Learning Payoff Functions in Infinite Games

Yevgeniy Vorobeychik, Michael P. Wellman, and **Satinder Singh**
University of Michigan
Artificial Intelligence Laboratory
1101 Beal Avenue
Ann Arbor, MI 48109-2110 USA
{ yvorobey, wellman, baveja }@umich.edu

Abstract

We consider a class of games with real-valued strategies and payoff information available only in the form of data from a given sample of strategy profiles. Solving such games with respect to the underlying strategy space requires generalizing from the data to a complete payoff-function representation. We address payoff-function learning as a standard regression problem, with provision for capturing known structure (symmetry) in the multiagent environment. To measure learning performance, we consider the relative utility of prescribed strategies, rather than the accuracy of payoff functions per se. We demonstrate our approach and evaluate its effectiveness on two examples: a two-player version of the first-price sealed-bid auction (with known analytical form), and a five-player market-based scheduling game (with no known solution).

Introduction

Game-theoretic analysis typically begins with a complete description of strategic interactions, that is, *the game*. We consider the prior question of determining what the game actually is, given a database of game experience rather than any direct specification. This is one possible target of *learning* applied to games (Shoham, Powers, & Grenager, 2003). When agents have few available actions and outcomes are deterministic, the game can be identified through systematic exploration. For instance, we can ask the agents to play each strategy profile in the entire joint strategy set and record the payoffs for each. If the joint action space is small enough, limited nondeterminism can be handled by sampling. Coordinating exploration of the joint set does pose difficult issues. Brafman and Tennenholtz, for example, address these carefully for the case of common-interest stochastic games (Brafman & Tennenholtz, 2003b), as well as the general problem of maintaining an equilibrium among learning algorithms (Brafman & Tennenholtz, 2003a).

Further difficulties are posed by intractably large (or infinite) strategy sets. We can make this problem tractable by reducing the number of profiles that agents are allowed to play, but this comes at the cost of transforming the game of interest into a different game entirely. Instead, we seek to identify the full game (or at least a less restrictive game) from limited data, entailing some generalization from observed instances. Approximating payoff functions using supervised learning (regression) methods allows us to deal with continuous agent strategy sets, providing a payoff for an arbitrary strategy profile. In so doing, we adopt functional forms consistent with prior knowledge about the game, and also admit biases toward forms facilitating subsequent game analysis (e.g., equlibrium calculation).

In this paper, we present our first investigation of approximating payoff functions, employing regression to low-degree polynomials. We explore two example games, both with incomplete information and real-valued actions. First is the standard first-price sealed bid auction, with two players and symmetric value distributions. The solution to this game is well-known (Krishna, 2002), and its availability in analytical form proves useful for benchmarking our learning approach. Our second example is a five-player market-based scheduling game (Reeves *et al.*, to appear), where time slots are allocated by simultaneous ascending auctions (Milgrom, 2000). This game has no known solution, though previous work has identified equilibria on discretized subsets of the strategy space.

Preliminaries

Notation

Our notation follows Mas-Colell, Whinston, & Green (1995). A generic normal form game is formally expressed as $[I, \{\Delta(S_i)\}, \{u_i(s)\}]$, where I refers to the set of players and $m = |I|$ is the number of players. S_i is the set of strategies available to player $i \in I$, and the set $\Delta(S_i)$ is the simplex of mixed strategies over S_i. Finally, $u_i(s) : S_1 \times \cdots \times S_m \rightarrow \mathcal{R}$ is the payoff function of player i when all players jointly play $s = (s_1, \ldots, s_m)$, with each $s_j \in S_j$. As is common, we assume von Neumann-Morgenstern utility, allowing an agent i's payoff for a particular mixed strategy profile to be

$$u_i(\sigma) = \sum_{s \in S} [\sigma_1(s_1) \cdots \sigma_m(s_m)] u_i(s),$$

where $\sigma_j : S_j \rightarrow [0, 1]$ is a mixed strategy of player i, assigning a probability to each pure strategy $s_j \in S_j$ such that all probabilities over the agent's strategy set add to 1 (i.e. $\sigma_j \in \{\Delta(S_j)\}$).

