

Hluboké sítě a reinforcement learning pro hraní videoher

Bc. Tomáš Trdla <tomas.trdla@tul.cz>, Ing. Karel Paleček, Ph.D.

ABSTRACT

The thesis describes issues of reinforcement learning with use of deep neural networks. The theoretical part contains an introduction to the algorithms of deep reinforcement learning and a more detailed analysis of the Deep Q-Network algorithm, which is then implemented in the practical part for learning an agent to play Atari video games.

ÚVOD

Motivací této práce bylo proniknutí do problematiky Reinforcement Learningu. Videohry zde slouží jako velice vhodný nástroj pro demonstrování schopností RL algoritmů a neuronových sítí, protože se skládají z extrémně komplexního, dynamického, nelineárního a sekvenčního stavového prostoru.

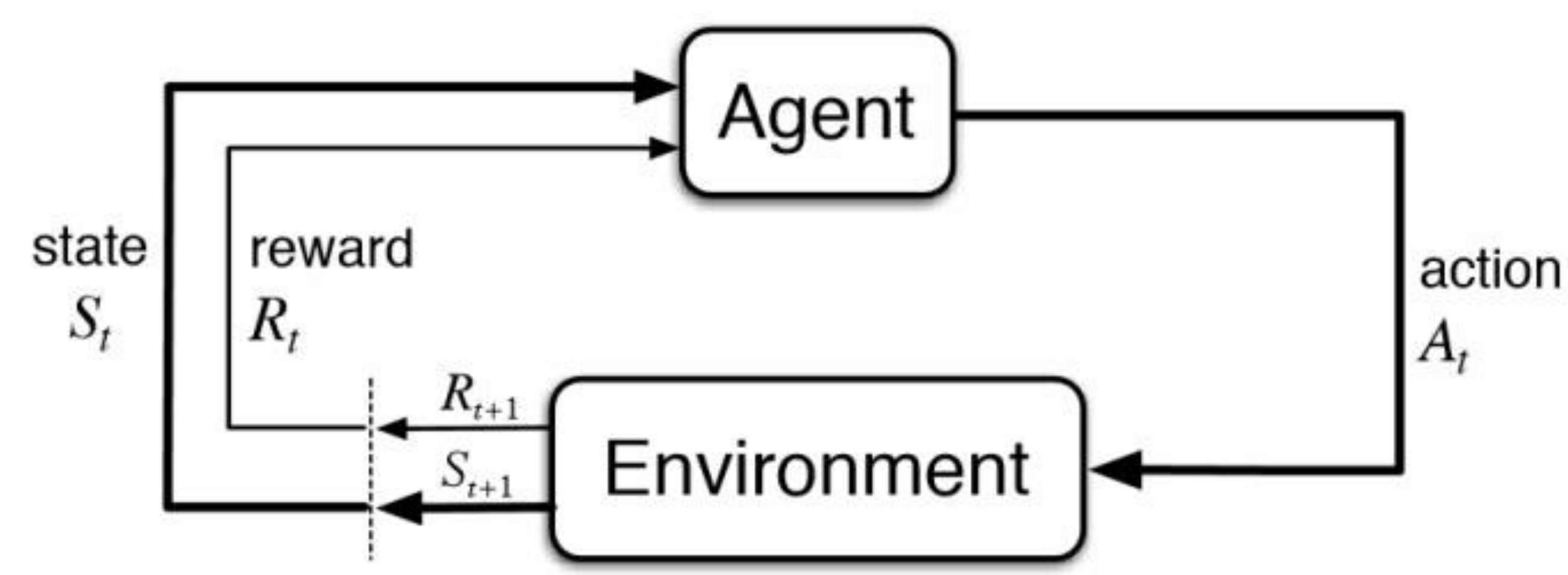
Využití RL lze nalézt například v nejnovějších algoritmech pro automatické hledání konfigurace struktury a hyperparametrů neuronových sítí (Neural Architecture Search), jež je implementováno v AutoML od společnosti Google, nebo různých korporátních výzkumech marketingu, robotiky, chemie a dalších.

METODIKA

Reinforcement learning (RL) [3] je jedna z forem strojového učení, ve které se využívá biologicky přirozené schopnosti organismů se učit z posbíraných zkušeností a RL je tedy umělou adaptací této vlastnosti. Neuronové sítě se zde používají kvůli jejich schopnosti zmapovat zmíněná komplexní a nelineární prostředí. Pro simulování podobných podmínek jako má člověk se zde používají konvoluční NN kvůli jejich aplikovatelnosti v oblasti strojového vidění.

Deep Q-Network (DQN) využívá algoritmus, známý také jako Q-learning [2] pro naučení neuronové sítě zmapovat mnohadimenzionální prostředí a pro každý stav vybrat vhodnou akci, která maximalizuje budoucí odměny.

Existuje několik algoritmů, které zlepšují učicí proces, a díky tomu dokáže umělá inteligence učeného agenta volit o poznání výhodnější strategii rozhodování. Něktěrymi těmito algoritmy jsou např. Double Deep Q-Network, Prioritized Experience Replay, Noisy Nets atd. V současné době je za state-of-art považován algoritmus s názvem Rainbow [1], který je kombinací několika těchto algoritmů.

