

# 多重属性を持つラベル集合を用いた データの記述

葛西 正裕

- 1 はじめに
  - 2 ラベル集合の順序
  - 3 多重属性を持つラベル集合の順序
  - 4 多重ラベル集合の順序による記述
  - 5 多重ラベル集合の順序の性質
  - 6 おわりに
- 参考文献

## 【概要】

多様化したデータに対してデータの種類の問わず一元的に分析に供するためには、データを適切に構成しておく必要があり、分類に用いる属性ごとに階層的に分類しておくことが有用である。分類後に、分析対象となるデータを複数の属性で指定する場合、多重属性を持つラベル集合を与えてデータを記述することになる。データを記述の際、与えられたラベル集合とデータのラベル（集合ラベル）の順序を用いることで柔軟なデータの記述が可能になる。本稿では、単一属性で用いたラベル集合の順序を多重属性に拡張することで、多重属性を持つラベル集合による多重属性を持つ集合ラベルのデータに対する記述法を明らかにし、それが十分な記述能力を有することを示す。また、多重ラベル集合の上下関係とそれらによって記述されるデータの包含関係が一致するという性質と多重ラベル集合が異なれば記述されるデータも異なるという性質についても検討する。

## 【キーワード】

階層的分類, 多重属性, 多重ラベル集合の順序, データの記述, データベースモデル, 多次元データベース, データウェアハウス

## 1 はじめに

複雑さを増す経済現象に対して適切な分析を行うためには、収集したデータを高度に分析する手法が必要である。数値データ、画像、動画、テキストといった多様化したデータに対して、データの種別を問わず一元的に分析に供するためには、データを適切に構成しておく必要がある<sup>[2][10]</sup>、データを階層的に分類しておくことが有用である<sup>[4][23]</sup>。データの分類は、一般に分類に用いる属性ごとに分類され、分類階層において該当するカテゴリのラベルが付される<sup>[3][5][21]</sup>。分類後は、複数の分類階層を複合的に用いることによって、分析対象となるデータを指定することになる<sup>[11][17]</sup>。例えば、データを地域、業種といった属性ごとに分類を行った上で、地域に関する分類階層に分類されているデータと業種に関する分類階層に分類されているデータを要求することで、愛知県の自動車に関するデータといったものを指定し分析に供されることになる。

複数の属性で分析対象となるデータを指定するには、属性ごとに分類された分類階層のデータに対し、複数の属性（多重属性）を持つラベル集合を与えてデータを記述することになる。多重属性のラベル集合が記述するデータは、一般に、ラベル集合におけるすべてのラベルに対してそれ以下のラベルを持つようなデータであると解釈されるが、その他にも様々な解釈がある。例えば、{東海, 九州, 製造業}という多重属性のラベル集合が与えられた際、{愛知, 福岡, 自動車}といったラベルのデータが該当する。それ以外にも、{東海, 九州, 製造業}は地域と業種という2種類の属性を持つので、そのうちのどちらか一方の属性を持つデータであればよいがその属性であればすべてのラベルに関わっているデータと解釈する場合は、{愛知, 福岡}や{電機機器, 自動車}といったラベルのデータが該当することになる。また、地域と業種の両方の属性があれば他の属性を持っていてもよい場合などもあり、多重属性のラベル集合が記述するデータの種類には様々な解釈がある。

複数の属性を複合的に扱う研究では、データウェアハウスのデータモデルとして多次元データベースが考案されている。多次元データベースに関する研究は、数値データの集約方法やそれを効率的に行う手法の研究が中心であり<sup>[1][19]</sup>、分析対象のデータは数値のみを対象としているのに対して、本稿では生データを対象としている点で異なる。本稿が対象とするような生データを階層的に分類するといった研究も多いが、主な研究の目的は、分類過程の自動化に重点が置かれ、データにラベルを自動的に付す研究や分類に用いる分類階層の構成の自動化に関する研究が多い<sup>[7][9][20][24]</sup>。ラベル集合を用いて分析対象となるデータを指定する研究として、キーワードの集合を用いてデータを抽出する研究もあるが<sup>[6][8]</sup>、キーワードとデー

タの関連性を定量的に測定するものである。これらの研究では、多重属性のラベル集合によるデータの指定はラベル集合の各ラベルで記述されるデータの和集合や積集合といった単純な指定に留まっており、分析対象となるデータを指定するためにラベル集合でどのようなデータを記述できるのかについて精緻な検討を行っていない。

データの指定を精緻に行うという点では、述語論理等が多くに用いられる一方で、単一の属性の分類階層において複数のカテゴリに分類されているデータを対象とする場合、与えられたラベル集合とデータのラベル（集合ラベル）の上下関係、すなわち順序を用いてデータを記述することで柔軟なデータの指定が可能になる<sup>[13] [14] [18]</sup>。よって、本稿においても順序を用いたデータの記述について考えるものとし、単一属性でデータの記述に用いる順序を多重属性に拡張することで、多重属性を持つラベル集合による多重属性を持つ集合ラベルのデータに対する記述法を明らかにする。これにより、属性ごとの分類階層のデータの積集合や和集合では記述できない種類のデータも記述が可能になる。また、本稿で明らかにする記述法が分析対象となるデータの指定に十分な記述法であることについても述べる。さらに、ラベル集合を用いてデータの記述を行う際、ラベル集合の上下関係とそれらによって記述されるデータの包含関係が一致するという性質（健全性）とラベル集合が異なれば記述されるデータも異なるという性質（妥当性）が有用であり、これらの性質についても多重属性のラベル集合に拡張して検討する。

本稿は以下のように構成される。2節では、単一属性におけるデータの記述に必要な順序を導き、3節において多重属性に拡張する。4節では、3節で明らかにした多重属性のラベル集合の順序によって記述されるデータを示した上で、記述能力について述べる。5節では、単一属性のラベル集合を用いたデータの記述において有用な健全性と妥当性について、多重属性のラベル集合に拡張した場合を検討する。6節はまとめである。

## 2 ラベル集合の順序

データの分類はそれに用いる属性に基づいて階層的に行われる。例えば、業種という属性では、製造業、輸送機器、自動車といった階層的な分類が行われる。複数の属性に関する分類では、一般に属性ごとに分類した複数の分類階層を組み合わせで利用する。本稿は、属性ごとに分類階層が予め与えられたものとし、本節では、単一属性の下での分類を考え、次節以降において多重属性による分類に拡張する。

1件のデータをオブジェクト $o$ 、オブジェクトの分類に用いられるラベルを $L$ とする。 $L$ によって記述されるオブジェクト集合を $\bar{L}$ で表し、オブジェクト $o$ が分類

に用いられる属性に対して持つ意味を $\tilde{o}$ で表す。オブジェクトは、あらかじめ与えられた分類階層においてオブジェクトの意味に対応する最も下位のカテゴリ（重複分類では複数カテゴリ）に分類される<sup>[15]</sup>。よって、 $\tilde{o}$ は、 $o$ が分類される最も下位のカテゴリのラベル（重複分類ではラベル集合）である。オブジェクトは分類階層の最下位でないカテゴリに分類されることもある<sup>[12] [22]</sup>。例えば、自動車のみならず航空機や船舶を含む輸送機器全般に関するオブジェクトのラベルは、最下位のカテゴリが自動車や航空機といったより詳細な産業分類があるような分類階層を用いて分類を行った場合、そのオブジェクトのラベルは最下位レベルではない。

ラベル $L_1$ と $L_2$ に対して、 $L_2$ が $L_1$ の上位概念のラベルならば、 $L_2$ は $L_1$ の上位（ $L_1$ は $L_2$ の下位）であり、 $L_1 \prec L_2$ で示す。また、 $L_2$ が $L_1$ の上位概念または等しい概念のラベルならば、 $L_2$ は $L_1$ 以上であり、 $L_1 \preceq L_2$ で表す。オブジェクトのラベルが単数ならば、オブジェクトが $\bar{L}$ の要素であることは、オブジェクトのラベルによって決定され、 $\tilde{o} \preceq L$ ならば $o$ は $\bar{L}$ の要素となる。すなわち、 $\bar{L} = \{o \mid \tilde{o} \preceq L\}$ である。

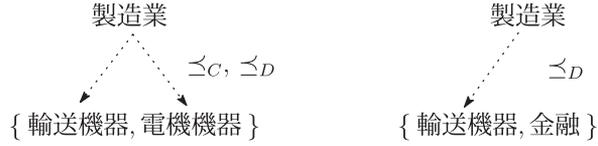
オブジェクトが複数のカテゴリに分類される時、オブジェクトのラベルはラベル集合になり、そのラベルを集合ラベルと呼ぶ。集合ラベルのオブジェクト $o$ に対して、 $\bar{L}$ を決めるには、単一ラベルとラベル集合の順序が必要である。ラベル集合を $L$ で表すと、 $L$ は一般に積または和で解釈され、それに応じた順序となる。

1. 積 (Conjunction) の順序：ラベル $L$ とラベル集合 $L$ に対し、 $L$ におけるすべてのラベルが $L$ 以下ならば、 $L$ は $L$ の下位であり、 $L \preceq_C L$ で表す。
2. 和 (Disjunction) の順序：ラベル $L$ とラベル集合 $L$ に対し、 $L$ におけるあるラベルが $L$ 以下ならば、 $L$ は $L$ の下位であり、 $L \preceq_D L$ で表す。

[例 1] 図1は、ラベル集合 {輸送機器, 電機機器} と {輸送機器, 金融} が“製造業”の下位であることを示している。“製造業”から“輸送機器”と“電機機器”への矢印は、分類階層で“製造業”が“輸送機器”や“電機機器”の上位概念であることを表している。積の順序では、“輸送機器”と“電機機器”は“製造業”の下位なので {輸送機器, 電機機器} は“製造業”の下位になるが、{輸送機器, 金融} の“金融”は“製造業”の下位ではないので {輸送機器, 金融} は“製造業”の下位ではない。和の順序では、{輸送機器, 電機機器} と {輸送機器, 電機機器} はともに“製造業”の下位のラベルを含むので“製造業”の下位である。

