Reinforcement Learning-based UAVs Trajectory Control to assist VANETs

Md. Mahmudul Islam, Malik Muhammad Saad, Ru Yang, Muhammad Toaha Raza Khan, Junho Seo, Dongkyun Kim
School of Computer Science and Engineering, Kyungpook National University, South Korea
{mislam,maliksaad,yr0818,toaha,junhoseo,dongkyun}@knu.ac.kr

***Abstract*—Intelligent transportation system (ITS) provides an efficient solution to road safety traffic. Enabling ITS requires connectivity among vehicles. However, dynamic vehicular networks cause link disruption among vehicles. To overcome this, unmanned aerial vehicles (UAVs) are deployed to provide connectivity between vehicles that are beyond the communication range. UAVs act as a relay node in providing connectivity among vehicles. We propose multi-agent reinforcement learning to control the trajectory for the deployment of UAVs. Each UAV acts as an agent to find the optimal position, from where it can provide coverage to the maximum number of vehicles. The proposed scheme is expected to increase the packet delivery ratio (PDR) and throughput while reducing the end-to-end delay.**

***Index Terms*—Unmanned Aerial Vehicles(UAVs), Reinforcement Learning (RL), VANET, Trajectory.**

I. Introduction

With the advent of the internet and the rapid growth of technology, the internet of vehicles (IoVs) play a vital role in enabling advance vehicular applications such as collision avoidance, vehicle platooning, remote sensing, and infotainment to increase safety and comfort. These applications come under the intelligent transportation system (ITS). Enabling ITS, require connectivity among vehicles. The vehicle-to-everything (V2X) communication technology is the pivotal technology in ITS. V2X supports communication between vehicle-to-network (V2N), vehicle-to-vehicle (V2V), and vehicle-to-pedestrian (V2P).

The two enabling vehicular technologies are known as dedicated short-range communication (DSRC) and cellular-vehicle-to-everything (C-V2X). IEEE 802.11p/DSRC was the first standard introduced for vehicular communication in 2010 [1], whereas, in 2017, third-generation partnership project (3GPP) introduces C-V2X in its Rel 14. C-V2X is developed over long-term evolution (LTE) device-to-device (D2D) module Rel 12 [2]. Most of the ITS applications rely on the connectivity among vehicles, once the vehicle is beyond the communication range, the link will be broken. For the sake of infotainment application, let suppose a vehicle is downloading content from another vehicle, once the vehicle left the communication range, it will lead to link disruption. To overcome this, researchers are finding alternatives such as unmanned aerial vehicles (UAVs) assistance to extend the communication range between isolated vehicles.

Recent technology advancements made in UAVs can assist the vehicles in downloading the content from the far vehicle via relay link. UAVs equipped with onboard units can act as a relay node to make communication more real-time and improve communication quality. At the same time, in areas where the network is sparse, UAVs can also be used to enhance wireless network coverage, thereby improving the reliability of communication [3]. The most advantage of using UAVs is that of their flexible deployment and scheduling. UAVs being portable relay nodes can be deployed anywhere anytime upon the demand of traffic. So, by cooperating with vehicles, UAVs can improve V2V connectivity, network information collection ability, and network efficiency. However, controlling the trajectories of UAVs efficiently to improve the V2V connectivity between isolated vehicles is a challenging task. In [4], UAVs, i.e., drones are deployed in the urban vehicular ad hoc network (VANET) to enhance the vehicular connectivity and to provide alternative routing paths in congested scenarios. In [5], UAVs are deployed in the uncovered region based on the predicted vehicular traffic.

Reinforcement learning (RL) is considered a powerful tool to solve real-time problems. RL consists of an agent, action state, and environment. Agent observes the current state and takes an action on the environment which in turn gets the reward and transit into a new state. In this paper, we use RL to find the optimal trajectories of UAVs to provide connectivity between isolated vehicles.

The remainder of this paper is as follows. The proposed scheme is discussed in Section II. Finally, Section III concludes the paper.

II. RL based UAVs trajectory control

In our work, the vehicular network is divided into grids and in each grid, a UAV is deployed to provide connectivity between isolated vehicles by working as a relay node. We use Q-learning which is a widely used RL algorithm to find the optimal trajectories for each UAV. The Q-learning agent is deployed inside the UAV and acts interactively with the vehicular environment through observations, actions, and rewards. The agent at each epoch n, observes a state from the state spaces $s_n \in S$, take an action a_n from the action spaces A according to the policy π and in return the agent receives a reward r_n, and translate to the next state s_{n+1}. The state

transition probability is modeled as a Markov decision process (MDP). The transition probability $P(s_{n+1}|s_n, a)$ defines the next state s_{n+1} when the agent receives a reward r_n by taking action a_n in the state s_n. The objective of the agent is to find the optimal policy π^* to move from the current state to the next state which will maximize the cumulative discounted future reward. If the reward at each state is $r(s_n, \pi(a_n))$ which is obtained after selecting the policy π then the cumulative discounted future reward is given by equation (1).

$$\mathrm{R}_n = \sum_{n=1}^{N} \gamma^{n-1} r(s_n, \pi(a_n)) \quad (1)$$

where $\gamma \in [0, 1]$ is the discount factor which defines how much preference should be given to the future rewards.

