

Learning in one-shot strategic form games*

Alon Altman, Avivit Bercovici-Boden, and Moshe Tennenholtz

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

alon@vipe.technion.ac.il, avivitb@tx.technion.ac.il,
moshet@ie.technion.ac.il

Abstract. We propose a machine learning approach to action prediction in *one-shot* games. In contrast to the huge literature on learning in games where an agent’s model is deduced from its previous actions in a multi-stage game, we propose the idea of inferring correlations between agents’ actions in different one-shot games in order to predict an agent’s action in a game which she did not play yet. We define the approach and show, using real data obtained in experiments with human subjects, the feasibility of this approach. Furthermore, we demonstrate that this method can be used to increase payoffs of an adequately informed agent. This is, to the best of our knowledge, the first proposed and tested approach for learning in one-shot games, which is the most basic form of multi-agent interaction.

1 Introduction

Learning in the context of multi-agent interaction has attracted the attention of researchers in cognitive psychology, experimental economics, machine learning, artificial intelligence, and related fields for quite some time (see e.g. [1,2,3,4,5,6,7,8]). Most of this work uses repeated games and stochastic games as models of such interactions. Roughly speaking, one can distinguish between two types of multi-agent learning: reinforcement learning and model-based/belief-based learning. In reinforcement learning an agent adapts its behavior based on feedback obtained in previous interactions, while model-based/belief-based learning is mostly concerned with inferring an “opponent model” from past observations. The aim of this paper is to tackle the problem of learning/predicting opponent actions, and thus our study can be viewed as part of model-based/belief-based learning. However, our study introduces the first general machine learning approach for tackling the prediction of an agent’s action in general **one-shot** games, i.e. without having access to any information on how this agent has played this particular game in the past. This is in difference to the extensive literature on learning in repeated and extensive form games [9,10]. Although some work in experimental economics has been devoted to modeling agent behavior in one-shot games [11,12], that work did not present an effective general learning approach to action prediction/selection in such games. Recent work in AI [13] deals with the idea of learning population statistics in order to predict agent behavior in a specific game, namely the

* We thank Ido Erev for helpful comments and illuminating discussions regarding this paper.

Nash Bargaining game. In difference to these lines of research we are after a general machine learning approach for action prediction in general one-shot games.

In order to tackle the above challenge we suggest to consider agents' interactions in a population of one-shot games. Given information on how different agents (including our "opponent") played in different games, our aim is to learn connections and correlations between agents' behaviors in different games, that will allow us to predict the opponent's behavior in one game (which he did not play yet) based on his or her action in another game. In a sense, what we offer is to try and learn association rules among games, ones that will allow to improve upon prediction of an agent's action in a given game. Our main contribution is by suggesting this approach, and proving its (perhaps surprising) feasibility using real data obtained in experiments involving human subjects.

The economics literature refers to population learning[14]. In population learning we aim at predicting an agent's action based on statistics on how the population played a game in the past. Population learning is considered a good predictor for an agent's behavior in a game, and therefore will be used as a benchmark. What we suggest is a mixed approach: given information about how the population played different one-shot games we aim at predicting an agent's action in a game she did not play based on the population statistics and rules which we try to infer about correlations between games. Our benchmark for success will be highly competitive, and coincide with the population statistics mentioned above: We aim at showing cases where association rules between games can be learned and lead to better predictions than the ones obtained by only using the statistics on how the population played in the game.

As we mentioned, to the best of our knowledge, there was no attempt to provide a machine learning approach for the general task of predicting agents' actions in one-shot games. Here we exploit a simple machine learning technique in order to offer an approach for addressing this issue. In a sense, our work is related to the relatively recent work on case-based decision making[15]. In case-based decision making, the idea of case-based reasoning, which is a classical topic in AI, is exploited to introduce an alternative approach to decision making in strategic contexts. This approach is based on similarity measure between different decision problems. Our work suggests to learn such similarity/association between decision problems, in order to improve upon opponent prediction in games, ultimately improving payoffs. We show that this machine learning approach is highly useful for that context. We have also experimented with other machine learning techniques (such as ID3 and KNN), but they were found to be less efficient for our objectives.

The paper is self contained in introducing the approach, the experimental setup, and the way in which real data has been obtained. Unfortunately, relying on simulations and on reasoning under abstract assumptions is known to be somewhat problematic in game-theoretic settings, and therefore we appealed to experimental design in this work. Our experiments use two standard sources from which the games we have experimented with have been selected: the first experiment uses games found in [16] , a standard game-theory book, and the second experiment uses games from the GAMUT site[17], a standard source for games used by CS researchers.

