

KOKKUVÕTE

Käesoleva lõputöö eesmärk on võrrelda tugevdatud masinõppe algoritme, autonoomse sõiduki kontrollimise ülesanne lahendamisel. Selleks on loodud 2-mõõtmeline keskkond kus sõiduk peab liikuda eesmärgini ja samal ajal vältima liikuvaid objekte. Andmeid objektidest sai sõiduk punktidenärv koordinaadistikus, analoogiliselt lidari andmetega. Sõiduki kontrollimise algoritm valis kiirendust ja nurkkiirust.

Enne võrdlemise algust valiti sobivad algoritmid rühmad: DQN, Poliitika optimeerimine ja Aktor-Kriitik. Antud rühmadest olid valitud QR-DQN, PPO, TRPO, DDPG, SAC, TD3 ja A2C algoritmid. Võrdluste tulemuste põhjal, moodustavad kasutatud algoritmid 3 rühma. PPO ja A2C ei näidanud positiivseid tulemusi. Teist rühma moodustavad algoritmid TD3, SAC, TRPO õppisid ainult lühiajalisi ülesandeid täitma. Kolmanda rühma mudelid QR-DQN ja DDPG õppisid täitma nii lühi, kui ka pikaajalisi ülesandeid. Objektide vältimine on lühiajaline ja eesmärgini liikumine on pikaajaline ülesanne.

Pikaajaliste ülesannete täitmine on levinud probleem masinõppes ja selle lahendamiseks eksisteerib HER algoritm. Selle abil olid parandatud TD3 ja SAC mudelite tulemused.

Kõige edukam algoritm oli DDPG mis valiti edaspidiseks optimeerimiseks. Optimeerimise käigus tehtud katsete põhjal valiti parim õpisamm, aktiveerimisfunktsioon, mudeli kuju ja muutujate algseisundi initsialiseerija. Samuti tehti katsed mälu neuronitega, aga tulemust see ei parandanud.

Autori arvates said kõik püstitatud ülesanded täidetud ja töö tulemus vastas ootustele. Töö tulemusena selgitati välja, millised algoritmid on perspektiivikad autonoomsete sõidukite kontrollimiseks ja kuidas selliste algoritmide efektiivsust tõsta. Antud uuring on hea alus edaspidisteks eksperimentideks kitsamates valdkondades. Järgmise sammuna võiks tuua keskkonna ja juhtimisalgoritmi tegelikkusele lähemale ning läbi viia uuringud päris sõidukiga.

SUMMARY

The aim of this study is to compare reinforced machine learning algorithms in the case of autonomous vehicle control problem. For this purpose, a 2-dimensional environment has been developed where the vehicle must move to the target and at the same time vehicles should avoid moving objects. The data were acquired from objects as points in an angle coordinate system, analogous to lidar data. The vehicle control algorithm should determine the acceleration and angular velocity.

Prior to the experiment, the appropriate groups of algorithms were selected: DQN, Policy Optimization, and Actor-Critic. The QR-DQN, PPO, TRPO, DDPG, SAC, TD3 and A2C algorithms were selected from these groups. Based on the results of the experimental data, the algorithms used form 3 groups. PPO and A2C did not show positive results. The algorithms TD3, SAC, TRPO forming the second group learned to perform only short-term tasks. The third group models QR-DQN and DDPG learned to perform both short-term and long-term tasks. Herein, avoiding objects is a short-term task and moving to the goal is a long-term task.

Performing long-term tasks is a common problem in machine learning, and it can be solved with the HER algorithm. This improved the results of the TD3 and SAC models.

The most successful algorithm was DDPG, which was chosen for further optimization. Based on the experimental study performed during the optimization, the best learning rate, the activation function, the shape of the model and variable initializer were selected. Experiments with memory units were also performed, but this did not improve the result.

According to the author's opinion, all the set tasks were completed and the result of the work met expectations. As a result of the work, it was found out which algorithms are promising for controlling autonomous vehicles and how to increase the efficiency of such algorithms. This study is a good basis for further experiments in narrower areas. The next step could be to bring the environment and the control algorithm closer to reality and to carry out research with a real vehicle.