可変精度ラフ集合モデルにおける決定ルールの確信度および被覆度の平均値に関する一考察

A Note on Averages of Certainty and Coverage of Decision Rules in Variable Precision Rough Set Models

工藤 康生 村井 哲也
Yasuo Kudo Tetsuya Murai
室蘭工業大学 北海道大学
Mutoran Institute of Technology Hokkaido University

Abstract: In this paper, we derive two equations to calculate averages of certainty and coverage of decision rules in variable precision rough set (for short, VPRS) models proposed by Ziarko. VPRS is an extension of Pawlak's rough set theory, and it provides a theoretical basis to treat inconsistent or probabilistic information in the framework of rough sets. Many kinds of attribute reduction in VPRS have also been proposed, however, studies of evaluation criteria of reducts in VPRS have not been widely investigated as far as we know. On the other hand, the authors have proposed an evaluation criterion of relative reducts in Pawlak's rough set by using averages of coverage of decision rules constructed from relative reducts. Thus, we intend that the derived two equations in the paper are basis to apply the authors' evaluation criterion of relative reducts to many kinds of reducts in VPRS, and the derived equations also include the similar equations in the case of Pawlak's rough set that the authors have proposed.

1 はじめに

Pawlak が提唱した**ラフ集合** (rough set) [6, 7] は，カテゴリカルなデータに対する近似および規則性の抽出に関する数学的な基礎概念を与えている．特に，ラフ集合による規則性の抽出に関しては，データを正しく分類するために最小限必要となる項目の集合 (相対縮約) およびデータに含まれる if-then 形式のルール (決定ルール) の抽出について，理論と応用の両面から幅広く研究が進められている (詳細は例えば [5])．その一例として，著者らは相対縮約から得られたすべての決定ルールにおける被覆度の平均値を用いて，相対縮約を評価する手法を提案している [4].

また，Ziarko が提案した**可変精度ラフ集合モデル** (variable precision rough set model) [8] は，ラフ集合における近似の定義を緩め，矛盾を含む情報をラフ集合の枠組みで扱う基礎的な手法を与えている．可変精度ラフ集合モデルにおける縮約の定義に関しては，下近似による近似の質を保存する縮約 [1]，下近似の構造を保存する縮約 [2] など多様な提案がなされている．しかし，これら多様な縮約に対する評価手法に関する研究は，あまり行われていない.

本研究では，可変精度ラフ集合モデルにおける各種の縮約に対する評価尺度の構築を目指し，その基礎として，可変精度ラフ集合モデルにおける決定ルールの確信度および被覆度の平均値に関する性質について考察する．具体的には，予め設定した精度 β $(0 \leq \beta < 0.5)$ の下で，任意の空でない条件属性集合および決定属性集合から得られる，確信度が β より大きいすべての決定ルールの確信度の平均値および被覆度の平均値に関する性質を導出する.

2 ラフ集合

本節ではラフ集合に関する基礎的な内容について概説する．なお，本節の内容は文献 [2, 3] に基づく.

2.1 情報表と識別不能関係

ラフ集合によるデータ解析では，一般的に，複数の属性とその値の組として表されるカテゴリカルなデータを対象とする．このようなデータは以下の 4 項組で定義される**決定表** (decision table) として表される.

$$(U, C \cup D, V, \rho),$$

ここで，U は対象の空でない有限集合，C は条件属性の空でない有限集合，D は決定属性の空でない有限集合であり，$C \cap D = \emptyset$ とする．すべての属性の集合を $AT \stackrel{\mathrm{def}}{=} C \cup D$ と表す．V は各属性 $a \in AT$ の値の集合，$\rho : U \times AT \to V$ は対象 x の属性 a での値 $\rho(x, a) \in V$ を表す関数である.