It will often be convenient to refer to the strategy (pure or mixed) of player i separately from that of the remaining players. To accomodate this, we use s_{-i} to denote the joint strategy of all players other than player i.

Nash Equilibrium

In this paper, we are concerned with one-shot normal-form games, in which players make decisions simultaneously and accrue payoffs, upon which the game ends. This single-shot nature may seem to preclude learning from experience, but in fact repeated episodes are allowed, as long as actions cannot affect future opportunities, or condition future strategies. Game payoff data may also be obtained from observations of other agents playing the game, or from simulations of hypothetical runs of the game. In any of these cases, learning is relevant despite the fact that the game is to be played only once.

Faced with a one-shot game, an agent would ideally play its best strategy given those played by the other agents. A configuration where all agents play strategies that are best responses to the others constitutes a *Nash equilibrium*.

Definition 1 *A strategy profile $s = (s_1, \ldots, s_m)$ constitutes a (pure strategy) Nash equilibrium of game $[I, \{S_i\}, \{u_i(s)\}]$ if for every $i \in I$, $s'_i \in S_i$,*

$$u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i}).$$

A similar definition applies when mixed strategies are allowed.

Definition 2 *A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_m)$ constitutes a mixed strategy Nash equilibrium of game $[I, \{\Delta(S_i)\}, \{u_i(s)\}]$ if for every $i \in I$, $\sigma'_i \in \{\Delta(S_i)\}$,*

$$u_i(\sigma_i, \sigma_{-i}) \geq u_i(\sigma'_i, \sigma_{-i}).$$

In this study we devote particular attention to games that exhibit symmetry with respect to payoffs.

Definition 3 *A game $[I, \{\Delta(S_i)\}, \{u_i(s)\}]$ is symmetric if $\forall i, j \in I$,*

- *$S_i = S_j$, and*
- *$u_i(s_i, s_{-i}) = u_j(s_j, s_{-j})$ whenever $s_i = s_j$ and $s_{-i} = s_{-j}$.*

Symmetric games have relatively compact descriptions and may present associated computational advantages (Cheng *et al.*, 2004). Given a symmetric game, we may focus on the subclass of symmetric equilibria, which are arguably most natural (Kreps, 1990), and avoid the need to coordinate on roles.[1] In fairly general settings, symmetric games do possess symmetric equilibria (Nash, 1951).

Payoff Function Approximation

Problem Definition

We are given a set of data points (s, v), each describing an instance where agents played strategy profile s and realized value $v = (v_1, \ldots, v_m)$. For deterministic games of complete information, v is simply u. With incomplete information or stochastic outcomes, v is a random variable, more specifically an independent draw from a distribution function of s, with expected value $u(s)$.

The *payoff function approximation task* is to select a function $\hat{u}$ from a candidate set $\mathcal{U}$ minimizing some measure of deviation from the true payoff function u. Because the true function u is unknown, of course, we must base our selection on evidence provided by the given data points.

Our goal in approximating payoff functions is typically not predicting payoffs themselves, but rather in assessing strategic behavior. Therefore, for assessing our results, we measure approximation quality not directly in terms of a distance between $\hat{u}$ and u, but rather in terms of the *strategies dictated by* $\hat{u}$ evaluated with respect to u. For this we appeal to the notion of approximate Nash equilibrium.

