Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Artificial Intelligence 171 (2007) 440–447

Artificial Intelligence

www.elsevier.com/locate/artint

A hierarchy of prescriptive goals for multiagent learning

Martin Zinkevich[a,*], Amy Greenwald[b], Michael L. Littman[c]

[a] *University of Alberta, Edmonton, AB, Canada*
[b] *Brown University, Providence, RI, USA*
[c] *Rutgers University, New Brunswick, NJ, USA*

Received 13 May 2006; received in revised form 15 February 2007; accepted 15 February 2007

Available online 30 March 2007

Abstract

A great deal of theoretical effort in multiagent learning involves either embracing or avoiding the inherent symmetry between the problem and the solution. Regret minimization is an approach to the prescriptive, non-cooperative goal that explicitly breaks this symmetry, but, since it makes no assumptions about the adversary, it achieves only limited guarantees. In this paper, we consider a hierarchy of goals that begins with the basics of regret minimization and moves towards the utility guarantees achievable by agents that could also guarantee converging to a game-theoretic equilibrium.

1. Introduction

The prescriptive, non-cooperative goal set forth in Shoham et al. is to design intelligent agents that perform well in the presence of other intelligent agents. Much of the research and analysis in this domain involves either bypassing the circularity of this objective (as in regret minimization) or embracing it (as in equilibrium or self-play analyses).

In this paper, we will try to unravel this objective by consciously breaking the symmetry between the *agent*, whose behavior we can prescribe, and the *environment* (other players[1]) over which we have no control. Much of the regret minimization literature also speaks from this perspective: in fact, concepts like calibration have origins in predicting the weather, hardly a multiagent problem!

Nonetheless, without making some assumptions about an agent's environment, there are some guarantees that simply cannot be achieved, such as convergence to the set of Nash or correlated equilibria. This observation has prompted others in the past to study self-play, where the environment is assumed to exhibit the same behavior as the agent. We believe that self-play is an example of a restriction upon the environment that limits the applicability of many results. The fact that an agent performs well in self-play says nothing about how that agent might interact with humans or agents designed by others.

Instead, we believe a more useful focus is on characterizations of environments that can be made in an agent-independent way, as described below.

[*] Corresponding author.
E-mail addresses: maz@cs.ualberta.ca (M. Zinkevich), amy@cs.brown.edu (A. Greenwald), mlittman@cs.rutgers.edu (M.L. Littman).

[1] We will use *player* to refer to the agent or the environment.

0004-3702/$ – see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.artint.2007.02.005

Different learning agents can be considered to be trading off between the breadth of the set of environments they work on and how well they work on that set. Or, to focus on the set of environments apart from an agent, three relevant attributes of an environment class come to mind:

- *Workability*: *What type of guarantees can be achieved against the class?* By definition, agents care about utility (although how they should go about earning utility is a matter of some debate). In this paper, for concreteness, we focus on a particular form of utility guarantees, and then discuss the environment sets in which it is possible to achieve those guarantees.
- *Breadth*: *How broad is the class?* As MAL researchers, we wish to develop agents that can achieve guarantees for as broad an environment class as possible.
  Consider the result that no-internal regret[2] agents [2,3] (we call this set A_{NIR}) converge (in the empirical frequency of joint actions) to the set of correlated equilibria when playing against *any* environment that is no-internal regret (we call this set E_{NIR}). An interesting question arises: what if the environment is not no-internal regret? Moreover, what if we only cared about getting the *utility* of a stationary correlated equilibrium instead of converging to the equilibrium itself?
  For instance, we could add to E_{NIR} (although not to A_{NIR}) those environments that play stationary Nash equilibria, and the NIR agents would still get a high utility.[3] Oddly, we could not add altruistic environments that play as if their opponent's utilities were their own and have no-internal regret: for instance, if the environment has a dominant strategy, it might lead to easy cooperation, whereas both players even with identical utilities might have a difficult time making a common choice. Thus, as we expand this environment set, not only do we get a stronger guarantee, but we also learn more about what makes the original result work in the first place.
- *Saliency*: *Are there agents in the class that perform well on the class? Are there agents that could be considered intelligent?* Throughout our description of environment sets, we avoid circularity by not addressing saliency directly. However, we hope to find the environments that are ultimately derived (using the principles of workability and breadth) to be a *superset* of the environments we consider intelligent, or at least to overlap significantly. Whether this outcome prevails or not will be the primary measure of the success of our proposed agent/environment split.

