

Inverzna kinematika manipulatora ostvarena pomoću dubokog potpornog učenja

Autor: Josip Torić

Mentor: izv. prof. dr. sc. Marija Seder

Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu

10. srpnja 2020.

1 Uvod

2 Teorijske
osnove potpornog
učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične
primjene
potpornog učenja

4 Algoritmi
optimizacije
politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna
kinematika
robotske ruke
Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Sadržaj

Uvod

Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

Praktične primjene potpornog učenja

Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

Inverzna kinematika robotske ruke Jaco

Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

Zaključak

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Sadržaj

Uvod

Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

Praktične primjene potpornog učenja

Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

Inverzna kinematika robotske ruke Jaco

Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

Zaključak

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Uvod

Živimo u svijetu prepunom informacija, zadataka i problema.

Znanstvenici su nakon Drugog svjetskog rata krenuli razvijati umjetnu inteligenciju iz koje su se kasnije razvili strojno učenje, duboko učenje i potporno učenje.

Cilj ovog diplomskog rada je naučiti inverznu kinematiku robotske ruke Jaco uz primjenu dubokog potpornog učenja.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Sadržaj

Uvod

Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

Praktične primjene potpornog učenja

Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

Inverzna kinematika robotske ruke Jaco

Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

Zaključak

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog

gradijenta politike

Algoritam optimizacije

politike uz regije

povjerenja

Algoritam proksimalne

optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Koncept potpornog učenja

Potporno učenje je grana strojnog učenja koja proučava agente i kako oni uče na temelju pokušaja i pogreške.

Osnovna ideja potpornog učenja je da postoji agent u svojem okruženju.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

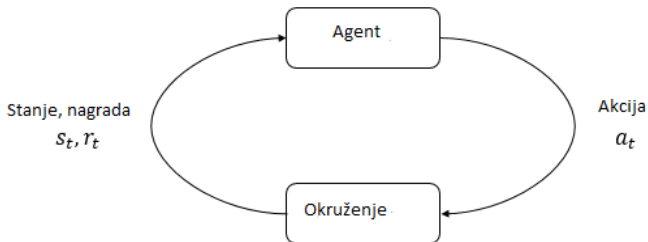
Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Odnos agenta i okruženja



1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Duboko potporno učenje

Politike odabira akcija u potpornom učenju se spremaju u funkcije.

No, kao i svaka funkcija ona je ograničena s parametrima i klasom funkcija koje pripada. Zbog toga u prošlosti, potporno učenje je imalo limitirane primjene.

Pravi proboj u potpornom učenju se dogodio kada su počeli umjesto funkcija aproksimirati dubokim neuronskim mrežama.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Osnovni pojmovi potpornog učenja

- ▶ Prostor opservacija
- ▶ Prostor akcija
- ▶ Politika
- ▶ Putanja
- ▶ Nagrada

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Problem potpornog učenja

Problem potpornog učenja je odabrati optimalnu politiku koja će maksimizirati očekivanu nagradu.

$$P(\tau|\pi) = \rho_0(s_0) \prod_{t=0}^{T-1} P(s_{t+1}|s_t, a_t) \pi(a_t|s_t) \quad (1)$$

$$J(\pi) = \int_{\tau} P(\tau|\pi) R(\tau) = \underset{\tau \sim \pi}{E} [R(\tau)] \quad (2)$$

$$\pi^* = \arg \max_{\pi} J(\pi) \quad (3)$$

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Vrijednosne funkcije

$$V^{\pi}(s) = E_{\tau \sim \pi} [R(\tau) | s_0 = s] \quad (4)$$

$$Q^{\pi}(s, a) = E_{\tau \sim \pi} [R(\tau) | s_0 = s, a_0 = a] \quad (5)$$

$$V^*(s) = \max_{\pi} E_{\tau \sim \pi} [R(\tau) | s_0 = s] \quad (6)$$

$$Q^*(s, a) = \max_{\pi} E_{\tau \sim \pi} [R(\tau) | s_0 = s, a_0 = a] \quad (7)$$

