Recurrent Neural Networks for Reinforcement Learning: an Investigation of Relevant Design Choices
J.T. van Dijk

Abstract

When solving real-world problems like traffic control with reinforcement learning, the observation of the environment is often noisy or simply incomplete. Traditional forms of Q-learning are ill-equipped to handle this kind of problems called Partially Observable Markov Decision Problems (POMDPs). When aiming to solve a POMDP with function approximation a lot of research has made use of recurrent neural networks. With the recent successes of Deep Q-Networks it was inevitable that Deep Recurrent Q-Networks would be introduced. Even though the results of this model are very promising, most researchers make arbitrary design choices when it comes to the recurrent networks. This thesis aims to better investigate such design choices in the context of a traffic control problem. The second contribution of this thesis is a new architecture called Action Motivation DRQN (AMDRQN). This model appends the probability of the previously chosen action to the observation input. This enables an agent to make decisions based on sequences of observation-action pairs but is also creates a new path of gradients over which the parameters can be updated using information about the action probabilities.