確率学習による適合度評価機構を有する遺伝的アルゴリズム (2)
3H－8　— 戦略獲得への応用 —

冨川裕樹　棟朝雅晴　高井昌彰　佐藤義治
北海道大学工学部

1　はじめに

従来の遺伝的アルゴリズム (Genetic Algorithm, 以下 GA と略す) [2] を、適合度値の評価に時間を要する問題や確率環境への適応学習に適用することは困難である。このような問題に対して、確率学習による適合度評価機構を有する遺伝的アルゴリズム StGA (Stochastic Genetic Algorithm)[1] が提案されている。

StGA では、適合度の評価に確率学習オートマトン SLA (Stochastic Learning Automata)[3] を用いている。SLA には、状態空間のサイズが非常に大きい場合に収束が著しく遅くなるという欠点がある。StGA はこの点を改善し、問題に適応する形で状態空間を圧縮することを目的としている。

我々は StGA が状態空間の圧縮を行なうという点に着目し、これを戦略の種類が非常に多いゲームにおける戦略の獲得に応用できるのではないかと考えた。本論文では、StGA と SLA をゲームにおける戦略の獲得を行なう手段としてインプリメントして対戦を行ない、状態空間のサイズが大きい場合における StGA の有効性の検証を行なう。

2　StGA の概要

StGA では、モデルの枠組として GA の集団と学習の対象である環境が与えられる。集団は、環境への入力である行動 (Action) を文字列としてコーディングした個体から成る。各個体には、行動がなされる時にその個体が選択される確率である適合度値が定義される。

行動の結果を用いて、集団内の適合度値の評価が行なわれる。この評価には SLA における linear reward-penalty scheme (L_{R-P}) が使用される。ある個体が選択され、その時の行動が成功した場合にはその個体の適合度値を高くし、その分その他の個体の適合度値を一定割合で低くする。また失敗した場合にはその個体以外のすべての個体の適合度値を一定割合で高くし、その分失敗を引き起こした個体の適合度値を低くする。

A Genetic Algorithm which has a Fitness Evaluation Mechanism by Stochastic Learning (2) — An Application to Strategy Acquisition of Games —
Yuki Tomikawa, Masaharu Munetomo, Yoshiaki Takai, and Yoshiharu Sato
Hokkaido University, Sapporo 060 JAPAN.

行動が失敗した場合には、一定確率で集団に対して遺伝的操作である交叉・突然変異を適用し、状態空間内の探索を行なう。これにより行動の成功確率を最大化する。

3　ゲームについて

プレイヤー p_1, p_2 の対戦によるゲームのルールを次のように定める。

- 各プレイヤーの取り得る戦略の種類はそれぞれ n_{p_1}, n_{p_2} 通りある。
- p_1 が戦略 s_1, p_2 が戦略 s_2 を選択した時、利得として p_1 へ $a^{p_1}_{s_1,s_2}$、p_2 へ $a^{p_2}_{s_1,s_2}$ を与える利得行列

$$A^{p_1} = [a^{p_1}_{i,j}](0 \le i < n_{p_1})(0 \le j < n_{p_2})$$

$$A^{p_2} = [a^{p_2}_{k,l}](0 \le k < n_{p_1})(0 \le l < n_{p_2})$$

  がある（利得行列 A^{p_1}, A^{p_2} の内容はプレイヤーからは不可視）。
- 繰り返し対戦を行ない、利得の累積値を相手プレイヤーより多くすることを目的とする。

4　実験

プレイヤー p_1 を StGA による学習を行なうプレイヤー、プレイヤー p_2 を SLA による学習を行なうプレイヤーとして対戦を行なった。プレイヤーの取り得る戦略の種類を $n = n_{p_1} = n_{p_2} = 1024$ 通りとした。利得行列として $A^{p_1} = -A^{p_2}$ である図1のような行列を用いた。

この行列は、$n/2$ 番目の戦略を取った時に勝つ（正の利得を得る）可能性が一番高くなっているが、この場合でも勝つ可能性は約 50%になっている。このため、この $n/2$ 番目の戦略を取り続けたとしても勝ち続けることはできない。

StGA における個体は、戦略の番号を二進表現したものを用いた。また、個体数は 20、遺伝的操作である突然変異・交叉はともに行動が失敗した時に 20% の確率で行なわれるものとした。

