Kursinio darbo projektas

Skatinamojo mokymo algoritmų palyginimas

Numerical investigation of Reinforcement Learning algorithms

Arnoldas Čiplys

Darbo tikslas ir uždaviniai

Tikslas:

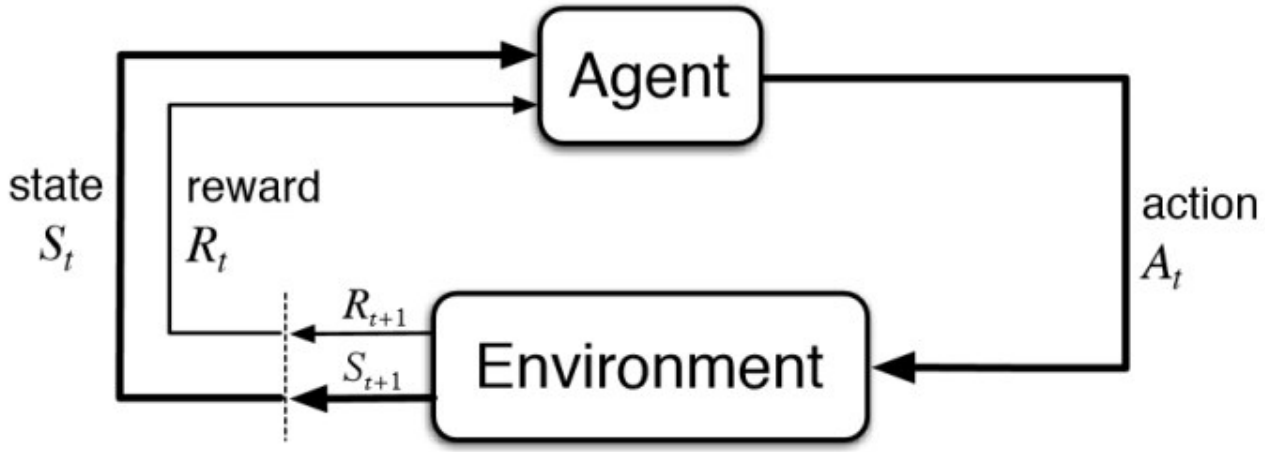
Palyginti skirtingus skatinamojo mokymo algoritmus paprastoje aplinkoje.

Uždaviniai:

- Pasirinkti treniravimo aplinką agentams
- Pasirinkti keletą populiarių skatinamojo mokymo algoritmų, kuriuos lyginsime
- Atlikti eksprerimentus su pasirinktais algoritmais pasirinktoje aplinkoje
- Palyginti gautus rezultatus ir padaryti išvadas

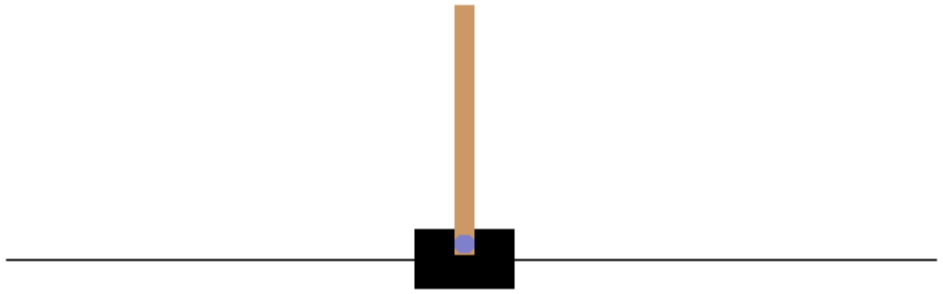
Skatinamasis mokymas

Mašininio mokymosi sritis, kurioje sprendžiama kaip agentas turi elgtis duotoje aplinkoje, kad gautų didžiausią atlygį.

