

Classifier System を用いた対戦によるゲームプレイルールの学習

梅原 聡 河口 信夫 外山 勝彦 稲垣 康善 (名古屋大学大学院工学研究科)

1 はじめに

ゲームの指し手の選択ルール(ゲームプレイルール)の学習に関する研究がなされている[堀之内 96]は、局面評価関数を遺伝的プログラミングを用いて生成することにより、ゲームプレイルールの学習する手法を提案した。しかし、この手法では、局面的事例ベースのみを用いて学習を行っているため、ゲームプレイルールの学習において重要な要因である対戦結果をその学習に反映させていない。本稿では、対戦結果に基づく学習を行うために if-then 型(条件-行動型)ルールの学習システムである classifier system を取り上げる。tic-tac-toe ゲームにおいて、パーフェクトプレイヤーとの対戦による学習実験を行い、対戦結果からゲームプレイルールの学習がすすめられることを確認した。

2 Classifier System による tic-tac-toe ゲームの学習

Classifier system は、if-then 型ルールの学習システムであり [Booker 89]、ルールの集合を遺伝的アルゴリズム (GA) を使って進化させることにより学習を行う。ここで、一つのルールは classifier という次元のストリングで表現され、また、有用であるかどうかを表す信頼値を付加する。

tic-tac-toe ゲームの一局面を図 1 に示す。O は先手、X は後手のコマで、B は空マスを表す。本稿で用いる classifier は、条件部に tic-tac-toe ゲームの局面、行動部に次の指し手を持つ。例えば、学習システム側が O の手番を受け持ったとき、図 1 のような局面で右下(マス 9)に O を打つルールは以下のように表せる。

$$XBOBXOBBB \Rightarrow 9$$

ただし、複数のルールをまとめて表現するため、条件部に記号 O, X, B の 3 種のうち、2 種ないしは 3 種すべてにマッチする記号を導入する。

以下に他プレイヤーとの対戦を通じた学習の手続きを示す。

1. (classifier の生成)

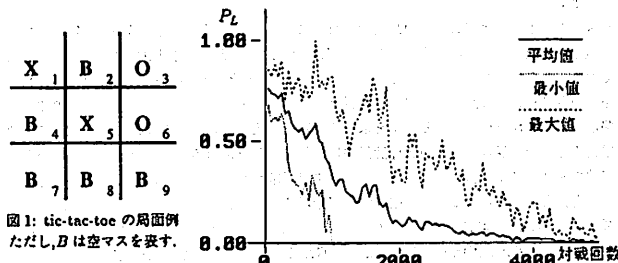
定められた数だけランダムに classifier を生成し、classifier の集合 (classifier list) を形成する。それぞれの classifier の信頼値は一定値とする。

2. (相手との対戦)

他プレイヤーとの対戦を行う。学習システムの手番では、classifier list の中から条件部にマッチするすべての classifier を挙げる。その中から信頼値に比例した確率で classifier を一つ選び、その行動部に示された手を打つ。この操作を勝負がつくまで繰り返す。

3. (信頼値の更新)

勝敗に応じ、いままでの指し手を選択したすべての classifier の信頼値を更新する。



4. (GA の介入)

一定回数の対戦をこなしたら、GA のプロセスを導入させ、信頼値の低い classifier を淘汰し、また、交叉や突然変異によって、新たな classifier list を形成する。

5. 2 に戻る。

3 学習実験及び考察

今回は、学習システム側には先手のみを担当させ、後手のパーフェクトプレイヤーとの対戦により学習をすすめた。学習システム側は、classifier の数を 3000、信頼値の初期値は 20 とし、信頼値の更新では、負けた際は信頼値を $1/2$ とし、引き分けの際には信頼値を $1/2$ したものに 30 を加えるものとした。また、対戦 50 回毎に GA のプロセスを導入させた。その際、最近 50 回の負け率 P_L に応じ、交叉率を $0.2P_L$ 、突然変異率を $0.05P_L$ に設定した。

図 2 に、乱数の値を変えて、10 回の学習実験をした結果を示す。なお、横軸は対戦回数、縦軸は 10 回のうちの P_L の最大値、最小値、平均値を示す。この図から P_L が 0 に収束していることが分かる。すなわち、対戦の結果を反映させながら学習をすすめることで、パーフェクトプレイヤーに対し、負けないルール集合が生成された。学習効率の向上や他ゲームへの応用は今後の課題である。

参考文献

- [Booker 89] L.B.Booker, D.E.Goldberg, J.H.Holland: Classifier Systems and Genetic Algorithms, *Artificial Intelligence* 40, pp.235-282 (1989).
[堀之内 96] 堀之内 剛史, 河口 信夫, 稲垣 康善: 遺伝的プログラミングを用いたゲームの局面評価関数の学習, 信学技報, KDBSE-96-35, pp.17-24 (1997).