The vehicles periodically send beacon messages to UAVs which contain their geographical locations, current speeds, and directions. By utilizing the information of the beacon messages, UAVs can construct the distribution of vehicles in the environment and can estimate how long the vehicles will stay inside the coverage range of UAVs. Then according to the policy, the agent will take the appropriate action (new location of the UAV which includes the direction and distance from the current position). Based on the action the agent takes, a reward will be given by the vehicular network immediately. The proposed method for UAVs trajectory control is shown in Fig. 1. The key definitions of RL to control UAVs trajectories are given as follows.

State Space: The state consists of multiple parameters and defined as $S_n = [X_n, V_n, D_n]$. Firstly, the instantaneous position of the vehicles X_n is included in the state, which will represent the vehicular distribution of the network. Secondly, the speed V_n and direction of the vehicles D_n will help the UAVs to estimate how long the vehicles will stay inside the UAVs coverage range which has a direct impact on the decisions of UAVs next optimal locations.

Action Space: At each epoch, the agent executes an action (UAVs next location which includes distance and direction from the current position). The UAV can travel to any new location within the vehicular network or hover in the current position. Note that, the new distance should not exceed the maximum distance that the UAV can travel in a time step, thus, some constraints have to be considered while taking the actions.

Reward: In this paper, our goal is to provide connectivity between the isolated vehicles so that network performance of VANET improves. So, if the PDR is more than 80 percent then the reward will be set to +1, +0.5 will be given if the PDR is less than 80 and greater than or equal to 30, and for less than 30 the reward will be -1.

To find the optimal trajectories for the UAVs we use Q-learning which solves the problem by using the time difference method. In Q-learning, the agents will update the Q table during the training period. The Q table tells the agent which action should be carried out that will maximize the discounted future reward. The simplest form of Q-learning is given by equation (2).

$$\mathrm{Q}(s_n, a_n) \leftarrow Q(s_n, a_n) + \alpha[r_{n+1} + \gamma \max_a Q(s_{n+1}, a) - Q(s_n, a_n)] \quad (2)$$

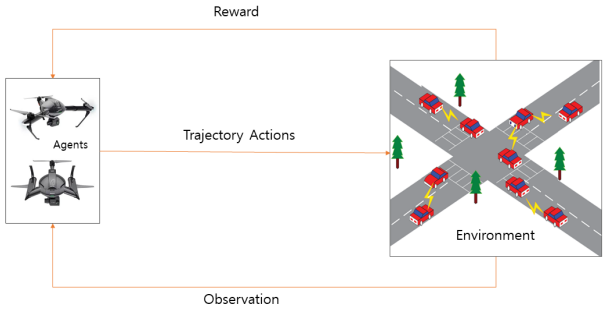

Fig. 1. RL-based proposed method to control UAVs trajectory.

where $\gamma \in [0, 1]$ is the discount factor that sets the weights of the future rewards. $\alpha \in [0, 1]$ is the learning rate that defines the weights of new and old information. $\alpha = 1$ denotes that old information will not be considered while calculating the reward. The objective of the Q-function is to find the optimal policy π that will maximize the discounted cumulative future rewards over time.

III. Conclusions

In this paper, we propose RL based trajectory control mechanism for UAVs to support V2V communication in VANETs. UAVs act as relay nodes to provide connectivity among vehicles that are beyond the communication range. Each UAV acts as an agent that finds the optimal position at each epoch. In the future, we will verify our proposed scheme by observing various performance parameters obtained from simulation and practical results.

Acknowledgement

This study was supported by the BK21 FOUR project (AI-driven Convergence Software Education Research Program) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (4199990214394), and in part by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2016R1D1A3B01015510).

References

[1] J. B. Kenney, "Dedicated short-range communications (dsrc) standards in the united states," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1162–1182, 2011.

[2] R. Molina-Masegosa and J. Gozalvez, "Lte-v for sidelink 5g v2x vehicular communications: A new 5g technology for short-range vehicle-to-everything communications," *IEEE Vehicular Technology Magazine*, vol. 12, no. 4, pp. 30–39, 2017.

[3] W. Shi, H. Zhou, J. Li, W. Xu, N. Zhang, and X. Shen, "Drone assisted vehicular networks: Architecture, challenges and opportunities," *IEEE Network*, vol. 32, no. 3, pp. 130–137, 2018.

[4] O. S. Oubbati, N. Chaib, A. Lakas, P. Lorenz, and A. Rachedi, "Uav-assisted supporting services connectivity in urban vanets," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3944–3951, 2019.

[5] N. Lin, L. Fu, L. Zhao, G. Min, A. Al-Dubai, and H. Gacanin, "A novel multimodal collaborative drone-assisted vanet networking model," *IEEE Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4919–4933, 2020.