Chapter XII

Reinforcement Learning

Introduction

Just as there are many different types of supervised and unsupervised learning, so there are many different types of reinforcement learning. Reinforcement learning is appropriate for an AI or agent which is actively exploring its environment and also actively exploring what actions are best to take in different situations. Reinforcement learning is so-called because, when an AI performs a beneficial action, it receives some reward which reinforces its tendency to perform that beneficial action again. An excellent overview of reinforcement learning (on which this brief chapter is based) is by Sutton and Barto (1998).

There are two main characteristics of reinforcement learning:

1. **Trial-and-error search.** The AI performs actions appropriate to a given situation without being given instructions as to what actions are best. Only subsequently will the AI learn if the actions taken were beneficial or not.

2. **Reward for beneficial actions.** This reward may be delayed because the action, though leading to a reward (the AI wins the game), may not be (and typically is not) awarded an immediate reward.

Thus the AI is assumed to have some goal which in the context of games is liable to be: win the game, drive the car as fast as possible, defeat the aliens, find the best route, and so on. Since the AI has a defined goal, as it plays, it will learn that some actions are more beneficial than others in a specific situation. However this raises the exploitation/exploration dilemma: Should the AI continue to use a particular action in a specific situation or should it try out a new action in the hope of doing even better? Clearly the AI would prefer to use the best action it knows about for responding to a specific situation but it does not know whether this action is actually optimal unless it has tried every possible action when it is in that situation. This dilemma is sometimes solved by using ε-greedy policies which stick with the currently optimal actions with probability 1-ε but investigate an alternative action with probability ε.

Henceforth we will call the situation presented to the AI: the state of the environment. Note that this state includes not only the passive environment itself but also any changes which may be wrought by other agencies (either other AIs or humans) acting upon the environment. This is sometimes described as the environment starts where the direct action of the AI stops, that is, it is everything which the AI cannot directly control. Every state has a value associated with it. This value is a function of the reward which the AI gets from being in that state but also takes into account any future rewards which it may expect to get from its actions in moving from that state to other states which have their own associated rewards. We also create a value function for each action taken in each state.

In the next section, we will formally define the main elements of a reinforcement learning system.

The Main Elements

Formally, we can identify four main elements of a reinforcement learning system (Sutton & Barto, 1998):

1. **A policy:** This determines what action the agent can take in each state. It provides a mapping from the perceived state of the environment to actions which can be taken in that state. It may be deterministic such as a simple look-up table or it may be stochastic and associate probabilities with actions which can be taken in that state.