GAMES AND ECONOMIC BEHAVIOR 9, 1–7 (1995)

Unlearning by Not Doing: Repeated Games with Vanishing Actions

REINOUD JOOSTEN, HANS PETERS, AND FRANK THUIJSMAN

Departments of Mathematics and Economics, MERIT, University of Limburg, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Received October 16, 1991

We examine two-person zero-sum repeated games in which the players' action choices are restricted in the following way. Let $r_1, r_2 \in \mathbb{N}$, where $\mathbb{N}$ also represents the set of stages of the game. If, at any stage τ, player $k \in \{1, 2\}$ did not select action i at any of the preceding r_k stages, then action i will vanish from his set of actions and will no longer be available in the remaining play. For several (r_1, r_2)-cases we show the existence of optimal strategies for limiting average optimal play. *Journal of Economic Literature* Classification Numbers: C72, C73. © 1995 Academic Press, Inc.

1. INTRODUCTION

Learning-by-doing is widely recognized as an important phenomenon in technical change (cf. Arrow, 1962). Unlearning-by-not-doing could be defined as the loss of relevant knowledge caused by ceasing to perform certain activities. The following question arises: Which actions (e.g., skills, activities, technological possibilities) should be maintained by doing, and which actions will consequently be dropped, or unlearned, by not-doing? In a one-person decision situation the obvious answer to this question is to keep the most profitable actions. In a multi-person setting the answer is not so obvious. In this first paper on the subject, we take a zero-sum infinitely repeated game model to study this question. Normally, in two-person repeated games with complete information the players face the same payoff matrix at all stages and they can select from the same set of actions at any of those stages. In this new setup, however, the actions are to be used frequently in order for them to remain available. In our model unlearning happens in a rather abrupt way: If a player has not

0899-8256/95 $6.00

performed some particular action during a given fixed period of time, then this action instantaneously disappears from his action set.

This paper is a first attempt to model aspects of unlearning and does not provide an exhaustive analysis of the matter. The organization is as follows. In the next section we give our formal model. In Section 3 we examine games where (at least) one of the players already loses actions after not selecting them for two consecutive stages. In Section 4 both players unlearn actions in three stages. Section 5 has concluding remarks on solving games with longer unlearning periods.

2. The Model

A repeated game with vanishing actions is given by an $(m \times n)$-matrix of reals $A = [a_{ij}]_{i=1}^{m}, {}_{j=1}^{n}$ and by two natural numbers r_1, r_2. The game is played as follows. Initially players 1 and 2 have pure action sets $\{1, 2, \ldots, m\}$ and $\{1, 2, \ldots, n\}$, respectively. Players are allowed to randomize over their pure actions. If action i of player k has not been realized during r_k consecutive stages, then action i is removed from player k's action set. Thus play continues on a submatrix of A. Whenever entry (i, j) is being selected, player 2 has to pay the stage payoff a_{ij} to player 1. Play continues forever and players evaluate the infinite stream of stage payoffs R_t as a limiting average reward $\liminf_{T\to\infty} (1/T) \sum_{\tau=1}^{T} R_\tau$. Player 1 wishes to maximize his expected average reward, while player 2 is minimizing the same.

Such a game is a special type of stochastic game with finite state and action spaces. It is well known that these games have a value, both for discounted rewards (cf. Shapley, 1953) and for limiting average rewards (cf. Mertens and Neyman, 1981). To compute discounted value and optimal strategies, several algorithms are available (cf. Raghavan and Filar, 1991), so in the case of discounting there is not really a problem in solving a repeated game with vanishing actions. On the other hand, little is known about how to solve limiting average reward stochastic games (except for some very specially structured subclasses). Generally there are no limiting average optimal strategies and only history dependent limiting average ε-optimal strategies ($\varepsilon > 0$) (cf. Blackwell and Ferguson, 1968). Repeated games with vanishing actions do not belong to any of the subclasses of limiting average reward stochastic games for which solution methods are known. We therefore focus on such rewards for our model.

