



Workshop 10 – Markov Decision Process

Advanced Analytics and Applications [AAA]

Multiple Choice Questions

|| Calculation

||| Programming

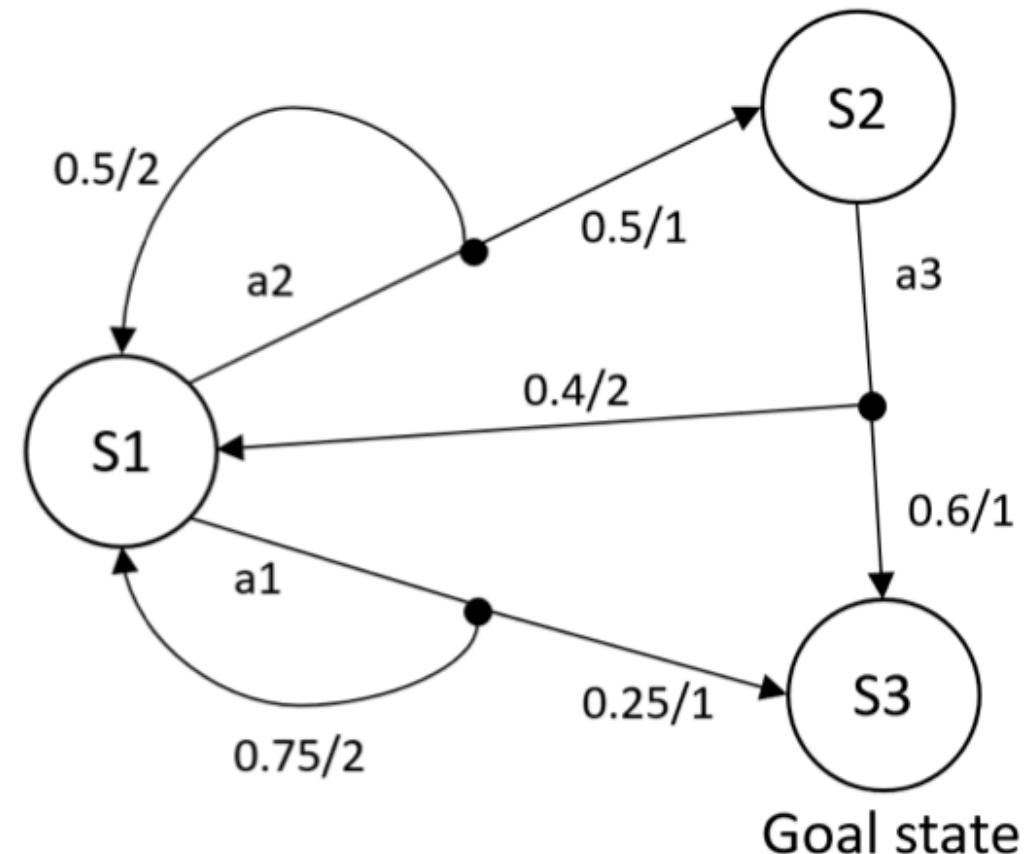
Multiple Choice Questions

||| Calculation

||| Programming

Question 2.1: Markov Decision Process

An MDP with a single goal state (S_3) is given below. a) Given the expected goal distances $c(S_1) = 7$, $c(S_2) = 4.2$, and $c(S_3) = 0$, calculate the optimal policy. b) Suppose that we want to follow a policy where we pick action a_2 in state S_1 and action a_3 in state S_2 . Calculate the expected goal distances of S_1 and S_2 for this policy.



Question 2.2: Monte Carlo Learning

Suppose there's an environment where we have 2 states – A and B. Let's say we observed 2 sample episodes:

$A + 3 \rightarrow A + 2 \rightarrow B - 4 \rightarrow A + 4 \rightarrow B - 3 \rightarrow \text{terminate}$

$B - 2 \rightarrow A + 3 \rightarrow B - 3 \rightarrow \text{terminate}$

$A+3 \Rightarrow A$ indicates a transition from state A to state A, with a reward +3.
Find out the value function using both first visit and every visit methods.

I | Multiple Choice Questions

II | Calculation

III | Programming

Question 2: A Multi-armed Bandit program

- Install Gym AI library.
- Choose an episodic environment.
- Write an algorithm which randomly takes actions and record the final cumulative reward at the end of each episode.
- Apply a Monte Carlo algorithm to find the optimal policies. (this is optional because we have not covered it yet).

Contact



For general questions and enquiries on **research**, **teaching**, **job openings** and new **projects** refer to our website at www.is3.uni-koeln.de



For specific enquiries regarding this course contact us by sending an email to the **IS3 teaching** address at is3-teaching@wiso.uni-koeln.de