

クラシファイヤシステムの引き込み(汎化)能力と 二倍体モデルの適用について

5E-3

丹羽 竜哉, 樋口 哲也, 古谷 立美
電子技術総合研究所

1 はじめに

近年, 新しい学習パラダイムとしてクラシファイヤシステムが注目されている[1]. これはプロダクションシステムをベースにしており, ルールベース自体が遺伝的アルゴリズムにより, 環境に合わせて適応, 進化していく所に特徴がある. これによりプロダクションシステムのネックである知識獲得の問題が回避されると共に, 学習能力についてもニューラルネットに匹敵する性能が得られることが医療データ分野で既に明らかになっている[4]. この学習能力に加え, 学習された知識がルールという人間に理解しやすい形で表現されることはニューラルネットにはない大きな利点である. ただし, ニューラルネットの特徴である引き込み(汎化)能力については, クラシファイヤシステムに関してはまだよく解っていない部分がある. このため, 本稿ではBOOLE[2]と呼ぶクラシファイヤシステムを例にとり, その引き込み能力について検証し, さらにその能力の強化法として染色体の減数分裂のモデル, すなわち二倍体モデルの適用を提案する.

2 クラシファイヤシステム: BOOLE

2.1 クラシファイヤシステム

クラシファイヤシステムとは, 遺伝的アルゴリズムを使ってプロダクションシステムのルールを自動的に学習していくシステムである. クラシファイヤシステムは三つの構成部分(ルールとメッセージシステム, 強さの割り当てシステム, 遺伝的アルゴリズム)よりなる. ルールとメッセージシステムは, プロダクションシステムと同様にルールに従ってメッセージを出力する. メッセージが競合する場合の出力は各々のルールが持つ強さに従う. 強さの割り当てシステムは, 各々のルールが持っている強さを, その正否に応じて増減する. 遺伝的アルゴリズムはルールに対して交叉や突然変異の操作を行なってルールを更新する.

2.2 BOOLE

BOOLE はブール関数を学習するクラシファイヤシステムである. ルールを構成している入力にマッチするビット列(ただし'0' '1'のいずれにもマッチするワイルドカード'#'を含む)および出力が遺伝的アルゴリズムの操作対象となる染色体に相当する. 文献[2]では例題として6-multiplexer問題が取り上げられている. これは, アドレス部2ビット, データ部4ビットの計6ビットを入力し, 2ビットのアドレス部で示されるデータ部の桁(ビット位置)のビット値を出力とする問題である.

3 BOOLE による引き込み

3.1 方法

表1の様な6ビットの入力を4つに分類する例題を作り学習させ, さらに入力に対して1ビットの誤りを加えたデータについてもっとも近い分類に引き込むかを実験した. 曖昧さを避けるため, 各々の入力間のハミング距離は, 全て4となるように例題を設定した.

表1 引き込みの学習パターン

入力	出力
000111	0 (00)
011001	1 (01)
101010	2 (10)
110100	3 (11)

3.2 結果

BOOLEのアルゴリズムをそのまま使用した場合, 1ビットの誤りを含むデータに対して約70%の引き込みが見られた. また, ルールの強さに対する評価方法を, 強さに比例する確率的方法から最も強いものを選ぶ方法に変更すると, 100%の引き込みを観察することができる. この様子を図1に示す. 縦軸は正答率, 横軸は学習回数で, 実線は確率的選択法(BOOLEそのまま), 破線は最強選択法を示す.

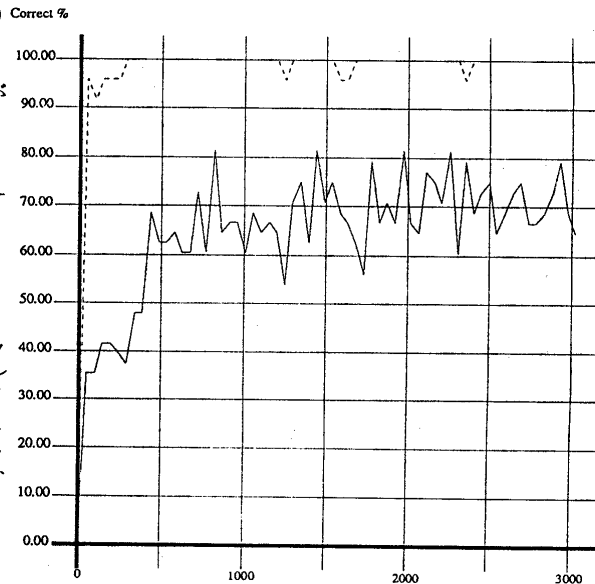


図1 BOOLE による引き込み

4 二倍体モデルの適用

4.1 BOOLE の問題点

BOOLE は集団の多様性を減少させる方向に進む性質を持っている。この性質はできたルールを少数にまとめる働きを示し、ルールをより解りやすいものにしていく。ところがこの性質のために、学習していないパターンにマッチするルールが減少していくことになる。つまり、多様性を減少させることで引き込みに有効なルールを失っている。逆に言えば、多様性を維持することによって引き込みを強くすることができると考えられる。

