情報処理学会論文誌　データベース　Vol. 3　No. 3　11–19 (Sep. 2010)

集合ラベルを持つデータの集約範囲の記述

古川 哲也[†1]　葛西 正裕[†2]

収集したデータを多面的に分析するためには，様々な範囲でデータを集約する必要がある．データの属性値を用いて集約を行う際，属性値に基づいてデータを階層的に分類しておくことが有用である．データは分類階層における複数のカテゴリのラベルに対応することが多く，属性値はラベル集合になる．そのようなデータはラベル集合を用いて集約されるため，本論文では，ラベル集合の順序を導入することで，ラベル集合が記述するデータ集合を明らかにする．ラベル集合は，ラベル集合のいずれかのラベルを持つのかすべてのラベルを持つのか，ラベル集合の範囲内かどうかの組合せで 4 種類のデータを記述する．記述されたデータを集約し分析に利用する際には，異なるラベル集合は異なるデータを記述し，ラベル集合のデータはそれより上位のラベル集合によっても記述されなければならない．ラベル集合の順序はこのような性質を満たしているかどうかについても議論する．

Specifying Aggregation Range of Data with a Set of Labels

Tetsuya Furukawa[†1] and Masahiro Kuzunishi[†2]

Various range aggregations of colleded data enable flexible and advanced analysis of data. Hierarchical classification based on the an attribute value is an efficient approach to aggregate data. When each data is classified to multiple categories or annotated with a set of labels, multi-labeled data are aggregated by giving a set of labels. This paper shows which data is expressed by a set of labels by introducing orders for sets of labels. A set of labels specifies four types of data, which are characterized by whether the labels of expressed data includes every label of the given set of labels within the range of the set. To utilize data for analysis by agregation, different sets of labels should specify different data and data should be specified by the higher set of labels. This paper slso discusses those properties for the orders of sets of labels.

1. はじめに

複雑さを増す現象を分析し有益な情報を得るには，一面的なデータの検討だけなく，様々な視点からデータを分析する必要がある．すなわち，様々な範囲でデータを集約し多面的にデータを分析しなければならない．データはそれが何の値であるのかをその属性によって説明されているものとし，属性値を用いて集約することで分析のための基礎データを得る．データはその属性値で階層的に分類しておくことが有用であり[1),2),12),17)]，そのようなデータの構成は，検索エンジンのカテゴリ分類などにも用いられている．階層的分類では，データには分類されたカテゴリのラベルが付される．すなわち，ラベルを用いてデータの集約範囲を記述する．

一般にはデータはただ 1 つのカテゴリに分類され，データには単一のラベルが付される[2),3)]．たとえば，ニュース記事を階層的に分類している *Newsgroups* では，それぞれの記事はただ 1 つのカテゴリに分類される[15)]．データの整理が目的であれば，このような分類でのデータの検索は単純で分かりやすく，データの構成法としても容易である．しかし，データはただ 1 つのカテゴリに分類できる単純なものばかりではない．たとえば，東京とニューヨークの両方に関係するデータを地域で分類する場合，“東京” または “ニューヨーク” のいずれかに分類するのは適切ではなく，両方のカテゴリに分類することで，東京とニューヨークの両方に関係するデータとして分析される．複数のカテゴリにデータを分類する場合には，通常，複数のラベルをデータに付する分類が行われる[11),14),15)]．そのような分類では，東京とニューヨークの両方に関係したデータは，“東京” と “ニューヨーク” の両方のカテゴリに分類され，{ 東京, ニューヨーク } のラベルが付される．

ラベルが付されたデータを集約するには，ラベルが記述するデータの集約範囲を明確にしておかなければならない．ラベルが記述するデータは，与えられたラベルと同じラベルが付されているデータ，あるいは，与えられたラベルの概念と同じか下位の概念のラベルが付されているデータと定義される[9)]．たとえば，“日本” というラベルが記述するデータは，“日本” というラベルが付されたデータ，あるいは “日本”，“関東”，“福岡” などのラベルが付されたデータである．階層的に分類されたデータが対象のときには，一般に後者が用いられ

†1 九州大学大学院経済学研究院
Faculty of Economics, Kyushu University
†2 愛知学院大学商学部
Faculty of Business and Commerce, Aichi Gakuin University

ており，与えられたラベルはそのラベルの概念以下となるラベルが付されたデータを記述するものとする．単一のラベルを付する分類ではラベルが記述するデータは明らかであるが，複数のラベルを付する分類では，ラベルの集合が記述するデータをデータの集約目的に応じて考える必要がある．

