CS7.301 Machine, Data, and Learning

Assignment 05A: Partially Observable Markov Decision Process

Naren Akash R J

Parameters Table

x	0.79
y	2

POMDP Formulation Table

Number of States	Five
Possible States	S1, S2, S3, S4, S5
Number of Actions	Two
Possible Actions	Left (L), Right (R)
Number of Observations	Two
Possible Observations	Green, Red
Discount Factor	One (1.0)
Action Probability	0.79

Observation State Probability Table

P (Observation = Red | State = Red)	0.85
P (Observation = Green | State = Red)	0.15
P (Observation = Green | State = Green)	0.90
P (Observation = Red | State = Green)	0.10

States Table

States	**S1**	**S2**	**S3**	**S4**	**S5**
Color	Red	Red	Green	Green	Red

Probability Transition Matrix

	S1, L	**S1, R**	**S2, L**	**S2, R**	**S3, L**	**S3, R**	**S4, L**	**S4, R**	**S5, L**	**S5, R**
S1	0.79	0.21	0.79	0.21	0	0	0	0	0	0
S2	0.21	0.79	0	0	0.79	0.21	0	0	0	0
S3	0	0	0.21	0.79	0	0	0.79	0.21	0	0
S4	0	0	0	0	0.21	0.79	0	0	0.79	0.21
S5	0	0	0	0	0	0	0.21	0.79	0.21	0.79

Initial State: One of the red states i.e. S1 S2, or S5

States	**S1**	**S2**	**S3**	**S4**	**S5**
Belief State	0.33	0.33	0	0	0.33

Action 01: Agent took the action Right.

Observation 01: Agent observed Red.

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.33	0.33	0	0	0.33
p = P (S1 | s, R)	0.21	0.21	0	0	0
p * b(s)	0.0693	0.0693	0	0	0

b(S1) := P (Observation = Red | State = Red) * ∑ (p * b(s)) = 0.85 * 0.1386 = 0.11781

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.33	0.33	0	0	0.33
p = P (S2 | s, R)	0.79	0	0.21	0	0
p * b(s)	0.2607	0	0	0	0

b(S2) := P (Observation = Red | State = Red) * ∑ (p * b(s)) = 0.85 * 0.2607 = 0.221595

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.33	0.33	0	0	0.33
p = P (S3 | s, R)	0	0.79	0	0.21	0
p * b(s)	0	0.2607	0	0	0

b(S3) := P (Observation = Red | State = Green) * ∑ (p * b(s)) = 0.10 * 0.2607 = 0.02607

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.33	0.33	0	0	0.33
p = P (S4 | s, R)	0	0	0.79	0	0.21
p * b(s)	0	0	0	0	0.0693

b(S4) := P (Observation = Red | State = Green) * ∑ (p * b(s)) = 0.10 * 0.0693 = 0.00693

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.33	0.33	0	0	0.33
p = P (S5 | s, R)	0	0	0	0.79	0.79
p * b(s)	0	0	0	0	0.2607

b(S5) := P (Observation = Red | State = Red) * ∑ (p * b(s)) = 0.85 * 0.2607 = 0.221595

Current Belief State: After Action 01

States	**S1**	**S2**	**S3**	**S4**	**S5**
b Value Before Normalization	0.11781	0.221595	0.02607	0.00693	0.221595
Belief State	0.198333333	0.373055556	0.043888889	0.011666667	0.373055556

Action 02: Agent took the action Left.

Observation 02: Agent observed Green.

