社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2005－AL－102 (4)
2005／9／16

区間灰数演算によるルール抽出と縮約アルゴリズムの提案

山口 大輔† 李 国棟‡ 水谷 晃三‡
永井 正武§ 北岡 正敏§

本論文は区間灰数演算に基づく新しいルール抽出方法と縮約アルゴリズムを提案する。本論文は灰色理論における区間灰数の概念と，区間解析における区間演算の概念を合わせて，区間灰数演算を定義・拡張し，区間灰数に同値関係を明確にすることで，与えられたデータの空間をラフ集合の同値類に拡張する。すなわち，灰色理論とラフ集合理論の結合を試みることにより，本手法は入力データを複数の既存の可能性カテゴリ，未知の可能性カテゴリ，および既存の必然性カテゴリに分類を行うルールを抽出する。必然性だけを必要とする実例については，ルール縮約アルゴリズムによりルールを簡潔にすることが可能となる。2つのデータ識別問題に本手法を適用する。

Rule Extraction and Reduction Algorithm Based on Grey Lattice Classification

Daisuke YAMAGUCHI† Guo-Dong LI‡ Kozo MIZUTANI‡
Masatake NAGAI§ Masatoshi KITAOKA§

This paper proposes a decision rule of extraction and reduction that is based on the grey lattice classification. This proposal method becomes from joining between rough set theory and grey theory as an approximation algorithm. The grey lattice operations are defined by combining interval grey number in grey theory with interval lattice operations in interval algebra. By defining the equivalents in interval grey number, given data space is correspondent to equivalents of rough set. This proposal method classifies the each data set into 3-patterns from given training samples, as existing possibility class, newly made possibility class and existing necessity class. As given examples which require only necessity class, decision rule is simplified by reduction procedure.

1. はじめに

データ処理，たとえば画像認識や文字（特に漢字）認識，データベースに対する情報検索などにおいて，パターン識別アルゴリズムは常に重要な役割を果たす。適用事例によってさまざまな手法が開発されているが，近年ではPawlak[1, 2]によるラフ集合が注目されている。

Pawlakは，データとカテゴリの因果関係を少ない属性で説明する縮約ルールの抽出方法を提案した。ただし，属性が多くなると計算量が膨大となるため，効率的に縮約ルールを獲得する方法がSkowron,Shan and Ziarkoらなどにより研究されている[3, 4, 5, 6, 7, 8]。

本論文は縮約ルールを効率的に獲得するための1解決方法として，灰色理論[11, 12, 13, 14]とラフ集合の結合に基づくパターン識別アルゴリズムを提案する。

灰色理論は鄧[11]により提唱された基礎理論であり，灰数という区間を持つ数と灰数に対する算術演算[12, 13]が定義されているのが特徴である。灰色理論に基づくクラスタリング手法やパターン識別手法として，例えば，灰色SOFM[10]，灰色分類ネットワーク[15]などの手法が多数報告されている。ただし，これまでの灰色理論による識別アルゴリズムは類似度を評価する灰色分析を中心とするもので，灰数の算術演算を用いない方法である。

本論文ではまず灰数区間演算という灰色理論による新しい演算を定義する。この灰数区間演算によるルール抽出方法と縮約アルゴリズムを提案する。灰数に対する算術演算だけでは本提案アルゴリズムを実現するには不十分である。そこで，筆者らは「区間解析 (interval analysis)[16, 17]」において定義されている区間演算を導入する。

区間解析はMoore[18]が精度保証付き数値計算[20, 21]に適用し，上で丸めた近似値と下で丸めた近似値をまとめて区間とする新しい数値処理方法である。区間に対する演算 (Interval Lattice) が既に定義されているため，筆者らはそれらのうちいくつか重要な定義を使用する。そして，筆者らは灰数と区間演算を結合した灰数区間演算を本論文にて提案する。特に，区間の否定論理の追加により，ド・モ

†神奈川大学大学院工学研究科，Graduate School of Engineering, Kanagawa University
‡帝京大学大学院理工学研究科，Graduate School of Sciense and Engineering, Teikyo University
§神奈川大学工学部，Faculty of Engineering, Kanagawa University

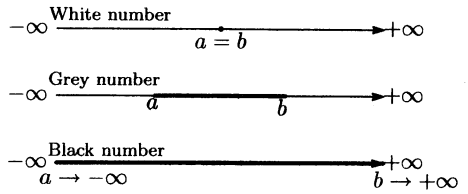

図 1 定義 1 のイメージ

ルガンの法則の拡張が実現可能となる。

本提案アルゴリズムとしての特徴は以下の 2 点である。(1) 与えられたデータを「既存の複数の可能性カテゴリ」,「未知の可能性カテゴリ」,「既存の必然性カテゴリ」に識別する。(2) 必然性だけを取り扱う識別処理では，少ない属性で識別処理を実現するためのルールの縮約を行う。

本論文で定義した灰数区間演算は同値関係を持ち，それがラフ集合の同値類に相当する。この提案により，ラフ集合理論と灰色理論が結合される。

2. 灰数と区間解析

はじめに，区間灰数と呼ばれる灰数の概念と数理表現を明確に定義する。次に，須永 [19] による区間演算を定義する。筆者らは区間演算と区間灰数を結合し，新たな定義と定理を加える。