This paper is organized as follows: In section 2 we present the experimental setting. Sections 3 and 4 introduce and discuss the use of a simple machine learning technique

(learning simple association rules) for successful action prediction in one-shot games. Section 5 extends upon the above technique (by considering aggregation of association rules), and shows the applicability of the machine learning approach to improving strategy selection in one-shot games.

2 The Experimental Setting

In this section we will describe the two experiments we performed. In the first experiment we have tested the idea and observed its (surprising) success, against a highly competitive benchmark. The second experiment has been carried out to further validate our observations, and test that our findings are robust for a variety of cases.

2.1 The first experiment

For the purpose of the first experiment, 16 strategic form games with payoffs between 0 and 10 were constructed, based on [16]. 11 of the games are symmetric, and the remaining 5 asymmetric games were duplicated with the roles of the players interchanged. These games were printed on forms in random order, with a basic explanation on how the form is graded, which consisted of a brief explanation of the concept of a strategic form game. All games tested can be found in Figure 1.

Twenty-six students with basic knowledge of game theory were given these forms. The students were told to select a strategy for the row player in each of the games, and were told that they will be given bonus points to their exam grade based on the result of running one of the games against some other student. It was emphasized that there is no “correct” answer, and the students were encouraged to use any idea they may have in order to try and maximize their revenue.

Additional data were gathered by asking 36 friends and faculty members with a knowledge of game theory to participate in the experiment. These participants were similarly informed.

2.2 The second experiment

This experiment was performed on two subject populations. The main population (55 subjects) were undergraduate students who got paid according to their performance. Additional data were gathered from 38 faculty members and graduate students with a knowledge of game theory that were asked to play as if they were being similarly paid.

The subjects were initially allocated 150 points, each point worth about 2.5 cents (US). In the experiment there were three sections. In each of the sections the subjects could gain or lose points. The three sections were as follows:

- **Auctions section:** In this section we study the participants’ behavior in first, second, and third price sealed bid auctions. In a k -price auction all agents submit bids simultaneously, the highest bidder wins, and pays the value of the k th highest bid (see [18]). Here, the subjects were presented with three identical tables, one table for each auction, See Table 1 for an example. For each subject, the valuations in the

First Experiment

		Opponent		
Your 1.	9\9	0\10		
Action 2.	10\0	5\5		

		Opponent		
Your 1.	6\6	4\9		
Action 2.	9\4	5\5		

		Opponent		
Your 1.	9\9	0\10		
Action 2.	10\0	1\1		

		Opponent		
Your 1.	10\6	5\0	6\7	
Action 2.	8\8	6\9	7\7	

		Opponent		
Your 1.	10\10	0\3		
Action 2.	3\0	2\2		

		Opponent		
Your 1.	10\10	0\9		
Action 2.	9\0	8\8		

		Opponent		
Your 1.	10\10	0\0	0\0	
Action 2.	0\0	10\10	0\0	
3.	0\0	0\0	9\9	

		Opponent		
Your 1.	10\10	0\0		
Action 2.	0\0	10\10		

		Opponent		
Your 1.	0\0	7\6		
Action 2.	6\7	0\0		

		Opponent		
Your 1.	0\0	7\6		
Action 2.	6\7	1\1		

		Opponent		
Your 1.	6\8	9\6		
Action 2.	5\7	8\9		

		Opponent		
Your 1.	4\8	7\6		
Action 2.	5\7	8\9		

		Opponent		
Your 1.	7\7	7\7		
Action 2.	8\6	5\5		

		Opponent		
Your 1.	0\0	3\4	6\0	
Action 2.	4\3	0\0	0\0	
3.	0\6	0\0	5\5	

		Opponent		
Your 1.	7\7	10\8		
Action 2.	8\10	7\7		

		Opponent		
Your 1.	10\2	0\0	0\0	
Action 2.	0\0	1\4	0\0	
3.	0\0	0\0	3\3	

Second Experiment

		Opponent		
Your 1.	100\100	0\200	0\150	
Action 2.	200\0	150\150	0\100	
3.	150\0	100\0	50\50	