Definition 4 *A strategy profile $\sigma = (\sigma_1, \ldots, \sigma_m)$ constitutes an ϵ-Nash equilibrium of game $[I, \{\Delta(S_i)\}, \{u_i(s)\}]$ if for every $i \in I$, $\sigma'_i \in \{\Delta(S_i)\}$,*

$$u_i(\sigma_i, \sigma_{-i}) + \epsilon \geq u_i(\sigma'_i, \sigma_{-i}).$$

We propose using ϵ in the above definition as a measure of approximation error of $\hat{u}$, and employ it in evaluating our learning methods. When u is known, we can compute ϵ in a straightforward manner. Let s_i^* denote i's *best response function*, defined by

$$s_i^*(\sigma_{-i}) = x \in \arg\max_{s_i} u_i(s_i, \sigma_{-i}).$$

For clarity of exposition, let us assume that $s_i^*(\sigma_{-i})$ is single-valued. Let $\hat{\sigma}$ be a solution (e.g., a Nash equilibrium) of game $[I, \{\Delta(S_i)\}, \{\hat{u}_i(s)\}]$. Then $\hat{\sigma}$ is an ϵ-Nash equilibrium of the true game $[I, \{\Delta(S_i)\}, \{u_i(s)\}]$, for

$$\epsilon = \max_{i \in I} \left[u_i(s_i^*(\hat{\sigma}_{-i}), \hat{\sigma}_{-i}) - u_i(\hat{\sigma}_i, \hat{\sigma}_{-i}) \right].$$

Since in general u will either be unknown or not amenable to this analysis, we developed a method for estimating ϵ from data. We will describe it in some detail in a later section.

Polynomial Regression

For the remainder of this report, we focus on a special case of the general problem, where action sets are real-valued intervals, $S_i = [0, 1]$, and the class $\mathcal{U}$ of payoff-function hypotheses includes only polynomial expressions. Moreover, we restrict attention to symmetric games, and further limit the number of variables in payoff-function hypotheses by using some form of aggregation of other agents' actions. [2] The assumption of symmetry allows us to adopt the convention for the remainder of the paper that payoff $u(s_i, s_{-i})$ is to the agent playing s_i.

[1] Contention may arise when there are disparities among payoffs in asymmetric equilibrium. Even for symmetric equilibria, coordination issues may still be present with respect to equilibrium selection.

[2] Although none of these restrictions are inherent in the approach, one must of course recognize the tradeoffs in complexity of the hypothesis space and generalization performance. Thus, we strive to build in symmetry to the hypothesis space whenever applicable. The choice of simple polynomial forms and restrictions on the number of variables were adopted purely for convenience in this initial investigation.

One class of models we consider are the *nth-degree separable polynomials*:

$$u(s_i, \phi(s_{-i})) = a_n s_i^n + \cdots + a_1 s_i + \\ + b_n \phi^n(s_{-i}) + \cdots + b_1 \phi(s_{-i}) + d, \quad (1)$$

where $\phi(s_{-i})$ represents some aggregation of the strategies played by agents other than i. For two-player games, ϕ is simply the identity function. We refer to polynomials of the form (1) as separable, since it lacks terms combining s_i and s_{-i}. We also consider models with such terms, for example, the *non-separable quadratic*:

$$u(s_i, \phi(s_{-i})) = a_2 s_i^2 + a_1 s_i + b_2 \phi^2(s_{-i}) + \\ + b_1 \phi(s_{-i}) + c s_i \phi(s_{-i}) + d. \quad (2)$$

Note that (2) and (1) coincide in the case $n = 2$ and $c = 0$. In the experiments described below, we employ a simpler version of non-separable quadratic that takes $b_1 = b_2 = 0$.

One advantage of the quadratic form is that we can analytically solve for Nash equilibrium. Given a general non-separable quadratic (2), the necessary first-order condition for an interior solution is

$$s_i = -\frac{a_1 + c\phi(s_{-i})}{2a_2}.$$

This reduces to

$$s_i = -\frac{a_1}{2a_2}$$

in the separable case. For the non-separable case with additive aggregation, $\phi(s_{-i}) = \sum_{j \neq i} s_j$, we can derive an explicit first-order condition for *symmetric* equilibrium:

$$s_i = -\frac{a_1}{2a_2 + (m-1)c}.$$

In the experiments that follow, whenever our approximation yields no pure strategy Nash equilibrium, we randomly select a symmetric profile from the joint strategy set. If, on the other hand, we find multiple pure equilibria, a sample equilibrium is selected arbitrarily.

