WS 2021-2022, SMT-ROB, DISAL-SP155 Project Summary

Simulation-Based Policy Improvement for Automatic Design of Behavioral Arbitrators for Khepera IV Robots

Jiaxuan Zhang

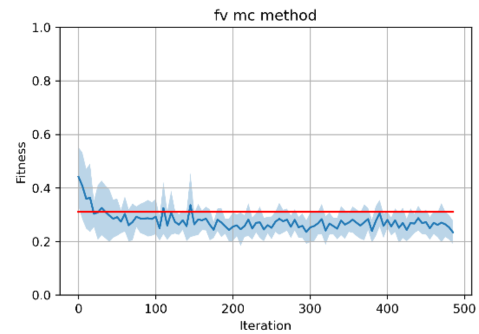
Professor : Alcherio Martinoli
Assistant(s) : Cyrill Baumann

The control components of robots can be expressed as Probabilistic Finite State Machines (PFSM) or Finite State Machines (FSM). How to optimize the structure and tune the parameters of the FSM-based controller remains a big problem. Previous researchers tend to use evolutionary algorithms (EA)., while EAs only use the fitness value from simulations and ignore other information that simulations can provide. Recently, policy improvements methods and reinforcement learning methods are increasingly popular in robotics control problems. Based on Markovian Decision Process (MDP), these methods can learn an optimal policy from trajectories. We believe information like simulation traces may be useful for optimizing the structure of the FSM-based controller. Therefore, we designed two solutions to convert between FSM and MDP. And we tested several classical Simulation-Based Policy Improvement methods on our framework.

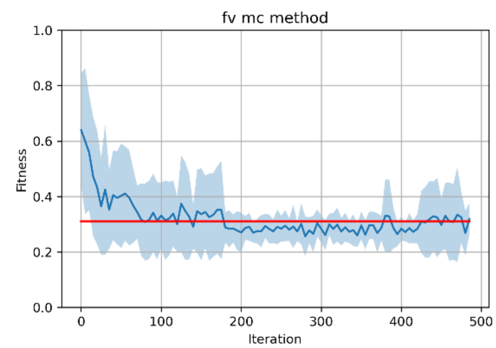
The first stage of the project is to find suitable ways to convert between FSM and MDP. Based on different intuitions, two MDP solutions are proposed. In solution 1, states consist of condition combinations and actions consist of robot behaviors. In solution 2, states consist of behavior tuples (current behavior, next behavior) and actions represent condition combinations. In addition, the process of converting a policy to an FSM and generating traces from simulations are also proposed.

Based on these two frameworks, we have tested several classical SBPI methods: Q-learning, SARSA, First Visit Monte Carlo Method (FVMC), Every Visit Monte Carlo Method (EVMC), and Sample Path Sharing Method. To overcome the conversion solutions' limitations, two specializations are realized: 1. Augmented Initial State, which is used to decide the initial behavior of the FSM. 2. Similarity Mechanism, which aims to use the similarity of condition combinations to speed up the learning process.

Then, experiments are done on the MicroSimulator.



First Visit Monte Carlo Method with Solution 2



First Visit Monte Carlo Method with Solution 1

The Figures show the convergence curves of the First Visit Monte Carlo Method with Solution 1 and Solution 2. These two configurations converge quickly and finally reach a solution very close to the reference FSM controller. The experiments show that FVMC with solution 2 has the most stable performance. It also shows that Monte Carlo methods always have better performance than Q-learning and SARSA in our scenarios. This may be because Monte Carlo Methods may be less harmed by violations of the Markov Properties. However, more experiments in more complex scenarios and some statistical tests are needed to test the effectiveness of these methods. Overall, our test is a quite innovative attempt and the result shows an encouraging convergence trend.