Multi-Agent Learning I: Problem definition

Yoav Shoham and Rob Powers
Stanford University
{shoham,powers}@cs.stanford.edu

January 16, 2010

1 Definition

Multi-agent learning refers to settings in which multiple agents learn simultaneously. Usually defined in a game theoretic setting, specifically in repeated games or stochastic games, the key feature that distinguishes multi-agent learning from single-agent learning is that in the former the learning of one agent impacts he learning of others. As a result neither the problem definition for mutli-agent learning, nor the algorithms offered, follow in a straightforward way from the single-agent case. In this first of two entries on the subject we focus on the problem definition.

2 Background

The topic of multiagent learning (MAL henceforth), has a long history in game theory, almost as long as the history of game theory itself.[1] In Artificial Intelligence (AI) the history of *single*-agent learning is of course as rich if not richer; one need not further than this Encyclopedia for evidence. And while it is only in recent years that AI has branched into the multi-agent aspects of learning, it has done so with something of a vengeance. If in 2003 one could describe the AI literature on MAL by enumerating the relevant articles, today this is no longer possible. The leading conferences routinely feature articles on MAL, as do the journals.[2]

While the AI literature maintains a certain flavor that distinguishes it from the game theoretic literature, the commonalities are greater than the differences. Indeed, alongside the area of mechanism design, and perhaps the computational questions surrounding solution concepts such as the Nash equilibrium, MAL is today arguably one of the most fertile interaction grounds between computer

[1] Another more recent term for the area within game theory is *interactive learning*.

[2] We acknowledge a simplification of history here. There is definitely MAL work in AI that predates the last few years, though the relative deluge is indeed recent. Similarly, we focus on AI since this is where most of the action is these days, but there are also other areas in computer science that feature MAL material; we mean to include that literature here as well.

science and game theory. They key aspect of MAL, which ties the work together, and which distinguishes it from single-agent learning, is the fact that in MAL one cannot separate the process of learning from the process of teaching. The learning of one agent causes it to change its behavior; this causes other agents to adapt their behavior, which in turn causes the first agent to keep adapting too. Such reciprocal – or interactive – learning calls not only for different types of learning algorithms, but also for different yardsticks by which to evaluate learning. For this reason the literature on MAL can be confusing. Not only are the learning techniques varies, but the goal of learning and the evaluation measures are diverse, and often left only implicit.

We will couch our discussion in the formal setting of *stochastic games* (aka *Markov games*). Most of the MAL literature adopts this setting, and indeed most of it focuses on the even more narrow class of *repeated games*. Furthermore, stochastic games also generalize *Markov Decision Problems (MDPs)*, the setting from which much of the relevant learning literature in AI originates. These are defined as follows.

A stochastic game can be represented as a tuple: $(N, S, \vec{A}, \vec{R}, T)$. N is a set of agents indexed $1, \ldots, n$. S is a set of n-agent stage games. $\vec{A} = A_1, \ldots, A_n$, with A_i the set of actions (or pure strategies) of agent i (note that we assume the agent has the same strategy space in all games; this is a notational convenience, but not a substantive restriction). $\vec{R} = R_1, \ldots, R_n$, with $R_i : S \times \vec{A} \rightarrow \mathcal{R}$ giving the immediate reward function of agent i for stage game S. $T : S \times \vec{A} \rightarrow \Pi(S)$ is a stochastic transition function, specifying the probability of the next stage game to be played based on the game just played and the actions taken in it.

We also need to define a way for each agent to aggregate the set of immediate rewards received in each state. For finitely repeated games we can simply use the sum or average, while for infinite games the most common approaches are to use either the limit average or the sum of discounted awards $\sum_{t=1}^{\infty} \delta^t r_t$, where r_t is the reward received at time t.

A repeated game is a stochastic game with only one stage game, while an MDP is a stochastic game with only one agent.

[Note: While most of the MAL literature lives happily in this setting, we would be remiss not to acknowledge the literature that does not. Certainly one could discuss learning in the context of extensive-form games of incomplete and/or imperfect information. Even farther afield, interesting studies exist of learning in large population games and evolutionary models, and particularly *replicator dynamics (RD)* and *evolutionary stable strategies (ESS)*.]

What is there to learn in stochastic games? Here we need to be explicit about some aspects of stochastic games that were glossed over so far. Do the agents know the stochastic game, including the stage games and the transition probabilities? If not, do they at least know the specific game being played at each stage, or only the actions available to them? What do they see after each stage game has been played – only their own rewards, or also the actions played by the other agent(s)? Do they perhaps magically see the other agent(s)' mixed strategy in the stage game? And so on.

In general, games may be known or not, play may be observable or not, and so on. We will focus on known, fully observable games, where the other agent's strategy (or agents' strategies) is not known a priori (though in some case there is a prior distribution over it). In our restricted setting there two possible things to learn. First, the agent can learn the opponent's (or opponents') strategy (or strategies), so that the agent can then devise a best (or at least a good) response. Alternatively, the agent can learn a strategy of his own that does well against the opponents, without explicitly learning the opponent's strategy. The first is sometimes called *model-based learning*, and the second *model-free learning*.