2.1 灰色理論における灰数の定義

定義 1，定義 2 は文献 [13] による。

定義 1. A を実数の部分集合 $A \subseteq \boldsymbol{R}$ とする。任意の実数 $a, b(a = \inf A, b = \sup A)$ を区間として備える数 $\underset{\otimes}{x}$ について，以下の定義を行う（**図 1**）。

1. $a \to -\infty, b \to +\infty$ のとき，$\underset{\otimes}{x}$ を黒数という
2. $a = b$ のとき，$\underset{\otimes}{x}$ は白数といい，$\otimes x$ で表す
3. 上記のいずれでもないとき，$\underset{\otimes}{x} = [a, b]$ を灰数という

灰数は任意の一点ではなく，任意の二点間を含む区間数 (interval number) を取り扱う。灰数は数値データの集合である。これは，おおよその範囲は既知であるが，明確な値は未知であることを表している。白数だけが灰数の区間内の 1 点としたときの値である。

定義 2. 定義 1 にて示した実数 a, b を 1 つの区間とするとき，任意の実数 x が

$$G_1(x) = \begin{cases} 1 & x \in [a, b] \\ 0 & x < a, b < x \end{cases} \tag{1}$$

$$G_2(x) = \begin{cases} [0, 1] & x \in [a, b] \\ 0 & x < a, b < x \end{cases} \tag{2}$$

となる $G_1(x)$ を区間灰数，$G_2(x)$ を鄧の灰数 (Deng's grey number) とする。

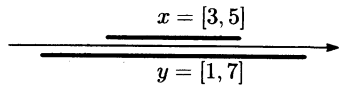

図 2 定義 4 のイメージ

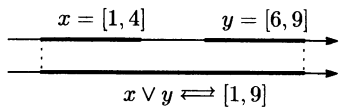

図 3 定義 7 のイメージ

灰色理論では上記の 2 つの定義による灰数が存在するが，本論文では前者の区間灰数を用いる。

2.2 区間解析における諸定義

区間演算の諸定義を文献 [19] より与えられる。

定義 3. x を実数の集合 $\boldsymbol{R}$ の要素とするとき，

$$[\underline{x}, \overline{x}] = \{x \in \boldsymbol{R} | \underline{x} \le x \le \overline{x}\} \tag{3}$$

を区間 (interval) とする。

定義 4. 区間 $x = [a, b]$ が他の区間 $y = [c, d]$ 要素であれば，

$$x \to y \qquad (c \le b, b \le d) \tag{4}$$

このとき，区間 x は y に含まれる (include) という（**図 2**）。

集合論の "$\subseteq$" と混同を避けるため，区間解析では "$\to$" を用いる。

定義 5. 区間 x, y が $x \to y$ かつ $y \to x$ であれば，x, y は一致 (coincidence) しているといい，以下に表す。

$$x \rightleftarrows y \qquad (a = c, b = d) \tag{5}$$

定義 6. 定義 4，5 より，区間解析では区間 x, y, z について以下の同値関係が成立する。

$$\left.\begin{array}{l} x \to x \\ x \to y, y \to x \Leftrightarrow x \rightleftarrows y \\ x \to y, y \to z \Rightarrow x \to z \end{array}\right\} \tag{6}$$

定義 7. 2 つの区間 $x = [a, b]$，$y = [c, d]$ の結合 (join) を以下に示す（**図 3**）。

$$x \vee y \rightleftarrows [a, d] \qquad (a \le d) \tag{7}$$

区間ではなく，任意の点 $a, b(a \le b)$ では，

$$a \vee b \rightleftarrows [a, b] \tag{8}$$

定義 8. 2 つの区間 $x = [a, b]$，$y = [c, d]$ の共通 (meet) を以下に示す（**図 4**）。

$$x \wedge y \rightleftarrows \begin{cases} [c, d] & (c \le b) \\ \varnothing & (b < c) \end{cases} \tag{9}$$

任意の点 a, b では，

$$a \wedge b \rightleftarrows \begin{cases} a & (a = b) \\ \varnothing & (a \ne b) \end{cases} \tag{10}$$

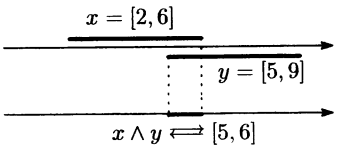

図 4 定義 8 のイメージ

定義 9. 区間 x, y, z について，区間解析では以下の性質が成り立つ。

1. $x \vee x \rightleftarrows x, x \wedge x \rightleftarrows x$
2. $x \vee y \rightleftarrows y \vee x, x \wedge y \rightleftarrows y \wedge x$
3. $(x \vee y) \vee z \rightleftarrows x \vee (y \vee z), (x \wedge y) \wedge z \rightleftarrows x \wedge (y \wedge z)$
4. $(x \vee y) \wedge x \rightleftarrows x, (x \wedge y) \vee x \rightleftarrows x$

