This document lists mandatory literature and the topics that are going to be elaborated by the students of the project group “Learning Agents in Dynamic Environments II”. The list gives abstracts of the cited papers to provide a first insight into each topic. Full papers are available online from within the university-wide network.

During our first meeting we will briefly present these topics before we assign one topic to each student. The listed works should be understood as a starting point for your own literature research in the designated area.

1 Organization

- Assignment of topics: 14/10/2011, 9:15h, Room O4:267
- Meet your supervisor until 28/10/2011
- Hand in the structure of your seminar paper: 09/11/2011
- Length of seminar paper: 12 pages
- Deadline for the final seminar paper: 24/11/2011
- The final papers then are distributed among the students for a double-blind peer-review on 25/11/2011
- Deadline for reviews: 30/11/2011
- The seminar will take place on one or two days between 07/12/2011 and 13/12/2011.
- Presentation: 30 min. + 10 min. discussion
- Deadline for revised version of the final seminar papers including comments from the reviewers and discussions from presentation: 18/12/2011
Hand in your texts via eMail to your supervisor until 11:59 PM of the respective date

2 Readings

Since all topics deal with reinforcement learning (RL) resp. multiagent reinforcement learning (MARL), we assume that each student reads thoroughly the following texts:

- For a general introduction in RL: Chapter 13 of Tom Mitchell’s book [Mit97].
- For general introduction in MARL: [BBDS08]

Additionally to these mandatory texts, there are other well known works, including:

- Comprehensive book on RL: [SB98]
- Survey on multiagent learning: [PL05]

3 Topics

Single Agent Reinforcement Learning [Sze10]:

Reinforcement learning is a learning paradigm concerned with learning to control a system so as to maximize a numerical performance measure that expresses a long-term objective. What distinguishes reinforcement learning from supervised learning is that only partial feedback is given to the learner about the learner’s predictions. Further, the predictions may have long term effects through influencing the future state of the controlled system. Thus, time plays a special role. The goal in reinforcement learning is to develop efficient learning algorithms, as well as to understand the algorithms’ merits and limitations. Reinforcement learning is of great interest because of the large number of practical applications that it can be used to address, ranging from problems in artificial intelligence to operations research or control engineering. In this book, we focus on those algorithms of reinforcement learning that build on the powerful theory of dynamic programming. We give a fairly comprehensive catalog of learning problems, describe the core ideas, note a large number of state of the art algorithms, followed by the discussion of their theoretical properties and limitations.

See also: e.g. [SB98]

Bayesian Multiagent Reinforcement Learning [CB03]: Much emphasis in multiagent reinforcement learning (MARL) research is placed on ensuring that MARL algorithms (eventually) converge to desirable equilibria. As in standard reinforcement learning, convergence generally requires sufficient exploration of strategy space. However, exploration often comes at a price in the form of penalties or foregone opportunities. In multiagent settings, the problem is exacerbated by the need for agents to “coordinate” their policies on equilibria. We propose a Bayesian model
for optimal exploration in MARL problems that allows these exploration costs to be weighed against their expected benefits using the notion of value of information. Unlike standard RL models, this model requires reasoning about how one’s actions will influence the behavior of other agents. We develop tractable approximations to optimal Bayesian exploration, and report on experiments illustrating the benefits of this approach in identical interest games.

**Reward Shaping** [DK11]: Potential-based reward shaping has previously been proven to both be equivalent to Q-table initialisation and guarantee policy invariance in single-agent reinforcement learning. The method has since been used in multi-agent reinforcement learning without consideration of whether the theoretical equivalence and guarantees hold. This paper extends the existing proofs to similar results in multi-agent systems, providing the theoretical background to explain the success of previous empirical studies. Specifically, it is proven that the equivalence to Q-table initialisation remains and the Nash Equilibria of the underlying stochastic game are not modified. Furthermore, we demonstrate empirically that potential-based reward shaping affects exploration and, consequentially, can alter the joint policy converged upon.

**Cooperative Multiagent Reinforcement Learning** [LR00]: The article focuses on distributed reinforcement learning in cooperative multiagent-decision-processes, where an ensemble of simultaneously and independently acting agents tries to maximize a discounted sum of rewards. We assume that each agent has no information about its teammates’ behaviour. Thus, in contrast to single-agent reinforcement-learning each agent has to consider its teammates’ behaviour and to find a cooperative policy. We propose a model-free distributed Q-learning algorithm for cooperative multi-agent-decision-processes. It can be proved to find optimal policies in deterministic environments. No additional expense is needed in comparison to the non-distributed case. Further there is no need for additional communication between the agents.

