Available online at www.sciencedirect.com
ScienceDirect
Artificial Intelligence 171 (2007) 448–452
Artificial Intelligence
www.elsevier.com/locate/artint

Learning equilibrium as a generalization of learning to optimize ☆

Dov Monderer *, Moshe Tennenholtz

Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa 32000, Israel

Received 15 May 2006; received in revised form 20 October 2006; accepted 21 December 2006

Available online 26 January 2007

Abstract

We argue that learning equilibrium is an appropriate generalization to multi-agent systems of the concept of learning to optimize in single-agent setting. We further define and discuss the concept of weak learning equilibrium.

Keywords: Learning; Machine learning; Learning equilibrium

1. Preface

In [16], Shoham, Powers, and Grenager (SPG) present five distinct agendas in multi-agent learning. In this manuscript we discuss their third agenda—the normative approach. SPG mention that the requirement that learning algorithms would be an equilibrium may serve as a synonyms to this approach. We claim that the equilibrium approach is indeed the right one if the question is: what should a *mediator* who makes recommendations to *all* players, recommend. The equilibrium property seems to be a necessary ingredient in such recommendations. As economists do not tend to consider mediators (competing firms do not go to a central mediator to determine their, say, pricing policy)[1] it is of no surprise that the equilibrium agenda is not a major issue in the learning literature in economics. However, when, say eBay provides the participants a proxy service, then it actually plays the role of a mediator (and not only of an organizer).

The theory of learning in multi-agent systems inherits all the conceptual and practical difficulties of learning in single-agent settings, as well as all difficulties of analyzing behavior in multi-agent settings. Therefore, in order to define and understand the equilibrium approach to learning in multi-agent systems we phrase it as an extension of work on learning in single-agent systems.[2]

2. Learning in single-agent systems

Roughly speaking, one could partition work on learning in single-agent systems into two major but not necessarily independent categories:

☆ Both authors thank the Israeli Science Foundation and the Fund for the Promotion of Research at the Technion for the support of their research.
* Corresponding author.
E-mail address: dov@ie.technion.ac.il (D. Monderer).

[1] Some well-known exceptions to this statement can be found in the literature on correlated equilibrium [4], and on communication equilibrium [8,15]. However, the general theme in economics is that there is no mediator in the system that recommends behavior to the agents.

[2] For the sake of exposition we introduce all notations in this paper with pure strategies.

0004-3702/$ – see front matter
doi:10.1016/j.artint.2007.01.002

- *Descriptive theory-prediction*: Given examples of past behavior of a system, and some background information, we would like to predict the future behavior of the system. Classical work in statistics, inductive inference and supervised learning in AI fit into this category, as well as much work in data mining and the classification of subjects in psychology.
- *Normative theory-optimization*: Given partial information about a system, our aim is to devise algorithms for optimizing an agent's behavior in the system. Typically, the interleaving of exploration and exploitation is needed. Work on reinforcement learning in AI, as well bandits problems fit into this category.

We only discuss our view about the extension to multi-agent systems of the normative approach. That is, we use the word learning in a single-agent setting as a synonym for optimization in dynamic situations with incomplete information.

Consider a single agent facing a dynamic decision problem with incomplete information, D, defined by a set of dynamic decision problems D_ω with a parameter $\omega \in \Omega$. Each $\omega \in \Omega$ is called a *state of nature*. Nature chooses ω, but the agent does not know ω. However, he possesses some initial information about the state of nature, and he acquires additional partial information after every stage. For example, every D_ω can be a Markov decision problem. If we have a prior probability distribution over the set Ω we call the problem a *Bayesian dynamic decision problem*. If we want to stress the fact that such a prior probability does not exist we call the problem a *Pre-Bayesian dynamic decision problem*.[3]

A strategy of the decision maker at each D_ω is called a *policy*. A strategy of the agent in the decision problem with incomplete information, D, is called in this paper an *algorithm*. We assume that had the agent known ω he would have chosen an optimal policy, which would have given him the long-run value, $v(\omega)$, of D_ω. More precisely, let U_ω be the long-run reward function of the problem D_ω, and let $S(\omega)$ be the set of possible policies for this problem. A policy $f^\omega \in S(\omega)$ is an *optimal policy* at D_ω if

$$\max_{g^\omega \in S(\omega)} U_\omega(g^\omega) = U_\omega(f^\omega). \tag{1}$$

The *value* of D_ω is the real-valued function v defined on Ω as follows:

$$v(\omega) = \max_{g^\omega \in S(\omega)} U_\omega(g^\omega) \quad \forall \omega \in \Omega. \tag{2}$$

Ideally, in a dynamic decision problem with incomplete information an optimizing agent would use an algorithm that guarantees $v(\omega)$ for every ω. This approach is mainly taken in machine learning. We accept this view; in our view the right notion for a learning-to-optimize algorithm for a decision problem with incomplete information is the following: *It is an algorithm that yields an optimal policy at every* ω.[4] More precisely:

Let f be an algorithm for D. For every ω we denote by f_ω, the policy induced by f on D_ω.

f is a *learning-to-optimize algorithm*in D if for every w, f_ω is an optimal algorithm in D_ω, that is

$$\max_{g^\omega \in S(\omega)} U_\omega(g^\omega) = U_\omega(f_\omega) \quad \forall \omega \in \Omega. \tag{3}$$

An equivalent definition will be useful in the sequel.

