I007 Osnove umjetne inteligencije

Tema: Planiranje uz nepouzdanost - vježbe

28. 4. 2021.

predavač: Darija Marković

asistent: Darija Marković

Planiranje uz nepouzdanost

- koristimo formule:

$$V^*(s) = \max_a Q^*(s,a)$$

$$Q^*(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$$

$$V^*(s) = \max_a \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$$

pri čemu su:

- $V^*(s)$ - očekivana dobit ukoliko se počinje u stanju s i djeluje optimalno
- $Q^*(s, a)$ - očekivana dobit ukoliko se u stanju s napravi akcija a i nakon toga djeluje optimalno
- $T(s, a, s')$ - funkcija prijelaza; sadrži vjerojatnost da a iz s vodi u s'
- $R(s, a, s')$ - funkcija nagrade
- $\gamma \in \langle 0, 1]$ - faktor umanjenja

- cilj planiranja: doći do optimalne strategije π^*, tj. do optimalne akcije u stanju s

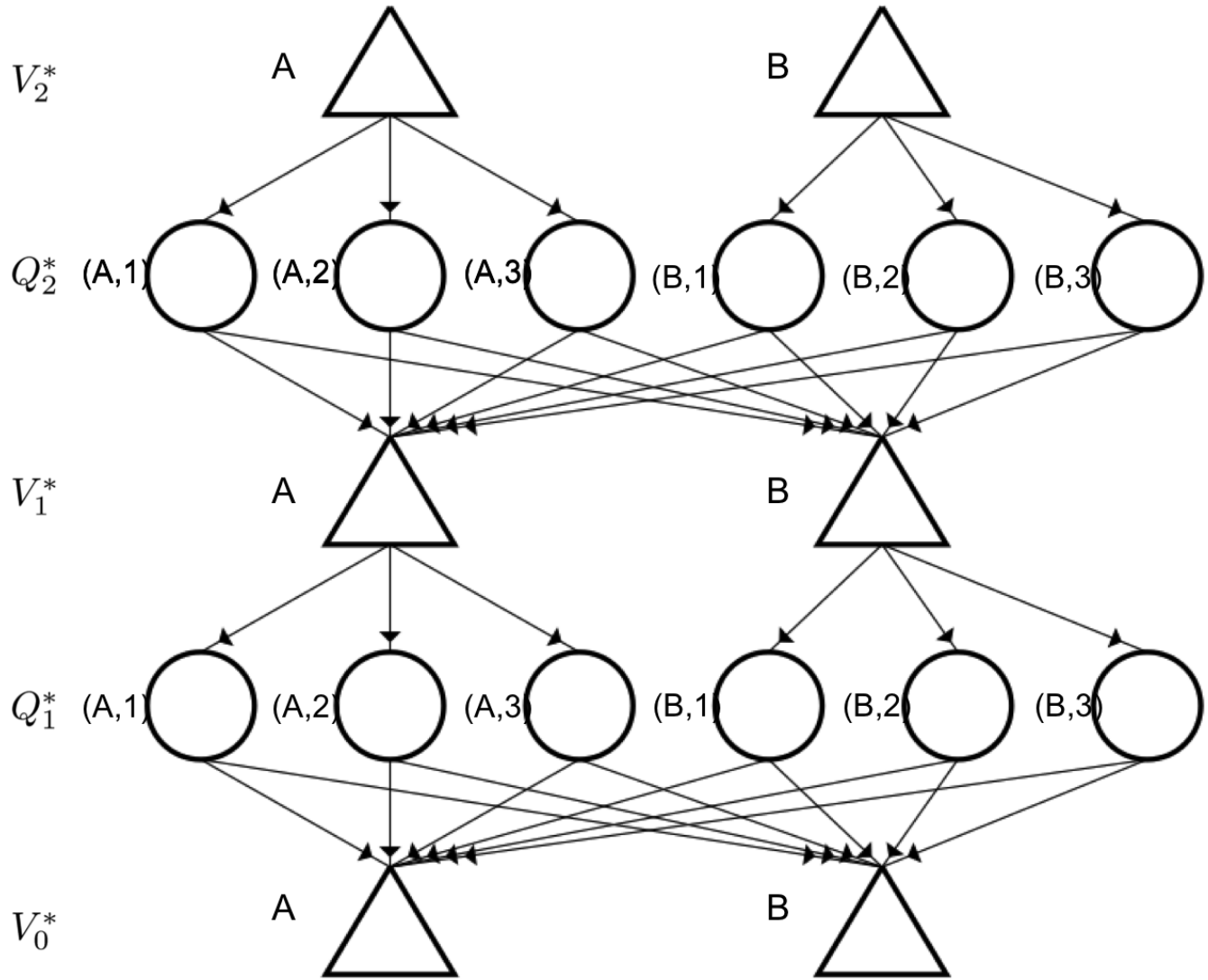
- strategiju izvodimo iz q-vrijednosti: $\pi^*(s) = \arg\max_a Q^*(s, a)$

Zadatak 1.

