Sensors Data Collection Scheme based UAV-Trajectory Optimization using Reinforcement Learning

Silvirianti*, Soo Young Shin

Department of IT Convergence Engineering, Kumoh National Institute of Technology

Gumi, South Korea

Abstract

In this paper, data collection of sensors is considered under trajectory optimization of unmanned aerial vehicle (UAV) utilize reinforcement learning. An optimized trajectory is learned to reach the goal point while collecting sensors data as much as possible using reinforcement learning under mentioned conditions. State-action-reward-state-action (SARSA) and Q-learning based UAV trajectory optimization algorithms are utilized to maximize the data collection during finite flight time. The simulation result shows that Q-learning outperformed SARSA and random movement strategy.

Keywords: Trajectory optimization, data collection, UAV, reinforcement learning.

1. Introduction

The applications of UAVs have drawn huge attention from researchers in order to expand the communication coverage and inscription of on-demand connectivity [1]. In general, the communication of UAV to ground devices are dominated by line-of-sight (LoS) channels [2]. Thus, it can be very beneficial for UAV application in data collection from ground devices such as sensors or ground users [3]. In that particular scenario, an optimized trajectory becomes a crucial point for collecting data from sensors due to limited UAV battery energy and a finite flight time. Machine learning particularly reinforcement learning based trajectory design for UAV network has been introduced in [4]. In this paper, an extension work of [5] is proposed by considering larger set of possible actions of UAV that can be taken to accomplish its mission within a finite flight time using reinforcement learning.

2. Proposed Scheme

A UAV is considered communicating with K sensors while flying from initial point L_I to goal point L_F at a constant altitude H meters (m) with speed of w as can be seen in Fig. 1. The sensors are randomly distributed on the environment which denoted by $D = \{D_1, D_2, ..., D_K\}$. The goal of UAV is to collect data from the sensors as much as possible while minimize its energy consumption by optimizing the trajectory. It is assumed that UAV applies time division multiple access (TDMA) for collecting data from sensors. There are B_k bits data stored at $D_k, D_k \in D$ which will be uploaded to UAV.

2.1. Signal Model

The location of UAV at time t can be denoted as $L(t) = (x_t, y_t, H)$. Those locations are projected to the ground plane that can be written as $p_k = (a_k, b_k)$, $l_0 = (x_0, y_0)$, $l_F = (x_F, y_F)$, and $l(t) = |x(t), y(t)|$, respectively. The distance between UAV and sensor D_k can be denoted as

$$d_k(t) = \sqrt{||l(t) - p_k||^2 + H^2}. \quad (1)$$

It is assumed that the communication between UAV and sensors are dominated by LoS link. The channel gain between UAV and sensor D_k is presented by

$$h_k(t) = \alpha_0 d_k(t)^{-}\beta = \frac{\alpha_0}{\left(||l(t) - p_k||^2 + H^2\right)^{\zeta_{pl}}}, \quad (2)$$

*Corresponding author

Email addresses: silvirianti@gmail.com (Silvirianti), wdragon@kumoh.ac.kr (Soo Young Shin)

where α_0 represents the power loss of channel at $d_0 =$ 1m with d_0 being a reference distance, while ζ_{pl} is the pathloss exponent.

2.2. Uploading-data Transmissions

A single sensor can only communicated with UAV at time instant t for uploading its data. Therefore, for each time $t \in [0, T]$ a schedule uploading-data transmission of sensor D_k to UAV can be presented as

$$\rho_k(t) = \begin{cases} 1, & C_k(t) \neq 0 \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

$$\sum_{k=0}^{K} \rho_k(t) \leq 1, \forall t \in (0, T_t], \tag{4}$$

where $\rho_k(t)$ is indicator to schedule transmission while $C_k(t)$ denotes received data rate of UAV at time t. The received rate of UAV is given by

$$C_k(t) = B \log_2 \left(1 + \frac{P_k |h_k(t)|^2}{\sigma^2}\right), \tag{5}$$

$$C_k(t) = B \log_2 \left(1 + \frac{\gamma_0}{[(x(t) - a_k)^2 + (y(t) - b_k)^2 + H^2]^{\zeta_{pl}}}\right), \tag{6}$$

where B denotes the channel bandwidth, P_k represents transmission power of D_k, σ^2 denotes the noise power, and γ_0 represents signal-to-noise-ratio (SNR), respectively. Furthermore, the total amount of collected data can be presented as,