定義 10. 区間 $x = [a, b]$ の直径 (diameter)$d(x)$ を以下に表記する *[21]*。

$$d(x) = b - a \tag{11}$$

2.3 筆者らによる諸定義

これまでの定義のほかに，筆者らはさらに以下の定義を追加する。

定義 11. 定義 *3* における区間の否定を以下に定義する。

$$\neg[\underline{x}, \overline{x}] = \{x \in \boldsymbol{R} | x < \underline{x}, \overline{x} < x\} \tag{12}$$

公理 1. 区間 x の否定を用いると，以下の補元律が成り立つ。

$$x \wedge \neg x \rightleftarrows \varnothing \tag{13}$$
$$x \vee \neg x \rightleftarrows \boldsymbol{R} \tag{14}$$

定理 1. *2* つの区間 x, y について，ド・モルガンの法則 (De Morgan's raw) が成立する。

$$\begin{aligned} \neg(x \vee y) &\rightleftarrows \neg x \wedge \neg y \\ \neg(x \wedge y) &\rightleftarrows \neg x \vee \neg y \end{aligned} \tag{15}$$

証明. 定義 7,8,9,11，公理 1 により，ブール代数と同等の諸定義が得られている。ここで，結合を論理和，共通を論理積とすれば，ブール代数におけるド・モルガンの法則と同じように，式 (15) を導出することができる。 □

定義 12. ブール代数における排他的論理和に相当する演算として，*2* つの区間 x, y における排他的結合 (Exclusive join) を定義する。

$$\begin{aligned} x \oplus y &= (x \vee y) \wedge (\neg x \vee \neg y) \\ x \oplus y &= (x \vee y) \wedge \neg(x \wedge y) \end{aligned} \tag{16}$$

さらに，本論文にて区間灰数と区間演算の結合を行う。

定義 13. $\underset{\otimes}{x}, \underset{\otimes}{y}$ をそれぞれ区間灰数とするとき，定義 *4,5,6* を区間灰数について再定義する。

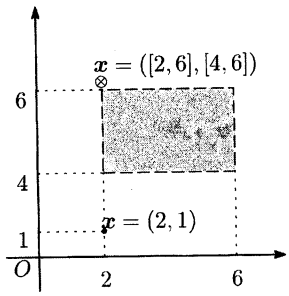

図 5 定義 16 のイメージ

定義 14. $\underset{\otimes}{x}, \underset{\otimes}{y}$ をそれぞれ区間灰数とするとき，定義 *7,8,10,11,12* にて示した演算を再定義する。これらの演算を区間灰数 $\underset{\otimes}{x}, \underset{\otimes}{y}$ に施すとき，定義 *7,8,10,11,12* は区間灰数演算とする。

定義 15. $\underset{\otimes}{x}, \underset{\otimes}{y}$ を区間灰数とするとき，定義 *9*, 公理 *1* を再定義する。

定理 2. $\underset{\otimes}{x}, \underset{\otimes}{y}$ を区間灰数とするとき，ド・モルガンの法則が成立する。

$$\begin{aligned} \neg(\underset{\otimes}{x} \vee \underset{\otimes}{y}) &\rightleftarrows \neg \underset{\otimes}{x} \wedge \neg \underset{\otimes}{y} \\ \neg(\underset{\otimes}{x} \wedge \underset{\otimes}{y}) &\rightleftarrows \neg \underset{\otimes}{x} \vee \neg \underset{\otimes}{y} \end{aligned} \tag{17}$$

証明. 定義 13,14,15 および定理 1 により，区間灰数についてもド・モルガンの法則が導くことができる。 □

また，須永は n 次元区間を文献 [19] で定義している。筆者らはそれをベクトルとして新たに定義する。

定義 16. 灰数を成分とするベクトル $\underset{\otimes}{\boldsymbol{x}}$ は，以下のように表記する (**図 *5***)。

$$\underset{\otimes}{\boldsymbol{x}} = \left(\underset{\otimes}{x}_1, \underset{\otimes}{x}_2, \cdots, \underset{\otimes}{x}_n \right)^T \tag{18}$$

ただし，$\underset{\otimes}{x}_n \rightleftarrows [\underline{x}_n, \overline{x}_n]$ とする。

定義 17. 区間 y について，$x \to y$ となる要素 x の総数を y の基数 (濃度：cardinal number) といい，$|y|$ で表す。同様に，$x \to \underset{\otimes}{y}, \underset{\otimes}{x} \to \underset{\otimes}{y}$ となる要素 $x, \underset{\otimes}{x}$ の総数を $|\underset{\otimes}{y}|$ で表す (**図 *6***)。

定義 18. *2* つの区間 x, y において，

$$v(x, y) = \frac{d(x \wedge y)}{d(x \vee y)} \tag{19}$$

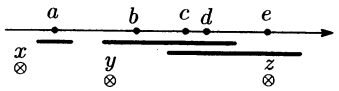

図 6 定義 17 のイメージ

となる $v(x,y)$ を重複度 (overlap degree) という。ただし,

$$x \wedge y = \varnothing \Leftrightarrow v(x,y) = 0 \tag{20}$$

$$x \rightleftharpoons y \Leftrightarrow v(x,y) = 1 \tag{21}$$

区間灰数 $\underset{\otimes}{x}, \underset{\otimes}{y}$ についても同様の演算が可能である。

定義 19. *2* つの区間 x, y において,

$$\delta(x,y) = \frac{|x \wedge y|}{|x \vee y|} \tag{22}$$

となる $\delta(x,y)$ を *2* つの区間 x, y の密度 (density) という。区間灰数 $\underset{\otimes}{x}, \underset{\otimes}{y}$ についても同様の演算が可能である。

