COMP9444 Project-3 Report

zID:z5127440

name:YIFAN ZHAO

In this project ,we need to run a game named "CartPole" in gym.
Define Network Graph:

```python
# TODO: Define Network Graph
c_name = ['eval_net_params', tf.GraphKeys.GLOBAL_VARIABLES]
w1 = tf.Variable(tf.random_normal(shape=[STATE_DIM, HIDDEN_NODES], seed=1), collections=c_name)
b1 = tf.Variable(tf.random_normal(shape=[1, HIDDEN_NODES], seed=1), collections=c_name)
output_layer1 = tf.nn.relu(tf.matmul(state_in, w1) + b1)

w2 = tf.Variable(tf.random_normal(shape=[HIDDEN_NODES, 1], seed=1), collections=c_name)
b2 = tf.Variable(tf.random_normal(shape=[1, 1], seed=1), collections=c_name)
value = tf.matmul(output_layer1, w2) + b2

w3 = tf.Variable(tf.random_normal(shape=[HIDDEN_NODES, ACTION_DIM], seed=1), collections=c_name)
b3 = tf.Variable(tf.random_normal(shape=[1, ACTION_DIM], seed=1), collections=c_name)
advantage = tf.matmul(output_layer1, w3) + b3

out = tf.nn.relu(value + (advantage - tf.reduce_mean(advantage, axis=1, keep_dims=True)))
```

Network outputs:

```python
# TODO: Network outputs
q_values = out
q_action = tf.reduce_sum(tf.multiply(q_values, action_in), reduction_indices=1)
```

update Q value:

```python
# TO IMPLEMENT: set the target_val to the correct Q value update
target_val = reward_batch[i] + GAMMA * np.max(Q_value_batch[i])
```

HyperParameters:

```python
global GAMMA
GAMMA = 0.9  # discount f
INITIAL_EPSILON = 0.6  # s
FINAL_EPSILON = 0.1  # fir
EPSILON_DECAY_STEPS = 100
HIDDEN_NODES = 20
REPLAY_SIZE = 10000  # ex
BATCH_SIZE = 200  # size
global replay_buffer
replay_buffer = []
```

Result:

```
episode: 100 epsilon: 0.28118600893024853 Evaluation Average Reward: 200.0
episode: 200 epsilon: 0.16631993905469117 Evaluation Average Reward: 200.0
episode: 300 epsilon: 0.12427524256528653 Evaluation Average Reward: 200.0
episode: 400 epsilon: 0.10888552387114742 Evaluation Average Reward: 200.0
episode: 500 epsilon: 0.10325238910599527 Evaluation Average Reward: 200.0
episode: 600 epsilon: 0.10119047959919901 Evaluation Average Reward: 200.0
```