Task 1:

Problem:

Wait time at traffic intersection

Description:

Decide the duration of traffic lights(red or green) at a 2-way intersection(south to north and west to east) to maximize the number of passing cars(there are sensors to count cars). A time step is the duration of traffic lights and we are deciding change or not change the traffic light in each time step.

States:

The color of traffic lights. The duration of traffic lights. The number of cars in each direction.

Actions:

Change or not change the traffic lights.

Rewards:

Number of passing cars in the next time step - number of cars blocked in the other direction.

Transition:

For example, we have a condition of a state(time step that has duration of x1 seconds) like this.

    From south to north:
        red light
        7 cars
    From west to east:
        green light
        9 cars
Change the light brings us to this state (time step that has duration of x2 seconds).
    From south to north:
        green light
        3 cars
    From west to east:
        red light
        12 cars
Not change the light brings us to this state (time step that has duration of x1+x2 seconds).
    From south to north:
        red light
        12 cars
    From west to east:
        green light
        4 cars

Task 2:

Problem:

Profitable automated stock trading. Because the stock market is complex and dynamic, it is very hard to design a profitable strategy but RL can automatically find a profitable strategy which maximize the investment return.

Project:

Paper:

Hongyang Yang, Xiao-Yang Liu, Shan Zhong, and Anwar Walid. 2020. Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy. In ICAIF '20: ACM International Conference on AI in Finance, Oct. 15–16, 2020, Manhattan, NY. ACM, New York, NY, USA.

Code:

https://github.com/AI4Finance-Foundation/FinRL-Live-Trading

Description:

Their Solution:

Ensemble Deep Reinforcement Learning Trading Strategy

This strategy includes three actor-critic based algorithms: Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C), and Deep Deterministic Policy Gradient (DDPG). It combines the best features of the three algorithms, thereby robustly adjusting to different market conditions. They test their algorithms on the 30 Dow Jones stocks that have adequate liquidity. The performance of the trading agent with different reinforcement learning algorithms is evaluated using Sharpe ratio and compared with both the Dow Jones Industrial Average index and the traditional min-variance portfolio allocation strategy. The proposed deep ensemble strategy is shown to outperform the three individual algorithms and two baselines in terms of the risk-adjusted return measured by the Sharpe ratio.