属性の任意の部分集合 $A \subseteq AT$ に対して，U 上の関係 R_A を次式で定義する.

$$xR_Ay \stackrel{\mathrm{def}}{\iff} \begin{array}{l} \text{すべての } a \in A \text{ に対して} \\ \rho(x, a) = \rho(y, a). \end{array} \quad (1)$$

表 1: 決定表の例

U	c_1	c_2	c_3	c_4	c_5	c_6	d
x_1	1	1	1	1	1	2	1
x_2	2	2	1	1	1	2	1
x_3	2	3	2	1	2	1	2
x_4	2	2	2	2	2	1	2
x_5	2	2	3	1	1	2	1
x_6	1	2	1	1	2	2	3

関係 R_A が同値関係となることは容易に確かめられる．R_A による x の同値類 $[x]_A$ は，A に含まれるどの属性についても x と区別できない対象の集合となる．このような関係 R_A を**識別不能関係** (indiscernibility relation) と呼ぶ．決定属性集合 D に基づく識別不能関係はサンプルの全体集合の分割 $\mathcal{D} = \{D_1, \cdots, D_m\}$ を与え，各 D_i は**決定クラス** (decision class) と呼ばれる．

各決定クラス D_i に対して，識別不能関係 R_A による**下近似** (lower approximation) $\underline{A}(D_i)$ および**上近似** (upper approximation) $\overline{A}(D_i)$ をそれぞれ以下の通り定義する．

$$\underline{A}(D_i) \overset{\text{def}}{=} \{x \in U \mid [x]_A \subseteq D_i\}, \qquad (2)$$
$$\overline{A}(D_i) \overset{\text{def}}{=} \{x \in U \mid [x]_A \cap D_i \neq \emptyset\}. \qquad (3)$$

識別不能関係 R_A の定義より，D_i の下近似 $\underline{A}(D_i)$ は，A に含まれる属性の値によって，確実に D_i に分類される対象の集合となる．これに対して，上近似 $\overline{A}(D_i)$ は D_i に分類される可能性がある対象の集合となる．すべての決定クラス D_i に対して $\underline{C}(D_i) = D_i = \overline{C}(D_i)$ となる決定表を無矛盾な決定表と呼び，そうでない決定表を矛盾を含む決定表と呼ぶ．

例 1 決定表の例を表 1 に示す．表 1 は議論の対象となる要素の集合 $U = \{x_1, \cdots, x_6\}$，条件属性集合 $C = \{c_1, \cdots, c_6\}$ 決定属性集合 $D = \{d\}$ などで構成され，$\rho(x_i, d) = i$ となる要素の集合を決定クラス D_i とすると，3 個の決定クラス $D_1 = \{x_1, x_2, x_5\}$ および $D_2 = \{x_3, x_4\}$，$D_3 = \{x_6\}$ が得られる．

2.2 決定ルールとその評価指標

本稿では条件属性集合 $B \subseteq C$ および決定属性の集合 D，対象 $x \in U$ から得られる決定ルールを

$$(B, x) \to (D, x)$$

と表す．決定ルールの評価指標として，以下の確信度

$$Cer((B, x) \to (D, x)) \overset{\text{def}}{=} \frac{|[x]_B \cap D_i|}{|[x]_B|} \qquad (4)$$

および被覆度

$$Cov((B, x) \to (D, x)) \overset{\text{def}}{=} \frac{|[x]_B \cap D_i|}{|D_i|} \qquad (5)$$

を用いる．ここで，$|X|$ は集合 X の要素の個数である．また，集合 D_i は対象 $x \in U$ が含まれる決定クラスである．確信度は決定ルールの正確さを，被覆度は決定ルールの一般性を表す指標である．なお，被覆度は C.I. 値とも呼ばれる．

例 2 表 1 において，$B = \{c_2, c_3\}$ および $D = \{d\}$，対象 x_1 から構成される決定ルール $(B, x_1) \to (D, x_1)$ は

$$(c_2 = 1) \wedge (c_3 = 1) \to (d = 1)$$

となり，確信度は 1，被覆度は $\frac{1}{3}$ である．

2.3 相対縮約

データから規則性を見出す観点から，できるだけ少ない属性数で，条件属性 C をすべて用いた識別不能関係 R_C による分類と同等な分類を与え，すべての決定クラスを近似できることが望ましい．そのような性質を満たす条件属性の集合 $A \subseteq C$ を相対縮約と呼ぶ．形式的には，分割 $\mathcal{D}$ の C に関する**相対縮約** (relative reduct) とは，すべての決定クラス $D_i \in \mathcal{D}$ $(i = 1, \cdots, m)$ に対して以下の 2 条件を満たす条件属性の部分集合 $A \subseteq C$ である．

