

# **Агентно-базирани системи**

## **Домашна задача 1**

Владимир Христовски – 223030

### **1. Нова верзија на Phone или не?**

#### **Состојби:**

- P – позитивно расположение/сентимент кон компанијата
- SS – така-така (неутрално) расположение/сентимент кон компанијата
- N – негативно расположение/сентимент кон компанијата

#### **Дозволени акции:**

- Нова верзија
- Причекај

#### **Награди:**

- $R(P) = +2$
- $R(SS) = 0$
- $R(N) = -1$

#### **Фактор на намалување:**

$$\gamma = 1$$

s	a	T(s, a, s')		
		s' = P	s' = SS	s' = N
P	Нова верзија	0.1	0.9	0
	Причекај	0.2	0.8	0
SS	Нова верзија	0.1	0.9	0
	Причекај	0	0.3	0.7
N	Нова верзија	0.9	0	0.1
	Причекај	0	0.5	0.5

**A:**

	P	SS	N
$V^*_0(s)$	0	0	0
$Q^*_1(s, \text{Нова верзија})$	0.2	0.2	1.7
$Q^*_1(s, \text{Причекај})$	0.4	-0.7	-0.5
$V^*_1(s)$	0.4	0.2	1.7
$Q^*_2(s, \text{Нова верзија})$	0.42	0.42	2.23
$Q^*_2(s, \text{Причекај})$	0.64	0.55	0.45
$V^*_2(s)$	0.64	0.55	2.23

$V^*_0(P) = 0$  - Иницијализација

$V^*_0(SS) = 0$  - Иницијализација

$V^*_0(N) = 0$  - Иницијализација

$$Q^*_1(P, \text{Нова верзија}) = (0.1 * [2 + 1 * 0]) + (0.9 * [0 + 1 * 0]) + (0 * [-1 + 1 * 0]) =$$

$$(0.1 * 2) + (0.9 * 0) + (0 * -1) = 0.2 + 0 + 0 = 0.2$$

$$Q^*_1(P, \text{Причекај}) = (0.2 * [2 + 1 * 0]) + (0.8 * [0 + 1 * 0]) + (0 * [-1 + 1 * 0]) =$$

$$(0.2 * 2) + (0.8 * 0) + (0 * -1) = 0.4 + 0 + 0 = 0.4$$

$$Q^*_1(SS, \text{Нова верзија}) = (0.1 * [2 + 1 * 0]) + (0.9 * [0 + 1 * 0]) + (0 * [-1 + 1 * 0]) =$$

$$(0.1 * 2) + (0.9 * 0) + (0 * -1) = 0.2 + 0 + 0 = 0.2$$

$$Q^*_1(SS, \text{Причекај}) = (0 * [2 + 1 * 0]) + (0.3 * [0 + 1 * 0]) + (0.7 * [-1 + 1 * 0]) =$$

$$(0 * 2) + (0.3 * 0) + (0.7 * -1) = 0 + 0 + -0.7$$

$$Q^*_1(N, \text{Нова верзија}) = (0.9 * [2 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.1 * [-1 + 1 * 0]) = \\ (0.9 * 2) + (0 * 0) + (0.1 * -1) = 1.8 + 0 + -0.1 = 1.7$$

$$Q^*_1(N, \text{Причекај}) = (0 * [2 + 1 * 0]) + (0.5 * [0 + 1 * 0]) + (0.5 * [-1 + 1 * 0]) = \\ (0 * 2) + (0.5 * 0) + (0.5 * -1) = 0 + 0 + -0.5 = -0.5$$

$$V^*_1(P) = \max (Q^*_1(P, \text{Нова верзија}), Q^*_1(P, \text{Причекај})) = \max(0.2, 0.4) = 0.4$$

$$V^*_1(SS) = \max (Q^*_1(SS, \text{Нова верзија}), Q^*_1(SS, \text{Причекај})) = \max(0.2, -0.7) = 0.2$$