In broader settings there is more to learn. In particular, with unknown games, one can learn the game itself. Some will argue the restricted setting is not a true learning setting, but (a) much of the current work on MAL, particularly in game theory, takes place in this setting, and (b) the foundational issues we wish to tackle surface already here. In particular, our comments are intended to also apply to the work in the AI literature on games with unknown payoffs, work which builds on the success of learning in unknown MDPs. We will have more to say about the nature of 'learning' in the setting of stochastic games in the following sections.

3 Problem definition

When one examines the MAL literature one can identify several distinct agendas at play, which are often left implicit and conflated. A prerequisite for success in the field is to be very explicit about the problem being addressed. Here we list five distinct coherent goals of MAL research. They each have a clear motivation and a success criterion. They can be caricatured as follows:

1. Computational
2. Descriptive
3. Normative
4. Prescriptive, cooperative
5. Prescriptive, non-cooperative

The first agenda is computational in nature. It views learning algorithms as an iterative way to compute properties of the game, such as solution concepts. As an example, fictitious play was originally proposed as a way of computing a sample Nash equilibrium for zero-sum games, and replicator dynamics has been proposed for computing a sample Nash equilibrium in symmetric games. These tend not to be the most efficient computation methods, but they do sometimes constitute quick-and-dirty methods that can easily be understood and implemented.

The second agenda is descriptive – it asks how natural agents learn in the context of other learners. The goal here is to investigate formal models of learning that agree with people's behavior (typically, in laboratory experiments), or possibly with the behaviors of other agents (for example, animals or organizations). This problem is clearly an important one, and when taken seriously calls for strong justification of the learning dynamics being studied. One approach is to apply the experimental methodology of the social sciences.

The centrality of equilibria in game theory underlies the third agenda we identify in MAL, which for lack of a better term we called normative, and which focuses on determining which sets of learning rules are in equilibrium with each other. More precisely, we ask which repeated-game strategies are in equilibrium; it just so happens that in repeated games, most strategies embody a learning rule of some sort. For example, we can ask whether fictitious play and Q-learning, appropriately initialized, are in equilibrium with each other in a repeated Prisoner's Dilemma game.

The last two agendas are prescriptive; they ask how agents *should* learn. The first of these involves distributed control in dynamic systems. There is sometimes a need or desire to decentralize the control of a system operating in a dynamic environment, and in this case the local controllers must adapt to each other's choices. This direction, which is most naturally modelled as a repeated or stochastic common-payoff (or 'team') game. Proposed approaches can be evaluated based on the value achieved by the joint policy and the resources required, whether in terms of computation, communication, or time required to learn the policy. In this case there is rarely a role for equilibrium analysis; the agents have no freedom to deviate from the prescribed algorithm.

In our final agenda, termed 'prescriptive, non-cooperative', we ask how an agent should act to obtain high reward in the repeated (and more generally, stochastic) game. It thus retains the design stance of AI, asking how to design an optimal (or at least effective) agent for a given environment. It just so happens that this environment is characterized by the types of agents inhabiting it, agents who may do some learning of their own. The objective of this agenda is to identify effective strategies for environments of interest. An effective strategy is one that achieves a high reward in its environment, where one of the main characteristics of this environment is the selected class of possible opponents. This class of opponents should itself be motivated as being reasonable and containing opponents of interest. Convergence to an equilibrium is not a goal in and of itself.

4 Recommended reading

Requisite background in game theory can be obtained from the many introductory texts, and most compactly from [Leyton-Brown and Shoham, 2008]. Game theoretic work on multiagent learning is covered in [Fudenberg and Levine, 1998] and [Young, 2004]. An expanded discussion of the problems addressed under the header of MAL cab be found in [Shoham et al., 2007], and the responses to it in [Vohra and (eds.), 2007]. Discussion of MAL algorithms, both traditional and more novel ones, can be found in the above references, as well as in [Greenwald and (Eds.), 2007].

References

[Fudenberg and Levine, 1998] Fudenberg, D. and Levine, D. (1998). *The Theory of Learning in Games*. MIT Press.

[Greenwald and (Eds.), 2007] Greenwald, A. and (Eds.), M. L. L. (2007). Special issue on learning and computational game theory. *Machine Learning*, 67(1-2).

[Leyton-Brown and Shoham, 2008] Leyton-Brown, K. and Shoham, Y. (2008). *Essentials of Game Theory*. Morgan and Claypool Publishers.

[Shoham et al., 2007] Shoham, Y., Powers, W. R., and Grenager, T. (2007). If multiagent learning is the answer, what is the question? *Artificial Intelligence*, 171(1):365–377. Special issue on Foundations of Multiagent Learning.

[Vohra and (eds.), 2007] Vohra, R. and (eds.), M. P. W. (2007). Special issue on foundations of multiagent learning. *Artificial Intelligence*, 171(1).

[Young, 2004] Young, H. P. (2004). *Strategic Learning and its Limits*. Oxford University Press.