□

図1 単一ラベルとラベル集合の順序



オブジェクトの記述に用いるラベルをラベル集合に拡張する。ラベル集合  $\mathbf{L}$  は一般に  $\mathbf{L}$  の要素によって記述されるオブジェクト集合の積集合 (Intersection) や和集合 (Union) として解釈される。よって、ラベル集合  $\mathbf{L}$  の要素が積の順序で記述するオブジェクト集合の積集合を  $\overline{\mathbf{L}}^{CI} = \bigcap_{L \in \mathbf{L}} \{o \mid \tilde{o} \preceq_C L\}$ , 和集合を  $\overline{\mathbf{L}}^{CU} = \bigcup_{L \in \mathbf{L}} \{o \mid \tilde{o} \preceq_C L\}$ ,  $\mathbf{L}$  の要素が和の順序で記述するオブジェクト集合の積集合を  $\overline{\mathbf{L}}^{DI} = \bigcap_{L \in \mathbf{L}} \{o \mid \tilde{o} \preceq_D L\}$ , 和集合を  $\overline{\mathbf{L}}^{DU} = \bigcup_{L \in \mathbf{L}} \{o \mid \tilde{o} \preceq_D L\}$  とする。

ラベル集合  $\mathbf{L}$  で記述されるオブジェクト集合は  $\mathbf{L}$  以下の集合ラベルのオブジェクトからなるとするには、ラベル集合の順序が必要になる。ラベル集合  $\mathbf{L}$  に対し、 $\overline{\mathbf{L}}^{CI}$ ,  $\overline{\mathbf{L}}^{CU}$ ,  $\overline{\mathbf{L}}^{DI}$ ,  $\overline{\mathbf{L}}^{DU}$  に含まれるオブジェクトは、それぞれラベル集合の順序を用いて表すことができる。

[定義 1] ラベル集合  $\mathbf{L}_1, \mathbf{L}_2$  に対し、

$$\forall L_2 \in \mathbf{L}_2, \forall L_1 \in \mathbf{L}_1, L_1 \preceq L_2 \iff \mathbf{L}_1 \preceq_{CI} \mathbf{L}_2$$

$$\exists L_2 \in \mathbf{L}_2, \forall L_1 \in \mathbf{L}_1, L_1 \preceq L_2 \iff \mathbf{L}_1 \preceq_{CU} \mathbf{L}_2$$

$$\forall L_2 \in \mathbf{L}_2, \exists L_1 \in \mathbf{L}_1, L_1 \preceq L_2 \iff \mathbf{L}_1 \preceq_{DI} \mathbf{L}_2$$

$$\exists L_2 \in \mathbf{L}_2, \exists L_1 \in \mathbf{L}_1, L_1 \preceq L_2 \iff \mathbf{L}_1 \preceq_{DU} \mathbf{L}_2$$

である。 □

[定理 1] ラベル集合  $\mathbf{L}$  に対し、

$$\overline{\mathbf{L}}^{CI} = \{o \mid \tilde{o} \preceq_{CI} \mathbf{L}\}$$

$$\overline{\mathbf{L}}^{CU} = \{o \mid \tilde{o} \preceq_{CU} \mathbf{L}\}$$

$$\overline{\mathbf{L}}^{DI} = \{o \mid \tilde{o} \preceq_{DI} \mathbf{L}\}$$

$$\overline{\mathbf{L}}^{DU} = \{o \mid \tilde{o} \preceq_{DU} \mathbf{L}\}$$

である。 □

(証明) 証明は文献<sup>[13]</sup>による。

(証明終)

一方、ラベル集合で単一ラベルによるオブジェクトの記述を考えた後に集合ラベルのオブジェクトに拡張する方法もある。単一ラベルでオブジェクトを記述するラベル集合は、オブジェクトの積集合と和集合、すなわち、 $\bigcap_{L \in \mathbf{L}} \overline{L} \cup \bigcup_{L \in \mathbf{L}} \overline{L}$  として解釈される。

積集合の解釈を、集合ラベルのオブジェクトに拡張するとき、オブジェクト  $o$  に対して、 $\tilde{o}$  のすべてのラベル  $L'$  が  $\mathbf{L}$  に対する積集合の性質を満たす、す

なわち、 $\bar{L}' \subseteq \bigcap_{L \in \mathbf{L}} \bar{L} = \bigcap_{L \in \mathbf{L}} \{o \mid \tilde{o} \preceq L\}$ であれば、 $o$ は $\mathbf{L}$ によって記述されるオブジェクトである。このようなオブジェクトの集合を $\bar{\mathbf{L}}^{IC}$ とすると、 $\bigcap_{L \in \mathbf{L}} \{o \mid \forall L' \in \tilde{o}, L' \preceq L\}$ となる。

同様に、ラベル集合 $\mathbf{L}$ を積集合とし、オブジェクトの集合ラベルを和とするオブジェクト集合、また、ラベル集合 $\mathbf{L}$ を和集合とし、オブジェクトの集合ラベルを積や和とするオブジェクト集合は、それぞれ、 $\bar{\mathbf{L}}^{ID} = \bigcap_{L \in \mathbf{L}} \{o \mid \exists L' \in \tilde{o}, L' \preceq L\}$ 、 $\bar{\mathbf{L}}^{UC} = \bigcup_{L \in \mathbf{L}} \{o \mid \forall L' \in \tilde{o}, L' \preceq L\}$ 、 $\bar{\mathbf{L}}^{UD} = \bigcup_{L \in \mathbf{L}} \{o \mid \exists L' \in \tilde{o}, L' \preceq L\}$ となる。

ラベル集合 $\mathbf{L}$ によって記述されるオブジェクト集合は $\mathbf{L}$ 以下の集合ラベルのオブジェクトなので、 $\bar{\mathbf{L}}^{IC}$ 、 $\bar{\mathbf{L}}^{ID}$ 、 $\bar{\mathbf{L}}^{UC}$ 、 $\bar{\mathbf{L}}^{UD}$ に対応する順序は次のものとなる。

[定義 2] ラベル集合 $\mathbf{L}_1$ と $\mathbf{L}_2$ に対し、

$$\begin{aligned} \forall L_1 \in \mathbf{L}_1, \forall L_2 \in \mathbf{L}_2, L_1 \preceq L_2 &\iff \mathbf{L}_1 \preceq_{IC} \mathbf{L}_2 \\ \exists L_1 \in \mathbf{L}_1, \forall L_2 \in \mathbf{L}_2, L_1 \preceq L_2 &\iff \mathbf{L}_1 \preceq_{ID} \mathbf{L}_2 \\ \forall L_1 \in \mathbf{L}_1, \exists L_2 \in \mathbf{L}_2, L_1 \preceq L_2 &\iff \mathbf{L}_1 \preceq_{UC} \mathbf{L}_2 \\ \exists L_1 \in \mathbf{L}_1, \exists L_2 \in \mathbf{L}_2, L_1 \preceq L_2 &\iff \mathbf{L}_1 \preceq_{UD} \mathbf{L}_2 \end{aligned}$$

である。 □

[定理 2] ラベル集合 $\mathbf{L}$ に対し、

$$\begin{aligned} \bar{\mathbf{L}}^{IC} &= \{o \mid \tilde{o} \preceq_{IC} \mathbf{L}\} \\ \bar{\mathbf{L}}^{ID} &= \{o \mid \tilde{o} \preceq_{ID} \mathbf{L}\} \\ \bar{\mathbf{L}}^{UC} &= \{o \mid \tilde{o} \preceq_{UC} \mathbf{L}\} \\ \bar{\mathbf{L}}^{UD} &= \{o \mid \tilde{o} \preceq_{UD} \mathbf{L}\} \end{aligned}$$

である。 □

(証明) 証明は文献<sup>[13]</sup>による。

(証明終)

8通りの順序を導出したが、 $\preceq_{ID}$ 、 $\preceq_{IC}$ 、 $\preceq_{CI}$ 、 $\preceq_{CU}$ は以下の理由でオブジェクトを記述する順序として議論する対象とはしない。ラベル集合 $\mathbf{L}_1$ と $\mathbf{L}_2$ に対して、 $\mathbf{L}_1 \preceq_{ID} \mathbf{L}_2$ または $\mathbf{L}_1 \preceq_{IC} \mathbf{L}_2$ であれば、 $\mathbf{L}_1$ のあるラベルまたは $\mathbf{L}_1$ のすべてのラベルが $\mathbf{L}_2$ のすべてのラベルの下位である。 $\mathbf{L}_2$ が互いに順序のないラベル $L_{21}$ と $L_{22}$  ( $L_{21} \not\prec L_{22}$ かつ $L_{22} \not\prec L_{21}$ )を含むとき、 $L \prec L_{21}$ かつ $L \prec L_{22}$ であるようなラベル $L$ は存在せず、 $\mathbf{L}_2$ によって記述されるオブジェクトは存在しない。 $\mathbf{L}_2$ にそのようなラベルがないとき、 $\mathbf{L}_2$ は $\mathbf{L}_2$ のうち最も下位の1つのラベルとすることができる。

$\preceq_{ID}$ や $\preceq_{IC}$ によってラベル集合 $\mathbf{L}$ でオブジェクトを記述する場合には、 $\bar{\mathbf{L}}^{ID} \neq \phi$ かつ $\bar{\mathbf{L}}^{IC} \neq \phi$ ならば $\mathbf{L}$ を1つのラベルに縮約できる。 $|\mathbf{L}|=1$ であれば、 $\bar{\mathbf{L}}^{ID} = \bar{\mathbf{L}}^{DI}$ 、 $\bar{\mathbf{L}}^{IC} = \bar{\mathbf{L}}^{UC}$ であり、 $\bar{\mathbf{L}}^{ID}$ と $\bar{\mathbf{L}}^{IC}$ は、それぞれ $\bar{\mathbf{L}}^{DI}$ と $\bar{\mathbf{L}}^{UC}$ の特殊な場合と考えることができる。したがって、 $\preceq_{ID}$ と $\preceq_{IC}$ は $\preceq_{DI}$ や $\preceq_{UC}$ とは異なる順序とする必要はない。また、 $\preceq_{CI}$ は $\preceq_{IC}$ と等しいので、同様に議論の対象とはしな

