進化型多目的最適化手法を用いたファジィルール選択*

山本 隆[†]・石渕 久生[†]

Fuzzy Rule Selection Using Evolutionary Multiobjective Optimization Methods*

Takashi YAMAMOTO[†] and Hisao ISHIBUCHI[†]

One advantage of evolutionary multiobjective optimization (EMO) algorithms over classical approaches is that many non-dominated solutions can be simultaneously obtained by their single run. This paper shows how this advantage can be utilized in genetic rule selection for the design of fuzzy rule-based classification systems. Our genetic rule selection is a two-stage approach. In the first stage, a pre-specified number of candidate rules are extracted from numerical data using a data mining technique. In the second stage, an EMO algorithm is used for finding non-dominated rule sets with respect to three objectives. Since one of the three objectives is to maximize a classification rate on training patterns, the evolution of rule sets tends to overfit to training patterns. The question is whether the other two objectives with respect to complexity work as a safeguard against the overfitting. In this paper, we examine the effect of the three-objective formulation on the generalization ability of obtained rule sets through computational experiments where many non-dominated rule sets are generated using an EMO algorithm for a number of high-dimensional pattern classification problems. Finally, we demonstrate that an ensemble of generated fuzzy rule-based systems leads to high generalization ability.

1. はじめに

一般に、ファジィ理論を用いた if-then 形式のルールはファジィルールとよばれ、前件部と後件部をファジィ命題により構成する言語的表現が可能なファジィルールや、モデルの推定などで用いられる後件部を回帰モデルで構成するファジィルールがある。これらのファジィルールにより構成されるファジィシステムは、ニューラルネットワークと同様に、高い近似能力を持つ非線形関数の近似器であることが知られている。そのため、ファジィシステムは、制御、関数近似、パターン識別など、様々な分野で適用されている [22]。ファジィシステムの設計における主な目的は性能の最大化であるが、最近の研究ではファジィシステムの解釈性が考慮されるようになってきた [2,3,24]。また、ファジィシステムやニューラルネッ

トワークなどの学習システムを用いて識別器の設計を行う場合、一般に学習用データに対する性能の最大化が行われる。しかし、学習用データに対する性能の最大化は、必ずしも未知のデータに対する識別器の汎化能力の最大化を意味しない。すなわち、学習用データに対する性能の最大化は、学習用データへの過剰適合により識別器の性能低下を引き起こす可能性がある。このような過剰適合を防止するために、識別器の複雑性と学習用データに対する性能を同時に考慮した識別器の設計方法が提案されている [4]。しかし、多くの研究では、性能と複雑性という異なる評価基準を同時に考慮してはいるが、最終的には加重和最大化手法などを用いることにより、ファジィシステムの設計問題を単一目的最適化問題へ帰着させている。多目的最適化の枠組みでファジィシステムの設計を議論した最初の研究として、2 目的ルール選択 [10] が挙げられる。目的関数をさらに加えることにより、3 目的最適化問題への拡張も行われている [13]。

本研究では、文献 [13] と同様に、ファジィ識別システムの設計の際に三つの目的（識別精度、ファジィルールの数、ファジィルールの長さの総和）を考慮する。ここで、ファジィルールの長さとはルールの条件部に含まれ

* 原稿受付 2003 年 7 月 15 日

[†] 大阪府立大学 大学院 工学研究科 Department of Industrial Engineering, Osaka Prefecture University; 1-1 Gakuencho, Sakai city, Osaka 599-8531, JAPAN

Key Words: rule selection, evolutionary multiobjective optimization methods, fuzzy rule-based systems, association rules, classification.

る条件の数である。通常、人間が大量のルールを一度に理解することは困難である。したがって、ファジイルールの数は解釈性に密接に関係している。しかし、少数のルールから構成されるファジイ識別システムは、いつも解釈性に優れているとは限らない。なぜなら、多くの条件を持つ長いルールを直感的に理解することができないからである。したがって、ルールの長さもまた解釈性と密接に関係している。進化型多目的最適化手法を用いて、これらの三つの目的に関する非劣解集合の探索を行う。

通常、個々の問題に対してファジイルールを生成するための適切なファジイ分割は与えられていない。また、一つのファジイ識別システムにおいて、一般的なルールは粗いファジイ分割、特殊なルールは細かなファジイ分割を必要とするかもしれない。このような粒度の異なるファジイ分割が同時に存在するという状況を扱うために、各ファジイルールの条件部ファジイ集合を、複数のファジイ分割のなかから選択することによって決定する。本研究では、Fig. 1 で示す粒度の異なる四つのファジイ分割を用いる。少数の条件部を持つ短いファジイルールを生成するために、Fig. 1 の14個のファジイ集合に加えて、Fig. 2 で示す“don't care”を用いる。したがって、各属性に対する条件部ファジイ集合は15個のファジイ集合のなかから選択される。このとき、その組合せの総数は、 n 次元パターン識別問題に対しては 15^n 通りになる。

