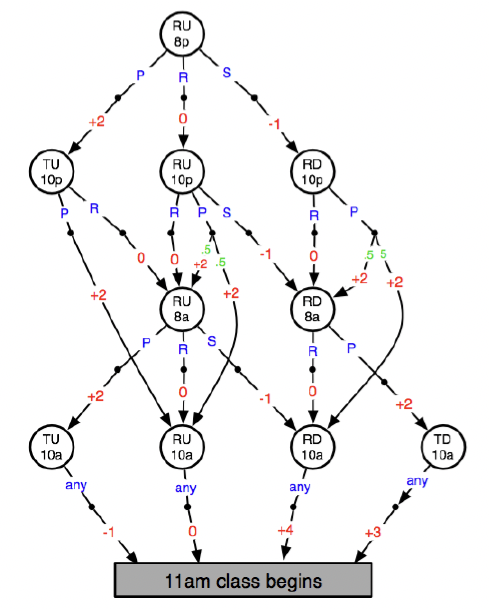
Programming assignment 3

Reinforcement learning – party problem

Yanbing Wang

11/7/2019

The figure below describes a Markov decision process based on life as a student and the decisions one must make to both have a good time and remain in good academic standing.



A.

1. Sequences of experiences from each episode, including the return observed.

episode 1

state RU8p, action R, reward 0

state RU10p, action P, reward 2

state RU8a, action R, reward 2

state RU10a, action R, reward 2

11am class begins

episode 2

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action S, reward 4

11am class begins

episode 3

state RU8p, action R, reward 0

state RU10p, action P, reward 2

state RU8a, action R, reward 2

state RU10a, action S, reward 2

11am class begins

episode 4

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action S, reward 4

11am class begins

episode 5

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action R, reward 4

11am class begins

episode 6

state RU8p, action S, reward -1

state RD10p, action R, reward -1

state RD8a, action R, reward -1

state RD10a, action S, reward 3

11am class begins

episode 7

state RU8p, action R, reward 0

state RU10p, action P, reward 2

state RU8a, action S, reward 1

state RD10a, action R, reward 5

11am class begins

episode 8

state RU8p, action R, reward 0

state RU10p, action S, reward -1

state RD8a, action R, reward -1

state RD10a, action P, reward 3

11am class begins

episode 9

state RU8p, action R, reward 0

state RU10p, action P, reward 2

state RU10a, action R, reward 2

11am class begins

episode 10

state RU8p, action R, reward 0

state RU10p, action R, reward 0

state RU8a, action S, reward -1

state RD10a, action R, reward 3

11am class begins

episode 11

state RU8p, action R, reward 0

state RU10p, action P, reward 2

state RU8a, action P, reward 4

state TU10a, action S, reward 3

11am class begins

episode 12

state RU8p, action R, reward 0

state RU10p, action R, reward 0

state RU8a, action S, reward -1

state RD10a, action R, reward 3

11am class begins

episode 13

state RU8p, action S, reward -1

state RD10p, action P, reward 1

state RD8a, action R, reward 1

state RD10a, action S, reward 5

11am class begins

episode 14

state RU8p, action P, reward 2

state TU10p, action R, reward 2

state RU8a, action R, reward 2

state RU10a, action R, reward 2

11am class begins

episode 15

state RU8p, action S, reward -1

state RD10p, action R, reward -1

state RD8a, action P, reward 1

state TD10a, action R, reward 4

11am class begins

episode 16

state RU8p, action S, reward -1

state RD10p, action R, reward -1

state RD8a, action R, reward -1

state RD10a, action S, reward 3

11am class begins

episode 17

state RU8p, action P, reward 2

state TU10p, action R, reward 2

state RU8a, action R, reward 2

state RU10a, action R, reward 2

11am class begins

episode 18

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action R, reward 4

11am class begins

episode 19

state RU8p, action S, reward -1

state RD10p, action R, reward -1

state RD8a, action R, reward -1

state RD10a, action P, reward 3

11am class begins

episode 20

state RU8p, action S, reward -1

state RD10p, action R, reward -1

state RD8a, action R, reward -1

state RD10a, action R, reward 3

11am class begins

episode 21

state RU8p, action R, reward 0

state RU10p, action S, reward -1

state RD8a, action R, reward -1

state RD10a, action R, reward 3

11am class begins

episode 22

state RU8p, action S, reward -1

state RD10p, action P, reward 1

state RD10a, action S, reward 5

11am class begins

episode 23

state RU8p, action R, reward 0

state RU10p, action S, reward -1

state RD8a, action R, reward -1

state RD10a, action R, reward 3

11am class begins

episode 24

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action S, reward 4

11am class begins

episode 25

state RU8p, action S, reward -1

state RD10p, action P, reward 1

state RD8a, action R, reward 1

state RD10a, action P, reward 5

11am class begins

episode 26

state RU8p, action S, reward -1

state RD10p, action R, reward -1

state RD8a, action P, reward 1

state TD10a, action S, reward 4

11am class begins

episode 27

state RU8p, action S, reward -1

state RD10p, action R, reward -1

state RD8a, action R, reward -1

state RD10a, action S, reward 3

11am class begins

episode 28

state RU8p, action R, reward 0

state RU10p, action P, reward 2

state RU8a, action P, reward 4

state TU10a, action S, reward 3

11am class begins

episode 29

state RU8p, action P, reward 2

state TU10p, action R, reward 2

state RU8a, action S, reward 1

state RD10a, action S, reward 5

11am class begins

episode 30