Pretpostavite da je zadan Markovljev proces odlučivanja s funkcijom prijelaza i funkcijom nagrade prikazanim u sljedećim tablicama:

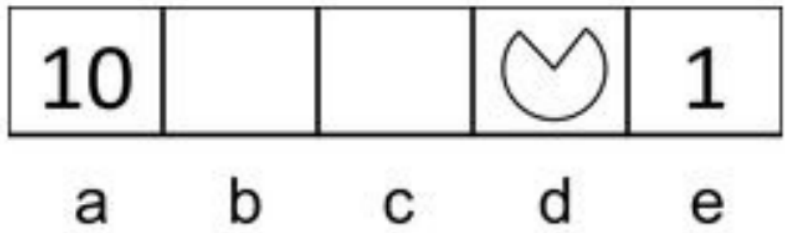
s	a	s'	T(s,a,s')	R(s,a,s')	s	a	s'	T(s,a,s')	R(s,a,s')
A	1	A	0	0	B	1	A	0.5	10
A	1	B	1	0	B	1	B	0.5	0
A	2	A	1	1	B	2	A	1	0
A	2	B	0	0	B	2	B	0	0
A	3	A	0.5	0	B	3	A	0.5	2
A	3	B	0.5	0	B	3	B	0.5	4

Pretpostavite da je faktor umanjenja $\gamma = 1$, tj. nema umanjenja.
Izračunajte vrijednosti V_0^*, V_1^*, V_2^*, Q_1^*, Q_2^* i ucrtajte ih na predloženi graf.
Neka je $\pi_i^*(s)$ optimalna strategija u stanju s ako igra traje i koraka.
Prikažite tablično funkciju $\pi_i^*(\cdot)$.


V_2^*
A
B
Q_2^*
(A,1)
(A,2)
(A,3)
(B,1)
(B,2)
(B,3)
V_1^*
A
B
Q_1^*
(A,1)
(A,2)
(A,3)
(B,1)
(B,2)
(B,3)
V_0^*
A
B

Zadatak 2.

Razmotrimo sljedeću mrežu:



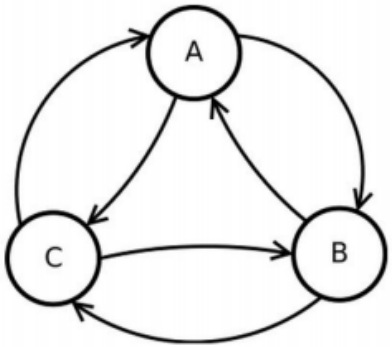
Na raspolaganju imamo akcije *lijevo* $(\leftarrow)$ i *desno* $(\rightarrow)$ koje su 100% uspješne. Dodatno u polju a imamo na raspolaganju akciju *izlaz (exit)* koja je također uvijek uspješna i donosi nagradu 10. Analogno u polju e imamo na raspolaganju akciju *izlaz (exit)* koja je također uvijek uspješna i donosi nagradu 1.

(a) Uz faktor umanjenja $\gamma = 1$, odredite sljedeće vrijednosti: $V_0(d)$, $V_1(d)$, $V_2(d)$, $V_3(d)$, $V_4(d)$ i $V_5(d)$.

(b) Uz faktor umanjenja $\gamma = 0.9$ za istu mrežu, odredite sljedeće vrijednosti: $V^*(a)$, $V^*(b)$, $V^*(c)$, $V^*(d)$ i $V^*(e)$.

Zadatak 3.

Razmotrimo sljedeći dijagram prijelaza, funkciju prijelaza i funkciju nagrade za MPO. Faktor umanjenja je $\gamma = 0.5$.



s	a	s'	$T(s,a,s')$	$R(s,a,s')$
A	$-$	B	0.6	2
A	$-$	C	0.4	2
A	$+$	C	1	1
B	$-$	A	0.2	-2
B	$-$	C	0.8	-2
B	$+$	A	0.8	1
B	$+$	C	0.2	1
C	$-$	A	0.6	2
C	$-$	B	0.4	0
C	$+$	A	0.4	2
C	$+$	B	0.6	0

(a) Pretpostavimo da nakon k iteracija imamo sljedeće vrijednosti za V_k:

$V_k(A)$	$V_k(B)$	$V_k(C)$
2.540	1.920	2.000

Odredite $V_{k+1}(A)$, $V_{k+1}(B)$ i $V_{k+1}(C)$.

