Spodbujevano učenje na impulznih nevronskih mrežah

Matjaž Pogačnik

Univerza v Ljubljani
Fakulteta za računalništvo in informatiko

mp24170@student.uni-lj.si

Mentor: prof. dr. Zoran Bosnić

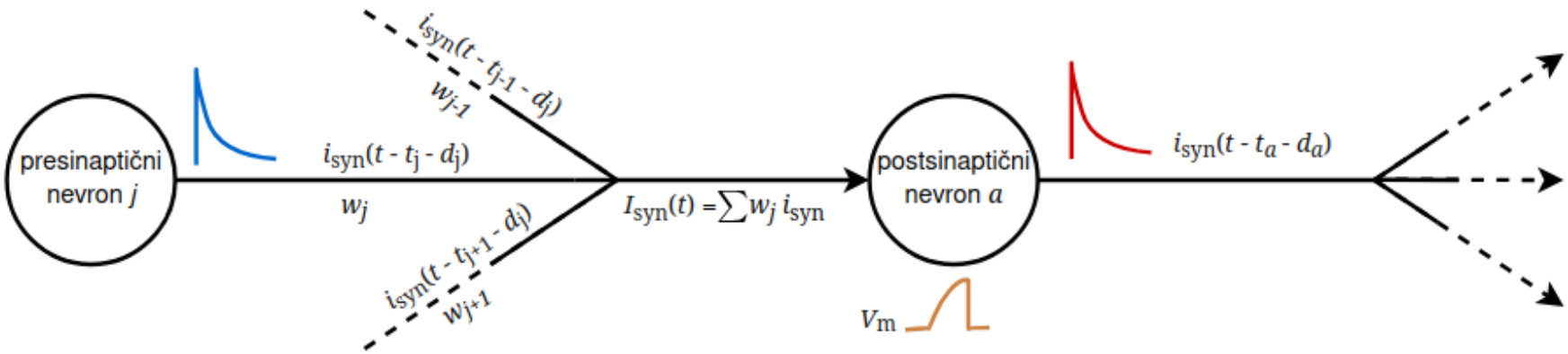
January 20, 2026

Overview

Uvod

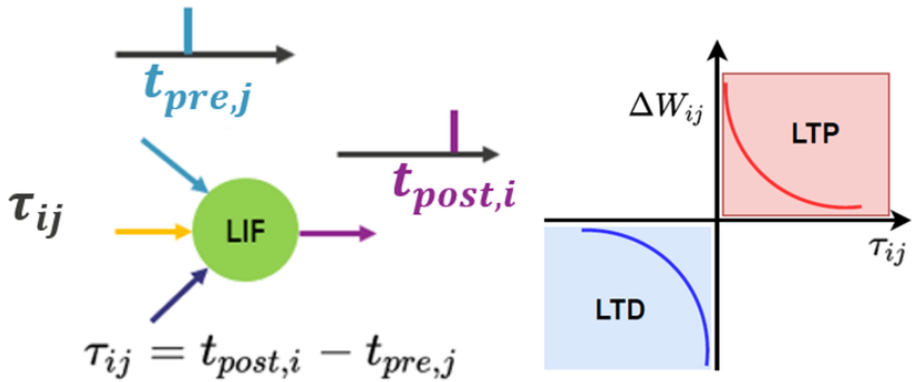
- Impulzne nevronske mreže SNN združujejo čas, energijsko učinkovitost in biološko realističnost.
- Informacija je kodirana v zaporedju in času impulzov.
- Pri ANN čas zanemarjen ali obravnavan v diskretnih korakih.
- Učenje preko lokalnih pravil namesto gradientov.

- **Problem:** kako izvajati spodbujevano učenje na impulznih nevronskih mrežah.
- **Cilj:** razviti biološko smiselno arhitekturo, ki omogoča učenje tudi pri oddaljenih nagradah.



