Proceedings of the 2006 WSEAS International Conference on Mathematical Biology and Ecology, Miami, Florida, USA, January 18-20, 2006 (pp249-253)

Learning strategies for a predator operating in variable model-mimic-alternative prey environments

A.TSOULARIS
IIMS,
MASSEY UNIVERSITY,
ALBANY,
PRIVATE BAG 102 904,
AUCKLAND,
NEW ZEALAND

Abstract: - In this paper we analyse the interactions of four biological species, a predator and two types of prey, the models and mimics. The models are noxious prey that must be avoided by the predator and the mimics are palatable prey and primary source of food for the preadator, that resemble in appearance the models, thus escaping consumption. The alternative prey represents a secondary food source for the predator. We identify the predator as a learning automaton with two actions, consume prey or ignore prey that elicit favourable and unfavourable probabilistic responses from the environment. Two kinds of environment are considered, stationary with fixed penalty probabilities and nonstationary with variable penalty probabilities. All prey are assumed to grow logistically. A benefit function is constructed for the predator that measures the consumption level at each stage of predation. Finally, strategies for increasing consumption are derived in terms of the parameters of the learning process.

Keywords:-learning automaton, reinforcement learning, mimics, models, alternative prey.

1. Introduction

In this paper we analyse in detail a linear reinforcement learning algorithm designed to allow a predator (the learning automaton) to operate efficiently in terms of acceptable prey consumption in an environment occupied by palatable and unpalatable prey and characterized by a penalty probability for each predator action. The predator chooses to either ignore prey or consume prey. Our present work builds on the framework laid out in a previous article [1].

A brief description of the concept of the **learning automaton** is given in [1]. For a comprehensive introduction the book by Narendra and Thathachar [2] is recommended. Linear reinforcement algorithms are based on the simple premise of increasing the probability of that action that elicits a favourable response by an amount proportional to the total value of all other action probabilities. Otherwise, it is decreased by an amount proportional to its current value. In this work we also adopt the 2-action **Linear Reward-Penalty** (L_{R-P}) scheme as the predator's learning algorithm., with actions a_1 (ignore) and a_2 (eat). A penalty is associated with either ignoring mimics or alternative prey, or consuming models. The penalty probabilities on these actions are defined as follows:

$$c_1 = probability(mimic \mid a_1)$$
$$f_1 = probability(alternative\ \ prey \mid a_1)$$
$$c_2 = probability(model \mid a_2)$$

The L_{R-P} algorithm is described below:

$$\left.\begin{aligned} p_1(k+1) &= p_1(k) + \alpha[1 - p_1(k)] \\ p_2(k+1) &= (1-\alpha)\, p_2(k) \end{aligned}\right\}$$
$$a(k) = a_1,\ response\ is\ \ favourable,\ 0 < \alpha < 1 \qquad (1)$$
$$\left.\begin{aligned} p_1(k+1) &= (1-\beta)\, p_1(k) \\ p_2(k+1) &= p_2(k) + \beta[1 - p_2(k)] \end{aligned}\right\}$$
$$a(k) = a_1,\ response\ is\ unfavourable,$$
$$0 < \beta < 1,\ (\beta \neq \alpha)$$

The expectation of the consumption probability, $p_2(k+1)$, conditioned on $p_2(k)$, is given by:

$$\begin{aligned} &\bar{p}_2(k+1) = E[p_2(k+1) \mid p_2(k)] = \\ &(c_2 - c_1 - f_1)(\alpha - \beta) p_2^2(k) \\ &+ [1 + \alpha(c_1 + f_1 - c_2) - 2\beta(c_1 + f_1)] p_2(k) \\ &+ \beta(c_1 + f_1) \end{aligned} \qquad (2)$$

2. Prey population growth

The mimic, model and alternative prey populations, X, M and A respectively, grow logistically as follows:

$$\begin{aligned} X(k+1) &= X(k) + r_1 X(k)\left(1 - \frac{X(k)}{K_1}\right) - \\ &d_2 p_2(k) X(k) \\ A(k+1) &= A(k) + r_2 A(k)\left(1 - \frac{A(k)}{K_2}\right) - \\ &f_2 p_2(k) A(k) \\ M(k+1) &= M(k) + r_3 M(k)\left(1 - \frac{M(k)}{K_3}\right) - \\ &c_2 p_2(k) M(k) \end{aligned} \qquad (3)$$

where r_1, r_2, r_3 are the intrinsic growth rates, K_1, K_2, K_3 are the carrying capacities, and d_2, f_2, c_2 are the prey consumption probabilities $(d_2 + f_2 + c_2 = 1)$.

3. The benefit function

The net expected benefit to the predator is assessed in terms of capturing a palatable mimic and the unnecessary energy expended in capturing an unpalatable model [3]. If *b*, *a* and *c* are the parameters associated with the consumption of a single mimic, alternative prey and model respectively, the expected net change in benefit at stage *k* is given by

$$\begin{aligned} &\Delta\bar{B}(k) = \\ &\bar{p}_2(k)[bX(k)d_2 + aA(k) f_2 - cM(k)c_2] \end{aligned} \qquad (4)$$

The objective of the predator is to optimize its next stage benefit by adjusting accordingly the learning parameters α and β at the current stage.