$$V^*_1(N) = \max (Q^*_1(N, \text{Нова верзија}), Q^*_1(N, \text{Причекај})) = \max(1.7, -0.5) = 1.7$$

$$Q^*_2(P, \text{Нова верзија}) = (0.1 * [2 + 1 * 0.4]) + (0.9 * [0 + 1 * 0.2]) + (0 * [-1 + 1 * 1.7]) = \\ (0.1 * 2.4) + (0.9 * 0.2) + (0 * 0.7) = 0.24 + 0.18 + 0 = 0.42$$

$$Q^*_2(P, \text{Причекај}) = (0.2 * [2 + 1 * 0.4]) + (0.8 * [0 + 1 * 0.2]) + (0 * [-1 + 1 * 1.7]) = \\ (0.2 * 2.4) + (0.8 * 0.2) + (0 * 0.7) = 0.48 + 0.16 + 0 = 0.64$$

$$Q^*_2(SS, \text{Нова верзија}) = (0.1 * [2 + 1 * 0.4]) + (0.9 * [0 + 1 * 0.2]) + (0 * [-1 + 1 * 1.7]) = \\ (0.1 * 2.4) + (0.9 * 0.2) + (0 * 0.7) = 0.24 + 0.18 + 0 = 0.42$$

$$Q^*_2(SS, \text{Причекај}) = (0 * [2 + 1 * 0.4]) + (0.3 * [0 + 1 * 0.2]) + (0.7 * [-1 + 1 * 1.7]) = \\ (0 * 2.4) + (0.3 * 0.2) + (0.7 * 0.7) = 0 + 0.06 + 0.49 = 0.55$$

$$Q^*_2(N, \text{Нова верзија}) = (0.9 * [2 + 1 * 0.4]) + (0 * [0 + 1 * 0.2]) + (0.1 * [-1 + 1 * 1.7]) = \\ (0.9 * 2.4) + (0 * 0.2) + (0.1 * 0.7) = 2.16 + 0 + 0.07 = 2.23$$

$$Q^*_2(N, \text{Причекај}) = (0 * [2 + 1 * 0.4]) + (0.5 * [0 + 1 * 0.2]) + (0.5 * [-1 + 1 * 1.7]) = \\ (0 * 2.4) + (0.5 * 0.2) + (0.5 * 0.7) = 0 + 0.1 + 0.35 = 0.45$$

$$V^*_2(P) = \max (Q^*_2(P, \text{Нова верзија}), Q^*_2(P, \text{Причекај})) = \max(0.42, 0.64) = 0.64$$

$$V^*_2(SS) = \max (Q^*_2(SS, \text{Нова верзија}), Q^*_2(SS, \text{Причекај})) = \max(0.42, 0.55) = 0.55$$

$$V^*_2(N) = \max (Q^*_2(N, \text{Нова верзија}), Q^*_2(N, \text{Причекај})) = \max(2.23, 0.45) = 2.23$$

## Б:

Според досегашните пресметки, доколку агентот е во позиција  $V^*_2(SS)$  неговата оптимална политика која што треба да ја преземи е “Причекај”, бидејќи резултатите од пресметаните  $Q$  вредности ни даваат до знаење дека политиката “Причекај” има поголема веројатност од политиката “Нова верзија” ( $0.42 < 0.55$ ). Доколку направиме пресметки за  $Q$  вредностите за следниот чекор би добиле:

$$\begin{aligned} Q^*_3(SS, \text{Нова верзија}) &= (0.1 * [2 + 1 * 0.64]) + (0.9 * [0 + 1 * 0.55]) + (0 * [-1 + 1 * 2.23]) = \\ &= (0.1 * 2.64) + (0.9 * 0.55) + (0 * 1.23) = 0.264 + 0.495 + 0 = 0.759 \end{aligned}$$

$$\begin{aligned} Q^*_3(SS, \text{Причекај}) &= (0 * [2 + 1 * 0.64]) + (0.3 * [0 + 1 * 0.55]) + (0.7 * [-1 + 1 * 2.23]) = \\ &= (0 * 2.64) + (0.3 * 0.55) + (0.7 * 1.23) = 0 + 0.165 + 0.861 = 1.026 \end{aligned}$$

$$V^*_3(SS) = \max (Q^*_3(SS, \text{Нова верзија}), Q^*_3(SS, \text{Причекај})) = \max(0.759, 1.026) = 1.026$$