**Transfer Learning** [TS09]: The reinforcement learning paradigm is a popular way to address problems that have only limited environmental feedback, rather than correctly labeled examples, as is common in other machine learning contexts. While significant progress has been made to improve learning in a single task, the idea of transfer learning has only recently been applied to reinforcement learning tasks. The core idea of transfer is that experience gained in learning to perform one task can help improve learning performance in a related, but different, task. In this article we present a framework that classifies transfer learning methods in terms of their capabilities and goals, and then use it to survey the existing literature, as well as to suggest future directions for transfer learning work.

**Credit Assignment Problem** [AT06]:

Coordinating multiple agents that need to perform a sequence of actions to maximize a system level reward requires solving two distinct credit assignment prob-
lems. First, credit must be assigned for an action taken at time step $t$ that results in a reward at time step $t' \neq t$. Second, credit must be assigned for the contribution of agent $i$ to the overall system performance. The first credit assignment problem is typically addressed with temporal difference methods such as Q-learning. The second credit assignment problem is typically addressed by creating custom reward functions. To address both credit assignment problems simultaneously, we propose the “Q Updates with Immediate Counterfactual Rewards-learning” (QUICR-learning) designed to improve both the convergence properties and performance of Q-learning in large multi-agent problems. QUICR-learning is based on previous work on single-time-step counterfactual rewards described by the collectives framework. Results on a traffic congestion problem shows that QUICR-learning is significantly better than a Qlearner using collectives-based (single-time-step counterfactual) rewards. In addition QUICR-learning provides significant gains over conventional and local Q-learning. Additional results on a multi-agent grid-world problem show that the improvements due to QUICR-learning are not domain specific and can provide up to a ten fold increase in performance over existing methods.

See also: e.g. [AT04]

Models used in RL/MARL [SZ08]: Over the last 5 years, the AI community has shown considerable interest in decentralized control of multiple decision makers or “agents” under uncertainty. This problem arises in many application domains, such as multi-robot coordination, manufacturing, information gathering, and load balancing. Such problems must be treated as decentralized decision problems because each agent may have different partial information about the other agents and about the state of the world. It has been shown that these problems are significantly harder than their centralized counterparts, requiring new formal models and algorithms to be developed. Rapid progress in recent years has produced a number of different frameworks, complexity results, and planning algorithms. The objectives of this paper are to provide a comprehensive overview of these results, to compare and contrast the existing frameworks, and to provide a deeper understanding of their relationships with one another, their strengths, and their weaknesses. While we focus on cooperative systems, we do point out important connections with game-theoretic approaches. We analyze five different formal frameworks, three different optimal algorithms, as well as a series of approximation techniques. The paper provides interesting insights into the structure of decentralized problems, the expressiveness of the various models, and the relative advantages and limitations of the different solution techniques. A better understanding of these issues will facilitate further progress in the field and help resolve several open problems that we identify.

See also: e.g. [KKB11]

State space representations [PTT10]: In this paper we survey the basics of reinforcement learning, generalization and abstraction. We start with an introduction to the fundamentals of reinforcement learning and motivate the necessity for gen-
eralization and abstraction. Next we summarize the most important techniques available to achieve both generalization and abstraction in reinforcement learning. We discuss basic function approximation techniques and delve into hierarchical, relational and transfer learning. All concepts and techniques are illustrated with examples.

See also: e.g. [MGLF11]

Function Approximation [PDJ00]: Many mobile robot tasks can be most efficiently solved when a group of robots is utilized. The type of organization, and the level of coordination and communication within a team of robots affects the type of tasks that can be solved. This paper examines the tradeoff of homogeneity versus heterogeneity in the control systems by allowing a team of robots to coevolve their high-level controllers given different levels of difficulty of the task. Our hypothesis is that simply increasing the difficulty of a task is not enough to induce a team of robots to create specialists. The key factor is not difficulty per se, but the number of skill sets necessary to successfully solve the task. As the number of skills needed increases, the more beneficial and necessary heterogeneity becomes. We demonstrate this in the task domain of herding, where one or more robots must herd another robot into a confined space.

References