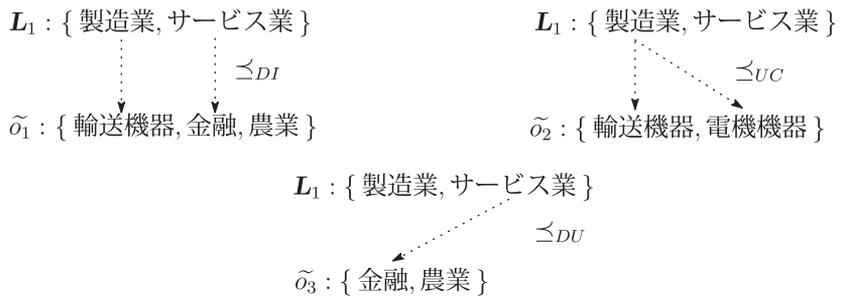
い。

$\preceq_{CU}$ については、ラベル集合 $\mathbf{L}$ におけるラベル $L_1$ と $L_2$ に対して、 $L_1 \not\preceq L_2$ かつ $L_2 \not\preceq L_1$ ならば $\overline{\{L_1\}}^{CU} \cap \overline{\{L_2\}}^{CU} = \phi$ であり、 $L_1 \preceq L_2$ ならば $\overline{\{L_1\}}^{CU} \subseteq \overline{\{L_2\}}^{CU}$ である。よって、 $\overline{\mathbf{L}}^{CU} = \bigcup_{L \in \mathbf{L}} \overline{\{L\}}^{CU}$ は、 $\mathbf{L}$ 中のラベル $L_1$ と $L_2$ が $L_1 \not\preceq L_2$ かつ $L_2 \not\preceq L_1$ であるような $\mathbf{L}$ であれば、 $\mathbf{L}$ の各ラベルで表わされるオブジェクトの直和になる。したがって、 $\mathbf{L}$ は個々のラベルごとに扱うことができ、 $|\mathbf{L}| = 1$ のとき $\overline{\{L\}}^{CU} = \overline{\{L\}}^{UC}$ なので、 $\preceq_{CU}$ は $\preceq_{UC}$ で考えることができる。よって、8通りの順序は、順序 $\preceq_{DI}$ 、 $\preceq_{UC}$ 、 $\preceq_{DU}$ (= $\preceq_{UD}$ )に集約される。

ラベル集合 $\mathbf{L}$ によって記述されるオブジェクト $o$ は、3通りの順序によって決定されるが、 $\mathbf{L}$ と $\tilde{o}$ の中には $\mathbf{L}$ の要素の決定に関係のないラベルがある。

[例 2] ラベル集合 $\mathbf{L}_1$ とオブジェクト $o_1$ のラベル $\tilde{o}_1$ をそれぞれ {製造業, サービス業} と {輸送機器, 金融, 農業} とする (図 2)。 $\tilde{o}_1$ は $\mathbf{L}_1$ におけるすべてのラベルについて、それ以下のラベルを含むので、 $o_1$ は $\overline{\mathbf{L}_1}^{DI}$ の要素であり、 $\tilde{o}_1$ の“農業”は要素の決定に関与しない。一方で、 $\overline{\mathbf{L}_1}^{UC}$ 中のオブジェクトの集合ラベルには $\mathbf{L}_1$ のラベルに対応するラベルが含まれていない場合もある。例えば、{輸送機器, 電機機器} という集合ラベルのオブジェクト $o_2$ は $\mathbf{L}_1$ の“サービス業”に対するそれ以下のラベルを含まないにもかかわらず、 $o_2$ は $\overline{\mathbf{L}_1}^{UC}$ の要素である。また、 $\overline{\mathbf{L}_1}^{DU}$ では、 $\overline{\mathbf{L}_1}^{DU}$ 中のオブジェクトの集合ラベルと $\mathbf{L}_1$ の両方に要素の決定に関与しないラベルが含まれる。 □

図 2 順序の決定に関与しないラベル



ラベル集合 $\mathbf{L}_1$ と $\mathbf{L}_2$ に対し、 $\mathbf{L}_2$ のすべてのラベルについてそのラベル以下のラベルが $\mathbf{L}_1$ にあれば $\mathbf{L}_1 \preceq_{DI} \mathbf{L}_2$ である。よって、 $\mathbf{L}_2$ には $\mathbf{L}_1$ のラベルの上位ではないラベルは含まれていない。すなわち、 $\mathbf{L}_1 \preceq_{DI} \mathbf{L}_2$ は上位の $\mathbf{L}_2$ に制限がある。同様に、 $\mathbf{L}_1 \preceq_{UC} \mathbf{L}_2$ は下位の $\mathbf{L}_1$ に制限があり、 $\mathbf{L}_1 \preceq_{DU} \mathbf{L}_2$ (= $\mathbf{L}_1 \preceq_{UD} \mathbf{L}_2$ )は、上位と下位のどちらにも制限はない。順序 $\preceq_{DI}$ 、 $\preceq_{UC}$ 、 $\preceq_{DU}$ (= $\preceq_{UD}$ )は定義3の順序で表すことができる。

[定義 3] ラベル集合  $\mathbf{L}_1$  と  $\mathbf{L}_2$  に対し,

$$\forall L_2 \in \mathbf{L}_2, \exists L_1 \in \mathbf{L}_1, L_1 \preceq L_2 \iff \mathbf{L}_1 \preceq_{RU} \mathbf{L}_2$$

$$\forall L_1 \in \mathbf{L}_1, \exists L_2 \in \mathbf{L}_2, L_1 \preceq L_2 \iff \mathbf{L}_1 \preceq_{RL} \mathbf{L}_2$$

$$\exists L_1 \in \mathbf{L}_1, \exists L_2 \in \mathbf{L}_2, L_1 \preceq L_2 \iff \mathbf{L}_1 \preceq_{RN} \mathbf{L}_2$$

である。 □

順序  $\preceq_{RU}$ ,  $\preceq_{RL}$ ,  $\preceq_{RN}$  を用いてラベル集合  $\mathbf{L}$  で記述されるオブジェクト集合をそれぞれ  $\overline{\mathbf{L}}^{RU}$ ,  $\overline{\mathbf{L}}^{RL}$ ,  $\overline{\mathbf{L}}^{RN}$  で表す, すなわち,  $\overline{\mathbf{L}}^{RU} = \overline{\mathbf{L}}^{DI}$ ,  $\overline{\mathbf{L}}^{RL} = \overline{\mathbf{L}}^{UC}$ ,  $\overline{\mathbf{L}}^{RN} = \overline{\mathbf{L}}^{DU} (= \overline{\mathbf{L}}^{UD})$  である。

次に, 順序の組合せを考えることで, オブジェクトを記述する上でその他に必要な順序がないかを検討する。  $\mathbf{L}_1 \preceq_x \mathbf{L}_2$  かつ  $\mathbf{L}_1 \preceq_y \mathbf{L}_2$  ( $x, y \in \{CI, CU, DI, DU, IC, ID, UC, UD\}$ ) であるとき, ラベル集合  $\mathbf{L}_1$  は  $\mathbf{L}_2$  以下とする順序を考える。  $x = DI$ ,  $y = UC$  で定義される順序以外は  $\preceq_x$  または  $\preceq_y$  のどちらかの順序に等しい。例えば,  $x = CI$ ,  $y = CU$  で定義される順序は  $\preceq_{CI}$  に等しい。

順序  $\preceq_{DI}$  と  $\preceq_{UC}$  はそれぞれ順序  $\preceq_{RU}$  と  $\preceq_{RL}$  なので,  $x = DI$ ,  $y = UC$  で定義される順序は  $\preceq_{RU}$  と  $\preceq_{RL}$  の性質を持つ順序であり, 上位と下位のラベル集合の両方に制限がある順序として  $\preceq_{RB}$  で表す。順序  $\preceq_{RB}$  でラベル集合  $\mathbf{L}$  によって記述されるオブジェクト集合を  $\overline{\mathbf{L}}^{RB}$  とする。  $\overline{\mathbf{L}}^{RB}$  は  $\overline{\mathbf{L}}^{RB} = \{o \mid \tilde{o} \preceq_{RB} \mathbf{L}\} = \{o \mid \tilde{o} \preceq_{RU} \mathbf{L}, \tilde{o} \preceq_{RL} \mathbf{L}\}$  と表すことができるので, 順序  $\preceq_{RB}$  は以下のように定義される。

[定義 4] ラベル集合  $\mathbf{L}_1$  と  $\mathbf{L}_2$  に対し,

$$\forall L_2 \in \mathbf{L}_2, \exists L_1 \in \mathbf{L}_1, L_1 \preceq L_2 \text{ and } \forall L_1 \in \mathbf{L}_1, \exists L_2 \in \mathbf{L}_2, L_1 \preceq L_2$$

$$\iff \mathbf{L}_1 \preceq_{RB} \mathbf{L}_2 \text{ である。}$$

□

### 3 多重属性を持つラベル集合の順序

2 節では, 単一属性に限定してラベル集合によるオブジェクト集合の記述のためにラベル集合の順序を導出した。複数の属性を複合的に扱うためには, 単一属性で議論したラベル集合の順序を複数の属性に拡張する必要がある。本節では, 多重属性のラベル集合を導入して, 多重属性のラベル集合の順序について議論する。

オブジェクトを複数の属性で記述する場合, 属性ごとの分類によって付されるラベル集合を複合的に扱う必要があるので, 属性ごとのラベル集合を要素とするラベル集合を定義する。

[定義 5]  $n (\geq 1)$  個の属性からなるラベル集合  $\hat{\mathbf{L}}$  は, 各属性のラベル集合を要素とするラベル集合, すなわち,  $\hat{\mathbf{L}} = \{\hat{L}_1, \hat{L}_2, \dots, \hat{L}_n\}$ ,  $\hat{L}_i = \{L_{i1}, L_{i2}, \dots, L_{ini}\}$  とし,  $\hat{\mathbf{L}}$  を多重ラベル集合と呼ぶ。また, オブジェクトのラベルが多重ラベル集合

である場合、そのラベルを多重集合ラベルという。 □

[例 3] 地域で分類した際のラベル集合が {日本, 米国} であり、業種で分類した際のラベル集合が {輸送機器, 金融} であるような多重ラベル集合は {{日本, 米国}, {輸送機器, 電気機器}} となる。また、輸送機器と金融に関する愛知とデトロイトを比較した文献を業種と地域という 2 つの属性で分類した場合、オブジェクトのラベルは {{輸送機器, 金融}, {愛知, デトロイト}} という多重集合ラベルになる。 □

本稿では議論を簡単にするために、異なる属性は共通のドメインを有しないと仮定する。すなわち、多重ラベル集合  $\hat{L}$  における要素  $\hat{L}_i, \hat{L}_j \in \hat{L} (i \neq j)$  に対して、 $\hat{L}_i \cap \hat{L}_j = \phi$  とする。

単一属性で定義した順序を多重ラベル集合に拡張する。単一属性におけるラベル集合の順序は、ラベル集合の要素であるラベルの上下関係で決定される。多重ラベル集合の要素はラベル集合なので、単一属性のラベル集合の順序を適用するには要素をラベルにする必要がある。よって、多重ラベル集合の要素であるラベル集合に対して和集合をとることで、多重ラベル集合の要素がラベルになるように展開する。多重ラベル集合  $\hat{L}$  の要素の和集合を  $e(\hat{L}) = \bigcup_{\hat{L} \in \hat{L}} \hat{L}$  とし、単一属性のラベル集合の順序を多重ラベル集合に適用する。