We end this section with a few obvious results on (r_1, r_2)-restricted games. If the matrix game A has a saddlepoint at (i^*, j^*), then the value of the matrix game A, denoted by val(A), as well as the value v_{r_1, r_2} of the (r_1, r_2)-restricted game $A_{r_1 r_2}$ is equal to $a_{i^* j^*}$. If $r_1 = r_2 = 1$, the actions realized at the initial stage are to be used at all stages. So the restricted game is equivalent to a (one-shot) matrix game and $v_{1,1} = \text{val}(A)$. If $r_1 = 1$ and $r_2 \geq 2$, then at stage 2 player 1 has only one action left, while player 2 has all actions left. Hence $v_{1,r_2} = \underline{v} = \max_i \min_j a_{ij}$ (and similarly $v_{r_1,1} = \bar{v} = \min_j \max_i a_{ij}$ for $r_1 \geq 2$). Asymptotically we have $\lim_{r_1,r_2\to\infty} v_{r_1,r_2} = \text{val}(A)$ for any $(m \times n)$-matrix A because, as r_1 increases, player 1 can use his matrix game optimal mixed action for a fraction of time growing to 1, now and then selecting pure actions that are about to vanish.

3. (2, r)-Restricted Games

Theorem 3.1. *Let A be an $(m \times n)$-matrix game and let $r \geq 3$. The value of the* (2, r)*-restricted game $A_{2,r}$ equals $v_{2,r} = \underline{v}$.*

Proof. Obviously, player 1 can achieve at least $\underline{v}$ in the (2, r)-restricted game. Player 2 can guarantee $\underline{v}$ by playing as follows without loss of generality: At stage 1 let player 2 choose action 1 and suppose that player 1 also chooses action 1. At stage 2 let player 2 choose action 2 with $a_{12} = \min_j a_{1j}$ and suppose that player 1 chooses action 2 as well. At stage 3 let player 2 choose action 3 with $a_{23} = \min_j a_{2j}$. At stage 3 player 1 can only choose 1 or 2, since he has lost all other actions. If player 1 chooses 2 (and loses 1) at stage 3, then player 2 can obviously enforce payoff a_{23}. If player 1 chooses 1 at stage 3, then player 2 can continue by playing actions 2 and 3 alternately, starting with 3 at stage 4. This would give payoffs a_{23} and a_{12} in cyclic order as long as player 1 is keeping both actions "alive." ■

Theorem 3.2. *For a matrix A of size $m \times 2$ the* (2, 2)*-restricted game $A_{2,2}$ has value $v_{2,2}$ given by $v_{2,2} = \frac{1}{2}\,\underline{v} + \frac{1}{2}\,\bar{v}$.*

Proof. At stage 1 player 1 can put probability $\frac{1}{2}$ on a row containing the first column maximum, as well as probability $\frac{1}{2}$ on a row containing the second column maximum. Then, with probability at least $\frac{1}{2}$, the initial payoff will be at least $\bar{v}$ (the minimum of these maxima). If this happens indeed, then player 1 can obtain an average reward of at least $\bar{v}$ because, as long as player 2 has two actions, player 1 can alternate between the two rows, each time receiving at least $\bar{v}$; if player 2 loses an action, player 1 can get the maximum of the remaining column. If the initial payoff is below $\bar{v}$, then player 1 can get at least $\underline{v}$ by choosing a maxmin action. Since player 2 can apply an analogous strategy, we have $v_{2,2} = \frac{1}{2}\,\underline{v} + \frac{1}{2}\,\bar{v}$. ■

We now give an example to illustrate how any (2,2)-restricted game of size $m \times n$ can be solved using backward induction.

