Reinforcement Learning Assignment 2

Yuankun Jiang(TA), Nuowen Kan(TA) Prof. Zou

2019-3-29

1 Introduction

The goal of this assignment is to do experiments with Monte-Carlo(MC) Learning and Temporal-Difference(TD) Learning. MC and TD methods learn directly from episodes of experience without knowledge of MDP model. TD method can learn after every step, while MC method requires a full episode to update value evaluation. Your goal is to implement MC and TD methods and test them in the small gridworld.
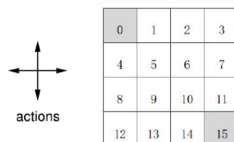
2 Small Gridworld

Figure 1: Gridworld

As shown in Fig.1, each grid in the gridwold represents a certain state. Let s_t denotes the state at grid t. Hence the state space can be denoted as $S = \{s_t | t \in 0, .., 15\}$. S_0 and S_{15} are terminal states, where S_1 to S_{14} are non-terminal states and can move one grid to north, east, south and west. Hence the action space is $A = \{n, e, s, w\}$. Note that actions leading out of the grid leave state unchanged. Each movement get a reward of -1 until the terminal state is reached.

3 Experiment Requirments

- Programming language: python3

- You should implement both first-visit and every-visit MC method and TD(0) and to evaluate state value in small grid world.

4 Report and Submission

- Your reports and source files (.py) should be compressed and named after "studentID+name".

- The files should be emailed to TA (yuankunjiang@126.com) before 2019.04.04 24:00.