遺伝的アルゴリズムに基づいたルール選択 [13] では、まず、大量の候補ルールが条件部ファジイ集合の組合せから生成される。つぎに、3目的に関する非劣ルール集合を見つけるために、生成された候補ルールの部分集合が遺伝的アルゴリズムを用いて調べられる。しかし、すべての属性に対して単一のファジイ分割（例えば、Fig. 1

Fig. 1 Four fuzzy partitions

Fig. 2 Membership function of “don't care”

右下の5分割)だけが考慮されている。この場合、“don't care”を含む条件部ファジイ集合の組合せの総数は、 n 次元問題に対して $(5+1)^n$ 通りとなり、本研究で考慮する 15^n よりもずっと少ない。候補ルールの数が増加するにつれて、非劣ルール集合を見つけるための探索空間の大きさは指数関数的に増加し、遺伝的アルゴリズムの効率は急激に悪化する [16]。したがって、候補ルールの総数を減らす必要がある。そこで本研究では、ファジイ関連ルールに対するルール評価基準を用いて候補ルールの選択を行う。ファジイ関連ルールとは、データマイニングの分野で用いられる関連ルール [1] をファジイルールに拡張したものである。候補ルールの選択は、各クラスに対して行われ、事前に設定した数のファジイルールが選択される。つぎに、選択されたファジイルールに対して進化型多目的最適化手法を適用し、様々なファジイ識別システムの獲得を行う。

2. ファジイ識別システム

2.1 ファジイルール選択問題

m 個の学習用パターン $\mathbf{x}_p = (x_{p1}, \dots, x_{pn})$, $p = 1, 2, \dots, m$ が n 次元実数空間内に与えられている M クラス識別問題を考える。ここで、 x_{pi} は p 番目の学習用パターンに対する i 番目の属性値である。一般性を失うことなく、入力区間が n 次元超立方体 $[0,1]^n$ であると仮定することができる。つまり、すべての属性値が区間 $[0,1]$ の実数値であると仮定する。 n 次元 M クラス識別問題に対して、以下の形式のファジイルールを用いる。

$$\text{Rule } R_q: \text{If } x_1 \text{ is } A_{q1} \text{ and } \dots \text{ and } x_n \text{ is } A_{qn} \text{ then Class is } C_q \text{ with } CF_q \quad (1)$$

ここで、 R_q は q 番目のファジイルールを示すルールラベル、 A_{qi} は i 番目の属性に対する条件部ファジイ集合、 C_q は結論部クラス、 CF_q はルールの重みである。条件部ファジイ集合 A_{qi} は、Fig. 1 と Fig. 2 に示す15個のファジイ集合のなかの一つである。ルールの重み CF_q は、 $[0,1]$ 区間内の実数値である。なお、重みを用いないファジイルールや結論部クラスが複数存在するファジイルールの使用も議論されている [5,21]。

本研究の目的は、高い識別能力を持つ理解しやすいファジイ識別システムを設計することである。つまり、高い識別能力を持つ少数のわかりやすいファジイルールを見つけることである。そのようなファジイ識別システムを設計するために、遺伝的ルール選択が提案されている [15]。文献 [13] では、ファジイ識別システムの設計を3目的最適化問題として取り扱っている。本研究でも同様に、ルール選択問題を以下の3目的最適化問題として取り扱う。

$$\text{Maximize } f_1(S), \text{ minimize } f_2(S), f_3(S) \quad (2)$$

ここで、 S は選択されたルール集合、 $f_1(S)$ はルール集合

S によって正しく識別された学習用パターン数, $f_2(S)$ は S に含まれるファジールール数, $f_3(S)$ は S に含まれるファジールールの長さの総和である.

通常, 三つの目的すべてに関する最適なルール集合は存在しない. したがって, 3目的最適化問題を解くことは, 他のルール集合によって優越されていない複数のルール集合を見つけることになる. あるルール集合 S が他のどのルール集合にも優越されていないとき, S は(2)式の3目的最適化問題のパレート最適解であるといわれる. 我々の3目的最適化問題の探索空間は 15^n 個のファジールールのべき集合であるが, この空間はあまりに広すぎて厳密なパレート最適解を見つけることは不可能に近い. したがって, 我々は近似パレート最適解を見つけることにする. 本研究では, 候補ルールを選択することによって探索空間を縮小し, 進化的多目的最適化手法を用いて縮小された空間内で近似パレート最適解の探索を行う.

2.2 ファジィ相関ルール

n 次元問題に対して, 条件部ファジィ集合の組合せは 15^n 通りである. n が小さい場合には, 条件部ファジィ集合のすべての組合せを調べることが可能であり, 生成されたすべてのルールを候補ルールとして用いることができる. しかしながら, n が大きい場合 (たとえば, ソナーデータでは $n=60$), すべての組合せを調べるには膨大な時間がかかり, 生成されたすべてのルールを候補ルールとして用いることは不可能である. そこで, 我々はデータマイニングにおけるルール評価基準を用いた候補ルールの選択を提案する.