3. 灰数区間演算による識別処理の提案

3.1 基本的な処理

本手法は教師データを使ったデータの識別を行う。その原理は**図 7**に示すように, 教師データの結合により, 各カテゴリの持つ領域を構成する。

灰数の概念はカテゴリが存在する領域を定義するのに適している。教師データおよび識別したいデータは常に同じ値になるとは限らない。そのため, 教師データの値の差を考慮し, 識別したいデータを正しいカテゴリに受け入れるための領域を灰数として定める。

カテゴリが持つ領域の構成を, 区間演算により定義することは重要である。なぜならば, 区間は定義 6 により, 同値関係が明確に定義されている。識別のために与えられたデータの空間はラフ集合の同値類 (rough equarity) に相当するためである。

識別処理を行う前処理として, 教師データから各カテゴリの領域を構築する方法を定義する。

定義 20. 白数である教師データを $\boldsymbol{x}$ とすると, n 個の属性を持つベクトルであれば,

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T \tag{23}$$

定義 21. 教師データが属しているカテゴリ (category, class) を $C_j (j = 1, 2, \cdots, M)$ とすると, その代表的なベクトルは

$$C_j = (c_{j1}, c_{j2}, \cdots, c_{jn})^T \tag{24}$$

または,

$$\underset{\otimes}{C}_j = \left(\underset{\otimes}{c}_{j1}, \underset{\otimes}{c}_{j2}, \cdots, \underset{\otimes}{c}_{jn}\right)^T \tag{25}$$

$\underset{\otimes}{C}_j$ をカテゴリ C_j の存在領域 (ED : Existing Domain) と定義する。

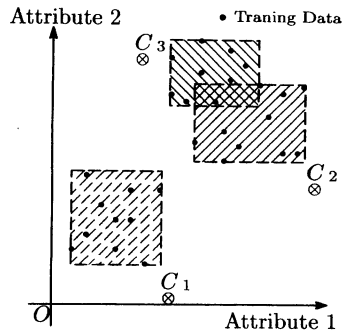

図 7 教師データを用いた存在領域の構築イメージ

定義 22. 存在領域 $\underset{\otimes}{C}_j$ の属性 $\underset{\otimes}{c}_{jk} (1 \leq k \leq n)$ は, 教師データ $\boldsymbol{x}$ を用いて次式にて算出する。

$$\underset{\otimes}{c}_{jk} = \bigvee_{\boldsymbol{x} \in C_j} x_k \tag{26}$$

定義 21 に示すように, C_j の成分が白数であれば, 代表的なベクトルは GCN[15] の結合ベクトルと等価である。本論文は, C_j を白数から灰数に拡張したのが特徴である。

教師データによって構築された存在領域 $\underset{\otimes}{C}_j$ により, 識別処理を定義する。識別処理の原則は, 入力データ $\boldsymbol{x}$ が C_j の存在領域内かどうかで判定する。入力データの識別は可能性, 必然性から行う。

定義 23. 入力データ $\boldsymbol{x}$ が,

$$P^{\exists}(\boldsymbol{x}) = \left\{ C_j | \exists k, x_k \rightarrow \underset{\otimes}{c}_{jk} \right\} \tag{27}$$

を満たす $P^{\exists}(\boldsymbol{x})$ を可能解集合とする。

定義 24. 入力データ $\boldsymbol{x}$ が,

$$P^{\forall}(\boldsymbol{x}) = \left\{ C_j | \forall k, x_k \rightarrow \underset{\otimes}{c}_{jk} \right\} \tag{28}$$

を満たす $P^{\forall}(\boldsymbol{x})$ を必然解集合とする。

定義 25. 入力データ $\boldsymbol{x}$ が,

$$P^{!}(\boldsymbol{x}) = P^{\exists}(\boldsymbol{x}) \cap P^{\forall}(\boldsymbol{x}) \tag{29}$$

となる $P^{!}(\boldsymbol{x})$ を可能必然解集合とする。

定義 26. 入力データ $\boldsymbol{x}$ について,

$$C^* = \left\{ \boldsymbol{x} | P^{!}(\boldsymbol{x}) = \varnothing \right\} \tag{30}$$