Со овие пресметки ние можеме да утврдиме дека и во следните 2 чекори, кои се последни, оптималната политика за агентот би била “Причекај”. Во претпоследниот чекор веројатноста за акцијата “Причекај” е поголема од веројатноста за акцијата “Нова верзија” ( $0.42 < 0.55$ ), истото важи и за последниот чекор каде веројатноста за акцијата “Причекај” е поголема од веројатноста за акцијата “Нова верзија” ( $0.759 < 1.026$ ).

$$\pi^*(SS) = \text{Причекај}$$

## 2. Во потрага по богатство

### Состојби:

- S0
- S1
- S2
- S3
- S4
- S5
- Hole

### Дозволени акции:

L – лево (веројатност 1.0)

R – десно (веројатност 0.5)

Exit – излез (веројатност 1 ако агентот се наоѓа на едно од двете целни полиња, инаку 0)

### Награди:

- R(Exit) = -4 (при акција R ако агентот се лизне и падне во бездна)

- R(Exit) = +100 (при акција Exit ако агентот се наоѓа на едно од двете целни полиња)

### Фактор на намалување:

$$\gamma = 1$$

s	a	T(s, a, s')						
		s' = S0	s' = S1	s' = S2	s' = S3	s' = S4	s' = S5	s' = Hole
S0	L	0	0	0	0	0	0	0
	R	0	0	0	0	0	0	0
	Exit	1	0	0	0	0	0	0
S1	L	1	0	0	0	0	0	0
	R	0	0	0.5	0	0	0	0.5
	Exit	0	0	0	0	0	0	0
S2	L	0	1	0	0	0	0	0
	R	0	0	0	0.5	0	0	0.5
	Exit	0	0	0	0	0	0	0
S3	L	0	0	1	0	0	0	0
	R	0	0	0	0	0.5	0	0.5
	Exit	0	0	0	0	0	0	0
S4	L	0	0	0	1	0	0	0
	R	0	0	0	0	0	0.5	0.5
	Exit	0	0	0	0	0	0	0
S5	L	0	0	0	0	0	0	0
	R	0	0	0	0	0	0	0
	Exit	0	0	0	0	0	1	0

**A:**

	S0	S1	S2	S3	S4	S5
$V^*_0(s)$	0	0	0	0	0	0
$Q^*_1(s, L)$	/	0	0	0	0	/
$Q^*_1(s, R)$	/	-2	-2	-2	-2	/
$Q^*_1(s, \text{Exit})$	100	/	/	/	/	100
$V^*_1(s)$	100	0	0	0	0	100
$Q^*_2(s, L)$	/	0	0	0	0	/
$Q^*_2(s, R)$	/	-2	-2	-2	-2	/
$Q^*_2(s, \text{Exit})$	200	/	/	/	/	200
$V^*_2(s)$	200	0	0	0	0	200

$V^*_0(S0) = 0$  - Иницијализација

$V^*_0(S1) = 0$  - Иницијализација

$V^*_0(S2) = 0$  - Иницијализација

$V^*_0(S3) = 0$  - Иницијализација

$V^*_0(S4) = 0$  - Иницијализација

$V^*_0(S5) = 0$  - Иницијализација

$Q^*_1(S0, L) =$  Невозможна акција

$Q^*_1(S0, R) =$  Невозможна акција

$$Q^*_1(S0, \text{Exit}) = (1 * [100 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (1 * 100) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 100 + 0 + 0 + 0 + 0 + 0 + 0 = 100$$

$$Q^*_1(S1, L) = (1 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (1 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$Q^*_1(S1, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0.5 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$Q^*_1(S1, \text{Exit}) =$  Невозможна акција

