강화학습을 이용한 무인항공기 기반 통신-레이더 융합 시스템 최적화

우수연, 김수민, 김준수*
한국공학대학교
jysy2125@tukorea.ac.kr, suminkim@tukorea.ac.kr, *junsukim@tukorea.ac.kr

Optimization of UAV based Joint Communication and Radar System using Reinforcement Learning

Soo Yeon Woo, Su Min Kim, and Junsu Kim*
Tech University of Korea

요 약

본 논문에서는 무인항공기(UAV)를 중계기로 활용하는 Joint Communication and Radar(JCR) 시스템에서 빠른 속도로 변화하는 Air-To-Ground (A2G) 환경에서 통신 및 레이더 시스템의 성능을 동시에 최적화할 수 있는 UAV 배치 문제를 해결하고자 한다. 시스템의 성능 지표는 throughput과 Cramer-Rao lower bound(CRLB)의 가중치 합으로 나타낸 utility function을 사용하였으며, 성능을 최대화하기 위해 Q-learning 기반의 강화학습 기법을 이용하여 레이더 신호를 통해 얻은 UAV의 위치와 레이더 신호의 전송 주기를 제어하고 모델을 학습시켰다. 결과적으로 UAV가 임의의 위치에 배치되어 있을 때, 학습 모델을 통해 시스템의 성능이 최대가 되는 최적의 위치로 UAV의 궤적을 제어할 수 있음을 확인하였다.

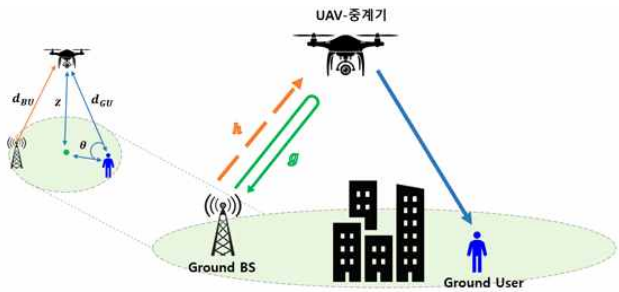
Ⅰ. 서 론

무선 단말기들의 기하급수적인 증가로 인해 과밀화되고 부족한 주파수 자원 문제를 해결하기 위해 통신과 레이더가 동일한 주파수를 공유하는 Joint Communication and Radar(JCR) 시스템에 관한 연구가 활발하게 이루어지고 있다[1]. 또한, 무선통신 시스템에서 Unmanned Aerial Vehicle(UAV)의 사용은 3차원 공간에서 자유로운 이동을 가능하게 하여 전파의 가시거리(Line of Sight, LoS) 채널을 확보할 수 있게 한다. 이를 통해 서비스의 품질이 향상되고 커버리지를 확장할 가능성이 있으며 이를 위해 3차원 공간에서 UAV의 위치적 적절히 제어해야 한다.

본 논문은 UAV를 중계기로 활용하는 JCR 시스템에서 통신과 레이더 시스템의 성능을 동시에 최적화하기 위해 Q-learning 기반의 강화학습 기법을 적용하여 UAV의 위치를 제어한다. Air-To-Ground (A2G) 환경에서 급격하게 바뀌는 환경에 대한 유동적인 채널 모델을 적용하고 UAV의 위치와 파일럿 전송 시간을 최적화함으로써 통신 성능과 레이더 성능이 동시에 향상되는 것을 시뮬레이션을 통해 확인하였다.

Ⅱ. 시스템 모델

그림 1은 본 논문에서 고려하는 시스템 모델이다. 지상의 기지국은 UAV를 중계기로 활용하여 지상의 사용자와 통신을 수행한다. 이때 기지국은 신호의 일부를 레이더로 활용하여 UAV의 위치를 추정한다. 그림 2는 프레임 구조를 나타낸다. 프레임은 레이더 신호 전송을 위한 파일럿과 통신 신호 전송을 위한 데이터 부분으로 구분된다. 파일럿 부분은 UAV의 위치를 추정하는 레이더 기능에 사용되고 데이터 부분은 지상의 기지국이 UAV를 통해 사용자에게 통신 서비스를 제공할 때 데이터의 전송에 사용된다. 각 신호의 전송 시간은 T_p와 T_d이고, 총 전송 시간은 $T = T_p + T_d$로 나타낼 수 있다. 신호의 전체 프레임의 에너지는 PT로 정의되며, 파일럿과 데이터 신호의 전력은 P로 동일하다고 가정한다.



