A Note on Strategic Learning in Policy Space

Steven O. Kimbrough[1], Ming Lu[2], and Ann Kuo[3]

[1] University of Pennsylvania, Philadelphia, PA, USA,
kimbrough@wharton.upenn.edu
[2] University of Pennsylvania, Philadelphia, PA, USA,
milu@wharton.upenn.edu
[3] University of Pennsylvania, Philadelphia, PA, USA,
kuo2_99@yahoo.com

Abstract. We report on a series of computational experiments with artificial agents learning in the context of games. Two kinds of learning are investigated: (1) a simple form of associative learning, called Q-learning, which occurs in state space, and (2) a simple form of learning, which we introduce here, that occurs in policy space. We compare the two methods on a number of repeated 2×2 games. We conclude that learning in policy space is an effective and promising method for learning in games.

1 Introduction

In what follows we report on a series of computational experiments with artificial agents learning in the context of games (situations of interdependent decision making). We discuss two kinds of learning. The first is a form of simple associative learning, called Q-learning (a form of reinforcement learning) in the machine learning literature.[1] This sort of reinforcement learning by agents in games has been investigated previously by a number of researchers.[2] We introduce, in our discussion of the second series of experiments, a new variety of reinforcement learning in the context of games. This kind of learning appears to be both very effective from the point of view of the agents and cognitively plausible.

We begin by providing, in the next two sections, some essential context and background.

2 Background: Games and Decisions

Decision contexts faced by agents may be distinguished into those that are and those that are not *strategic*. In *non-strategic contexts,* decisions by other agents do not need to be taken into account. When, for example, one decides

[1] Cf., [KLM96,SB98].
[2] E.g., [Cam03], [KL03], [KL04], and [SC95].

whether to dress for rain on a given day, what matters most is whether it will in fact rain.[3] This fact, whichever way it turns out, is not in any way dependent on the decisions of another agent. Nature, no matter how we joke otherwise, does not care about and does not have interests in whether we get wet or not. Similarly, the decisions made by an animal in navigating towards a goal do not (at least in many cases) need to take into account decisions by other animals. Constructing and exploiting a map gets the agent/animal home, not, e.g., negotiation with others.

Strategic decisions—the subject of the theory of games—are those for which the outcome depends on the agent's choice and the non-agentive environment (as in non-strategic decisions), as well as on decisions made by other agents. Encountering someone approaching on the sidewalk, one has to decide whether to swerve left or right. The success of the maneuver will typically depend upon a corresponding decision made by the other agent. The two agents are playing a game—are interacting strategically—in which both are rewarded maximally if each swerves right or if each swerves left; and both are punished, or receive a lesser reward, if one swerves right and the other swerves left.

	C_L (Cooperate)	C_R (Defect)
R_U (Cooperate)	(R,R)*	(S,T)*
R_D (Defect)	(T,S)*	(P,P)#

Fig. 1. Prisoner's Dilemma. T>R>P>S. 2R>T+S. #=Nash equilibrium. * = Pareto-optimal outcome.

To illustrate (with a different game), Figure 1 shows the generic and famous Prisoner's Dilemma game in strategic form. There are are two players, Row and Column. Row has a choice between R_U (mnemonic: up) and R_D (down), while Column chooses between C_L (left) and C_R (right). The outcome (R_D, C_R), in which both players receive P, is said to be a Nash equilibrium because neither player could do better by unilaterally changing its choice of strategy. Row's only alternative choice is R_U, which would yield Row a return of S<P, and similarly for Column. Classical game theory predicts that game outcomes will occur at Nash equilibria. An outcome is said to be (strictly) Pareto-optimal if there is no other outcome at which all (here, both) players can do better. In the Prisoner's Dilemma all of the outcomes *except* the Nash equilibrium are Pareto-optimal. The outcome (R_U, C_L), said to be the result of mutual cooperation, is especially attractive from the players' perspective, since *both* do better than they do at the Nash equilibrium.

[3] So we shall assume for the sake of the example. Nothing is ever so simple. One's loss function—the value one places on staying dry—does matter as much as whether it rains or not.