Normally, when researchers formulate an environment class, they either consider the set of all environments or they begin with some concept of saliency (such as self-play).[4] Between these two extremes in multiagent learning—guarantees that can be achieved in every environment and guarantees that can be achieved in self-play—there lies a hierarchy:

Definition 1. A *hierarchy of prescriptive goals* is:

(1) A hierarchy of environment sets $\{S_1, \dots, S_k\}$, each one contained in its predecessor.
(2) A hierarchy of guarantees $\{G_1, \dots, G_k\}$,

such that there exists a single agent that, for every i, satisfies G_i with every environment in S_i.

In the large margin structural risk-minimization literature, there is a hierarchy of hypothesis spaces, with large margin hypotheses nearer to the top of the hierarchy and smaller margin hypotheses nearer to the bottom. Assume that the data agrees with some hypothesis, given a supervised-learning problem. If it agrees only with a small margin hypothesis, then a large amount of training data is required for good generalization performance. This guarantee

[2] They are referred to as no-regret agents by Foster and Vohra [2]. We use the terminology of Greenwald and Jafari [4] to distinguish no-internal and no-external regret.

[3] In particular, the agents in A_{NIR} would still CEV work with the environments that play stationary Nash equilibria (see Section 3). However, the empirical joint action frequency might no longer converge to the set of correlated equilibria. Moreover, the agents that play stationary Nash equilibria do not work with E_{NIR}, because that the E_{NIR} algorithms will converge to a best response to a Nash equilibrium, not a Nash equilibrium itself.

[4] Another popular choice is the set of stationary environments. The advantage of this set is that the stationarity assumption is made in classification tasks, and therefore supervised-learning techniques can carry over. However, there is no reason to believe that intelligent agents should exhibit stationary behavior. They, too, should learn.

near the bottom of the hierarchy is weak. On the other hand, if a large margin hypothesis agrees with the data, then the generalization performance is excellent with only a few training examples. A single algorithm—a support-vector machine—can achieve all of these guarantees. If one such algorithm did not exist, then this hierarchy of generalization guarantees and hypotheses spaces could only be viewed as a collection of goals, and could not be interpreted as a single, overarching goal. Similarly, we view our proposed hierarchy of prescriptive goals as a single, overarching goal for multiagent learning.

Our perspective is similar to the objectives specified by Bowling [1], who suggests that agents should be designed to achieve two different guarantees (convergence to Nash equilibrium and best response) against two different classes of environments (self-play and stationary). Even more closely related to what we are proposing here are two results about no-internal regret agents: namely, that they minimize internal regret in an arbitrary environment and that their empirical joint action frequency converges to the set of correlated equilibria. In this work, we will focus on the implications upon utility of these two guarantees, and we will analyze whether a broader class of environments can be handled at each level.

It is not only the theoretical elegance of structural risk minimization that has excited machine-learning researchers; what is most impressive is its practical implications. The advantage to an empiricist of using such a technique is that it allows her to give a soft boundary on what she expects to happen. Thus, an algorithm can use the data to decide exactly how accurate the empiricist's assumptions are and to relax them as necessary to deal with real-world occurrences. Whether through regret methods or the hierarchies described here, *soft assumptions* which, if true, result in a performance guarantee, but, if false, do not preclude the algorithm from performing reasonably well,[5] are useful for developing agents that perform well in practice.

2. A model of interaction: Repeated bimatrix games

In this paper, we restrict our attention to repeated bimatrix games, as they are sufficient to model many of the fundamental issues in multiagent learning.[6] Moreover, we assume that the game, including its utility functions, is known to both players. Although this assumption will bias our discussion, we believe it is a good place to begin.

We will talk about a game G, which formally is a tuple (A_1, A_2, u_1, u_2), A_1 being the actions for the first player, A_2 for the second, $u_1 : A_1 \times A_2 \to \mathbf{R}$ and $u_2 : A_1 \times A_2 \to \mathbf{R}$ the utilities for the first and second players, respectively. We define $S_{\mathrm{CE}}(G)$ to be the set of correlated equilibria of the single-shot variant of G and $S_{\mathrm{NE}}(G)$ to be the set of Nash equilibria of the single-shot variant of G. As usual, we will refer to the history of the game as $h \in (A_1 \times A_2)^\infty$, with h_t being the joint action at round t.

We will refer to the first player and its behavior as the *agent*, and the second player and its behavior as the *environment*.

