Centro de Informática
UFPE

Universidade Federal de Pernambuco
Centro de Informática
Graduação em Engenharia da Computação

Evaluation of time limits in reinforcement learning applied to robot soccer simulation

Student: Juliana do Nascimento Damurie da Silva (jnds@cin.ufpe.br)

Advisor: Hanseclever de França Bassani (hfb@cin.ufpe.br)

Domain: Reinforcement learning/Robotic

14 de Junho de 2021.

1. Abstract

In reinforcement learning, the agent's objective is to optimize the reward accumulated along with an episode. However, we can classify the tasks as episodic tasks and continuous tasks, besides tasks that are naturally limited by time and tasks unlimited by time. All these characteristics impact how we should accumulate these rewards, how we should observe the terminal states, and how we should calculate the value function. Therefore, this work aims to perform an analysis on the impact of agent modeling from the characteristics of the tasks applied in robot soccer, thus extracting information on which are the best modeling for each environment.

2. Resumo

No aprendizado por reforço, o objetivo do agente é otimizar a recompensa acumulada ao longo de um episódio, porém podemos dividir as tarefas em tarefas episódicas e tarefas continuas, além disso em tarefas que são naturalmente limitadas pelo tempo e tarefas ilimitadas pelo tempo. Todas essas características impactam na forma em que devemos acumular essas recompensas, como devemos observar os estados terminais e como devemos calcular a função de valor. Portanto, o objetivo desse trabalho é realizar um análise sobre o impacto da modelagem do agente a partir das características das tarefas aplicadas no futebol de robôs, assim extraindo informações de quais são as melhores modelagem para cada ambiente.

3. Introduction

The Brazilian robotics competition is an event that occurs annually, and one of their categories is the IEEE Very Small Size Soccer, where two teams of three robots up to $80x80x80mm$ play a soccer match. Currently, there is a simulated version of this category in which there are the same challenges as the real robot but with easier access, because we only need a simulator. In this same competition and in the RoboCup, we have the Small Size League category, also known as F-180, where two teams with six robots up to $180mm$ in diameter confront each other in a soccer match. In this category, there is also a simulator.

Reinforcement Learning is a subcategory of machine learning in which learning happens from interactions between the agent and the environment, and the agent learns how to map observations from the environment into actions [1]. Therefore, at each interaction t, the agent sends an action A_t to the environment, the environment performs the action and then returns to the agent a new state S_{t+1} and also a reinforcement signal the reward R_{t+1} that represents the impact of that action on the environment, which can have been positive, negative or neutral.

In the last years, reinforcement learning methods have been in evidence in the literature for their expressive results in the ability to learn complex activities and performing better than humans, such as games like Chess [2], Go [3], Atari [4], Starcraft 2 [5], Dota [6].

In robot soccer simulation, reinforcement learning has been present in RoboCup for years, as we can see in Stone and Sutton [1] application. It is still a challenge to apply these methods in real robots, so we have the creation of simulators and frameworks that can provide environments closer to the real world, and this is a step that is being done in order to apply the simulated methods in real robots, as an example we have the Very Small Size Soccer competition and the Small Size League competition, which are real competitions. However, we see several simulated applications and also works that already perform the transfer of learning [7].

Therefore, as in most traditional machine learning algorithms, the choice of hyper-parameters and the architecture design directly impact the results obtained, in reinforcement learning, it is no different. So, defining and characterizing the task that the agent needs to do and the architecture is essential to obtain an optimal policy. Pardo *et al.* defines in [8] two types of tasks: those that must be optimized in a fixed period of time and those that must be optimized over an indefinite period of time and the time limit is only used during training to diversify the experiments, in this paper as a way to solve these tasks Pardo *et al.* defined two types of agent: time-aware and partial episode bootstrapping, one to solve the problems presented in the time-limited tasks and the other for the time-unlimited tasks.

Therefore, the focus of this work is to apply the methods developed by Pardo *et al.* and variations of this method in the context of robot soccer in order to create different types of agents and conduct a study on their performance and policies created, giving a baseline of which tasks in robot soccer have a better behavior depending on the agent we use.