$$A^{\pi}(s, a) = Q^{\pi}(s, a) - V^{\pi}(s) \quad (8)$$

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Bellmanove jednadžbe

$$V^\pi(s) = E_{s' \sim P}[r(s, a) + \gamma V^\pi(s')] \quad (9)$$

$$Q^\pi(s, a) = E_{s' \sim P}[r(s, a) + \gamma E_{a' \sim \pi}[Q^\pi(s', a')]] \quad (10)$$

$$V^*(s) = \max_a E_{s' \sim P}[r(s, a) + \gamma V^*(s')] \quad (11)$$

$$Q^*(s, a) = E_{s' \sim P}[r(s, a) + \gamma \max_{a'} Q^*(s', a')] \quad (12)$$

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina
potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

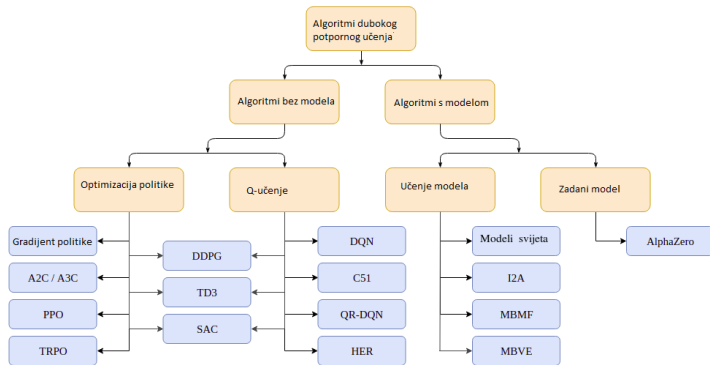
Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Podjela algoritama potpornog učenja



1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Sadržaj

Uvod

Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

Praktične primjene potpornog učenja

Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

Inverzna kinematika robotske ruke Jaco

Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

Zaključak

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Primjene u igranju šaha

Postoji otprilike 10^{120} mogućih šahovskih partija.

Većina današnjih računalnih programa za igranje šaha počiva na nekom principu pretraživanja prostora.

Algoritmi bazirani na dubokom potpornom učenju, kao što su AlphaZero i Leela Chess Zero, uspjeli su kombinirati računalnu snagu s ljudskom intuicijom.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Primjene u igranju računalnih igara

Računalne igre, za razliku od igara na ploči, su korak bliže realnom svijetu.

Vjerojatno jedna od najpoznatijih primjena dubokog potpornog učenja u zadnje vrijeme je kada je skupina znanstvenika okupljena u OpenAI timu razvila botove koji su postali bolji od svjetskih prvaka u računalnoj igri Dota 2.

Također, jedan od čestih benchmarkova za evaluiranje rada algoritama je igranje igara na konzoli Atari 2600.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Atari 2600



1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Primjene u robotici

Kod robota okruženje je stvarni svijet i njihove akcije imaju posljedice u stvarnom svijetu.

Robotu su najčešće dostupne informacije o položaju u kojem se nalazi, zatim o kutevima zglobova koji ga pokreću, trenutnoj brzini, itd.

Jedan od najvećih problema vezano za robotiku je nepraktičnost treniranja robota.

Glavni zadatak ovog diplomskog rada je naučiti inverznu kinematiku robotske ruke Jaco prilikom dohvaćanja loptice iz okruženja, prvo bez prepreka te zatim s preprekama

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Sadržaj

Uvod

Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

Praktične primjene potpornog učenja

Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

Inverzna kinematika robotske ruke Jaco

Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

Zaključak

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Pregled algoritama optimizacije politike

Najosnovniji algoritam optimizacije politike je algoritam jednostavnog gradijenta politike, ali problem kod njega je prevelika nestabilnost.