EXAMPLE 3.3. Consider the following matrix game without saddlepoints:

$$A = \begin{pmatrix} 2 & 4 & 6 \\ 3 & \frac{4}{5} & 5 \\ 7 & \frac{5}{2} & 0 \end{pmatrix}.$$

Now suppose that at stage 1 entry (1, 1) is being selected. Then at stage 2 choosing action 1 is, for both players, weakly dominated by choosing something else. So, after stage 2 we are in one of the following submatrices with a history of playing the main diagonal from up/left to bottom/right:

$$\begin{array}{c} \\ 1 \\ 2 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} 2 & 4 \\ 3 & \frac{4}{5} \end{pmatrix} \end{array}, \quad \begin{array}{c} \\ 1 \\ 2 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 1 & 3 \end{array} \\ \begin{pmatrix} 2 & 6 \\ 3 & 5 \end{pmatrix} \end{array}, \quad \begin{array}{c} \\ 1 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} 2 & 4 \\ 7 & \frac{5}{2} \end{pmatrix} \end{array}, \quad \begin{array}{c} \\ 1 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 1 & 3 \end{array} \\ \begin{pmatrix} 2 & 6 \\ 7 & 0 \end{pmatrix} \end{array}.$$

From the arguments in the proof of Theorem 3.2 one can conclude that the resulting average rewards will respectively be 2, 3, $\frac{5}{2}$, and 2. Hence at stage 2 the players would like to play optimally in the matrix game

$$\begin{array}{c} \\ 2 \\ 3 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 2 & 3 \end{array} \\ \begin{pmatrix} 2 & 3 \\ \frac{5}{2} & 2 \end{pmatrix} \end{array},$$

which has value $\frac{7}{3}$. Similarly, we find the other entries of the matrix game representing the initial stage situation

$$A^* = \begin{pmatrix} \frac{7}{3} & 4 & 4 \\ 3 & \frac{84}{37} & \frac{25}{7} \\ 4 & \frac{5}{2} & \frac{52}{21} \end{pmatrix}.$$

The unique optimal mixed actions in A^* are $(\frac{1836}{3895}, \frac{37}{3895}, \frac{2022}{3895})$ for player 1 and $(\frac{489}{1025}, \frac{74}{1025}, \frac{462}{1025})$ for player 2, while the (2,2)-restricted value is given by $v_{2,2} = \text{val}(A^*) = \frac{3285}{1025} \approx 3.2049$. (The matrix game A has $\text{val}(A) = \frac{46}{13}$ and optimal mixed action $(\frac{9}{13}, 0, \frac{4}{13})$ for player 1 and $(\frac{3}{13}, \frac{10}{13}, 0)$ for player 2.)

4. (3,3)-RESTRICTED GAMES

THEOREM 4.1. *Consider*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

without saddlepoints. Let $\tilde{v} = \frac{1}{4}(a_{11} + a_{12} + a_{21} + a_{22})$. *Then* $v_{3,3} = \text{median}\{\underline{v}, \tilde{v}, \overline{v}\}$.

Proof. Assume that $\underline{v} \leq \tilde{v} \leq \overline{v}$ (the other cases follow similarly) and let player 1 use the following strategy:

a. (Initialization) Until a payoff of at least $\overline{v}$ is obtained, play $(\frac{1}{2}, \frac{1}{2})$ except when it is necessary to prevent an action from vanishing; in that case choose that action.

Once a payoff of at least $\overline{v}$ is obtained, continue play according to:

b. If the previous payoff is at least $\overline{v}$ and player 2 has both actions available, then choose the action different from the previous one.

c. If the previous payoff is at most $\underline{v}$ and player 2 has both actions available, then choose the previous action again.

d. If player 2 has only one action left, select the action giving the highest remaining payoff.

Now the worst thing that can happen for player 1 is that play will cycle around all entries of the matrix, giving average reward $\frac{1}{4}(a_{11} + a_{12} + a_{21} + a_{22})$. ■

Without proof we state the following result.

THEOREM 4.2. *If* $A = [a_{ij}]$ *is a matrix game of size* $m \times 2$, *then* $v_{3,3}(A) = \max\{v_{3,3}(A') : A'$ *a* (2×2)*-submatrix of* $A\}$.

