Project #2: TD Learning with Eligibility Traces and Planning

I. Overview

The goal of this project assignment is to experiment with eligibility traces and planning as TD learning techniques. The assignment comprises of writing a Matlab-based simulation, the results of which are to be analyzed and explained in detail. The work is to be done in groups of up to 2 students. The deadline for submitting the project report is Nov. 3, 2015.

II. Problem Statement: Gridworld

You are required to construct a 20×20 gridworld, in which there is a fixed starting state (position) and a fixed goal state (to be set at a sufficiently distant location from the starting state). The task is to be considered continuous, whereby each time the agent reaches the goal state it is relocated to a new, random, starting state. All states visited by the agent offer zero rewards, other than the goal state which carries a reward of +1. The gridworld should contain a reasonable number of obstacles, which are in the form of states that cannot be moved into by the agent (see reference in figure below). The agent’s actions are the standard four: up, down, right and left.

(a) You are required to use one eligibility traces learning method to facilitate the agent’s learning process.

(b) For the method used, show the path taken by the agent as well as the change in state values as the algorithm progressed. An example of such results is depicted below. In this example, the values were all initially 0, and all rewards were zero, except for a positive reward at the location marked by ‘*’. The arrows in the other two panels show which action values were strengthened as a result of taking this path. Any method of illustrating such changes can be used.

Figure 1: Examples for the required illustration of the simulation results.
(c) The problem statement provided above is intentionally open-mindedly defined. You are expected to select your own learning scheme and determine the associated parameters. As a result, you are required to derive conclusions pertaining to the performance attributes (accuracy, speed of convergence, general behavior, etc.) for the learning scheme used.

**Grading Policy**

The project report should be coherent and clear. The following is a suggested outline for the report structure:

- Cover page and Abstract
- Introduction and Background
- Design (suggested outline)
  - Design objectives
  - Design challenges
  - Technical approach
  - Experiments and Results
- Summary
  - What have you learned from this project? What have you achieved?
- Appendix A: *Source code listings* - the implementation (this item is judged based on the combination of your design and implementation).

Students are encouraged to discuss the project with each other, however **individual work is to be submitted by each student/pair**.