例 1　ラベル集合 $\boldsymbol{L}$ を { 日本, 米国 } とする．$\boldsymbol{L}$ が記述するデータを「日本と米国のみに関するデータ」とすると，{ 東京, ニューヨーク } といったラベルが付されたデータが該当する．一方で，「日本と米国に関するデータ」として，日本や米国に無関係なラベルを持つデータ，たとえば，{ 東京, ニューヨーク, パリ } のようにパリを持つデータも含めたい場合がある．また，「日本あるいは米国のどちらかのみに関するデータ」や「日本あるいは米国に関するデータ」として，米国に関するラベルを持たないデータ，たとえば，{ 東京 } や { 東京, パリ } のようなデータも該当させたい場合がある．

ラベル集合が記述するデータの範囲には様々あるが，それに関する議論はあまりなされていない．近年の分類に関する研究では，ウェブページやテキストの分類を複数のラベルで行うものもあるが[8),13),18)]，研究の主な目的はデータにラベルを自動的に付することであり，分類後のデータはラベル集合の各ラベルに対応するデータの和集合や積集合として求められる．ラベル集合の利用に関する研究では，キーワードの集合，すなわちラベル集合を用いてデータを抽出する研究があるが[4)–6)]，それらは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．そのような研究では，ラベル集合で記述されるデータは，「すべてのラベルに関係するデータ」あるいは「いくつかのラベルに関係するデータ」として利用されるのが一般的で，ラベル集合が記述するデータの範囲を精緻に議論したものではない．

複数のラベルが付されたデータを高度に利用するためには，ラベル集合に基づいてデータを記述することになる．たとえば，経済分析において産業別にデータを分析する際には，複数の業種を統合する必要があり，ラベル集合が記述するデータを集約する．その際，ラベル集合によってどのようなデータの記述が可能か，記述されたデータはどのようなものかを明確にしておかなければならない．本論文の目的は，ラベル集合が付されたデータを集約する際に，どのような集約が可能かを明らかにすることである．

ラベル集合が記述するデータは，ラベル集合の順序を定義することによって明確にされる[10)]．すなわち，ラベル集合はそれ以下のラベルが付されたデータを記述する．異なる集約範囲の記述は，ラベル集合の順序をどのように定義するかによって形式化される．

本論文は以下のように構成される．2 章でラベル集合の順序を導き，3 章でラベル集合が記述するデータにはどのようなものがあるかについて検討する．4 章および 5 章では，ラベル集合で記述したデータを利用する際の性質について議論する．6 章はまとめである．

2. ラベル集合の順序

データの分類はそれに用いる属性ごとに階層的に行われるものとする．たとえば，地域という属性で，国，都市といった分類が行われる．複数の属性に対する分類は論文 7), 13) などで議論されており，本論文では特定の属性でのデータの分類を対象とする．また，分類に用いる分類階層はオントロジなどによりあらかじめ与えられており，データの分類は木構造の分類階層に基づいてなされるものとする．

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 $\overline{L}$ で表し，分類によってオブジェクト o に付されたラベルを $\widetilde{o}$ で表す．オブジェクトは，あらかじめ与えられた分類階層で対応する最も下位のカテゴリに分類される[1),8),11)]．$\widetilde{o}$ は，o が分類される最も下位のカテゴリのラベル（複数のカテゴリに分類されている場合はラベル集合）である．オブジェクトは分類階層の最下位ではないカテゴリに分類されることもある[8),9),16)]．たとえば，日本に関するオブジェクトのラベルは国レベルであり，最下位のカテゴリが都市レベルの分類階層で分類を行った場合，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，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，$L_1 \prec L_2$ で表す．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．すなわち，$\prec$ は分類階層におけるラベルの順序を示している．オブジェクトのラベルが単一のとき，オブジェクトが $\overline{L}$ 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，$\widetilde{o} \preceq L$ ならば o は $\overline{L}$ の要素となる．すなわち，$\overline{L} = \{o \mid \widetilde{o} \preceq L\}$ である．

オブジェクトが分類階層での複数のカテゴリへ分類される場合は，オブジェクトのラベルは複数（ラベル集合）になる．そのようなラベルを集合ラベルという．集合ラベルのオブジェクトに対して，単一のラベルによって記述されるオブジェクト集合は，単一ラベルとラベル集合の順序によって決定される．ラベル集合は一般に積または和で解釈され，それに対応した順序となる．

(1)　積（Conjunction）の順序：ラベル L とラベル集合 $\boldsymbol{L}$ に対し，$\boldsymbol{L}$ におけるすべてのラベルが L 以下ならば，$\boldsymbol{L}$ は L の下位であり，$\boldsymbol{L} \preceq_C L$ で表す．

(2)　和（Disjunction）の順序：ラベル L とラベル集合 $\boldsymbol{L}$ に対し，$\boldsymbol{L}$ におけるあるラベルが L 以下ならば，$\boldsymbol{L}$ は L の下位であり，$\boldsymbol{L} \preceq_D L$ で表す．

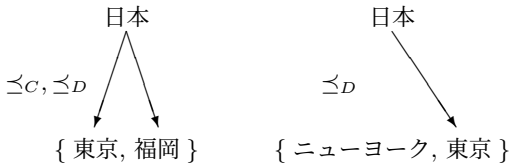

図 1　単一ラベルとラベル集合の順序
Fig. 1　An order between a label and a set of labels.