1. $\underline{A}(D_i) = \underline{C}(D_i)$.
2. 任意の $B \subset A$ に対して $\underline{B}(D_i) \neq \underline{C}(D_i)$.

ここで，$B \subset A$ は集合 B が集合 A の真部分集合であることを意味する．

条件 1 は，相対縮約 A に基づく下近似が，すべての条件属性 C を用いた場合の下近似と一致することを要請している．条件 2 は，A に含まれる属性を 1 つでも取り除くと，C を用いた下近似と一致しなくなることを要請している．これらの条件より，相対縮約は，条件属性 C と同等な分類を与える必要最小限の属性集合となる．

例 3 表 1 には以下の 7 種類の相対縮約が存在する:
1: $\{c_1, c_5\}$, 2: $\{c_3, c_5\}$, 3: $\{c_5, c_6\}$, 4: $\{c_1, c_2, c_3\}$, 5: $\{c_1, c_2, c_4\}$, 6: $\{c_1, c_2, c_6\}$, 7: $\{c_2, c_4, c_5\}$.

2.4 可変精度ラフ集合モデル

可変精度ラフ集合モデルは，Pawlak による近似の定義を緩め，矛盾を含む情報および例外を含むルールなどをラフ集合の枠組みで扱う基礎的な手法を与えるモデルである．

決定表 $(U, C \cup D, V, \rho)$ が与えられているとする．任意の部分集合 $X, Y \subseteq U$ に対して，誤判別度 $c(X, Y)$ を以下の式で定義する．

$$c(X,Y) \stackrel{\text{def}}{=} \begin{cases} 1 - \dfrac{|X \cap Y|}{|X|}, & |X| > 0 \text{ のとき}, \\ 0, & \text{それ以外}. \end{cases} \tag{6}$$

誤判別度 $c(X, Y)$ は，X のすべての要素を Y に分類した場合に生じる誤分類の割合を表す．誤判別度 $c(X, Y)$ を用いると，X が Y に完全に包含されるのは $c(X, Y) = 0$ のとき，かつそのときのみであることがわかる．

$$X \subseteq Y \Longleftrightarrow c(X, Y) = 0. \tag{7}$$

よって，許容する誤分類の度合いを表す**精度** (precision) β $(0 \leq \beta < 0.5)$ を用いることで，包含関係を以下のように拡張することができる．

$$X \overset{\beta}{\subseteq} Y \overset{\text{def}}{\Longleftrightarrow} c(X, Y) \leq \beta. \tag{8}$$

R_A を集合 $A \subseteq C$ に基づく識別不能関係，U/R_A を R_A による同値類の集合とする．決定クラス D_i の R_A による **β-下近似** (β-lower approximation) $\underline{A}_\beta(D_i)$ および **β-上近似** (β-upper approximation) $\overline{A}_\beta(D_i)$ をそれぞれ以下の通り定義する．

$$\begin{aligned} \underline{A}_\beta(D_i) &\stackrel{\text{def}}{=} \bigcup \{[x]_A \in U/R_A \mid [x]_A \overset{\beta}{\subseteq} D_i\} && (9) \\ &= \{x \in U \mid c([x]_A, D_i) \leq \beta\}, && (10) \\ \overline{A}_\beta(D_i) &\stackrel{\text{def}}{=} \{x \in U \mid c([x]_A, D_i) < 1 - \beta\}. && (11) \end{aligned}$$

Pawlak による下近似および上近似の定義は，この定義における $\beta = 0$ の場合に相当する．よって，可変精度ラフ集合モデルは Pawlak によるラフ集合の拡張と見なすことができる．