となる集合 C^* を未知データ集合と定義する。

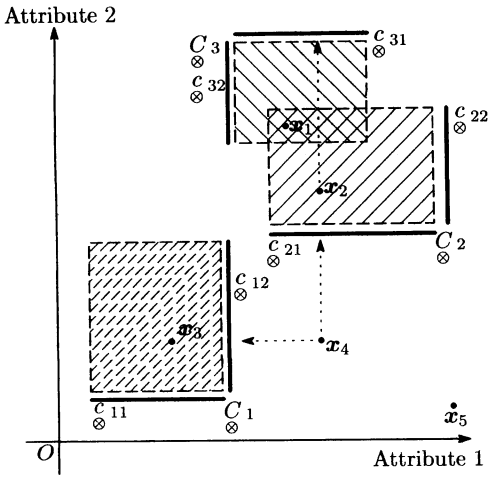

図 8 識別処理の例

定義 23 にて示した可能解集合とは，教師データにより構成された存在領域のうち，1 つの属性でも区間内に存在するカテゴリを解とする集合である。$P^{\exists}(\boldsymbol{x})$ は Rough 集合の上近似 (upper approximation) に相当し，既存のカテゴリに $\boldsymbol{x}$ を識別処理する。

定義 24 にして示した必然解集合とは，教師データにより構成された存在領域のうち，すべての属性が区間内に存在するカテゴリを解とする集合である。$P^{\forall}(\boldsymbol{x})$ は Rough 集合の下近似 (lower approximation) に相当し，既存のカテゴリに $\boldsymbol{x}$ を識別処理する。

本論文ではさらに定義 25，26 を与える。定義 23，24 より，

$$P^{\forall}(\boldsymbol{x}) \subseteq P^{\exists}(\boldsymbol{x}) \tag{31}$$

である。あるデータ $\boldsymbol{x}$ について必然解集合が存在すれば可能解集合も存在する。もし $P^{!}(\boldsymbol{x}) = \varnothing$ であれば，入力データ $\boldsymbol{x}$ は可能解集合だけを持つか，または可能解集合も必然解集合も持たない。$P^{!}(\boldsymbol{x}) = \varnothing$ となるデータ $\boldsymbol{x}$ は既存のカテゴリでは説明できないが，新しくカテゴリを作れば可能解として説明できる。すなわち $\boldsymbol{x}$ は未知のカテゴリに識別処理する。

3.2 物理的意味

図 8の例を用いて可能解集合，必然解集合の物理的意味を示す。図 8 の例は 3 カテゴリ (C_1, C_2, C_3) と 5 データ $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_5)$ がある。各データの解は図 8 より，

$$\begin{array}{ll} P^{\exists}(\boldsymbol{x}_1) = \{C_2, C_3\} & P^{\exists}(\boldsymbol{x}_2) = \{C_2, C_3\} \\ P^{\forall}(\boldsymbol{x}_1) = \{C_2, C_3\} & P^{\forall}(\boldsymbol{x}_2) = \{C_2\} \\ P^{!}(\boldsymbol{x}_1) = \{C_2, C_3\} & P^{!}(\boldsymbol{x}_2) = \{C_2\} \\ & \because x_{21} \rightarrow c_{\otimes 31} \end{array}$$

$$\begin{array}{ll} P^{\exists}(\boldsymbol{x}_3) = \{C_1\} & P^{\exists}(\boldsymbol{x}_4) = \{C_1, C_2, C_3\} \\ P^{\forall}(\boldsymbol{x}_3) = \{C_1\} & P^{\forall}(\boldsymbol{x}_4) = \varnothing \\ P^{!}(\boldsymbol{x}_3) = \{C_1\} & P^{!}(\boldsymbol{x}_4) = \varnothing \end{array}$$

$$\begin{array}{ll} P^{\exists}(\boldsymbol{x}_5) = \varnothing & C^* = \{\boldsymbol{x}_4, \boldsymbol{x}_5\} \\ P^{\forall}(\boldsymbol{x}_5) = \varnothing & \\ P^{!}(\boldsymbol{x}_5) = \varnothing & \end{array}$$

これらの解より，$\boldsymbol{x}_3$ は C_1 を唯一の解とする。

$\boldsymbol{x}_2$ は C_2 の存在領域に入っている。しかし，属性 1 だけに注目すると C_3 の存在領域に入っているため，C_3 は可能解とする。$\boldsymbol{x}_1$ は C_2, C_3 の両方の存在領域に含まれているため，可能必然解となる。これらのデータは 2 つのカテゴリに存在する可能性がある。例えば，パターン認識として誤認識する可能性とみるか，情報検索結果として複数の解釈が可能とみるかは適用事例に依存する。

$\boldsymbol{x}_4$ は属性 1 だけに注目したときは C_2, C_3 に入り，属性 2 だけに注目したときは C_1 の区間内である。しかし，$\boldsymbol{x}_4$ は必然解を持たない。$\boldsymbol{x}_5$ については可能解と必然解を両方とも持たない。これらのデータは既存のカテゴリで全てを説明するのは難しい。例えば，動植物の調査に適用したとすると，これらのデータは既存のカテゴリの突然変異とみるか，または新種の発見かもしれない。工学分野ではしばしば厄介と思われる，これらのデータへの新たな解釈を本手法は与える。

4. ルール抽出と縮約アルゴリズム

4.1 基本概念

人物認証や文字認識といったパターン認識などにおいて，必然解 $P^{\forall}$ だけを必要とする工学分野では，識別処理を迅速にかつ正確に行う必要がある。そのような分野に本手法を適用するのであれば，以下に提案するルール抽出アルゴリズムと縮約アルゴリズムが有益である。