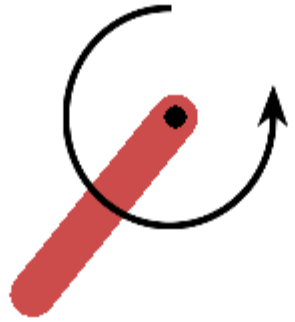


Aplinka

- 2 aplinkos – diskreti (discrete) ir ištisinė (continuous) veiksmų erdvės
- 20 milijonų žingsnių



CartPole

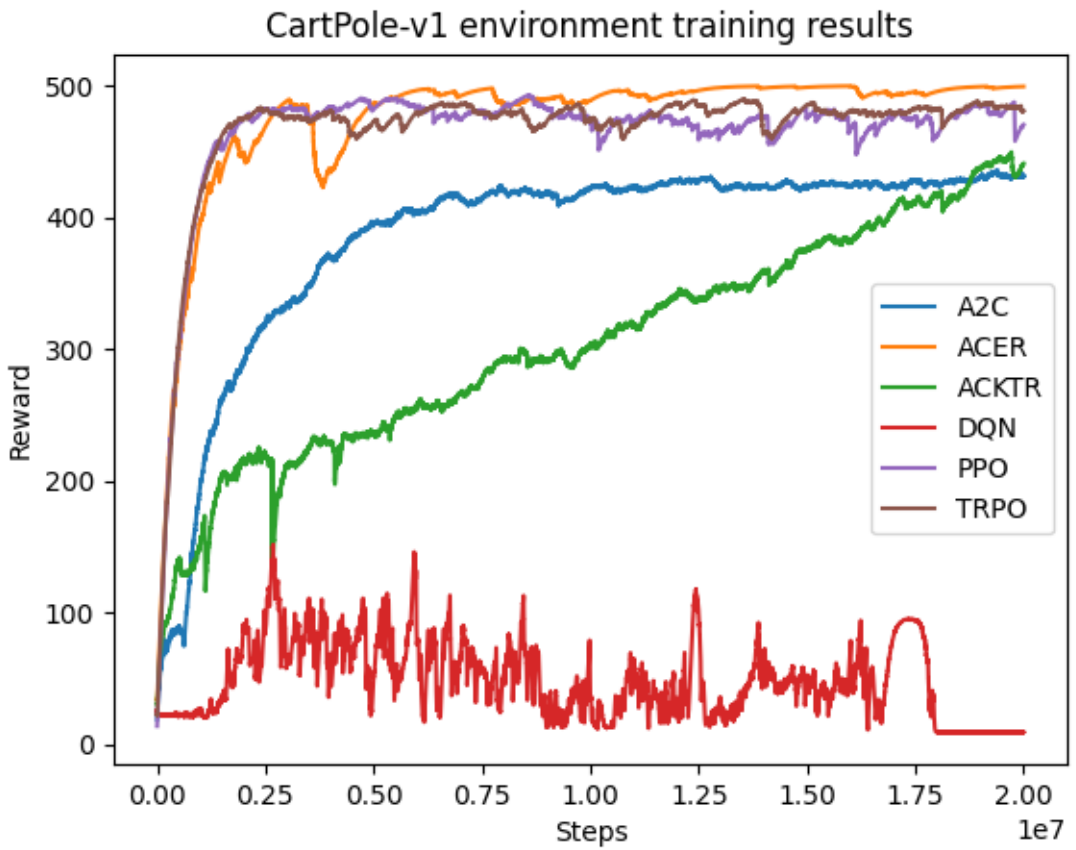


