

Lab Report - Reinforcement Learning Project

Janosch Moos · Kay Hansel · Cedric
Derstroff

Received: date / Accepted: date

Abstract Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

Keywords Furuta Pendulum · Cartpole · Inverted Pendulum · Ball Balancer · Classical Control · Artificial Intelligence · Reinforcement Learning · NPG · NES

1 Introduction

The field of Reinforcement Learning is progressing further and further. A lot of algorithms have been published over the past years using various approaches to solve blackbox problems. We have been assigned the task to implement two of those algorithms and verify their performance on different platforms. The purpose of this report is to present our results of the Natural Policy Gradient (npg) and Natural Evolution Strategies (nes) algorithm on the cartpole swing up, the Double Pendulum and the Furuta Pendulum.

1.1 Algorithms

UNSATISFIED

The npg or natural policy gradient algorithm works on the trajectory space using policy gradients to improve. To increase the convergence the KL-divergence is added as a constraint to regularize the gradients direction. This way the "vanilla" gradient becomes the natural gradient. The explorative policy is improved by these gradients of the policy for each action performed during a trajectory. Each episode new trajectories are run with the current policy.

C. Derstroff
E-mail: cedric.derstroff@stud.tu-darmstadt.de

J. Moos
E-mail: janosch.moos@stud.tu-darmstadt.de

K. Hansel
E-mail: kay.hansel@stud.tu-darmstadt.de

In contrast the nes, called natural evolution strategies, utilizes a search distribution over the parameters. Each episode a batch of sample policy is drawn from this distribution. These samples are used to run a trajectory which is evaluated with a fitness function. The distribution is afterwards updated using the samples weighted by their respective fitness. Due to using a distribution over parameters the policy does not need any exploration itself.

Should be rewritten with more detail and maybe optimization function. Add pseudo-code? Add citations!

1.2 Platforms

The algorithms are tested on at least three different platforms. All of them are popular platforms from classical control engineering to demonstrate linear and nonlinear controls.

The cartpole consists of two main parts. The first is a sledge sitting on a rail which can move left and right. Connected to the sledge is a rigid rod swinging freely when the sledge moves. There are several variations of the cartpole problem including stabilization or the swing up of one or more rods where the rods are stacked on top of each other. Two variations of the cartpole had to be solved. The first is the single pendulum swing up and the second the double pendulum stabilization.

The furuta pendulum works a bit differently from the cartpole swing up. Instead from a moving cart the pendulum is connected perpendicular to a link which can be rotated in a horizontal plane by a motor. By doing so the pendulum will swing up.

MODELS?

2 Results

MUHAHAHA THE ALMIGHTY CHAPTER

2.1 Simulation

2.2 Real Platform

References

1. Katsuhisa Furuta, Masaki Yamakita, Seiichi Kobayashi, Swing Up Control of Inverted Pendulum, IEEE, 1991
2. Katsuhisa Furuta, Masaki Yamakita, Seiichi Kobayashi, Swing-Up Control of Inverted Pendulum using Pseudo-State Feedback, Journal of Systems and Control Engineering, 206, 263-269, 1992
3. Andrew Careaga Houck, Robert Kevin Katzschmann, Joao Luiz Almeida Souza Ramos, Furuta Pendulum, MIT, 2013
4. Magnus Gfvert, Modelling the Furuta Pendulum, Lund Institute of Technology, 1998
5. Pavol Seman, Boris Rohal'-Ilkiv, Martin Juhs, Michal Salaj, Swinging up the Furuta Pendulum and its stabilization via model predictive control, Journal of Electrical Engineering, Vol. 64, 152-158, 2013

-
6. Benjamin Seth Cazzolato and Zebb Prime, On the Dynamics of the Furuta Pendulum, Journal of Control Science and Engineering, Vol. 2011, , 2011
 7. Wierstra et. al., Natural Evolution Strategies, Journal of Machine Learning Research, Vol. 15, 949-980, 2014