4. Consumption strategies for Stationary prey environments

A stationary pey environment is one in which the penalty probabilities remain. constant. In this case the consumption probability given in (2) converges to the asymptotic value which is the fixed point solution of (2).

The maximum rate of net benefit change is determined by the sign of the derivatives, $\frac{\partial\Delta\bar{B}(k)}{\partial\alpha}$, $\frac{\partial\Delta\bar{B}(k)}{\partial\beta}$, and consequently by the sign of the derivatives, $\frac{\partial\bar{p}_2(k)}{\partial\alpha}$, $\frac{\partial\bar{p}_2(k)}{\partial\beta}$. The optimal strategies for the predator is identical to the those in Table 1 of our previous work [1], with c_1 in that table replaced by $c_1 + f_1$ to account for the presence of alternative prey.

5. Nonstationary prey environments

In this section we analyse the performance of the learning algorithm of the last section when each penalty probability, $c_1 + f_1$, c_2 is a monotonically increasing function of the respective action probability, a_i, i =1,2. We base our decision on the reasonable assumption that if the predator is ignoring all prey with a certain frequency, palatable prey amongst them are essentially ignored at a less frequent rate, and by the same token, we extend this assumption to the frequency of consumption. Thus at each stage k:

$$c_1(k) + f_1(k) = g_1 p_1(k), \quad 0 < g_1 < 1$$
$$c_2(k) = g_2 p_2(k), \quad 0 < g_2 < 1$$

The two coefficients, g_1 and g_2, can be interpreted respectively as the fraction of falsely avoided mimics and alternative prey in the proportion of overlooked prey, and the fraction of falsely consumed models in the proportion of consumed prey. Values of either factor close to 0 indicate that the predator commits either penalty infrequently, whereas values close to 1 indicate a large penalty frequency. The complementary expressions, $1-g_1$ and $1-g_2$, may be thought of as the predatory efficiency in avoiding the wrong prey and consuming the right prey respectively.

The expectation of the action probability, $p_2(k)$, conditioned on $p_2(k\text{-}1)$, is a third-order polynomial in $p_2(k\text{-}1)$:

$$\begin{aligned}&\bar{p}_2(k) = E[p_2(k) \mid p_2(k-1)] = \\ &(g_1+g_2)(\alpha-\beta)p_2^3(k-1) + \\ &(3\beta g_1 - 2\alpha g_1 - \alpha g_2)p_2^2(k-1) + \\ &(1+\alpha g_1 - 3\beta g_1)p_2(k-1) + \\ &\beta g_1\end{aligned} \tag{5}$$

The asymptotic probability can be found as one of the three roots of the resulting cubic polynomial, based on the work of Cardan [4]. For algebraic convenience we shall confine ourselves to the case $\alpha = \beta$, in which case:

$$\begin{aligned}&\bar{p}_2(k) = \alpha(g_1 - g_2)p_2^2(k-1) \\ &+ (1 - 2\alpha g_1)p_2(k-1) + \alpha g_1\end{aligned} \tag{6}$$

with $g_1 \neq g_2$. The scheme admits the asymptotically stable probability:

$$\bar{p}_2^* = \frac{\sqrt{g_1}}{\sqrt{g_1} + \sqrt{g_2}} \tag{7}$$

The expected net change in benefit is now

$$\Delta\bar{B}(k) = \bar{p}_2(k)[bX(k)d_2(k) + aA(k)f_2(k)] - \bar{p}_2(k)cM(k)c_2(k)$$

Let

$$f_2(k) = \gamma(1 - c_2(k))$$
$$d_2(k) = (1-\gamma)(1 - c_2(k))$$

where γ, $0 < \gamma < 1$, is a parameter reflecting the fraction of the prey consumption frequency dedicated to the alternative prey. Note again that $f_2(k) + d_2(k) + c_2(k) = 1$, for all k. The benefit change is rewritten as

$$\begin{aligned}&\Delta\bar{B}(k) = \bar{p}_2(k)[bX(k)(1-\gamma) + aA(k)\gamma] - \\ &g_2\bar{p}_2^2(k)[bX(k)(1-\gamma) + aA(k)\gamma + cM(k)]\end{aligned} \tag{8}$$

We treat the learning parameter, α, as the decision variable at each stage, k, that influences the magnitude of the expected change in the net benefit, at the next stage. To test whether the expected benefit is continually increasing we consider the partial derivative of the benefit change with respect to α. Since the dependence of $\Delta\bar{B}(k)$ on α is implicit only through $\bar{p}_2(k)$, which is a linear function of α, the derivative will be simply the slope of $\Delta\bar{B}(k)$. Thus the optimal action will

depend on the choice of parameters that yield the maximum slope. We write the derivative:

$$\frac{\partial \Delta \overline{B}(k)}{\partial \alpha} = \left[\left(bX(k)(1-\gamma)+aA(k)\gamma\right)(1-2g_2\overline{p}_2(k)) - 2cM(k)g_2\overline{p}_2(k)\right]\frac{\partial \overline{p}_2(k)}{\partial \alpha} \quad (9)$$

