

Spodbujevano učenje na impulznih nevronske mrežah

Matjaž Pogačnik

Univerza v Ljubljani
Fakulteta za računalništvo in informatiko

mp24170@student.uni-lj.si

Mentor: prof. dr. Zoran Bosnić

January 18, 2026

Overview

1 Uvod

2 R-STDP model

- Igra Pong
- Mrežni svet

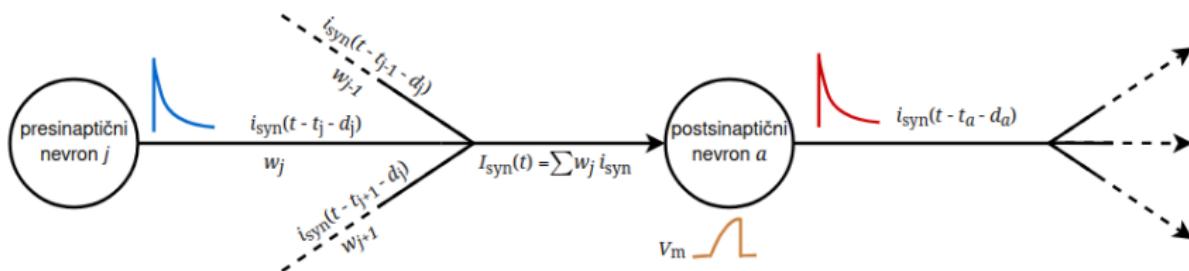
3 TD-učenje in model akter-kritik

- Rezultati

4 Zaključek

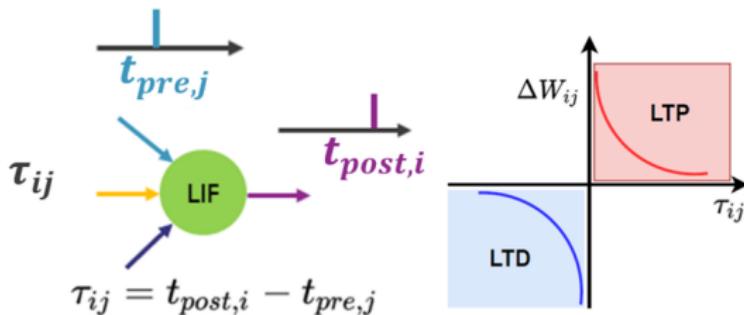
Uvod

- 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 nevronskeih mrežah.
- **Cilj:** razviti biološko smiselno arhitekturo, ki omogoča učenje tudi pri oddaljenih nagradah.



Lokalno pravilo za učenje

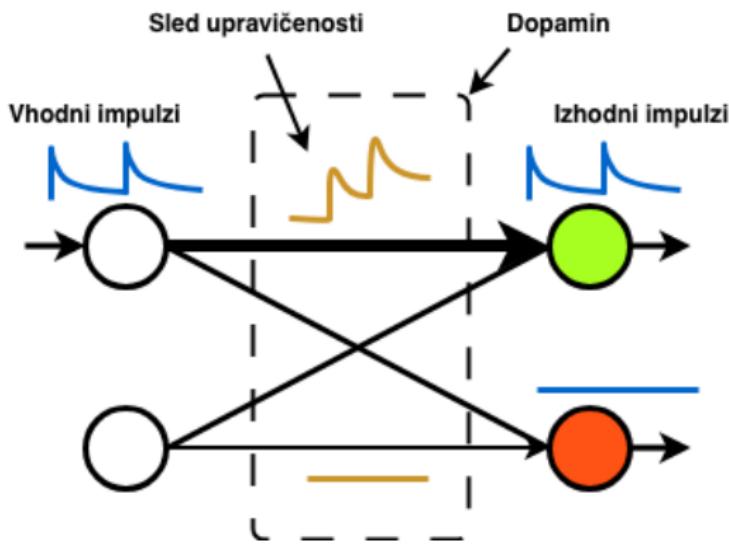
- STDP je lokalno pravilo učenja, kjer se sinaptične uteži spreminjajo glede na relativni čas proženja pre- in postsinaptičnih nevronov.
- **Zakaj STDP?**
 - biološko smiselno,
 - ne potrebuje gradientov,
 - deluje naravno s časom impulzov,
- Ne vključuje informacije o nagradi. Razširimo z dopaminsko modulacijo in kasneje s TD učenjem.



Source: [Safa, 2024]

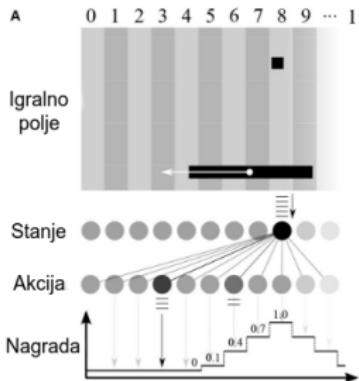
Nevromodulirana STDP

- 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.