3 同値類による分割に基づく相対縮約の評価

本節では，著者ら [4] による同値類による分割を用いた相対縮約の評価手法について概説する．相対縮約の評価の方針として，「粗い」分割を与える相対縮約ほど高く評価する．分割を用いた相対縮約の評価手法では，相対縮約から得られる決定ルールに対する評価基準を用いて，相対縮約の評価を行う．

まず，条件属性集合 $B \subseteq C$ に基づく識別不能関係 R_B による同値類 $[x]_B$ に対して，以下の集合を定義する:

$$Dec([x]_B) \stackrel{\text{def}}{=} \{D_i \in U/R_D \mid D_i \cap [x]_B \neq \emptyset\}. \tag{12}$$

集合 $Dec([x]_B)$ は，決定表から抽出できる決定ルールにおいて，式 (B, x) を条件部に持つ決定ルールの結論部に対応する．よって，すべての同値類 $[x]_B$ に対して集合 $Dec([x]_B)$ を求め，その要素の個数を合計した値

$$N_R \stackrel{\text{def}}{=} \sum_{[x]_B \in U/R_B} |Dec([x]_B)| \tag{13}$$

は，条件属性集合 B および決定属性集合 D から得られる決定ルールの総数となる．B および D から得られる決定ルールの確信度および被覆度の平均値について以下の定理が得られる [4].

定理 1 与えられた決定表における，任意の空でない条件属性集合 $B \subseteq C$ および決定属性集合 D に対して，属性集合 B から得られるすべての決定ルール $(B, x) \rightarrow (D, x)$ $(\forall x \in U)$ の確信度の平均値 $ACer(B)$ は次式で求められる:

$$ACer(B) = \frac{|U/R_B|}{N_R}. \tag{14}$$

また，被覆度の平均値 $ACov(B)$ は次式で求められる:

$$ACov(B) = \frac{|U/R_D|}{N_R}. \tag{15}$$

定理 1 で条件属性集合として相対縮約を用いると，無矛盾な決定表の場合，確信度の平均値はどの相対縮約でも 1 となる．一方，被覆度の平均値は，相対縮約を用いて構成された分割に含まれる同値類の個数が少ない (分割が「粗い」) ほど，高い値を取る．以上より，相対縮約から得られる決定ルールの被覆度の平均値を，その相対縮約の評価値として用いる．

例 4 表 1 の相対縮約 $\{c_1, c_5\}$ からは以下の 4 個の決定ルールが得られる:

- $(c_1 = 1) \wedge (c_5 = 1) \rightarrow (d = 1)$, 確信度 1, 被覆度 $\frac{1}{3}$,
- $(c_1 = 2) \wedge (c_5 = 1) \rightarrow (d = 1)$, 確信度 1, 被覆度 $\frac{2}{3}$,
- $(c_1 = 2) \wedge (c_5 = 2) \rightarrow (d = 2)$, 確信度 1, 被覆度 1,
- $(c_1 = 1) \wedge (c_5 = 2) \rightarrow (d = 3)$, 確信度 1, 被覆度 1.

確信度の平均値は $(1+1+1+1)/4 = 1$ となり，定理 1 による「同値類の個数/決定ルールの個数」の値と一致する．また，被覆度の平均値は $(\frac{1}{3} + \frac{2}{3} + 1 + 1)/4 = \frac{3}{4}$ となり，定理 1 による「決定クラスの個数/決定ルールの個数」の値と一致する．以上より，相対縮約 $\{c_1, c_5\}$ の評価値は $\frac{3}{4}$ である．すべての相対縮約について確信度および被覆度の平均値を求めた結果を表 2 に示す．この結果より，相対縮約 $\{c_5, c_6\}$ が最も「粗い」分割を与える相対縮約と見なされる．実際に，$\{c_5, c_6\}$ から得られる分割は，決定属性集合 D による分割 (決定クラスの集合) U/R_D と一致する．