Strategy Aggregation

As noted above, we consider payoff functions on two-dimensional strategy profiles in the form

$$u(s_i, s_{-i}) = f(s_i, \phi(s_{-i})).$$

As long as the output of $\phi(s_{-i})$ is the same for different permutations of the same strategies in s_{-i}, the payoff function is symmetric. Since the actual payoff functions for our example games are also known to be symmetric, we constrain that $\phi(s_{-i})$ preserve the symmetry of the underlying game.

In our experiments, we compared three variants of $\phi(s_{-i})$. First and most compact is the simple sum, $\phi_{sum}(s_{-i}) = \sum_{j \neq i} s_j$. Second is the ordered pair (ϕ_{sum}, ϕ_{ss}), where $\phi_{ss}(s_{-i}) = \sum_{j \neq i} (s_j)^2$. The third variant, $\phi_{identity}(s_{-i}) = s_{-i}$, simply takes the strategies in their direct, unaggregated form. To enforce the symmetry requirement in this last case, we sort the strategies in s_{-i}.

First-Price Sealed-Bid Auction

In the standard first-price sealed-bid (FPSB) auction game (Krishna, 2002), agents have private valuations for the good for sale, and simultaneously choose a bid price representing their offer to purchase the good. The bidder naming the highest price gets the good, and pays the offered price. Other agents receive and pay nothing. In the classic setup first analyzed by Vickrey (1961), agents have identical valuation distributions, uniform on $[0, 1]$, and these distributions are common knowledge. The unique (Bayesian) Nash equilibrium of this game is for agent i to bid $\frac{m-1}{m} x_i$, where x_i is i's valuation for the good.

Note that strategies in this game (and generally for games of incomplete information), $b_i : [0, 1] \rightarrow [0, 1]$, are functions of the agent's private information. We consider a restricted case, where bid functions are constrained to the form

$$b_i(x_i) = k_i x_i, \; k_i \in [0, 1].$$

This constraint transforms the action space to a real interval, corresponding to choice of parameter k_i. We can easily see that the restricted strategy space includes the known equilibrium of the full game, with $s_i = k_i = \frac{m-1}{m}$ for all i, which is also an equilibrium of the restricted game in which agents are constrained to strategies of the given form.

We further focus on the special case $m = 2$, with corresponding equilibrium at $s_1 = s_2 = 1/2$. For the two-player FPSB, we can also derive a closed-form description of the actual expected payoff function:

$$u(s_1, s_2) = \begin{cases} 0.25 & \text{if } s_1 = s_2 = 0, \\ \frac{(s_1 - 1)\left[(s_2)^2 - 3(s_1)^2\right]}{6(s_1)^2} & \text{if } s_1 \geq s_2, \\ \frac{1 - s_1}{3 s_2} & \text{otherwise.} \end{cases} \quad (3)$$

The availability of known solutions for this example facilitates analysis of our learning approach. Our results are summarized in Figure 1. For each of our methods (classes of functional forms), we measured average ϵ for varying training set sizes. For instance, to evaluate the performance of separable quadratic approximation with training size N, we independently draw N strategies, $\{s^1, \ldots, s^N\}$, uniformly on $[0, 1]$. The corresponding training set comprises N^2 points: $((s^i, s^j), u(s^i, s^j))$, for $i, j \in \{1, \ldots, N\}$, with u as given by (3). We find the best separable quadratic fit $\hat{u}$ to these points, and find a Nash equilibrium corresponding to $\hat{u}$. We then calculate the least ϵ for which this strategy profile is an ϵ-Nash equilibrium with respect to the actual payoff function u. We repeat this process 200 times, averaging the results over strategy draws, to obtain each value plotted in Figure 1.