以上の条件下でシミュレーション実験を行なった。このときのプレイヤー p_1 (StGA) が得た利得の累積

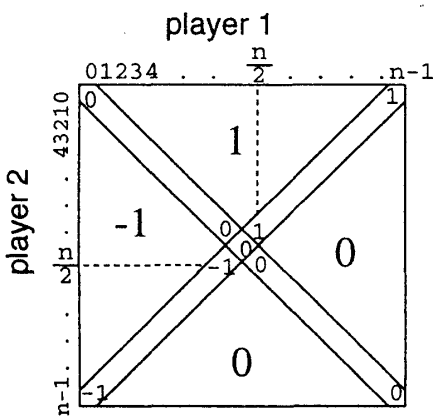

図 1: 利得行列 $A^{p_1} = -A^{p_2}$ の内容

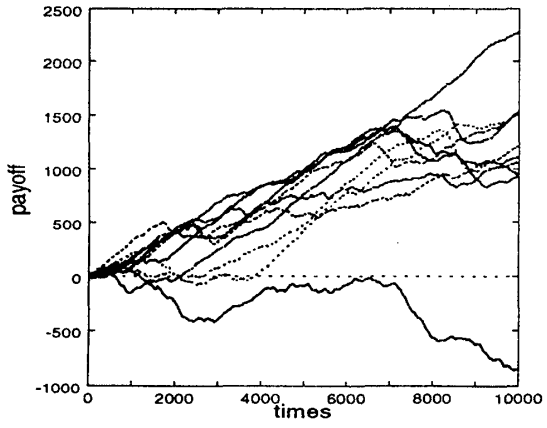

図 2: StGA の 利得の 累積値の推移

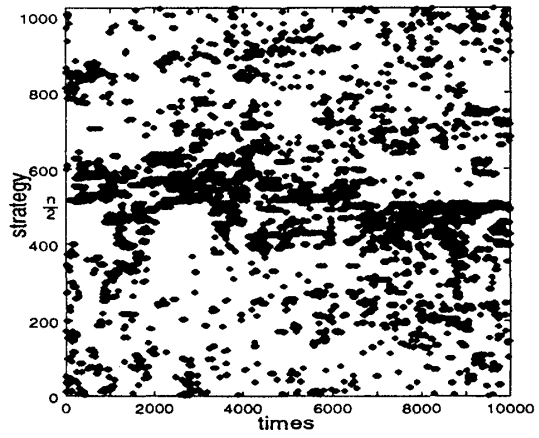

図 3: StGA の 戦略の推移

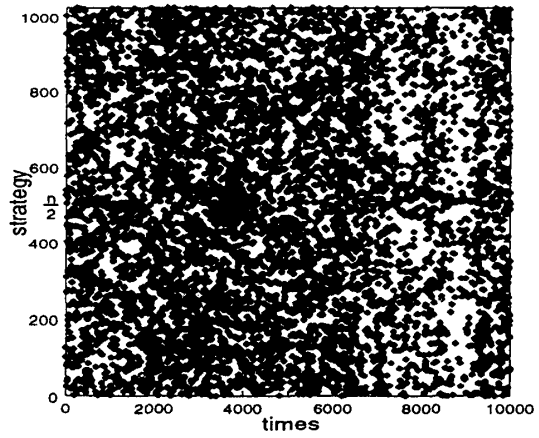

図 4: SLA の 戦略の推移

値の推移を図 2 に示す。連続 10000 回の対戦実験を 10 回行なった結果である。(利得行列を $A^{p_1} = -A^{p_2}$ としたために両者の利得値の和は零になり、グラフは利得零に関して対称になる。このため、SLA の利得の累積値の推移のグラフは省略した。) この図より、StGA が 10 回中 9 回はほぼ全域に渡って勝っていることがわかる。

図 3, 4 に StGA, SLA それぞれの戦略の推移を示す。図 3 より、StGA の場合は状態空間が圧縮されていることもあるが、比較的良いであろう $n/2$ に近い戦略を多く選択している様子が確認できる。一方、SLA の方は図 4 より、戦略が全域に散らばっており、学習がうまく行なわれていないことがわかる。

5 おわりに

本論文では、確率学習による適合度評価機構を有する遺伝的アルゴリズム StGA の応用として、ゲームにおける戦略獲得への適用を行なった。非常に多くの戦略を有するゲームの実験結果から、StGA は SLA よりも環境への適応能力が優れていることがわかった。

今後の課題としては、利得行列の内容が対戦の繰り返しの過程で次々と変化する、つまり動的に状態が変化する環境においても StGA による学習が有効であることを実験により確認することがあげられる。また、今回は比較的単純な利得行列を用いたが、具体的なゲームを想定した利得行列を用いた実験も行ないたいと考えている。

参考文献

[1] 棟朝雅晴, 高井昌彰, 佐藤義治: "確率学習による適合度評価機構を有する遺伝的アルゴリズム (1) – 基本モデル –", 情報処理学会第 49 回全国大会講演論文集 (1994).

[2] D. E. Goldberg: *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley (1989).

[3] K. S. Narendra and M. A. L. Thathachar: "Learning automata – a survey", *IEEE Transactions on System, Man, and Cybernetics* , Vol. 4, No. 4, pp.323-334 (1974).