무인 항공기 경로 제어를 위한 심층 강화 학습 알고리즘과 실험

문지선, 김효원, 김선우
한양대학교 전자컴퓨터통신공학과
{jiseonmoon, khw870511, remero}@hanyang.ac.kr

Experimental Results of Deep Reinforcement Learning based UAV Trajectory Control

Jiseon Moon, Hyowon Kim, and Sunwoo Kim
Department of Electronics and Computer Engineering, Hanyang University

요 약

본 논문에서는 지상 단말 추적을 위한 심층 강화 학습 기반의 다중 무인 항공기 경로 설정 알고리즘을 제안한다. 무인 항공기는 단말의 위치를 추정하며, 무인 항공기 경로 제어를 통해 단말의 위치 추정 오차 감소를 목표로 한다. 심층 신경망은 무인 항공기의 위치와 지상 단말의 위치 정보를 학습하여, 무인 항공기의 이동 경로를 결정한다. 무인 항공기의 측정 잡음 모델 및 운용 환경을 고려한 단말 위치의 크래머-라오 하한을 통해 심층 강화 학습의 보상 함수를 설계하며, 전체 시스템의 성능과 시스템에 대한 각 무인 항공기의 기여도를 고려하는 보상함수를 통해 각 무인 항공기에 보상값을 부여한다. 실험을 통해 제안하는 심층 강화 학습 알고리즘은 단말 위치의 크래머-라오 하한 및 단말 추정 오차를 감소시키는 것을 확인하였다.

Ⅰ. 서 론

무인 항공기(Unmanned Aerial Vehicle, UAV)는 군사, 운송, 구조 활동 등의 분야에서 활용되고 있다. 무인 항공기의 이동성, 비용 측면에서 효과적인 무선 통신 네트워크를 구성할 수 있으며, 특히 인프라가 존재하지 않는 재난 상황에서 다양한 임무를 수행할 수 있다[1]. 본 논문에서는 심층 강화 학습을 통해 다중 무인 항공기가 지상 단말을 추적하는 알고리즘을 제안한다. 추정치의 성능 지표인 크래머-라오 하한(Cramér-Rao lower bound, CRLB)을 보상함수로 설계하여, 단말 위치 추정 오차를 감소시키는 경로를 설정하며, 기존 방식과의 성능 비교를 통해 제안한 알고리즘을 평가한다[2].

Ⅱ. 단말 추적을 위한 심층 강화 학습 기반의 무인 항공기 경로 제어 기법

시간 k에서 무인 항공기의 상태 벡터는 무인 항공기의 위치 $\mathbf{u}_k=[u_x,u_y,u_z]_k$로 표현되며, 지상에서 움직이는 단말의 상태 벡터는 $\mathbf{x}_k=[\mathbf{c}_k,\dot{\mathbf{c}}_k]^\top$이고, $\mathbf{c}_k=[x_k,y_k,z_k]$는 단말의 위치, $\dot{\mathbf{c}}_k$는 단말의 속도이다. 단말의 움직임을 위해 선형 이동 모델[2]을 적용한다.

$$\mathbf{x}_k=\Phi\,\mathbf{x}_{k-1}+\Gamma\boldsymbol{\nu}_k,\qquad \Phi=\begin{bmatrix}\mathbf{I}_3 & \Delta T\times\mathbf{I}_3\\ 0_3 & \mathbf{I}_3\end{bmatrix},\quad \Gamma=\begin{bmatrix}0.5\times\Delta T^2\times\mathbf{I}_3\\ \Delta T\times\mathbf{I}_3\end{bmatrix} \tag{1}$$

$\boldsymbol{\nu}_k$는 단말의 프로세스 잡음, ΔT는 샘플링 주기, I_3는 3×3 단위 행렬, 0_3은 3×3 영행렬이다. M개의 무인 항공기는 샘플링 주기마다 N개의 단말과의 거리를 측정한다. i번째 무인 항공기가 j번째 단말로부터 수집한 거리 측정치는 다음과 같다.

