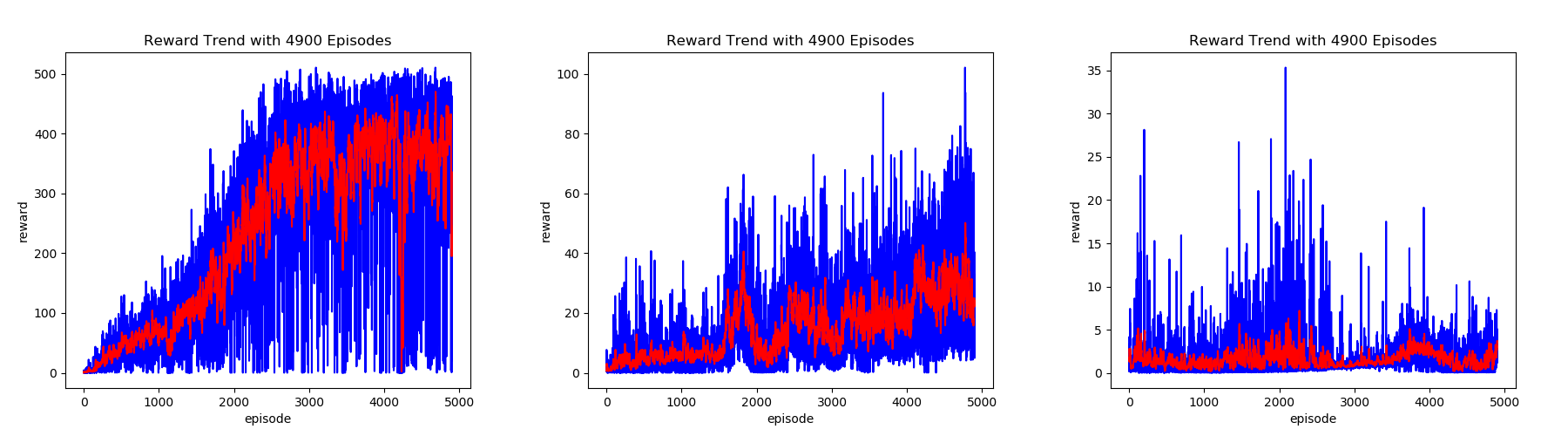
Cube Hyperparameters Optimization

To explore the impact of different hyperparameters, we design and experiment with 10 different pairs of hyperparameters while using the variable-controlling approach. The hyperparameters of the experiments are showed in Table 1. and we can label them into 4 different groups ( (1, 2, 3), (1, 4, 5), (1, 6, 7), (6, 8, 9) ) .

  
Table 1. Hyperparameters of each experiment ( Platform: Qube, Algorithm: DQN)

### Learning Rate

The experiments of number 1, 2, 3 are designed to explore the impact of “Learning Rate” while changing it from 10-3 to 10-5. From Figure 1. , the reward trend of each experiment in this group within 5000 episodes is plotted in blue and red. Obviously, the algorithm with a larger “Learning Rate” could train the model in a higher speed and fast create a model with an acceptable performance but not perfect (a higher reward value means a better performance). However, a larger “Learning Rate” might make the algorithm’s convergence worse since there is a relatively larger fluctuation of the plot on the left.

  
Figure 1. From left to right are the graphs of reward trend of number 1, 2, 3 experiments. The values plotted in blue are the reward value of each episode and the ones in red are the average reward value of the last 10 episodes of that time.