

Spodbujevano učenje pri igranju namiznih iger (angl. *Reinforcement learning in board games*)

Tim Kalan

Fakulteta za matematiko in fiziko

27. oktober 2020

Strojno učenje

- ▶ Nadzorovano učenje
- ▶ Nenadzorovano učenje
- ▶ **Spodbujevano učenje**

Motivacija: instrumentalno pogojevanje

- ▶ Tu bo slika
- ▶ Lepa psihološko motivirana podlaga
- ▶ **Nagrade in kazni**

Spodbujevano učenje

- ▶ **Okolje, agent, nagrada, (model)**
- ▶ Pomemben je čas
- ▶ Ne poznamo »pravilnih« akcij
- ▶ Raziskovanje in izkoriščanje
- ▶ Vrednostna funkcija

Primer: Križci in krožci 1

- ▶ tu slika tistega loopa
- ▶ **Stanje:** Kje je prazno, kje »X« in kje »O«
- ▶ **Agent:** Program, ki se odloča, kako igrati
- ▶ **Okolje:** Agentu sporoča nagrade in stanje
- ▶ **Nagrada:** Pozitivna za zmago, negativna za poraz

Primer: Križci in krožci 2

- ▶ Agent igra igre, posodablja svoje vrednosti stanj glede na odgovor okolja
- ▶ Kako naj to stori?

Primer: Križci in krožci 3

- ▶ Enostavna ideja:

Primer: Križci in krožci 3

- ▶ Enostavna ideja:

$$V(s) \leftarrow V(s) + \alpha * [\gamma * V(s') - V(s)]$$

Primer: Križci in krožci 3

- ▶ Enostavna ideja:

$$V(s) \leftarrow V(s) + \alpha * [\gamma * V(s') - V(s)]$$

- ▶ s je trenutno stanje

Primer: Križci in krožci 3

- ▶ Enostavna ideja:

$$V(s) \leftarrow V(s) + \alpha * [\gamma * V(s') - V(s)]$$

- ▶ s je trenutno stanje
- ▶ V je vrednostna funkcija

Primer: Križci in krožci 3

- ▶ Enostavna ideja:

$$V(s) \leftarrow V(s) + \alpha * [\gamma * V(s') - V(s)]$$

- ▶ s je trenutno stanje
- ▶ V je vrednostna funkcija
- ▶ α je velikost koraka (hitrost učenja)

Primer: Križci in krožci 3

- ▶ Enostavna ideja:

$$V(s) \leftarrow V(s) + \alpha * [\gamma * V(s') - V(s)]$$

- ▶ s je trenutno stanje
- ▶ V je vrednostna funkcija
- ▶ α je velikost koraka (hitrost učenja)
- ▶ γ je diskontni faktor (pomemben je čas)

Primer: Križci in krožci 3

- ▶ Enostavna ideja:

$$V(s) \leftarrow V(s) + \alpha * [\gamma * V(s') - V(s)]$$

- ▶ s je trenutno stanje
- ▶ V je vrednostna funkcija
- ▶ α je velikost koraka (hitrost učenja)
- ▶ γ je diskontni faktor (pomemben je čas)
- ▶ s' je stanje, ki sledi s

Demonstracija: Križci in krožci

Morda kakšna slika/grafikon