The set of all algorithms for D is denoted by S. For every $f \in S$ and for every $\omega \in \Omega$ define $U(\omega, f) = U_\omega(f_\omega)$. Obviously, f is a learning-to-optimize algorithm in D if and only if

$$\max_{g \in S} U(\omega, g) = U(\omega, f) \quad \forall \omega \in \Omega. \tag{4}$$

Unfortunately, there exist dynamic decision problems for which a learning-to-optimize algorithms do not exist.

In Bayesian dynamic decision problems it is customary to look for algorithms that maximize the long-run expected reward of the agent. Such algorithms generally exist. We call such an algorithm an *optimal Bayesian algorithm*.

[3] In economics, it is customary to relate to a Bayesian model as a model with incomplete information. Until recently, a model without priors was not given a special name. Recently, such games have received several titles in various papers. In this paper we follow the terminology of [11], and we refer to such games as pre-Bayesian.

[4] Practically, the definition would be more elaborate, and would refer to various accuracy parameters. Notice that we refer here to the long-run value mentioned above.

That is, f is an optimal Bayesian algorithm if

$$\max_{g\in S}\int_{\Omega} U(\omega, g)\,\mathrm{d}\mu(\omega) = \int_{\Omega} U(\omega, f)\,\mathrm{d}\mu(\omega), \tag{5}$$

where μ is the prior probability on Ω.[5]

It is important to note that in a Bayesian dynamic decision problem, every learning-to-optimize algorithm is also an optimal Bayesian algorithm, but the converse does not necessarily hold.[6]

3. Learning in multi-agent settings

We take the position of a mediator who is about to assign algorithms to a set of selfish agents, $N = \{1, 2, \ldots, n\}$ who are engaged in a multistage game with incomplete information, G.[7] This game, G is defined by a set of multi-stage games with complete information, G_ω with a parameter $\omega \in \Omega$. Nature chooses a state of nature ω, but the agents do not know ω. However, each agent possesses some initial private information about the state of nature, and he acquires additional partial information after every stage. For example, every G_ω can be a repeated game. If we have a prior probability distribution over the set Ω we say that G is a *Bayesian multi-stage game*. If we want to stress the fact that such a prior probability does not exist we call G a *pre-Bayesian multi-stage game*. In the complete information case, when dealing with the multi-agent setting, the term policy used in the single agent setting is replaced by the term *strategy*. As in the single agent setting, a strategy of an agent in G is called an *algorithm*.

The first question is what is the analogous concept of an "optimal policy" in the single-agent setup in the game G_ω, in which the agents know ω. This is one of the most important conceptual issues dealt with in game theory. We take the position that in the *presence of a mediator*, optimality means equilibrium. That is, the strategy given to every agent i is optimal if all other agents are using their strategies in the profile.[8] Hence, the analogous definition to an optimal policy in the single agent decision problem with complete information is: A profile of strategies, $\mathbf{f}^\omega = (f_1^\omega, f_2^\omega, \ldots, f_n^\omega)$ is an *equilibrium profile* in G_ω if

$$U_i^\omega(\mathbf{f}^\omega) = \max_{g_i^\omega \in S_i(\omega)} U_i^\omega(g_i^\omega, f_{-i}^\omega) \quad \forall i \in N. \tag{6}$$

Note, however, that except for two-person zero-sum games, the concept of a value function does not have a well-defined meaning in the multi-agent model.

It is important to stress again the existence of a mediator in order to understand our approach. We do not claim that economic agents play in equilibrium. We do not claim that an agent who is facing a multi-agent decision problem should use an equilibrium strategy; This is because we cannot be sure that other agents would use an equilibrium strategy, and even if they do they may stick to another equilibrium. However, a reliable mediator who provides all agents with algorithms can expect players to use the algorithms only if the profile of algorithms is in equilibrium.

Hence, we assume that had the agents known ω they would have chosen an optimal profile of strategies, i.e. they would have behaved according to an equilibrium profile of G_ω.