$$Q^*_1(S2, L) = (0 * [0 + 1 * 0]) + (1 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (1 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$Q^*_1(S2, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (0.5 * 0) + (0 * 0) + (0 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$Q^*_1(S2, \text{Exit}) =$  Невозможна акција

$$Q^*_1(S3, L) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (1 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (1 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$Q^*_1(S3, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0.5 * 0) + (0 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$Q^*_1(S3, \text{Exit}) = \text{Невозможна акција}$

$$Q^*_1(S4, L) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (1 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (1 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$Q^*_1(S4, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0.5 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0.5 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$Q^*_1(S4, \text{Exit}) = \text{Невозможна акција}$

$Q^*_1(S5, L) = \text{Невозможна акција}$

$Q^*_1(S5, R) = \text{Невозможна акција}$

$$Q^*_1(S5, \text{Exit}) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (1 * [100 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (1 * 100) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 100 + 0 = 100$$

$$V^*_1(S0) = \max(Q^*_1(S0, \text{Exit})) = \max(100) = 100$$

$$V^*_1(S1) = \max(Q^*_1(S1, L), Q^*_1(S1, R)) = \max(0, -2) = 0$$

$$V^*_1(S2) = \max(Q^*_1(S2, L), Q^*_1(S2, R)) = \max(0, -2) = 0$$

$$V^*_1(S3) = \max(Q^*_1(S3, L), Q^*_1(S3, R)) = \max(0, -2) = 0$$

$$V^*_1(S4) = \max(Q^*_1(S4, L), Q^*_1(S4, R)) = \max(0, -2) = 0$$

$$V^*_1(S5) = \max(Q^*_1(S5, \text{Exit})) = \max(100) = 100$$



$Q^*_2(S0, L) = \text{Невозможна акција}$

$Q^*_2(S0, R) = \text{Невозможна акција}$

$$Q^*_2(S0, \text{Exit}) = (1 * [100 + 1 * 100]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (1 * 200) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 200 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 200$$

$$Q^*_2(S1, L) = (1 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (1 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$Q^*_2(S1, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0.5 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$Q^*_2(S1, \text{Exit}) = \text{Невозможна акција}$

$$Q^*_2(S2, L) = (0 * [0 + 1 * 0]) + (1 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (1 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$Q^*_2(S2, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (0.5 * 0) + (0 * 0) + (0 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$Q^*_2(S2, \text{Exit}) = \text{Невозможна акција}$

$$Q^*_2(S3, L) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (1 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (1 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$Q^*_2(S3, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0.5 * 0) + (0 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$Q^*_2(S3, \text{Exit}) = \text{Невозможна акција}$

$$Q^*_2(S4, L) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (1 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (1 * 0) + (0 * 0) + (0 * 0) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$Q^*_2(S4, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0.5 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0.5 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$$Q^*_2(S4, \text{Exit}) = \text{Невозможна акција}$$

$$Q^*_2(S5, L) = \text{Невозможна акција}$$

$$Q^*_2(S5, R) = \text{Невозможна акција}$$

$$Q^*_2(S5, \text{Exit}) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (0 * [0 + 1 * 0]) + (1 * [100 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (0 * 0) + (1 * 200) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 200 + 0 = 200$$

