Reinforcement Learning (tutorial talk)

Ann Nowé

Vrije Universiteit Brussel, Belgium
ann.nowe@como.vub.ac.be

Abstract. Reinforcement learning (RL) is an important learning paradigm that can be applied in a wide variety of applications. It allows an agent to learn to behave optimally in an a priori unknown environment, which is generally stochastic. Instead being told what to do. i.e. which action to take in which state, the agent gradually learns which actions to perform in order to optimise a reward signal that is provided by the user.
In this talk I will introduce the basics of single-agent reinforcement learning and the associated theoretical convergence guarantees related to Markov decision processes. I will discuss how the agent?s learning can be improved by incorporating prior knowledge through reward shaping, and how to extend the RL framework to handle multiple objectives simultaneously. We then move from single-agent to multi-agent RL, and outline why convergence guarantees no longer hold in a setting where multiple agents learn. We will explain practical approaches on how to scale single agent reinforcement learning to these situations where multiple agents influence each other and introduce a framework, based on game theory and evolutionary game theory, that allows to analyse the dynamics of multi-agent learning. The theoretical concepts with be illustrated with applications.

Fundamentals of reinforcement learning

The tutorial starts with an introduction of the program and the speakers. The remainder of the first slot will be spent on the basics of reinforcement learning (RL), forming the foundation for the extensions discussed later on. Within the topic of RL we discuss Markov decision processes (MDPs), policy and value iteration (including algorithms such as Learning Automata and Q-learning), and we briefly outline various well-know extensions to the basic algorithms such as eligibility traces.

Reward shaping

Reward shaping is a method to incorporate external knowledge in reinforcement learning. The attendees will be famil- iarized with one of the big challenges of reinforcement learning: its often prohibitively large sample complexity. Reward shaping is introduced as an approach to addressing this problem by using external knowledge to guide a learning agent?s exploration. Various knowledge sources and how they can be used will be discussed: expert knowledge, experiences in previous tasks (transfer), (human) demonstrations/advice, etc.

Multi-objective RL

As many real-world problems involve the optimization of multiple, possibly conflicting objectives we will also provide an overview of the state-of-the-art of multi-objective reinforcement learning (MORL). We will discuss several approaches including adaptive linear scalarisation methods which are suitable if the Pareto front is convex. However, when the Pareto front can not be assumed to be convex, more advanced methods are needed which through clever bootstrapping allow to learn all policies that belong to the Pareto front in parallel.

From single-agent to multi-agent RL

Here we discuss how RL can be applied in a multi-agent setting (MARL). The presence of other learning agents gives rise to partial observability and a non-stationary environment, violating the assumptions on which traditional RL algorithms are based. These key challenges motivate the need for a new theoretical framework to study MARL. In particular, we discuss Markov games as a multi-agent extension of MDPs, as well as three approaches to learning in this extended setting: independent learning, joint action learning, and gradient based methods.

Evolutionary dynamics of multi-agent RL

Evolutionary game theory (EGT) provides tools for the analysis of strategic interactions between multiple agents. We dis- cuss (normal form) games, best response sets, Nash equilibria, Pareto optimality, replicator dynamics, and evolutionarily stable strategies. The replicator dynamics are formally linked to MARL and form the basis of the evolutionary framework for studying learning dynamics of MARL.