		Opponent		
Your 1.	30\170	30\170		
Action 2.	50\100	0\200		

		Opponent		
Your 1.	70\120	150\30	80\90	
Action 2.	90\20	200\30	90\90	
3.	150\120	90\150	0\120	

		Opponent		
Your 1.	20\0	70\70	100\100	
Action 2.	70\70	0\20	70\70	
3.	100\100	70\70	20\20	

		Opponent		
Your 1.	40\40	0\ - 70		
Action 2.	-70\0	100\100		

		Opponent		
Your 1.	20\20	100\50		
Action 2.	50\100	40\40		

		Opponent		
Your 1.	70\70	-10\ - 10		
Action 2.	-100\ - 10	100\100		

		Opponent		
Your 1.	0\0	100\50	100\50	
Action 2.	50\100	0\0	100\50	
3.	50\100	50\100	0\0	

		Opponent		
Your 1.	0\ - 50	100\ - 100		
Action 2.	50\50	50\70		

Fig. 1. List of games used in both experiments

Table 1. Example of an auction table

The product's value for you	Your bid
239	
419	
668	
947	

Table 2. Discretization of Auctions

Range	1 st Price	2 nd Price	3 rd Price
1	$\alpha < 0.91$	$\alpha < 1$	$\alpha < 1$
2	$\alpha \geq 0.91$	$\alpha = 1$	$\alpha = 1$
3		$\alpha > 1$	$1 < \alpha < 1.125$
4			$\alpha \geq 1.125$

tables were uniformly distributed in four ranges between 1 and 1000. The subjects were told the rules of the various auctions, and were asked to provide a bid for each auction and for each of the valuations in the table. The subjects were told they will be partitioned into groups of 10 subjects each, and that for each of the auctions we will choose randomly one of their values and perform the auction. Detailed discussions with the subjects were carried out in order to verify they understand the auction rules and parameters.

- **The Centipede Game section:** This section has been devoted to the Centipede game following the description in [17]. The subjects were partitioned randomly into pairs, where one of each pair is randomly selected to play the role of Player 1 (who moves first), and the other plays the role of Player 2. The game is described as follows. There are many piles of points on the table, where each pile contains 60 points. The players alternate in their moves. At each move the player needs to decide whether to take one pile or two piles of points from the table. If a player decides to take two piles of points then the game ends. Otherwise, the game is over after 6 turns (3 for each subject). The subjects were given a detailed explanation of the game, and were asked to mark their selected moves (i.e. take one pile or take two piles) if they were playing the odd turns of the game. Similarly, given this game the participants were asked to mark their selected moves if they were playing the even turns.
- **Strategic form games section:** For the purpose of this section, 9 strategic form games with payoffs between -100 and 200 were constructed, based on games found in the Gamut web site. Four of the games are symmetric, and the remaining 5 asymmetric games were duplicated with the roles of the players interchanged. These games were printed on forms in random order. The subjects were explained the notion of a strategic form game, and their understanding was verified by solving a sample game with an obvious dominant strategy. The subjects were told to select a strategy for the row player in each of the games, and were told that they will be paid based on the result of running one of the games against some other subject. Two control games with obvious dominant strategies were also mixed in with the other games, and were used to verify the subjects' understanding. All games tested can be found in Figure 1.

3 Learning Algorithm and Evaluation

The experiments on the data were conducted using leave-one-out cross validation. In each iteration the play of one game g by one player x was hidden, and was to be pre-

dicted by the remaining data. We used the prediction based on most frequent play in g by the other players as a baseline for comparison. Using play data of the remaining players, rules were learned that map a strategy choice in games $g' \neq g$ where the play is known to a strategy choice in g . The association rule with the highest confidence level which was satisfied was applied to select a strategy choice for x .

Our learning system is based on the identification of association rules with high confidence and support. Association rules specify that given that a certain set of strategies is played by a player in a specific set of games, then the player will play a specific strategy in the game we are trying to predict with a given probability. That probability is called the confidence level of the association rule. The support of an association rule is the percentage of the population who confirm the rule. That is, satisfy both the preconditions and postcondition of the rule.

Formally, let $N = \{1, \dots, n\}$ be the set of participants and let $G = \{1, \dots, m\}$ be the set of games. Let $S_g = \{s_g^1, \dots, s_g^{n_g}\}$ be the set of strategies for the row player in game g , and let $p_g^x \in S_g$ be the action selected by player x in game g . The baseline prediction for p_g^x is defined as

$$\hat{p}_g^x \in \operatorname{argmax}_{s \in S_g} |\{i | i \neq x, p_g^i = s\}|.$$

If the maximum is not unique, the prediction is chosen arbitrarily among the maxima.

