Structured and Continuous Reinforcement Learning

Abstract

In *reinforcement learning*, an agent tries to learn optimal behavior in an unknown environment by evaluating feedback – usually some quantifiable and comparable reward – to his actions. As the learner's actions may pay off not immediately, he must be able to learn also from delayed feedback, for example by accepting short-term discouraging feedback to achieve a long-term goal giving large positive feedback. Thus, in typical reinforcement learning applications like robotics, control, or game playing, the learner will get rewarding feedback only when a given task is finished after a series of coordinated actions which individually give no or even misleading feedback. While various reinforcement learning algorithms have been developed, these methods have been denied a major breakthrough in practice. One of the major problems with application of reinforcement learning algorithms to real world problems is that typical algorithms are not efficient in large domains. Thus, while many potential applications could be handled by reinforcement learning algorithms in principle, from the practical point of view they are too costly, as their complexity and *regret* (the total lost reward with respect to an optimal strategy) grow linearly or even polynomially with the size of the underlying domain. One of the reasons for this is that – unlike humans – reinforcement learning algorithms usually are not able to exploit similarities and structures in the domain of a problem.

In a precursor project, together with scientists from the SequeL team at Inria Lille, an interdisciplinary center for reinforcement learning, we were able to define very general similarity structures for reinforcement learning problems in finite domains and to achieve improved theoretical regret bounds when the underlying similarity structure is known. The developed techniques and algorithms also led to the first theoretical regret bounds for reinforcement learning in continuous domains. The proposed project wants to take the research on continuous reinforcement learning – a setting which is of particular importance for applications – a step further, not only by improving over the known bounds, but also by the development of efficient algorithms. Moreover, we also want to investigate in more general settings where the learner does not have direct access to the domain information, but only to a set of possible models. Also for this setting, the precursor project has produced first theoretical results, assuming finite domains and that the set of possible models contains the correct model. In the proposed project, we aim at generalizing this to infinite domains and loosening the assumption on the model set, which shall not necessarily contain the correct model, but only a good approximation of it.