3. A utility guarantee as a goal

Before we discuss possible behaviors for agent and environment, we will fix a concept of what it means to perform well. In a repeated bimatrix game, goals are often stated in terms of utility: best response, regret minimization, etc. Because we are attempting to develop objectives first so that we can determine in which environments they can be achieved later, we will choose a utility objective that is independent of the behavior of the environment or the agent that produces this behavior.

Definition 2. Given a game G, and a real number v, an agent σ (G, v)-*works* with an environment ρ if, assuming the history h is the outcome of σ and ρ playing G repeatedly, with probability 1,

$$\liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} u_1(h_t) \geqslant v. \tag{1}$$

We say that ρ is (G, v)-*workable*.

[5] An example of a soft assumption is: the true hypothesis has a large margin.

[6] For instance, the bimatrix game formulation can represent arbitrary extensive-form games, and the notion of repetition facilitates learning.

There are many natural values that agents have been proven to achieve:

(1) The *safe value* v of a game G is the minimax value (to player 1) of the game:

$$\mathrm{safe}(G) = \max_{a_1 \in A_1} \min_{s_2 \in \Delta(A_2)} u_1(a_1, s_2). \tag{2}$$

(2) The *correlated equilibrium value* (CEV) of a game G is the minimum value (to player 1) across the set of correlated equilibria of the game:

$$\mathrm{CEV}(G) = \min_{s \in S_{\mathrm{CE}}(G)} u_1(s). \tag{3}$$

(3) The *Nash equilibrium value* (NEV) of a game G is the minimum value (to player 1) across the set of Nash equilibria of the game:

$$\mathrm{NEV}(G) = \min_{s \in S_{\mathrm{NE}}(G)} u_1(s). \tag{4}$$

Fact 1. *The* NEV *of a game is no less than the* CEV*, and the* CEV *is no less than the safe value.*

Fact 2. *There is a game where the* NEV *exceeds the* CEV*, and a game where the* CEV *exceeds the safe value.*

In other words, an agent works (at a certain level v) if the average utility eventually approaches or exceeds v. Let us relate this concept to others in game theory.

(1) If an agent σ plays a best response to an environment ρ in a game G (in terms of immediate utility on every round), then σ $(G, \mathrm{safe}(G))$-works with ρ.
(2) If an agent σ minimizes external regret [5] in G, for any environment ρ, then σ $(G, \mathrm{safe}(G))$-works with ρ.
(3) If an agent σ minimizes internal regret in a game G, for any environment ρ that minimizes internal regret, σ $(G, \mathrm{CEV}(G))$-works with ρ.
(4) If an agent σ is a best response to an environment ρ in a game G (in terms of immediate utility on every round), and ρ is a best response to σ, then σ $(G, \mathrm{NEV}(G))$-works with ρ.

From the first point, we note that a best response can be guaranteed to achieve only the safe value when facing an *arbitrary* environment. From the fourth point, however, if the environment is such that it always plays a best response as well, the (no smaller and possibly larger) Nash equilibrium value is achieved. Thus, the natural concepts of "learnable" (i.e., like PAC-learnable or learnable in the statistical query model) do not apply to these domains, because *perfectly predicting the environment is not enough to guarantee good performance*. The level of performance achieved is contingent, in part, on properties of the environment itself. Moreover, most machine-learning research focuses on learning independent and identically distributed sequences, or at most some stationary Markov process, whereas the environments of interest here can themselves learn and are therefore likely to be more difficult to model.[7]

Observe that we have constructed a hierarchy: for a given game G, a given ρ, an agent σ that $(G, \mathrm{NEV}(G))$ works with an environment ρ also $(G, \mathrm{CEV}(G))$ works with ρ, and an agent σ that $(G, \mathrm{CEV}(G))$ works with ρ also $(G, \mathrm{safe}(G))$ works with ρ. While it is a hierarchy of guarantees, these are guarantees on a single environment, not a class. Therefore, we extend the concept of workable to sets.

Definition 3. Given a game G, and a real number v, an agent σ (G, v)-*works* with a class of environments $\mathcal{C}$ if for all $\rho \in \mathcal{C}$, σ (G, v)-*works* with ρ. $\mathcal{C}$ is said to be (G, v)-workable.

To connect this definition to the machine-learning literature, it is easiest to think of environment classes as hypothesis classes. "Workable" is analogous to the concept of "learnable".

Given a game G, the hierarchy of goals we consider here, while it is certainly not the only one possible, is:

[7] We do not deny the fact that being able to predict the environment's behavior is a powerful tool. It is quite possible that by formalizing these environment classes, they may be used as the basis for priors of learning agents.