表 2: 相対縮約から得られる決定ルールの確信度および被覆度の平均値

相対縮約	確信度の平均値	被覆度の平均値
$\{c_1, c_5\}$	1	0.75
$\{c_3, c_5\}$	1	0.75
$\{c_5, c_6\}$	1	1
$\{c_1, c_2, c_3\}$	1	0.5
$\{c_1, c_2, c_4\}$	1	0.6
$\{c_1, c_2, c_6\}$	1	0.6
$\{c_2, c_4, c_5\}$	1	0.6

4 可変精度ラフ集合モデルにおける確信度および被覆度の平均値に関する性質

本節では，前節で述べた Pawlak ラフ集合における確信度および被覆度の平均値に関する性質を，可変精度ラフ集合モデルに拡張することを試みる．

4.1 問題設定

式 (12) が示すように，著者らは Pawlak ラフ集合において決定ルール $(B, x) \to (D, x)$ が決定表から抽出可能であることを，

$$\text{確信度が 0 より大きい} \Leftrightarrow x \in \overline{B}(D_i)$$

と設定した．β-上近似の定義より以下の性質:

$$x \in \overline{B}_\beta(D_i) \Longleftrightarrow \frac{|[x]_B \cap D_i|}{|[x]_B|} > \beta$$

が成り立つことに注意すると，可変精度ラフ集合モデルにおいて，精度 β の下で式 (B, x) を条件部に持つ抽出可能な決定ルールは，「確信度が β より大きい決定ルール」に相当する．以上より，可変精度ラフ集合モデルにおいて前述の定理 1 に相当する性質を導くことは，以下の問題を解くことに帰着させることができる．

> **問題**　与えられた決定表および精度 β $(0 \leq \beta < 0.5)$ の元で，空でない条件属性集合 $B \subseteq C$ および決定属性集合 D から得られる，確信度が β より大きいすべての決定ルールの確信度および被覆度の平均値を求めよ．

4.2 定理 1 の可変精度ラフ集合モデルへの拡張

定理 1 で示した確信度および被覆度の平均値に関する性質は，以下の性質が成り立つことに依拠している．

命題 1 与えられた決定表における，任意の空でない条件属性集合 $B \subseteq C$ および決定属性集合 D，対象 $x \in U$ に対して以下の性質が成り立つ:

- すべての $y \in [x]_B$ に対して，決定ルール $(B, y) \to (D, y)$ の確信度の和は 1 になる．
- すべての $z \in [x]_D$ に対して，決定ルール $(B, z) \to (D, z)$ の被覆度の和は 1 になる．

しかし，上述の問題設定の下で精度が β より大きいルールに限定すると，同様の性質は一般的に成り立たない．

例 5 表 1 において条件属性集合 $B = \{c_3\}$ および決定属性集合 $D = \{d\}$，対象 x_6 から $[x_6]_B = \{x_1, x_2, x_6\}$ および $[x_6]_D = \{x_6\}$ を得るが，精度 $\beta = \frac{1}{3}$ とすると以下の 2 個のルール

- $(c_3 = 1) \to (d = 1)$, 確信度 $\frac{2}{3}$, 被覆度 $\frac{2}{3}$ (x_1，x_2 に対応),
- $(c_3 = 1) \to (d = 3)$, 確信度 $\frac{1}{3}$, 被覆度 1 (x_6 に対応)

ではなく，x_1，x_2 に対応するルールしか抽出されないため，式 $(c_3 = 1)$ を条件部とする決定ルールの確信度の総和は 1 にならない．同様に，式 $(d = 3)$ を結論部とする決定ルールの被覆度の総和も 1 にならない．

以上の例より，可変精度ラフ集合モデルにおいて決定ルールの確信度および被覆度の平均値を考慮する場合は，Pawlak ラフ集合の場合と比較して，抽出可能な決定ルールの個数の減少および確信度，被覆度の総和の減少を考慮する必要がある．そのため，これらに対してそれぞれ減少の程度を表す係数を導入する．

