

UNIVERZA NA PRIMORSKEM  
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN  
INFORMACIJSKE TEHNOLOGIJE

Zaključna naloga  
Učenje iz interakcije  
Learning from interaction

Ime in priimek: Rok Breulj  
Študijski program: Računalništvo in informatika  
Mentor: doc. dr. Peter Rogelj

Koper, Avgust 2013

## Ključna dokumentacijska informacija

## Key words documentation

## Zahvala

# Kazalo

<b>1</b>	<b>Uvod</b>	<b>7</b>
1.1	Okrepiteveno učenje . . . . .	7
1.2	Primeri . . . . .	7
1.3	Elementi okrepitvenega učenja . . . . .	7
<b>2</b>	<b>Problem</b>	<b>8</b>
2.1	Ocenjevanje povratne informacije . . . . .	8
2.2	Celoten problem okrepitvenega učenja . . . . .	8
<b>3</b>	<b>Rešitve</b>	<b>9</b>
3.1	Dinamično programiranje . . . . .	9
3.2	Predvidevanje - vrednost stanja . . . . .	9
3.2.1	Monte Carlo metode . . . . .	9
3.2.2	Učenje na podlagi časovne razlike - TD(0) . . . . .	9
3.2.3	Združitev metod - TD( $\lambda$ ) . . . . .	9
3.3	Krmiljenje - vrednost dejanja . . . . .	9
3.3.1	Monte Carlo metode . . . . .	9
3.3.2	Učenje na podlagi časovne razlike - TD(0) . . . . .	9
3.3.3	Združitev metod - TD( $\lambda$ ) . . . . .	9
<b>4</b>	<b>Posploševanje in funkcijska aproksimacija</b>	<b>10</b>
4.1	Predvidevanje - vrednost stanja . . . . .	10
4.2	Krmiljenje - vrednost dejanja . . . . .	10
<b>5</b>	<b>Učenje na namizni igri Hex</b>	<b>11</b>
5.1	Ozadje . . . . .	11
5.2	Implementacija . . . . .	11
<b>6</b>	<b>Zaključek</b>	<b>12</b>
<b>7</b>	<b>Literatura</b>	<b>13</b>
<b>8</b>	<b>Priloge</b>	<b>14</b>

## Tabele

## Slike

# **1 Uvod**

## **1.1 Okrepiteveno učenje**

## **1.2 Primeri**

## **1.3 Elementi okrepitevenega učenja**



## **2 Problem**

### **2.1 Ocenjevanje povratne informacije**

### **2.2 Celoten problem okrepitevenega učenja**

## 3 Rešitve

### 3.1 Dinamično programiranje

### 3.2 Predvidevanje - vrednost stanja

#### 3.2.1 Monte Carlo metode

#### 3.2.2 Učenje na podlagi časovne razlike - TD(0)

#### 3.2.3 Združitev metod - TD( $\lambda$ )

### 3.3 Krmiljenje - vrednost dejanja

#### 3.3.1 Monte Carlo metode

#### 3.3.2 Učenje na podlagi časovne razlike - TD(0)

#### 3.3.3 Združitev metod - TD( $\lambda$ )

## **4 Posploševanje in funkcijska aproksimacija**

### **4.1 Predvidevanje - vrednost stanja**

### **4.2 Krmiljenje - vrednost dejanja**

## 5 Učenje na namizni igri Hex

### 5.1 Ozadje

### 5.2 Implementacija

## 6 Zaključek

## 7 Literatura

## 8 Priloge