Lokalno pravilo za učenje

- **Problem:** katere povezave so bile odgovorne za proženje izhoda? (problem pripisovanja odgovornosti)
- Sinaptična plastičnost odvisna od časovne razporeditve impulzov (STDP).
- Sledi upravičenosti (eligibility traces)
- **Zakaj STDP?**
  - biološko smiselno,
  - ne potrebuje gradientov,
  - deluje naravno s časom impulzov,

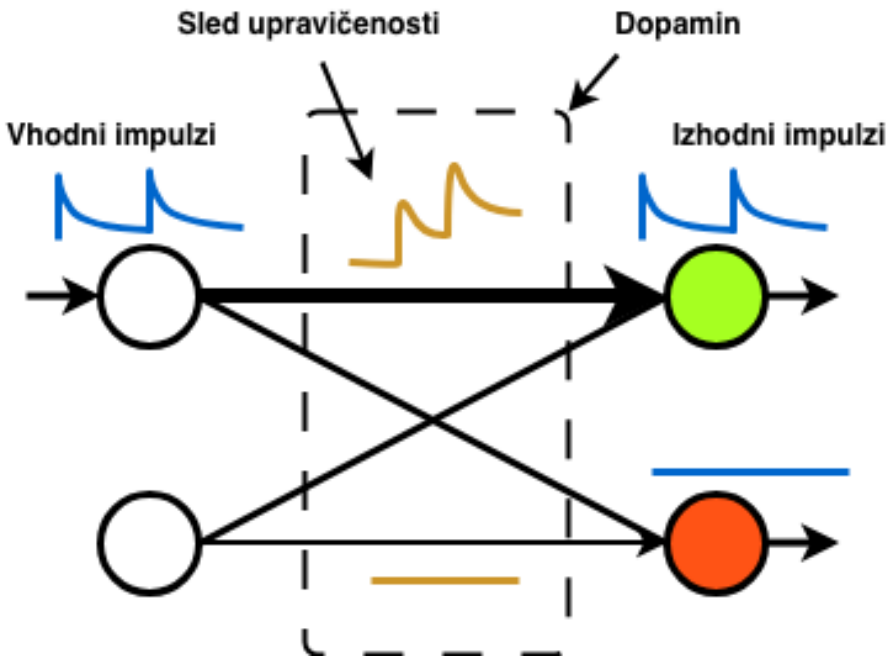


Source: [?]

Nevromodulirana STDP

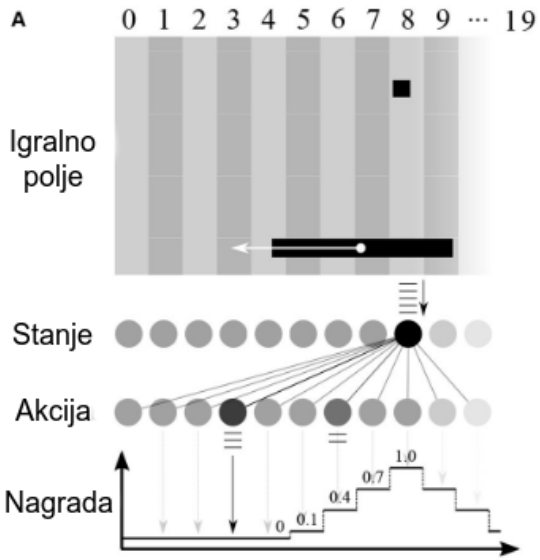
- STDP ne vključuje informacije o nagradi. Razširimo z dopaminsko modulacijo.
- Dopamin globalno modulira plastičnost sinaps (R-STDP).
- Sinapse si zapomnijo preteklo aktivnost (sled upravičenosti).

- **Problem:** če nagrada pride prepozno, R-STDP ne ve več, katera odločitev je bila prava.