To find $v_{3,3}$ for an $(m \times n)$-game we propose the following algorithm:

a. Find a lower and an upper bound for $v_{3,3}$ by examining what players can achieve by focusing on one action each:

$$\underline{v} \leq v_{3,3}(A) \leq \overline{v}.$$

b. If the previous lower and upper bounds are not equal, then find better lower and upper bounds for $v_{3,3}$ by examining what players can achieve by focusing on two actions each:

$$\max_{i_1,i_2} \min_{j_1,j_2} v_{3,3}(A_{i_1,i_2,j_1,j_2}) \leq v_{3,3}(A) \leq \min_{j_1,j_2} \max_{i_1,i_2} v_{3,3}(A_{i_1,i_2,j_1,j_2}),$$

where A_{i_1,i_2,j_1,j_2} is the subgame of A that consists of rows i_1, i_2 and of columns j_1, j_2.

c. If the previous lower and upper bounds are not equal, then find better lower and upper bounds for $v_{3,3}$ by examining what players can achieve by focusing on three actions each:

$$\max_{i_1,i_2,i_3} \min_{j_1,j_2,j_3} \tfrac{1}{3} \sum_{t=1}^{3} a_{i_t j_t} \leq v_{3,3}(A) \leq \min_{j_1,j_2,j_3} \max_{i_1,i_2,i_3} \tfrac{1}{3} \sum_{t=1}^{3} a_{i_t j_t}.$$

d. If the previous lower and upper bounds are not equal, then apply backward induction to find $v_{3,3}(A)$. This backward induction proceeds in a manner similar to that for (2, 2)-restricted games (cf. Example 3.3).

Examples illustrating this algorithm are provided in Joosten *et al.* (1991).

5. Concluding Remarks

In the previous sections we have seen how to solve (r_1, r_2)-restricted games for some specific values of r_1 and r_2. Solution methods also exist for (r, ∞)-restricted games. These games are also known as single-controller stochastic games and solutions for such games can be found by solving a single linear program and its dual (cf. Vrieze, 1981; Filar and Raghavan, 1984).

To extend these results to other (r_1, r_2)-values appears to be a non-trivial problem. In contrast with Theorem 3.1, for example, it is not true that $v_{3,r} = \underline{v}$ for any matrix A. However, an algorithm like that at the end of the previous section would obviously work for any (r, r)-restricted game A with arbitrary r under the assumption that one knows how to compute $v_{r,r}$ for all real submatrices of A. Thus, a first problem seems to be to find a procedure for computing v_{r_1,r_2}, with r_1, r_2 arbitrary, for (2×2)-matrices. Such a procedure seems to call for a kind of induction on r, but it is not clear to us how this should work.

Needless to say, in view of these difficulties, finding a solution for the non-zero-sum model, which would obviously be much more appropriate for economic applications, will be even more problematic.

Acknowledgment

The idea of Theorem 4.1 is due to Koos Vrieze. We gratefully thank him for stimulating discussions.

References

Arrow, K. J. (1962). "The Economic Implications of Learning by Doing," *Rev. Econ. Stud.* **29,** 155–173

Blackwell, D., and Ferguson, T. S. (1968). "The Big Match," *Ann. Math. Statist.* **39,** 159–163.

Filar, J. A., and Raghavan, T. E. S. (1984). "A Matrix Game Solution of the Single-Controller Stochastic Game," *Math. Oper. Res.* **9,** 356–362.

Joosten, R., Peters, H., and Thuijsman, F. (1991). "Unlearning by Not Doing: Repeated Games with Vanishing Actions," Report M 91-14, Department of Mathematics, University of Limburg, Maastricht.

Mertens, J.-F., and Neyman, A. (1981). "Stochastic Games," *Int. J. Game Theory* **10,** 53–66.

Raghavan, T. E. S., and Filar, J. A. (1991). "Algorithms for Stochastic Games, a Survey," *Z. Oper. Res.* **33,** 437–472.

Shapley, L. S. (1953). "Stochastic Games," *Proc. Nat. Acad. Sci. U.S.A.* **39,** 1095–1100.

Vrieze, O. J. (1981). "Linear Programming and Undiscounted Stochastic Games in Which One Player Controls Transitions," *OR Spektrum* **3,** 29–35.