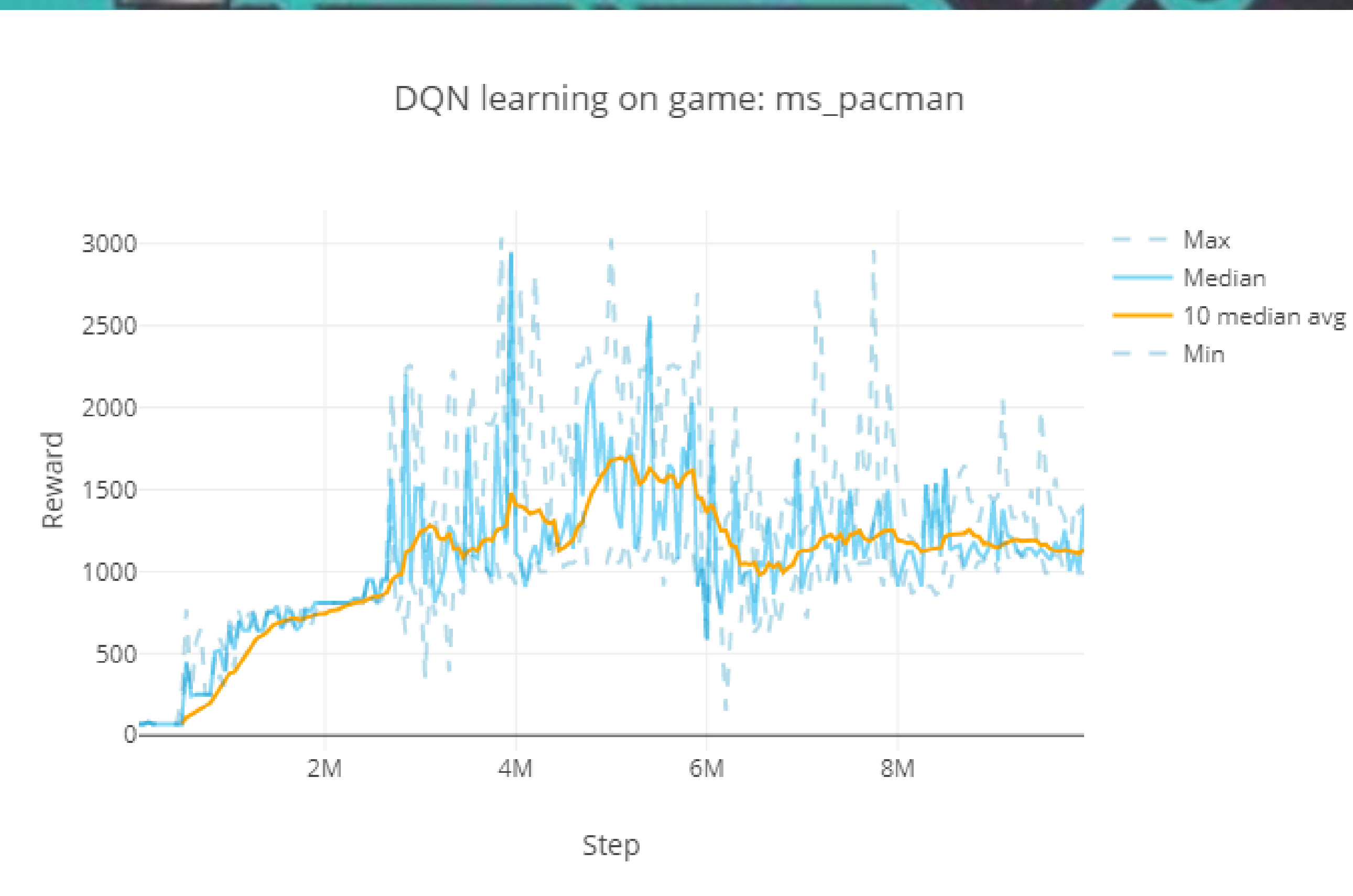


Obrázek 1 – Smyčka algoritmu RL učení

VÝSLEDKY A ZÁVĚR

V rámci tohoto diplomového projektu byl implementován DQN algoritmus v jazyce python a s využitím knihovny PyTorch pro práci s NN. Jako videohry byly zvoleny staré ATARI hry, konkrétně Space Invaders, Pacman, Seaquest a pong. Agenti naučení pomocí DQN algoritmu nejsou schopni se v rámci těchto her rozhodovat natolik kvalitně, aby se dalo mluvit o podobnosti s rozhodováním člověka.

V porovnání s oficiálními výzkumy v této oblasti dosáhl učený agent podobných výsledků ve většině zvolených her. Důvodem pro nezaznamenané známky učení ve hře pong byl mimo nedostatek času s největší pravděpodobností nevhodný návrh hyperparametrů a struktury neuronové sítě. Výsledky by bylo možné zlepšit za pomoci algoritmů pro zlepšení učicího procesu, které v práci byly krátce popsány.



Obrázek 2 – Získané odměny v průběhu učení hry pacman



Obrázek 3 – Screenshots prostředí (zleva pacman, seaquest, space invaders)

REFERENCE

- [1] HESSEL, Matteo, Joseph MODAYIL, Hado HASSELT, et al. Rainbow: Combining Improvements in Deep Reinforcement Learning [online]. 6 Oct 2017 [cit. 26-05-2019]. Dostupné z: <https://arxiv.org/abs/1710.02298>
- [2] Watkins Christopher. - Learning from delayed rewards, PhD. thesis, Cambridge University, 1989
- [3] Online kurz "Reinforcement Learning", <https://eu.udacity.com/course/reinforcementlearning--ud600>