まず，条件属性集合 $B \subseteq C$ に基づく識別不能関係 R_B による同値類 $[x]_B$ に対して，以下の集合を定義する:

$$Dec_\beta([x]_B) \stackrel{\text{def}}{=} \left\{ D_i \in U/R_D \;\middle|\; \frac{|[x]_B \cap D_i|}{|[x]_B|} > \beta \right\}. \tag{16}$$

集合 $Dec_\beta([x]_B)$ は，式 (B, x) を条件部に持ち，かつ確信度が β より大きい決定ルールの結論部に対応する．

集合 $Dec_\beta([x]_B)$ および式 (12) を用いて，決定ルールの個数の減少を表す係数 L_R を以下の通り定義する:

$$L_R \stackrel{\text{def}}{=} \sum_{[x]_B \in U/R_B} |Dec([x]_B) - Dec_\beta([x]_B)|. \tag{17}$$

式 (12) および式 (16) より明らかに

$$Dec_\beta([x]_B) \subseteq Dec([x]_B),\ \forall [x]_B \in U/R_B$$

が成り立つため，$N_R \geq L_R$ を得る．ここで，N_R は式 (13) で定義される値であり，条件属性集合 B および決定属性集合 D から得られる確信度が 0 より大きい決定ルールの総数である．

例5における対象 x_6 のように，精度 β の下では，空でない条件属性集合 B および決定属性集合 D を用いて対応する決定ルールを抽出することができない対象の集合を，次式で定義する:

$$DC(B,D) \stackrel{\text{def}}{=} \left\{ y \in U \;\middle|\; \begin{array}{l} \forall [x]_B \in U/R_B, \\ D_i \in Dec_\beta([x]_B) \Rightarrow y \notin D_i \end{array} \right\}. \tag{18}$$

集合 $DC(B,D)$ は，確信度が β より大きいいかなる決定ルールにおいても，該当する結論部が存在しない対象の集まりである．この定義より，$y \in DC(B,D)$ である対象 y について，決定ルール $(B,y) \to (D,y)$ の確信度は β 以下となるため，この決定ルールは4.1節で述べた問題設定では対象外となる．よって，確信度および被覆度の総和を計算する際に，このようなルールの確信度および被覆度は除外する必要があるため，これらの総和をそれぞれ確信度の減少を表す係数 L_{Cer} および被覆度の減少を表す係数 L_{Cov} として以下の通り定義する:

$$\begin{aligned} L_{Cer} &\stackrel{\text{def}}{=} \sum_{[y]_B \in \mathcal{X}} \sum_{[y]_D \in \mathcal{Y}} Cer((B,y) \to (D,y)) && (19) \\ &= \sum_{[y]_B \in \mathcal{X}} \sum_{[y]_D \in \mathcal{Y}} \frac{|[y]_B \cap [y]_D|}{|[y]_B|} && (20) \\ &= \sum_{[x]_B \in U/R_B} \frac{|[x]_B \cap DC(B,D)|}{|[x]_B|}. && (21) \end{aligned}$$

ここで，集合 $\mathcal{X} \stackrel{\text{def}}{=} \{[y]_B \in U/R_B \mid y \in DC(B,D)\}$ は除外する必要がある決定ルールの条件部に対応する同値類の集合，集合 $\mathcal{Y} \stackrel{\text{def}}{=} \{[y]_D \in U/R_D \mid y \in DC(B,D)\}$ は除外する必要がある決定ルールの結論部に対応する同値類の集合をそれぞれ表す.

$$\begin{aligned} L_{Cov} &\stackrel{\text{def}}{=} \sum_{[y]_B \in \mathcal{X}} \sum_{[y]_D \in \mathcal{Y}} Cov((B,y) \to (D,y)) && (22) \\ &= \sum_{[y]_B \in \mathcal{X}} \sum_{[y]_D \in \mathcal{Y}} \frac{|[y]_B \cap [y]_D|}{|[y]_D|} && (23) \\ &= \sum_{D_i \in U/R_D} \frac{|D_i \cap DC(B,D)|}{|D_i|}. && (24) \end{aligned}$$