データマイニングの分野では, 信頼度 (confidence) と支持度 (support) とよばれる基準が相関ルール [1] を評価するために用いられている. (1) 式のファジールールは, $A_q \Rightarrow C_q$ という相関ルールの一種とみなすことができ, このルールをファジィ相関ルールとよぶ. 以下では, これら二つの基準をファジィ相関ルール $A_q \Rightarrow C_q$ に対する基準へ拡張する方法を述べる [19]. 同様のファジィ相関ルールへの拡張が, 文献 [9] で提案されている.

与えられた m 個の学習用パターン $x_p = (x_{p1}, \dots, x_{pn})$, $p = 1, 2, \dots, m$ の集合を D とする. データ集合 D は m 個のパターンから構成されているので, D の濃度は $|D|$ によって表現され, $|D|=m$ となる. 相関ルール $A_q \Rightarrow C_q$ の信頼度は以下のように定義される [1].

$$c(A_q \Rightarrow C_q) = \frac{|D(A_q) \cap D(C_q)|}{|D(A_q)|} \quad (3)$$

ここで, $|D(A_q)|$ は条件部 A_q に適合しているパターン数であり, $|D(A_q) \cap D(C_q)|$ は条件部 A_q と結論部 C_q の両方に適合しているパターン数である. 信頼度 c は相関ルール, $A_q \Rightarrow C_q$ の妥当性を示している. つまり, 条件部 A_q と適合しているパターンの $c \times 100\%$ が, 結論部 C_q とも適合していることを表わしている. 通常の

相関ルールでは, A_q と C_q はファジィ集合ではないので, $|D(A_q)|$ と $|D(A_q) \cap D(C_q)|$ の計算は適合しているパターン数を数えるだけでよい. しかし, $A_q \Rightarrow C_q$ がファジィ相関ルールの場合, 条件部 A_q に対して各パターンは異なる適合度 $\mu_{A_q}(x_p)$ を持つ. 結論部はファジィ集合ではないので, 条件部の適合度を考慮することで, (3) 式は以下のように書き変えることができる [19].

$$c(A_q \Rightarrow C_q) = \frac{\sum_{p \in \text{Class } C_q} \mu_{A_q}(x_p)}{\sum_{p=1}^m \mu_{A_q}(x_p)} \quad (4)$$

本研究では, 学習用パターン x_p と条件部 $A_q = (A_{q1}, \dots, A_{qn})$ との適合度 $\mu_{A_q}(x_p)$ を積演算によって以下のように定義する.

$$\mu_{A_q}(x_p) = \mu_{A_{q1}}(x_{p1}) \times \dots \times \mu_{A_{qn}}(x_{pn}) \quad (5)$$

ここで, $\mu_{A_{qi}}(x_{pi})$ は条件部ファジィ集合 A_{qi} のメンバシップ関数である (Fig. 1 および Fig. 2 の15個のファジィ集合のなかの一つ).

同様に, 相関ルール $A_q \Rightarrow C_q$ の支持度は以下のように定義される [1].

$$s(A_q \Rightarrow C_q) = \frac{|D(A_q) \cap D(C_q)|}{|D|} \quad (6)$$

支持度 s は, 相関ルール $A_q \Rightarrow C_q$ の適用範囲を示している. つまり, 全パターンの $s \times 100\%$ が, 相関ルール $A_q \Rightarrow C_q$ と適合している (条件部 A_q と結論部 C_q の両方に適合している) ことを表わしている. (4) 式と同様に, (6) 式は以下のように書き変えることができる [19].

$$s(A_q \Rightarrow C_q) = \frac{\sum_{p \in \text{Class } C_q} \mu_{A_q}(x_p)}{m} \quad (7)$$

2.3 結論部クラスとルールの重み

条件部 A_q を持つファジールール R_q の結論部クラス C_q は, 以下のように決定される.

$$c(A_q \Rightarrow C_q) = \max_h \{c(A_q \Rightarrow \text{Class } h)\} \quad (8)$$

つまり, 結論部クラスは, 最大の信頼度を持つクラスに決定される. 信頼度の代わりに支持度を用いた場合も, 条件部 A_q に対して同じ結論部クラスが得られる. 通常, 条件部ファジィ集合の組合せに対して結論部クラスを一意に決定することができるが, すべてのパターンが条件部と適合しない場合のように複数のクラスが同じ最大値を持つ可能性がある. このような状況では, A_q に対して結論部クラス C_q を決定することができず, 対応するファジールール R_q を生成しない.

