Learning Agents in a Homo Egualis Society

Ann Nowé
Computational Modeling Lab
Vrije Universiteit Brussel,
Belgium
asnowe@info.vub.ac.be

Katja Verbeeck
Computational Modeling Lab
Vrije Universiteit Brussel,
Belgium
kaverbee@vub.ac.be

Tom Lenaerts
Computational Modeling Lab
Vrije Universiteit Brussel,
Belgium
tlenaert@vub.ac.be

ABSTRACT

Currently we are studying how distributed agents can learn how to share common resources in a fair way, with only a limited amount of communication.

We are looking at multi-agent problems where local competition between agents arises, but where there is also the necessity of global cooperation. The kind of applications we have in mind is for example routing in telecommunication networks. A natural multi-agent approach to this problem is to put an agent at each router. An agent is responsible for sending the packets to the destination specified by selecting an appropriate neighboring router, which in turn becomes responsible for getting the packet to the destination. Clearly, agents have to collaborate to perform the global task of routing, in the sense that a sequence of actions by different agents has to be optimized. This can be achieved by a classical reinforcement learning approach, provided resource availability, i.e. bandwidth, is plentiful. If on the other hand resources are scarce, competition arises between the agents and collaboration becomes necessary in order to perform the global task as well as possible. The problems we address are thus characterized by the fact that the performance of the global system can only be as well as that of the weakest agent in the system.

For learning in a multi-agent system, two extreme approaches can be recognized. On the one hand, the presence of other agents, who are possibly influencing the effects a single agent experiences, can be completely ignored. Thus a single agent is learning as if the other agents are not around. This usually leads to oscillating policies or unfair solutions. On the other hand, the presence of other agents can be modeled explicitly. This results in a joint action space approach, which recently received quite a lot of attention. In it's pure form joint action spaces assume agents are taking their actions synchronously and that agents have access to the information available to other agents, which implies free communication and feedback is given to all the agents at the same time. That this is unrealistic for most real-world problems is clear, consider for instance the routing problem mentioned[1]. Moreover in general sum games it is not always clear which policies should be learned. Playing pure Nash equilibria is often unfair to at least one of the players, while playing a mixed strategy doesn't give any guarantee for coordination and usually results in a sub-optimal payoff for all agents.

We therefore introduce the notion of periodical policies. In a periodical policy players alternate between playing different Nash equilibria if these equilibria are optimal for some players but suboptimal for others. We show how a periodical policy can be learned by social reinforcement learners, which are inspired on the Homo egualis society from sociology. Our algorithm is based on results from Learning Automata theory that assures that players in an N-person nonzero-sum game who use a reward-inaction update scheme with an arbitrarily small step size will always converge to one of the equilibrium points depending on the initial conditions. After convergence to one of the Nash equilibria, communication takes places only to compare performances. If necessary agents reduce their action spaces to go find alternative Nash equilibria in other subspaces. In case the pure Nash equilibria don't give the same payoff to all the players, the algorithm proposed is able to learn the length of the period each agent needs to equalize the average payoff. The algorithm has been applied on 2-agent and 3-agent games with conflicting equilibrium points. Synchronous action taking with immediate feedback, as well as asynchronous action taking with delayed feedback are tested. In all cases the agents where able to learn a periodical policy such that their overall payoff was equalized optimally.

1. REFERENCES

[1] Nowé, A., Verbeeck, K.,: Distributed Reinforcement learning, Loadbased Routing a case study. Proceedings of the Neural, Symbolic and Reinforcement Methods for sequence Learning Workshop at ijcai99. (1999).

[2] Nowé, A., Verbeeck, K., Lenaerts T.,: Learning Agents in a Homo Egualis Society. Technical report. Como/03/2001. Vrije Universiteit Brussel (2001).

[3] Nowé, A., Parent, J., Verbeeck, K., : Social Agents Playing a Periodical Policy. Technical report. Como/04/2001. Vrije Universiteit Brussel (2001).

[1] For some classes of games the joint action space approach is overkill. For instance cooperative maximum games are dominance solvable on the unique Pareto-dominant outcome.