EVALUATING REINFORCEMENT LEARNING ON ROBOCUP SOCCER SIMULATION 2D

Por

MATEUS GONÇALVES MACHADO

Proposta de Trabalho de Graduação

RECIFE/2019
Resumo

A liga de simulação 2D de futebol do RoboCup é uma das mais maduras da competição, iniciada em 1996. Aprendizagem supervisionada e algoritmos determinísticos são as técnicas mais usadas pelas 5 melhores equipes. Pesquisas recentes usando Aprendizagem Profunda por Reforço (APR) para treinar agentes autônomos em sistemas multiagentes superaram os agentes com base em algoritmos supervisionados ou determinísticos. Um exemplo é a equipe CYRUS2019 produziu jogadores defensivos treinados com o APR, alcançando em terceiro lugar na RoboCup 2019. Este trabalho pretende comparar três técnicas de APR em agentes defensivos com base no CYRUS2019, adaptando o *Half Field Offensive* para ser um ambiente semelhante aos da OpenAI GYM e aplicar a melhor técnica aos agentes do time RoboClIn2d.
Abstract

The Simulation 2D Football league of RoboCup is one of the most matures leagues of the competition having started in 1996. Supervised Learning and deterministic algorithms are the usual techniques used by the TOP 5 teams. Recent researches using Deep Reinforcement Learning to train autonomous agents in multi-agent systems have shown that it outperforms agents based on supervised or deterministic algorithms. The team CYRUS2019 has released defensive players trained with Deep Reinforcement Learning achieving in third place on RoboCup2019. This work intents to compare three DRL techniques for defensive agents based on CYRUS2019 adapting the Half Field Offensive environment to an OpenAI GYM like environment and apply the best technique on RoboCln2d’s agents.
1

Introduction

The Simulation 2D Football league of ROBOCUP (2019) (Sim2d) is one of the most matures leagues of the competition having started in 1996. Supervised Learning and deterministic algorithms are the usual techniques used by the TOP 5 teams. Recent researches using Deep Reinforcement Learning to train autonomous agents in multi-agent systems have shown that it outperforms agents based on supervised or deterministic algorithms. The team CYRUS2019, ZARE et al. (2019), has released defensive players trained with Deep Reinforcement Learning achieving in third place on RoboCup2019. This work intents to compare three DRL techniques for defensive agents based on CYRUS2019 adapting the Half Field Offensive environment to an OpenAI GYM, Brockman et al. (2016), environment and apply the best technique on ROBOCIN (2019) agents.

DeepMind (2014) researchers MNIH et al. (2013) have shown that Deep Q Learning was so effective on Atari games that it outplays human controlled agents. In 2015 Deepmind’s AlphaGo agent, SILVER et al. (2016), won from the European Go Champion Fan Hui and later on 2016 from Lee Sedol, one of the best worldwide. Since then OpenAI and DeepMind has invested a lot of effort on Reinforcement Learning and Autonomous Agents systems. Recently OpenAI (2018) has become the best agent on Dota2 (2011) on 1v1 games and it has a great teamwork performance on 5v5 games winning from most of amateur teams and a few professionals.

HAUSKNECHT (2015) developed an environment to train SARSA (state, action, reward, next state, next action) agents based on the usual OpenAI environments. It provides 2 spaces of states and 3 spaces of actions:

- Low-Level Features and Actions- Uses raw features from Sim2d server and provides raw actions.
- Mid-Level Actions - Uses raw features but some complex and chained actions.
- High-Level Features and Actions - Uses processed features and only complex or chained actions.

Using SILVER et al. (2014) and the Half Field Offensive environment with High Level Features and actions, ZARE et al. (2019) got a reduction of 20% of goals taken. Our intent on this work
is replicate the paper, use two other DRL techniques and optimize the defenders for teams from Brazil to use on the Latin America Robot Competition (2019).
Objectives

The main objective of this work is to analyze the performance of three DRL algorithms on RoboCup’s Simulation 2D environment. Deep Q Network (DQN), MNIH et al. (2013), Dueling Double Q Network (DDQN), WANG; FREITAS; LANCTOT (2015), and Deep Deterministic Policy Gradients (DDPG), SILVER et al. (2014), algorithms were chosen to be compared. This choice was based on ZARE et al. (2019) work which applied DDPG on its defensive agents.

Before the train and test of the algorithms all the environment has to be set doing the following minors objectives:

- Adapt HFO server to our context.
- Prepare the environment to be like OpenAI GYM.
- Choose the features to use on the NNs.
- Creation of RoboCin’s biased dataset without PER.
- Creation of RoboCin’s biased dataset with PER.
3

Schedule

<table>
<thead>
<tr>
<th>Activity</th>
<th>August</th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature Review</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algorithm studies</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Memories creation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiments</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Presentation preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Possíveis Avaliadores

Hansenclever de França Bassani
Assinaturas

O aluno e o orientador assinam abaixo, comprometendo-se com o desenvolvimento do projeto descrito neste documento.

__________________________________________

Tsang Ing Ren (Orientador)

__________________________________________

Mateus Gonçalves Machado (Aluno)
Referências


HAUSKNECHT, M. RoboCup 2D Half Field Offense. 2015. https://github.com/LARG/HFO.


OpenAI. OpenAI Five. 2018.