生成されたファジールールを用いて未知パターンを識

別するために、本研究では、単一勝利ルールに基づくファジィ推論法 [14] を用いる。なお、文献 [5,12,21] では、他のファジィ推論法についても議論されている。\$S\$ をファジィ識別システムに含まれるファジイルールの集合とする。ある入力パターン \$\mathbf{x}_p\$ に対して、勝利ルールは以下のように選択される。

$$\mu_{Aw}(\mathbf{x}_p) \cdot CF_w = \max_{R_q \in S} \{ \mu_{Aq}(\mathbf{x}_p) \cdot CF_q \} \quad (9)$$

勝利ルールは入力パターンとの適合度とルール重みの積が最大のルールであり、勝利ルールの結論部クラスが入力パターンの識別結果として出力される。したがって、条件部ファジィ集合の修正を必要とせずに、ファジイルールのルール重みだけを調整することで高い識別精度を得ることができる [11]。なお、勝利ルールが複数存在し、それらのルールの結論部クラスが異なる場合、入力パターンの識別は拒否される。

文献 [5] では、ファジイルールの重みとして信頼度を用いられている。一方、単一勝利ルールに基づく推論方法を用いる場合、信頼度を重みとして用いるよりも以下の定義から良い結果が得られることが示されている [17]。

$$CF_q = c(\mathbf{A}_q \Rightarrow C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^M c(\mathbf{A}_q \Rightarrow \text{Class } h) \quad (10)$$

3. ファジイルールの選択

3.1 候補ルールの選択

データマイニングの分野では、信頼度と支持度が相関ルールの選択基準として用いられている。文献 [18] では、ファジイルールの選択基準として7種類の基準が用いられ、それぞれの基準を用いて設計されたファジィ識別システムの性能の比較が行われている。本研究では、それらの基準のなかから、定義が簡単でありながら良好な性能を示した以下の選択基準を用いることにする。

$$f(R_q) \Rightarrow s(\mathbf{A}_q \Rightarrow C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^M s(\mathbf{A}_q \Rightarrow \text{Class } h) \quad (11)$$

これは、SLAVE[3,8] とよばれる繰り返しファジィ遺伝的機械学習アルゴリズムにおいて用いられている基準を変形したものである。

生成されたすべてのファジイルールから、(11) 式のルール選択基準を用いて候補ルールを選択する。しかし、複数のファジイルールが等しい基準値を持つ可能性がある。そのような場合、次の二つの基準を用いて順位付けを行う。一つ目は、ファジイルールの条件部の数である。条件部に含まれる条件の数が少ないルール、すなわち、短いルールが選択される。この基準でルールの順位付けができなかった場合は、二つ目の基準としてファジイルールの条件部ファジィ集合が覆う面積を用いる。この面積は、それぞれの条件部ファジィ集合の三角型メンバ

シップ関数の面積を足すことで計算される。大きな面積を持つファジイルール、すなわち、より大きい空間を覆うルールが選択される。二つ目の基準を用いてもルールの順位付けができなかった場合は、ランダムにルールを選択する。本研究では、各クラスから \$N/M\$ 個のルールを選択し、全体として \$N\$ 個のファジイルールを選択する。しかし、\$n\$ 次元パターン識別問題に対して、条件部ファジィ集合の組合せは \$15^n\$ 通りである。そのため、\$n\$ が大きい場合、すべての組合せを調べて \$N\$ 個の候補ルールを選択するのは非常に時間がかかり現実的ではない。また、理解しやすいファジィ識別システムを設計するという点から、長いファジイルールは候補ルールとして適切ではない。したがって、短いファジイルールだけが候補ルールとして調べられる。このルールの長さにおける制約は、3目的ルール選択問題の3番目の目的と一致している。長さ \$L\$ のファジイルールの数は、\$nC_L \times 14^L\$ である。\$n\$ が大きい場合でも、小さな \$L\$ の値に対しては、\$nC_L \times 14^L\$ は大きくはならない。つまり、生成可能なファジイルールの総数が膨大になっても、短いファジイルールの数は大きくならないことを意味している。

なお、すべてのファジイルールを探索した場合においても、ルール選択基準により獲得される候補ルールの長さは短いものになる。たとえば、ワインデータに対しては、最も長い候補ルールの長さが3となった。

3.2 GA によるルール選択

\$N\$ 個のファジイルールが候補ルールとして選択されたとする。\$N\$ 個の候補ルールの部分集合 \$S\$ は、進化型多目的最適化 (EMO) アルゴリズムにおいて個体として扱われ、長さ \$N\$ のバイナリー記号列として表現される。

$$S = s_1 s_2 \cdots s_j \cdots s_N \quad (12)$$

ここで、\$s_j = 1\$ は \$j\$ 番目の候補ルールがルール集合 \$S\$ に含まれることを表わし、\$s_j = 0\$ は \$S\$ に含まれないことを表わしている。

近年、高い探索能力を持つ様々な進化型多目的最適化アルゴリズム [6,20,27] が提案されているが、従来の研究 [10,13,16] では、加重和を用いたスカラー適応度関数に基づく多目的遺伝アルゴリズムが用いられていた。しかし、ルール選択問題においては、候補ルールの部分集合はバイナリー記号列で表現されるため、任意の EMO アルゴリズムの適用が可能である。本研究では、EMO アルゴリズムとして NSGA-II を用いることにする。これは、文献 [6] において探索能力の高さが十分に議論されており、その実装も比較的容易なためである。数値実験では、本来の NSGA-II の遺伝的操作に加えて、問題特有の二つのヒューリスティックを組み込む。一つは、バイアスのある突然変異操作である。この操作では、0 から 1 への突然変異確率と 1 から 0 への確率が異なっている。具体的には、ルールを減らす方向に大きな確率が割り当