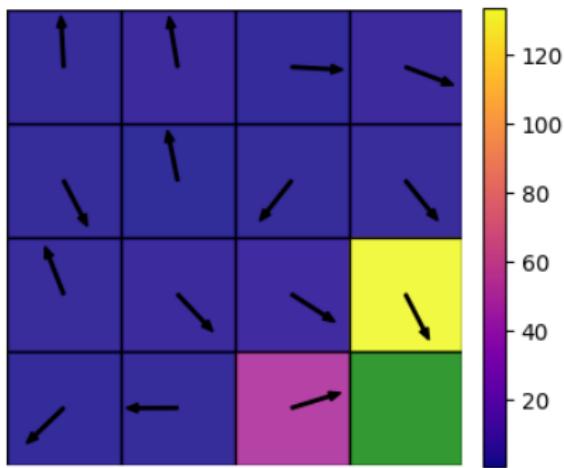


Source: [Wunderlich et al., 2019]



Mrežni svet

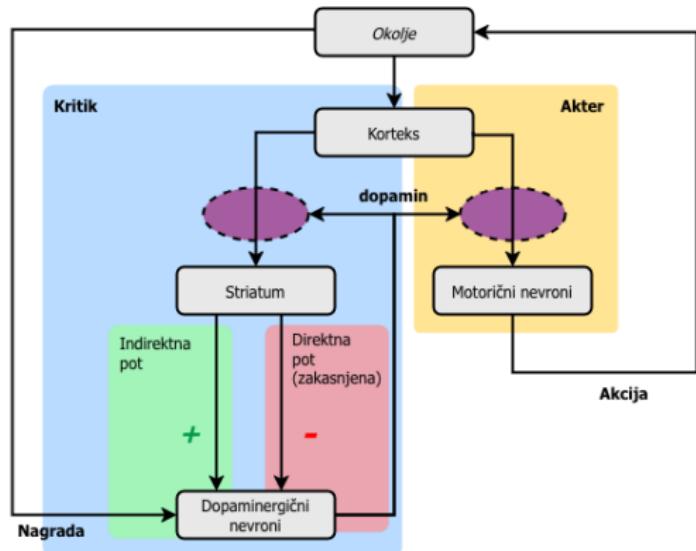
- Verjetnost izbire akcije a v stanju i $\pi(a|i)$ - utež med vhodnim nevronom (stanje) in izhodnim (akcija).
- Rezultat učenja je zvišana sinaptična utež za pravilno akcijo.
- **Problem zakasnjene nagrade.**



TD-učenje in model akter-kritik

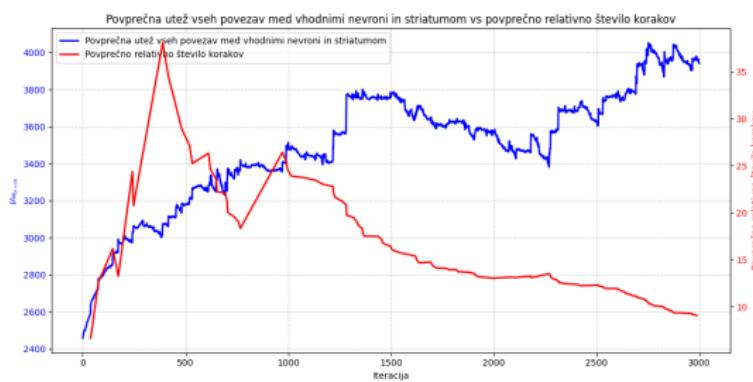
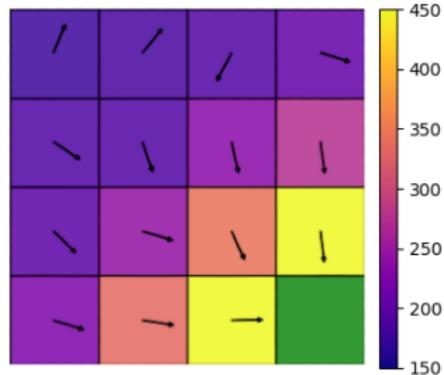
- TD omogoča postopno propagacijo nagrade nazaj skozi stanja,
- Kritik ocenjuje pričakovano nagrado stanja,
- uporabimo zakasnjene in vzbujajoče/inhibitorne povezave.
- Akter izbira akcije,
- biološka povezava: bazalni gangliji + dopamin.

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



Rezultati

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



Zaključek

- Uspešna izvedba spodbujevanega učenja na impulznih nevronskeih 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*).

Reference



A. Safa (2024).

Continual Learning in Bio-plausible Spiking Neural Networks with Hebbian and Spike Timing Dependent Plasticity: A Survey and Perspective.

arXiv preprint.

<https://doi.org/10.48550/arXiv.2407.17305>



IBM Think Blog (2023).

Reinforcement Learning.

<https://www.ibm.com/think/topics/reinforcement-learning>



T. Wunderlich in sod. (2019).

Demonstrating Advantages of Neuromorphic Computation: A Pilot Study.

Frontiers in Neuroscience, 13:260.

<https://doi.org/10.3389/fnins.2019.00260>