Available online at www.sciencedirect.com
ScienceDirect
Applied Soft Computing 7 (2007) 1188–1196
Applied Soft Computing
www.elsevier.com/locate/asoc

Chaotic time series prediction for the game, Rock-Paper-Scissors

Franco Salvetti [a,*], Paolo Patelli [b], Simone Nicolo [a]

[a] Department of Computer Science, University of Colorado at Boulder, 430 UCB, Boulder, CO 80309–0430, USA
[b] T-13 Complex System Group, Theoretical Division, CNLS Los Alamos National Laboratory, Mail Stop B-213, Los Alamos, NM 87545, USA

Available online 28 February 2006

Abstract

Two players of Rock-Paper-Scissors are modeled as adaptive agents which use a reinforcement learning algorithm and exhibit chaotic behavior in terms of trajectories of probability in mixed strategies space. This paper demonstrates that an external super-agent can exploit the behavior of the other players to predict favorable moments to play against one of the other players the symbol suggested by a sub-optimal strategy. This third agent does not affect the learning process of the other two players, whose only goal is to beat each other. The choice of the best moment to play is based on a threshold associated with the Local Lyapunov Exponent or the Entropy, each computed by using the time series of symbols played by one of the other players. A method for automatically adapting such a threshold is presented and evaluated. The results show that these techniques can be used effectively by a super-agent in a game involving adaptive agents that exhibit collective chaotic behavior.

Keywords: Time series prediction; Chaos theory; Game theory; Local Lyapunov Exponent; Entropy filtering; Reinforcement learning; Agent irrationality; Rock paper scissors game

1. Introduction

Game theory [12] and the related concept of Nash equilibrium [11] have produced important practical and theoretical results using the idealization of a *rational agent* [9] which seeks only to maximize its utility. Despite these results, however, it is evident that a rational agent cannot be a completely accurate model of a real agent [14,7]. As a consequence we are observing an increasing interest in the modeling of agent-*irrationality*: ‘‘Standard models in economics stress the role of intelligent agents which maximize utility. However, there may be situations where, for some purposes, constraints imposed by market institutions dominate intelligent agent behavior’’ [6]. A possible alternative to rationality is ‘‘bounded-rationality’’, where the agent always takes the action that is expected to optimize its performance measure, given some informational and computational constraints [4]. Also, an agent which makes decisions by using simple heuristics or rules of thumb is considered a bounded-rational agent [8].

Real players of Rock-Paper-Scissors[1] (RPS) always use some impulse or inclination to choose a throw, and will therefore settle into unconscious but nonetheless predicable patterns [15]. We expect that such patterns are weaknesses in their behavior.

Chaos in the probability space trajectory has been demonstrated in recent studies of RPS [16,17], where the players change the probability of playing each symbol by using reinforced learning [20] based on coupled Ordinary Differential Equations (ODE). Even with a more *realistic* adaptive agent which uses reinforced learning based on micro-founded heuristics [8] there is still chaos in the probability space trajectories [13]. We conjecture that although local instability on attractors prohibits accurate long-term predictions, short-term predictions can be made with varying degrees of accuracy [10].

Given a time-series which represents the behavior of a system, the usual aim is to predict its behavior in the future [22]. Here we are interested in showing that by means of the Local Lyapunov Exponent (LLE) [1,2] it is possible for a super-agent to predict the behavior of one player [5] of RPS in order to ameliorate its own performance. We consider the game RPS only as a metaphor of, for example, some real market where the super-agent represents a speculator.

Our goal is not to suggest the optimal symbol to be played, but to develop a *theory of the best moment for playing* the symbol suggested by a sub-optimal strategy. We base our analysis on the observable behavior of one of the players, and show that it is possible for the super-agent to improve its

* Corresponding author. Tel.: +1 303 817 9805; fax: +1 303 565 6883.
E-mail addresses: franco.salvetti@colorado.edu (F. Salvetti), paolo@lanl.gov (P. Patelli), simone.nicolo@colorado.edu (S. Nicolo).

[1] See http://www.worldrps.com for an exhaustive explanation of the game.

1568-4946/$ – see front matter
doi:10.1016/j.asoc.2006.01.006

performance by playing rounds intermittently on the basis of indicators such as the LLE and the Entropy.

2. Framework

The RPS game is a two-person, zero-sum game whose payoff matrix is presented in Table 1.

The RPS game has a Nash equilibrium in the mixed strategies space. In the Nash equilibrium each player randomizes the pure strategies with probability 1/3. At equilibrium no player has an incentive to deviate from the Nash strategy.

We study a repeated version of RPS where the same players play the same game iteratively. Unlike a single-shot game, a repeated game allows players to use strategies contingent on past moves by exploiting, if possible, any weakness in the strategy of the other player. Learning and adaptation become fundamental components of the players' interaction.