Association rules can be written in the form $S \Rightarrow s_g^k$, where

$$S \subseteq \{s_{g'}^i | g' \in G \setminus \{g\}, s_g^i \in S_g\}$$

and $\forall s_{g_1}^i \neq s_{g_2}^j \in S : g_1 \neq g_2$. Such a rule means that a player who has played the strategies in S is likely to play the strategy s_g^k . In this paper, we only consider rules in which $|S| \leq 1$.

For each player x , association rules of the form $\{s_g^i\} \Rightarrow s_{g'}^j$ were learned. These rules mean that if a player has played strategy s_g^i in game g , she is likely to play strategy $s_{g'}^j$ in game g' . For each rule, the support level was calculated as follows:

$$S(\{s_g^i\} \Rightarrow s_{g'}^j) = \frac{|\{k | k \neq x, p_g^k = s_g^i, p_{g'}^k = s_{g'}^j\}|}{n}.$$

Rules with $S(r) < 0.2$ were discarded. The support threshold of 0.2 has been chosen to fit the size of data considered in such experiments (similar results are obtained with slightly different support). The confidence level was calculated as follows:

$$C(\{s_g^i\} \Rightarrow s_{g'}^j) = \frac{|\{k | k \neq x, p_g^k = s_g^i, p_{g'}^k = s_{g'}^j\}|}{|\{k | k \neq x, p_g^k = s_g^i\}| + 1}.$$

One was added to the denominator to account for the uncertainty on the action selection of the agent being predicted.

Furthermore, for every game g , a baseline rule $\emptyset \Rightarrow s_g^i$ was added to the rule set with support 1 and confidence

$$C(\emptyset \Rightarrow s_g^i) = \frac{|\{k | k \neq x, p_g^k = s_g^i\}|}{n}.$$

Let R_x be the set of rules generated for player x .

For every game g , and for every player x , the association rules are tested in declining order of confidence. Thus, the predicted strategy \tilde{p}_g^x is the strategy s for which the following maximum is obtained:

$$\max\{C(S \Rightarrow s) \mid (S \Rightarrow s) \in R_x \wedge \forall s_{g'}^i \in S : p_{g'}^x = s_{g'}^i\}$$

The auctions were evaluated as follows: For each auction we assumed a linear correlation between valuations and bids and calculated the slope α of the linear regression line for the 4 observations. We then split α into ranges as depicted in Table 2. These range designations were interpreted as strategies in the auction games.

4 Results

Our results show that the use of association rules has significantly improved our prediction of the strategies selected by the agents. This has been found in both experiments. We now illustrate some of the most effective association rules which have been learned, which have led to most significant improvements in prediction.

4.1 First experiment

It is interesting to note that the improvements of results have been independently obtained for both sub-populations, as well as for the entire group.

The results of the experiment can be seen in Fig. 2. The average increase in prediction was 1% with a standard deviation of 0.05. Specifically, we have found a significant improvement in prediction in games 7, 20 and 21. The declines in prediction are an artifact of the leave-one-out cross validation technique, where removing one agent changes the learned association rules for the worst. The low average increase is expected, and less relevant, as we expect to improve prediction in only a few of the games. The improvements in prediction are a result of learning and application of several meaningful and nontrivial association rules discussed below¹:

$$- s_3^1 \Rightarrow s_7^1$$

9\9	0\10	⇒	10\10	0\9
10\0	1\1		9\0	8\8

This rule is responsible for a 15% increase in prediction accuracy, up to 75% accuracy in game 7.

This rule means that players who played *cooperate* in the extreme version of the prisoner's dilemma (game 3), tended to play the *trust* strategy in the extreme version of the trust game (game 7).

We see that players who hope for the 9\9 result in the prisoner's dilemma also try for the risky 10\10 in the trust game. We capture a tendency to improve the total welfare of both players, even over the benefit of playing a dominant strategy and the risk of playing the trust strategy.

¹ Selected strategies are shaded.

This rule also captures players who strongly believe in symmetry, and try to grab the best payoff assuming the other player does the same.