For $g_1 \neq g_2$:

$$\frac{\partial \overline{p}_2(k)}{\partial \alpha} = (g_1-g_2)\left(p_2-\frac{\sqrt{g_1}}{\sqrt{g_1}+\sqrt{g_2}}\right)\left(p_2-\frac{\sqrt{g_1}}{\sqrt{g_1}-\sqrt{g_2}}\right)$$

Since

$$\frac{\partial \Delta \overline{B}^2(k)}{\partial \alpha^2} = -2g_2\left(\frac{\partial \overline{p}_2(k)}{\partial \alpha}\right)^2 (bX(k)(1-\gamma)+aA(k)\gamma+cM(k)) < 0$$

the benefit change can only be maximal with respect to the learning parameter, α, when the consumption probability reaches its asymptotic value, (7), or also prior to that when it crosses the critical value at some k:

$$\hat{p}_2(k) = \frac{bX(k)(1-\gamma)+aA(k)\gamma}{2g_2\left[bX(k)(1-\gamma)+aA(k)\gamma+cM(k)\right]} \quad (10)$$

Note that $\hat{p}_2(k)$ may not always exist.

The optimal strategy for the predator's consumption frequency will be determined upon the results of the comparison between the values of g_2 and the value of the ratio $\frac{bX(k)(1-\gamma)+aA(k)\gamma}{2\left[bX(k)(1-\gamma)+aA(k)\gamma+cM(k)\right]}$.

For $g_1 = g_2$:

$$\frac{\partial \overline{p}_2(k)}{\partial \alpha} = g_1(1-2p_2)$$

In this simpler scenario with equal efficiency measures, the predator's optimal consumption strategy will be dictated upon comparison of the consumption probabilities $\frac{1}{2}$ and if it exists, $\hat{p}_2(k)$, as given by (10).

Figure 1 displays the improvement in the benefit change by lowering the learning parameter from $\alpha = 0.9$ to $\alpha = 0.1$.

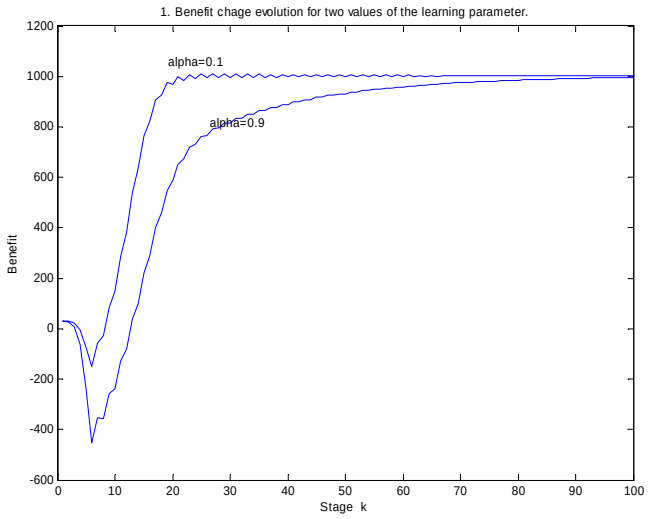

Figure 1. Benefit change growth for $g_1 = 0.5, g_2 = 0.1, p_2(0) = 0.9, a = 1, b = 1, c = 0.9, \gamma = 0.1, r_1 = 1, r_2 = 2, r_3 = 2, X(0) = 50, M(0) = 200, A(0) = 100, K_1 = 5000, K_2$=10000, K_3=20000.

6. Discussion

In this paper we have explored the concept of a predator as a learning automaton feeding on prey that can be broadly categorized as either palatable (the mimics and alternative prey) or unpalatable (the models). The predator's actions is to either attack the prey or simply ignore it. Each action elicits a probabilistic response from the

environment that is classified as favourable or unfavourable. A response is deemed favourable if the prey consumed is of the palatable type or if the prey ignored is unpalatable and deemed unfavourable if the prey ignored is palatable or the prey consumed is unpalatable. This distinction made when ignoring prey is related to the predator's ability to discriminate effectively against models. If the predator senses that the prey ignored is of palatable nature it will decrease the frequency of avoidance and vice versa. A suitable function has been constructed to take into account the net energetic benefit to predator. Conditions for maximal increase in benefit have been derived dependent upon the prey populations, and the efficiency coefficients g_1 and g_2. The present work effectively extends the theoretical framework presented in [1] by including a third type of prey.

References:

[1] A.Tsoularis, Reinforcement learning in predator-prey interactions, *WSEAS Transactions on Biology and Biomedicine*, Vol. 2, Issue 2, April 2005, pp. 141-146.

[2] K.Narendra, M.A.L.Thathachar, *Learning Automata: An Introduction*, Prentice Hall, Englewood Cliffs NJ, 1989.

[3] G.F.Estabrook, D.C.Jespersen, Strategy for a predator encountering a model-mimic system, *The American Naturalist*, Vol. 108 No. 962, July-August 1974, pp. 443-457.

[4] W.L.Ferrar, *Higher Algebra*, Oxford University Press, Oxford, 1962.