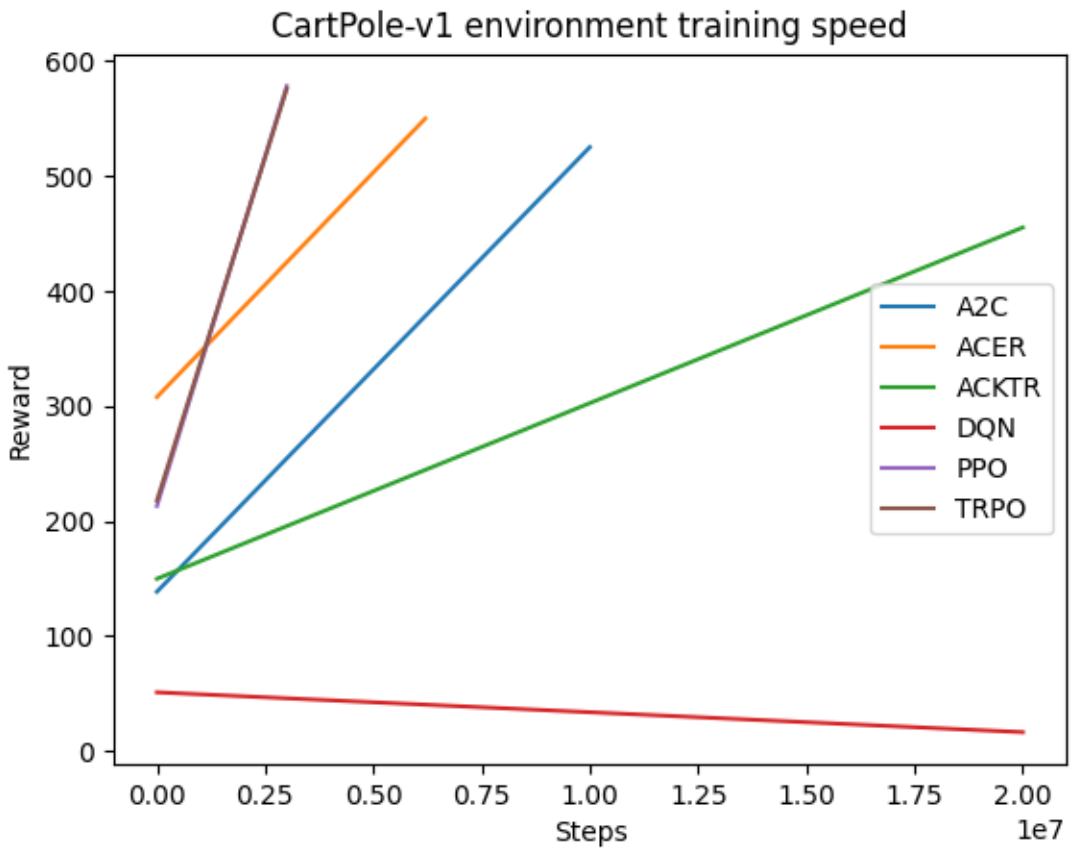
Pendulum

Algoritmai

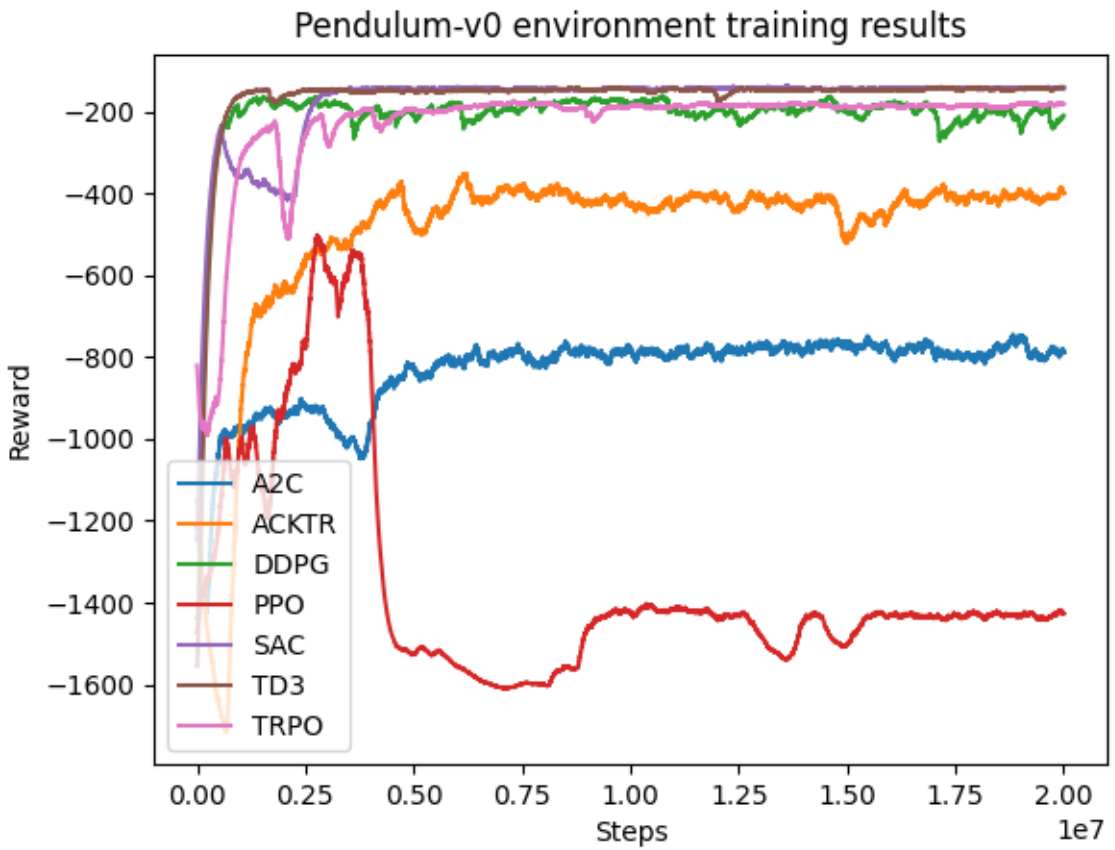
- A2C
- ACER
- ACKTR
- DDPG
- DQN
- PPO
- SAC
- TD3
- TRPO