(1) The agent is no-internal regret.[8]
(2) The agent $(G, \mathrm{CEV}(G))$-works.
(3) The agent $(G, \mathrm{NEV}(G))$-works.

The first two levels are the two guarantees about no-internal regret algorithms in a utility guarantee form. The third represents the utilitarian part of a much discussed objective: converging to a stationary Nash equilibrium (in utility, at least).

Throughout the remainder of the paper, we will focus on this hierarchy of guarantees, and investigate the hierarchy of environment sets that matches it well. We call this goal hierarchy the *Hierarchy of Equilibrium Utilities*. In the next section, we formalize the environment set that corresponds to this hierarchy. We give the full definition of this hierarchy at the end of the next section.

4. Developing a hierarchy of environment sets

Let us begin with the broadest possible class—the set of all environments for S_1, because there are agents that have no internal regret in any environment.

It is natural to consider as the next environment class one that is $(G, \mathrm{CEV}(G))$-workable. One likely candidate would be the no-internal regret environments (NIR). Observe that an agent that can $(G, \mathrm{CEV}(G))$-work with the NIR environments could also be NIR with all environments. The existence of such an agent is an example of achieving a hierarchy of prescriptive goals.

Now, let us forget about its origins, and consider the NIR environments as a hypothesis space. If we consider the principal property of this set to be that it is $(G, \mathrm{CEV}(G))$-workable, can we expand it to achieve a larger hypothesis space/environment class?

This metaphor is quite crucial: when machine-learning researchers realize that they can broaden their horizons virtually for free, they do so without hesitation. Such an expansion may involve hypotheses that are unlikely or even downright nonsensical, but it is a small price to pay for incorporating additional hypotheses. In our work here and in the work of others, most formally defined environment classes that are sufficiently broad will have both reasonable and unreasonable environments. Arguably, the guiding principle of machine learning is to get the best bounds (or practical performance) with the broadest hypothesis space. Correspondingly, a critical part of our theory here will be the *maximal* $(G, \mathrm{CEV}(G))$-*workable sets*: environment sets for which there exists no strict superset that is also a $(G, \mathrm{CEV}(G))$-workable set.

Now, there are two problems with such a concept, one technical and one philosophical:

(1) In general, it is not obvious that one such maximal set exists: it could be that one could always add one more environment and still be $(G, \mathrm{CEV}(G))$-workable, although we have not constructed such an example.
(2) By constructing a maximal set, we may go too far. For instance, it may be that some environments are workable only with agents whose first three actions are “left”, “right”, “right”, whereas other environments might be workable only with agents that begin by choosing “right”, “left”, “left”. Thus, whereas we are trying to design a canonical maximal set, we may find a plethora of maximal sets, not one environment class for which to develop prescriptive agents.

Consider an example of this last point, a game where the utility functions are equal and each player has a strictly dominant strategy. If one agent plays foolishly, then the other will also pay the penalty. It could be that an environment will play its dominant strategy only if it sees a certain sequence of actions in the beginning, say “left”, “right”, “right”. Otherwise, it will play a dominated action. Such a game may seem contrived, but it serves to highlight an issue that exists but is more subtle in more ordinary games—an agent may choose to not play its dominant action for arbitrary reasons.

Hence, if we want a general purpose class of environments, we may not wish it to be maximal, possibly enabling an agent to focus on a peculiarity of a small subset of environments (even if the maximal set is much larger). Loosely

[8] We include no-internal regret because it is stronger than $(G, \mathrm{safe}(G))$-working and can still be applied to all environments.

speaking, it may be more productive to consider the environments that are common to all maximal sets—the intersection. More formally:

Definition 4. An environment ρ is in the canonical (G, v)-workable set if for every (G, v)-workable set S, $S \cup \{\rho\}$ is also a (G, v)-workable set.

Suppose you have some agent σ that (G, v)-works with a set of environments S. If you add some element ρ from the canonical (G, v)-workable set, then if your agent does not (G, v)-work with $S \cup \{\rho\}$, there will be another agent σ' that does. Therefore, given a workable set, adding an environment from the canonical set will not make the set unworkable.

This canonical concept avoids those environments that force a particular line of play: if one environment requires an agent to always go left, and the other to always go right, neither will be in the canonical set.

One final, more practical tweak to building workable sets is to build the hierarchy of goals from the bottom upwards. For instance, we can say for a given existing hierarchy that there is a certain set of agents A that can satisfy all the properties of that hierarchy. When we design higher levels of the hierarchy, we should restrict the agents we consider included in this set to A, otherwise we will develop an unachievable set of goals.