Zbog toga ću iznijeti još dva algoritma, algoritam optimizacije politike uz regije povjerenja i algoritam proksimalne optimizacije politike.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Algoritam jednostavnog gradijenta politike

- 1: Ulaz: Inicijalni parametri politike θ_0 , inicijalni parametri funkcije vrijednosti ϕ_0
- 2: **for** $k = 0, 1, 2, \dots$ **do**
- 3: Prikupimo listu putanja $\mathcal{D}_k = \{\tau_i\}$ prateći politiku $\pi_k = \pi(\theta_k)$ u okruženju.
- 4: Izračunajmo nagrade \hat{R}_t .
- 5: Izračunajmo funkciju prednosti \hat{A}_t koristeći trenutnu funkciju vrijednosti V_{ϕ_k} .
- 6: Izračunajmo gradijent prema formuli:

$$\hat{g}_k = \frac{1}{|\mathcal{D}_k|} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) |_{\theta_k} \hat{A}_t.$$

- 7: Izračunajmo nove parametre uz pomoć gradijentnog spusta:

$$\theta_{k+1} = \theta_k + \alpha_k \hat{g}_k,$$

- 8: Izračunajmo nove parametre funkcije vrijednosti

$$\phi_{k+1} = \arg \min_{\phi} \frac{1}{|\mathcal{D}_k| T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T \left(V_{\phi}(s_t) - \hat{R}_t \right)^2,$$

- 9: **end for**

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Algoritam optimizacije politike uz regije povjerenja

- 1: Ulaz: Inicijalni parametri politike θ_0 , inicijalni parametri funkcije vrijednosti ϕ_0
- 2: Hiperparametri: Limit KL udaljenosti, koeficijent traženja unazad α , maksimalan broj koraka traženja unazad K
- 3: **for** $k = 0, 1, 2, \dots$ **do**
- 4: Prikupimo listu putanja $\mathcal{D}_k = \{\tau_i\}$ prateći politiku $\pi_k = \pi(\theta_k)$ u okruženju.
- 5: Izračunajmo nagrade \hat{R}_t .
- 6: Izračunajmo funkciju prednosti \hat{A}_t koristeći trenutnu funkciju vrijednosti V_{ϕ_k} .
- 7: Izračunajmo gradijent prema formuli:

$$\hat{g}_k = \frac{1}{|\mathcal{D}_k|} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) |_{\theta_k} \hat{A}_t.$$

- 8: Uz pomoć algoritma konjugatnog gradijenta izračunajmo:

$$\hat{x} \approx \hat{H}_k^{-1} \hat{g}_k$$

- 9: Izračunajmo nove parametre uz pomoć pretrage unatrag

$$\theta_{k+1} = \theta_k + \alpha^j \sqrt{\frac{2\delta}{\hat{x}_k^T \hat{H}_k \hat{x}_k}} \hat{x}_k$$

- 10: Izračunajmo nove parametre funkcije vrijednosti

$$\phi_{k+1} = \arg \min_{\phi} \frac{1}{|\mathcal{D}_k| T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T (V_{\phi}(s_t) - \hat{R}_t)^2$$

- 11: **end for**

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Algoritam proksimalne optimizacije politike

- 1: Ulaz: Inicijalni parametri politike θ_0 , inicijalni parametri funkcije vrijednosti ϕ_0
- 2: **for** $k = 0, 1, 2, \dots$ **do**
- 3: Prikupimo listu putanja $\mathcal{D}_k = \{\tau_i\}$ prateći politiku $\pi_k = \pi(\theta_k)$ u okruženju.
- 4: Izračunajmo nagrade \hat{R}_t .
- 5: Izračunajmo funkciju prednosti \hat{A}_t koristeći trenutnu funkciju vrijednosti V_{ϕ_k} .
- 6: Izračunajmo nove parametre politike

$$\phi_{k+1} = \arg \min_{\phi} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T L(s, a, \theta_k, \theta),$$

gdje je

$$L(s, a, \theta_k, \theta) = \min \left(\frac{\pi_{\theta}(a|s)}{\pi_{\theta_k}(a|s)} A^{\pi_{\theta_k}}(s, a), \text{clip} \left(\frac{\pi_{\theta}(a|s)}{\pi_{\theta_k}(a|s)}, 1 - \epsilon, 1 + \epsilon \right) A^{\pi_{\theta_k}}(s, a) \right),$$

- 7: Izračunajmo nove parametre funkcije vrijednosti

$$\phi_{k+1} = \arg \min_{\phi} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T \left(V_{\phi}(s_t) - \hat{R}_t \right)^2$$

- 8: **end for**

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Sadržaj

Uvod

Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

Praktične primjene potpornog učenja

Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

Inverzna kinematika robotske ruke Jaco

Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

Zaključak

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Robotski manipulatori

Robotski manipulatori sastoje se od čvrstih tijela koji su međusobno povezanih zglobovima.

