Cooperative Multi-Agent Systems from the Reinforcement Learning Perspective — Challenges, Algorithms, and an Application

Dagstuhl Seminar on
"Algorithmic Methods for Distributed Cooperative Systems"

Thomas Gabel

Machine Learning Lab
Department of Computer Science
University of Freiburg
Georges-Köhler-Allee 079
79110 Freiburg im Breisgau
Germany
tgabel@informatik.uni-freiburg.de

Abstract. Reinforcement Learning has established as a framework that allows an autonomous agent for automatically acquiring – in a trial and error-based manner – a behavior policy based on a specification of the desired behavior of the system. In a multi-agent system, however, the decentralization of the control and observation of the system among independent agents has a significant impact on learning and it complexity. In this survey talk, we briefly review the foundations of single-agent reinforcement learning, point to the merits and challenges when applied in a multi-agent setting, and illustrate its potential in the context of an application from the field of manufacturing control and scheduling.

Key words: reinforcement learning, multi-agent systems, decentralized control, job-shop scheduling, multi-agent learning and coordination

Decentralized decision-making has become an active research topic in artificial intelligence [32]. In a distributed system, a number of individually acting agents coexist. If they strive to accomplish a common goal, i.e. if the multi-agent system is a cooperative one, then the establishment of coordinated cooperation between the agents is of utmost importance [18]. With this in mind, our focus is on multi-agent reinforcement learning methods which allow for automatically acquiring cooperative policies based solely on a specification of the desired joint behavior of the whole system. Most of the content presented in this survey is based on the author's work on learning in cooperative multi-agent systems [5].

Research in distributed systems has pointed out that the decentralization of the control of the system and of the observation of the system among independent agents has a significant impact on the complexity of solving a given problem [3]. Therefore, we have addressed the intricacy of learning and acting in multi-agent systems by the two following complementary approaches.

Many practical problems exhibit some structure whose exploitation may ease the task of finding solutions [2, 17]. For this reason, we have identified a subclass of general decentralized decision-making problems, the class of decentralized Markov decision processes with changing action sets and partially ordered transition dependencies, which features certain regularities in the way the agents interact with one another [14]. We have shown that the complexity of optimally solving a problem instance from this class is provably lower than solving a general one [6].

Even though a lower complexity class may be entered by sticking to certain subclasses of a general multi-agent problem [26], the computational complexity may be still so high that optimally solving it is infeasible. This holds, in particular, when intending to tackle problems of larger size that are of relevance for practical problems. Given these facts, our goal has not been to develop optimal solution algorithms that are applicable to small problems only [15], but to look for techniques capable of quickly obtaining approximate solutions in the vicinity of the optimum [19, 33]. To this end, we have developed and successfully utilized various model-free reinforcement learning approaches [13, 22]. In contrast to offline planning algorithms which aim at finding optimal solutions in a model-based manner, reinforcement learning [29] allows for employing independently learning agents and, hence, for a full decentralization of the problem [10].

As a matter of fact, many large-scale applications are well-suited to be formulated in terms of spatially or functionally distributed entities [24, 31, 23, 7, 16, 21]. Thus, multi-agent approaches are of high relevance to various real-world problems [4, 1, 27, 30, 28, 25]. Job-shop scheduling [20] is one such application stemming from the field of factory optimization and manufacturing control. It has been our particular goal to interpret job-shop scheduling problems as distributed sequential decision-making problems [8] and to employ the multi-agent reinforcement learning algorithms we propose for solving such problems [9, 12]. Moreover, we have successfully evaluated the performance of our learning approaches in the scope of various established scheduling benchmark problems [11].

References

1. A. Barto and R. Crites. Improving Elevator Performance using Reinforcement Learning. In *Advances in Neural Information Processing Systems 8 (NIPS 1995)*, pages 1017–1023, Denver, USA, 1996. MIT Press.
2. R. Becker, S. Zilberstein, V. Lesser, and C. Goldman. Solving Transition Independent Decentralized MDPs. *Journal of Artificial Intelligence Research*, 22:423–455, 2004.