$$V^*_2(S0) = \max(Q^*_2(S0, \text{Exit})) = \max(200) = 200$$

$$V^*_2(S1) = \max(Q^*_2(S1, L), Q^*_2(S1, R)) = \max(0, -2) = 0$$

$$V^*_2(S2) = \max(Q^*_2(S2, L), Q^*_2(S2, R)) = \max(0, -2) = 0$$

$$V^*_2(S3) = \max(Q^*_2(S3, L), Q^*_2(S3, R)) = \max(0, -2) = 0$$

$$V^*_2(S4) = \max(Q^*_2(S4, L), Q^*_2(S4, R)) = \max(0, -2) = 0$$

$$V^*_2(S5) = \max(Q^*_2(S5, \text{Exit})) = \max(200) = 200$$

**Б:**

$$\pi(S1) = R$$

$$\pi(S2) = L$$

$$\pi(S3) = L$$

$$\pi(S4) = R$$

$$V(S1) = -4$$

$$V(S2) = -4$$

$$V(S3) = 8$$

$$V(S4) = 10$$

	S1	S2	S3	S4
$V^*_0(s)$	-4	-4	8	10
$Q^*_1(s, L)$	/	-4	-4	/
$Q^*_1(s, R)$	-4	/	/	-2
$V^*_1(s)$	-4	-4	-4	-2
$Q^*_2(s, L)$	/	-4	-4	/
$Q^*_2(s, R)$	-4	/	/	-2
$V^*_2(s)$	-4	-4	-4	-2

$V^*_0(S1) = -4$  – Искуство од претходен трагач

$V^*_0(S2) = -4$  - Искуство од претходен трагач

$V^*_0(S3) = 8$  - Искуство од претходен трагач

$V^*_0(S4) = 10$  - Искуство од претходен трагач

$$Q^*_1(S1, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0.5 * [0 + 1 * -4]) + (0 * [0 + 1 * 8]) + (0 * [0 + 1 * 10]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (0.5 * -4) + (0 * 8) + (0 * 10) + (0 * 0) + (0.5 * -4) = 0 + 0 + -2 + 0 + 0 + 0 + -2 = -4$$

$$Q^*_1(S2, L) = (0 * [0 + 1 * 0]) + (1 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * 8]) + (0 * [0 + 1 * 10]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (1 * -4) + (0 * -4) + (0 * 8) + (0 * 10) + (0 * 0) + (0 * -4) = 0 + -4 + 0 + 0 + 0 + 0 + 0 = -4$$

$$Q^*_1(S3, L) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (1 * [0 + 1 * -4]) + (0 * [0 + 1 * 8]) + (0 * [0 + 1 * 10]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (1 * -4) + (0 * -4) + (0 * 8) + (0 * 10) + (0 * 0) + (0 * -4) = 0 + 0 + -4 + 0 + 0 + 0 + 0 = -4$$

$$Q^*_1(S4, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * 8]) + (0 * [0 + 1 * 10]) + (0.5 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (0 * -4) + (0 * 8) + (0 * 10) + (0.5 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$$V^*_1(S1) = \max (Q^*_1(S1, R)) = \max(-4) = -4$$

$$V^*_1(S2) = \max (Q^*_1(S2, L)) = \max(-4) = -4$$

$$V^*_1(S3) = \max (Q^*_1(S3, L)) = \max(-4) = -4$$

$$V^*_1(S4) = \max (Q^*_1(S4, R)) = \max(-2) = -2$$

$$Q^*_2(S1, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0.5 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (0.5 * -4) + (0 * -4) + (0 * -2) + (0 * 0) + (0.5 * -4) = 0 + 0 + -2 + 0 + 0 + 0 + -2 = -4$$

$$Q^*_2(S2, L) = (0 * [0 + 1 * 0]) + (1 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (1 * -4) + (0 * -4) + (0 * -4) + (0 * -2) + (0 * 0) + (0 * -4) = 0 + -4 + 0 + 0 + 0 + 0 + 0 = -4$$

$$Q^*_2(S3, L) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (1 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (1 * -4) + (0 * -4) + (0 * -4) + (0 * -2) + (0 * 0) + (0 * -4) = 0 + 0 + -4 + 0 + 0 + 0 + 0 = -4$$

$$Q^*_2(S4, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0.5 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (0 * -4) + (0 * -4) + (0 * -2) + (0.5 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$$V^*_2(S1) = \max (Q^*_2(S1, R)) = \max(-4) = -4$$