(b) Pretpostavimo da nakon konvergencije dobijemo sljedeće vrijednosti:

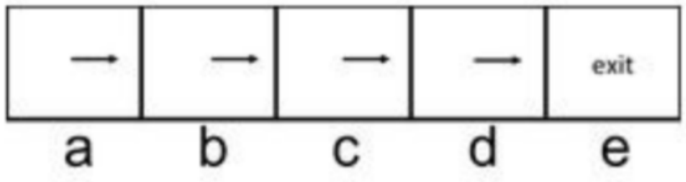
$V^*(A)$	$V^*(B)$	$V^*(C)$
3.324	2.601	2.717

Izračunajte $Q^*(C, +)$ i $Q^*(C, -)$. Koja je optimalna akcija u stanju C?

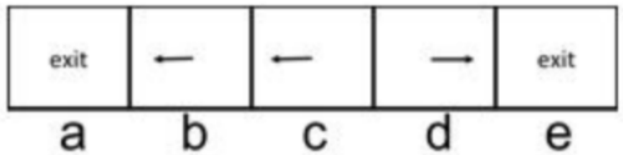
Zadatak 4.

Razmotrimo mrežu iz zadatka 2.

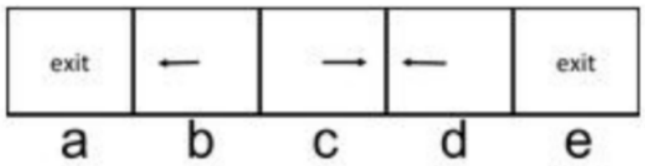
(a) Uz $\gamma = 1$, odredite vrijednost za strategiju π_1:



(b) Uz $\gamma = 1$, odredite vrijednost za strategiju π_2:



(c) Uz $\gamma = 0.9$, odredite vrijednost za strategiju π_3:

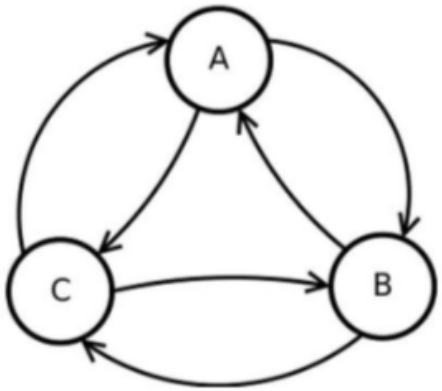


(d) Kako bi izgledalo poboljšanje strategije π_3?

Zadatak 5.

Zadatak za vježbu

Razmotrimo sljedeći dijagram prijelaza, funkciju prijelaza i funkciju nagrade za MPO. Faktor umanjenja je $\gamma = 0.5$.



s	a	s'	$T(s,a,s')$	$R(s,a,s')$
A	$-$	B	0.6	0
A	$-$	C	0.4	-1
A	$+$	B	0.2	-2
A	$+$	C	0.8	-1
B	$-$	A	0.4	2
B	$-$	C	0.6	1
B	$+$	A	0.8	2
B	$+$	C	0.2	-2
C	$-$	A	1	1
C	$+$	A	0.2	1
C	$+$	B	0.8	0

Procjenjujemo sljedeću strategiju π:

A	B	C
$-$	$+$	$+$

Nakon k koraka imamo sljedeću procjenu:

$V_k^\pi(A)$	$V_k^\pi(B)$	$V_k^\pi(C)$
0	1.060	0.640

(a) Izračunajte $V_{k+1}^\pi(A)$, $V_{k+1}^\pi(B)$ i $V_{k+1}^\pi(C)$.

(b) Pretpostavimo da nakon konvergencije imamo sljedeću tablicu:

$V^\pi(A)$	$V^\pi(B)$	$V^\pi(C)$
0.150	1.335	0.749

Izračunajte $Q^\pi(B,+)$ i $Q^\pi(B,-)$. Koji bi bio izbor akcije u stanju B ukoliko određujemo poboljšanje strategije π?

Zadatak 6.

Zadatak za vježbu

U igri *Micro - Blackjack* izvlače se karte (s vraćanjem) 2, 3 i 4 s jednakom vjerojatnošću. Ukoliko je ukupna suma vrijednosti karata manja od 6, možete nastaviti s izvlačenjem ili završiti igru. Ako završite igru, vaša nagrada je ukupna suma vrijednosti izvučenih karata, ukoliko ta suma iznosi najviše 5, inače je 0. Ako nastavite izvlačenje, ne primate nagradu za tu akciju. Ne postoji umanjenje nagrade ($\gamma = 1$).

(a) Navedite stanja i akcije za ovu igru ako ju modelirate kao MPO.

(b) Navedite funkciju prijelaza i funkciju nagrade za ovaj MPO.

(c) Odredite optimalnu strategiju za ovaj MPO.

(d) Koji je najmanji broj iteracija vrijednosti za ovaj MPO nakon kojeg se uočava konvergencija (ukoliko konvergencija postoji)?