Extending upon the single-agent perspective, in our opinion the right notion for a learning-to-optimize algorithm profile in a pre-Bayesian multi-stage game is the following notion of *learning equilibrium*: It is a profile of algorithms that yield an equilibrium profile of strategies at the multi-stage game, G_ω, for every ω. That is, $\mathbf{f}$ is a *learning equilibrium* if

$$\Big[\max_{g_i^\omega \in S_i(\omega)} U_i^\omega(g_i^\omega, f_{-i}^\omega) = U_i^\omega(\mathbf{f}^\omega) \quad \forall i \in N\Big] \quad \forall \omega \in \Omega. \tag{7}$$

[5] In models in which U_ω is defined as the limit of averages, it is some times customary to take the limit before the integral in (5).

[6] Some of the literature in single-agent learning deals with the issue of when a given optimal Bayesian algorithm is also a learning-to-optimize algorithm.

[7] Recall that this approach is heavily based on the existence of a mediator. The reader may consult [1] in order to see one of the authors' approach to learning to optimize in the absence of such a mediator.

[8] In particular models one can focus on particular refinements of Nash equilibrium like sub-game perfect equilibrium, sequential equilibrium or dominated strategy equilibrium.

Alternatively, one can use the analogous definition to (5). That is, **f** is a *learning equilibrium* if and only if

$$\left[\max_{g_i \in S_i} U_i(\omega, g_i, \mathbf{f}_{-i}) = U_i(\omega, \mathbf{f}^{\omega}) \quad \forall i \in N\right] \quad \forall \omega \in \Omega. \tag{8}$$

Notice that condition (7) has a local flavor in the sense that in order to show that **f** is not a learning equilibrium one has to find for some player i, a state ω and a strategy for i in the game G_ω that violates (7). On the other hand, the requirement in (8) is a global one in the sense that in order to show that **f** is not a learning equilibrium, one has to find for some player i, an algorithm in G that violates (8) for some ω. The global definition is just the classical definition of ex-post equilibrium. Hence, the terms learning equilibrium and ex post equilibrium coincide for multi-stage games with incomplete information.[9]

In the context of Bayesian multi-stage game, the equilibrium approach implies the use of a profile of algorithms that form a *Bayesian equilibrium*. Recall that **f** is a Bayesian equilibrium if

$$\max_{g_i \in S_i} \int_{\Omega} U_i(\omega, g_i, \mathbf{f}_{-i})\, \mathrm{d}\mu(\omega) = \int_{\Omega} U_i(\omega, \mathbf{f}^{\omega})\, \mathrm{d}\mu(\omega) \quad \forall i \in N. \tag{9}$$

Like in the single agent setting, in a Bayesian multi-stage game, every learning equilibrium is also a Bayesian equilibrium, but the converse does not necessarily hold.

While in general Bayesian equilibrium exists (with mixed strategies), and learning equilibrium may not exist, recent work have shown that (surprisingly) learning equilibrium does exist in several rich settings [3,5,6].

When a learning equilibrium does not exist one may use other notions of equilibrium in pre-Bayesian games in order to define learning. For example, minimax-regret equilibrium [12] and safety level equilibrium [2] can be used. These other notions have the advantage of existence, but they will not be state-wise optimal in the sense of yielding a Nash equilibrium at every state of nature.

4. Weak learning equilibrium

For ease of exposition we discuss in this section only a simple type of multi-stage game with incomplete information—repeated games with incomplete information. Hence, for every ω there is a one-shot game $G(\omega)$ such that G_ω is the dynamic game, in which $G(\omega)$ is played in every stage. In $G(\omega)$ the players choose among possible *actions*. Hence, every strategy profile f^{ω} in G_ω generates an infinite path of action profiles in $G(\omega)$. Moreover, we only discuss the long-run utility function defined by the limit of averages of payoffs in the one stage games.

As mentioned, in the Bayesian setting, if a learning equilibrium does not exist, it is natural to require that the algorithm profile satisfies the global optimality required by a Bayesian equilibrium. In such an equilibrium an agent who considers his expected long-run payoff would not deviate from the algorithm designed for him. However, a profile of algorithms, which forms a Bayesian equilibrium might not satisfy the state-wise optimality requirement of being in equilibrium in every state of nature.

Kalai and Lehrer [13] defined a notion of "weak" state-wise optimality-like property, which may be satisfied by a profile of algorithms. Their notion inspires the following definition. An algorithm profile **f** is a *weak learning equilibrium* if it is a Bayesian equilibrium that satisfies the following property: the path of actions generated at every state of nature is an equilibrium path from a certain stage on. That is, for every ω there exists an equilibrium strategy profile $\mathbf{g}^{\omega}$ at $G(\omega)$ and an integer $T(\omega)$ such that the profile of actions generated by $\mathbf{f}^{\omega}$ and $\mathbf{g}^{\omega}$ at stage T coincide for every $T \geqslant T(\omega)$.