Given a game G, we can design agents that have no-internal regret with all environments. However, it is not enough to have a $(G, \mathrm{CEV}(G))$-workable set at the next level. It must be $(G, \mathrm{CEV}(G))$-workable with an NIR agent, otherwise no agent will be able to satisfy both goals. Thus, we only consider environment sets to be $(G, \mathrm{CEV}(G))$-workable if they work with an agent that is NIR.

Definition 5. Given a game G, a tuple of goals $(g_1 \dots g_k)$, a tuple of environment sets $(S_1 \dots S_k)$, an agent is $((g_1 \dots g_k), (S_1 \dots S_k))$ *viable* if for all i, it achieves guarantee g_i with set S_i. Given a new goal value $v' \in \mathbf{R}$, the *relative canonical set* S' with respect to $((g_1, \dots, g_k), (S_1 \dots S_k))$ is the set of all environments ρ such that for any set S' where S' is (G, v')-workable with a viable agent σ, there exists a viable agent σ' that (G, v') works with $S' \cup \{\rho\}$.

One way to envision this construction is that we are creating a hierarchy from the bottom up, one level at a time. Before we add a new environment to the highest level, we first make sure it does not interfere with our existing structure.

Definition 6. The *Hierarchy of Equilibrium Utilities* is the hierarchy of guarantees:

(1) g_1: the agent is no-internal regret.
(2) g_2: the agent $(G, \mathrm{CEV}(G))$ works.
(3) g_3: the agent $(G, \mathrm{NEV}(G))$ works.

and the hierarchy of environments:

(1) S_1: all environments.
(2) S_2: relative canonical $(G, \mathrm{CEV}(G))$ set with respect to $((g_1), (S_1))$.
(3) S_3: relative canonical $(G, \mathrm{NEV}(G))$ set with respect to $((g_1, g_2), (S_1, S_2))$.

5. Salience

After a hierarchy is developed, one can measure its practicality in part by noticing where existing algorithms and models of intelligence fit in. In particular, if one comes up with a hierarchy with no-internal regret environments near the top, that would provide evidence that the classes are salient.

Another important aspect of salience involves the agents that actually achieve the hierarchy of goals. Are some of them in the smallest environment classes? If they are, then an emergent (not deliberate) result follows: they achieve the highest guarantee in the hierarchy in self-play.

Theorem 7. *In the Hierarchy of Equilibrium Utilities, for any game G, all* NIR *environments are in S_2, and all stationary Nash equilibrium environments are in S_3.*

Proof. Here, we prove the first result: the second result is derived similarly. The (g_1, S_1) viable agents are exactly the NIR agents. Therefore, if an environment set is S' $(G, \mathrm{CEV}(G))$ workable with a viable agent σ, then σ being viable is an NIR agent, and if ρ is an NIR environment, then σ $(G, \mathrm{CEV}(G))$ works with ρ and therefore σ $(G, \mathrm{CEV}(G))$ works with $S' \cup \{\rho\}$. This establishes that the environment set S_2 is salient. □

One might also consider behaviors in the higher environment levels as candidate agents. These environments can be lifted over the artificial divide we have constructed between agents and environments, and their performance as agents can be analyzed.

6. A continuum of guarantees

Although CEV and NEV are the utility guarantees corresponding to existing guarantees of play, interesting utility levels vary from game to game. This point brings up the challenge of how to design a hierarchy of environments for the hierarchy of guarantees consisting of *all* utility guarantees for a game. More mathematical difficulties arise than in the finite-guarantee case; nonetheless, such a hierarchy would be of interest for two reasons. First, this hierarchy of goals could be construed as defining a new concept of rationality in the game. Second, if an agent achieved these hierarchical goals, the ensuing behavior of a pair of these agents (one for each player) would yield a new type of equilibrium.

7. Open problems

What we have presented here is a mostly framework to guide the development of multiagent-learning algorithms. Nonetheless, there are several open mathematical problems that can be derived from the examples we give here. For instance, the following prescriptive goals could be addressed:

Open Problem 8. Can an agent be constructed for any bimatrix game that satisfies all the goals in the Hierarchy of Equilibrium Utilities?

Open Problem 9. Where do no-internal regret environments reside in the Hierarchy of Equilibrium Utilities for different games? Are they always in the NEV-workable level of the hierarchy? Could no-internal regret agents CEV work with the CEV-workable level of the hierarchy?