Upravljanje robotima se dijeli na direktnu i inverznu kinematiku.

Jaco je učvršćen za podlogu, ima šest zglobova na čijem vrhu se nalazi efektor, a robotom se upravlja tako da se zadaju brzine zglobova.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Robotska ruka Jaco



1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina
potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Problem dohvaćanja objekta

Cilj problema je efektorom robotske ruke doći do objekta, prvo u prostoru bez prepreka, onda s preprekama.

Prostor opservacija će se sastojati od položaja efektora, cilja i prepreka, a prostor akcija će biti šest brzina zglobova.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Implementacija manipulatora u simulatoru

Robotska ruka Jaco ostvarena je u simulatoru s programskim jezikom Python i njegovim bibliotekama.

Fizički model ostvaren je uz pomoć PyBulleta i Pyb-Manipulatora, a algoritmi su ostvareni uz pomoć Gyma i Baselinesa.

Ovo je sve trebalo povezati u okruženje za treniranje te pronaći odgovarajuću funkciju nagrade, odnosno gubitka.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

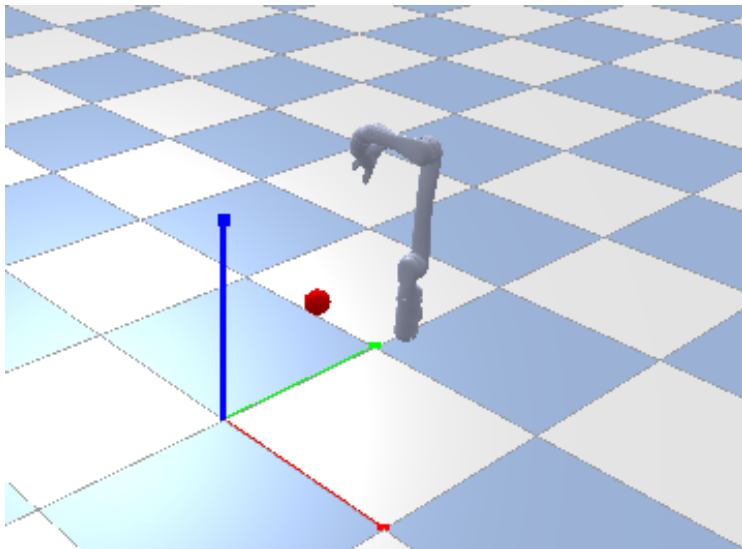
Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Problem bez prepreka



1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

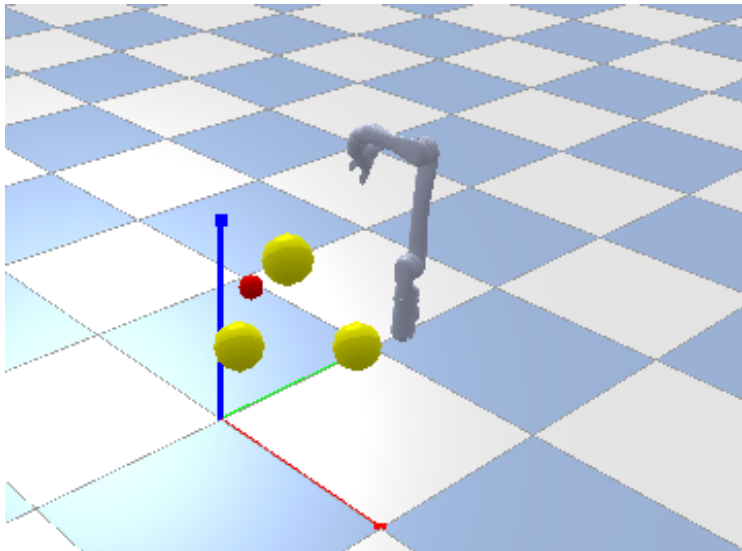
Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Problem s preprekama