As we can see, both second-degree polynomial forms we tried do quite well on this game. For $N < 20$, quadratic regression outperforms the model labeled "sample best", in which the payoff function is approximated by the discrete training set directly. The derived equilibrium in this model is simply a Nash equilibrium over the discrete strategies in the training set. At first, the success of the quadratic model may be surprising, since the actual payoff function (3) is only piecewise differentiable and has a point of discontinuity. However, as we can see from Figure 2, it appears quite

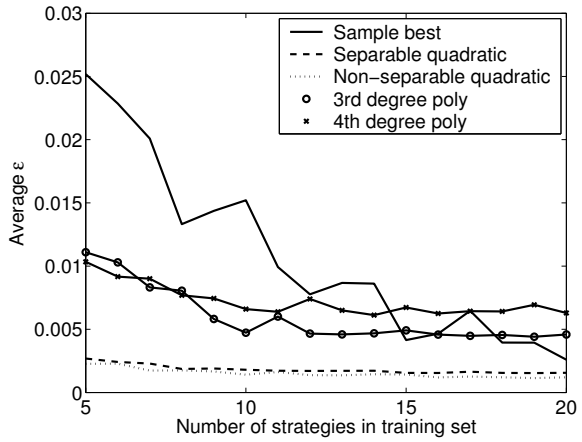

Figure 1: Epsilon versus number of training strategy points for different functional forms.

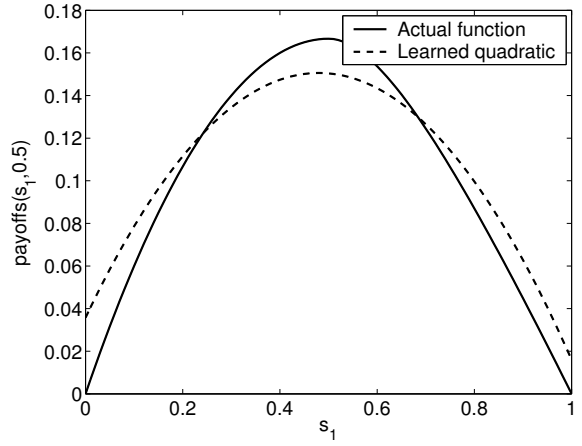

Figure 2: Learned and actual payoff function when the other agent plays 0.5. The learned function is the separable quadratic, for a particular sample with $N = 5$.

smooth and well approximated by a quadratic polynomial. The higher-degree polynomials apparently overfit the data, as indicated by their inferior learning performance displayed in this game.

The results of this game provide an optimistic view of how well regression might be expected to perform compared to discretization. This game is quite easy for learning since the underlying payoff function is well captured by our lower-degree model. Moreover, our experimental setup eliminated the issue of noisy payoff observations, by employing the actual expected payoffs for selected strategies.

Market-Based Scheduling Game

The second game we investigate presents a significantly more difficult learning challenge. It is a five-player symmetric game, with no analytic characterization, and no (theoretically) known solution. The game hinges on incomplete information, and training data is available only from a simulator that samples from the underlying distribution.

The game is based on a market-based scheduling scenario (Reeves *et al.*, to appear), where agents bid in simultaneous auctions for time-indexed resources necessary to perform their given jobs. Agents have private information about their job lengths, and values for completing their jobs by various deadlines. Note that the full space of strategies is quite complex: it is dependent on multi-dimensional private information about preferences as well as price histories for all the time slots. As in the FPSB example, we transform this policy space to the real interval by constraining strategies to a parametrized form. In particular, we start from a simple myopic policy—*straightforward bidding* (Milgrom, 2000), and modify it by a scalar parameter (called "sunk awareness", and denoted by k) that controls the agent's tendency to stick with slots that it is currently winning. Although the details and motivation for sunk awareness are inessential to the current study, we note that $k \in [0, 1]$, and that the optimal setting of k involves tradeoffs, generally dependent on other agents' behavior.