$$V^*_2(S2) = \max (Q^*_2(S2, L)) = \max(-4) = -4$$

$$V^*_2(S3) = \max (Q^*_2(S3, L)) = \max(-4) = -4$$

$$V^*_2(S4) = \max (Q^*_2(S4, R)) = \max(-2) = -2$$

**B:**

	S1	S2	S3	S4
$V^*_2(s)$	-4	-4	-4	-2
$Q^*_3(s, L)$	0	-4	-4	-4
$Q^*_3(s, R)$	-4	-4	-3	-2
$V^*_3(s)$	0	-4	-3	-2

$$V^*_2(S1) = -4 - \text{Искуство од претходно барање, чекор } i = 2$$

$$V^*_2(S2) = -4 - \text{Искуство од претходно барање, чекор } i = 2$$

$$V^*_2(S3) = -4 - \text{Искуство од претходно барање, чекор } i = 2$$

$$V^*_2(S4) = -0.5 - \text{Искуство од претходно барање, чекор } i = 2$$

$$Q^*_3(S1, L) = (1 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (1 * 0) + (0 * -4) + (0 * -4) + (0 * -4) + (0 * -2) + (0 * 0) + (0 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 = 0$$

$$Q^*_3(S1, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0.5 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (0.5 * -4) + (0 * -4) + (0 * -2) + (0 * 0) + (0.5 * -4) = 0 + 0 + -2 + 0 + 0 + 0 + -2 = -4$$

$$Q^*_3(S2, L) = (0 * [0 + 1 * 0]) + (1 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (1 * -4) + (0 * -4) + (0 * -4) + (0 * -2) + (0 * 0) + (0 * -4) = 0 + -4 + 0 + 0 + 0 + 0 + 0 = -4$$

$$Q^*_3(S2, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0.5 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (0 * -4) + (0.5 * -4) + (0 * -2) + (0 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + -2 + 0 + 0 + -2 = -4$$

$$Q^*_3(S3, L) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (1 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (1 * -4) + (0 * -4) + (0 * -2) + (0 * 0) + (0 * -4) = 0 + 0 + -4 + 0 + 0 + 0 + 0 = -4$$

$$Q^*_3(S3, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0.5 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (0 * -4) + (0 * -4) + (0.5 * -2) + (0 * 0) + (0.5 * -4) = 0 + 0 + 0 + 0 + -1 + 0 + -2 = -3$$

$$Q^*_3(S4, L) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (1 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0 * [0 + 1 * 0]) + (0 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (0 * -4) + (1 * -4) + (0 * 0) + (0 * -2) + (0 * -4) = 0 + 0 + 0 + -4 + 0 + 0 + 0 = -4$$

$$Q^*_3(S4, R) = (0 * [0 + 1 * 0]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -4]) + (0 * [0 + 1 * -2]) + (0.5 * [0 + 1 * 0]) + (0.5 * [-4 + 1 * 0]) = (0 * 0) + (0 * -4) + (0 * -4) + (0 * -4) + (1 * 0) + (0 * -2) + (0.5 * -4) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + -2 = -2$$

$$V^*_3(S1) = \max(Q^*_1(S2, L), Q^*_1(S1, R)) = \max(0, -4) = 0$$

$$V^*_3(S2) = \max(Q^*_1(S2, L), Q^*_1(S1, R)) = \max(-4, -4) = -4$$

$$V^*_3(S3) = \max(Q^*_1(S2, L), Q^*_1(S1, R)) = \max(-4, -3) = -3$$

$$V^*_3(S4) = \max(Q^*_1(S2, L), Q^*_1(S1, R)) = \max(-4, -2) = -2$$

Со овие пресметки заклучуваме дека после еден чекор има измена во дел од политиките и тоа:

$\pi^*(S1) = L$  – подобрена политика за разлика од претходната каде  $\pi^*(S1) = R$

$\pi^*(S2) = L/R$  – политиката во овој чекор е непроменета, но според правилата на игра јас би ја поставил  $\pi^*(S2) = L$  поради сигурноста од веројатност 1

$\pi^*(S3) = R$  – подобрена политика за разлика од претходната каде  $\pi^*(S3) = L$

$\pi^*(S4) = R$  – политиката остана иста