Table 1 Data sets used in our computer simulations

Data sets	Number of attributes	Number of patterns	Number of classes	C4.5 error rates [%][7]	
				Best	Worst
Wisconsin Breast Cancer Database	9	683*	2	5.1	6.0
Pima Indians Diabetes Database	8	768	2	25.0	27.2
Glass Identification Database	9	214	6	27.3	32.2
Cleveland Heart Disease Database	13	297*	5	46.3	47.9
Sonar Database	60	208	2	24.6	35.8
Wine Recognition Data	13	178	3	5.6	8.8

てられており、ルール集合に含まれるルールを効率的に減らすことができる。もう一つは、不必要なルールの削除である。入力パターンを識別する際に、単一勝利ルールに基づく推論法を用いているので、ルール集合 S 内のいくつかのルールは、どのパターンに対しても勝利ルールにならない可能性がある。この場合、識別性能（1番目の目的）に影響を与えることなく、それらのルールを削除することで、ルール数（2番目の目的）とルールの長さの総和（3番目の目的）を減らすことができる。したがって、どのパターンに対しても勝利ルールになっていないすべてのルールを削除する。

4. 数値実験

数値実験では、UCI Machine Learning Repository (<http://www.ics.uci.edu/~mllearn/>) から利用可能なデータ集合を用いた。本研究では、多数のデータ集合のうち Table 1 に示す 6 種類のデータ集合に対する数値実験結果を報告する。Table 1 では、欠測値を含むデータ集合は * で示されており、欠測値を含むパターンを除いたパターン数が示されている。文献 [7] では、UCI MLR (Machine Learning Repository) に含まれているデータ集合を用いて、C4.5 アルゴリズム [23] の 6 種類の変種の性能評価が行われている。各アルゴリズムの性能は、10-fold cross-validation(10CV) 手法 [26] の 10 回の繰り返しによって評価されている。Table 1 では、各データ集合に対して、6 種類の変種のなかで最良の結果と最悪の結果を引用している。なお、数値実験では、ソナーデータに対しては長さが 2 以下、その他のデータに対しては長さが 3 以下のファジィルールだけを候補ルールとして調べた。

10CV 手法は、評価用データに対する識別性能を評価する手法であり、機械学習の分野でしばしば用いられている。10CV 手法は、まずデータ集合を大きさの等しい 10 個の部分集合に分割する。そのなかで 9 個の部分集合を学習用データとして使い、残り 1 個の部分集合を評価用データとして用いて識別性能を評価する。10 個の部分集合のそれぞれが評価用データとして用いられるように、学習・評価の手順を 10 回繰り返す。10 回の実行における識別性能を平均したものが、1 回の 10CV 操作の

結果となる。10 個の部分集合への分割はランダムに行われるが、10CV 手法の結果はそのランダムな分割に依存する。そのため、一般に 10CV は複数回繰り返され、その平均性能が汎化性能として用いられる。

4.1 EMO 手法によるファジィルール選択

まず、(11) 式のルール選択基準を用いて、各クラスから 300 個のルールを候補ルールとして選び出した。したがって、 M クラス問題において、 $300M$ 個のファジィルールが選択される。つぎに、選択された $300M$ 個の候補ルールに対して NSGA-II を適用し、非劣解集合の探索を行った。数値実験では、NSGA-II のパラメータを以下のように設定した。

個体群サイズ: 200 個体

交叉確率: 0.8

突然変異確率: $p_m(0 \rightarrow 1) = 1/300M$
 $p_m(1 \rightarrow 0) = 0.1$

終了条件: 5000 世代

獲得されたルール集合の汎化能力を調べるために、文献 [7] と同様に、10CV 手法の 10 回繰り返しを用いた。個々の実行において、我々の 2 段階ルール選択手法は学習用パターンに対して適用され、獲得された非劣ルール集合の性能が評価用パターンによって評価された。したがって、ルール選択手法が各データ集合（テスト問題）に対して、100 回適用された。各データ集合に対して、NSGA-II の 1 回実行で複数の非劣ルール集合が得られる。そこで、数値実験結果をまとめるために、100 回の実行から得られた多数のルール集合の性能を、ルール数とルール長が同じもので平均した。

Wisconsin Breast Cancer Database

乳癌データに対する実験結果を Table 2 に示す。ここで、平均ルール長とはファジィルール一つ当たりのルールの長さの平均値を表わしており、ルールの長さの総和は、ルール数と平均ルール長の積に等しい。最後の列は、該当するルール集合が、100 回の実行のうちで何回獲得されたかを示している。ルール数と平均ルール長の様々な組合せのルール集合が得られたが、30 回未満しか獲得

