文本

描述已自动生成MSC Project Outline

Student name: Yi Yan

Student ID Number.:11451917

Project ID: 295

Supervisor: Dr. Zhongguo Li

Chosen topics

Multi-agent Path Planning with active exploration based on Explorer-Actor-Critic

Introduction

Multi-agent path planning is a key technology to study the coordinated actions of multiple agents in a shared environment to achieve their respective goals. With the development of artificial intelligence and automation technology, multi-agent systems have been widely used in areas such as drone formation, warehouse robot scheduling, traffic management and computer games. The goal of multi-agent path planning is to ensure that all agents can reach their respective destinations efficiently and conflict-free, while maximizing the overall system performance. Multi-agent path planning is a research field full of challenges and opportunities, and its technological development is of great significance in promoting the application of intelligent systems and improving the efficiency of multi-agent working together.

Motivations

Traditional path planning algorithms suffer from dimensionality limitations when encountering large-scale multi-agent path planning, so we would like to solve this problem using deep reinforcement learning.

In reinforcement learning algorithms, uncertainty affects the convergence of the system, so we used a deep ensemble approach to estimate the uncertainty of the system. In the deep ensemble algorithm, we use multiple neural networks with the same structure and different hyper-parameters to combine and train to derive the uncertainty of the system, but how to determine these hyper-parameters is a key problem, so on the basis of deep integration, we introduced hyper-parameter deep integration, which is learned to obtain these hyper-parameters, so as to make the system model more accurate.

Secondly, sampling efficiency makes a key issue in reinforcement learning, in traditional reinforcement learning algorithms, the epsilon-greedy policy is usually used to explore the environment, but this exploration strategy is inefficient, so we aim to develop an uncertainty-driven active exploration reinforcement learning algorithm, we add an explorer network based on the A2C algorithm and introduce action mixing to achieve the balance between exploration and exploitation.

Possible methodologies

First, to achieve active exploration of agent in the environment, state uncertainty needs to be quantified, and in this project, uncertainty quantification is considered to be performed using a deep ensemble approach. This project plans to use deep ensemble for calculating the posterior possibility distribution, and the prior possibility distribution has been already calculated by the actor network, so we can use KL divergence to describe the degree of the uncertainty. In addition to this, how many groups of neural networks to choose and how to choose the hyperparameters of the neural networks to achieve the best training effect is a key issue, so this project introduces the hyperparameter deep ensemble, which determines the hyperparameters of the deep ensemble algorithm through iterative learning.

Secondly, sampling efficiency is a key issue in reinforcement learning algorithms, and effective exploration is the key to solve this problem, so based on the A2C algorithm, this project plans to add a network of explorers, which do not consider global rewards, but only consider performing actions with high uncertainty in order to explore the environment, and in the previous section we quantified the uncertainty using KL divergence, so we can use KL divergence as a target for training the explorer network, we want agents to move in the direction of high uncertainty, so we can define the loss function of the explorer network with,

thus making the explorer perform actions with high uncertainty. Secondly, in terms of agent action policies, we will perform action mixing, defining the policy of the actor as , the policy of the explorer as , and the action mixing policy as,

The p-value can be gradually reduced as the learning time step increases.

Expected outcomes

The project is planned to be developed in python and the model is built using pytorch. The result that achieved by optimized algorithm will be compared with that of the original algorithm to discuss the performance and the expected results can prove the effectiveness of the optimized algorithm.