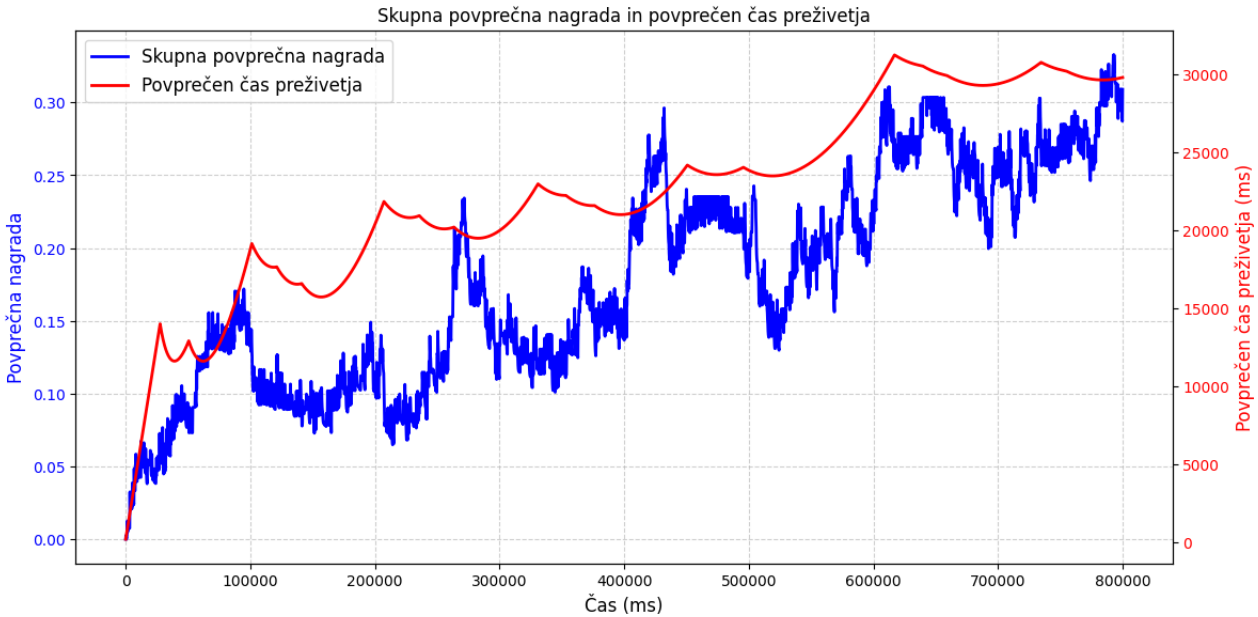


Igra Pong

- Stanje - presinaptični nevron.
- Akcija - postsinaptični nevron.
- Nagrada - koncentracija dopamina.
- Gradient aproksimiramo preko sinaps z visoko sledjo upravičenosti.

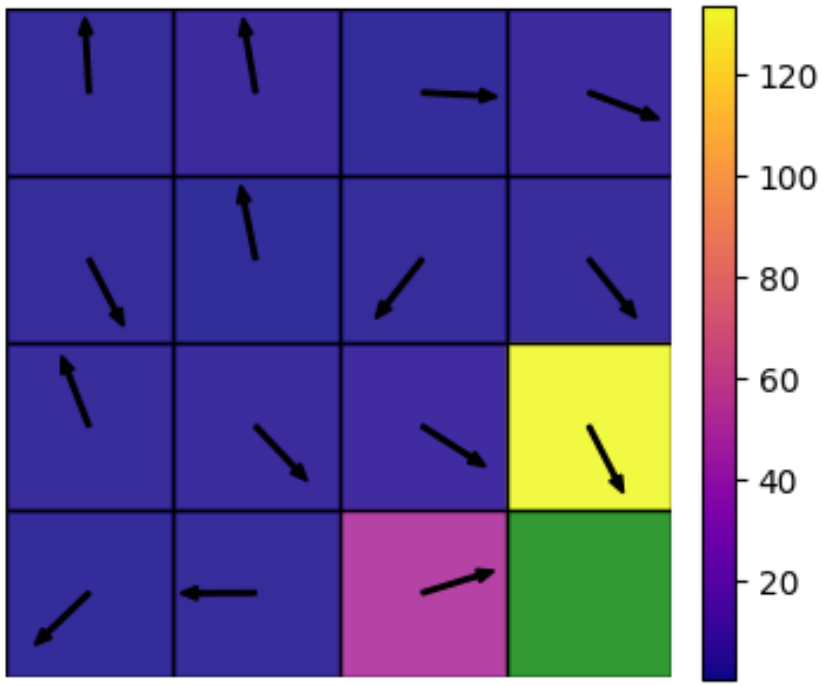


Source: [?]