例 2　**図 1** は，ラベル集合 { 東京, 福岡 } と { ニューヨーク, 東京 } が日本の下位であることを示している．日本から東京と福岡への矢印は，分類階層で日本が東京や福岡の上位概念であることを表している．積の順序では，東京と福岡は日本の下位なので { 東京, 福岡 } は日本の下位になるが，{ ニューヨーク, 東京 } のニューヨークは日本の下位ではないので { ニューヨーク, 東京 } は日本の下位ではない．和の順序では，{ 東京, 福岡 } と { ニューヨーク, 東京 } はともに日本の下位のラベルを含むので日本の下位である．

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 $\boldsymbol{L}$ で記述されるオブジェクト集合を $\overline{\boldsymbol{L}}$ で表す．ラベル集合 $\boldsymbol{L}$ は一般に $\boldsymbol{L}$ 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．

まず，積の順序をラベル集合の順序に拡張する．ラベル集合 $\boldsymbol{L}$ 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 $\overline{\boldsymbol{L}}^{CI}$，和集合を $\overline{\boldsymbol{L}}^{CU}$ で表す．すなわち，

$$\overline{\boldsymbol{L}}^{CI} = \bigcap_{L \in \boldsymbol{L}} \{o \mid \widetilde{o} \preceq_C L\}$$
$$\overline{\boldsymbol{L}}^{CU} = \bigcup_{L \in \boldsymbol{L}} \{o \mid \widetilde{o} \preceq_C L\}$$

である．

ラベル集合 $\boldsymbol{L}$ によって記述されるオブジェクト集合は $\boldsymbol{L}$ 以下の集合ラベルを持つオブジェクトからなるとするためには，ラベル集合の順序が必要になる．$\overline{\boldsymbol{L}}^{CI}$ と $\overline{\boldsymbol{L}}^{CU}$ に対応する順序は以下のように定義できる．

定義 1　ラベル集合 $\boldsymbol{L}_1$，$\boldsymbol{L}_2$ に対し，

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

である．

$\overline{\boldsymbol{L}}^{CI}$，$\overline{\boldsymbol{L}}^{CU}$ は，それぞれラベル集合の順序 $\preceq_{CI}$ と $\preceq_{CU}$ を用いて表すことができる．

定理 1　ラベル集合 $\boldsymbol{L}$ に対し，$\overline{\boldsymbol{L}}^{CI} = \{o \mid \widetilde{o} \preceq_{CI} \boldsymbol{L}\}$，$\overline{\boldsymbol{L}}^{CU} = \{o \mid \widetilde{o} \preceq_{CU} \boldsymbol{L}\}$ である．

（証明）　$\overline{\boldsymbol{L}}^{CI} = \bigcap_{L \in \boldsymbol{L}} \{o \mid \widetilde{o} \preceq_C L\}$ の要素は $\forall L \in \boldsymbol{L}$，$\widetilde{o} \preceq_C L$ となるオブジェクト o であり，$\preceq_C$ の定義より $\forall L \in \boldsymbol{L}$，$\forall L' \in \widetilde{o}$，$L' \preceq L$ である．よって，$\overline{\boldsymbol{L}}^{CI}$ は $\{o \mid \forall L \in \boldsymbol{L}, \forall L' \in \widetilde{o}, L' \preceq L\}$ と表せるので，$\preceq_{CI}$ の定義より，$\overline{\boldsymbol{L}}^{CI} = \{o \mid \widetilde{o} \preceq_{CI} \boldsymbol{L}\}$ である．同様に $\overline{\boldsymbol{L}}^{CU}$ は $\{o \mid \exists L \in \boldsymbol{L}, \forall L' \in \widetilde{o}, L' \preceq L\}$ と表せるので，$\preceq_{CU}$ の定義より，$\overline{\boldsymbol{L}}^{CU} = \{o \mid \widetilde{o} \preceq_{CU} \boldsymbol{L}\}$ でる．　（証明終）

積の順序と同様に和の順序についてもラベル集合に拡張すると，それらは

$$\overline{\boldsymbol{L}}^{DI} = \bigcap_{L \in \boldsymbol{L}} \{o \mid \widetilde{o} \preceq_D L\}$$
$$\overline{\boldsymbol{L}}^{DU} = \bigcup_{L \in \boldsymbol{L}} \{o \mid \widetilde{o} \preceq_D L\}$$