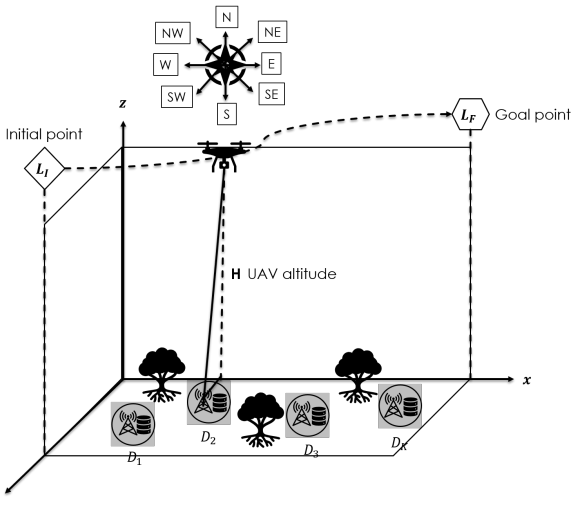

Figure 1. Data collection of IOT sensors using UAV

$$\hat{C}(t)(\{\rho_k(t), q(t)\}, T_t) = \int_0^{T_t} \rho_k(t) C_k(t)\, dt \tag{7}$$

where $\hat{C}(t)$ denotes the total of collected data during the flight under trajectory $q(t)$ and T_t represents required time duration of UAV to collect the data.

This work considers maximization of cumulative collected data in a long-term by designing optimized trajectory of UAV, which can be expressed by

$$\max_{\{(x(t),y(t)) \in q(t)\}} \int_0^{T_t} \rho_k(t) C_k(t)\, dt. \tag{8}$$

3. Reinforcement Learning based Trajectory Optimization

In order to solve the problem formulation which mentioned in Eq.(8), the trajectory optimization of UAV is modelled as Markov Decision Process.

3.1. State Set

The state is modelled by discretizing the flight time of UAV T into M time slots with step size $\delta = \frac{T}{M}$. Thus, the flight area Ψ can be divided into $M_1 = \frac{L_1}{\delta w}$ by $M_2 = \frac{L_2}{\delta w}$ small tiles. Based on the considered area, the state set of UAV can be presented as $S = \{s(1,1), s(1,2), ..., s(M_1, M_2)\}$.

3.2. Action Set

The action set is presented as $A = \{N, NE, E, SE, S, SW, W, NW\}$. For ease clarification, N, NE, E, SE, S, SW, W and NW denotes North, North East, East, South East, South, South West, West, and North West, respectively. The actions represent direction of UAV.

3.3. Reward Formulation

The reward is devised to find optimal solution that satisfy agent on environment in reinforcement learning. The reward is equal to the total amount of collected data as presented by

$$R(t) = \begin{cases} \hat{C}(t), & C_k(t) \geq r_0 \\ & \sum_{t=t_0}^{t} C_k(t) \leq B_k \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where r_0 is the minimum data rate of UAV.

4. Simulation Results

The result in Fig. 2 shows that Q-learning (*blue*) has highest cumulative rewards compared to SARSA (*green*), and random movement (*red*).

5. Conclusion

This paper utilized reinforcement learning for optimizing trajectory of UAV while maximize collected sensors data within a finite flight time. The simulation result shows UAV trajectory under Q-learning received highest rewards compared to SARSA and random movement.

6. Acknowledgement

This work was supported by Priority Research Centers Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology"(2018R1A6A1A03024003).

References

[1] T. 38.811 (2019) Study on new radio (NR) to support non-terrestrial networks (Release 15). [On-line] Available: https://portal.3gpp.org/desktopmodules/SPecifications/SpecificationDetails.aspx?specificationId=3234

[2] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," IEEE Wireless Commun. Lett., vol.3, no.6, pp.569-572, Dec.2014.

[3] X. Lin, V. Yajnanarayana, S.D. Muruganathan, S. Gao, H. Asplund, H. Maattanen, M. Bergstorm, S. Euler, and Y..E. Wang, "The sky is not the limit: LTE for unmanned aerial vehicles," IEEE Commun. Mag., vol. 56, no.4, pp. 204-210, Apr.2018.

[4] J. Cui, Z. Ding, Y. Deng, and A. Nallanathan, "Model-free based automated trajectory optimization for UAVs toward data transmission," in IEEE Proc. of Global Commun. Conf. (GLOBECOM), Dec. 2019.

[5] J. Cui, Z. Ding, Y. Deng, A. Nallanathan and L. Hanzo, "Adaptive UAV-Trajectory Optimization Under Quality of Service Constraints: A Model-Free Solution," in IEEE Access, vol. 8, pp. 112253-112265, 2020.

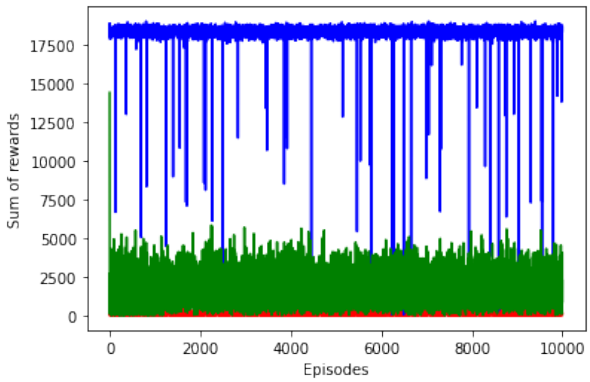

Figure 2. Sum of rewards received by different algorithms during episodes with $\gamma = 0.98$