3. D. Bernstein, D. Givan, N. Immerman, and S. Zilberstein. The Complexity of Decentralized Control of Markov Decision Processes. *Mathematics of Operations Research*, 27(4):819–840, 2002.
4. J. Boyan and M. Littman. Packet Routing in Dynamically Changing Networks – A Reinforcement Learning Approach. In *Advances in Neural Information Processing Systems 6 (NIPS 1993)*, pages 671–678, Denver, USA, 1994. Morgan Kaufmann.
5. T. Gabel. *Learning in Cooperative Multi-Agent Systems: Distributed Reinforcement Learning Algorithms and their Application to Scheduling Problems.* Südwestdeutscher Verlag für Hochschulschriften, Saarbrücken, Germany, 2009.
6. T. Gabel. *Multi-Agent Reinforcement Learning Approaches for Distributed Job-Shop Scheduling Problems.* Ph.D. Thesis, University of Osnabrück, 2009.
7. T. Gabel, R. Hafner, S. Lange, M. Lauer, and M. Riedmiller. Bridging the Gap: Learning in the RoboCup Simulation and Midsize League. In *Proceedings of the 7th Portuguese Conference on Automatic Control (Controlo 2006)*, Porto, Portugal, 2006. Portuguese Society of Automatic Control.
8. T. Gabel and M. Riedmiller. Adaptive Reactive Job-Shop Scheduling with Learning Agents. *International Journal of Information Technology and Intelligent Computing*, 2(4), 2007.
9. T. Gabel and M. Riedmiller. On a Successful Application of Multi-Agent Reinforcement Learning to Operations Research Benchmarks. In *Proceedings of the IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning (ADPRL 2007)*, pages 68–75, Honolulu, USA, 2007. IEEE Press.
10. T. Gabel and M. Riedmiller. Scaling Adaptive Agent-Based Reactive Job-Shop Scheduling to Large-Scale Problems. In *In Proceedings of the IEEE Symposium on Computational Intelligence in Scheduling (CI-Sched 2007)*, pages 259–266, Honolulu, USA, 2007. IEEE Press.
11. T. Gabel and M. Riedmiller. Evaluation of Batch-Mode Reinforcement Learning Methods for Solving DEC-MDPs with Changing Action Sets. In *Proceedings of the 8th European Workshop on Reinforcement Learning (EWRL 2008)*, pages 82–95, Lille, France, 2008. Springer.
12. T. Gabel and M. Riedmiller. Gradient Descent Policy Search for Distributed Job-Shop Scheduling Problems. In *Online Proceedings of the 18th International Conference on Planning and Scheduling (ICAPS 2008)*, Sydney, Australia, 2008. AAAI Press.
13. T. Gabel and M. Riedmiller. Joint Equilibrium Policy Search for Multi-Agent Scheduling Problems. In *Proceedings of the 6th Conference on Multiagent System Technologies (MATES 2008)*, pages 61–72, Kaiserslautern, Germany, 2008. Springer.
14. T. Gabel and M. Riedmiller. Reinforcement Learning for DEC-MDPs with Changing Action Sets and Partially Ordered Dependencies. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 1333–1336, Estoril, Portugal, 2008. IFAAMAS.
15. C. Goldman and S. Zilberstein. Decentralized Control of Cooperative Systems: Categorization and Complexity Analysis. *Journal of Artificial Intelligence Research*, 22:143–174, 2004.
16. J. Marecki and M. Tambe. On Opportunistic Techniques for Solving Decentralized Markov Decision Processes with Temporal Constraints. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007)*, page 219, Honolulu, USA, 2007. IFAAMAS.

17. R. Nair, P. Varakantham, M. Tambe, and M. Yokoo. Networked Distributed POMDPs: A Synthesis of Distributed Constraint Optimization and POMDPs. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI 2005)*, pages 133–139, Pittsburgh, USA, 2005. AAAI Press.
18. L. Panait and S. Luke. Cooperative Multi-Agent Learning: The State of the Art. *Autonomous Agents and Multi-Agent Systems*, 11(3):387–434, 2005.
19. L. Peshkin, K. Kim, N. Meuleau, and L. Kaelbling. Learning to Cooperate via Policy Search. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 489–496, Stanford, USA, 2000. Morgan Kaufmann.
20. M. Pinedo. *Scheduling. Theory, Algorithms, and Systems*. Prentice Hall, USA, 2002.
21. J. Pita, M. Jain, J. Marecki, F. Ordóñez, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus. Deployed ARMOR protection: the application of a game theoretic model for security at the Los Angeles International Airport. In *Industry Track Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 125–132, Estoril, Portugal, 2008. IFAAMAS.
22. M. Riedmiller, T. Gabel, R. Hafner, and S. Lange. Reinforcement Learning for Robot Soccer. *Autonomous Robots*, 27(1), 2009.
23. H. Santana, G. Ramalho, V. Corruble, and B. Ratitch. Multi-Agent Patrolling with Reinforcement Learning. In *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004)*, pages 1122–1129, New York, USA, 2004. IEEE Computer Society.
24. J. Schneider, J. Boyan, and A. Moore. Value Function Based Production Scheduling. In *Proceedings of the 15th International Conference on Machine Learning (ICML 1998)*, pages 522–530, Madison, USA, 1998. Morgan Kaufmann.
25. G. Settembre, P. Scerri, A. Farinelli, K. Sycara, and D. Nardi. A Decentralized Approach to Cooperative Situation Assessment in Multi-Robot System. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, pages 31–38, Estoril, Portugal, 2008. IFAAMAS.
26. J. Shen, R. Becker, and V. Lesser. Agent Interaction in Distributed POMDPs and Implications on Complexity. In *Proceedings of the Fifth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2006)*, pages 529–536, Hakodate, Japan, 2006. ACM Press.
27. S. Singh and D. Bertsekas. Reinforcement Learning for Dynamic Channel Allocation in Cellular Telephone Systems. In *Advances in Neural Information Processing Systems 9 (NIPS 1996)*, pages 974–980, Denver, USA, 1997. MIT Press.
28. M. Smith, S. Lee-Urban, and H. Muñoz-Avila. RETALIATE: Learning Winning Policies in First-Person Shooter Games. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence (AAAI 2007)*, pages 1801–1806, Vancouver, Canada, 2007. AAAI Press.
29. R. Sutton and A. Barto. *Reinforcement Learning. An Introduction*. MIT Press/A Bradford Book, Cambridge, USA, 1998.
30. G. Tesauro, R. Das, W. Walsh, and J. Kephart. Utility-Function-Driven Resource Allocation in Autonomic Systems. In *Proceedings of the Second International Conference on Autonomic Computing (ICAC 2005)*, pages 342–343, Seattle, USA, 2005. IEEE Computer Society.
31. K. Tumer and J. Lawson. Collectives for Multiple Resource Job Scheduling Across Heterogeneous Servers. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2003)*, pages 1142–1143, Victoria, Australia, 2003. ACM Press.

32. G. Weiss. Intelligent Agents. In Gerhard Weiss, editor, *Multi Agent Systems*. MIT Press, 1999.
33. D. Yagan and C. Tham. Coordinated Reinforcement Learning for Decentralized Optimal Control. In *Proceedings of the IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning (ADPRL 2007)*, pages 296–302, Honolulu, USA, 2007. IEEE Press.