Mrežni svet

- Verjetnost izbire akcije a v stanju i $\pi(a|i)$ - utež med vhodnim nevronom (stanje) in izhodnim (akcija).
- Rezultat učenja je zvišana sinaptična utež za pravilno akcijo.
- **Problem zakasnjene nagrade.**

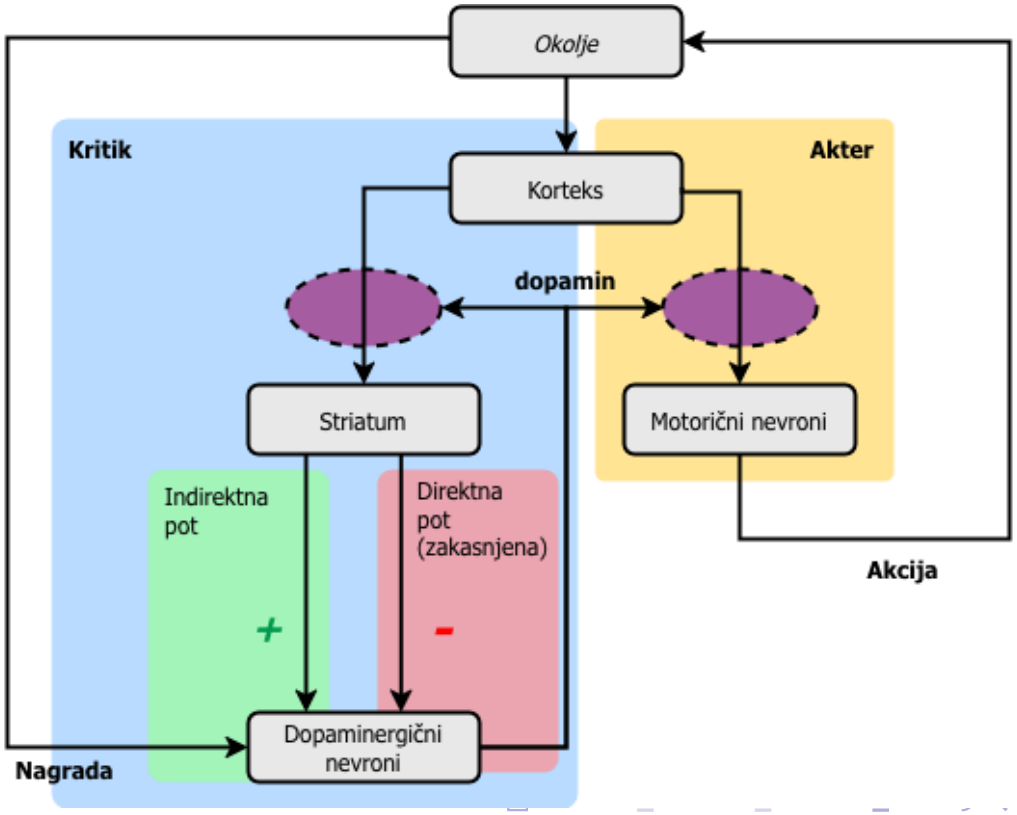


TD-učenje in model akter-kritik

- TD omogoča postopno propagacijo nagrade nazaj skozi stanja.
- Kritik ocenjuje pričakovano nagrado stanja.
- Uporabimo zakasnjene in vzbujajoče/inhibitorne povezave.
- Akter izbira akcije.
- Biološka povezava: bazalni gangliji + dopamin.