となる．

定義 2　ラベル集合 $\boldsymbol{L}_1$，$\boldsymbol{L}_2$ に対し，

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

である．

ラベル集合 $\boldsymbol{L}$ に対し，$\overline{\boldsymbol{L}}^{DI}$ と $\overline{\boldsymbol{L}}^{DU}$ に含まれるオブジェクトは，それぞれラベル集合の順序 $\preceq_{DI}$ と $\preceq_{DU}$ を用いて表すことができる．

定理 2　ラベル集合 $\boldsymbol{L}$ に対し，$\overline{\boldsymbol{L}}^{DI} = \{o \mid \widetilde{o} \preceq_{DI} \boldsymbol{L}\}$，$\overline{\boldsymbol{L}}^{DU} = \{o \mid \widetilde{o} \preceq_{DU} \boldsymbol{L}\}$ である．

（証明）　定理 1 の証明と同様に，$\overline{\boldsymbol{L}}^{DI}$ と $\overline{\boldsymbol{L}}^{DU}$ は，それぞれ $\{o \mid \forall L \in \boldsymbol{L}, \exists L' \in \widetilde{o}, L' \preceq L\}$ と $\{o \mid \exists L \in \boldsymbol{L}, \exists L' \in \widetilde{o}, L' \preceq L\}$ と表すことができ，それらは定義 2 より $\{o \mid \widetilde{o} \preceq_{DI} \boldsymbol{L}\}$，$\{o \mid \widetilde{o} \preceq_{DU} \boldsymbol{L}\}$ である．　（証明終）

単一ラベルを用いて集合ラベルのオブジェクトを記述するための順序をラベル集合による記述に拡張した．一方で，ラベル集合を用いた単一ラベルのオブジェクトの記述を集合ラベルのオブジェクトに拡張することもできる．

単一ラベルのオブジェクトを記述するラベル集合は，単一のラベルで記述されるオブジェクト集合の積集合 $\bigcap_{L \in \boldsymbol{L}} \overline{L}$ あるいは和集合 $\bigcup_{L \in \boldsymbol{L}} \overline{L}$ を表す．

積集合を表す場合を集合ラベルのオブジェクトに拡張する．オブジェクトの集合ラベルを積とするとき，オブジェクト o は，$\widetilde{o}$ のすべてのラベル L' が $\boldsymbol{L}$ に対する積集合の性質を満たせば $\boldsymbol{L}$ 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

$$\overline{\boldsymbol{L}}^{IC} = \bigcap_{L \in \boldsymbol{L}} \{o \mid \forall L' \in \widetilde{o}, L' \preceq L\}$$

とする．

同様に，ラベル集合 $\boldsymbol{L}$ を積集合，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 $\boldsymbol{L}$ を和集合，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，次のように表すことができる．

$\overline{\boldsymbol{L}}^{ID} = \bigcap_{L \in \boldsymbol{L}} \{o \mid \exists L' \in \widetilde{o}, L' \preceq L\}$

$\overline{\boldsymbol{L}}^{UC} = \bigcup_{L \in \boldsymbol{L}} \{o \mid \forall L' \in \widetilde{o}, L' \preceq L\}$

$\overline{\boldsymbol{L}}^{UD} = \bigcup_{L \in \boldsymbol{L}} \{o \mid \exists L' \in \widetilde{o}, L' \preceq L\}$

ラベル集合 $\boldsymbol{L}$ によって記述されるオブジェクト集合を $\boldsymbol{L}$ 以下のラベルのオブジェクトと考えると，$\overline{\boldsymbol{L}}^{IC}$，$\overline{\boldsymbol{L}}^{ID}$，$\overline{\boldsymbol{L}}^{UC}$，$\overline{\boldsymbol{L}}^{UD}$ に対応する順序は次のものとなる．

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

$\forall L_1 \in \boldsymbol{L}_1, \forall L_2 \in \boldsymbol{L}_2, L_1 \preceq L_2 \iff \boldsymbol{L}_1 \preceq_{IC} \boldsymbol{L}_2$

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

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

$\exists L_1 \in \boldsymbol{L}_1, \exists L_2 \in \boldsymbol{L}_2, L_1 \preceq L_2 \iff \boldsymbol{L}_1 \preceq_{UD} \boldsymbol{L}_2$

である．

定理 3　ラベル集合 $\boldsymbol{L}$ に対し，$\overline{\boldsymbol{L}}^{IC} = \{o \mid \widetilde{o} \preceq_{IC} \boldsymbol{L}\}$，$\overline{\boldsymbol{L}}^{ID} = \{o \mid \widetilde{o} \preceq_{ID} \boldsymbol{L}\}$，$\overline{\boldsymbol{L}}^{UC} = \{o \mid \widetilde{o} \preceq_{UC} \boldsymbol{L}\}$，$\overline{\boldsymbol{L}}^{UD} = \{o \mid \widetilde{o} \preceq_{UD} \boldsymbol{L}\}$ である．

