DAI2022 SMARTS Competition Track 1 Technical Report

Jingbo Sun[1,2], Xing Fang[1,2], Zhiyuan Zhang[1,2]

1. The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
2. University of Chinese Academy of Sciences, Beijing, China

Abstract: In the competition, our final solution consists of three parts: state representation, strategy learning and action modification. In the state representation part, we designed a multi-mode representation of network to percept different types of information. In strategy learning part, we designed a multi scene training method based on the PPO algorithm to solve the problem that multi scene training is difficult to converge. In the action revision part, we have corrected unreasonable actions. Our solution finally ranks 6th in track1.

1 Introduction

Deep reinforcement learning (DRL) combines the representation ability of deep learning with the decision-making ability of reinforcement learning, and is widely used in intelligent games, automatic driving and other fields. Therefore, we use the Proximal Policy Optimization(PPO) algorithm to make decisions.

In the task, we first designed multi-mode state observation to facilitate the perception network to better understand the current state. After that, we designed the action space and reward function which accord with the reality. Later, in order to learn the complex road conditions, we designed a perception network that includes multi-layer perceptron, convolutional neural network and Transformer network. At the same time, in order to solve the problem that the model is difficult to train in different scenarios, we designed a scene classification auxiliary task. Finally, we designed a action revision module to regulate the vehicle action.

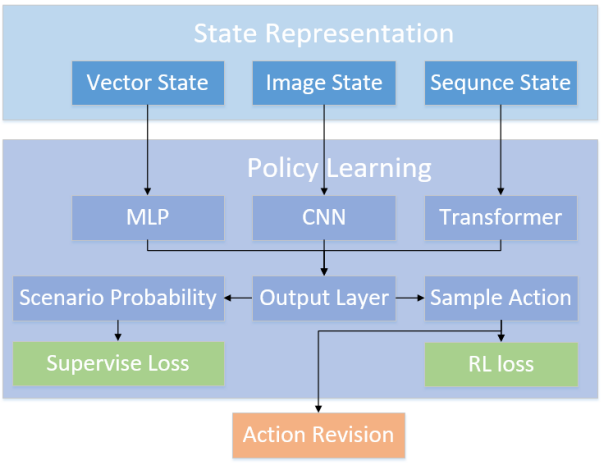

Fig. 1: Figure 1: The overall framework

2 Technical Implementation

2.1 Overall Framework

In this competition, our vehicle is expected to perform different tasks such as left turning at intersections, left turning at T-junctions, cruising tasks, merging tasks, etc. On the one hand, the algorithm needs to be able to distinguish different scenes, on the other hand, it needs to perceive the surrounding road conditions. We expect to design a model that can be applied to different tasks. Therefore, The overall framework (illustrated in Figure 1) of our final solution consists of three parts: multi-modes state representation, strategy learning with auxiliary tasks, and action modification.

2.2 State Representation

In our setting, the state space consists of vector observation, image observation and sequence observation. Among them, vector observation is mainly used to extract waypoints information. Vector observation (68 dimensions) includes the absolute and relative positions and angles (60 dimensions) of 10 waypoints, the positions, angles and speeds (7 dimensions) of ego vehicle, and the distance from the vehicle to the road center. The vector state is input into the multi-layer perceptron network to obtain 256 dimensional features.

Image observation is mainly used to extract road condition information. The image observation is a four channel image obtained by superimposing the aerial view of three channels and the relative position point map of the target of one channel. The image state is input into the convolutional neural network, and 256 dimensional state features are obtained after feature extraction.

Sequence observation is mainly used to extract the information of adjacent vehicles. Sequence observation includes the absolute and relative position, speed and direction (11 * 9 dimensions) of ego vehicle and ten adjacent vehicles. The sequence observations are input into the Transformer network to extract the relationship features between vehicles.

2.3 Action Space

The action of the environment in the game is targetpose, that is, the position and direction of the next moment. Therefore, the action space we choose includes the arc change between 0-0.1 based on the current heading angle and the speed selection between 0-27 m/s.

2.4 Reward Design

Our reward design consists of six parts:

$$r = r_{cen} + r_e + r_d + r_s + r_{env} \tag{1}$$

where r_{cen} is the distance penalty between the vehicle and the lane center; r_e is the empty lane reward; r_d is the distance

penalty from the vehicle in front; r_s is the penalty of high speed; r_{env} is the environment reward.

Rewards are defined as follows:

$$r_{cen} = -abs(d_{cen}) \tag{2}$$

$$r_e = \begin{cases} -0.05 * r_{env} & if\ exit\ empty\ lane \\ 0 & else \end{cases} \tag{3}$$

$$r_d = -0.1 * (d_{safe} - d_{lane}) \tag{4}$$

$$r_s = \begin{cases} -(v - v_l)/v_l & if\ v > v_l \\ 0 & else \end{cases} \tag{5}$$

where d_{safe} is the safe distance, d_{lane} is the distance from the vehicle ahead, v is the speed of the ego vehicle, and vl is the lane limit speed.

2.5 Training Design

Since our learning goal is to learn an effective model in different scenarios, we set multiple parallel environments for training. Each parallel environment is a different scenario. but during the training process, we found that the algorithm is difficult to converge. To solve this problem, we designed an auxiliary task to obtain the classification probability of the scene after extracting features from the Actor network and Critical network, and add the supervision training loss for scene classification. By adding auxiliary tasks, the algorithm can get convergence.

2.6 Action Revision

In order to ensure the security of the form, we have designed an action revision module. If the distance between ego vehicle and the vehicle in front is less than the safe distance, force the brake; if the main vehicle is at the intersection and the speed of the surrounding vehicles is less than 2.7m/s, continue to move forward.

3 Results and Analysis

At present, our method has achieved the sixth place. We think that the current problem of our method is that the collision avoidance function of the intersection is not good enough, and the generalization performance for multiple scenarios needs to be further improved.