Open Problem 10. How do Bayesian techniques (such as fictitious play) fit into this framework? Where do they fit as environments? As agents?

One could ask such questions about any technique that has been applied to bimatrix games in the past. Also, with regards to saliency, there are two descriptive goals:

Open Problem 11. In the NEV-workable and CEV-workable environments, are there behaviors that would serve to model humans?

Open Problem 12. Would the agents that work with these environment sets serve to model humans?

From a more philosophical perspective, questions could be posed about other hierarchies of guarantees besides the Hierarchy of Equilibrium Utilities. One could also consider higher level guarantees, such as achieving a particular utility quickly. For instance, one could develop an uncountably infinite set of environment classes depending upon how quickly $\mathrm{CEV}(G)$ is approached.

8. Discussion

Some AI researchers have ideas about what human learning is like and design learning algorithms guided by their ideas. Since our ideas of what human behavior is really like are much more vague than our ideas about how a car moves or how light travels, this design process can easily go astray.

If we consider the prescriptive goal to be to design algorithms that work well with *humans*, the most natural solution to this problem would be to develop a descriptive model of human behavior independent of attempting to solve any prescriptive goals. This tack, however, requires us to first understand what learning could be like. Since humans are at their very essence social animals, to model how they behave we must first understand what behaviors are relevant in social environments, and then discover which of these best represents humans. However, this path returns us the very essence of the prescriptive goal.

Thus, in the end we are left with a problem: to address descriptive goals in multiagent learning, we must first address prescriptive goals, and to attack the prescriptive goals, we must first deal with the descriptive ones.

In this paper, we discussed how to bootstrap this process. We gave an example of a hierarchy of objectives for prescriptive agents. These objectives were not meant to be indicative of what intelligence is, but rather what intelligence *can deal with*. Thus, at the lowest level of this process, we consider the classes of behaviors (environments) against which an objective can be achieved. Given this beginning, we can move naturally towards what intelligence is, and then hopefully to a better understanding of human intelligence and, more generally, social intelligence.

The hypothesis of this paper is that an important aspect of multiagent learning is the development and understanding of classes of agents that may not all be intelligent in and of themselves but can be “handled” by intelligent agents. We believe that such an objective rests at the foundation of multiagent learning.

9. Conclusion

Often when multiagent-learning researchers consider a set of environments, they are guided by concerns of salience and workability: i.e., they select an environment that seems intelligent and then try to build an agent to work with it, tweaking agent and environment until some (hopefully) useful property emerges. What we would like to highlight that is missing from this process is a desire for *breadth.* At the extreme, one can consider the set of all environments. While some interesting guarantees (e.g., regret minimization) can be obtained under this assumption, one cannot derive a blanket guarantee of utility (or achieve many other prescriptive goals) across all environments beyond the safe value.

The main point of this paper is that, when working on the prescriptive, non-cooperative agenda of multiagent learning, there is a fundamental difference between the way one should consider algorithms one is designing and algorithms one is facing. Whereas saliency is desirable in the algorithms one designs, the environments one considers should be designed according to workability and breadth more than saliency. In particular, the perspective when considering an agent and the perspective when considering an environment are asymmetric.

Hence, in this paper, we have argued for a hierarchy of prescriptive goals, analogous to the structural risk minimization framework in binary classification. We believe that developing such a hierarchy of goals is crucial for grasping the basic nature of social intelligence: i.e., the ability to interact effectively in a multiagent setting.

Acknowledgements

We would like to thank Michael Bowling for some fundamental guidance in the direction of this paper. This research was supported by NSF career grant #IIS-0133689, and by Alberta Ingenuity through the Alberta Ingenuity Center for Machine Learning.

References

[1] M. Bowling, Multiagent learning in the presence of agents with limitations, Ph.D. thesis, Carnegie Mellon University, 2003.

[2] D. Foster, R. Vohra, Calibrated learning and correlated equilibrium, Games and Economic Behavior 21 (1) (1997) 40–55.

[3] D. Foster, R. Vohra, Regret in the on-line decision problem, Games and Economic Behavior 29 (1) (1999) 7–35.

[4] A. Greenwald, A. Jafari, A general class of no-regret algorithms and game-theoretic equilibria, in: Proceedings of the 2003 Computational Learning Theory Conference, August 2003, pp. 1–11.

[5] J. Hannan, Approximation to Bayes risk in repeated play, Contribution to the Theory of Games 3 (1957) 97–139.