$$y^{ij}=h(\mathbf{c}^j,\mathbf{u}^i)+w^{ij}, \tag{2}$$

함수 $h(\mathbf{c}^j,\mathbf{u}^i)=\|\mathbf{c}^j-\mathbf{u}^i\|_2$는 무인 항공기와 단말 간의 유클리디안 거리이며, $w^{ij}\sim\lambda^{ij}N(0,\sigma_{LoS}^2)+(1-\lambda^{ij})N(\mu_{NLoS},\sigma_{NLoS}^2)$는 측정 잡음이다. $\lambda^{ij}=1/\{1+\alpha\exp(-\beta(\Theta^{ij}-\alpha))\}$는 무인 항공기 i가 단말 j로부터 LoS를 수신할 확률이고[3], Θ^{ij}는 무인 항공기 i와 단말 j 사이의 고도각이며, α,β는 무인 항공기의 운용 환경에 따른 매개변수이다.

CRLB는 추정치 공분산의 이론적 하한으로서, 추정치의 성능을 나타내는 지표이다[4].

$$\mathrm{var}(\widehat{\mathbf{F}}_k)\geq\mathrm{tr}(\mathbf{J}^{-1}(\mathbf{F}_k)), \tag{4}$$

$\mathbf{F}_k=[\mathbf{c}_k^1,\ldots,\mathbf{c}_k^N]$은 N개 단말의 위치를 포함한 벡터이며, $\mathbf{J}(\mathbf{F}_k)$는 피셔 정보 행렬(Fisher information matrix, FIM)로 다음과 같다.

$$\mathbf{J}(\mathbf{F}_k)=-\mathrm{E}\left\{\frac{\partial^2\ln p(\mathbf{Y}_k|\mathbf{X}_k)}{\partial\mathbf{F}_k^2}\right\},\qquad \mathrm{p}(\mathbf{Y}_k|\mathbf{X}_k)=\prod_{i=1}^{M}\prod_{j=1}^{N}N(y_k^{ij}\,|\,h(\mathbf{c}_k^j,\mathbf{u}_k^i)+\mu^{ij},(\sigma_k^{ij})^2), \tag{5}$$

$\mathbf{Y}_k(i,j)=y^{ij}$, $i\in\{1,\ldots,M\}$, $j\in\{1,\ldots,N\}$는 무인 항공기가 단말로부터 수신한 측정치이며, $\mathbf{X}_k(j)=\mathbf{x}^j$, $j\in\{1,\ldots,N\}$는 단말의 상태이다.

i번째 무인 항공기에 대한 심층 신경망의 입력은 i번째 무인 항공기의 절대 위치 $\mathbf{u}^i$, i번째 무인 항공기를 제외한 무인 항공기 i'와의 상대 위치 $\mathbf{p}^{ii'}$, j번째 단말과의 상대 위치 $\mathbf{q}^{ij}$로 구성된다.

$$\mathbf{s}_k^i=[\mathbf{u}_k^i,\mathbf{p}_k^{i1},\ldots,\mathbf{p}_k^{iM},\mathbf{q}_k^{i1},\ldots,\mathbf{q}_k^{iN}]\in\mathbb{R}^{3(M+N)},\qquad \mathbf{p}^{ii'}=\mathbf{u}^{i'}-\mathbf{u}^i,\quad i,i'=1,\ldots,M,\quad i\neq i',\qquad \mathbf{q}^{ij}=\mathbf{c}^j-\mathbf{u}^i,\quad j=1,\ldots,N, \tag{6}$$

심층 신경망의 출력은 무인 항공기의 이동 방향이며, 무인 항공기가 취할 수 있는 행동은 위, 아래, 앞, 뒤, 좌, 우 6개로 한정된다. 각 무인 항공기는 행동에 대한 보상을 받게 되며, 보상함수는 다음과 같다.