The inverse rule was not learned, as many more players have played *trust* in the trust game than *cooperate* in prisoner’s dilemma. This is predictable, due to the fact that *defect* is a dominant strategy in prisoner’s dilemma, while no such dominant strategy exists in the trust game.

– $s_{20}^1 \Rightarrow s_{21}^2$

1\0	2\0	0\0	0\0	⇒	2\1	0\0	0\0
0\0	1\4	0\0			0\0	4\1	0\0
0\0	0\0	3\3			0\0	0\0	3\3

This rule is responsible for a 10% increase in prediction accuracy, up to 55% accuracy in the inverse of game 20.

This rule means that in game 20, players that select the row with the highest payoff for themselves when playing rows, tend to do the same when playing columns.

This rule captures players who play a “bully” strategy. That is, try to maximize their own payoffs, under the assumption that the other player will play their best response.

In game 20, this bully strategy is the most common strategy, played by 57% of the participants. In its inverse only 31% of the population do so, many of them have “bullied” in game 20 as well.

In this case, the inverse rule is trivial, as s_{20}^1 is already the most common strategy in game 20.

– $s_{21}^3 \Rightarrow s_{20}^3$

2\1	0\0	0\0		⇒	1\0	2\0	0\0
0\0	4\1	0\0			0\0	1\4	0\0
0\0	0\0	3\3			0\0	0\0	3\3

This rule is responsible for a 15% increase in prediction accuracy, up to 70% accuracy in game 20.

This rule captures the fact that players who try to reach a middle ground in game 21 tend to do the same in the original game 20.

This rule captures players who believe in symmetry in the sense that both players should get the same payoffs, and consistency in the sense that strategy selection in the rows and columns should be in some kind of equilibrium when possible.

Again, the inverse rule is trivial, since s_{21}^3 is the most common strategy in the inverse of game 20.

4.2 Second experiment

The results of running the second experiment on the full corpus of 93 players are in Figure 2. For every game (see Figure 1) there are two bars representing the portion of correct predictions for the baseline(grey) and association rule(black). The results for each of the sub-populations were similar.

We see a significant improvement in prediction for the Centipede, game 10 and the Third price auction. In a few other games, there were less significant differences between the quality of predictions when using the learning rule compared to the baseline predictions. In all games, excluding one, the learning rule predictions were equal or

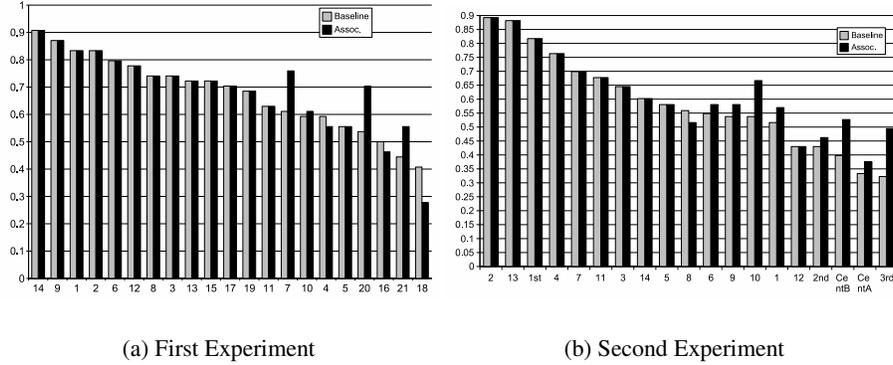


Fig. 2. Prediction results for the first and second experiments

superior to the baseline. In general the average increase in prediction was 3% with standard deviation of 0.07, which is expected as in the first experiment. The significant improvement in prediction is a result of learning and application of several meaningful and nontrivial association rules as illustrated below.

The following association rules learned by our algorithm resulted in improved predictions:

- $s_{CentipedeA}^1 \Rightarrow s_{CentipedeB}^1$
 This rule means that players who took 2 piles of 60 points in the first turn of the Centipede game, when playing the odd turns, tended to take 2 piles in their first turn when they were playing the even turns. We expose here a tendency of the players to be self coherent and take the money in the first opportunity they can, rather than take the chance of highly improving both players' payoffs.
- $s_{CentipedeA}^4 \Rightarrow s_{CentipedeB}^3$
 This rule means that players who took only one pile of 60 points in all of their turns when playing the odd turns, tended to take two piles of 60 points only on the 6th turn when playing the even turns. We see that players who hope that the other player will "cooperate" with them in order to accumulate more points, take two piles of points only in their last (6th) turn. We capture here a tendency to improve the total welfare of both players, even over the benefit of playing a dominant strategy and the risk of trusting the other player.