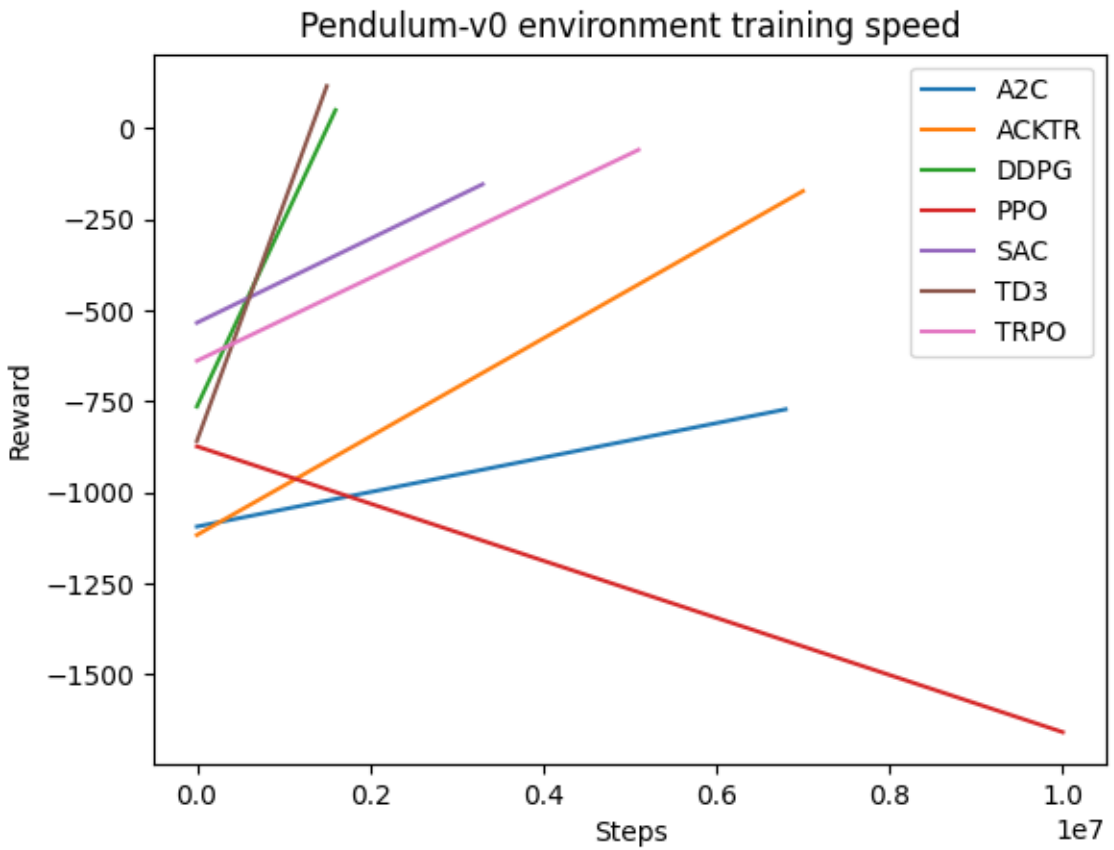
Diskrečios aplinkos rezultatai





Ištisinės aplinkos rezultatai





Išvados

- Algoritmo pasirinkimas turi didelę įtaką galutiniam rezultatui
- Specifinio algoritmo rezultatai gali žymiai skirtis priklausomai nuo aplinkos veiksmų erdvės
- Diskrečioje aplinkoje geriausiai pasirodė PPO, TRPO ir ACER, iš kurių ACER treniravosi šiek tiek lėčiau, bet turėjo geriausią galutinį rezultatą
- Ištisinėje aplinkoje geriausiai pasirodė TD3, SAC, DDPG ir TRPO, iš kurių TD3 ir SAC turėjo šiek tiek geresnius ir pastovesnius rezultatus