$$R_1^i=R_1+R_2+R_3^i,\qquad R_1=\eta_1\times\left(\frac{\Delta_\psi-\Delta_m}{\Delta_M-\Delta_m}+\tau_1\right),\qquad R_2=\eta_2\times\left(\exp(-\delta\times\psi_{k+1})+\tau_2\right),\qquad R_3^i=\eta_3\times D_i, \tag{7}$$

ψ_k는 시간 k에서의 CRLB이고, $\Delta_\psi=\psi_k-\psi_{k+1}$는 CRLB의 변화량이다. R_1는 CRLB 값의 크기에 따른 보상, R_2는 CRLB의 감소에 따른 보상이며, 두 보상함수는 무인 항공기와 단말이 구성하는 전체 시스템에 대한 보상(global reward)이다. R_3^i은 i번째 무인 항공기가 전체 시스템에 기여하는 정도에 대한 보상(difference reward)이다[5].

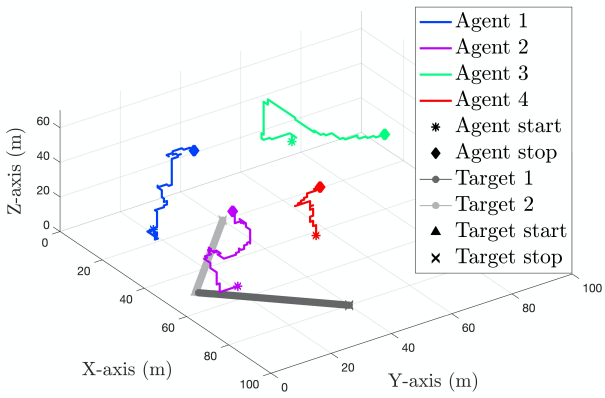

그림 1. 무인 항공기와 단말의 이동 경로

$$j^{\star} = \arg\max_{j\in\{1,\ldots,N\}} |\mathbf{J}_j^{-1} - (\mathbf{J}_j^{-1})_{-i}|,$$
$$D_i = \frac{|\mathbf{J}_{j^{\star}}^{-1} - (\mathbf{J}_{j^{\star}}^{-1})_{-i}|}{(\mathbf{J}_{j^{\star}}^{-1})_{-i}} = 1 - \frac{\mathbf{J}_{j^{\star}}^{-1}}{(\mathbf{J}_{j^{\star}}^{-1})_{-i}}, \quad (8)$$

$\mathbf{J}_j^{-1}$은 j번째 단말과 전체 무인 항공기에 의한 CRLB, $(\mathbf{J}_j^{-1})_{-i}$는 j번째 단말과 i번째 무인 항공기를 제외한 나머지 무인 항공기 의한 CRLB를 뜻한다.

Ⅲ. 실험 결과

실험은 4대의 무인 항공기와 2개의 단말이 있는 상황을 가정한다. 4대의 무인 항공기의 초기 위치는 [30m, 10m, 20m], [70m, 10m, 20m], [10m, 70m, 20m], [50m, 50m, 10m]이며, 단말의 초기 상태 벡터는 [50m, 10m, 0m, 0.3m/s, 0.3m/s, 0m/s], [50m, 10m, 0m, -0.3m/s, 0.3m/s, 0m/s]이다. 실제 무인 항공기 운용 상황에서 일어날 수 있는 고장 상황을 가정하며, 4번째 무인항공기가 시간 50 직후에 고장나게 된다. 그림 1은 심층 강화 학습에서 Q-value의 최댓값에 해당하는 행동만을 취하여 무인 항공기를 조종한 경로이다. 이 경로에 따른 CRLB를 계산한 결과는 그림 2와 같다. 시간에 따라 CRLB가 감소하는 경향을 보이며, 4번째 무인 항공기의 고장 직후인 시간 51에서 CRLB가 상승한다. 그림 3은 2000개의 파티클을 이용하여 100번의 몬테카를로 실험을 한 결과로, 추정한 단말의 위치와 실제 단말 위치 사이의 오차를 나타낸다. 추정 오차는 그림 2의 CRLB와 같은 경향을 보이는 것을 확인할 수 있다. 제안한 알고리즘은 탐욕 알고리즘(greedy algorithm)의 일종인 CRLB-based control과 비교하여 비슷한 수준의 성능을 가지며, 이 실험을 통해 CRLB가 감소함으로써 무인 항공기의 단말 추정 성능이 향상됨을 확인할 수 있다.