4. Objectives

The general objective of this work is to study and evaluate robot football environments by simulation and in-game tasks, which are the best methods to be used from the characteristics of these activities and through tests and comparisons of the methods for each possibility of application of the time limit [1] [8].

The specific objectives of this work are:

- Test with the GridWorld environment the results of the agent when we impose time limits and compare with a ground truth.
- Test in the environment of very small size soccer simulation agents of types: standard, time-aware and partial-episode bootstrapping.
- Test in the environment of small size league simulation agents of types: standard, time-aware and partial-episode bootstrapping.
- Compare the results and extract characteristics of each environment, and define which agent provides us with better behavior.

5. Methodology

The literature review will be the first step to be taken in this work, the review will be done on reinforcement learning, reinforcement learning applied to robotics, and robot soccer. In this way we will have a better view of other ways that we can apply our research.

After the literature review, we will choose which methods we will apply to the chosen environments of robot soccer simulation, such as IEEE Very Small Size Soccer and Small Size Çeague. RSoccer Gym, a framework for the Small Size League and IEEE Very Small Size environments, will be used in the tests because it offers us the environment configuration required for the study in reinforcement learning.

We will apply initially the methods proposed by Pardo *et al.* [8], the first method is time awareness for time-limited tasks, where it consists in adding the remaining time in the observation of the environment. The second method proposed is Partial-episode bootstapping for tasks without time limit, which consists of presenting to the agent that episodes that are artificially terminated by the time limit do not represent a terminal state. Then for both cases, we will define which environments and situations in the two available competitions are time-limited or non-time-limited and perform tests in order to evaluate whether we will have advantages using this method instead of the standard one.

As described earlier, the proposed methods do not directly change the agents but rather how the agents view the state endings or the agent's observation state to the environment. Therefore, to train the agents, we will have the possibility to train the agents with Deep Deterministic Policy Gradient (DDPG) and Deep Q-Network (DQN) because they are networks that we already know have a good performance for the framework that we will be using.

6. Schedule

Activity	May	June	July	Aug
Literature review	X	X		
Algorithm study		X		
Experiments		X	X	
Analyse of results			X	
Dissertation writing			X	X
Presentation preparation			X	X

References

[1] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press, 2018.

[2] M. Campbell, A. J. Hoane Jr, and F.-h. Hsu, "Deep blue," *Artificial intelligence*, vol. 134, no. 1-2, pp. 57–83, 2002.

[3] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, *et al.*, "Mastering the game of go without human knowledge," *nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[5] O. Vinyals, T. Ewalds, S. Bartunov, P. Georgiev, A. S. Vezhnevets, M. Yeo, A. Makhzani, H. Küttler, J. Agapiou, J. Schrittwieser, *et al.*, "Starcraft ii: A new challenge for reinforcement learning," *arXiv preprint arXiv:1708.04782*, 2017.

[6] C. Berner, G. Brockman, B. Chan, V. Cheung, P. Dębiak, C. Dennison, D. Farhi, Q. Fischer, S. Hashme, C. Hesse, *et al.*, "Dota 2 with large scale deep reinforcement learning," *arXiv preprint arXiv:1912.06680*, 2019.

[7] H. F. Bassani, R. A. Delgado, J. N. de O. Lima Junior, H. R. Medeiros, P. H. M. Braga, M. G. Machado, L. H. C. Santos, and A. Tapp, *A framework for studying reinforcement learning and sim-to-real in robot soccer*, 2020. arXiv: `2008.12624 [cs.RO]`.

[8] F. Pardo, A. Tavakoli, V. Levdik, and P. Kormushev, *Time limits in reinforcement learning*, 2018. arXiv: `1712.00378 [cs.LG]`.

7. Possible Evaluators

The possible evaluators are the professors:

- Tsang Ing Ren (tir@cin.ufpe.br)
- Cleber Zanchettin (cz@cin.ufpe.br)

8. Signatures

Hansenclever de França Bassani
Advisor

Juliana do Nascimento Damurie da Silva

Juliana do Nascimento Damurie da Silva
Student