（証明）　$\overline{\boldsymbol{L}}^{IC}$ は $\{o \mid \forall L' \in \widetilde{o}, \overline{L'} \in \bigcap_{L \in \boldsymbol{L}} \overline{L}\}$ なので，$\{o \mid \forall L' \in \widetilde{o}, \forall L \in \boldsymbol{L}, L' \preceq L\}$ である．よって，$\preceq_{IC}$ の定義より，$\overline{\boldsymbol{L}}^{IC} = \{o \mid \widetilde{o} \preceq_{IC} \boldsymbol{L}\}$ である．同様に，$\overline{\boldsymbol{L}}^{UC}$ は $\{o \mid \forall L' \in \widetilde{o}, \exists L \in \boldsymbol{L}, L' \preceq L\}$ であり，$\preceq_{UC}$ の定義より $\{o \mid \widetilde{o} \preceq_{UC} \boldsymbol{L}\}$ である．$\overline{\boldsymbol{L}}^{ID}$ と $\overline{\boldsymbol{L}}^{UD}$ についての証明はそれぞれ $\overline{\boldsymbol{L}}^{IC}$ と $\overline{\boldsymbol{L}}^{UC}$ についての証明と同様である．（証明終）

3. ラベル集合が記述する集約範囲

2 章ではラベル集合の順序を導いた．本章では，それらの順序がどのようなオブジェクトを記述しているかについて検討し，オブジェクトの記述のための順序を求め，ラベル集合による集約範囲を明らかにする．

ラベル集合 $\boldsymbol{L}$ によって記述されるオブジェクト o は $\boldsymbol{L}$ と $\widetilde{o}$ の順序によって決定されるが，$\boldsymbol{L}$ と $\widetilde{o}$ の中には順序の決定に関与しないラベルが存在しうる．

例 3　ラベル集合 $\boldsymbol{L}_1$ とオブジェクト o_1 のラベル $\widetilde{o_1}$ をそれぞれ { 日本, 米国 } と { 東京, ニューヨーク, パリ } とする．$\widetilde{o_1}$ は $\boldsymbol{L}_1$ 中の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 $\overline{\boldsymbol{L}_1}^{DI}$ の要素であり，$\widetilde{o_1}$ のパリは要素の決定に関与しない（**図 2** (a)）．すなわち，$\widetilde{o_1}$ は $\boldsymbol{L}_1$ 中のすべてのラベルに対してそれ以下のラベルを含むが，$\boldsymbol{L}_1$ と無関係なラベルを含んでいてもよい．一方で，$\overline{\boldsymbol{L}_1}^{UC}$ に含まれるオブジェクトの集合ラベルには $\boldsymbol{L}_1$ 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．たとえば，集合ラベルが { 東京, 福岡 } であるオブジェクト o_2 は $\boldsymbol{L}_1$ の米国以下のラベルを含まないにもかかわらず，o_2 は $\overline{\boldsymbol{L}_1}^{UC}$ の要素である（図 2 (b)）．$\overline{\boldsymbol{L}_1}^{DU}$ では，$\overline{\boldsymbol{L}_1}^{DU}$ 中のオブジェクトの集合ラベルと $\boldsymbol{L}_1$ 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 (c)）．

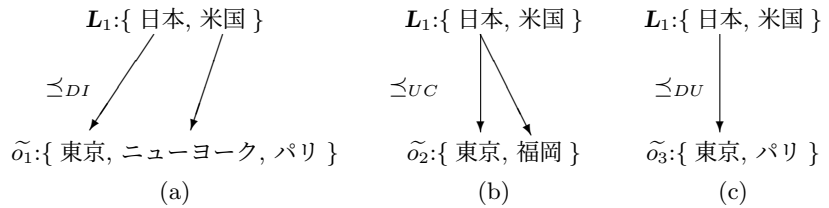

図 2　順序の決定に関与しないラベル
Fig. 2　Labels unrelated to the order of sets of labels.

ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対し，$\boldsymbol{L}_2$ のすべてのラベルについてそのラベル以下のラベルが $\boldsymbol{L}_1$ に含まれており $\boldsymbol{L}_1 \preceq_{DI} \boldsymbol{L}_2$ であれば，$\boldsymbol{L}_2$ には $\boldsymbol{L}_1$ 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，$\boldsymbol{L}_1 \preceq_{DI} \boldsymbol{L}_2$ では上位のラベル集合 $\boldsymbol{L}_2$ に制限がある．同様に，$\boldsymbol{L}_1 \preceq_{UC} \boldsymbol{L}_2$ は下位のラベル集合 $\boldsymbol{L}_1$ に制限があり，$\boldsymbol{L}_1 \preceq_{DU} \boldsymbol{L}_2$（$= \boldsymbol{L}_1 \preceq_{UD} \boldsymbol{L}_2$）は，上位と下位のどちらにも制限はない．$\preceq_{DI}$，$\preceq_{UC}$，$\preceq_{DU}$（$=\preceq_{UD}$）を，それぞれ $\preceq_{RU}$，$\preceq_{RL}$，$\preceq_{RN}$ で表すこととする．

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

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

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

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

$\preceq_{ID}$，$\preceq_{IC}$，$\preceq_{CI}$，$\preceq_{CU}$ 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 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対し，$\boldsymbol{L}_1 \preceq_{ID} \boldsymbol{L}_2$ または $\boldsymbol{L}_1 \preceq_{IC} \boldsymbol{L}_2$ であれば，$\boldsymbol{L}_1$ のあるラベルまたは $\boldsymbol{L}_1$ のすべてのラベルが $\boldsymbol{L}_2$ のすべてのラベルの下位である．$\boldsymbol{L}_2$ 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，$\boldsymbol{L}_2$ によって記述されるオブジェクトは存在しない．$\boldsymbol{L}_2$ にそのようなラベルがないとき，$\boldsymbol{L}_2$ は $\boldsymbol{L}_2$ のうち最も下位の 1 つのラベルとすることができる．

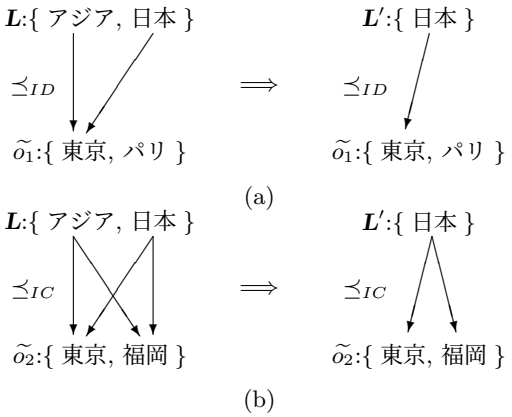

図 3　ラベルの縮約
Fig. 3　Reduction of labels.

例 4　ラベル集合 $\boldsymbol{L}=\{$ アジア, 日本 $\}$ に対して，**図 3** (a)，(b) はそれぞれ $\overline{\boldsymbol{L}}^{ID}$ のオブジェクト o_1 と $\overline{\boldsymbol{L}}^{IC}$ のオブジェクト o_2 を示している．$\widetilde{o_1}$ の東京と $\widetilde{o_2}$ の東京，福岡は $\boldsymbol{L}$ のすべてのラベルに対して下位である．日本は $\boldsymbol{L}$ における最も下位のラベルであり，$\boldsymbol{L}$ は 1 つのラベル日本からなる $\boldsymbol{L}'$ とすることができる．

$\preceq_{ID}$ や $\preceq_{IC}$ によってラベル集合 $\boldsymbol{L}$ でオブジェクトを記述する場合には，$\overline{\boldsymbol{L}}^{ID}\neq\phi$ かつ $\overline{\boldsymbol{L}}^{IC}\neq\phi$ ならば $\boldsymbol{L}$ を 1 つのラベルに縮約できる．$|\boldsymbol{L}|=1$ であれば，$\overline{\boldsymbol{L}}^{ID}=\overline{\boldsymbol{L}}^{RU}$，$\overline{\boldsymbol{L}}^{IC}=\overline{\boldsymbol{L}}^{RL}$ であり，$\overline{\boldsymbol{L}}^{ID}$ と $\overline{\boldsymbol{L}}^{IC}$ は，それぞれ $\overline{\boldsymbol{L}}^{RU}$ と $\overline{\boldsymbol{L}}^{RL}$ 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 $\preceq_{RU}$ や $\preceq_{RL}$ 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 $\boldsymbol{L}$ 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，$\overline{\boldsymbol{L}}^{CU}=\bigcup_{L\in\boldsymbol{L}}\overline{\{L\}}^{CU}$ は，$\boldsymbol{L}$ 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 $\boldsymbol{L}$ であれば，$\boldsymbol{L}$ の各ラベルで表されるオブジェクトの直和になる．したがって，$\boldsymbol{L}$ は個々のラベルごとに扱うことができ，$|\boldsymbol{L}|=1$ のとき $\overline{\{L\}}^{CU}=\overline{\{L\}}^{RL}$ なので，$\preceq_{CU}$ は $\preceq_{RL}$ で考えることができる．