[그림 1] 시스템 모델

[그림 2] Signal frame 구조

JCR 시스템의 성능은 통신 성능과 레이더 성능을 동시에 고려하기 위해 두 지표의 가중치 합을 이용하여 나타내었다. 통신 시스템의 성능 지표는 Throughput을 사용하였고, 레이더 성능 지표는 Cramer-Rao lower bound(CRLB)를 사용하였다. 따라서, JCR 시스템의 성능 지표는 다음과 같은 utility function으로 나타낼 수 있다[2].

$$U = w_c(Thr) - w_r \log_{10}(CRLB). \quad (1)$$

이때, 성능의 가중치는 $w_c = 1/B$, $w_r = 0.5$, $B = 200MHz$로 고정하였다. A2G 환경에서 UAV의 채널은 주변의 환경에 따라서 LoS가 있을 수도 있고, 없을 수도 있다. 따라서 정확한 환경에 대한 정보가 존재하지 않을 때,

LoS와 Non-LoS를 가질 확률이 모두 고려되어야 한다. 먼저, A2G 환경에서 UAV 채널의 대규모 페이딩(Large-scale fading)을 고려한 Path Loss 모델은 다음과 같이 나타낼 수 있다[3].

$$PL(dB) = 20\log_{10}(\frac{4\pi d f_0}{c}) + P_{LoS}\eta_{LoS} + P_{NLoS}\eta_{NLoS}, \quad (2)$$

$$P_{LoS} = \frac{1}{1+\alpha\exp(-\beta(\theta-\alpha))}. \quad (3)$$

식 (2)에서 P_{LoS}는 통신 간에 장애물이 없는 LoS가 보장될 확률이고, P_{NLoS}는 장애물이 있을 확률로 $P_{LoS} = 1 - P_{NLoS}$와 같다. η_{LoS}와 η_{NLoS}는 추가적인 감쇄 수치이다. 식 (3)에서 α와 β는 환경에 따른 상수이다. 다음으로, 소규모 페이딩(small-scale fading) 모델은 Rician K-factor을 UAV의 고도h_{UAV}의 함수로 나타낸 모델을 사용하였다[4].

$$L_{ss}(dB) = 3.53 + 0.65h_{UAV} \quad (4)$$

식 (2)와 식 (4)를 이용하여 UAV 채널의 SNR(신호 대 잡음비)을 구할 수 있고, 이를 통해 Throughput과 CRLB를 나타낼 수 있다. 본 논문에서는 지상의 기지국과 사용자는 고정된 위치에 있고, UAV만 움직이는 상황을 가정하여 성능을 최대화하도록 UAV를 제어하였다. UAV와 지상의 기지국 및 사용자 간의 거리에 따른 SNR을 고려하여 성능 지표를 나타냈으므로, 성능을 최대화하기 위해서는 기지국과 UAV 사이의 거리, UAV와 통신 사용자 사이의 거리를 고려해야 하고, 추가로 파일럿 신호의 전송 시간을 조절함으로써 성능이 향상되는 것을 확인하였다.

Ⅲ. Q-learning을 통한 최적화

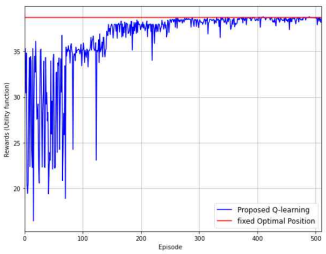
강화학습 중 Q-learning은 Agent가 어떤 State에서 어떠한 Action을 취하고 그로 인한 다음 state에 대한 보상을 받는 과정을 반복하여, 한 Episode가 종료되었을 때의 보상이 최대가 되는 action들을 취하도록 학습시키는 방법이다. 본 논문에서는 A2G 환경에서 JCR 시스템의 성능을 최대화하는 것을 목적으로 하므로, UAV의 위치와 파일럿 신호의 전송 시간 T_p를 조절하여 성능을 확인하고 성능이 최대가 되는 최적의 UAV 위치와 파일럿 신호의 전송 시간을 찾기 위해서 강화학습 환경을 구성하고 Q-learning을 진행하였다.

