Author：Liu Jian

Time：2021-07-10

Outline of Tabular Solution Methods

In this part of the book we describe almost all the core ideas of reinforcement learning algorithms in their simplest forms: that in which the state and action spaces are small enough for the approximate value functions to be represented as arrays, or tables. In this case, the methods can often find exact solutions, that is, they can often find exactly the optimal value function and the optimal policy. This contrasts with the approximate methods described in the next part of the book, which only find approximate solutions, but which in return can be applied effectively to much larger problems.

- The first chapter of this part of the book describes solution methods for the special case of the reinforcement learning problem in which there is only a single state, called bandit problems.
- The second chapter describes the general problem formulation that we treat throughout the rest of the book—finite Markov decision processes—and its main ideas including Bellman equations and value functions.
- The next three chapters describe three fundamental classes of methods for solving finite Markov decision problems: dynamic programming, Monte Carlo methods, and temporal-difference learning.
  - Dynamic programming methods are well developed mathematically, but require a complete and accurate model of the environment.
  - Monte Carlo methods don’t require a model and are conceptually simple, but are not well suited for step-by-step incremental computation.
  - Temporal-difference methods require no model and are fully incremental, but are more complex to analyze.

  The methods also differ in several ways with respect to their efficiency and speed of convergence.
- The remaining two chapters describe how these three classes of methods can be combined to obtain the best features of each of them.
  - In one chapter we describe how the strengths of Monte Carlo methods can be combined with the strengths of temporal-difference methods via multi-step bootstrapping methods.
  - In the final chapter of this part of the book we show how temporal-difference learning methods can be combined with
    model learning and planning methods (such as dynamic programming) for a complete and unified solution to the tabular reinforcement learning problem.
