

The Value Iteration Algorithm

Alice Gao

Lecture 22

Readings: R & N 17.1.2, 17.2

Based on work by K. Leyton-Brown, K. Larson, and P. van Beek

Outline

Learning Goals

Revisiting the Learning goals

Learning Goals

By the end of the lecture, you should be able to

- ▶ Trace the execution of and implement the value iteration algorithm for solving a Markov Decision Process.

CQ: Determine optimal action given true utility

CQ: What is the optimal action for state s_{13} ?

(A) Up (B) Down (C) Left (D) Right

$$U(s, a) = \sum_{s'} P(s'|s, a) U(s')$$

$$\pi^*(s) = \arg \max_a U(s, a).$$

The true utilities $U(s)$ are given below.

	1	2	3	4
1	0.705	0.655	0.611	0.388
2	0.762	X	0.660	-1
3	0.812	0.868	0.918	+1

CQ: Solve the Bellman equations efficiently

CQ: Can we solve the system of Bellman equations efficiently?

(A) Yes

(B) No

(C) I don't know

The Bellman equation for $U(s_{11})$:

$$U(s_{11}) = -0.04 + \gamma \max \begin{bmatrix} 0.8U(s_{12}) + 0.1U(s_{21}) + 0.1U(s_{11}), \\ 0.9U(s_{11}) + 0.1U(s_{12}), \\ 0.9U(s_{11}) + 0.1U(s_{21}), \\ 0.8U(s_{21}) + 0.1U(s_{12}) + 0.1U(s_{11}) \end{bmatrix}.$$

CQ: Value iteration

CQ: What is $U_1(s_{23})$?

- (A) $(-\infty, 0)$ (B) $[0, 0.25)$ (C) $[0.25, 0.5)$
(D) $[0.5, 0.75)$ (E) $[0.75, 1]$

$U_0(s)$:

	1	2	3	4
1	0	0	0	0
2	0	X	0	-1
3	0	0	0	+1

CQ: Value iteration

CQ: What is $U_1(s_{33})$?

- (A) 0.26 (B) 0.36 (C) 0.46
(D) 0.56 (E) 0.76

$U_0(s)$:

	1	2	3	4
1	0	0	0	0
2	0	X	0	-1
3	0	0	0	+1

CQ: Value iteration

CQ: What is $U_2(s_{33})$?

(A) 0.822

(B) 0.832

(C) 0.842

(D) 0.852

(E) 0.862

CQ: Value iteration

CQ: What is $U_2(s_{23})$?

(A) 0.464

(B) 0.466

(C) 0.468

(D) 0.470

(E) 0.472

CQ: Value iteration

CQ: What is $U_2(s_{32})$?

(A) 0.16

(B) 0.36

(C) 0.56

(D) 0.76

(E) 0.96

Revisiting the Learning Goals

By the end of the lecture, you should be able to

- ▶ Trace the execution of and implement the value iteration algorithm for solving a Markov Decision Process.