1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Sadržaj

Uvod

Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

Praktične primjene potpornog učenja

Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

Inverzna kinematika robotske ruke Jaco

Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

Zaključak

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Izlaz agenta

Ime parametra	Vrijednost
eplenmean	5e+03
eprewmean	-5.09e+03
fps	5.54e+03
loss/approxkl	0.0064
loss/clipfrac	0.0686
loss/policy_entropy	8.43
loss/policy_loss	-0.0045
loss/value_loss	104
misc/explained_variance	0.00614
misc/nupdates	3
misc/serial_timesteps	6.14e+03
misc/time_elapsed	141
misc/total_timesteps	7.86e+05

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Udaljenost i brzina

```
observation = self.get_observation()  
distance_vector = observation[:3] - self.target  
distance_reward = - np.linalg.norm(distance_vector)  
speed_reward = - np.linalg.norm(action)  
reward = distance_reward + speed_reward
```

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene

potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Udaljenost i brzina

eplenmean	eprewmean	misc/total_timesteps
5001.00	-4644.67	1.57e+06
5001.00	-4405.25	2.88e+06
4934.95	-4400.78	4.19e+06
5001.00	-4487.64	5.51e+06
4976.26	-4392.46	6.82e+06
4860.11	-4157.88	8.13e+06
4916.93	-4253.58	9.44e+06
4879.62	-4589.74	1.07e+07
4953.33	-4476.76	1.21e+07
4911.33	-4511.79	1.34e+07
4862.73	-4430.52	1.47e+07
4928.81	-4643.92	1.60e+07
4973.91	-4636.84	1.73e+07
5001.00	-4764.21	1.86e+07
4968.79	-4585.88	1.99e+07

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Napredak i brzina

```
observation = self.get_observation()
distance_vector = observation[:3] - self.target
distance_reward = - np.linalg.norm(distance_vector)
progress_reward = - 300 * (-distance_reward +
                           self.distance_reward_old)
self.distance_reward_old = distance_reward
speed_reward = - np.linalg.norm(action)
reward = progress_reward + speed_reward
```

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina
potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Napredak i brzina

eplenmean	eprewmean	misc/total_timesteps
4558.74	-2145.22	1.57e+06
3990.53	-1814.93	2.88e+06
1933.10	-884.44	4.19e+06
1260.32	-582.99	5.51e+06
1300.25	-597.80	6.82e+06
1016.17	-468.28	8.13e+06
885.16	-409.56	9.44e+06
824.18	-381.70	1.07e+07
928.05	-428.19	1.21e+07
890.89	-412.20	1.34e+07
793.56	-358.08	1.47e+07
854.25	-394.39	1.60e+07
711.99	-326.41	1.73e+07
746.49	-337.91	1.86e+07
791.45	-361.71	1.99e+07

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Udaljenost, napredak i brzina

```
observation = self.get_observation()
distance_vector = observation[:3] - self.target
distance_reward = - np.linalg.norm(distance_vector)
progress_reward = - 300 * (-distance_reward +
                           self.distance_reward_old)
self.distance_reward_old = distance_reward
speed_reward = - np.linalg.norm(action)
reward = distance_reward + progress_reward + speed_reward
```

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina
potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Udaljenost, napredak i brzina

eplenmean	eprewmean	misc/total_timesteps
3819.96	-3031.81	1.57e+06
3578.17	-2906.71	2.88e+06
2264.80	-1710.63	4.19e+06
1791.30	-1360.40	5.51e+06
1497.91	-1119.63	6.82e+06
1110.11	-774.56	8.13e+06
971.30	-714.09	9.44e+06
977.26	-753.41	1.07e+07
1102.77	-832.46	1.21e+07
830.59	-669.11	1.34e+07
779.17	-604.76	1.47e+07
773.30	-599.63	1.60e+07
807.47	-645.58	1.73e+07
719.24	-580.12	1.86e+07
701.33	-577.36	1.99e+07

