

G 3 遺伝的アルゴリズムを用いた Poker Player 戦略の進化

小松 謙二郎† 元木 達也††
 †新潟大学大学院自然科学研究科 ††新潟大学工学部

1 はじめに

本研究では遺伝的アルゴリズム (GA) を用いて Texas Hold'em と呼ばれる Poker ゲームの戦略を追及することを目的とする。Poker は相手の持っている手札が分からない不確定な状況下で意思決定を行うゲームで、身近な知的ゲームである故にこれまでも様々な研究が行われてきている。N.V.Findler[1] はドローポーカーについて、人間の意思決定アプローチをモデル化する実験を行い、D.Billingsら[2] は Texas Hold'em についてワールドクラスの Poker Player を目指すための手札の評価方法や、敵のモデリングの方法を提案している。L.Barone&L.While[3][4][5] は Texas Hold'em について進化的な手法を用いて強い Poker Agent を探索している。本研究では L.Barone&L.While の進化構造を変えて、競合的進化を用いて戦略を進化させてみる。

2 Texas Hold'em

Texas Hold'em は Poker ゲームの一種で、自分の手を見てお金を賭け (bet という) 続けるかどうかを判断しながら Player 全員から出された賭金 (ポットと呼ばれる) を勝ち取るのを目指すゲームである。全 Player に共通な場のカードと自分のカードの組合せで Poker の役の強さを競う。ゲームには次の4つの進行状況があり、最終的にはゲームを降りずに残った Player 全員が自分のカード 2 枚と場のカード 5 枚で役の強さを競う。

- pre-flop : 手札が 2 枚 (場のカード無し)
- post-flop : 手札が 2 枚と場のカード 3 枚
- post-turn : 手札が 2 枚と場のカード 4 枚
- post-river : 手札が 2 枚と場のカード 5 枚

Texas Hold'em において各 Player は次の 3 つの行動を選択することができる。全員の賭金が同額になって親に順番が来るとゲームは次の段階へと進む。

- fold : ゲームを降りる。
- call : 前に賭けたプレイヤーの賭け金に金額を合わせる。既に金額が前のプレイヤーと同じ場合は check と呼ばれる。
- raise : 前のプレイヤーの賭け金を超えて bet をすること。ポットをつり上げる選択になる。

3 強い Poker Player の探索

3.1 L.Barone&L.While の手法

Poker Player の動作 L.Barone&L.While[4][5] は自分の手札から割り出した勝率 r を基に、fold、call、raise する割合を与える関数 fold()、call()、raise() を次の様に定義した。

$$\begin{cases} \text{fold}(r) = \exp(-b_1 * (r - a_1)) \\ \text{call}(r) = c_2 * \exp(-b_2^2 * (r - a_2)^2) \\ \text{raise}(r) = \exp(b_3 * (r + a_3 - 1)) \end{cases}$$

$$a_i \in [0, 1], b_i \in [0, \infty), c_i \in [0, \infty)$$

そして、これら 3 つの関数を用いて Poker Agent を次の様に確率的に動作させることにした。

$$\begin{cases} \text{foldする確率} = \frac{\text{fold}(r)}{\text{fold}(r) + \text{call}(r) + \text{raise}(r)} \\ \text{callする確率} = \frac{\text{call}(r)}{\text{fold}(r) + \text{call}(r) + \text{raise}(r)} \\ \text{raiseする確率} = \frac{\text{raise}(r)}{\text{fold}(r) + \text{call}(r) + \text{raise}(r)} \end{cases}$$

Poker Player の戦略の進化 実際の強い Poker Player の動作は自分の手の強さ、自分の座席の位置、賭けの状況等によって異なるので、関数 fold()、call()、raise() の最適設定 (i.e. 最適な係数 a_i 、 b_i 、 c_i の選び方) はこれらの環境情報に依存すると考えられる。そこで、L.Barone&L.While[4][5] は可能な状況を 12 個~48×(敵の数) 個に区分し、これらの状況毎に 7 個の係数 $a_1 \sim a_3$ 、 $b_1 \sim b_3$ 、 c_2 の組を、人工的に構成した Poker Player との対戦を通じて進化的に学習させた。

3.2 競合的進化

本研究においては、Poker Player の動作方式、個体表現については L.Barone&L.While[4][5] の手法に倣い、進化手法については独自に構成してみた。L.Barone&L.While との主な違いは次の 2 点である。

- 競合的進化 : position 毎に集団を用意して互いに競わせ、position 毎に 7 個の係数 $a_1 \sim a_3$ 、 $b_1 \sim b_3$ 、 c_2 の最適化を図る。
- 交叉あり : L.Barone&L.While の手法では mutation のみを用いていたが、ここでは交叉も用いる。