Table 2 Obtained rule sets for the breast cancer data

Number of rules	Average rule length	Error rates [%]		Runs
		Train	Test	
0	0.00	100.00	100.00	100
1	1.00	34.99	35.21	100
2	1.00	5.25	6.03	100
2	1.50	3.34	3.47	100
2	2.00	3.15	3.79	92
3	1.33	2.85	4.21	80
3	1.67	2.64	4.25	92
4	1.50	2.42	4.44	58
4	1.75	2.30	4.60	37
4	2.00	2.28	4.49	31
5	1.60	2.09	4.53	54
5	1.80	2.06	4.50	41
6	1.50	1.91	4.20	37
6	1.67	1.95	3.17	35

Fig. 3 Error rates for the breast cancer data set

Fig. 4 Error rates of rule sets with the same number of rules for the breast cancer data set

できなかったルール集合は、この表からは省略されている。Table 1 と Table 2 との比較から、多数のルール集合が C4.5 アルゴリズムの最良の結果 (5.1%) よりも良い汎化性能を持つことがわかる。

ルール集合の性能と複雑性の関係を明らかにするために、Fig. 3 に学習用データに対する性能 (左) と評価用データに対する性能 (右) を示す。ルールの長さの総和という 3 番目の目的関数が存在するために、同じルール数のルール集合が複数個存在する。なお、図中の黒丸は、同じルール数のルール集合のなかで、学習用データに対

して最小の誤識別率を持つルール集合を示している。つまり、3 番目の目的を考慮せず、2 目的最適化問題として扱った場合には、黒丸で示されるルール集合しか獲得できないことになる。Fig. 3 の左図から、学習用データに対する性能とルール数の間のトレードオフが明確である。スカラー適応度関数として、ルール集合の性能とルール数の加重和を用いたとすれば、重みの設定に応じて、左図の黒丸の一つが最適解として得られる。なお、評価用データに対する性能を示している右側の図の黒丸は、左側の図の黒丸に対応している。たとえば、学習用データに対する誤識別率の最小化を行った場合、一番右の黒丸に対応するルール集合が得られる。この場合、ルール数の増加により、学習用データに対する過剰適合が発生し、評価用データに対する性能は必ずしも高くない。また、この黒丸に対応するルール集合の汎化能力は 3 目的最適化問題から得られた同じルール数を持つ他の非劣解の汎化能力に劣っている。つまり、ルール長の総和の増加により、学習用データに対する過剰適合が発生している場合が存在する。ルール長の総和の増加による過剰適合をより明らかに示すために、Fig. 4 に形式を変えた図を示す。この図では、平均ルール長が横軸になっており、左側の図はルール数が 2 であるルール集合の性能を、右側の図はルール数が 4 であるルール集合の性能をそれぞれ示している。これらの図から、ルール長の総和の増加によって、評価用データに対する誤識別率が増加している場合が存在することがわかる。

Pima Indians Diabetes Database

乳癌データと同様の形式で、糖尿病データに対して得られた非劣ルール集合を Fig. 5 と Fig. 6 に示す。Table 1 と比較すると、得られたルール集合は C4.5 の最良の結果 (25.0%) より若干悪く、最悪の結果 (27.2%) より良い汎化能力を持つことがわかる。Fig. 5 の左図からは、学習用データに対する性能とルール数の間のトレードオフが明らかである。一方、評価用データに対する誤識別率に対しては、ルール数の増加による学習用データへの過剰適合が発生している。また、ルール長の総和の増加による学習用データへの過剰適合も発生している (Fig. 6)。

Fig. 5 Error rates for the diabetes data set

Glass Identification Database

ガラスデータに対する結果を、Fig. 7 と Fig. 8 に示す。Fig. 8 は、ルール数が 6 であるルール集合と、ルール数が

Fig. 6 Error rates of rule sets with the same number of rules for the diabetes data set

Fig. 7 Error rates of rule sets for the glass data set

Fig. 8 Error rates of rule sets with same number of rules for the glass data set

8であるルール集合の性能をそれぞれ示している。Fig. 7において、学習用データに対する性能とルール数の間のトレードオフは明らかであるが、過剰適合は明らかではない。この結果は、ファジィルール数の増加やメンバーシップ関数の調整により、汎化能力を改善することができる可能性を示唆している。そこで、得られたファジィルール集合に対して、ルールの重みの学習を行った。しかし、ほとんど改良は見られなかったため、結果は省略する。これは、重みの値は既に適切に設定されていたということを示している。評価用データに対する性能は、C4.5 アルゴリズムの最良の結果 (27.3%) にかなり劣っている。一方、ルール長の総和の増加による過剰適合は明らかである。このデータ集合に対して、文献 [25] では、平均 8.5 個のルールを持つファジィ識別システムで 42.1% の誤識別率が報告されている。つまり、(1) 式の形式のファジィルールと Fig. 1 のようなファジィ集合を用いたファジィ識別システムでは、ガラスデータをうまく識別することが困難であることを示している。

Fig. 9 Error rates for the heart disease data set

Fig. 10 Error rates of rule sets with the same number of rules for the heart disease data set