[定義 6] 多重ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対し、

$$\exists L_1 \in e(\hat{L}_1), \exists L_2 \in e(\hat{L}_2), L_1 \preceq L_2 \iff \hat{L}_1 \preceq_{RN'} \hat{L}_2$$

$$\forall L_2 \in e(\hat{L}_2), \exists L_1 \in e(\hat{L}_1), L_1 \preceq L_2 \iff \hat{L}_1 \preceq_{RV'} \hat{L}_2$$

$$\forall L_1 \in e(\hat{L}_1), \exists L_2 \in e(\hat{L}_2), L_1 \preceq L_2 \iff \hat{L}_1 \preceq_{RL'} \hat{L}_2$$

$$\forall L_2 \in e(\hat{L}_2), \exists L_1 \in e(\hat{L}_1), L_1 \preceq L_2 \text{ and } \forall L_1 \in e(\hat{L}_1), \exists L_2 \in e(\hat{L}_2), L_1 \preceq L_2 \\ \iff \hat{L}_1 \preceq_{RB'} \hat{L}_2$$

である。 □

2 節では 8 通りのラベル集合の順序を 3 通りの順序に集約した。多重ラベル集合の順序に関しても、同様の集約が可能である。まず、単一の属性における順序  $\preceq_{ID}$  と  $\preceq_{IC} (= \preceq_{CI})$  について、それぞれ、多重ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対して、 $\exists L_1 \in e(\hat{L}_1), \forall L_2 \in e(\hat{L}_2), L_1 \preceq L_2 \iff \hat{L}_1 \preceq_{ID'} \hat{L}_2$  と  $\forall L_1 \in e(\hat{L}_1), \forall L_2 \in e(\hat{L}_2), L_1 \preceq L_2 \iff \hat{L}_1 \preceq_{IC'(CI')} \hat{L}_2$  とする。要素が単数であれば、前節で議論した通りオブジェクトを記述する際の順序として考慮する必要はない。要素が複数である場合には、属性が共通するドメインを持たないので記述されるオブジェクトは存在しない。例えば、{日本} と {輸送機器} という両方の要素に共通する下位概念のラベルは存在しない。よって、これらの順序は多重ラベル集合に拡張する際に考慮する必要はない。残りの順序  $\preceq_{ID}$  についても、多重ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対して、 $\exists L_2 \in e(\hat{L}_2), \forall L_1 \in e(\hat{L}_1), L_1 \preceq L_2 \iff \hat{L}_1 \preceq_{ID'} \hat{L}_2$  とした場

合、多重集合ラベルの要素が複数であるオブジェクト、すなわち属性が複数あるような多重集合ラベルのオブジェクトは記述されないため考慮する必要はない。

多重ラベル集合の順序について要素をラベルに展開してラベル集合の定義を適用した定義 6 は、命題 1 の通り、多重ラベル集合の要素をラベル集合とする定義に変換できる。

[命題 1] 多重ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対し、

$$\begin{aligned} \exists \hat{L}_1 \in \hat{L}_1, \exists \hat{L}_2 \in \hat{L}_2, \hat{L}_1 \preceq_{RN} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RN'} \hat{L}_2 \\ \forall \hat{L}_2 \in \hat{L}_2, \exists \hat{L}_1 \in \hat{L}_1, \hat{L}_1 \preceq_{RU} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RU'} \hat{L}_2 \\ \forall \hat{L}_1 \in \hat{L}_1, \exists \hat{L}_2 \in \hat{L}_2, \hat{L}_1 \preceq_{RL} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RL'} \hat{L}_2 \\ \forall \hat{L}_2 \in \hat{L}_2, \exists \hat{L}_1 \in \hat{L}_1, \hat{L}_1 \preceq_{RB} \hat{L}_2 \text{ and } \forall \hat{L}_1 \in \hat{L}_1, \exists \hat{L}_2 \in \hat{L}_2, \hat{L}_1 \preceq_{RB} \hat{L}_2 &\iff \\ \hat{L}_1 \preceq_{RB'} \hat{L}_2 \end{aligned}$$

である。 □

(証明)  $\exists \hat{L}_1 \in \hat{L}_1, \exists \hat{L}_2 \in \hat{L}_2, \hat{L}_1 \preceq_{RN} \hat{L}_2$  について、 $\preceq_{RN}$  の定義より、 $\hat{L}_1 \preceq_{RN} \hat{L}_2$  は  $\exists L_1 \in \hat{L}_1, \exists L_2 \in \hat{L}_2, L_1 \preceq L_2$  と書き換えられる。 $L_1$  と  $L_2$  は、それぞれ  $e(\hat{L}_1)$  と  $e(\hat{L}_2)$  に含まれるので、 $\exists \hat{L}_1 \in \hat{L}_1, \exists \hat{L}_2 \in \hat{L}_2, \hat{L}_1 \preceq_{RN} \hat{L}_2$  は  $\exists L_1 \in e(\hat{L}_1), \exists L_2 \in e(\hat{L}_2), L_1 \preceq L_2$  と書き換えられる。よって、定義 6 の  $\preceq_{RN'}$  の定義より  $\exists \hat{L}_1 \in \hat{L}_1, \exists \hat{L}_2 \in \hat{L}_2, \hat{L}_1 \preceq_{RN} \hat{L}_2 \iff \hat{L}_1 \preceq_{RN'} \hat{L}_2$  である。 $\preceq_{RU'}$ 、 $\preceq_{RL'}$ 、 $\preceq_{RB'}$  についても同様である。 (証明終)

命題 1 より、多重ラベル集合の順序は要素をラベル集合として定義でき、要素間の順序にラベル集合の順序を用いることができる。ラベル集合の順序は同じ属性においてのみ比較可能なので、多重ラベル集合の順序は、同じ属性の要素の有無により以下の 4 通りに分類できる。

[定義 7] ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対し、

$$\begin{aligned} \exists \hat{L}_{1i} (\hat{L}_{1i} \in \hat{L}_1, \hat{L}_{1i} \neq \phi), \hat{L}_{2i} \neq \phi \text{ であることを } \hat{L}_1 \preceq_{RN} \hat{L}_2 \\ (\exists \hat{L}_{2i} (\hat{L}_{2i} \in \hat{L}_2, \hat{L}_{2i} \neq \phi), \hat{L}_{1i} = \phi \text{ であることも同義}) \\ \forall \hat{L}_{2i} (\hat{L}_{2i} \in \hat{L}_2, \hat{L}_{2i} \neq \phi), \hat{L}_{1i} \neq \phi \text{ であることを } \hat{L}_1 \preceq_{RU} \hat{L}_2 \\ \forall \hat{L}_{1i} (\hat{L}_{1i} \in \hat{L}_1, \hat{L}_{1i} \neq \phi), \hat{L}_{2i} \neq \phi \text{ であることを } \hat{L}_1 \preceq_{RL} \hat{L}_2 \\ \hat{L}_1 \preceq_{RU} \hat{L}_2 \text{ かつ } \hat{L}_1 \preceq_{RL} \hat{L}_2 \text{ を } \hat{L}_1 \preceq_{RB} \hat{L}_2 \end{aligned}$$

と表し、 $\preceq$  を型という。また、 $\preceq_{RN}$  を RN 型、 $\preceq_{RU}$  を RU 型、 $\preceq_{RL}$  を RL 型、 $\preceq_{RB}$  を RB 型と呼ぶ。 □

多重ラベル集合の順序は、同じ属性の要素間について上下関係を比較することになるので、多重ラベル集合の要素間をラベル集合の順序で定義する。

[定義 8] 多重ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対し、 $\forall \hat{L}_{1i}, \hat{L}_{2i} (\hat{L}_{1i} \in \hat{L}_1, \hat{L}_{1i} \neq \phi, \hat{L}_{2i} \in \hat{L}_2, \hat{L}_{2i} \neq \phi)$  について、 $\hat{L}_{1i} \preceq_x \hat{L}_{2i} (x \in \{RN, RU, RL, RB\})$  であることを  $\hat{L}_1 \preceq_x \hat{L}_2$  と表す。

命題1における多重ラベル集合の順序は、定義7の型を用いて要素間をラベル集合の順序で定義できる。

[定理3] 多重ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対し、

$$\begin{aligned} \hat{L}_1 \preceq_{RN} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RN} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RN'} \hat{L}_2 \\ \hat{L}_1 \preceq_{RU} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RU} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RU'} \hat{L}_2 \\ \hat{L}_1 \preceq_{RL} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RL} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RL'} \hat{L}_2 \\ \hat{L}_1 \preceq_{RB} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RB} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RB'} \hat{L}_2 \end{aligned}$$

である。 □

(証明) 多重ラベル集合の順序  $\preceq_{RN'}$  に対して、 $\hat{L}_1 \preceq_{RN} \hat{L}_2$  かつ  $\hat{L}_1 \preceq_{RN} \hat{L}_2$  は  $\exists \hat{L}_1 \in \hat{L}_1, \exists \hat{L}_2 \in \hat{L}_2, \hat{L}_1 \preceq_{RN} \hat{L}_2$  と同義なので、命題1より  $\hat{L}_1 \preceq_{RN'} \hat{L}_2$  である。よって、 $\hat{L}_1 \preceq_{RN} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RN} \hat{L}_2 \iff \hat{L}_1 \preceq_{RN'} \hat{L}_2$  が成り立つ。 $\preceq_{RU'}$ ,  $\preceq_{RL'}$ ,  $\preceq_{RB'}$  についても同様である。 (証明終)

定理3より、多重ラベル集合の順序は型と要素間のラベル集合の順序で決定されることが明らかになった。ゆえに、多重ラベル集合は型とラベル集合の順序に分けて考えることができる。

## 4 多重ラベル集合の順序による記述

### 4.1 多重ラベル集合の順序で記述されるオブジェクト

多重ラベル集合の順序は、型と要素間のラベル集合の順序によって決まる。したがって、4通りの型と4通りのラベル集合の順序の組合せが考えられるので、多重集合ラベルでオブジェクトを記述する多重ラベル集合の順序は以下の通りである。