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene

potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Linearno kažnjavanje blizine i fiksne nagrade na kraju epizode

```
#reward calculation
observation = self.get_observation()
distance_vector = observation[:3] - self.target
distance_reward = - np.linalg.norm(distance_vector)
progress_reward = - 300 * (-distance_reward +
                           self.distance_reward_old)
self.distance_reward_old = distance_reward
speed_reward = - np.linalg.norm(action)
min_obstacle_distances = self.min_obstacle_distances()
obstacle_reward = -0.4/np.min(min_obstacle_distances)+1.2
reward = distance_reward + progress_reward +
        speed_reward + obstacle_reward

# check if the goal has been reached
# otherwise check if the episode is over
if np.linalg.norm(distance_vector) < 0.05:
    reward = 2000
    done = True
elif np.any(min_obstacle_distances < 0.15):
    reward = -1000
    done = True
elif self.number_of_steps > self.episode_length:
    done = True
else:
    done = False
```

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Linearno kažnjavanje blizine i fiksne nagrade na kraju epizode

eplenmean	eplewmean	misc/total_timesteps
4660.64	-4230.35	1.57e+06
4067.72	-3506.76	2.88e+06
3041.90	-2856.47	4.19e+06
3681.57	-3091.60	5.51e+06
3496.40	-3116.02	6.82e+06
3568.77	-3106.93	8.13e+06
3296.14	-2790.05	9.44e+06
3704.81	-2935.27	1.07e+07
3625.26	-3061.75	1.21e+07
3196.91	-2385.26	1.34e+07
2445.88	-1936.88	1.47e+07
2597.89	-1847.29	1.60e+07
2053.09	-1335.68	1.73e+07
1389.56	-628.74	1.86e+07
1786.80	-1134.73	1.99e+07

1 Uvod

2 Teorijske
osnove potpornog
učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične
primjene
potpornog učenja

4 Algoritmi
optimizacije
politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna
kinematika
robotske ruke
Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Linearno kažnjavanje blizine i bez završavanja prilikom udarca u prepreku

```
observation = self.get_observation()
distance_vector = observation[:3] - self.target
distance_reward = - np.linalg.norm(distance_vector)
progress_reward = - 300 * (-distance_reward +
                           self.distance_reward_old)
self.distance_reward_old = distance_reward
speed_reward = - np.linalg.norm(action)
min_obstacle_distances = self.min_obstacle_distances()
obstacle_reward = -0.4/np.min(min_obstacle_distances)+1.2
reward = distance_reward + progress_reward +
        speed_reward + obstacle_reward
# check if the goal has been reached or episode over
if np.linalg.norm(distance_vector) < 0.05:
    done = True
elif np.any(min_obstacle_distances < 0.15):
    done = False
elif self.numberOfSteps > self.episodeLength:
    done = True
else:
    done = False
```

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Linearno kažnjavanje blizine i bez završavanja prilikom udarca u prepreku

eplenmean	eplewmean	misc/total_timesteps
4853.53	-9865.14	1.57e+06
4960.64	-10915.79	2.88e+06
4919.95	-9538.70	4.19e+06
4625.72	-8614.35	5.51e+06
4683.32	-9372.60	6.82e+06
4557.16	-8919.95	8.13e+06
4165.51	-7997.34	9.44e+06
4299.81	-8486.60	1.07e+07
4245.40	-8459.33	1.21e+07
3756.71	-7258.29	1.34e+07
4494.71	-8596.00	1.47e+07
4400.12	-7979.35	1.60e+07
4069.01	-7443.17	1.73e+07
4042.78	-7213.07	1.86e+07
3930.52	-7177.77	1.99e+07

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Eksponencijalno kažnjavanje blizine i bez završavanja prilikom udarca u prepreku

```
observation = self.get_observation()
distance_vector = observation[:3] - self.target
distance_reward = - np.linalg.norm(distance_vector)
progress_reward = - 2 * 300 * (-distance_reward +
                                self.distance_reward_old)
self.distance_reward_old = distance_reward
speed_reward = - np.linalg.norm(action)
min_obstacle_distances = self.min_obstacle_distances()
obstacle_reward = -0.02/(np.min(min_obstacle_distances)**3)
reward = distance_reward + progress_reward +
        speed_reward + obstacle_reward

# check if the goal has been reached
# otherwise check if the episode is over
if np.linalg.norm(distance_vector) < 0.05:
    done = True
elif np.any(min_obstacle_distances < 0.15):
    done = False
elif self.numberOfSteps > self.episodeLength:
    done = True
else:
    done = False
```