このアルゴリズムは与えられた教師データから，変数選択法のように識別に使える属性だけを取り出し，カテゴリ識別ルールを IF-THEN 形式で表す。その結果，多くのデータを効率的に処理する最小識別ルールを実現する。

前掲の図 6 を例として，区間灰数の性質を用いて本論文で提案するルール抽出の原理を説明する。図 6 は，カテゴリ x, y, z が持つある 1 つの属性の特徴を存在領域で表している。

パターン認識などのルール抽出では他のカテゴリが持っていない特徴 (feature) をもっとも重要視する。我々が通常 2 つのもの，特に本物と偽物を識別するとき，共通部分よりも相違部分に注目する。

区間における相違部分とは，図 6 の x に示すように，他のカテゴリの区間と重複がないことを意味する。x は他の区間と重複がないため，図 6 に示す 1 つの属性を用いるだけで識別を行うことができる。

y, z は互いに重複する区間 (meet) が存在するため，全てのデータを 1 つの属性だけで識別すること

はできない。しかし，重複していない区間，すなわち $y \oplus z$ を用いれば，この属性で識別が可能なデータ（図 6 の b, e）は存在する。もし c, d のようなデータについて，この属性が重複している区間内に含まれているのであれば，他の属性を用いて識別を試みる。

4.2 ルール抽出アルゴリズム

定義 27. 任意の属性 $k(1 \leq k \leq n)$ において，存在領域 $\underset{\otimes}{C}_j, \underset{\otimes}{C}_h (j \neq h)$ の成分 $\underset{\otimes}{c}_{jk}, \underset{\otimes}{c}_{hk}$ の共通が，

$$\underset{\otimes}{c}_{jk} \wedge \underset{\otimes}{c}_{hk} = \varnothing \qquad \text{for all } h \tag{32}$$

となるとき，カテゴリ C_j は属性 k にて識別可能である。

定義 28. ある属性 $k(1 \leq k \leq n)$ において，カテゴリ C_j が識別可能または条件付識別可能な区間は，

$$C_j^{\forall}(x_k) \rightleftarrows \underset{\otimes}{c}_{jk} \wedge \bigoplus_{h=1}^{M} \underset{\otimes}{c}_{hk} \tag{33}$$

$C_j^{\forall}(x_k)$ を属性 k における識別可能区間と定義する。

定義 29. 識別可能区間の中で，

$$\sum_{h=1}^{M} |C_h^{\forall}(x_k)| = \max \tag{34}$$

となる属性 k を最高効率属性と定義する。

はじめに，定義 22 に従い，教師データを用いて各カテゴリの存在領域を獲得する。

次に，各存在領域から全ての属性について定義 27 に従い識別可能かどうか調べる。定義 28 に従い識別可能区間 $C_j^{\forall}(x_k)$ を抽出する。この区間内にデータが入っているか否かを，IF-THEN 形式でルール表現する。

識別ルールは属性単位で抽出する。効率のよいルールとは，1 つの識別可能区間で最も多くの教師データを識別できるルールのことである。効率的なルールとそうでないルールを区別するため，識別可能区間に順序を与える。最高効率属性は，識別可能区間に含まれる教師データの数が最も多い属性であるので，定義 29 を満たす属性をルールにする。

次に，今抽出した識別ルールにより識別可能な教師データ，カテゴリは次のルール抽出には不要のため取り除く。残ったカテゴリと教師データを使って引き続きルールを抽出する。

4.3 ルール縮約アルゴリズム

定義 27 に示す $C_j^{\forall}(x_k), C_h^{\forall}(x_k)$ が 2 つ以上存在するとき，定義 30 の条件を満たせば区間を結合してよい。本手法における識別ルールの縮約の特徴である。

定義 30. 定義 27 に示す $C_j^{\forall}(x_k)$ が 2 つ以上存在し，$C_j^{\forall}(x_k) \wedge C_h^{\forall}(x_k) = \varnothing (j \neq h)$ であれば，2 つの $C_j^{\forall}(x_k), C_h^{\forall}(x_k)$ は条件付識別可能である。

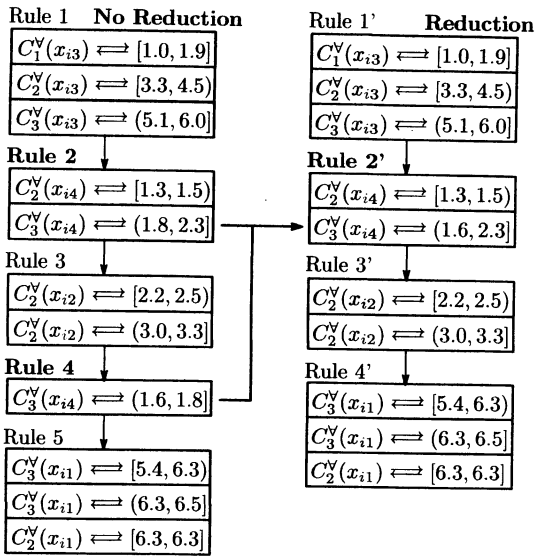

図 9 IRIS を識別する決定木

表 1 抽出されたルールと識別可能な教師データ数

使用したルール	識別率
1	54/75 (72%)
1 and 2	61/75 (81.3%)
1,2 and 3	65/75 (87%)
1 to 4	69/75 (92%)
1 to 5	75/75 (100%)

