Automated Equilibrium Analysis of Repeated Games with Private Monitoring: A POMDP Approach

(Extended Abstract)

YongJoon Joe[1], Atsushi Iwasaki[1], Michihiro Kandori[2], Ichiro Obara[3], and Makoto Yokoo[1]
1: Kyushu University, Japan, {yongjoon@agent., iwasaki@, yokoo@}inf.kyushu-u.ac.jp
2: University of Tokyo, Japan, kandori@e.u-tokyo.ac.jp,
3: UCLA, California, ichiro.obara@gmail.com

ABSTRACT

The present paper investigates repeated games with *imperfect private monitoring*, where each player privately receives a noisy observation (signal) of the opponent's action. Such games have been paid considerable attention in the AI and economics literature. Identifying pure strategy equilibria in this class has been known as a hard open problem. Recently, we showed that the theory of partially observable Markov decision processes (POMDP) can be applied to identify a class of equilibria where the equilibrium behavior can be described by a finite state automaton (FSA). However, they did not provide a practical method or a program to apply their general idea to actual problems. We first develop a program that acts as a wrapper of a standard POMDP solver, which takes a description of a repeated game with private monitoring and an FSA as inputs, and automatically checks whether the FSA constitutes a symmetric equilibrium. We apply our program to repeated Prisoner's dilemma and find a novel class of FSA, which we call k-period mutual punishment (k-MP). The k-MP starts with cooperation and defects after observing a defection. It restores cooperation after observing defections k-times in a row. Our program enables us to exhaustively search for all FSAs with at most three states, and we found that 2-MP beats all the other pure strategy equilibria with at most three states for some range of parameter values and it is more efficient in an equilibrium than the grim-trigger.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multi-agent systems*; J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Algorithms, Economics, Theory

Keywords

Game theory, repeated games, private monitoring, POMDP

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

1. INTRODUCTION

We consider repeated games with *imperfect private monitoring*, where each player privately receives a noisy observation (signal) of the opponent's action. This class of games represents long-term relationships among players and has a wide range of applications, e.g., secret price cutting and agent planning under uncertainty. Therefore, it has been paid considerable attention in the AI and economics literature. In particular, for the AI community, the framework has become increasingly important for handling noisy environments. In fact, Tennenholtz and Zohar consider repeated congestion games where an agent has limited capability in monitoring the actions of her counterparts [5].

Analytical studies on this class of games have not been quite successful. The difficulty comes from the fact that players do not share common information under private monitoring, and finding pure strategy equilibria in such games has been known as a hard open problem [4]. Under private monitoring, each player cannot observe the opponents' private signals, and he or she has to draw statistical inferences about the history of the opponents' private signals. The inferences quickly become very complicated over time, even if players adopt relatively simple strategies [1]. As a result, finding a profile of strategies which are mutual best replies after any history, i.e., finding an equilibrium, is a quite demanding task.

Quite recently, we show that the theory of the partially observable Markov decision process (POMDP) can be used to identify equilibria, when equilibrium behavior is described by a finite state automaton (FSA) [2]. This result is significant since it implies that by utilizing a POMDP solver, we can systematically determine whether a given profile of finite state automata can constitute an equilibrium. Furthermore, this result is interesting since it connects two popular areas in AI and multi-agent systems, namely, POMDP and game theory.

We first develop a program that acts as a wrapper of a standard POMDP solver. Furthermore, as a case study to confirm the usability of this program, we identify equilibria in an infinitely repeated prisoner's dilemma game, where each player privately receives a noisy signal about each other's actions.

2. REPEATED GAMES WITH PRIVATE MONITORING AND FSA

A finite state automaton (FSA) is a popular approach for

compactly representing the behavior of a player in repeated games. We focus on a *symmetric pure finite state equilibrium* (SPFSE), which is a pure strategy sequential equilibrium of a repeated game with private monitoring, where each player's behavior on the equilibrium path is given by an FSA. A sequential equilibrium is a refinement of Nash equilibrium for dynamic games of imperfect information.

We apply the POMDP technique to the prisoner's dilemma model analyzed by [2]. The stage game payoff is given as follows.

	$a_2 = C$	$a_2 = D$
$a_1 = C$	$1, 1$	$-y, 1+x$
$a_1 = D$	$1+x, -y$	$0, 0$

Each player's private signal is $\omega_i \in \{g, b\}$ (*good* or *bad*), which is a noisy observation of the opponent's action. For example, when the opponent chooses C, player i is more likely to receive the correct signal $\omega_i = g$, but sometimes an observation error provides a wrong signal $\omega_i = b$. Let us introduce the joint distribution of private signals $o(\boldsymbol{\omega} \mid \boldsymbol{a})$ for the prisoner's dilemma model. When the action profile is (C, C), the joint distribution is given as follows (when the action profile is (D, D), p and s are exchanged).

