Aplikácia strojového učenia založená na učení posilňovaním pre vybraný problém

Ing. Miloš Foltán, *Školiteľ: Ing. Lukáš Falát PhD.*
Fakulta riadenia a informatiky, Žilinská univerzita v Žiline

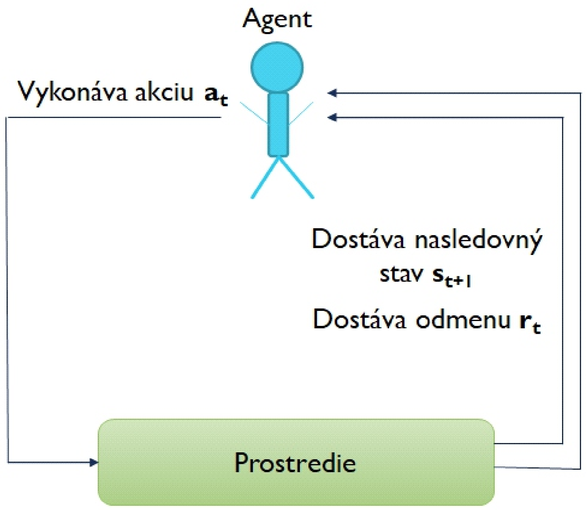
Motivácia

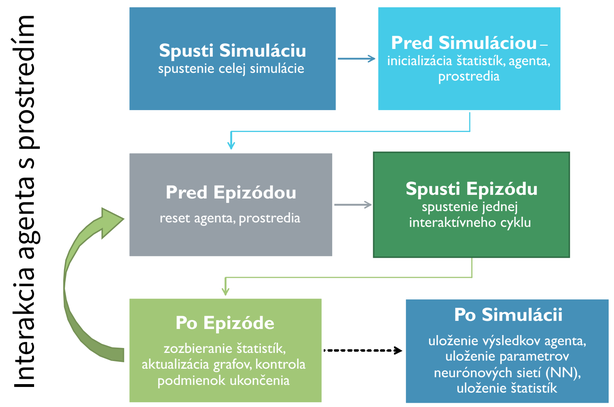
Učenie s posilňovaním (RL) je ďalšou formou umelej inteligencie. Je odklonom od štandardných metód, nakoľko sa pri učení neriadi chybovým ukazovateľom, ale získanou odmenou. Koncept je založený na vzájomnej interakcii s riešeným problémom. Priebeh učenia je teda ľahšie interpretovateľný na základe skúmania akcií a reakcií. Algoritmy založené na RL dokážu nájsť uplatnenie v rôznych simuláciách alebo hrách. Hry, ktoré musia algoritmy riešiť, sú jednoduchšie, ale aj zložitejšie. Pri hraní týchto hier bol skúmaný vplyv nastavenia parametrov na výsledky riešenia.
V aplikácii je možné trénovať viacero implementovaných agentov, nastavovať im parametre, meniť prostredia alebo prezentovať ich dosiahnuté výsledky. Na základe experimentov a empirických testovaní bol vytvorený vylepšený model DDPG. Tento model je otestovaný oproti štandardným modelom a je aplikovaný na prostredí s veľkou dimenziou akcií a stavov.

Ciele práce

- Vyriešenie viacerých typov problémov (hier) s implementáciou viacerých agentov rôznych typov
- Implementácia aplikácie so zobrazovaním sledovaných štatistík
- Experimentálne zlepšenie vlastností a výsledkov agentov
- Výsledky experimentov aplikovať do modelu pre extrémne ťažkú úlohu

Reinforcement Learning (RL - Učenie s posiľňovaním)





Implementované algoritmy Reinforcement Learning

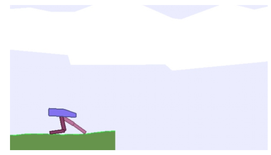
- Q - učenie
- DQN
- Aktér - kritik
- DDPG - Rozšírenie DQN na spojité pomocou AC

Testovacie prostredia

Prostredia, ktoré boli uspôsobené pre potreby RL. Každé z nich predstavuje určitý problém vo forme hry, ktoré je nutné vyriešiť. Všetky sú dostupné na zdroji ***https://gym.openai.com/envs/***

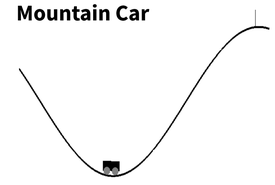
- ***diskrétne***
  - Cart Pole
  - Lunar Lander
  - Mountain car
- ***spojité***
  - Cart Pole
  - Lunar Lander
  - Pendulum

Lunar Lander



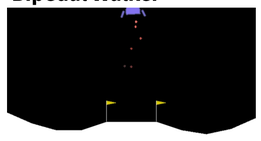
DDPG – 2480 epizód [128 hodín]

Mountain Car



DDPG – 549 epizód [11 minút]

Bipedal Walker



DDPG – 872 epizód [< 2hodiny]

Pendulum

DDPG – 458 epizód [< 1 hodina]

Experimenty

- **Stabilita učiaceho cyklu**
  - Zdrojová – cieľová neurónová sieť
  - Jemné aktualizovanie váh
- **Rýchlosť nájdenia riešenia**
  - Inicializácia váh neurónovej siete
  - Zdieľané parametre
- **Spôsob odmeňovania**
  - Klasické vs alternatívne
- **Hyper – parametre**
  - Koeficient učenia
  - Učiaca dávka
  - Počet neurónov
  - Diskontný faktor

Stabilita učiaceho cyklu

Stabilita učiaceho cyklu

- Zdieľanie parametrov NN
  - *Zrýchlenie v malých prostrediach*
  - *V spojitých môže narobiť problém – preskočenie riešenia*
- Inicializácia váh NN
  - *Žiadny vplyv na rýchlosť*

Spôsob ODMEŇOVANIA

- Klasické odmeňovanie prostredia
  - *Skrýva viac informácii (počítaná podľa polohy agenta k cieľu)*
- Pomalé kladné odmeňovanie
  - *Úlohy kontroly*
- Pomalé záporné odmeňovanie
  - *Podpora agentovej motivácie nájsť cieľ*
- Normovanie odmeny
  - *Stabilnejší gradient*
  - *Zmena agentovho cieľa v úlohe*
  - *Náchylnosť na lokálne minimá (uspokojí sa aj s čiastočným riešením)*

VÝSLEDKY Vylepšeného AGENTA DDPG S MODULMI