5. 例題

5.1 IRIS データ

ルール抽出と識別の例として，Fisher あやめデータ¶で分類を試みる。3 カテゴリ（C_1: setosa, C_2: versicolour, C_3: virginica）$\times 50 = 75$ のデータのうち，75 を教師データ，残りを識別テスト用とする。

定義 22 により得られた存在領域を用いてルール抽出を行うと，**図 9**左に示す 5 ルールを抽出することができる。抽出したルールと識別可能な教師データ数の推移は**表 1**により把握することができる。

図 9 によると，ルール 2 とルール 4 は定義 30 の条件を満たすため，あわせて 1 つのルールにすることができる。同じ属性 4 を使って識別するが，2 つのルールを 1 つにしても 2 つの存在領域が重複するといった矛盾が生じないからである。ルールを結合すると，図 9 右に示す 4 ルールとなり，Subclass 法で得られる決定木 [22] よりも少ないルールとなる。

¶データは UCI Machine Learning Repository (http://www.ics.uci.edu/mlearn/MLRepository.html) より使用

表 2 識別テスト用データの識別率

使用したルール	縮約なし	縮約あり
1	59/75 (78.7%)	59/75 (78.7%)
1 and 2	67/75 (89.3%)	69/75 (92%)
1,2 and 3	70/75 (93.3%)	72/75 (96%)
1 to 4	72/75 (96%)	73/75 (97.3%)
1 to 5	73/75 (97.3%)	—

表 3 各手法の識別率

	Proposed Method	GCN[15]	SM[22]
rate	97.3%	96%	97.3%

抽出したルールを識別テスト用データに適用すると，**表 2**に示す識別率を得ることができた。縮約しない場合とした場合では，全てのルールを使用したときの識別率は同じである。ただし，縮約したルールは97.3%の識別率に到達するまでのルール数が縮約なしに比べて少ない。これは，ルールの縮約が識別を効率的にすることを表している。

表 3は，提案手法，GCN[15]，本提案手法と同じくカテゴリを領域で識別する Subclass 法 [22](SM) の各識別率を示す。提案手法は SM よりも少ないルール数であるにもかかわらず，同じ識別率である。また，筆者らが以前に提案した GCN よりも識別率が向上している。

5.2 WINE データ

同様に，2 番目の適用例として，WINE データ‖の識別を行う。178 サンプルからなる WINE データは 3 カテゴリで 13 の属性を備えている。IRIS データと同様半分を教師データ，残りをテスト用とする。

ルール抽出と縮約アルゴリズムを実行することにより，**図 10**左に示す決定木が得られる。抽出されたルールと識別した教師データ数の推移を**表 4**に示す。

6 ルールを抽出したが，縮約すると 10 右に示す 4 ルールとなる。縮約してないルールと縮約したルールを用いてテストデータを識別したところ，**表 5**に示す識別率が得られている。

同じ教師データを使って GCN でも WINE データの識別を行った。識別結果の比較を**表 6**に示す。識別率はどの方法とも同程度である。しかし，GCN は 13 属性を全て使っているのに対し，本手法は高々 4 属性で GCN と同程度の識別率を確保している。この結果は，本提案手法が効率的に識別処理を実現していることを表している。本手法は属性の数を増やすか，またはアルゴリズムの改良によりさらに高精度化が望める。

‖このデータも UCI より使用。

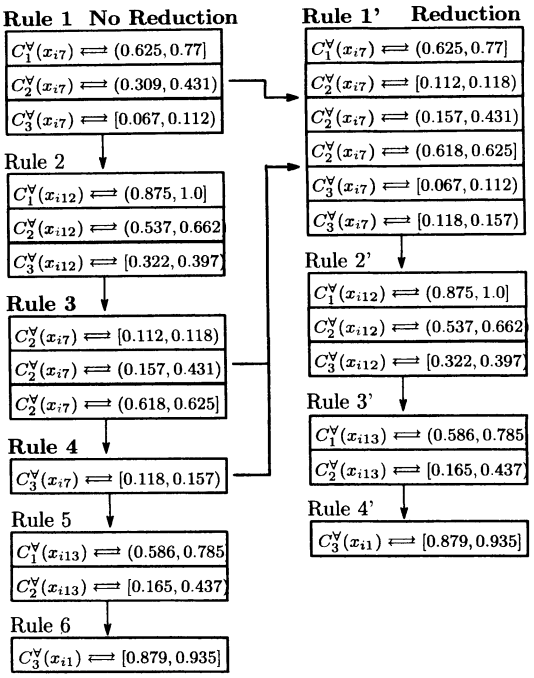

図 10 抽出されたルール (wine)

表 4 抽出されたルールと識別可能な教師データ数

使用したルール	識別率
1	30/88 (34.1%)
1 and 2	53/88 (60.2%)
1,2 and 3	62/88 (70.5%)
1 to 4	69/88 (77.3%)
1 to 5	83/88 (94.3%)
1 to 6	88/88 (100%)