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action S, reward 4

11am class begins

episode 31

state RU8p, action S, reward -1

state RD10p, action P, reward 1

state RD8a, action P, reward 3

state TD10a, action P, reward 6

11am class begins

episode 32

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action R, reward 4

11am class begins

episode 33

state RU8p, action S, reward -1

state RD10p, action P, reward 1

state RD8a, action P, reward 3

state TD10a, action R, reward 6

11am class begins

episode 34

state RU8p, action R, reward 0

state RU10p, action R, reward 0

state RU8a, action R, reward 0

state RU10a, action S, reward 0

11am class begins

episode 35

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action S, reward 4

11am class begins

episode 36

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action S, reward 4

11am class begins

episode 37

state RU8p, action P, reward 2

state TU10p, action R, reward 2

state RU8a, action S, reward 1

state RD10a, action P, reward 5

11am class begins

episode 38

state RU8p, action P, reward 2

state TU10p, action R, reward 2

state RU8a, action S, reward 1

state RD10a, action R, reward 5

11am class begins

episode 39

state RU8p, action S, reward -1

state RD10p, action P, reward 1

state RD10a, action R, reward 5

11am class begins

episode 40

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action S, reward 4

11am class begins

episode 41

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action S, reward 4

11am class begins

episode 42

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action P, reward 4

11am class begins

episode 43

state RU8p, action R, reward 0

state RU10p, action S, reward -1

state RD8a, action R, reward -1

state RD10a, action S, reward 3

11am class begins

episode 44

state RU8p, action P, reward 2

state TU10p, action P, reward 4

state RU10a, action S, reward 4

11am class begins

episode 45

state RU8p, action R, reward 0

state RU10p, action P, reward 2

state RU10a, action R, reward 2

11am class begins

episode 46

state RU8p, action S, reward -1

state RD10p, action R, reward -1

state RD8a, action P, reward 1

state TD10a, action P, reward 4

11am class begins

episode 47

state RU8p, action R, reward 0

state RU10p, action S, reward -1

state RD8a, action R, reward -1

state RD10a, action S, reward 3

11am class begins

episode 48

state RU8p, action P, reward 2

state TU10p, action R, reward 2

state RU8a, action P, reward 4

state TU10a, action R, reward 3

11am class begins

episode 49

state RU8p, action S, reward -1

state RD10p, action R, reward -1

state RD8a, action R, reward -1

state RD10a, action P, reward 3

11am class begins

episode 50

state RU8p, action R, reward 0

state RU10p, action R, reward 0

state RU8a, action P, reward 2

state TU10a, action S, reward 1

11am class begins

2. The value of each state using the Bellman equation

Iteration 1, delta = 4.0

Iteration 2, delta = 3.5

Iteration 3, delta = 2.625

Iteration 4, delta = 1.5972222222222219

Iteration 5, delta = 0

The value of state RU8p is 3.514

The value of state TU10p is 1.667

The value of state RU10p is 2.500

The value of state RD10p is 5.375

The value of state RU8a is 1.333

The value of state RD8a is 4.500

The value of state TU10a is -1.000

The value of state RU10a is 0.000

The value of state RD10a is 4.000

The value of state TD10a is 3.000

The value of state 11am class begins is 0.000

3. The average return from 50 episodes

Average return for 50 episodes is 3.540.

This value is similar to the value of the start state “RU8P” obtained fromthe Bellman equation, 3.514.

4. The source code and instructions are attached.

B. Policy iteration

1. The policy and value function for each iteration of the algorithm.

While iteration 0

State values: RU8p: 4.0 | TU10p: 2.0 | RU10p: 2.5 | RD10p: 6.5 | RU8a: 1.0 | RD8a: 5.0 | TU10a: -1.0 | RU10a: 0.0 | RD10a: 4.0 | TD10a: 3.0 | 11am class begins: 0.0 |

Policy: RU8p: S | TU10p: R | RU10p: S | RD10p: P | RU8a: S | RD8a: P | TU10a: P | RU10a: P | RD10a: P | TD10a: P |

While iteration 1

State values: RU8p: 5.5 | TU10p: 2.0 | RU10p: 4.0 | RD10p: 6.5 | RU8a: 3.0 | RD8a: 5.0 | TU10a: -1.0 | RU10a: 0.0 | RD10a: 4.0 | TD10a: 3.0 | 11am class begins: 0.0 |

Policy: RU8p: S | TU10p: R | RU10p: S | RD10p: P | RU8a: S | RD8a: P | TU10a: P | RU10a: P | RD10a: P | TD10a: P |

While iteration 2

State values: RU8p: 5.5 | TU10p: 3.0 | RU10p: 4.0 | RD10p: 6.5 | RU8a: 3.0 | RD8a: 5.0 | TU10a: -1.0 | RU10a: 0.0 | RD10a: 4.0 | TD10a: 3.0 | 11am class begins: 0.0 |

Policy: RU8p: S | TU10p: R | RU10p: S | RD10p: P | RU8a: S | RD8a: P | TU10a: P | RU10a: P | RD10a: P | TD10a: P |

2. Source code and instructions are attached.