	$w_2 = g$	$w_2 = b$
$w_1 = g$	p	q
$w_1 = b$	r	s

Similarly, when the action profile is (C, D), the joint distribution of private signals is given as follows (when the action profile is (D, C), v and u are exchanged).

	$w_2 = g$	$w_2 = b$
$w_1 = g$	t	u
$w_1 = b$	v	w

These joint distributions of private signals require only the constraints of $p + q + r + s = 1$ and $t + u + v + w = 1$.

We define a monitoring structure that is *nearly-perfect*. We say monitoring is nearly-perfect if each player is always likely to perfectly observe the opponent's action in each period, i.e., $p = v$, $q = r = t = w$, and $s = u = 1 - p - 2q$, where p is much larger than q or s. Although the monitoring structure is quite natural, systematically finding equilibria in such structure has not been possible without utilizing a POMDP solver. Alternatively, we say monitoring is *almost-public* if players are always likely to get the *same* signal (after (C, D), for example, players are likely to get (g, g) or (b, b)), i.e., $p + s = t + w \approx 1$ and $q = r = u = v \approx 0$.

Let us summarize the existing FSAs. First, grim-trigger (GT) is a well-known FSA under which a player first cooperates, but as soon as she observes defection, she defects forever. GT can often constitute an equilibrium. Second, tit-for-tat (TFT) is another well-known FSA in Fig. 1. It is well known that TFT does not prescribe mutual best replies after a deviation (hence it is *not* a subgame perfect Nash equilibrium (SPNE)). This problem does not arise under almost-public monitoring. Finally, 1-period mutual punishment (1-MP) in Fig. 2 is known as *Pavlov* [3]. According to this FSA, a player first cooperates. If her opponent defects, she also defects, but after one period of mutual defection, she returns to cooperation. It is well-known that Pavlov

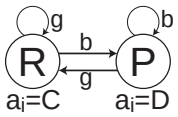

Figure 1: TFT

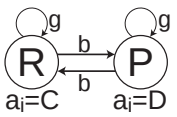

Figure 2: 1-MP

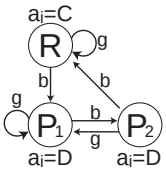

Figure 3: 2-MP

can constitute an SPNE under perfect monitoring. However, this has not been investigated well in the setting of private monitoring.

3. K-PERIOD MUTUAL PUNISHMENT

Let us first consider 1-MP. We can see that after one observation error occurs, players can quickly return to the mutual cooperation state RR. The expected probability (in the invariant distribution) that players are in state RR is about $p - 2q$. Unfortunately, 1-MP does not constitute an SPFSE in our parameterization, since it is too forgiving.

Therefore we generalize the idea of 1-MP to k-period mutual punishment (k-MP). Under this FSA, a player first cooperates. If her opponent defects, she also defects, but after k consecutive periods of mutual defection, she returns to cooperation. Figure 3 shows the FSAs of 2-MP. 2-MP is less forgiving than 1-MP, since it cooperates approximately once in every three periods to the opponent who always defects. By increasing k, we can make this strategy less forgiving. When $k = \infty$, this strategy becomes equivalent to GT.

Although it is somewhat counter-intuitive, requiring such mutual defection periods is beneficial in establishing a robust coordination among players under nearly-perfect monitoring. In contrast, under almost-public monitoring, TFT can better coordinate players' behavior; TFT can be an equilibrium, while k-MP is not. In both cases, GT can be an equilibrium. Accordingly, our program helps us to gain important insights into the way players coordinate their behavior under different private monitoring structures.

Furthermore, we exhaustively search for small-sized FSAs that can constitute an equilibrium under nearly-perfect monitoring. We enumerate all possible FSAs with at most three states, i.e., 5832 FSAs, which is obtained from the numbers of actions, signals, and states, and check whether they constitute an SPFSE. We found that only eleven FSAs (after removing equivalent ones) could be an SPFSE in a reasonably wide range of signal parameters. In addition, among them, 2-MP is the only FSA that is more efficient than GT.

4. REFERENCES

[1] M. Kandori. Repeated games. *Game theory*. pages 286–299. Palgrave macmillan, 2010.

[2] M. Kandori and I. Obara. Towards a Belief-Based Theory of Repeated Games with Private Monitoring: An Application of POMDP. mimeo, 2010.

[3] D. Kraines and V. Kraines. Pavlov and the prisoner's dilemma. *Theory and Decision*, 26:47–79, 1989.

[4] G. Mailath and L. Samuelson. *Repeated Games and Reputation*. Oxford University Press, 2006.

[5] M. Tennenholtz and A. Zohar. Learning equilibria in repeated congestion games. In *AAMAS*, pages 233–240, 2009.