VILNIAUS UNIVERSITETAS
INFORMATIKOS INSTITUTAS
PROGRAMŲ SISTEMOS

Skatinamojo mokymosi taikymas žaidimo agento valdymo programos kūrimui

Application of reinforcement learning to the software development for game agent management

Bakalauro baigiamasis darbas

Atliko:	Jokūbas Rusakevičius	(parašas)
Darbo vadovas:	vyresn. m.d. Virginijus Marcinkevičius	(parašas)
Darbo recenzentas:	j. asist. Linas Petkevičius	(parašas)

Vilnius – 2020

Santrauka

TODO: Santrauka

Raktiniai žodžiai: Skatinamasis mokymas, Sokoban žaidimas, aktorius-kritikas based metodai, raktinis žodis 4, raktinis žodis 5

Summary

TODO: summary

Keywords: Reinforcement learning, Sokoban game, actor-critic based methods, keyword 4, keyword 5

TURINYS

ĮVADAS 5

1. TEORIJA 6
   1.1. Skatinamasis mokymasis 6
      1.1.1. Markovo procesas 6
      1.1.2. Sustiprinto mokymosi strategijos ieškojimas 6
      1.1.3. Aktoriaus-kritiko principas 6
   1.2. Gilusis mokymas 6
   1.3. Konvoliuciniai neuroniniai tinklai 6
   1.4. LSTM 6

2. METODOLOGIJA 7
   2.1. Sokoban žaidimas 7
      2.1.1. OpenAI Gym 7
   2.2. Skatinamojo mokymosi bibliotekos parinkimas 7
      2.2.1. Stable Baselines architektūra 7
         2.2.1.1. A2C aprašymas 7
         2.2.1.2. ACER aprašymas 7
         2.2.1.3. POP2 aprašymas 7

3. EKSPERIMENTAI 8
   3.1. Eksperimentinė aplinka 8
      3.1.1. Eksperimentinės aplinkos specifikacijos 8
      3.1.2. Ekseperimentinės aplinkos paruošimas 8
   3.2. Eksperimento planas 8
         3.2.0.1. Pirmas eksperimentas: Geriausios strategijos ieškojimas (*angl. policy*) 8
      3.2.1. Eksperimentas 8

SANTRUMPOS 9

PRIEDAI 9

Įvadas

1. Teorija

1.1. Skatinamasis mokymasis

1.1.1. Markovo procesas

1.1.2. Sustiprinto mokymosi strategijos ieškojimas

1.1.3. Aktoriaus-kritiko principas

1.2. Gilusis mokymas

1.3. Konvoliuciniai neuroniniai tinklai

1.4. LSTM

2. Metodologija

2.1. Sokoban žaidimas

2.1.1. OpenAI Gym

2.2. Skatinamojo mokymosi bibliotekos parinkimas

2.2.1. Stable Baselines architektūra

2.2.1.1. A2C aprašymas

2.2.1.2. ACER aprašymas

2.2.1.3. POP2 aprašymas

3. Eksperimentai

Šiame skyriuje aprašomi bakalauro darbo metu atlikti eksperimentai bei jiems paruošta eksperimentinė aplinka.

3.1. Eksperimentinė aplinka

Eksperimentai atlikti naudojantis realia mašina su „Ubuntu“ OS. Minėtoje mašinoje įdiegta „Anaconda“ paketų valdymo ir dislokavimo sistema, naudojama aplinkų atskyrimui. Didžioji programinė dalis eksperimento atliekama „Jupyter Notebook“ programavimo aplinkoje naudojantis „Python“ kalba.

3.1.1. Eksperimentinės aplinkos specifikacijos

Eksperimentas atliekamas naudojantis realią „Ubuntu“ mašiną.

1. Kompiuterio techninė specifikacija:
   (a) Procesorius - „**Intel Core i5-9600K**“ (6 branduoliai, bazinis greitis 3.70 GHz).
   (b) Grafinė vaizdo plokštė - „**Nvidia GeForce RTX 2070 Super**“.
   (c) Operatyvioji atmintis - „**HyperX Predator Black**“ (**32GB**, 3200MHz, DDR4, CL16).
   (d) Pastovioji atmintis - „**Western Digital**“ (**1TB**).
2. Kompiuterio programinė įranga:
   (a) Operacinė sistema - „**Ubuntu 18.04 LTS**“ (versija: **18.04.4 LTS**).
   (b) Paketų ir aplinkų valdymo sistema - „**Anaconda**“ (versija: **2020.02**).
   (c) Programavimo kalba - „**Python**“ (versija: **3.7.6**).
   (d) Atviro kodo programa kintančio kodo, matematinių funkcijų, teksto bei duomenų vizualizavimui - „**Jupyter Notebook**“ (versija: **6.0.3**).
   (e) ... „**Tensorflow**“ (versija: **1.14.0**).
   (f) ... „**OpenAI Gym**“ (versija: **0.17.1**).
   (g) ... „**Stable Baselines**“ (versija: **2.10.1a0**).

3.1.2. Ekseperimentinės aplinkos paruošimas

3.2. Eksperimento planas

Darbo metu atliktas eksperimentas susideda iš trijų dalių. Šiame skyriuje yra aprašomi šių trijų eksperimentų planai: kaip bus atliekamas eksperimentas, kokia bus naudojama aplinka, kokių rezultatų yra tikimasi ir pan.

3.2.0.1. Pirmo eksperimentas: Geriausios strategijos ieškojimas planas

3.2.1. Eksperimentas

Santrumpos

Darbe naudojamų santrumpų paaiškinimai:

RL – (*angl. Reinforcement Learning*) skatinamasis mokymas.

MDP – (*angl. Markov Decision Processes*) Markovo procesas.