- $s_4^3 \Rightarrow s_{10}^2$

70\120	150\30	80\90
90\20	200\30	90\90
150\120	90\150	0\120

 \Rightarrow

70\70	-10\10
-100\10	100\100

This rule means that in game 4, players that select the row with the highest risk, tend to do the same when playing game 10.

This rule captures players who are risk seeking, and try to increase social welfare by selecting rows which may lead to significant loss (0 in the former case, and -100 in the latter), but contain the best sum for both players.

$$- s_8^2 \Rightarrow s_{10}^2$$

40\40	0\70	⇒	70\70	-10\10
-70\0	100\100		-100\10	100\100

This rule means that in game 8, players that select the row with the highest risk, tend to do the same when playing game 10.

This rule further captures players who are risk seeking .

$$- s_{Second}^1 \Rightarrow s_{Third}^1$$

The meaning of the above rule is that the subjects who underbid in the second price auction also underbid in the third price auction. This rule was highly useful: from among 39 subjects that underbid in the second price auction, 26 underbid in the third price auction.

This rule captures players who play “safe”, and did not notice the dominant strategy, and were fearful of winning the auction but not incurring profit (or incurring loss).

5 Improving Strategy Selection

We have seen that learning of association rules may improve prediction of strategy selection by players in one-shot strategic games. The question arises whether or not this prediction can be used to improve a player’s expected payoff when playing against a player whose actions in other games are known. That is, we wish to see whether by learning association rules an agent may obtain higher payoffs in games. Our results indicate that the answer is yes.

In order to do so, we consider an extension of the simple technique discussed in previous sections. Indeed, in order to select an appropriate action one may wish to have a rather accurate estimation of the probability distribution of his opponent strategies. In order to do so we need to exploit the information hidden in the set of association rules.

We use the population data and the 10 association rules with the highest confidence, combined using linear regression to estimate the probability distribution over our opponent’s strategies. This enables us to calculate the best response against this particular opponent².

The above method differs from choosing a best response to the predicted (pure) strategy by the fact that we compensate for the scale of our lost payoffs due to our mistakes. For example, in game 4 (in the first experiment) our data suggests that the best response to the mixed strategy defined by the population is for the column player to select the leftmost column, even though this strategy is not a best response to any pure strategy of the row player.

Assume we are trying to respond a player p ’s behavior in game g . The probability distribution over this player’s strategy is estimated as follows: For every player x and for every strategy s_g^i in g we compute the following three parameters on the training set with x removed:

1. F_g^i – the frequency of the strategy s_g^i in the population without x :

$$F_g^i = \frac{|\{k | k \neq x, p_g^k = s_g^i\}|}{n}$$

² A similar technique (of using the linear regression to combine a new model and experimental data, to better predict probabilities) has also been considered in economic literature, see [19].

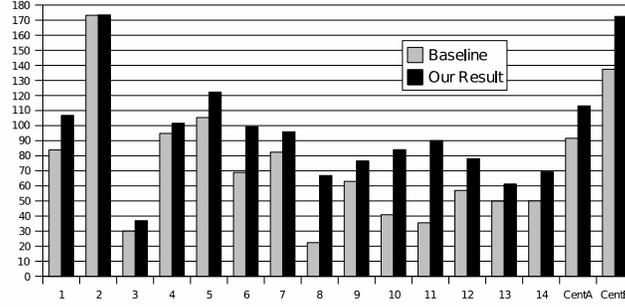


Fig. 3. Results of best response from the second experiment

2. C_g^i – the confidence of the association rule for g with the greatest confidence among the rules whose pre-condition is satisfied by player x if this rule predicts s_g^i , and 0 otherwise.
3. A_g^i – the average confidence among the 10 association rules for g with the greatest confidence, where a rule that does not predict s_g^i is assumed to have confidence 0. That is, if the 10 association rules for g with the greatest confidence predict s_1, s_2, \dots, s_{10} with confidence levels c_1, c_2, \dots, c_{10} respectively, then the average confidence will be

$$A_g^i = \frac{1}{10} \sum_{t=1}^{10} \begin{cases} c_t & s_t = s_g^i \\ 0 & \text{Otherwise} \end{cases}$$

Let R_g^i be 1 if player x played strategy s_g^i in game g , and 0 otherwise. For every strategy in g we take these three parameters and develop a linear regression line of the form $B_0 + B_1 * F_g^i + B_2 * C_g^i + B_3 * A_g^i$, where the B_i are selected to minimize

$$(B_0 + B_1 * F_g^i + B_2 * C_g^i + B_3 * A_g^i - R_g^i)^2$$

over all players. Given these B_i values, we calculate these parameters for the game g and player p over the entire training set, and calculate the probability for each move. Then we calculate a best response to the mixed strategy defined by this probability distribution.