Our agents use a reinforced learning algorithm to adapt their behavior in response to changes in the behavior of the other player. The reinforcement learning algorithm updates the probability associated with each pure strategy on the basis of the success of past strategies. If the player plays the strategy $s_i \in \{R, P, S\}$, then the complementary strategy for s_i, identified by $\bar{s}_i$, is the strategy that wins against s_i. For example, if the agent plays *Rock*, then the complementary strategy is *Scissors*. If a pure strategy is successful, then the player increases the probability associated with that strategy, and decreases the probability of playing the complementary strategy. Vice-versa, if the strategy is unsuccessful, the agent decreases the probability associated with that strategy and increases the probability of the complementary strategy. The agent updates the probability of playing strategy s_i, $P_{t+1}(s_i)$ at time $t+1$ in the following way:

$$P(s_i)_{t+1} = P(s_i)_t + w_i(\alpha(1 - P(s_i)_t)) \quad (1)$$

where w_i assume value 1 if the strategy outcome is positive, and -1 if negative. The parameter α is called the *learning rate,* and determines the agent's adaptation velocity. We update the probability associated with the complementary strategy in a similar way:

$$P(\bar{s}_i)_{t+1} = P(\bar{s}_i)_t + (1 - w_i)\alpha(1 - P(\bar{s}_i)_t) \quad (2)$$

We perform different experiments changing each agent's initial condition and learning rate parameter α. The projection of an agent's trajectory in the probability simplex for an experiment with $\alpha = 0.1$ and 20,000 repetitions is reported in Fig. 1.

The trajectory resembles a *random* set of points. However, if we project the trajectory within some short intervals of time, as shown in Fig. 2, we notice that the set of seemingly *random* points in Fig. 1 has a precise internal structure, typical of chaotic phenomena. These *scattered orbits* are frequently present in trajectories generated by using sets of heuristics employed to update the probability distribution of each player. The *orbits* in Fig. 2 are not as *smooth* as those proposed in [16,17] because the update of the probability is not controlled by an equation, but follows heuristics which should capture the behavior of a real player who, in a highly discrete way, can change the probability used to play a symbol.

For instance a real player can decide to halve the probability of playing *Rock* as a result of a sequence of losses, updating the other probability to preserve:

$$P_{\text{Rock}} + P_{\text{Paper}} + P_{\text{Scissors}} = 1 \quad (3)$$

Some of these heuristics capture the empirical observation that a player increases its confidence in a symbol after a win, and that the increase in confidence is somehow related with the current probability distribution.

Table 1
Payoff matrix of two-person Rock-Paper-Scissors

Player B	Player A		
	R	S	P
R	0, 0	1,−1	−1, 1
S	−1, 1	0, 0	1, −1
P	1, −1	−1, 1	0, 0

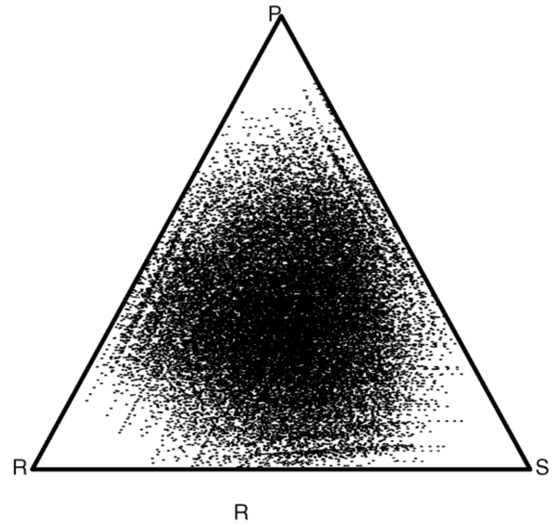

Fig. 1. Trajectory of 20,000 points in the probability simplex generated by two agents learning with $\alpha = 0.1$.

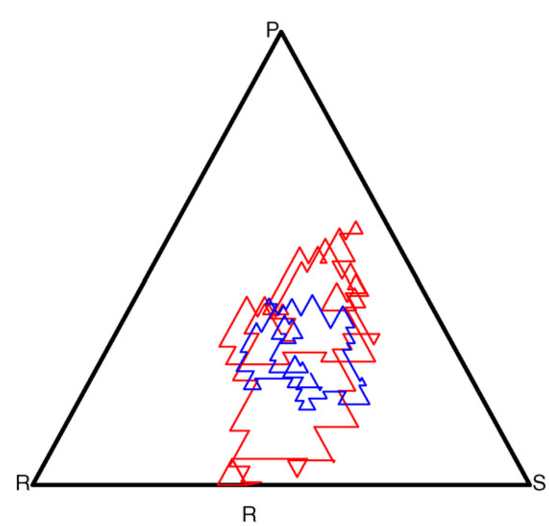

Fig. 2. Two portions of the trajectory of 20,000 points in the probability simplex generated by two agents learning with a probability update of $\alpha = 0.1$ as presented in Fig. 1.