

クラシファイヤシステムによる実数値データの学習に関する一考察

1 P-6

丹羽竜哉, 伊庭斉志, 樋口哲也

電子技術総合研究所

1. はじめに

クラシファイヤシステム (以下 CS) Boole [1] は, 論理関数を学習する手法として高い学習能力を持っている. 我々は, ノイズのある環境においても十分な学習能力を有していることを既に確認している [5]. しかし実際の応用面においては, 種々のデータが論理値でなく実数値データとして与えられることが多い.

この実数値データの学習については, 評価用データによるシミュレーションのみならず実際の応用面においてもニューラルネットが高い性能を示すことが, 従来より知られている.

本稿では, Boole を論理関数から実数値関数への拡張方法を提案し, 評価用データを用いたシミュレーションによってその性能を評価する.

2. 学習アルゴリズム

CS とは, プロダクションシステムの *if... then...* ルールを遺伝的アルゴリズムによって更新し, 適応させて学習していく方法である. 本稿で使用した学習アルゴリズムは, CS の一つである Boole の改良型の NewBoole [2] をもとに, 実数値型のデータに対応させるため [3] を参考にして変更を加えたものである.

Boole では, 各々のルールは入力ビット列にマッチするビット列であったが, 我々の実験ではこれを実数値データに対応させるため, 最大値と最小値を示す実数値の列とした. すなわち,

1. 入力データの条件部の各々の要素の全ての値がルールの条件部に示された値の範囲内にあることをもってマッチする (図 1 a). 一つでも範囲内にならない場合はマッチしないものとした.
2. 突然変異は, ルールの条件部の実数値の範囲をランダムに変更するものとし, 交叉は [1] と同じく 12% で生起する 1 点交叉とした. 図 1 b において, 縦線は交叉の点を示し, 下線は突然変異の発生を示す. (Boole では交叉によって子は 1 つしかできない.)
3. 過学習に関して検証するため, 集団サイズは様々に変えて実験を行なった.

4. 各々のルールの強さの割当のシステム等その他のアルゴリズムは, 基本的に NewBoole [2] と同じものとした.

ルール	条件部			出力部
	最大値	0.4	0.8 0.7	
	最小値	0.1	0.5 0.4	1
サンプル	(マッチ)	0.3	0.4 0.5	1

図 1 a : ルールとサンプルの例

親ルール	条件部			出力部
	最大値	0.4	0.8 0.7	
親ルール 1	最小値	0.1	0.5 0.4	1
親ルール 2	最大値	0.3	0.5 0.9	1
	最小値	0.2	0.4 0.1	
		↓		
子ルール	最大値	0.4	0.5 0.9	1
	最小値	0.1	0.4 0.2	

図 1 b : 交叉と突然変異

3. 学習能力の評価

学習能力の評価は, ニューラルネット等の評価でしばしば使われる IRIS (あやめ) のデータを用いて行なった. このデータは, 4次元の実数値を入力として3種類に分類する評価用データで, 3種類各々について 50 個ずつ合計 150 個のサンプルが与えられている. 我々は, 3種類の中から線形分離が可能でない組合せ一組 (2種類, 100 サンプル) を選んで, 学習 / 判別の実験を行なった.

IRIS のデータは 4次元各々の値域が異なっており, このままでは学習のシミュレーションが困難であるため, あらかじめ最小値-最大値が [0.1] になる様に正規化を行なった.

実験は, 100 サンプルのデータを学習サンプルと試験サンプルの 2 つに分け, 学習サンプルだけを学習させた後に試験サンプルを正しく分類する率によって学習能力を評価した.

学習は, 5000 世代で終了させた.

4. 結果および考察

集団サイズ 70, 学習サンプル数 80, 試験サンプル数 20 とした時の世代毎の学習の様子を図 2 に示す。横軸は世代数, 縦軸は正答率 (%) とし, グラフ中の実線は学習サンプル, 破線は試験サンプルに対する正答率を示す。学習サンプルは急速に (~1000 世代) 学習されてしまうが, 試験サンプルに対する正答率は数千世代かかって次第に上昇している。