[定義9] ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対し、

$$\begin{aligned} \hat{L}_1 \preceq_{RN} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RN} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RN/RN} (= \preceq_{RN'}) \hat{L}_2 \\ \hat{L}_1 \preceq_{RU} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RU} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RN/RU} \hat{L}_2 \\ \hat{L}_1 \preceq_{RL} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RL} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RN/RL} \hat{L}_2 \\ \hat{L}_1 \preceq_{RN} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RB} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RN/RB} \hat{L}_2 \\ \hat{L}_1 \preceq_{RU} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RN} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RU/RN} \hat{L}_2 \\ \hat{L}_1 \preceq_{RU} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RU} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RU/RU} (= \preceq_{RU'}) \hat{L}_2 \\ \hat{L}_1 \preceq_{RU} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RL} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RU/RL} \hat{L}_2 \\ \hat{L}_1 \preceq_{RU} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RB} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RU/RB} \hat{L}_2 \\ \hat{L}_1 \preceq_{RL} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RN} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RL/RN} \hat{L}_2 \\ \hat{L}_1 \preceq_{RL} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RU} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RL/RU} \hat{L}_2 \\ \hat{L}_1 \preceq_{RL} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RL} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RL/RL} (= \preceq_{RL'}) \hat{L}_2 \\ \hat{L}_1 \preceq_{RL} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RB} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RL/RB} \hat{L}_2 \end{aligned}$$

$$\begin{aligned} \hat{L}_1 \preceq_{RB} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RN} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RB/RN} \hat{L}_2 \\ \hat{L}_1 \preceq_{RB} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RU} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RB/RU} \hat{L}_2 \\ \hat{L}_1 \preceq_{RB} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RL} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RB/RL} \hat{L}_2 \\ \hat{L}_1 \preceq_{RB} \hat{L}_2 \text{ and } \hat{L}_1 \preceq_{RB} \hat{L}_2 &\iff \hat{L}_1 \preceq_{RB/RB} (= \preceq_{RB'}) \hat{L}_2 \end{aligned}$$

である。

□

多重ラベル集合  $\hat{L}$  に対して,  $\preceq_{x/y}$  ( $x, y \in \{= RN, RU, RL, RB\}$ ) の各順序で  $\hat{L}$  が記述するオブジェクト集合を  $\overline{\hat{L}}^{x/y} = \{o \mid o \preceq_{x/y} \hat{L}\}$  で表す。 $\overline{\hat{L}}^{x/y}$  は, 例 4 に示すようなオブジェクトを記述している。

- [例 4] 多重ラベル集合  $\hat{L} = \{\text{製造業, サービス業}, \{\text{日本, 米国}\}\}$  に対して,
- $\overline{\hat{L}}^{RN/RN}$  :  $\{\{\text{輸送機器, 農業, } \dots\}, \{\text{東京証券取引所, N.Y. 証券取引所, } \dots\}, \{\dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RN/RU}$  :  $\{\{\text{輸送機器, 金融, 農業, } \dots\}, \{\text{東京証券取引所, N.Y. 証券取引所, } \dots\}, \{\dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RN/RL}$  :  $\{\{\text{輸送機器}\}, \{\text{東京証券取引所, N.Y. 証券取引所, } \dots\}, \{\dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RN/RB}$  :  $\{\{\text{輸送機器, 金融}\}, \{\text{東京証券取引所, N.Y. 証券取引所, } \dots\}, \{\dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RU/RN}$  :  $\{\{\text{輸送機器, 農業, } \dots\}, \{\text{愛知, 上海, } \dots\}, \{\text{東京証券取引所, N.Y. 証券取引所, } \dots\}, \{\dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RU/RU}$  :  $\{\{\text{輸送機器, 金融, 農業, } \dots\}, \{\text{愛知, デトロイト, 上海, } \dots\}, \{\text{東京証券取引所, N.Y. 証券取引所, } \dots\}, \{\dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RU/RL}$  :  $\{\{\text{輸送機器}\}, \{\text{愛知}\}, \{\text{東京証券取引所, N.Y. 証券取引所, } \dots\}, \{\dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RU/RB}$  :  $\{\{\text{輸送機器, 金融}\}, \{\text{愛知, デトロイト}\}, \{\text{東京証券取引所, N.Y. 証券取引所, } \dots\}, \{\dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RL/RN}$  :  $\{\{\text{輸送機器, 農業, } \dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RL/RU}$  :  $\{\{\text{輸送機器, 金融, 農業, } \dots\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RL/RL}$  :  $\{\{\text{輸送機器}\}\}$  といった多重集合ラベルを持つオブジェクト集合
  - $\overline{\hat{L}}^{RL/RB}$  :  $\{\{\text{輸送機器, 金融}\}\}$  といった多重集合ラベルを持つオブジェクト集合

$\overline{\mathbf{L}}^{RB/RN}$  :  $\{\{\text{輸送機器, 農業, …}\}, \{\text{愛知, 上海, …}\}\}$  といった多重集合ラベルを持つオブジェクト集合

$\overline{\mathbf{L}}^{RB/RU}$  :  $\{\{\text{輸送機器, 金融, 農業, …}\}, \{\text{愛知, デトロイト, 上海, …}\}\}$  といった多重集合ラベルを持つオブジェクト集合

$\overline{\mathbf{L}}^{RB/RL}$  :  $\{\{\text{輸送機器}\}, \{\text{愛知}\}\}$  といった多重集合ラベルを持つオブジェクト集合

$\overline{\mathbf{L}}^{RB/RB}$  :  $\{\{\text{輸送機器, 金融}\}, \{\text{愛知, デトロイト}\}\}$  といった多重集合ラベルを持つオブジェクト集合

である。

□

多重ラベル集合の順序によって、多重ラベル集合は 16 通りのオブジェクト集合を記述することになる。この際、型によって属性値の有無を考慮したオブジェクト集合の記述がなされるが、オブジェクトがある分類属性に対してその値が  $\phi$  である場合、すなわち多重集合ラベルにおいてある属性値が  $\phi$  である場合の解釈は以下の 3 通りになる。

1. 属性値が空値である場合であり、オブジェクトが属性を持たない場合に用いられる
2. 属性値が null である場合であり、オブジェクトが未分類である場合に用いられる
3. 属性値が“0”や“無”といった場合であり、属性に対する値がない場合に用いられる

[例 5] 1. 愛知県の政治に関する文献といったオブジェクトであれば、業種という属性を持たないために業種の属性値は空値になる。

2. 愛知県の製造業に関する文献といったオブジェクトが業種という属性で分類されていない場合には業種の属性値は *null* である。

3. 未上場の愛知県の中小企業に関するオブジェクトを上場市場という属性で分類した場合には業種の属性値は無になる。

□

オブジェクトの多重集合ラベルの属性値が  $\phi$  であるという情報は、多重ラベル集合を用いて分析対象のオブジェクトを指定する際、オブジェクトが属性を持たない、オブジェクトが未分類、属性に対する値がないという理由で分析対象から除く場合に有用である。

#### 4. 2 多重ラベル集合の順序の記述能力

先行研究<sup>[16]</sup>では、複数の分類階層において任意のオブジェクトを記述することについて、論理式の積和標準形に対応する形式でオブジェクトを記述することが可能であることを示すことで十分な記述能力を証明している。しかし、この記述に

関しては、各ラベルを個別に扱い和集合や積集合のみで記述しているため、ラベルを集合として捉えた記述を考慮していない。これまでの多重ラベル集合が記述するオブジェクト集合と多重属性の各属性が記述するオブジェクト集合の和集合や積集合で記述されるオブジェクト集合を比較したものが性質1である。

[性質1] 多重ラベル集合  $\hat{L}$  に対し、

$$\bigcap_{\hat{L} \in \hat{L}} \overline{\hat{L}}^{RN} = \overline{\hat{L}}^{RU/RN}, \bigcup_{\hat{L} \in \hat{L}} \overline{\hat{L}}^{RN} = \overline{\hat{L}}^{RN/RN} = \overline{\hat{L}}^{RN'}$$

$$\bigcap_{\hat{L} \in \hat{L}} \overline{\hat{L}}^{RU} = \overline{\hat{L}}^{RU/RU} = \overline{\hat{L}}^{RU'}, \bigcup_{\hat{L} \in \hat{L}} \overline{\hat{L}}^{RU} = \overline{\hat{L}}^{RN/RU}$$

$$\bigcap_{\hat{L} \in \hat{L}} \overline{\hat{L}}^{RL} = \phi, \bigcup_{\hat{L} \in \hat{L}} \overline{\hat{L}}^{RL} \subseteq \overline{\hat{L}}^{RL/RL}$$

$$\bigcap_{\hat{L} \in \hat{L}} \overline{\hat{L}}^{RB} = \phi, \bigcup_{\hat{L} \in \hat{L}} \overline{\hat{L}}^{RB} \subseteq \overline{\hat{L}}^{RL/RB}$$

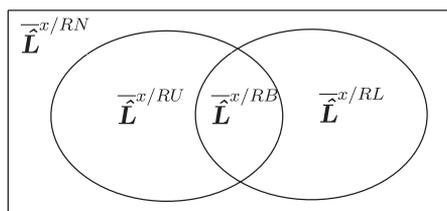
である。 □

性質1より、多重属性の各属性が記述するオブジェクト集合の和集合や積集合で記述した場合には、 $\overline{\hat{L}}^{RN/RL}$ ,  $\overline{\hat{L}}^{RN/RB}$ ,  $\overline{\hat{L}}^{RU/RL}$ ,  $\overline{\hat{L}}^{RU/RB}$ ,  $\overline{\hat{L}}^{RL/y}$ ,  $\overline{\hat{L}}^{RB/y}$  ( $y \in \{=RN, RU, RL, RB\}$ ) が記述できない。これは、ラベル集合の順序の  $\preceq_{RL}$  や  $\preceq_{RB}$  を用いた記述のように、ラベル集合の範囲を意識したオブジェクトの記述ができないためである。同時に、多重属性になることで考慮すべき型についても、各ラベルで記述されるオブジェクトの和集合や積集合では  $RN$  型の順序で記述されるオブジェクト集合や  $RU$  型の順序で記述されるオブジェクト集合をそれぞれ記述できたとしても、多重ラベル集合の属性の範囲を意識した  $RL$  型の順序や  $RB$  型の順序で記述されるオブジェクト集合は記述できない。

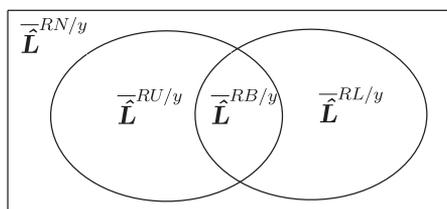
これより、多重ラベル集合の順序によって記述されるオブジェクト集合を整理し、それらが分析対象となるオブジェクト集合の指定に十分な記述であることを示す。多重ラベル集合によって記述されるオブジェクト集合は、要素間のラベル集合の順序と型に分けて考えられるので、図3(a)は要素間のラベル集合の順序によるオブジェクト集合の包含関係を、図3(b)は型の違いに起因する包含関係を示している。