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うえでそのほかに必要な順序がないかを検討する．ラベル集合 $\boldsymbol{L}_1$ は $\boldsymbol{L}_2$ 以下とする順序を，$\boldsymbol{L}_1\preceq_x\boldsymbol{L}_2$ かつ $\boldsymbol{L}_1\preceq_y\boldsymbol{L}_2$（$x,y\in\{CI,CU,DI,DU,IC,ID,UC,UD\}$）とする．$x=DI$，$y=UC$ あるいは $x=UC$，$y=DI$ 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一致する．たとえば，$x=CI$，$y=CU$ で定義される順序は，$\boldsymbol{L}_1\preceq_x\boldsymbol{L}_2$ かつ $\boldsymbol{L}_1\preceq_y\boldsymbol{L}_2$ であれば $\boldsymbol{L}_1\preceq_{CI}\boldsymbol{L}_2$ であり，$\preceq_{CI}$ に一致する．

$\preceq_{DI}$ と $\preceq_{UC}$ はそれぞれ $\preceq_{RU}$ 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．$\preceq_{RB}$ でラベル集合 $\boldsymbol{L}$ によって記述されるオブジェクト集合を $\overline{\boldsymbol{L}}^{RB}$ とする．$\overline{\boldsymbol{L}}^{RB}$ は $\overline{\boldsymbol{L}}^{RB}=\{o\mid\widetilde{o}\preceq_{RB}\boldsymbol{L}\}=\{o\mid\widetilde{o}\preceq_{RU}\boldsymbol{L},\widetilde{o}\preceq_{RL}\boldsymbol{L}\}$ と表すことができるので，$\preceq_{RB}$ は次のように定義される．

ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対し，$\boldsymbol{L}_2$ におけるすべてのラベルに対して $\boldsymbol{L}_1$ におけるあるラベルがそのラベル以下であり，かつ $\boldsymbol{L}_1$ におけるすべてのラベルに対して $\boldsymbol{L}_2$ におけるあるラベルがそのラベル以上であるとき，$\boldsymbol{L}_1$ は $\boldsymbol{L}_2$ 以下である，すなわち，

$$\forall L_2\in\boldsymbol{L}_2,\exists L_1\in\boldsymbol{L}_1,L_1\preceq L_2\ and\ \forall L_1\in\boldsymbol{L}_1,\exists L_2\in\boldsymbol{L}_2,L_1\preceq L_2$$
$$\iff\quad\boldsymbol{L}_1\preceq_{RB}\boldsymbol{L}_2$$

である．

ラベル集合で集約範囲を記述するための順序は $\preceq_{RN}$，$\preceq_{RU}$，$\preceq_{RL}$，$\preceq_{RB}$ である．ラベル集合 $\boldsymbol{L}$ に対し，$\overline{\boldsymbol{L}}^{RN}$ と $\overline{\boldsymbol{L}}^{RL}$ は $\boldsymbol{L}$ を和集合を表すものとしており，$\overline{\boldsymbol{L}}^{RU}$ と $\overline{\boldsymbol{L}}^{RB}$ は積集合を表すものとしている．また，$\overline{\boldsymbol{L}}^{RN}$ と $\overline{\boldsymbol{L}}^{RU}$ は $\boldsymbol{L}$ と無関係なラベルを持つオブジェクトも要素とするが，$\overline{\boldsymbol{L}}^{RL}$ と $\overline{\boldsymbol{L}}^{RB}$ はそのようなオブジェクトを含まない，すなわち，要素となるオブジェクトのラベル集合の範囲を限定している．

ラベル集合による集約範囲の記述では，すべてのラベルに関係するオブジェクトを集約範囲とするか，いずれかのラベルに関係するオブジェクトを集約範囲とするかが考えうる．また，集合ラベル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の選択もある．これらは，ラベル集合を積集合とするか和集合とするか，ラベル集合の範囲を限定するかしないかに対応しており，ラベル集合の 4 種類の順序で十分な記述ができる（**図 4**）．

例 5　ラベル集合 $\boldsymbol{L}=\{$ 日本, 米国 $\}$ に対して，$\overline{\boldsymbol{L}}^{RN}$ は $\{$ 東京 $\}$，$\{$ 東京, ニューヨーク $\}$，$\{$ 東京, ニューヨーク, パリ $\}$，$\overline{\boldsymbol{L}}^{RL}$ は $\{$ 東京 $\}$，$\{$ 東京, ニューヨーク $\}$ のような $\boldsymbol{L}$ の要素を和集合としているオブジェクトの集合であるのに対し，$\overline{\boldsymbol{L}}^{RU}$ は $\{$ 東京, ニューヨーク $\}$，$\{$ 東京, ニューヨーク, パリ $\}$，$\overline{\boldsymbol{L}}^{RB}$ は $\{$ 東京, ニューヨーク $\}$ のような積集合としているオブジェクトの集合である．また，$\overline{\boldsymbol{L}}^{RN}$ と $\overline{\boldsymbol{L}}^{RU}$ には $\{$ 東京, ニューヨーク, パリ $\}$

		範囲の制限	
		なし	あり
ラベル集合	和集合	RN	RL
	積集合	RU	RB

図 4　ラベル集合の解釈と範囲
Fig. 4　Interpretation and rage of sets of labels.

