Learning Automata as a Basis for Multi Agent Reinforcement Learning

Ann Nowé [a] Katje Verbeeck [a] Maarten Peeters [a]

[a] *Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel*

1 Article Summary

Learning Automata (LA) are adaptive decision making devices suited for operation in unknown environments [12]. Originally they were developed in the area of mathematical psychology and used for modeling observed behavior. In its current form, LA are closely related to Reinforcement Learning (RL) approaches and most popular in the area of engineering. LA combine fast and accurate convergence with low computational complexity, and have been applied to a broad range of modeling and control problems. However, the intuitive, yet analytically tractable concept of learning automata makes them also very suitable as a theoretical framework for Multi agent Reinforcement Learning (MARL).

Reinforcement Learning (RL) is already an established and profound theoretical framework for learning in stand-alone or single-agent systems. Yet, extending RL to multi-agent systems (MAS) does not guarantee the same theoretical grounding. As long as the environment an agent is experiencing is Markov, and the agent can experiment sufficiently, RL guarantees convergence to the optimal strategy. In a MAS however, the reinforcement an agent receives, may depend on the actions taken by the other agents acting in the same environment. Hence, the Markov property no longer holds. And as such, guarantees of convergence are lost. In the light of the above problem it is important to fully understand the dynamics of multi-agent reinforcement learning.

Although, they are not fully recognized as such, LA are valuable tools for current MARL research. LA are updated strictly on the basis of the response of the environment, and not on the basis of any knowledge regarding other automata, i.e. nor their strategies, nor their feedback. As such LA agents are very simple. Moreover, LA can be treated analytically. Convergence proofs do exist for a variety of settings ranging from a single automaton model acting in a simple stationary random environment to a distributed automata model interacting in a complex environment.

In this paper we argue that LA are very interesting building blocks for learning in multi agent systems. The LA can be viewed as policy iterators, who update their action probabilities based on private information only. This is a very attractive property in applications where communication is expensive. LA are in particular appealing in games with stochastic payoffs. Then we move to collections of learning automata, that can independently converge to interesting solution concepts. We study the single stage setting, including the analytical results. Then we generalize to interconnected learning automata, that can deal with multi agent multi-stage problems. We also show how Ant Colony Optimization can be mapped to the interconnected Learning Automata setting.

This extended abstract is a summary of an article published earlier [14].

References

[1] Eric Bonabeau, Marco Dorigo, and Guy Theraulaz. *Swarm Intelligence, From Natural to Artificial Systems*. Santa Fe Institute studies in the sciences of complexity. Oxford University Press, 1999.

[2] C. Boutilier. Planning, learning and coordination in multiagent decision processes. In *Proceedings of the 6th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 195 – 210, Renesse, Holland, 1996.

[3] C. Boutilier. Sequential optimality and coordination in multiagent systems. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, pages 478 – 485, Stockholm, Sweden, 1999.

[4] R.R. Bush and F. Mosteller. *Stochastic Models for Learning.* Wiley, New York, 1958.

[5] C. Claus and C. Boutilier. The dynamics of reinforcement learning in cooperative multiagent systems. In *Proceedings of the 15th National Conference on Artificial Intelligence*, pages 746 – 752, 1998.

[6] A. Colorni, M. Dorigo, F. Maffioli, V. Maniezzo, G. Righini, and M. Trubian. Heuristics from nature for hard combinatorial optimization problems. *International Transactions in Operational Research*, 1996.

[7] M. Dorigo, G. Di Caro, and L.M. Gambardella. Ant algorithms for discrete optimization. *Artificial Life*, 5 (2):137 – 172, 1999.

[8] Marco Dorigo and Gianno Di Caro. The ant colony optimization meta-heuristic. *D.Corne, M.Dorigo and F.Glover (Eds.), New Ideas In Optimization, Maidenhaid, UK: McGraw-Hill*, 1999.

[9] Marco Dorigo, Vittorio Maniezzo, and Alberto Colorni. The ant system: Optimization by a colony of cooperating agents. *IEE Transactions on Systems, Man, and Cybernetics*, 1996.

[10] Marco Dorigo and Thomas Stützle. *Ant Colony Optimization.* The MIT Press, 2004.

[11] M. Littman. Markov games as a framework for multi-agent reinforcement learning. In *Proceedings of the 11th International Conference on Machine Learning*, pages 322 – 328, 1994.

[12] K. Narendra and M. Thathachar. *Learning Automata: An Introduction.* Prentice-Hall International, Inc, 1989.

[13] K.S. Narendra and K. Parthasarathy. Learning automata approach to hierarchical multiobjective analysis. Technical Report Report No. 8811, Electrical Engineering Yale University, New Haven, Connecticut, 1988.

[14] Ann Nowé, Katja Verbeeck, and Maarten Peeters. Learning automata as a basis for multi agent reinforcement learning. *Learning and Adaptation in Multi-Agent Systems, Lecture Notes in Artificial Intelligence*, 3898:71–85, 2005.

[15] B.J. Oommen and T.D. Roberts. Continuous learning automata solutions to the capacity assignment problem. *IEEE Transactions on Computations*, 49:608 – 620, 2000.

[16] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, MA, 1998.

[17] M.L. Tsetlin. Automaton theory and modelling of biological systems. *Mathematics in Science and Engineering*, 102, 1973.

[18] C. Unsal, P. Kachroo, and J.S. Bay. Multiple stochastic learning automata for vehicule path control in an automated highway system. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 29:120 – 128, 1999.

[19] K. Verbeeck. *Coordinated Exploration in Multi-Agent Reinforcement Learning.* PhD thesis, Computational Modeling Lab, Vrije Universiteit Brussel, Belgium, 2004.

[20] K. Verbeeck, A. Nowé, K. Tuyls, and M.Peeters. Multi-agent reinforcement learning in stochastic single and multi-stage games. In *Kudenko et al (Eds): Adaptive Agents and Multi-Agent Systems II*, pages 275–294. Springer LNAI 3394, 2005.