図3(a)の  $\overline{\hat{L}}^{x/RN}$  は、 $\hat{L}$  の要素(ラベル集合)におけるいくつかのラベルに対して上下関係のあるラベルを含む要素を持つ多重集合ラベルのオブジェクトを記述し、 $\overline{\hat{L}}^{x/RU}$  は  $\hat{L}$  の要素(ラベル集合)のすべてのラベルにおいて上下関係があるラベルが存在するような要素を持つ多重集合ラベルのオブジェクト集合を記述する。 $\overline{\hat{L}}^{x/RL}$  は  $\hat{L}$  の要素(ラベル集合)の範囲を限定したオブジェクト集合を記述し、 $\overline{\hat{L}}^{x/RB}$  は  $\overline{\hat{L}}^{x/RU}$  と  $\overline{\hat{L}}^{x/RU}$  の両方の性質を満たすオブジェクトを記述している。要素

図3 多重ラベル集合の順序によって記述されるオブジェクト集合の包含関係



(a)



(b)

間の比較で分析対象となるオブジェクトを指定することを考えると、ラベル集合の順序によるオブジェクト集合の記述では、すべてのラベルに関するオブジェクトを対象とするか、いずれかのラベルに関するオブジェクトを対象とするかが考えられる。また、オブジェクトの多重集合ラベルの要素に無関係なラベルを含んでいるものを含めるか含めないかという選択もある。これらは、ラベル集合の4通りの順序によるオブジェクト集合の記述の種類に対応しているので、4種類のラベル集合の順序で十分な記述ができることを示している。

一方、図3 (b) の  $\overline{\mathbf{L}}^{RN/y}$  では、 $\hat{\mathbf{L}}$  のいくつかの属性に関係がある多重集合ラベルのオブジェクトを記述している。それに対して、 $\overline{\mathbf{L}}^{RU/y}$  は  $\hat{\mathbf{L}}$  のすべての属性に関するオブジェクト集合を記述し、 $\overline{\mathbf{L}}^{RL/y}$  は  $\hat{\mathbf{L}}$  が有する属性の範囲内のオブジェクト集合を記述し、 $\overline{\mathbf{L}}^{RB/y}$  は  $\overline{\mathbf{L}}^{RU/y}$  と  $\overline{\mathbf{L}}^{RL/y}$  の両方の性質を満たすオブジェクトを記述している。多重ラベル集合で分析対象となるオブジェクトの属性の種類を指定することを考えると、多重ラベル集合が持つ属性の集合と多重集合ラベルの持つ属性の集合における包含関係を考えればよい。まず、多重ラベル集合によって記述されるオブジェクトであれば、多重ラベル集合の属性の集合とそれによって記述されるオブジェクトにおける多重集合ラベルの属性の集合に共通集合がなければならず、これは  $RN$  型の順序で記述されるオブジェクトに該当する。次に、共通集合がある場合には、多重ラベル集合の属性の集合に対して、多重集合ラベルの属性の集合が、一致する場合、部分集合である場合、超集合である場合があり、それぞれ  $RB$  型の順序、 $RL$  型の順序、 $RU$  型の順序が対応する。したがって、多重ラベル集合

で分析対象となるオブジェクトの属性の種類指定においても、4 種類の型で十分な記述ができるといえる。

## 5 多重ラベル集合の順序の性質

### 5.1 多重ラベル集合の順序の健全性

多重ラベル集合を用いて分析対象となるオブジェクトを指定する際、ある多重ラベル集合に対する上位概念の多重ラベル集合で分析対象を拡張させたり、下位概念の多重ラベル集合で狭めることがある。すなわち、多重ラベル集合の上下関係とそれらによって記述されるオブジェクト集合の包含関係が一致すべきであり、この性質について議論する。

多重ラベル集合の順序が概念の順序に等しいと考えるならば、多重ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対して、 $\hat{L}_1$  が  $\hat{L}_2$  以下ならば  $\overline{\hat{L}_1}$  におけるオブジェクトは  $\overline{\hat{L}_2}$  の要素であることが求められる。

[定義 10] 多重ラベル集合  $\hat{L}_1$  と  $\hat{L}_2$  に対し、 $\hat{L}_1 \preceq_{x/y} \hat{L}_2$  ( $x, y \in \{RN, RU, RL, RB\}$ ) であることと  $\overline{\hat{L}_1}^{x/y} \subseteq \overline{\hat{L}_2}^{x/y}$  であることが等価であるならば、多重ラベル集合の順序  $\preceq_{x/y}$  は健全である。 □

多重ラベル集合の順序が推移律を満たす場合に限り、その順序は健全である。

[補題 1] 多重ラベル集合の順序が推移律を満たすことは、多重ラベル集合の順序が健全であることの必要十分条件である。 □

(証明) 多重ラベル集合の順序  $\hat{L}_1 \preceq_{x/y} \hat{L}_2$  ( $x, y \in \{RN, RU, RL, RB\}$ ) が推移律を満たすとする。多重ラベル集合  $\hat{L}_1$  とオブジェクト  $o$  に対して、 $\tilde{o} \preceq_{x/y} \hat{L}_1$  ならば、 $o$  は  $\overline{\hat{L}_1}^{x/y}$  に含まれる。また、 $\hat{L}_1 \preceq_{x/y} \hat{L}_2$  であるような多重ラベル集合  $\hat{L}_2$  に対して、 $\tilde{o} \preceq_{x/y} \hat{L}_1$  と  $\hat{L}_1 \preceq_{x/y} \hat{L}_2$  から  $\tilde{o} \preceq_{x/y} \hat{L}_2$  なので、 $o$  は  $\overline{\hat{L}_2}^{x/y}$  の要素である。 $\overline{\hat{L}_1}^{x/y}$  のすべてのオブジェクトは  $\overline{\hat{L}_2}^{x/y}$  のオブジェクトなので、 $\overline{\hat{L}_1}^{x/y} \subseteq \overline{\hat{L}_2}^{x/y}$  である。 $\overline{\hat{L}_1}^{x/y} \subseteq \overline{\hat{L}_2}^{x/y}$  ならば、 $\overline{\hat{L}_1}^{x/y}$  におけるオブジェクト  $o$  は  $\overline{\hat{L}_2}^{x/y}$  に含まれる。また、 $\tilde{o} = \hat{L}_1$  であるとき  $\tilde{o} \preceq_{x/y} \hat{L}_2$  であり  $\hat{L}_1 \preceq_{x/y} \hat{L}_2$  である。したがって、 $\preceq_{x/y}$  が推移律を満たせば  $\preceq_{x/y}$  は健全である。

$\hat{L}_1 \preceq_{x/y} \hat{L}_2$  かつ  $\hat{L}_2 \preceq_{x/y} \hat{L}_3$  であるような多重ラベル集合  $\hat{L}_1, \hat{L}_2, \hat{L}_3$  に対して、 $\preceq_{x/y}$  が健全ならば  $\overline{\hat{L}_1}^{x/y} \subseteq \overline{\hat{L}_2}^{x/y}$  かつ  $\overline{\hat{L}_2}^{x/y} \subseteq \overline{\hat{L}_3}^{x/y}$  である。 $\overline{\hat{L}_1}^{x/y} \subseteq \overline{\hat{L}_2}^{x/y} \subseteq \overline{\hat{L}_3}^{x/y}$  なので、 $\overline{\hat{L}_1}^{x/y}$  におけるオブジェクト  $o$  は  $\overline{\hat{L}_3}^{x/y}$  にも含まれる。 $\tilde{o} = \hat{L}_1$  であるとき、 $\tilde{o} \preceq_{x/y} \hat{L}_3$  かつ  $\hat{L}_1 \preceq_{x/y} \hat{L}_3$  である。よって、 $\preceq_{x/y}$  が健全ならば  $\preceq_{x/y}$  は推移律を満たす。 (証明終)

[補題 2] 多重ラベル集合の型と要素間のラベル集合の順序がともに推移律を満たすことは、多重ラベル集合の順序が推移律を満たすことの必要十分条件であ

る。

□

(証明)  $\hat{L}_1 \preceq_{x/y} \hat{L}_2$  かつ  $\hat{L}_2 \preceq_{x/y} \hat{L}_3$  ( $x, y \in \{RN, RU, RL, RB\}$ ) であるような多重ラベル集合  $\hat{L}_1, \hat{L}_2, \hat{L}_3$  に対して,  $\hat{L}_1 \preceq_x \hat{L}_3$  が成り立たなければ,  $\hat{L}_1$  と  $\hat{L}_3$  に共通する属性がないので, いずれの要素間において上下関係が成り立たないため  $\hat{L}_1 \preceq_{x/y} \hat{L}_3$  が成り立たない。また, ラベル集合の順序に推移律が成り立たなければ,  $\hat{L}_1 \not\preceq_y \hat{L}_3$  なので  $\hat{L}_1 \preceq_{x/y} \hat{L}_3$  が成り立たない。よって, 多重ラベル集合  $\hat{L}_1, \hat{L}_2, \hat{L}_3$  に対して,  $\hat{L}_1 \preceq_{x/y} \hat{L}_3$  ならば,  $\hat{L}_1 \preceq_x \hat{L}_3$  かつ  $\hat{L}_1 \preceq_y \hat{L}_3$  である。すなわち, 多重ラベル集合の順序が推移律を満たせば, 多重ラベル集合の型と要素間のラベル集合の順序はともに推移律を満たす。

また,  $\hat{L}_1 \preceq_x \hat{L}_2$  かつ  $\hat{L}_2 \preceq_x \hat{L}_3$  であり,  $\hat{L}_1 \preceq_y \hat{L}_2$  かつ  $\hat{L}_2 \preceq_y \hat{L}_3$  ( $x, y \in \{RN, RU, RL, RB\}$ ) をともに満たすような多重ラベル集合  $\hat{L}_1, \hat{L}_2, \hat{L}_3$  に対して, 型と要素間のラベル集合に推移律が成り立つので,  $\hat{L}_1 \preceq_x \hat{L}_3$  かつ  $\hat{L}_1 \preceq_y \hat{L}_3$  である。よって, 定理 3 より  $\hat{L}_1 \preceq_{x/y} \hat{L}_3$  が成り立つので, 多重ラベル集合の型と要素間のラベル集合の順序がともに推移律を満たせば, 多重ラベル集合の順序は推移律を満たす。 (証明終)

