Kuncheng Feng
CSC 466
MM Reading - Chapter 8

1). What is the primary method used by animal trainers?

Reward positive behaviors and ignore negative behaviors.

2). What is meant by the term “operant conditioning?”

A training approach called reinforcement learning, receives reward for acting a certain way in an environment, the reward acts as feedback for the learning agent.

3). TRUE/FALSE - Operant conditioning inspired an important machine-learning approach called reinforcement learning.

True.

4). TRUE/FALSE - Reinforcement learning requires labeled training examples

False.

5). TRUE/FALSE - In reinforcement learning, an agent – the learning program – performs actions in an environment (usually a computer simulation) and occasionally receives rewards from the environment. These intermittent rewards are the only feedback the agent uses for learning.

True

6). TRUE/FALSE - The technique of reinforcement learning is a relatively new addition to the AI toolbox.

False

7). TRUE/FALSE - Reinforcement learning played a central role in the program that learned to beat the best humans at the complex game of Go in 2016.

True

8). In just a few sentences, describe the “illustrative example” that MM used to communicate the basic concepts associated with reinforcement learning, in general, and the variant of reinforcement learning known as Q Learning, in particular.

When the agent receive an reward for acting a certain way in an environment, it will add value to the action that will give rewards in a certain state, for example, the action kick have a value of 10 when the state is 0 steps away from the ball, as the agent learned that this action gives the most return in that state. The rewards of actions are propagated backwards from immediately receiving rewards to the beginning of the task, and only 1 thing is learning per iteration as too much learning can be detrimental.

9). TRUE/FALSE - The promise of reinforcement learning is that the agent can learn flexible strategies on its own simply by performing actions in the world and occasionally receiving rewards (that is, reinforcement) without humans having to manually write rules or directly teach the agent every possible circumstance.

True

10). TRUE/FALSE - In general, the state of an agent in a reinforcement learning situation is the agent's perception of its current situation.

True

11). TRUE/FALSE - A crucial notion in reinforcement learning is that of the value of performing a particular action in a given state.

True

12). In reinforcement learning, what is the value of action A in state S?

The value of action A in state S is a number reflecting the agent's current prediction of how much reward it will eventually obtain if, when in state S, it performs action A.

13). What is the "Q-table" in Q-learning?

A table that keeps track of all the possible states, and in each state actions and values.

14). Why the name "Q-learning?"

This form of reinforcement learning is called Q-learning instead of V-learning because the letter V for value was used for something else in the original paper (for Q-Learning).

15). The Q-learning manifestation of reinforcement learning is a process that iterates over "episodes" until the learning is accomplished. What is an episode in this learning technique?

Each episode refers to the successful task completion for the agent, from starting the task to receiving the reward.

16). List a couple of issues, other than the “exploration versus exploitation balance” issue, that reinforcement-learning researchers face complex tasks.

In the real world, the agent's perception of its states is often uncertain, like how many steps from the ball, or which one is the ball, where is the ball. The effect of performing an action can also be uncertain, it could even endanger the agent performing the task.

17). Deciding how much to explore new actions and how much to exploit (that is, stick with) tried-and-true actions is called the exploration versus exploitation balance. Achieving the right balance is a core issue for making reinforcement learning successful. What real world example does MM use to illustrate the exploration versus exploitation balance?

When we go to a restaurant, do we always order the meal that we have already tried and found to be good, or do we always try something new because the menu might contain something better.

18). MM identifies two “stumbling blocks” to using reinforcement learning in the real world. Please briefly describe each of these stumbling blocks.

The first stumbling block is that in the real world the complexity of tasks might not all fit into a Q-table, for example if we were to use an Q-table for self-driving cars, it’s impossible to define a small set of “states” to fit in a Q-table, as each different image captured by car’s cameras or other values gathered by other sensors all count as a different state.

The second stumbling block is the difficulty for carrying out the learning process in the real world, it is very time consuming for both the agent and the human supervising it, also the agent could put itself in harm's way when exploring states.