To investigate learning for this game, we collected data for all strategy profiles over the discrete set of values $k \in \{0, 0.05, \ldots, 1\}$. Accounting for symmetry, this represents 53,130 distinct strategy profiles. For evaluation purposes, we treat the sample averages for each discrete profile as the true expected payoffs on this grid.

Since in this senario we do not have access to the actual underlying payoff function, we must estimate ϵ based on a test set. If the profile of interest, say $\hat{s}$, is in the test set, we can find the empirical approximation of ϵ by computing the maximum benefit from deviation within the data:

$$\epsilon_{emp} = \max_{i \in I} \max_{s_i \in S_i} \left[u_i(s_i, \hat{s}_{-i}) - u_i(\hat{s}) \right],$$

where S_i is the strategy set of player i represented within the data set. If the game is symmetric, the maximum over the players can be dropped, and all the agent strategy sets are identical.

In general, the profile of interest may not appear in the training set. For that case we developed a method for estimating ϵ for pure symmetric approximate equilibria in symmetric games. Let us designate the pure symmetric equilibrium strategy of the approximated game by $\hat{s}$. We first determine the closest neighbors to $\hat{s}$ in the symmetric strategy set S represented within the data. Let these neighbors be denoted by s' and s''. We define a mixed strategy α with support $\{s', s''\}$ as the probability of playing s', which we compute based on the relative distance of $\hat{s}$ from its neighbors:

$$\alpha = 1 - \frac{|\hat{s} - s'|}{|s' - s''|}.$$

Note that symmetry allows a more compact representation of a payoff function if agents other than i have only a choice of two strategies. Thus, we define $U(s_i, j)$ as the payoff to a (symmetric) player for playing strategy $s_i \in S$ when j other agents play strategy s'. If $m - 1$ agents each independently choose whether to play s' with probability α, then the probability that exactly j will choose s' is given by

$$\Pr(\alpha, j) = \binom{m-1}{j} \alpha^j (1 - \alpha)^{m-1-j}.$$

We can thus approximate ϵ of the mixed strategy α by

$$\epsilon_{emp} = \max_{s_i \in S} \sum_{j=0}^{m-1} \Pr(\alpha, j) \times \\ \times \ (U(s_i, j) - \alpha U(s', j) - (1-\alpha)U(s'', j)) \, .$$

The previous empirical study of this game by Reeves *et al.* (to appear) estimated the payoff function over a discrete grid of profiles assembled from the strategies $\{0.8, 0.85, 0.9, 0.95, 1\}$. We therefore generated a training set based on the data for these strategies, regressed to the quadratic forms, and calculated ϵ values with respect to the entire data set. From the results presented in Table 1, we see that the Nash equilibria for the learned functions are quite close to that reported by Reeves *et al.* (to appear), but with ϵ values quite a bit lower. (Since 0.876 is not a grid point, we determined its ϵ post hoc, by running further profile simulations with all agents playing 0.876, and where one agent deviates to any of the strategies in $\{0, 0.05, \ldots, 1\}$.)

Method	Equilibrium s_i	ϵ
Separable quadratic	0.876	0.0027
Non-separable quadratic	0.876	0.0027
Reeves *et al.* (to appear)	0.85	0.0262

Table 1: Values of ϵ for the symmetric pure-strategy equilibria of games defined by different payoff function approximation methods in the game with exponential preferences. The quadratic models were trained on profiles confined to strategies in {0.8,0.85,0.9,0.95,1}.

In a more comprehensive trial, we collected additional samples per profile, and ran our learning algorithms on 100 training sets, each uniformly randomly selected from the discrete grid $\{0, 0.05, \ldots, 1\}$. Each training set included between five and fifteen of the twenty-one agent strategies on the grid. We compared results from regression to the method which simply selects from the training set the symmetric pure profile with the smallest value of ϵ. We refer to this method as "sample best". Finally, our "global best" is the lowest ϵ value for a symmetric pure profile in the entire data set of twenty-one agent strategies and corresponding payoffs.

