Multi-agents Reinforcement Learning for a Type of Platoon Control Problem

April 13, 2020

Abstract

MSC Class:

Key Words: Autonomous vehicle, Platoon control, Reinforcement learning, Stackelberg game, Q-learning, Nash equilibrium

1 Background

With the continuous development and expansion of the world's automobile industry, its position in the world's economic construction is becoming more and more prominent, and automobile industry has gradually became a pillar industry in major automobile producing countries. Safety, energy saving and efficient passage are the eternal themes of automobile industry. In recent years, the continuous increase in vehicle ownership has exacerbated traffic congestion and environmental pollution, and led to an increase in the frequency of traffic accidents. The platoon control of vehicles can achieve a stable queue driving, thereby increasing the vehicle density of the road and improving the traffic efficiency. What's more, it can also enhance the safety of the transportation system and save chemical energy.

The queue control of vehicles is to form adjacent vehicles in a single lane, and automatically adjust the vehicle's motion state according to the information of adjacent vehicles, so that the adjacent vehicles maintain a stable distance and a consistent driving speed. Research on vehicle platoon control began with the PATH project in California in the 1980s. This project talked about many fundamental topics in platoon control, such as the allocation of control tasks, the layout of control architecture, technologies for perception and actuation, and longitudinal/lateral control laws. After that, plenty of attractive topics in the platoon control have been discussed, such as the choice of distance between vehicles, the impact of homogeneity and heterogeneity.

On the other hand, the information flow topology has a very important influence on platoon control. In the former studies, the information flow topology used was relatively single, which focused only on a few common structures, such as predecessor-following topology, predecessor-leader following topology, bidirectional topology and so on. Along with the development of communication technology, the communication between vehicles (Vehicle-to-Vehicle,V2V) is becoming more and more popular. A large number of different kinds of information flow topology structures can be produced in the vehicle platoon, including two-predecessor following topology and multiple-predecessor following topology. But there are some new challenges arise naturally when we consider the variety of topologies, such as time delay, packet loss and quantization error in the communications.

From the view of control, the platoon can be regarded as a dynamic system composed of multiple single vehicle nodes, where the controls can be formed for the individual vehicle through the information interaction between the nodes. Therefore, the platoon of vehicles can be regarded as a special multi-agent system which is a dynamic system formed by multiple agents with independent autonomy through the interaction of certain information topological structures. Under this perspective, a platoon can be decomposed into four interrelated sub-components: Node dynamic(ND), Information flow topology(IFT) , Formation geometry(FG) and Distributed controller(DC). The Node dynamic mainly describes the dynamic behavior of a single vehicle, including the position, velocity and acceleration of the vehicle. Information flow topology is a graph which can model the topological relation of information transfer between individual vehicle nodes. Formation geometry gives us the desired distance between adjacent vehicle. And distributed controller mainly reflects how can each vehicle adjusts its own behavior through the obtained information.

In this paper, we utilize the above framework and modify the models in [1]. As there are many uncertainties in the driving process and signal delay when receiving the information, we add the randomness to the Node dynamic. [1] mainly talks about the control of each vehicle to keep a stable distance and same speed between adjacent vehicles, which is a stability problem. We propose an optimization problem that gives a cost function for each vehicle in the platoon and try to find the Nash equilibrium. To achieve the goal, we adopt reinforcement learning approach and utilize the Stacklberg game theory to simplify the problem.

2 Problem Setup

We consider a platoon running on a flat road with $N+1$ vehicles, including a leading vehicle (LV, indexed by 0) and N following vehicles (FVs, indexed from 1 to N). Motivated by the Node dynamic model in [1], we neglects the inertial delay in powertrain dynamics and assume that the vehicle dynamics is ideal double integrators. But considering the uncertainties in the driving process and signal delay, we add the randomness to the Node dynamic model, for $i = 0, 1, \dots, N$,

$$\begin{cases} dp_i(t) = v_i(t)\,dt + \sigma_1\,dW_{i,t} \\ dv_i(t) = u_i(t)\,dt + \sigma_2\,dB_{i,t} \end{cases} \tag{1}$$

where $p_i(t)$ and $v_i(t)$ denote the position and velocity of vehicle i, σ_1 and σ_2 are two constants, $W_{i,t}$ and $B_{i,t}$ are Brownian motions and the control input $u_i(t)$ for this model is the acceleration of each vehicle. As we can not change the acceleration very sharply, we add a condition for the control input $-k \le u_i(t) \le k$, where k is a constant.

The goal of the platoon control is to maintain the best distance between adjacent vehicles, keep the speed of each following vehicle consistent with that of the leader vehicle, and make the traffic efficiency as high as possible, so for each vehicle in the platoon we can set the cost function as follows:

$$J_i(u) = \mathbb{E}\Big[\int_0^T (|p_i(t) - p_{i-1}(t) - d_{i-1,i}|^2 + |v_i(t) - v_0(t)|^2)\,dt - p_i^2(T)\Big], \tag{2}$$

where $d_{i-1,i}$ is the the desired space between node $i-1$ and node i. Our goal is to obtain the Nash equilibrium $(u_0^*, u_1^*, \dots, u_N^*)$, which can be achieved when

$$J_i(u_i, u_{-i}^*) \ge J_i(u_i^*, u_{-i}^*) \tag{3}$$

holds for any $i = 0, 1, \dots, N$. So, to arrive the Nash equilibrium, the optimal strategy sequence $(u_0^*, u_1^*, \dots, u_N^*)$ should satisfy $N+1$ inequalities. It is difficult for us to verify the condition of Nash equilibrium, thus we want to use the Stackelberg game theory to simplify this problem.