多重ラベル集合の順序  $\preceq_{RN/y}$  ( $y \in \{RN, RU, RL, RB\}$ ) の型は推移律を成り立たないため,  $RN$  型の順序は推移律を満たさない。例えば, 多重ラベル集合  $\hat{L}_1 = \{\{\text{日本}, \text{米国}\}\}$  に対して, 多重ラベル集合  $\hat{L}_2 = \{\{\text{東京}, \text{ニューヨーク}\}, \{\text{サービス業}\}\}$  は下位概念である。  $\overline{\hat{L}_2}^{RN/y}$  に含まれる  $\{\{\text{金融}\}\}$  という多重ラベル集合のオブジェクト  $o$  は,  $o \not\prec_{RN} \hat{L}_1$  であり  $o \not\preceq_{RN/y} \hat{L}_1$  なので,  $o$  は  $\overline{\hat{L}_1}^{RN/y}$  に含まれない。

一方, 多重ラベル集合の順序  $\preceq_{RU/y}, \preceq_{RL/y}, \preceq_{RB/y}$  の型には推移律が成り立つ。

[補題 3] 多重ラベル集合の順序  $\preceq_{RU/y}, \preceq_{RL/y}, \preceq_{RB/y}$  ( $y \in \{RN, RU, RL, RB\}$ ) の型, すなわち,  $\preceq_{RU}, \preceq_{RL}, \preceq_{RB}$  は推移律を満たす。 □

(証明)  $\hat{L}_1 \preceq_{RU} \hat{L}_2$  かつ  $\hat{L}_2 \preceq_{RU} \hat{L}_3$  であるような多重ラベル集合  $\hat{L}_1, \hat{L}_2, \hat{L}_3$  に対して,  $\forall \hat{L}_{3i} (\hat{L}_{3i} \in \hat{L}_3, \hat{L}_{3i} \neq \phi)$  が  $\hat{L}_{2i} \neq \phi$  であり,  $\forall \hat{L}_{2i} (\hat{L}_{2i} \in \hat{L}_2, \hat{L}_{2i} \neq \phi)$  が  $\hat{L}_{1i} \neq \phi$  なので,  $\hat{L}_3$  のすべての属性を  $\hat{L}_1$  が有する。すなわち,  $\hat{L}_1$  と  $\hat{L}_3$  に対して,  $\forall \hat{L}_{3i} (\hat{L}_{3i} \in \hat{L}_3, \hat{L}_{3i} \neq \phi)$  が  $\hat{L}_{1i} \neq \phi$  なので,  $\preceq_{RU}$  の定義より  $\hat{L}_1 \preceq_{RU} \hat{L}_3$  である。また,  $\preceq_{RL}$  と  $\preceq_{RB}$  についても  $\preceq_{RU}$  の証明と同様に推移律を満たす。 (証明終)

多重ラベル集合の順序が健全であるためには, 要素間の順序であるラベル集合の順序においても推移律を満たす必要があるが, ラベル集合の順序  $\preceq_{RN}$  は推移律を満たさない。例えば, ラベル集合  $L_1, L_2, L_3$  をそれぞれ  $\{\text{東京}\}, \{\text{日本}, \text{ニューヨーク}\}, \{\text{米国}\}$  とする。  $L_1 (= \{\text{東京}\}) \preceq_{RN} L_2 (= \{\text{日本}, \text{ニューヨーク}\})$  かつ  $L_2 (= \{\text{日本}, \text{ニューヨーク}\}) \preceq_{RN} L_3 (= \{\text{米国}\})$  であるが,  $L_1 (= \{\text{東京}\})$

$\hat{L}_{RN} \mathbf{L}_3 (= \{\text{米国}\})$ である。

一方、ラベル集合の順序 $\preceq_{RU}$ ,  $\preceq_{RL}$ ,  $\preceq_{RB}$ には推移律が成り立つ。

[補題 4] ラベル集合の順序 $\preceq_{RU}$ ,  $\preceq_{RL}$ ,  $\preceq_{RB}$ は推移律を満たす。 □

(証明) 証明は文献<sup>[13]</sup>による。 (証明終)

[定理 4] 多重ラベル集合の順序 $\preceq_{x/y} (x, y \in \{RU, RL, RB\})$ は健全である。 □

(証明) 補題3と4より、多重ラベル集合の順序 $\preceq_{x/y} (x, y \in \{RU, RL, RB\})$ において、多重ラベル集合の型と要素間のラベル集合の順序に推移律が成り立つ。よって、補題2より $\preceq_{x/y}$ には推移律が成り立つので、補題1から $\preceq_{x/y}$ は健全である。 (証明終)

## 5.2 多重ラベル集合の順序の妥当性

多重ラベル集合を用いて分析対象となるオブジェクトを記述する際には、多重ラベル集合の要素や要素のラベルを変えることで分析対象を変える操作が必要になる。例えば、多重ラベル集合  $\{\{\text{日本}, \text{米国}\}\}$  に対して、多重ラベル集合  $\{\{\text{日本}, \text{フランス}\}, \{\text{自動車}\}\}$  とした場合には、記述されるオブジェクトは異なってくる。しかし、 $\{\{\text{日本}, \text{愛知}\}\}$  といったような非排他的なラベル集合の要素を持つ多重ラベル集合を考えると、多重ラベル集合が異なっても分析対象が等しい場合がある。本節ではこの性質について議論する。

多重ラベル集合 $\hat{L}_1$ と $\hat{L}_2$ が異なれば $\hat{L}_1$ と $\hat{L}_2$ が記述するオブジェクトも異なることが求められる。

[定義11] 多重ラベル集合 $\hat{L}_1$ と $\hat{L}_2 (\hat{L}_1 \neq \hat{L}_2)$ に対し、 $\overline{L}_1^{x/y} \neq \overline{L}_2^{x/y} (x, y \in \{RN, RU, RL, RB\})$ ならば多重集合ラベルの順序は妥当である。 □

多重ラベル集合は、共通のドメインを持つ属性が存在しないので、多重ラベル集合の要素を新たに追加したり削除すれば、必ず異なるオブジェクトを記述することになる。例えば、多重ラベル集合  $\{\{\text{日本}, \text{米国}\}\}$  に対して、多重ラベル集合  $\{\{\text{日本}, \text{米国}\}, \{\text{自動車}\}\}$  とした場合には、記述されるオブジェクトは異なってくる。よって、多重ラベル集合の各要素が記述するオブジェクトのみを考えればよい。要素が記述するオブジェクトは、多重ラベル集合における要素と多重集合ラベルにおける要素の順序によって決まるので、これよりラベル集合の順序に限定して考える。

ラベル集合 $L_1$ と $L_2$ に対して、 $L$ を $L_1 - L_2$ 中のラベルとする。 $L_1$ によって記述されるオブジェクト集合は、一般に $L$ があるために $L_2$ によって記述されるオブジェクト集合とは異なる。しかし、 $L_1 \cap L_2$ 中に $L$ 以下のラベルが存在するならば、 $\preceq_{RU}$ では、 $\overline{L}_1^{RU}$ の要素であるが $\overline{L}_2^{RU}$ の要素ではないようなオブジェクトは、 $L$ があるために存在することはない。例えば、ラベル集合 $L_1$ と $L_2$ をそれぞれ

{アジア, 日本} と {日本} とする  $L_1 - L_2$  にアジアがあるが,  $L_1 \cap L_2$  の日本はアジア以下なので  $\overline{L_1}^{RU}$  の要素ではあるが  $\overline{L_2}^{RU}$  の要素ではないようなオブジェクトはない。

ラベル集合のラベル間に上下関係があるラベルが含まれていないラベル集合に限定した場合, ラベル集合は, その要素が異なればそれらによって記述されるオブジェクトも異なるので妥当である。ラベル集合中のラベル  $L_i$  と  $L_j$  が  $L_i \preceq L_j$  ではない場合, そのようなラベル集合を排他であると呼ぶことにする。よって, 多重ラベル集合の要素を排他である限定すれば妥当性が成り立つ。しかし, 排他ではない要素を持つ多重ラベル集合は, オブジェクトの分類において必要なものである。例えば, 自動車産業における日本とアジア全体を比較するような文献のラベルは {{アジア, 日本}, {自動車}} になり, このような要素を持つような多重ラベル集合を考える必要がある。

ラベル集合  $L_1$  と  $L_2$  ( $L_1 \neq L_2$ ) が排他ではないとき,  $L_1 \preceq_{RU} L_2$  かつ  $L_2 \preceq_{RU} L_1$  となるラベル集合  $L_1$  と  $L_2$  が存在し,  $L_1 \preceq_{RU} L_2$  かつ  $L_2 \preceq_{RU} L_1$  を  $L_1 \approx_{RU} L_2$  で表す。例えば, ラベル集合  $L_1$  と  $L_2$  をそれぞれ {日本, 米国} と {アジア, 日本, 米国} とする。  $L_1 \preceq_{RU} L_2$  かつ  $L_2 \preceq_{RU} L_1$  なので,  $L_1 \approx_{RU} L_2$  である。

$L_1 \preceq_{RU} L_2$  であるようなラベル集合  $L_1$  と  $L_2$  に対して,  $L_2 \in L_2$ ,  $L'_2 \notin L_2$ ,  $L_2 \preceq L'_2$  であるようなラベルを  $L_2$  と  $L'_2$  とする。  $L_1$  中の  $L_1 \preceq L_2$  であるようなラベル  $L_1$  に対して,  $L_1 \preceq L'_2$  なので  $L_1 \preceq_{RU} L_2 \cup \{L'_2\}$  ( $= L'_2$ ) が成り立つ。よって,  $\overline{L_1}^{RU}$  と  $\overline{L_2}^{RU}$  は同じオブジェクト集合である。すなわち,  $\overline{L}^{RU}$  では, ラベル集合  $L$  は  $L$  におけるすべてのラベルに対して上位ではないラベルで構成される  $L$  の部分集合に縮約される。

排他ではないラベル集合  $L$  について,  $L$  におけるすべてのラベルに対して上位ではないラベルで構成される  $L$  の部分集合を  $L$  の下限とし,  $l(L) = \{L \mid L \in L, \forall L' \in L \text{ s.t. } L' \neq L, L' \not\preceq L\}$  で表す。ラベル集合の順序  $\preceq_{RU}$  では, 排他ではないラベル集合  $L$  で記述されるオブジェクト集合は,  $L$  の下限で記述されるオブジェクト集合に等しい。

[命題 2] ラベル集合  $L$  に対し,  $\overline{L}^{RU} = l(\overline{L})^{RU}$  である。 □

(証明) 証明は文献<sup>[13]</sup>による。

(証明終)

