Bc. Marek Adamovič

**Posilňované učenie ako vzdelávací projekt – výskumný zámer**

Diplomová práca

Vedúci práce: Ing. Martin Komák PhD.

December 2022

# Zámer - analýza

Posilňované učenie (reinforcement learning), ako časť strojového učenia, je [1] programovanie a následné učenie agentov formou odmeny a trestu, vďaka čomu (teoreticky) nemusíme predom špecifikovať stratégiu a pravidlá, ako dosiahnuť čo najlepší výsledok v dynamickom prostredí. Agent sa musí sám naučiť, ktoré akcie ho dovedú k najvyššej odmene [5]. Takýto agent musí mať definovaný cieľ (čo má dosiahnuť, čo je odmenou), musí vedieť vnímať stav prostredia, v ktorom sa nachádza a taktiež musí mať k dispozícií akcie, pomocou ktorých ovplyvňuje prostredie.

Na rozdiel od „učenia pod dohľadom“ (supervised learning), ktoré patrí medzi jedno z najviac popísaných foriem učenia [5], nepotrebujeme dáta, ktoré by nám povedali, ktorá akcia je v danom momente najlepšia. Učenie pod dohľadom využíva takéto dáta na natrénovanie agenta, ktorý tieto poznatky následne využíva v situáciách mimo trénovacieho datasetu. Znamená to, že posilňované učenie je „učenie bez dohľadu“ (unupervised learning)? Nie tak celkom, keďže učenie bez dohľadu sa snaží hľadať rôzne vzory v neoznačených dátach (teda dáta nám nehovoria, ktoré akcie sú pre ne správne) Naopak, posilňované učenie nepotrebuje trénovacie dáta, v ktorých by hľadal logické prepojenie a vzory, ale učí sa priamo v prostredí na základe vlastnej skúsenosti. Keďže posilňované učenie nespadá pod učenie pod dohľadom a ani pod učenie bez dohľadu, môžeme ho považovať za tretiu paradigmu strojového učenia.

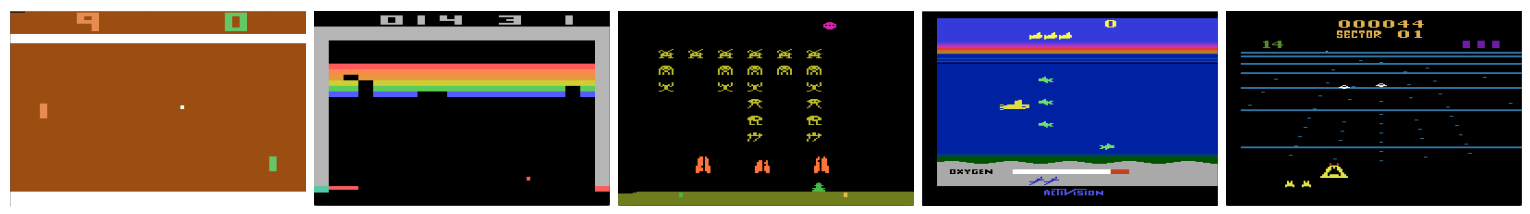
Existujú dve hlavné stratégie [1] pre riešenie problémov, na ktoré sa dá aplikovať posilňované učenie. Prvou je hľadanie najideálnejšieho správania, ktoré dosahuje dobré výsledky v konkrétnom prostredí. Táto stratégia je využívaná genetickými algoritmami a v genetickom programovaní a jej cieľom je najmä optimalizácia riešenia. Druhá stratégia, ktorá je charakteristická pre posilňované učenie, spočíva v odhadovaní užitočnosti jednotlivých krokov, ktoré agent vykoná pomocou štatistických techník a metód dynamického programovania.

Jedným z problémom, s ktorým sa stretávame pri posilňovanom učení [1, 2], je dlhšia odozva pripočítania skóre za vykonanie jednotlivých akcií. To znamená, že agent sa po vykonaní akcie nedozvedá hneď, či bola táto akcia správna alebo nie z pohľadu dosiahnutého skóre.

Ďalším problémom môže byť stochastické (náhodné) prostredie [2], ktoré je definované ako prostredie, ktoré sa ľahko analyzuje, ale ťažko predvída (teda je od povahy náhodné).

Rozšírením posilňovaného učenia je hlboké posilňované učenie (deep reinforcement learning), ktoré využíva viacvrstvové hlboké neurónové siete pre naučenie sa komplexných vzorov v danom prostredí.

Častým príkladom využitia posilňovaného učenia je tvorba agenta, ktorý bude vedieť hrať hry. Týmto využitím sa zaoberá aj štúdia [2], ktorá nechala agenta naučiť sa hrať arkádové hry na základe výstupu (pixelov) z týchto hier. Zaujímavé je, že autori zvolili využitie jednej neurónovej siete pre viacero hier.



Obrázok - ukážka arkádových hier, ktoré sa agent učil hrať

Dokopy sa tento agent učil hrať 7 hier. Pri troch hrách dosiahol dokonca lepšie skóre ako ľudskí experti.

# Plán práce

V prvej fáze projektu plánujeme vytvoriť návrh webovej aplikácie, ktorá bude slúžiť ako vzdelávací projekt v oblasti posilňovaného učenia. To zahŕňa návrh architektúry celej aplikácie a taktiež návrh samotného prostredia („hry“) s agentom, ktorý bude využívať posilňované učenie, prípadne hlboké posilňované učenie využívajúce neurónové siete. V prípade, že by sa jednalo o menej komplexné prostredie, je možné navrhnúť viacero jednoduchších prostredí a následne porovnať ich úspešnosť v rovine vzdelávacích projektov (ktoré prostredia najlepšie vysvetľovali posilňované učenie). Zatiaľ uvažujeme o 2D prostredí podobnému hre Asteroids, kde agent má k dispozícií vesmírnu loď a jeho úlohou je naučiť sa vyhýbať asteroidom, prípadne ich zostreľovať. Taktiež je možné vytvoriť 2 verzie prostredia, kde jedno bude stochastické (náhodné), teda asteroidy budú vždy pri spustení na inej pozícii, inej veľkosti a budú mať inú rýchlosť a smer. Druhé prostredie bude deterministické, teda pri spustení budú všetky parametre vždy rovnaké (použije sa rovnaký random seed). Na týchto dvoch prostrediach vieme následne zhodnotiť úspešnosť agenta a v rámci vzdelávacieho projektu poukázať na to, do akých prostredí je vhodné využiť posilňované učenie.

Druhou fázou projektu, po navrhnutí aplikácie, bude vytvorenie jednoduchého funkčného prototypu prostredia s agentom využívajúcom posilňované učenie podľa návrhu a ich následná evaluácia na menšej vzorke používateľov. Evaluácia bude obsahovať dotazník/test, vďaka ktorému budeme vedieť povedať, či používateľ problematike porozumel a taktiež dostaneme spätnú väzbu. Na základe spätnej väzby upravíme prototyp do finálnej podoby a znovu vykonáme evaluáciu, už na väčšej vzorke používateľov. Na záver práce tieto dáta vyhodnotíme.

# Zdroje

1. <https://www.jair.org/index.php/jair/article/view/10166>
2. <https://arxiv.org/abs/1312.5602>
3. <https://link.springer.com/content/pdf/10.1007/978-3-642-27645-3.pdf>
4. <https://www.javatpoint.com/agent-environment-in-ai>
5. <https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf>