CSCI 48700: Artificial Intelligence - Programming Assignment 2

Instructor: Dr. Snehasis Mukhopadhyay Due date: November 25, 2025

The objective of this programming assignment is to implement and experiment with two relatively simple reinforcement learning (RL) algorithms, i.e., the L-RI algorithm and the Pursuit Learning Algorithm (PLA) (see course notes on Brightspace). The L-RI is an example of a model-free learning algorithm, while the PLA is a model-based learning algorithm.

You need to simulate a learning automaton with 10 actions operating in a random, but stationary environment. The environment feedback is binary (0 indicates penalty, and 1 indicates reward). The environmental reward probabilities (i.e., probability of generating a feedback of 1 (reward)) when an action is chosen, are [0.19, 0.2, 0.21, 0.59, 0.6, 0.61, 0.72, 0.41, 0..39, 0.4] respectively for the 10 actions.

The RL agent operates in feedback with this environment according to the following algorithm:

1. The action probabilities for the 10 actions are all initialized to a value of 0.1.

2. In each trial (iteration), the agent chooses an action randomly by sampling the action probabilities. That is, the agent calls a Pseudo Random Number (PRN) Generator (a programming language dependent library function such as the function drand48() in C language) to get a PRN x in the range 0 to 1. Then,
If $0 < p_1 + p_2 + \ldots + p_{(i-1)} < x < p_1 + p_2 + \ldots + p_{(i-1)} + p_i$, then the action chosen is the i-th action a_i.

3. In the same iteration, the environment generates a feedback by randomly sampling the reward probability d_i for the chosen action a_i. That is, the environment calls the PRN Generator of get a PRN y in the range (0,1). If $y < d_i$, the environment generates a feedback of 1, else the feedback is 0.

4. In the same iteration, based on the environment feedback, the RL agent updates the action probabilities using the chosen RL algorithm.

5. This process (i.e., the trials or iterations) continues until the action probabilities converge, where convergence is defined as some action probability reaching a value of 0.9. If the corresponding action is the action with the highest reward probability, the convergence is accurate; otherwise, it is inaccurate.

Implement and simulate such an RL agent with each of the L-RI and PLA RL algorithms separately. Run each algorithm independently 100 times with different initialization seeds for the PRN Generator (for C language, the corresponding seed initialization function is srand48()). Measure the percentage accuracy (i.e., the number of times in 100 runs in which the RL agent converged to the correct action), as well as the average learning speed (i.e., the average number of iterations needed for convergence whether accurate or inaccurate).

Measure the accuracy and learning speed for each of the two RL algorithms with different learning rates (step sizes in the respective RL algorithm). You should do that with learning rates of 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, and 0.5.

The deliverables for the assignment are (a) your source code (please do not include your executable file) and (b) a report containing in your own words:
(i) A Statement of the learning problem
(ii) A description of the two RL algorithms
(iii) Two separate plots, one showing (comparing) the accuracy against the learning rate for both L_RI and PLA algorithms, and the other showing (comparing) the average learning speed against the learning rate for the two algorithms.
(iv) Your comments and conclusions.

This is an individual project. Discussions with the instructor/other students are allowed. However, your work must be your own and cannot be copy of others' works.