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.198333333	0.373055556	0.043888889	0.011666667	0.373055556
p = P (S1 | s, L)	0.79	0.79	0	0	0
p * b(s)	0.156683333	0.294713889	0	0	0

b(S1) := P (Observation = Green | State = Red) * ∑ (p * b(s)) = 0.15 * 0.451397222 = 0.067709583

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.198333333	0.373055556	0.043888889	0.011666667	0.373055556
p = P (S2 | s, L)	0.21	0	0.79	0	0
p * b(s)	0.04165	0	0.034672222	0	0

b(S2) := P (Observation = Green | State = Red) * ∑ (p * b(s)) = 0.15 * 0.076322222 = 0.011448333

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.198333333	0.373055556	0.043888889	0.011666667	0.373055556
p = P (S3 | s, L)	0	0.21	0	0.79	0
p * b(s)	0	0.078341667	0	0.009216667	0

b(S3) := P (Observation = Green | State = Green) * ∑ (p * b(s)) = 0.90 * 0.087558334 = 0.078802501

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.198333333	0.373055556	0.043888889	0.011666667	0.373055556
p = P (S4 | s, L)	0	0	0.21	0	0.79
p * b(s)	0	0	0.009216667	0	0.294713889

b(S4) := P (Observation = Green | State = Green) * ∑ (p * b(s)) = 0.90 * 0.303930556 = 0.2735375

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.198333333	0.373055556	0.043888889	0.011666667	0.373055556
p = P (S5 | s, L)	0	0	0	0.21	0.21
p * b(s)	0	0	0	0.00245	0.078341667

b(S5) := P (Observation = Green | State = Red) * ∑ (p * b(s)) = 0.15 * 0.080791667 = 0.01211875

Current Belief State: After Action 02

States	**S1**	**S2**	**S3**	**S4**	**S5**
b Value Before Normalization	0.067709583	0.011448333	0.078802501	0.2735375	0.01211875
Belief State	0.152630837	0.025806815	0.177636474	0.616607807	0.027318067

Action 03: Agent took the action Left.

Observation 03: Agent observed Green.

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.152630837	0.025806815	0.177636474	0.616607807	0.027318067
p = P (S1 | s, L)	0.79	0.79	0	0	0
p * b(s)	0.120578361	0.020387384	0	0	0

b(S1) := P (Observation = Green | State = Red) * ∑ (p * b(s)) = 0.15 * 0.140965745 = 0.021144862

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.152630837	0.025806815	0.177636474	0.616607807	0.027318067
p = P (S2 | s, L)	0.21	0	0.79	0	0
p * b(s)	0.032052476	0	0.140332814	0	0

b(S2) := P (Observation = Green | State = Red) * ∑ (p * b(s)) = 0.15 * 0.17238529 = 0.021144862

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.152630837	0.025806815	0.177636474	0.616607807	0.027318067
p = P (S3 | s, L)	0	0.21	0	0.79	0
p * b(s)	0	0.005419431	0	0.487120167	0

b(S3) := P (Observation = Green | State = Green) * ∑ (p * b(s)) = 0.90 * 0.492539598 = 0.443285638

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.152630837	0.025806815	0.177636474	0.616607807	0.027318067
p = P (S4 | s, L)	0	0	0.21	0	0.79
p * b(s)	0	0	0.03730366	0	0.021581273

b(S4) := P (Observation = Green | State = Green) * ∑ (p * b(s)) = 0.90 * 0.058884933 = 0.05299644

s	**S1**	**S2**	**S3**	**S4**	**S5**
b(s)	0.152630837	0.025806815	0.177636474	0.616607807	0.027318067
p = P (S5 | s, L)	0	0	0	0.21	0.21
p * b(s)	0	0	0	0.129487639	0.005736794

b(S5) := P (Observation = Green | State = Red) * ∑ (p * b(s)) = 0.15 * 0.135224433 = 0.020283665

Current Belief State: After Action 03

States	**S1**	**S2**	**S3**	**S4**	**S5**
b Value Before Normalization	0.021144862	0.021144862	0.443285638	0.05299644	0.020283665
Belief State	0.037836011	0.037836011	0.793202652	0.094830315	0.03629501

Reference(s)

CPS 570 Partially Observable Markov Decision Processes by Prof. Ron Parr; Duke University

http://db.cs.duke.edu/courses/compsci590.2/fall17/POMDPs.pdf