以上の議論をまとめると，以下の定理を得る．

定理 2 与えられた決定表および精度 β $(0 \le \beta < 0.5)$ に対して，任意の空でない条件属性集合 $B \subseteq C$ および決定属性集合 D から得られる，確信度が β より大きいすべての決定ルール $(B,x) \to (D,x)$ の確信度の平均値 $ACer_\beta(B)$ は次式で求められる:

$$ACer_\beta(B) = \begin{cases} \dfrac{|U/R_B| - L_{Cer}}{N_R - L_R}, & \text{if } N_R \neq L_R, \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

また，被覆度の平均値 $ACov_\beta(B)$ は次式で求められる:

$$ACov_\beta(B) = \begin{cases} \dfrac{|U/R_D| - L_{Cov}}{N_R - L_R}, & \text{if } N_R \neq L_R, \\ 0, & \text{otherwise.} \end{cases} \tag{26}$$

更に，$\beta = 0$ の場合は以下が成り立つ:

$$ACer_0(B) = ACer(B), \quad ACov_0(B) = ACov(B).$$

例 6 表1の決定表において，条件属性集合 $B = \{c_3, c_4\}$ および決定属性集合 $D = \{d\}$，精度 $\beta = \frac{1}{3}$ とする．定理2を用いて，B および D から得られる，確信度が β より大きいすべての決定ルールの確信度および被覆度の平均値を求める．

まず，同値類の集合 U/R_B の各同値類は以下のようになり，$|U/R_B| = 4$ となる:

$[x_1]_B = \{x_1, x_2, x_6\}$, $[x_3]_B = \{x_3\}$,
$[x_4]_B = \{x_4\}$, $[x_5]_B = \{x_5\}$.

また，決定クラスは例1と同様であり，$|U/R_D| = 3$ となる．

次に，各同値類 $[x]_B \in U/R_B$ に対して，集合 $Dec([x]_B)$ および $Dec_\beta([x]_B)$ をそれぞれ以下の通り構成する:

$$\begin{array}{ll} Dec([x_1]_B) = \{D_1, D_3\}, & Dec_\beta([x_1]_B) = \{D_1\}, \\ Dec([x_3]_B) = \{D_2\}, & Dec_\beta([x_3]_B) = \{D_2\}, \\ Dec([x_4]_B) = \{D_2\}, & Dec_\beta([x_4]_B) = \{D_2\}, \\ Dec([x_5]_B) = \{D_1\}, & Dec_\beta([x_5]_B) = \{D_1\}. \end{array}$$

これらの集合を用いて，式 (13) の N_R は以下の通り求められる:

$$N_R = \sum_{[x]_B \in U/R_B} |Dec([x]_B)| = 2 + 1 + 1 + 1 = 5.$$

また，式 (17) の L_R は以下の通り求められる:

$$\begin{aligned} L_R &= \sum_{[x]_B \in U/R_B} |Dec([x]_B) - Dec_\beta([x]_B)| \\ &= (2-1) + (1-1) + (1-1) + (1-1) \\ &= 1. \end{aligned}$$

更に，B および D では対応する決定ルールを抽出することができない対象の集合は，決定クラス D_1 および D_2 に含まれない対象の集合となるため，以下のようになる:

$$DC(B,D) = \{x_6\}.$$

得られた集合 $DC(B,D)$ を用いて，式 (21) および式 (24) に基づいて係数 L_{Cer} および L_{Cov} をそれぞれ以下の通

り求める:

$$\begin{aligned} L_{Cer} &= \sum_{[x]_B \in U/R_B} \frac{|[x]_B \cap DC(B,D)|}{|[x]_B|} \\ &= \frac{1}{3} + \frac{0}{1} + \frac{0}{1} + \frac{0}{1} \\ &= \frac{1}{3}, \\ L_{Cov} &= \sum_{D_i \in U/R_D} \frac{|D_i \cap DC(B,D)|}{|D_i|} \\ &= \frac{0}{3} + \frac{0}{1} + \frac{1}{1} \\ &= 1. \end{aligned}$$