From Figure 3 we see that regression to a separable quadratic produces a considerably better approximate equilibrium when the size of the training set is relatively small—specifically, when we sample payoffs for fewer than twelve agent strategies. Figure 4 shows that the non-separable quadratic performs similarly. The results appear relatively insensitive to the degree of aggregation applied to the representation of other agents' strategies.

Figures 5 and 6 illustrate the two error measures, ϵ and mean squared error. While the ϵ metric is more meaningful for our purposes, we are still interested in the quality of the functional fit in general. One interesting and somewhat surprising result that we see in Figure 5 is that separable quadratic actually outperforms the non-separable variety. If we recall that the non-separable quadratic, while including

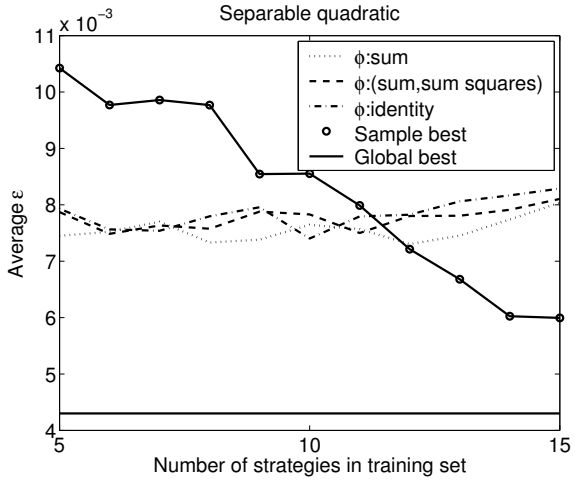

Figure 3: Effectiveness of learning a separable quadratic model with different forms of $\phi(s_{-i})$.

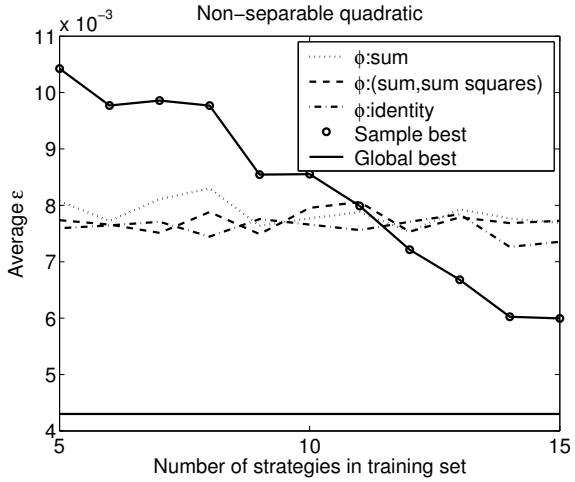

Figure 4: Effectiveness of learning a non-separable quadratic model with different forms of $\phi(s_{-i})$.

the strategy interaction term, omits several other terms from the general quadratic model, this counter-intuitive result is easily explainable: the omitted terms appear to have more relevance to the data set than the strategy interaction term.

Conclusion

While there has been much work in game theory attempting to solve particular games defined by some payoff functions, little attention has been given to approximating such functions from data. This work addresses the question of payoff function approximation by introducing a regression learning technique and applying it to representative games of interest. Our results in both the FPSB and market-based scheduling games suggest that when data is sparse, our methods can provide better approximations of the underlying game—at least in terms of ϵ-Nash equilibria—than discrete approxi-