We have tested this best response algorithm using leave-one-out cross validation on the data of the second experiment. The results are shown in Fig. 3. The average increase in payoff is 48%. We see a significant increase in payoffs across almost all the games, compared to the very competitive baseline of best-response to the empirical distribution of the players' actions.

6 Concluding Remarks

We have introduced a novel approach to machine learning in one-shot games by utilizing information about actions of the player in other games. Our experimental results are encouraging and confirm that this approach is in fact viable.

We have seen evidence that learning of association rules can improve by some level both prediction and play in one-shot strategic form games.

When larger data-sets are available, one may try to re-visit learning algorithms such as ID3, k -nearest neighbor, Bayesian inference, and Support Vector Machines, which require larger data sets in order to produce meaningful results. As we mentioned, with the current available data, these algorithms have been found less efficient than the technique previously discussed.

In conclusion, we have opened room for work in ML on action prediction and strategy selection in one-shot games, exploiting correlations between games which need not be explicitly specified.

References

1. Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: A survey. *Journal of AI Research* **4** (1996) 237–285
2. Erev, I., Roth, A.: Predicting how people play games: Reinforcement learning in games with unique strategy equilibrium. *American Economic Review* **88** (1998) 848–881
3. Claus, C., Boutilier, C.: The dynamics of reinforcement learning in cooperative multi-agent systems. In: *Proc. Workshop on Multi-Agent Learning*. (1997) 602–608
4. Fudenberg, D., Levine, D.: *The theory of learning in games*. MIT Press (1998)
5. Littman, M.L.: Markov games as a framework for multi-agent reinforcement learning. In: *Proc. 11th ICML*. (1994) 157–163
6. Hu, J., Wellman, M.: Multi-agent reinforcement learning: Theoretical framework and an algorithms. In: *Proc. 15th ICML*. (1998)
7. Brafman, R.I., Tennenholtz, M.: R-max – a general polynomial time algorithm for near-optimal reinforcement learning. In: *IJCAI'01*. (2001)
8. Carmel, D., Markovitch, S.: Exploration strategies for model-based learning in multiagent systems. *Autonomous Agents and Multi-agent Systems* **2**(2) (1999) 141–172
9. Kagel, J., Roth, A.: *The Handbook of Experimental Economics*. Princeton University Press (1995)
10. Roth, A., Erev, I.: Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* **8** (1995) 164–212
11. Camerer, C.F., Ho, T.H., Chong, J.K.: A cognitive hierarchy model of games. *The Quarterly Journal of Economics* **119**(3) (2004) 861–898
12. Costa-Gomes, M., Crawford, V.P., Broseta, B.: Cognition and behavior in normal-form games: An experimental study. *Econometrica* **69**(5) (2001) 1193–1235
13. Gal, Y., Pfeffer, A., Marzo, F., Grosz, B.J.: Learning social preferences in games. In: *Proc. of AAAI-04*. (2004) 226–231
14. Stahl, D.O.: Population rule learning in symmetric normal-form games: theory and evidence. *Journal of Economic Behavior and Organization* **1304** (2001) 1–14
15. Gilboa, I., Schmeidler, D.: Case-based decision theory. *Quarterly Journal of Economics* **110** (1995) 605–639
16. Fudenberg, D., Tirole, J.: *Game Theory*. MIT Press (1991)
17. Nudelman, E., Wortman, J., Shoham, Y., Leyton-Brown, K.: (Run the gamut: A comprehensive approach to evaluating game-theoretic algorithms) Available at <http://gamut.stanford.edu>.
18. Wolfstetter, E.: Auctions: An introduction. *Journal of Economic Surveys* **10**(4) (1996) 367–420
19. Erev, I., Roth, A., Slonim, R., Barron, G.: Learning and equilibrium as useful approximations: accuracy of prediction on randomly selected constant sum games. (2006)