進化の詳細については次の通りである。

勝率の計算 D.Billingsら [2] に倣って現時点の自分の手札に対する勝率を計算する。すなわち、敵の全てのカードの組合せを考え、最後までプレイを続行した時の勝ちと負けと引き分けの数を数え上げて、ランダムな相手に対する勝率 r を求める。

個体表現 可能な状況を次の3つの観点で区別する。

- risk management... 勝率と現在のポットから計算される勝つ時の利益の期待値。その値を4つの区間に分ける
- game stage... ゲームの進行状況。pre-flop, post-flop, post-turn, post-river の4つに区別する。
- position... プレイヤーの座席の位置。(早い順番かどうか)

そして risk management と game stage に関して可能な $4 \times 4 = 16$ 個の状況の各々に対して関数 fold, call, raise の7個の係数 $a_1 \sim a_3, b_1 \sim b_3, c_2$ の値を並べたものを1つの個体とする。さらに、こういった個体の集団を position 毎に用意する。(Figure 1)

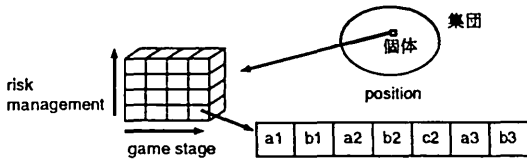


Figure 1: 個体表現

適応度の評価 それぞれの個体の善し悪しを計るために各 position の集団から出た個体同士で実際に Poker ゲームを行う。予め持ち点を与え、何回か繰り返し戦わせた後の持ち点をその個体の適応度とする。(Figure 2)

交叉・突然変異 交叉の方法としては一点交叉や多点交叉、状況の中の情報をまるごと交換する方法や状況の中の要素を交換する方法を用いる。突然変異の方法としては個体の要素それぞれにある確率でランダムな値に変える方法を用いる。また、扱う個体の要素が実数値ということで実数値 GA の考えも用いてみる。

4 今後の課題

まず、交叉確率や突然変異確率などの適切な数値を絞り込む。この実験ではある程度の計算時間がかかると思われ、1世代あたりの総当たり対戦にかかる時間を計測し、実験にかかる時間が適当

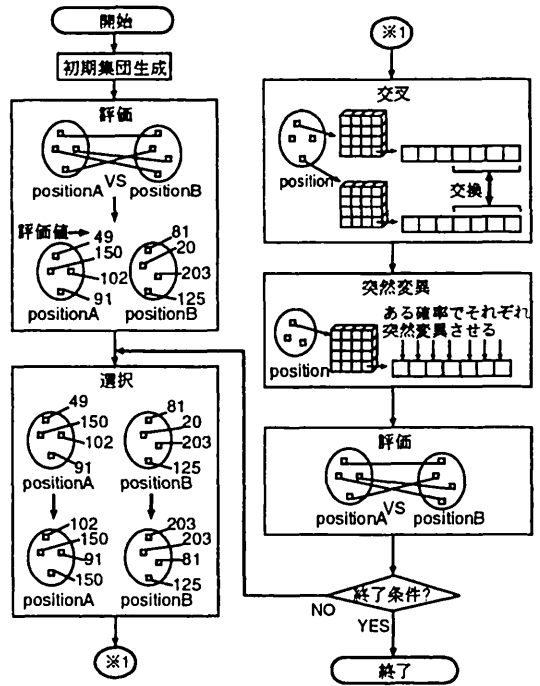


Figure 2: 競合的進化 (Position 数が2の場合)

と思われる範囲での設定や方法を模索しつつ、パラメータを変えて実験を行う。また、この実験では敵の戦略に対しての判断が何も行われぬ。そこで次に敵の行動の性格をモデリングする手法を取り入れた実験システムを構築し、敵の行動の性格をモデリングしない方法との比較実験も行いたい。

参考文献

- [1] Nicholas V. Findler, ポーカーをやるコンピューター, サイエンス, 1978 9月号 p.98-107 (Scientific American 日本語版 July 1978).
- [2] D.Billings, D.Papp, J.Schaeffer, and D.Szafron, Opponent Modelling in Poker, In "Fifteenth National Conference of the American Association for Artificial Intelligence", p.493-499, 1998.
- [3] L.Barone and L.While, Evolving adaptive play for simplified poker, in "the 1998 International Conference on Evolutionary Computation (ICEC '98)", p.108-113, IEEE Publications, 1998.
- [4] L.Barone and L.While, An Adaptive Learning Model for Simplified Poker Using Evolutionary Algorithms, In "the 1999 Congress on Evolutionary Computation (CEC '99)", p.153-160, IEEE Publications, 1999.
- [5] L.Barone and L.While, Adaptive Learning for Poker, In "Genetic and Evolutionary Computation Conference (GECCO-2000)", p.566-573, Morgan Kaufmann Publishers, 2000.