IMPLEMENT REINFORCEMENT LEARNING TO FIND DIRECTION ON A MAP

ABSTRACT

Motivation: To find the optimal solution in the least number of iteration

For comparison: we take three popular algorithm Q-learning, DQN and PPO as major comparison.

Customized environment: we create a scenario of helping Clint to find the optimal path in going home, the detail setting is in the following.

Q-learning

Using off-policy learning using Temporal Difference learning, It is an action-value function to calculate the value for each action at each state

DQN

Using off-policy learning, It can repeatly use the sample data as action and policy is not constantly related.

PPO

Using on-policy learning, It can repeatly use the sample data by doing the important sampling.

SET-UP

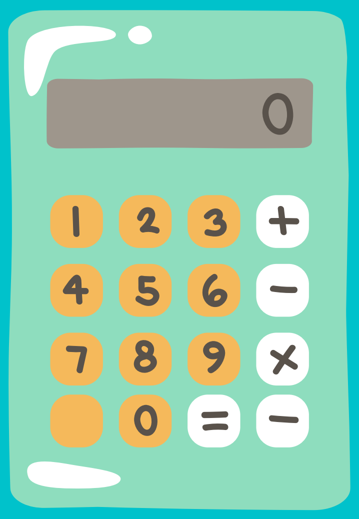
ACTION SPACE

5 discrete actions

- Stay = 0,
- North = 1,
- East = 2,
- South = 3, or
- West = 4

OBSERVATION SPACE

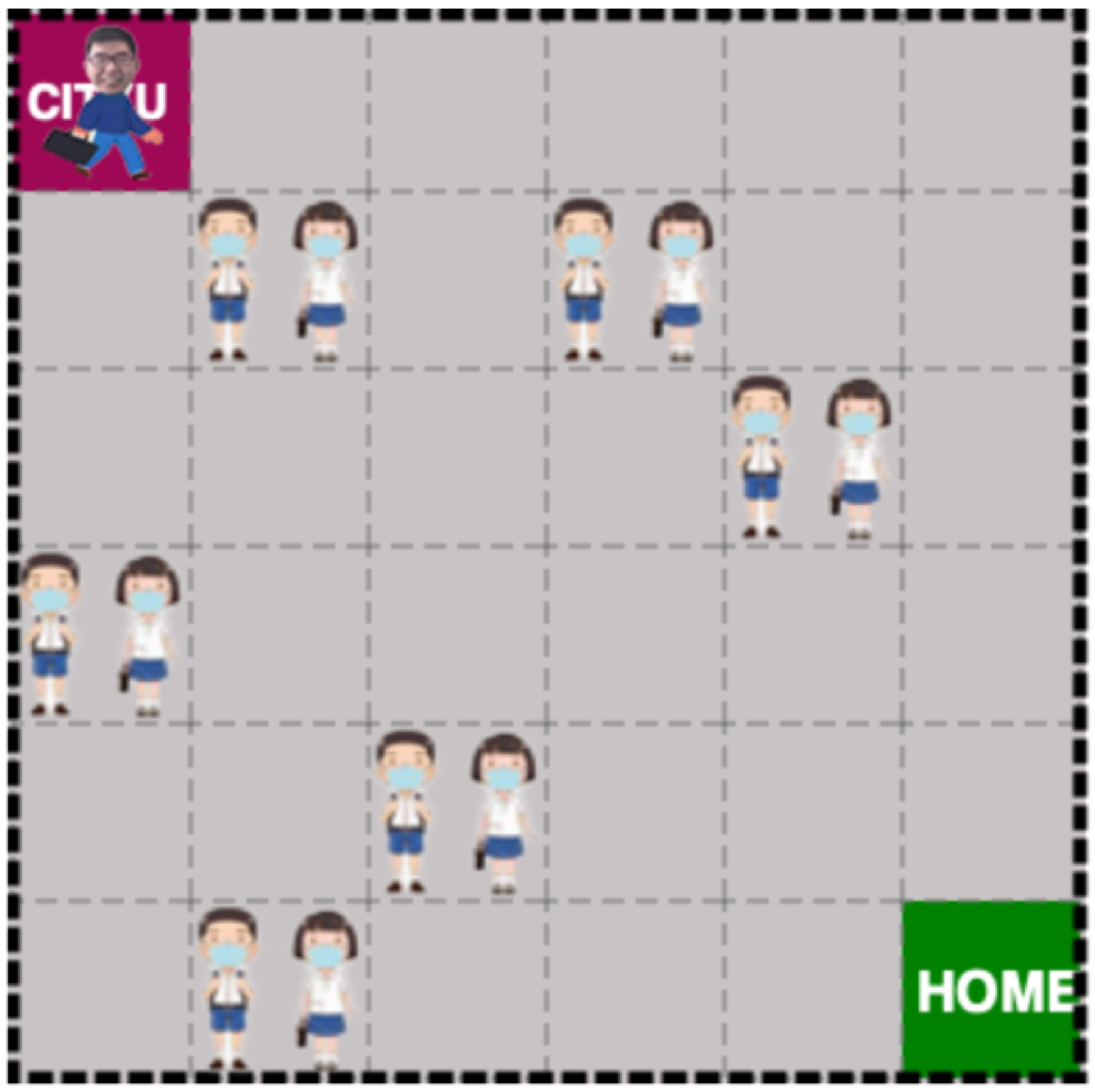
For the 6*6 diagram
The start point is [0,0]
The end point is [5,5]



EPSILSON GREEDY POLICY

Such policies do exploration by trying random action with probability epsilon, and do exploitation with probability 1-epsilon

ENVIRONMENT



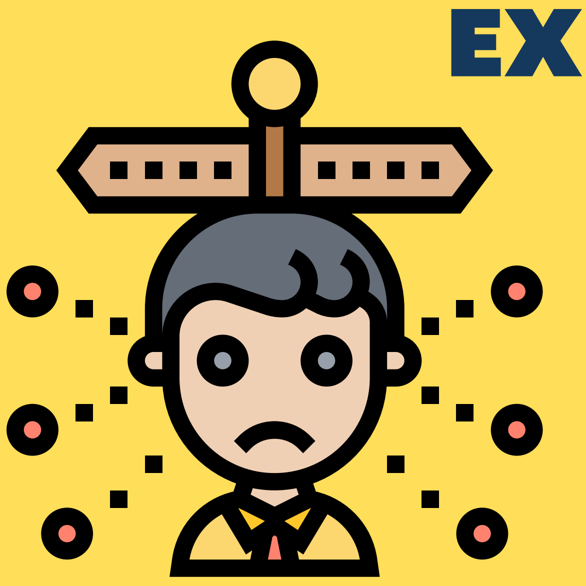
Our agent, the professor Clint, should try to go home from CityU by avoiding all students on his way

EVALUATION

	Q-Learning	DQN	PPO
Mean	-32.665	-6.302	-6.609
Std Dev	28.6892	0.7568	1.3821

DQN algorithm had the best result
Deep Reinforcement algorithms can handle the task better
Models are trained with 1000 episodes

EXPLANATION



DQN uses Convolutional Neural Networks and different tricks such as. Experience Replay, Fixed Q-training to stabilize the learning. Experience Replay helps to learn from one experience several times, making the training more efficient.

Group 2