

Autonomous Agents

Assignment 1: Single Agent Planning

21 September, 2012

By:

Paris Mavromoustakos

Georgios Methenitis

Patrick de Kok

Marios Tzakris

Introduction

The purpose of this first assignment was to implement a reinforcement learning task that satisfies the Markov property, known as Markov Decision Process (MDP). To achieve this we used one prey who is part of the environment and behaves in a known way, and one agent (predator) whose goal is to catch the prey. When that happens the episode ends. The environment we used is a 11 by 11 toroidal grid. Also we assumed that the entire MDP is known to our agent. The agent could thus determine the optimal policy even before interacting with the environment. To program the assignment we used the programming language Java.

1

In the first part, we simulated the environment keeping the policies of the predator and the prey separate. The predator starts from position (0,0) and the prey from (5,5) moving randomly and depending on the given probabilities. The time it takes on average for the predator to catch the prey with random policy is , and the standard deviation is .

2

In order to evaluate the random policy we computed the state-value function V^π and determined the following values for the given states:

$$V^\pi(\text{Predator}(0,0), \text{Prey}(5,5)) =$$

$$V^\pi(\text{Predator}(2,3), \text{Prey}(5,4)) =$$

$$V^\pi(\text{Predator}(2,10), \text{Prey}(10,0)) =$$

$$V^\pi(\text{Predator}(10,10), \text{Prey}(0,0)) =$$

It takes N times to converge.

3

mpla mpla mpla mpla god bless you diedrik

4

To find an optimal policy we computed the optimal value function V for all states which the prey is located at (5,5).

The converge speed (in number of iterations) for different discount factors (γ) are:

For $\gamma = 0.1$, converge speed=
For $\gamma = 0.5$, converge speed=
For $\gamma = 0.7$, converge speed=
For $\gamma = 0.9$, converge speed=

5

To iteratively improve the policy we implemented policy iteration V^π for all states which the prey is located at (5,5).

The converge speed (in number of iterations) for different discount factors (γ) are:

For $\gamma = 0.1$, converge speed=
For $\gamma = 0.5$, converge speed=
For $\gamma = 0.7$, converge speed=
For $\gamma = 0.9$, converge speed=
COMPARE THE RESULTS