CS の学習能力は内部のルール数である集団サイズに左右されるが, 集団サイズの過大による学習能力の低下 (過学習) の現象が見られないという特徴が観察されている [5]。過学習は, ニューラルネットでは問題となっているが, この現象が見られないという特徴は CS の大きな長所と考えられる。学習サンプルの数が一定 (70 サンプル) とし集団サイズを様々に変化させた場合について, 図 3 に示す。横軸は学習サンプル数に対する集団サイズの比, 縦軸は正答率 (%) で, 図中の縦線は各々の分散を示す。実数値の学習の場合も, 過学習の傾向は観察されない。

5. おわりに

今回の実験は NewBoole を単純に拡張したものであるためデータを 2 種類に分類する学習であったが, 現在 3 種類を同時に分類する CS について研究中である。更に複雑な評価用データに対する実験も試みている。また, 二倍体モデル等の工夫 [4] による能力の向上についても検討中である。

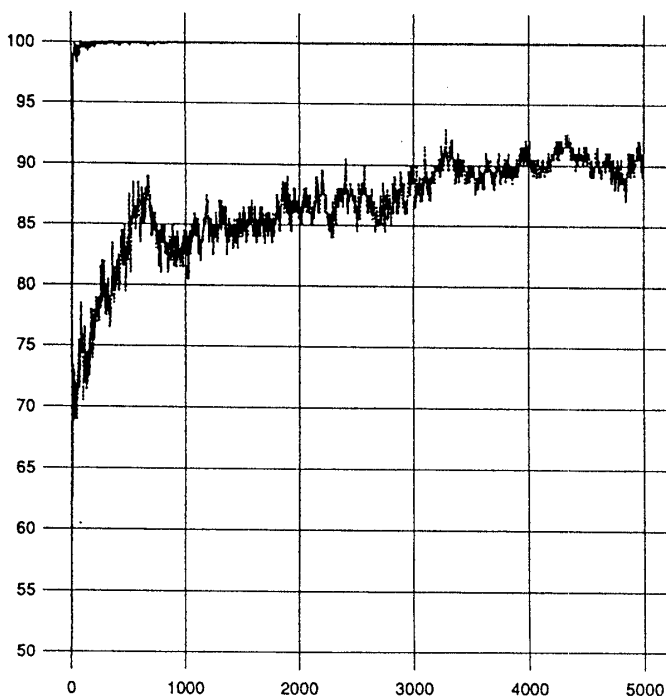


図 2 : 世代毎の学習の様子

謝辞

本研究に関し様々な議論や資料を下された電総研の栗田氏や RWC の古谷室長他の方々, また本研究の機会を与えて下さった太田部長他の方々に感謝の意を表す。

参考文献

- [1] S.W.Wilson, "Classifier Systems and the Animal Problem", *Machine Learning*, vol.2, no.3, 1987.
- [2] P.Bonelli, and A.Parodi, "An Efficient Classifier System and its Experimental Comparison with two Representative learning methods on three medical domains", *Proc. of 4th ICGA*, 288-295, 1991.
- [3] J.J.Grefenstette, C.L.Ramsey, and A.C.Schultz, "Learning Sequential Decision Rules Using Simulation Models and Competition", *Machine Learning*, vol.5, no.4, 1990.
- [4] 丹羽, 樋口, 古谷, "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", 情報処理学会第 45 回全国大会 5E-03, 1992.
- [5] 丹羽, 伊庭, 佐藤, 樋口, "ブール関数の進化論的手法による学習について", 電気学会情報処理研究会論文集 IP-93-14, 1993.

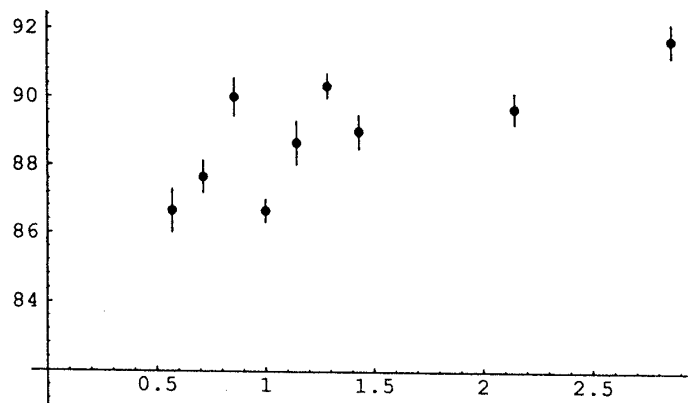


図 3 : 集団サイズに対する学習能力の変化