

Factored Temporal Difference Learning in the New Ties Environment*

Viktor Gyenes,[†] Ákos Bontovics,[†] and András Lőrincz^{†‡}

Abstract

Although reinforcement learning is a popular method for training an agent for decision making based on rewards, well studied tabular methods are not applicable for large, realistic problems. In this paper, we experiment with a factored version of temporal difference learning, which boils down to a linear function approximation scheme utilising natural features coming from the structure of the task. We conducted experiments in the New Ties environment, which is a novel platform for multi-agent simulations. We show that learning utilising a factored representation is effective even in large state spaces, furthermore it outperforms tabular methods even in smaller problems both in learning speed and stability, because of its generalisation capabilities.

Keywords: reinforcement learning, temporal difference, factored MDP

1 Introduction

Reinforcement learning (RL) [16] is a framework for training an agent for a given task based on positive or negative feedback called *immediate rewards* that the agent receives in response to its actions. Mathematically, the behaviour of the agent is characterised by a *Markov decision process* (MDP), which involves the *states* the agent can be in, *actions* the agent can execute depending on the state, a *state transition model*, and the *rewards* the agent receives.

For small, discrete state spaces well-studied tabular methods exist for solving the learning task. However, real world tasks include many variables, often continuous ones, for which the state space is very large, or even infinite, making these

*This material is based upon work supported partially by the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, under Contract No. FA-073029. This research has also been supported by an EC FET grant, the ‘New Ties project’ under contract 003752. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the European Office of Aerospace Research and Development, Air Force Office of Scientific Research, Air Force Research Laboratory, the EC, or other members of the EC New Ties project.

[†]Eötvös Loránd University, Department of Information Systems, E-mail: gyenesvi@inf.elte.hu, bontovic@elte.hu, andras.lorincz@elte.hu

[‡]Corresponding author