표 1. 시뮬레이션 및 심층 강화 학습 모델 파라미터

Parameters	Value	Parameters	Value
layer, node	3, 50	η_1, η_2, η_3	20, 10, 7
학습 횟수	10000	Δ_M, Δ_m	20, -20
learning rate	0.0001	τ_1, τ_2	-0.5
minibatch	128	δ	0.02
discount factor	0.9	σ_{LoS}	0.8m
α	0.7	σ_{NLoS}	10m
β	10	μ_{NLoS}	5m

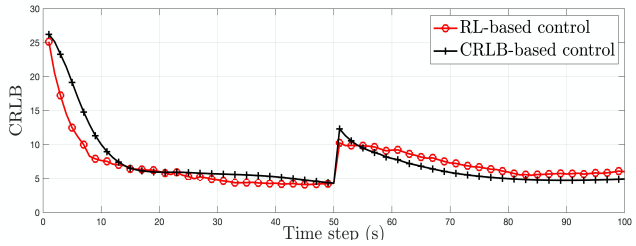

그림 2. 시간에 따른 CRLB 변화

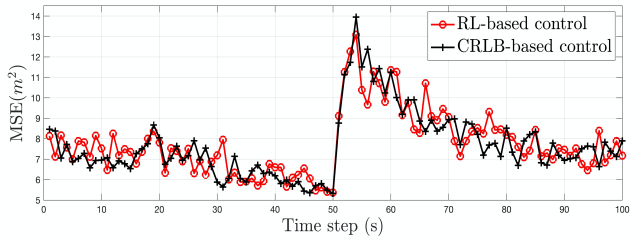

그림 3. 단말의 추정위치 오차

이 실험은 Intel i7-4790K CPU를 탑재한 컴퓨터로 수행되었으며, 하나의 행동을 취하기 위해 제안한 알고리즘은 0.05초, CRLB-based control은 0.86초로 시간적 측면에서 제안한 알고리즘이 효과적이다.

Ⅳ. 결 론

본 논문에서는 지상 단말 추적을 위한 심층 강화 학습을 기반 다중 무인 항공기의 경로 설정 알고리즘을 제안한다. 기존의 CRLB-based control에서 발생하는 연산 시간 문제를 해결하기 위해 심층 강화 학습을 도입하였다. 무인 항공기와 단말의 위치를 학습시켜 무인 항공기의 경로를 제어하며, 추정치의 성능 지표인 CRLB를 활용하여 보상함수를 설계하였다. 각 무인 항공기는 전체 시스템에 대한 CRLB와 시스템 성능에 기여한 정도에 따라 보상을 받는다. 실험 결과 CRLB가 감소하면서 무인 항공기의 단말 추정 오차가 감소하며, 기존의 방식과 비교하여 시간적 측면에서 유리한 것을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(소방청)의 재원으로 정보통신기획평가원(No.2019-0-01325, 재난현장 무선통신 추적기반 요구조자 및 소방관 위치정보시스템 개발), 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터 육성 지원사업(IITP-2020-2017-0-01637)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Y. Zeng *et al.*, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36-42, May 2016.

[2] S. Papaioannou *et al.*, "Coordinated CRLB-based control for tracking multiple first responders in 3D environments," in *Proc. 2020 IEEE Int. Conf. Unmanned Aircra. Syst. (ICUAS)*, Athens, Greece, Sep. 2020.

[3] A. Al-Hourani *et al.*, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp.569-572, Dec. 2014.

[4] Y. Bar-Shalom *et al.*, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[5] A. K. Agogino *et al.*, "Analyzing and visualizing multiagent rewards in dynamic and stochastic domains," *Autonomous Agents and Multi-Agent Systems*, 2008.