Cleveland Heart Disease Database

Fig. 9 と Fig. 10 に、クリーブランド心臓病データに対する実験結果を示す。Fig. 10 は、ルール数が 6 であるルール集合とルール数が 10 であるルール集合の性能を示している。Fig. 9 から、学習用データに対する性能とルール数の間の明確なトレードオフが明らかであり、ルール数の増加による学習用データへの過剰適合も明らかである。また Fig. 10 から、ルール長の総和の増加による過剰適合も明らかである。得られたルール集合のいくつかの汎化能力は、C4.5 の最良の結果 (46.3%) よりも優れている。

Sonar Database

ソナーデータに対する実験結果を Fig. 11 に示す。学習用データに対する性能とルール数の間に明らかなトレードオフが存在していることがわかる (左図)。しかし、ルール数の増加による過剰適合は明らかではない (右図)。一方、ルール数が 4 と 5 の場合には、ルール長の総和の増加による過剰適合が発生している。得られたルール集合のいくつかの汎化能力は、C4.5 の最良の結果

Fig. 11 Error rates for the sonar data set

Table 3 Classification performance of our ensemble classifier

Data sets	Ensemble systems [%]	Individual average [%]	Number of fuzzy systems	C4.5 error rate [%][7]	
				Best	Worst
Wisconsin Breast Cancer	3.75	4.31	11.5	5.1	6.0
Pima Indians Diabetes	25.22	26.30	31.2	25.0	27.2
Glass Identification	39.25	40.76	38.7	27.3	32.2
Cleveland Heart Disease	46.23	47.19	16.5	46.3	47.9
Sonar	22.79	25.06	13.1	24.6	35.8
Wine Recognition	4.78	6.28	12.0	5.6	8.8

Fig. 12 Error rates for the wine data set

Wine Recognition Data

ワインデータに対する実験結果を Fig. 12 に示す。学習用データに対する性能とルール数の間には、トレードオフが存在していることがわかる（左図）。しかし、ルール数増加による過剰適合は発生しておらず、ルール長の総和の増加による過剰適合も発生していない（右図）。C4.5 の最良の結果 (5.6%) よりも優れた汎化能力を持つルール集合が得られている。

既に述べたように、NSGA-IIの1回の実行で複数のファジィ識別システム（ルール集合）が得られる。本節では、獲得された複数のファジィ識別システムから一つのアンサンブルシステムを構築することにより、多目的最適化問題としてのルール選択問題の取り扱いの有効性を調べる。なお、最終的な識別は、複数のファジィ識別システムの識別結果に基づいて単純な多数決投票によって行った。複数のクラスが同点であった場合、それらのクラスからランダムに識別クラスを決定した。アンサンブルシステムの構築には、クラス数以上のルールを含むファジィ識別システムだけを用いた。たとえば、ガラスデータに対しては、6個以上のルールを含むファジィ識別器システムだけを用いた。アンサンブルシステムによる誤識別率をTable 3に示す。個々の識別器の平均とは、アンサンブルシステムの構築に用いたファジィ識別システムの誤識別率の平均を表わしている。ファジィ識別システムの数、アンサンブルを構成しているファジィ識別システムの数、アンサンブルを構成しているファジィ識別システムの数の平均を表わしている。また、C4.5の

最良と最悪の結果を再びこの表に引用する。Table 3 から、投票を用いたアンサンブルシステムの性能が、平均性能よりも良いことがわかる。また、ガラスデータを除いた他のデータ集合に対しては、C4.5 の最良の結果と同程度、あるいは良い性能を持つことがわかる。NSGA-II で獲得された複数のファジィ識別システムから一つを選択する場合では、評価用データに対してどのファジィ識別システムが最良であるかはわからず、ファジィ識別システムの選択という問題が残る。しかし、アンサンブルシステムを用いることにより、この問題は解決され、評価用データに対して高い性能を得ることができる。また、進化型多目的最適化手法の性能の評価は、パレート解への収束と非劣解集合の分布を考慮しなければならず非常に困難である。しかし、進化型多目的最適化手法の評価指標として、アンサンブルによる識別能力を用いることもできる。

テムから多数のルールから構成される複雑なファジィ識別システムまで、様々な非劣ルール集合を獲得することができた。このように、本研究では、同時に多数の非劣解を見つけだすことができる進化型多目的最適化手法の能力を用いて、ファジィ識別システムの汎化能力とその複雑性との関係の実験的な解析を行った。最後に、複数のファジィ識別システムを統合することによって、信頼性のあるファジィ識別システムを構築することができることを示した。

一般に、進化型多目的最適化手法の性能の評価は、パレートフロントへの収束性および獲得された解集合の広がりと同時に考慮しなければならず非常に困難である。しかし、進化型多目的最適化手法の評価指標として、アンサンブルによる識別能力を用いることもできる。数値実験結果は、知識発見やデータマイニングの分野における進化型多目的最適化手法の潜在的な有用性を示している。