以上をまとめると，定理 2 より，B および D から得られる，確信度が β より大きいすべての決定ルールの確信度および被覆度の平均値はそれぞれ以下の通り求められる:

$$\begin{aligned} ACer_\beta(B) &= \frac{|U/R_B| - L_{Cer}}{N_R - L_R} \\ &= \frac{4 - \frac{1}{3}}{5-1} \\ &= \frac{11}{12}, \\ ACov_\beta(B) &= \frac{|U/R_D| - L_{Cov}}{N_R - L_R} \\ &= \frac{3-1}{5-1} \\ &= \frac{1}{2}. \end{aligned}$$

実際に，属性集合 $\{c_3, c_4\}$ および $\beta = \frac{1}{3}$ からは以下の 4 個の決定ルールが得られる:

- $(c_3 = 1) \wedge (c_4 = 1) \rightarrow (d = 1)$, 確信度 $\frac{2}{3}$, 被覆度 $\frac{2}{3}$,
- $(c_3 = 2) \wedge (c_4 = 1) \rightarrow (d = 1)$, 確信度 1, 被覆度 $\frac{1}{2}$,
- $(c_3 = 2) \wedge (c_4 = 2) \rightarrow (d = 2)$, 確信度 1, 被覆度 $\frac{1}{2}$,
- $(c_3 = 3) \wedge (c_4 = 1) \rightarrow (d = 3)$, 確信度 1, 被覆度 $\frac{1}{3}$.

確信度の平均値は $(\frac{2}{3} + 1 + 1 + 1)/4 = \frac{11}{12}$ となり，定理 2 による値と一致する．また，被覆度の平均値は $(\frac{2}{3} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3})/4 = \frac{1}{2}$ となり，定理 2 による値と一致する．

5 おわりに

本研究では可変精度ラフ集合モデルにおいて，与えられた決定表および精度 β $(0 \leq \beta < 0.5)$ に対して，任意の空でない条件属性集合 $B \subseteq C$ および決定属性集合 D から得られる，確信度が β より大きいすべての決定ルールの確信度の平均値および被覆度の平均値に関する性質を導出した．導出した性質は同値類の個数および決定クラスの個数，得られる決定ルールの個数に加え，得られる決定ルールの個数の減少および確信度の総和の減少，被覆度の総和の減少に関する項を含み，$\beta = 0$ の場合は Pawlak ラフ集合に対する同様の性質と一致する．

今後の課題として，本研究で導出した性質の更なる精緻化および計算方法の計算量の考察，可変精度ラフ集合モデルにおける縮約の評価指標への応用，実データに対する応用事例の検証などが挙げられる．

参考文献

[1] Beynon, M.: Reducts within the variable precision rough set model: A further investigaton, *European Journal of Operations Research*, Vol. 134, pp. 592–605, 2001.

[2] Inuiguchi, M.: Several approaches to attribute reduction in variable precision rough set model, *Proceedings of MDAI 2005*, Springer-Verlag, pp.215–226, 2005.

[3] 工藤 康生, 村井 哲也: ラフ集合によるルール生成, 第 16 回あいまいと感性研究部会ワークショップ講演論文集, pp.18–23, 2006.

[4] 工藤 康生, 村井 哲也: 同値類による分割を用いた相対縮約の評価について, 感性フォーラム札幌 2008 講演論文集, 2008.

[5] 森 典彦, 田中 英夫, 井上 勝雄 (共編): ラフ集合と感性〜データからの知識獲得と推論〜, 海文堂出版, 2004.

[6] Pawlak, Z.: Rough Sets, *International Journal of Computer and Information Science*, Vol. 11, pp.341–356, 1982.

[7] Pawlak, Z.: *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, 1991.

[8] Ziarko, W.: Variable Precision Rough Set Model, *Journal of Computer and System Science*, Vol. 46, pp.39–59, 1993.

連絡先

工藤 康生

〒 050-8585 北海道室蘭市水元町 27-1

室蘭工業大学工学部情報工学科

Tel: 0143–46–5469, Fax: 0143–46–5499

E-mail: `kudo@csse.muroran-it.ac.jp`