[표 1] 강화학습 파라미터

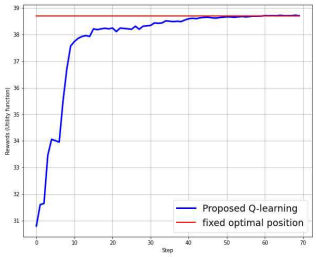
Component	Parameter
Agent	UAV
State	$x[t], y[t], z[t], T_p$
Action	$\pm\Delta x, \pm\Delta y, \pm\Delta z, \pm\Delta T_p$
Reward	$U = w_c(Thr) - w_r\log_{10}(CRLB)$

본 논문에서 설계한 강화학습 환경은 표 1과 같이 Agent, State, Action, Reward로 구성된다. Agent를 UAV로 하고, State는 UAV의 3차원 공간의 좌표(x, y, z)와 pilot 신호의 전송 시간(T_p)이 된다. Action은 UAV의 x, y, z축 공간에서 +/- 움직임과 전송 시간의 증가 및 감소로 총 8가지이다. Episode가 종료되었을 때의 보상은 마지막 step에서의 action과 그로 인한 다음 state에 대한 Reward가 된다. 한 번의 step에서 하나의 action을 취하여 받는 Reward는 식 (1)에서 나타낸 시스템의 성능 지표인 utility function을 사용하였다. Agent가 취할 action을 선택하는 방법으로 decaying epsilon greedy algorithm을 적용하였다. 학습 초반에는 탐험하는 확률변수인 epsilon을 비교적 크게 설정하고, 학습이 진행됨에 따라 점차 epsilon이 줄어들어 학습 후반에는 Q-value가 큰 action을 더 취하도록 설계하였다.

Ⅳ. 시뮬레이션 및 결과



[그림 3] Episode 진행에 따른 보상



[그림 4] 학습된 모델의 성능

그림 3은 본 논문에서 제안한 강화학습 모델의 학습 과정이다. Episode를 총 500번 반복하여 Reward가 최대가 되는 Q-learning 모델을 학습시켰을 때, Episode가 반복될수록 보상이 최적의 값으로 수렴하는 결과를 보여준다. 그림 4는 임의의 위치에 UAV를 배치하고 학습된 모델을 사용하여 UAV를 제어했을 때 결과이다. UAV의 궤적을 최적의 위치로 제어하여 Reward가 최적값에 수렴하게 하는 것을 확인할 수 있다.

Ⅴ. 결론

본 논문에서는 A2G 환경에서 JCR 시스템의 통신 및 레이더 성능을 동시에 최대화하기 위해 강화학습 기반의 Q-learning 기법을 사용하여 UAV의 궤적과 pilot 신호의 전송 시간을 제어하였다. 최적의 위치에 UAV를 고정하였을 때의 Reward와 비교하여 제안된 알고리즘이 최적의 값으로 수렴하는 것을 확인할 수 있었다. 향후 다수의 UAV와 사용자의 통신, 에너지 효율, 보안 등 본 논문에서 고려하지 않은 한계들을 추가로 고려하여 본 논문을 발전시킬 수 있을 것이다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) funded in part by the Korea government (MSIT) (No. 2021R1A2C1013150) and in part by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program (IITP-2023-RS-2022-00156326) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation).

참 고 문 헌

[1] Mazahir, Sana; Ahmed, Sajid; Alouini, Mohamed-Slim (2020): A Survey on Joint Communication-Radar Systems.

[2] J. M. Park, J. Cho, S. Noh and H. Yu, "Optimal Pilot and Data Power Allocation for Joint Communication-Radar Air-to-Ground Networks," IEEE Access, vol. 10, pp.52336-52342, 2022.

[3] ur Rahman, Shams, Geon-Hwan Kim, You-Ze Cho, and Ajmal Khan. "Positioning of UAVs for throughput maximization in software-defined disaster area UAV communication networks." Journal of Communications and Networks 20, no. 5 (2018): 452-463.

[4] X. Ye, X. Cai, X. Yin, J. Rodriguez-Pineiro, L. Tian, and J. Dou, "Air-to-ground big-data-assisted channel modeling based on passive sounding in LTE networks," in Proc.IEEE GC Wkshps, Singapore, Dec.2017.