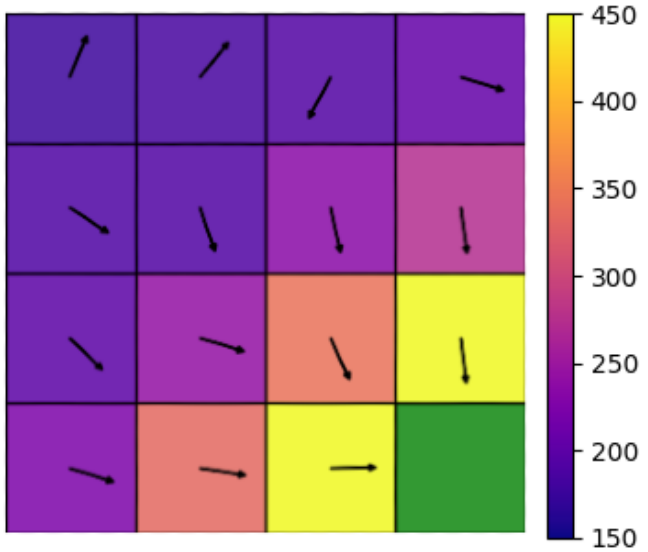
$$V(s_t) \leftarrow V(s_t) + \alpha\,\delta_t$$

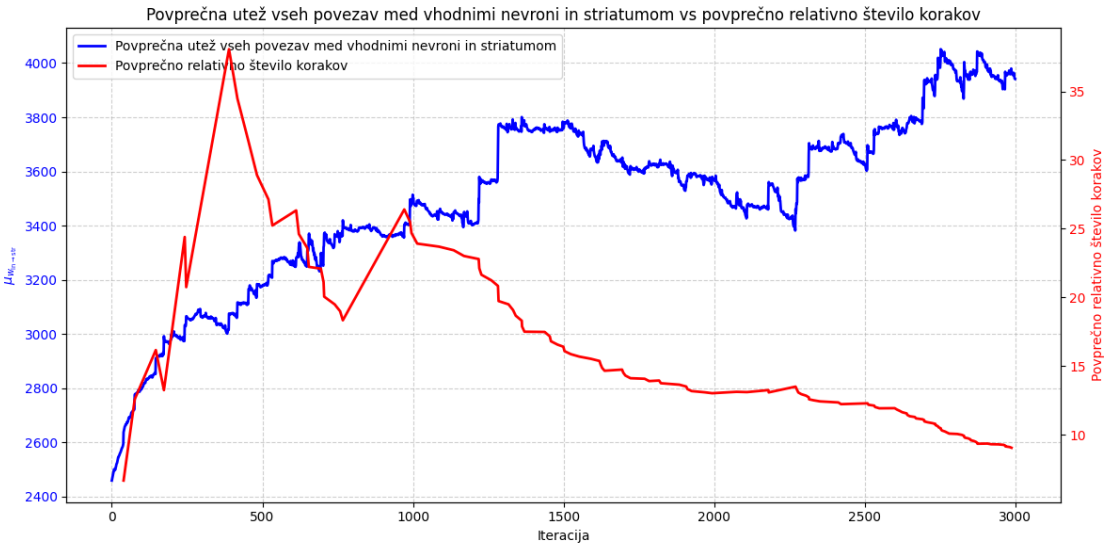
$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$$



Rezultati

- Rezultat učenja je zvišana sinaptična utež za pravilno akcijo in višanje pričakovane nagrade skozi čas.





Zaključek

- Uspešna izvedba spodbujevanega učenja na impulznih nevronskih mrežah.
- Model deluje, vendar je občutljiv na hiperparametre.
- Negativne nagrade niso bile uporabljene (nadaljnje delo: *Izognitveno obnašanje*).
- Vzvratne povezave niso bile uporabljene (nadaljnje delo: *Rekurenčne povezave*).