表 5 識別テスト用データの識別率

使用したルール	縮約なし	縮約あり
1	30/90 (33.3%)	47/90 (52.2%)
1 and 2	41/90 (45.6%)	51/90 (56.7%)
1,2 and 3	44/90 (48.9%)	74/90 (82.2%)
1 to 4	52/90 (57.8%)	77/90 (85.6%)
1 to 5	74/90 (82.2%)	—
1 to 6	78/90 (86.7%)	—

6. おわりに

本論文は灰数区間演算に基づくデータ識別アルゴリズムを提案した。

表 6 各手法との比較

	No Reduction	Reduction	GCN
rate	86.7%	85.6%	85.6%

本提案手法を提案するため，本論文では灰色理論における区間灰数の概念と，区間解析における区間演算の概念を合わせて，灰数区間演算を定義した。

データの識別処理に灰数区間演算を使用した。識別処理は可能解集合，必然解集合，未知データ集合にデータを分類できる。灰数区間演算の同値関係が，Rough 集合における近似アルゴリズムとして，本提案手法は灰色理論と Rough 集合理論のインタフェースになる。

必然性だけを必要とする事例に対応するため，ルール抽出アルゴリズムと縮約アルゴリズムを提案した。IRIS のデータを用いて 4 ルールで 97.3%，WINE データは 4 ルールで 85.6%の識別率であったことを報告する。

本論文では教師データを使用したが，教師データを必要としない場合のアルゴリズムが今後の課題である。

参 考 文 献

[1] Z. Pawlak, Rough Sets, *Int. J. of Computer and Information Sciences*, 11(5):341–356, 1982.

[2] Z. Pawlak, Rough Classification, *Int. J. Man-Machine Studies*, 20:469–483, 1984.

[3] N. Mori, KANSEI Engineering for Designing, *J. Japan Society for Fuzzy Theory and Systems*, 11(1):52–63, 1999.

[4] A. Nakamura, Rough set—Its theory and applications, *J. Japan Society for Fuzzy Theory and Systems*, 8(4):594–603, 1996.

[5] 日本ファジイ学会 (編)，ファジイとソフトコンピューティングハンドブック，pp.535–565，共立出版，2000.

[6] T. Okuzaki, S. Hirano, S. Kobayashi, Y. Hata and Y. Takahashi, A Rough Set Based Clustering Method by Knowledge Combination, *IEICE Trans. Inf. & Syst.*, E85-D(12):1898–1908, 2002.

[7] 井上勝雄，ラフ集合，多変量解析の使い方，pp.169–184，筑波出版会，2002.

[8] K. Okuhara, Y. Matsubara, K. Sugihara and H. IShii, Rule Selection by Rough Set Considering Ordinality in Attributes for Kansei Evaluation, *IEICE Trans. Fundamentals*, J87-A(7):1045–1053, 2004.

[9] 森 典彦，田中英夫，井上勝雄 (編)，ラフ集合と感性，海文堂，2004.

[10] Y. C. Hu, R. S. Chen, Y. T. Hsu and G. H. Tzeng, Grey self-organizing feature maps, *Neurocomputing*, 48(4):863–877, 2002.

[11] J. L. Deng, Control problems of grey systems, *Systems & Control Letters*, 1(5):288–294, Mar. 1982.

[12] J. L. Deng, *Grey Systems*, China Ocean Press, Beijing, China, 1988.

[13] X. Jun, Grey System Theory to Hydrology, Huanzhong University of Science and Technology Press, Wuhan, 2000.

[14] 永井正武，山口大輔，灰色理論と工学応用方法，共立出版，2004.

[15] 山口大輔，李 国棟，水谷晃三，永井正武，“灰色理論による学習ベクトル量子化法への一提案，” 情報処理学会研究報告，vol.2005, no.26, AL-100-2, pp.9–16, Mar. 2005.

[16] F. Bierbaum and K. P. Schwiertz, A Bibliography on Interval-Mathematics, *J. of Comput. and App. Math.*, 4(1):59–86, 1978.

[17] J. Gorloff and K. P. Schwiertz, A Bibliography on Interval-Mathematics (Continuation), *J. of Comput. and App. Math.*, 6(1):67–79, 1980.

[18] R. E. Moore, *Reliability in Coputiong*, Academic Press, San Diego, CA, 1988.

[19] T. Sunaga, Theory of interval algebra and its application to numerical analysis, *RAAG Memoirs*, 2:29–46, Gakujutsu Bunken Fukyukai, Tokyo, 1958.

[20] 大石進一，非線形解析入門，コロナ社，1997.

[21] 大石進一，精度保証付き数値計算，コロナ社，2000.

[22] M. Kudo and M. Shimbo, Optimal Subclasses with Dichotomous Variables for Feature Selection and Discrimination, *IEEE Trans. on Syst., Man, and Cybern.* 19(5):1194–1199, Sept. 1989.

[23] M. Kudo, Y. Torii, Y. Mori and M. Shimbo, Approximation of class regions by quasi convex hulls, *Pattern Recognition Letters*, 19:777–786, 1998.