以上のように、引き込みを有効に機能させるためには、集団の中の多様性を保つことが重要であると考えられる。多様性を保つ方法として、二倍体モデルや地理的隔離等が有効であることが知られている [5]。

また、BOOLE のようなクラシファイヤシステムの場合、交叉によってルールが破壊されることはきわめて無駄な探索を行なうことになるものと考えられる。実際、交叉を行なわないでシミュレーションをしてみると、交叉を行なった場合と学習速度の差は見られない。つまり、交叉は学習能力の向上にほとんど寄与していないことがわかる。図 2 において、縦軸は正答率、横軸は学習回数、実線は交叉あり、破線は交叉なしを示す。

本稿ではこれらの問題に対して二倍体モデルの適用を提案する。

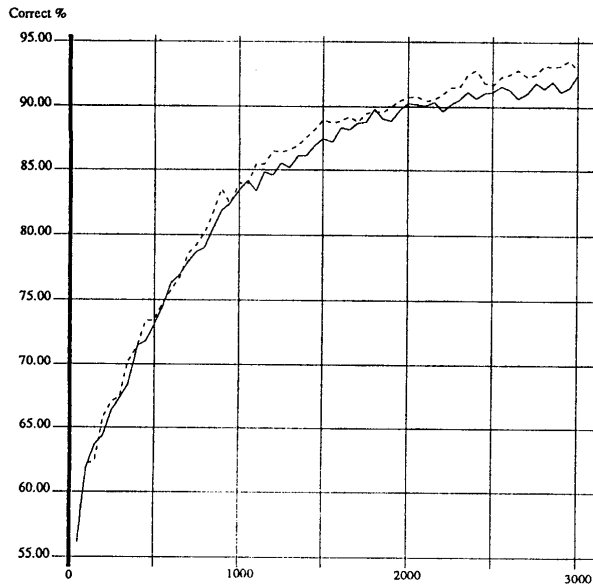


図 2 交叉の有無と学習速度

4.2 特徴

二倍体とは、一個体が両親からの染色体を各々一本ずつ、合計二本持つものである。交叉は、両親となる個体各々が持つ二本の染色体の間で行なわれる。後に示すように、ルールが破壊される可能性は別個体との間で交叉を行なう一倍体モデルに比べて小さくなると考えられる。

二倍体モデルの適用で多様性の維持による引き込み能力の強化の他、ルールの破壊の低減による学習の効率化が期待できる。

4.3 適用例

BOOLE に対する二倍体モデルの適用例を以下に示す。染色体は各々 0 または 1 の遺伝子を持つ。各々の桁において、二本の染色体の遺伝子が一致すれば一致した遺伝子を、一致しなければワイルドカード ' #' を表現形とする。ただし、ルールの右辺 (出力) は一致しなければならないものとする。

交叉前	遺伝子 1	1 0 0 1 1 1 0 : 1
	遺伝子 2	1 0 1 1 1 0 1 : 1
	表現形	1 0 # 1 # # : 1
交叉後	遺伝子 1'	1 0 0 1 0 1 : 1
	遺伝子 2'	1 0 1 1 1 0 : 1
	表現形'	1 0 # 1 # # : 1

図 3 二倍体モデルの例

図 3 に示される通り、交叉の前後で表現形に変化は起こらない。表現形が良性のもの (誤りの出力を出さない) であるならば、遺伝子 1、遺伝子 2、及びそれらの間に交叉を施した遺伝子 1'、遺伝子 2' の全てが良性のものとなる。すなわち、親が良性であれば子にもその形質が伝えられる可能性が、一倍体のものに比べて高いと言える。

5 まとめ

クラシファイヤシステム BOOLE の引き込みの機能について検証した。また、引き込みの機能や学習能力を強化する方法として、二倍体モデルの適用方法を提案した。現在、二倍体モデルについて性能評価中である。

謝辞

日頃貴重な御意見を下さる計算機研究室の皆様と、本研究の機会を与えて下さった弓場情報アーキテクチャ部部長をはじめとする多くの方々に、感謝の意を表します。

参考文献

- [1] Goldberg, D. E., (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley.
- [2] Wilson, S.W., (1987). Classifier systems and the Animal problem. *Machine Learning* 2, 199-228.
- [3] Bonelli, P., Parodi, A., Sen, S., Wilson, S., (1990). Newboole: A Fast GBML System. *In Proceedings of the International Conference on Machine Learning*, 153-159.
- [4] Bonelli, P., Parodi A., (1991). An Efficient Classifier System and its Experimental Comparison with two Representative learning methods on three medical domains. *ICGA 91*, 288-295.
- [5] 河内, 稻吉, 星野, (1992). 二倍体遺伝と地理的隔離を用いた遺伝的アルゴリズムによるニューラルネット構造の最適化. *人口知能学会誌*, Vol.7, 509-516.