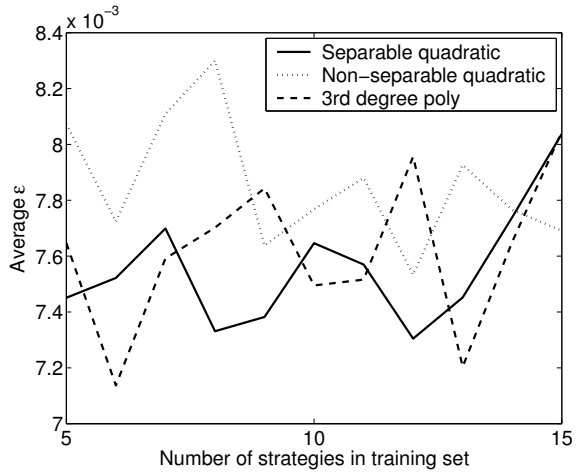

Figure 5: Relative effectiveness of different learning methods in terms of average ϵ.

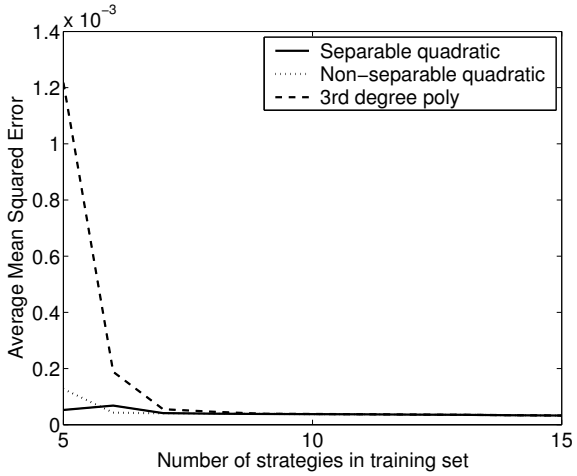

Figure 6: Relative effectiveness of different learning methods in terms of average mean squared error.

mations using the same data set.

Regression or other generalization methods offer the potential to extend game-theoretic analysis to strategy spaces (even infinite sets) beyond directly available experience. By selecting target functions that support tractable equilibrium calculations, we render such analysis analytically convenient. By adopting functional forms that capture known structure of the payoff function (e.g., symmetry), we facilitate learnability. This study provides initial evidence that we can sometimes find models serving all these criteria.

To date, we have explored only a few function-learning methods in only a few example games. In ongoing work, we are investigating alternative regression techniques applied to the market-based scheduling game, as well as other challenging domains.

References

Brafman, R. I., and Tennenholtz, M. 2003a. Efficient learning equilibrium. In Becker, S.; Thrun, S.; and Obermayer, K., eds., *Advances in Neural Information Processing Systems 15*. MIT Press. 1603–1610.

Brafman, R. I., and Tennenholtz, M. 2003b. Learning to coordinate efficiently: A model-based approach. *Journal of Artificial Intelligence Research* 19:11–23.

Cheng, S.-F.; Reeves, D. M.; Vorobeychik, Y.; and Wellman, M. P. 2004. Notes on equilibria in symmetric games. In *AAMAS-04 Workshop on Game-Theoretic and Decision-Theoretic Agents*.

Kreps, D. M. 1990. *Game Theory and Economic Modelling*. Oxford University Press.

Krishna, V. 2002. *Auction Theory*. Academic Press.

Mas-Colell, A.; Whinston, M. D.; and Green, J. R. 1995. *Microeconomic Theory*. New York: Oxford University Press.

Milgrom, P. 2000. Putting auction theory to work: The simultaneous ascending auction. *Journal of Political Economy* 108:245–272.

Nash, J. 1951. Non-cooperative games. *Annals of Mathematics* 54:286–295.

Reeves, D. M.; Wellman, M. P.; MacKie-Mason, J. K.; and Osepayshvili, A. to appear. Exploring bidding strategies for market-based scheduling. *Decision Support Systems*.

Shoham, Y.; Powers, R.; and Grenager, T. 2003. Multi-agent reinforcement learning: A critical survey. Technical report, Stanford University.

Vickrey, W. 1961. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance* 16:8–37.