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Eksponencijalno kažnjavanje blizine i bez završavanja prilikom udarca u prepreku

eplenmean	eprewmean	misc/total_timesteps
4877.34	-11253.18	1.57e+06
4679.16	-10129.53	2.88e+06
4122.46	-9429.34	4.19e+06
3730.75	-8085.56	5.51e+06
4014.43	-7986.74	6.82e+06
3106.13	-6309.29	8.13e+06
3120.73	-6476.75	9.44e+06
1998.13	-3773.81	1.07e+07
1702.80	-2987.62	1.21e+07
2300.74	-4048.41	1.34e+07
2318.29	-4525.41	1.47e+07
1770.62	-3393.89	1.60e+07
2229.36	-4106.26	1.73e+07
1683.42	-3848.00	1.86e+07
1239.36	-2636.43	1.99e+07

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Eksponencijalno kažnjavanje blizine i bez završavanja prilikom udarca u prepreku

eplenmean	eprewmean	misc/total_timesteps
4661.85	-9838.50	2.88e+06
4137.11	-8440.07	5.51e+06
3791.15	-7692.32	8.13e+06
2962.36	-5753.18	1.07e+07
2715.98	-5039.09	1.34e+07
2221.33	-4388.06	1.60e+07
1667.47	-3272.12	1.86e+07
1704.15	-3254.69	2.12e+07
1323.04	-2676.78	2.39e+07
1401.51	-2744.61	2.65e+07
1364.24	-2430.31	2.91e+07
1106.77	-2208.91	3.17e+07
1154.20	-2399.01	3.43e+07
1256.99	-2446.23	3.70e+07
1100.65	-2358.74	3.96e+07

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Eksponencijalno kažnjavanje blizine i bez završavanja prilikom udarca u prepreku

eplenmean	eprewmean	misc/total_timesteps
4887.21	-16079.80	2.88e+06
4490.02	-12362.30	5.51e+06
4715.55	-17295.69	8.13e+06
4389.97	-14387.87	1.07e+07
3545.18	-10016.69	1.34e+07
3276.90	-10959.46	1.60e+07
2771.84	-7745.69	1.86e+07
2524.67	-8585.83	2.12e+07
2393.39	-7386.62	2.39e+07
2620.81	-8626.77	2.65e+07
2077.17	-6748.79	2.91e+07
2084.14	-7185.85	3.17e+07
1545.79	-5538.39	3.43e+07
1880.94	-6492.34	3.70e+07
1663.24	-5263.44	3.96e+07

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina

potpornog učenja

Podjela algoritama

potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Sadržaj

Uvod

Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

Praktične primjene potpornog učenja

Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

Inverzna kinematika robotske ruke Jaco

Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

Zaključak

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Zaključak

Problem inverzne kinematike riješili smo pomoću dubokog potpornog učenja, no povećanjem težine zadatka, rješavanje problema postaje još teže.

Prije desetak godina, kada je umjetna inteligencija uzela maha, jako mnogo se pričalo o takozvanoj "tehnološkoj singularnosti", točki u vremenu kada će računala poprimiti vlastitu svijest i postati ravnopravna ljudima.

Mišljenja sam da ako će nas išta dovesti do "tehnološke singularnosti", onda je to duboko potporno učenje.

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina potpornog učenja

Podjela algoritama potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog gradijenta politike

Algoritam optimizacije politike uz regije povjerenja

Algoritam proksimalne optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak

Hvala na pažnji!

1 Uvod

2 Teorijske osnove potpornog učenja

Potporno učenje

Teoretska pozadina
potpornog učenja

Podjela algoritama
potpornog učenja

3 Praktične primjene potpornog učenja

4 Algoritmi optimizacije politike

Algoritam jednostavnog
gradijenta politike

Algoritam optimizacije
politike uz regije
povjerenja

Algoritam proksimalne
optimizacije politike

5 Inverzna kinematika robotske ruke Jaco

6 Rezultati

Izlaz agenta

Problem bez prepreka

Problem s preprekama

7 Zaključak