なお、本研究の一部はカシオ科学振興財団からの研究助成を受けて行われたものである。

参考文献

- [1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. I. Verkamo: Fast discovery of association rules; *Advances in Knowledge Discovery and Data Mining* (U. M. Fayyad et al.), AAAI Press, pp. 307–328 (1996)
- [2] J. Casillas, O. Cordon, F. Herrera and L. Magdalena: *Trade-off between Accuracy and Interpretability in Fuzzy Rule-Based Modelling*, Physica-Verlag (2002)
- [3] L. Castillo, A. Gonzalez and R. Perez: Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm; *Fuzzy Sets and Systems*, Vol. 120, pp. 309–321 (2001)
- [4] V. Cherkassky and F. Mulier: *Learning From Data: Concept, Theory, and Methods*, John Wiley & Sons (1998)
- [5] O. Cordon, M. J. Del Jesus and F. Herrera: A proposal on reasoning methods in fuzzy rule-based classification systems; *International Journal of Approximate Reasoning*, Vol. 20, pp. 21–45 (1999)
- [6] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan: A fast and elitist multiobjective genetic algorithm: NSGA-II; *IEEE Trans. on Evolutionary Computation*, Vol. 6, pp. 182–197 (2002)
- [7] T. Elomaa and J. Rousu: General and efficient multi-splitting of numerical attributes; *Machine Learning*, Vol. 36, pp. 201–244 (1999)
- [8] A. Gonzalez and R. Perez: SLAVE: A genetic learning system based on an iterative approach; *IEEE Trans. on Fuzzy Systems*, Vol. 7, pp. 176–191 (1999)
- [9] T. -P. Hong, C. -S. Kuo and S. -C. Chi: Trade-off between computation time and number of rules for fuzzy mining from quantitative data; *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 9, pp. 587–604 (2001)
- [10] H. Ishibuchi, T. Murata and I. B. Turksen: Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems; *Fuzzy Sets and Systems*, Vol. 89, pp. 135–149 (1997)
- [11] H. Ishibuchi and T. Nakashima: Effect of rule weights in fuzzy rule-based classification systems; *IEEE Trans. on Fuzzy Systems*, Vol. 9, pp. 506–515 (2001)
- [12] H. Ishibuchi, T. Nakashima and T. Morisawa: Voting in fuzzy rule-based systems for pattern classification problems; *Fuzzy Sets and Systems*, Vol. 103, pp. 223–238 (1999)
- [13] H. Ishibuchi, T. Nakashima and T. Murata: Three-objective genetics-based machine learning for linguistic rule extraction; *Information Sciences*, Vol. 136, pp. 109–133 (2001)
- [14] H. Ishibuchi, K. Nozaki and H. Tanaka: Distributed representation of fuzzy rules and its application to pattern classification; *Fuzzy Sets and Systems*, Vol. 52, pp. 21–32 (1992)
- [15] H. Ishibuchi, K. Nozaki, N. Yamamoto and H. Tanaka: Selecting fuzzy if-then rules for classification problems using genetic algorithms; *IEEE Trans. on Fuzzy Systems*, Vol. 3, pp. 260–270 (1995)
- [16] H. Ishibuchi and T. Yamamoto: Fuzzy rule selection by data mining criteria and genetic algorithms; *Proc. of Genetic and Evolutionary Computation Conference*, pp. 399–406 (2002)
- [17] H. Ishibuchi and T. Yamamoto: Comparison of heuristic rule weight specification methods; *Proc. of 11th IEEE International Conference on Fuzzy Systems*, pp. 908–913 (2002)
- [18] H. Ishibuchi and T. Yamamoto: Comparison of fuzzy rule selection criteria for classification problems; *Proc. of 2nd International Conference on Hybrid Intelligent Systems*, pp. 132–141 (2002)
- [19] H. Ishibuchi, T. Yamamoto and T. Nakashima: Fuzzy data mining: Effect of fuzzy discretization; *Proc. of 1st IEEE International Conference on Data Mining*, pp. 241–248 (2001)
- [20] J. D. Knowles and D. W. Corne: Approximating the nondominated front using the pareto archived evolution strategy; *Evolutionary Computation*, Vol. 8, pp. 149–172 (2000)
- [21] L. I. Kuncheva: *Fuzzy Classifier Design*, Physica-Verlag (2000)
- [22] C. T. Leondes: *Fuzzy Theory Systems: Techniques and Applications* (Vols. 1-4), Academic Press (1999)
- [23] J. R. Quinlan: *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publisher (1993)
- [24] H. Roubos and M. Setnes: Compact and transparent fuzzy models and classifiers through iterative com-

- plexity reduction; *IEEE Trans. on Fuzzy Systems*, Vol. 9, pp. 516–524 (2001)
- [25] L. Sanchez, I. Couso and J. A. Corrales: Combining GA operators with SA search to evolve fuzzy rule base classifiers; *Information Sciences*, Vol. 136, pp. 175–191 (2001)
- [26] S. M. Weiss and C. A. Kulikowski: *Computer Systems That Learn*, Morgan Kaufmann Publishers (1991)
- [27] E. Zitzler and L. Thiele: Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach; *IEEE Trans. on Evolutionary Computation*, Vol. 3, pp. 257–271 (1999)
-