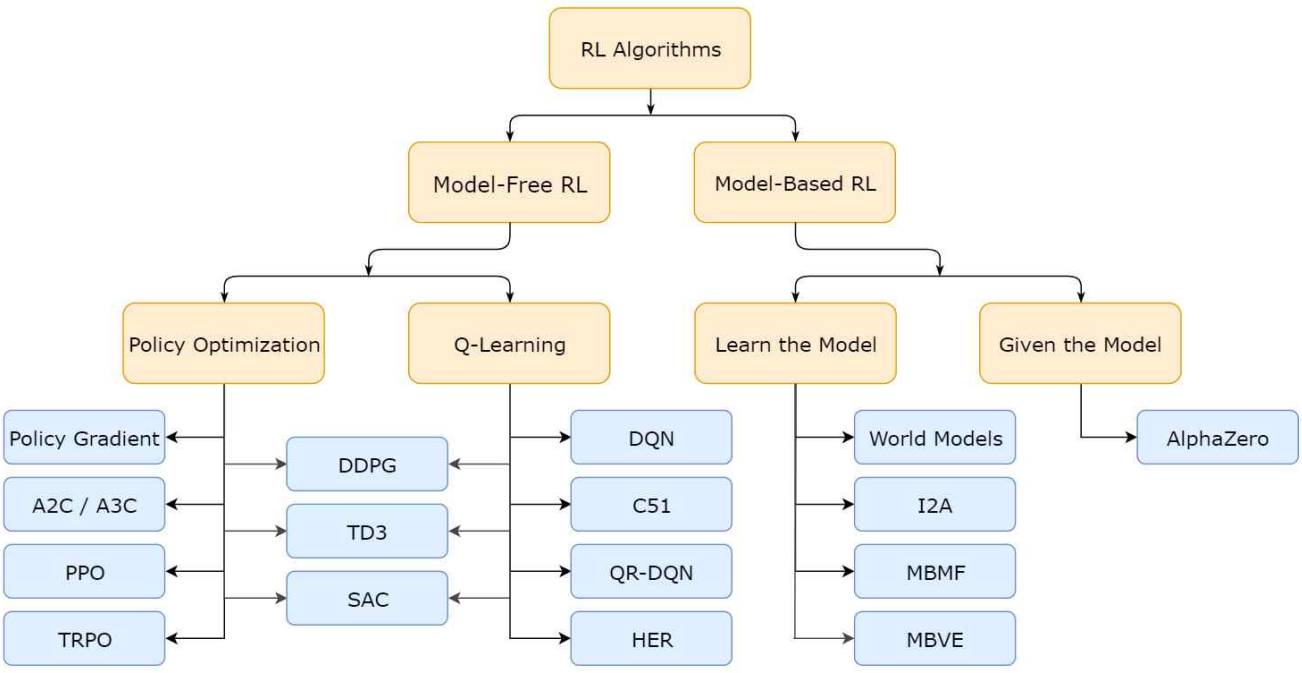

그림 1. 강화학습 분류

Ⅱ. 관련연구

2.1 강화학습

강화학습(Reinforcement Learning)은 머신러닝 분야 중 하나로 주어진 환경과 상호작용하여 최적의 보상을 받을 수 있는 방향으로 진행하는 알고리즘이다. 강화학습은 보통 상태(State), 에이전트(Agent), 행동(Action), 보상(Reward)과 같이 4개의 요소로 구분할 수 있다.

에이전트는 어떤 행동을 취하는 주체를 의미한다. 상태는 현재 에이전트가 어떤 상태인지를 나타낸다. 행동은 현재 상태에서 다음 상태로 취해야 하는 행위를 의미한다. 보상은 에이전트가 현재 상태에서 다음 상태로의 행동을 취했을 때 받을 수 있는 값을 의미한다.

강화학습은 Model의 존재 여부에 따라 그림 1과 같이 크게 Model-Free 강화학습과 Model-Based 강화학습으로 분류할 수 있다.

여기서 Model이란 에이전트가 행동을 취할 때 상태가 어떻게 변하는지에 대해 사전에 예측할 수 있도록 만들어 놓은 프로그램과 같은 것을 의미한다. 하지만 Mode-Based 강화학습의 단점으로는 급격하게 변하는 환경에서의 상태 예측에는 제한적이라는 것이다.

Model-Based 강화학습의 단점을 극복하기 위하여 Model-Free 강화학습 중 TRPO 기법을 활용한다.

무인항공기는 다른 무인항공기 및 갑작스럽게 발생할 수 있는 장애물과의 충돌을 방지하기 위하여 실시간으로 비행경로계획을 수립하여 임무를 수행하여야 한다.

Ⅲ. TRPO 알고리즘 기반 비행경로계획

3.1 TRPO 알고리즘

일반적인 강화학습에서는 현재 상태에서 최대 보상값만을 확인하여 다음 상태로 천이한다. 하지만 이러한 경우에는 과도한 상태 갱신이 발생할 수 있어 리소스 부족 등 다양한 문제가 발생할 수 있다. 이러한 문제점을 보완하기 위해 TRPO 알고리즘을 적용한다.

TRPO(Trust Region Policy Optimization) 알고리즘은 강화학습 기법 중 하나로 에이전트가 다음 상태로 천이할 때 범위 제한(trust region)을 두어 상태 갱신 시 과도한 갱신이 발생되는 문제점을 해결할 수 있는 알고리즘이다.[2]

3.2 TRPO 알고리즘 기반 비행경로계획

무인항공기는 현재 비행하고 있는 위치에서 다양한 방향으로의 이동경로를 따라 비행이 가능하다. 이때 제한된 연료 등으로 인해 최적의 비행경로계획 수립을 위해 TRPO 알고리즘을 적용한다.

강화학습에서 사용되는 용어들을 무인항공기 환경에 적용하면 다음과 같다. 행동을 취하는 에이전트는 무인항공기, 상태는 무인항공기의 현재 위치, 행동은 무인항공기가 다음 경로로 이동해야하는 행위, 보상은 다음 경로로 이동했을 때의 결과 값이다.

무인항공기는 가상의 범위를 제한하고, 현재 위치에서 다음 위치로 이동할 때 최대의 보상을 얻을 수 있는 위치를 탐색한다.

만약에 범위를 제한하지 않는다면 TRPO 알고리즘을 적용했을 때와 달리 과도한 컴퓨팅 리소스 소모가 발생될 것이다.

Ⅳ. 결론 및 향후 연구

본 논문은 강화학습 기법 중 하나인 TRPO 알고리즘을 활용하여 최적의 비행경로계획을 수립하는 방안을 제안했다. TRPO 알고리즘의 핵심인 범위 제한을 두어 무인항공기의 컴퓨팅 리소스 소모를 최소화 할 수 있을 것이다.

참 고 문 헌

[1] S. Aggarwal, and N. Kumar, “Path planning techniques for unmanned aerial vehicles: A review, solutions, and challenges,” Computer Communications, vol. 149, pp. 270-299, 2020.

[2] J. Schulman, S. Levine, P. Abbeel, M. Jordan, abd P. Moritz, “Trust region policy optimization,” In International conference on machine learning, pp. 1889-1897, June, 2015.