For example, if each $G(\omega)$ is a Prisoner's Dilemma game (where different ω's determine different payoff functions of the prisoners dilemma), then a Bayesian equilibrium algorithm profile that for every ω generates the cooperative outcome at every stage would be a weak learning equilibrium, because by the Folk theorem, for each ω there exists a repeated game equilibrium strategy profile that generates the path "cooperation at every stage".

Kalai and Lehrer [13] proved the existence of a mixed-strategy version of weak learning equilibrium for a Bayesian repeated game with a countable set of states of nature in which every player knows his own payoff, and only his own payoffs.

[9] The original definition of learning equilibrium in [5,6], and its generalizations, in particular to the concept of robust learning equilibrium in [3], used the global approach.

5. Some related literature from economics

The literature on learning in economics suggests a weakening of the notion of Bayesian equilibrium, by requiring self-confirming equilibrium. This notion has been implicitly defined by Fudenberg and Kreps [9] and further developed by Dekel, Fudenberg and Levine [7,10]. A related variant, titled subjective equilibrium, was defined and analyzed by Kalai and Lehrer [13,14]. In a self confirming equilibrium the beliefs of the players may not coincide off equilibrium path. This definition makes a lot of sense in economics environments, where the algorithms of the players are independently chosen , but makes less sense when the algorithms are suggested to the players by a mediator. Dekel, Fudenberg, and Levine [7] discussed situations under which self-confirming equilibrium satisfies the weak notion of local optimality suggested by [13], yielding in our terminology a mixed-strategy version of weak learning equilibrium.

6. Summary

We consider multi-stage games with incomplete information, in which a mediator provides the agents with algorithms. We focus on generalizing learning-to-optimize algorithms in the single-agent setting to learning equilibrium in a multi-agent setting. The existence of such a mediator is a major issue in our setting. Our generalization refers to two requirements that turn out to imply each other: the algorithms should be state-wise optimal, as well as globally optimal, where optimality in the multi-agent setting is captured by the notion of equilibrium. Hence, in a learning equilibrium, an agent does not attempt to learn the other players' algorithms but rather she takes these algorithms as given, and she tries to optimize. This optimization may yield gradual learning of the true state through stages of exploration and exploitation.

References

[1] A. Altman, A. Boden-Bercovici, M. Tennenholtz. Learning in one-shot strategic form games, in: Proceedings of ECML-06, 2006.
[2] I. Ashlagi, D. Monderer, M. Tennenholtz, Resource selection games with unknown number of players, in: Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-06), 2006.
[3] I. Ashlagi, D. Monderer, M. Tennenholtz, Robust learning equilibrium, in: Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence (UAI 2006), 2006.
[4] R.J. Aumann, Subjectivity and correlation in randomized strategies, Journal of Mathematical Economics 1 (1974) 67–96.
[5] R. Brafman, M. Tennenholtz, Optimal efficient learning equilibrium: Imperfect monitoring in symmetric games, in: Proceedings of AAAI-2005, AUAI Press, 2005.
[6] R.I. Brafman, M. Tennenholtz, Efficient learning equilibrium, Artificial Intelligence 159 (2004) 27–47.
[7] E. Dekel, D. Fudenberg, D.K. Levine, Payoff information and self-confirming equilibrium, Journal of Economic Theory 89 (2) (1999) 165–185.
[8] F. Forges, An approach to communication equilibrium, Econometrica 54 (6) (1986) 1375–1385.
[9] D. Fudenberg, D.M. Kreps, Learning in extensive games, I: Self-confirming equilibrium, Games and Economic Behavior 8 (1995) 20–55.
[10] D. Fudenberg, D.K. Levine, Self-confirming equilibrium, Econometrica 61 (1993) 523–546.
[11] R. Holzman, N. Kfir-Dahav, D. Monderer, M. Tennenholtz, Bundling equilibrium in combinatorial auctions, Games and Economic Behavior 47 (2004) 104–123.
[12] N. Hyafil, C. Boutilier, Regret minimizing equilibria and mechanisms for games with strict type uncertainty, in: Proceedings of the 20th Annual Conference on Uncertainty in Artificial Intelligence (UAI-04), Arlington, VA, AUAI Press, 2004, pp. 268–277.
[13] E. Kalai, E. Lehrer, Rational learning leads to Nash equilibrium, Econometrica 61 (5) (1993) 1019–1045.
[14] E. Kalai, E. Lehrer, Subjective equilibrium in repeated games, Econometrica 61 (5) (1993) 1231–1240.
[15] R.B. Myerson, Multistage games with communication, Econometrica 54 (2) (1986) 323–358.
[16] Y. Shoham, R. Powers, T. Grenager, If multi-agent learning is the answer, what is the question? Stanford University Discussion Paper, 2006.