Universidade:
presente!
UFRGS
PROPESQ
XXXI SIC
CONHECIMENTO FORMAÇÃO INOVAÇÃO
Salão UFRGS 2019
21. 25. OUTUBRO • CAMPUS DO VALE

| | |
|---|---|
| **Evento** | Salão UFRGS 2019: SIC - XXXI SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS |
| **Ano** | 2019 |
| **Local** | Campus do Vale - UFRGS |
| **Título** | Multiagent Deep Reinforcement Learning Context Detection |
| **Autor** | LUCAS NUNES ALEGRE |
| **Orientador** | ANA LUCIA CETERTICH BAZZAN |

# Multiagent Deep Reinforcement Learning Context Detection

Aluno: Lucas N. Alegre Orientadora: Ana L. C. Bazzan

Instituto de Informática, Universidade Federal do Rio Grande do Sul

Reinforcement learning (RL) algorithms have been successfully applied to solve various sequential decision-making problems for decades. In more recent years, deep reinforcement learning (DRL) has achieved remarkable results by combining RL with modern machine learning methods such as deep neural networks. Following this line, many works have also explored multiagent DRL (MADRL) scenarios, in which several challenges are introduced as multiple agents are interacting at the same time. The goal of a RL agent is to learn a policy--a mapping of states to actions--that maximizes a reward signal by interacting with a dynamic environment. Most algorithms are designed modeling the problem as Markov decision processes (MDPs), in which the environment changes following a fixed state transition function. However, many real world problems have a non-stationary nature, i.e., the environment's current dynamics (context) may unpredictably change over time. This implies that agents may need to learn different policies for different environment contexts, and also that they should be capable of predicting the current context.

In this work, we introduce Multiagent Deep Reinforcement Learning Context Detection (MADRL-CD), an extension of existing methods that also deal with non-stationary environments, but that are only applicable to environments with discrete states; in particular, we extend them to the multiagent case and to problems with continuous states. The method consists in learning different shared dynamics models and policies for each one of the environment contexts, so that each agent can (in real time) detect the current context and then follow and update the corresponding policy. In order to learn the dynamics model of each environment context, we used Mixture Density Networks (MDNs), a special kind of neural networks that can predict the parameters of a conditional density function represented as a mixture of Gaussian distributions. The best policy for each context can be learned using any DRL algorithm. Finally, our model uses the MDNs to predict the next state of the environment, and compare the predictions made by this model with actual next-state observations made in the environment. If these are sufficiently inconsistent, according to a context-change statistic that can be efficiently computed, our method identifies that the environment is now operating under a new context. In this case, our algorithm automatically identifies whether the new context has already been experienced (and is associated with a previously-trained dynamics prediction model and corresponding policy) or whether a new one needs to be created.

We are currently evaluating the proposed method on a suite of single-agent benchmarks of an extension of Gym, a well-known repository of RL environments. This extension provides variations of the default environments that we can define as different contexts. We successfully trained MDNs to predict the probability distribution of the next state given the current state and action taken in the Acrobot task. Also, we successfully learned control policies reproducing state-of-the-art performance for this task using the A2C algorithm. Our experiments indicate that by training context-specific MDNs, these networks are indeed capable of specializing in different contexts of the environment. As next steps, we will apply sequential statistical tests for the detection of the context changes and extend our analyses to multiagent scenarios, such as traffic networks with multiple traffic signal agents.