ラベル集合  $L_1$  と  $L_2$  ( $L_1 \neq L_2$ ) に対して,  $l(L_1) = l(L_2)$  ならば, 命題 2 より  $\overline{L_1}^{RU} = \overline{L_2}^{RU}$  が成り立つので, 多重ラベル集合の順序  $\preceq_{z/RU}$  は妥当ではない。

ラベル集合の下限と同様に, 排他ではないラベル集合  $L$  におけるすべてのラベルに対して下位ではないラベルで構成される  $L$  の部分集合を  $L$  の上限とし,  $u(L) = \{L \mid L \in L, \forall L' \in L \text{ s.t. } L' \neq L, L \not\preceq L'\}$  で表す。

[命題 3] ラベル集合  $\mathbf{L}$  に対し,  $\overline{\mathbf{L}}^{RL} = \overline{u(\mathbf{L})}^{RL}$ ,  $\overline{\mathbf{L}}^{RN} = \overline{u(\mathbf{L})}^{RN}$  である。□

(証明) 証明は文献<sup>[13]</sup>による。(証明終)

ラベル集合の順序  $\preceq_{RB}$  に対して,  $ul(\mathbf{L})$  を  $u(\mathbf{L}) \cup l(\mathbf{L})$  とすると,  $\mathbf{L}$  は  $ul(\mathbf{L})$  に縮約できる。

[命題 4] ラベル集合  $\mathbf{L}$  に対し,  $\overline{\mathbf{L}}^{RB} = \overline{ul(\mathbf{L})}^{RB}$  である。□

(証明) 証明は文献<sup>[14]</sup>による。(証明終)

多重ラベル集合の要素が排他ではない場合は, いずれの順序も妥当ではなく, 要素間のラベル集合の順序に応じて要素を上限または下限に縮約してオブジェクトを記述できる。多重ラベル集合  $\hat{\mathbf{L}}$  に対し,  $u(\hat{\mathbf{L}}) = \{\hat{L} \mid \hat{L} \in \hat{\mathbf{L}}, u(\hat{L})\}$ ,  $l(\hat{\mathbf{L}}) = \{\hat{L} \mid \hat{L} \in \hat{\mathbf{L}}, l(\hat{L})\}$ ,  $ul(\hat{\mathbf{L}}) = \{\hat{L} \mid \hat{L} \in \hat{\mathbf{L}}, ul(\hat{L})\}$  とする。

[定理 5] 多重ラベル集合  $\hat{\mathbf{L}}$  に対し,  $\overline{\hat{\mathbf{L}}}^{x/RN} = \overline{u(\hat{\mathbf{L}})}^{x/RN}$ ,  $\overline{\hat{\mathbf{L}}}^{x/RU} = \overline{l(\hat{\mathbf{L}})}^{x/RU}$ ,  $\overline{\hat{\mathbf{L}}}^{x/RL} = \overline{u(\hat{\mathbf{L}})}^{x/RL}$ ,  $\overline{\hat{\mathbf{L}}}^{x/RB} = \overline{ul(\hat{\mathbf{L}})}^{x/RB}$  ( $x \in \{RN, RU, RL, RB\}$ ) である。□

(証明) 命題 2 と 3 と 4 より明らか。(証明終)

多重ラベル集合の順序  $\preceq_{x/RU}$  で多重ラベル集合  $\hat{\mathbf{L}}$  が記述するオブジェクトは,  $\hat{\mathbf{L}}$  の要素の下限が記述するオブジェクトであり, 多重ラベル集合の順序  $\preceq_{x/RL}$  や  $\preceq_{x/RN}$  で多重ラベル集合  $\hat{\mathbf{L}}$  が記述するオブジェクトは,  $\mathbf{L}$  の要素の上限が記述するオブジェクトである。多重ラベル集合の順序  $\preceq_{x/RB}$  で多重ラベル集合  $\hat{\mathbf{L}}$  が記述するオブジェクトは,  $\hat{\mathbf{L}}$  の要素の上限と下限で構成されるラベル集合で記述されるオブジェクトである。よって, 多重ラベル集合の要素が排他でない場合には, 要素の上限または下限に縮約した多重ラベル集合を用いて記述すればよい。

## 6 おわりに

本稿では, ラベル集合の順序を多重属性のラベル集合に拡張することで, 多重ラベル集合の順序は型と要素間のラベル集合の順序で決定されることを明らかにした。これにより, 型と要素間のラベル集合の順序の組合せで記述されるオブジェクト集合は, 単一属性で記述されるオブジェクト集合の和集合や積集合で記述されるオブジェクト集合よりも多様な記述が可能になる。同時に, 多重ラベル集合の順序を用いた記述法が多重集合ラベルのオブジェクトの記述法として十分であることも示した。

単一属性のラベル集合の順序で成り立つ性質を多重ラベル集合の順序に拡張して検討した結果, 健全性に関しては, ラベル集合の順序が推移律を満たしても  $RN$  型の順序であれば健全ではない。また, 多重ラベル集合の型は妥当性の性質に関与しないが, 多重ラベル集合の要素が排他でない場合, 要素を上限または下限に縮約した多重ラベル集合を用いて記述すればよい。

多重属性を扱うデータに関するデータベースは、数値データを対象とした多次元データベースが開発され、データウェアハウスにおけるデータモデルとして多用されている。本稿で明らかにした多重ラベル集合によるオブジェクトの記述法は、オブジェクト、すなわち生データを対象とした多次元データベースに応用可能であり、生データに対する高度な分析ができるようになる。特に、分析対象となるデータの多様化に伴い、異なる属性を持つような不均一な生データを一元的に扱う必要性が高まっている中で、本稿で明らかにした型を意識した記述法は有用である。

### 参考文献

- [1] Agrawal, R., Gupta, A., and Sarawagi, S.: Modeling Multidimensional Databases, *Proc. the 13th Int'l Conf. on Data Engineering (ICDE'97)*, pp. 234-243 (1997).
- [2] Aho, A., Hopcroft, J., and Ullman, J.: *Data Structure and Algorithms*, Addison-Wesley Publishing Company (1987).
- [3] Adami, G., Avesani, P., and Sona, D.: Bootstrapping for Hierarchical Document Classification, *Proc. Int'l Conf. on Information and Knowledge Management (CIKM'03)*, pp. 295-302 (2003).
- [4] Bertino, E., Fan, J., Ferrari, E., Hachi, M., and Elamagarmid, A.: A Hierarchical Access Control Model for Video Database Systems, *ACM Transactions on Information Systems*, Vol. 21, No. 2, pp. 151-191 (2003).
- [5] Cardoso-Cachopo, A. and Oliveira, A.: Semi-supervised Single-label Text Categorization Using Centroid-based Classifiers, *Proc. Symposium on Applied Computing (SAC'07)*, pp. 844-851 (2007).
- [6] Chakrabarti, K., Ganti, V., Han, J., and Xin, D.: Ranking Objects by Exploiting Relationships: Computing Top-K over Aggregation, *Proc. ACM SIGMOD Int'l Conf. on Management of Data*, pp. 371-382 (2006).
- [7] Chuang, S. and Chien, L.: Taxonomy Generation for Text Segments: A Practical Web-Based Approach, *ACM Transactions on Information Systems*, Vol. 23, No. 4, pp. 363-396 (2005).
- [8] Dakka, W., Ipeirotis, P. G., and Wood, K. R.: Automatic Construction of Multifaceted Browsing Interfaces, *Proc. Int'l Conf. on Information and*

- Knowledge Management* (CIKM'05), pp. 768–775 (2005).
- [9] Dumais, S. and Chen, H.: Hierarchical Classification of Web Content, *Proc. Int'l Conf. on Research and Development in Information Retrieval* (SIGIR'00), pp. 256–263 (2000).
- [10] Elmasri, R. and Navathe, S.: *Fundamentals of Database Systems*, Addison-Wesley Publishing Company, Second Edition (1994).
- [11] Furukawa, T.: Multiple Classification Hierarchies in Cooperative Databases, *Advanced Database Syst. for Integration of Media and User Environments '98*, Advanced Database Research and Development Ser., World Scientific, Vol. 9, pp. 309–314 (1998).
- [12] Furukawa, T. and Kuzunishi, M.: Hierarchical Classification of Heterogeneous Data, *Proc. IASTED Int'l Conf. on Databases and Applications* (DBA'05), pp. 252–257 (2005).
- [13] 古川哲也, 葛西正裕: 不均一データの利用のための意味集合, データベースと Web 情報システムに関するシンポジウム論文集, Vol. 2006, No. 16, pp. 153–160 (2006).
- [14] Furukawa, T. and Kuzunishi, M.: Multi-Labeled Data Expressed by a Set of Labels, *Proc. World Academy of Science, Engineering and Technology*, Vol. 65, pp. 857–863 (2010).
- [15] Ghamrawi, N. and McMallum, A.: Collective Multi-Label Classification, *Proc. Int'l Conf. on Information and Knowledge Management* (CIKM'05), pp. 195–200 (2005).
- [16] 河野弘史, 古川哲也: 複合階層索引の質問記述能力, 情報処理学会研究報告, 98-DBS-119-52 (1999).
- [17] 河野弘史, 王磊, 古川哲也: 進化する多重階層索引の実現方式, 九州大学情報基盤センター年報, Vol. 2, pp.81–87 (2002).
- [18] 葛西正裕, 古川哲也: データ分析におけるラベル集合の性質と利用, 火の国情報シンポジウム 2010, A-3-2, 情報処理学会九州支部 (2010).
- [19] Liu, L. and Oezsu, T. (Editors-in-Chief): *Encyclopedia of Database Systems Volume 3*, Springer (2009).
- [20] Sun, A. and Lim, E.: Hierarchical Text Classification and Evaluation, *Proc. IEEE Int'l Conf. on Data Mining* (ICDM'01), pp. 521–528 (2001).
- [21] Tang, L., Rajan, S., and Narayanan, V.: Large Scale Multi-label Classification via MetaLabeler, *Proc. Int'l Conf. on World Wide Web* (WWW'09), pp. 211–220 (2009).

- [22] Wang, K., Zhou, S., and He, Y.: Hierarchical Classification of Real Life Documents, *Proc. SIAM Int'l Conf. on Data Mining*, pp. 1-16 (2001).
- [23] Wang, Y. and Oyama, K.: Web Page Classification Based on Surrounding Page Model Representing Connection Type and Directory Hierarchy, *IPSJ Transactions on Databases*, Vol. 2, No. 2, pp. 29-43 (2009).
- [24] Yang, B., Sun, J., Wang, T., Chen, Z.: Effective Multi-label Active Learning for Text Classification, *Proc. ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 917-926 (2009).