といった日本や米国とは無関係なパリを集合ラベルに含むオブジェクトも含まれるのに対し，$\overline{\boldsymbol{L}}^{RL}$ と $\overline{\boldsymbol{L}}^{RB}$ は { 東京, ニューヨーク } といった日本と米国の範囲内の集合ラベルのオブジェクトの集合である．

4. ラベル集合の縮約

ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対し，$\overline{\boldsymbol{L}_1}^{x}$ と $\overline{\boldsymbol{L}_2}^{x}$（$x \in \{RN, RU, RL, RB\}$）の集約結果を比較することで $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ の違いを分析することができる．しかし，$\overline{\boldsymbol{L}_1}^{x} = \overline{\boldsymbol{L}_2}^{x}$ であれば $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ の比較は意味をなさない．本章では，集約結果が必ず一致するラベル集合の性質を示す．

$\boldsymbol{L}_1 - \boldsymbol{L}_2$ 中のラベル L に対して，$\boldsymbol{L}_1$ によって記述されるオブジェクト集合は，一般に L が存在するために $\boldsymbol{L}_2$ 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，$\boldsymbol{L}_1 \cap \boldsymbol{L}_2$ 中に L 以下のラベルが存在するならば，$\preceq_{RU}$ では，$\overline{\boldsymbol{L}_1}^{RU}$ の要素であるが $\overline{\boldsymbol{L}_2}^{RU}$ の要素ではないようなオブジェクトは，L のために存在することはない．

例 6　ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ をそれぞれ { アジア, 日本 } と { 日本 } とする．$\boldsymbol{L}_1 \cap \boldsymbol{L}_2$ 中の日本は $\boldsymbol{L}_1 - \boldsymbol{L}_2$ 中のアジアの下位なので，$\overline{\boldsymbol{L}_1}^{RU}$ の要素ではあるが $\overline{\boldsymbol{L}_2}^{RU}$ の要素ではないようなオブジェクトは存在しない．

ラベル間に上下関係のないラベル集合に限定すれば，異なるラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対して，$\overline{\boldsymbol{L}_1}^{x}$ と $\overline{\boldsymbol{L}_2}^{x}$（$x \in \{RN, RU, RL, RB\}$）は異なる．ラベル集合 $\boldsymbol{L}$ 中のラベルに上下関係がないとき，$\boldsymbol{L}$ を排他であるとする．排他でない集合ラベルは，データの分類において必要なものである．たとえば，日本の経済状況をアジア全体の経済状況と比較するデータのラベルは { アジア, 日本 } になる．

ラベル集合 $\boldsymbol{L}_1$ または $\boldsymbol{L}_2$（$\boldsymbol{L}_1 \neq \boldsymbol{L}_2$）が排他でないとき，$\boldsymbol{L}_1 \preceq_{RU} \boldsymbol{L}_2$ かつ $\boldsymbol{L}_2 \preceq_{RU} \boldsymbol{L}_1$ となる場合がある．そのようなラベル集合は $\preceq_{RU}$ で等位であるといい，$\boldsymbol{L}_1 \approx_{RU} \boldsymbol{L}_2$ で表す．

例 7　ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ をそれぞれ { 日本, 米国 } と { アジア, 日本, 米国 } とする．$\boldsymbol{L}_1 \preceq_{RU} \boldsymbol{L}_2$ かつ $\boldsymbol{L}_2 \preceq_{RU} \boldsymbol{L}_1$ なので，$\boldsymbol{L}_1 \approx_{RU} \boldsymbol{L}_2$ である（**図 5**）．

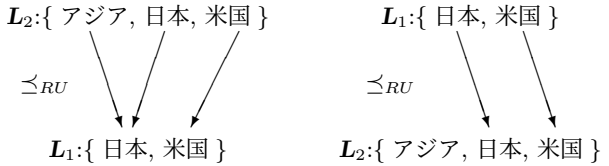

図 5　$\preceq_{RU}$ での等位性
Fig. 5　Equivalence by $\preceq_{RU}$.

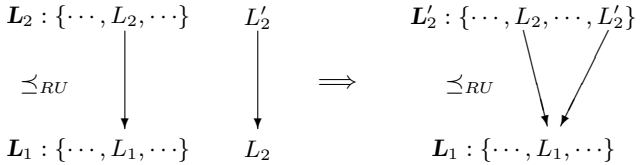

図 6　$\preceq_{RU}$ での冗長なラベル
Fig. 6　Redundant labels for $\preceq_{RU}$.

$\boldsymbol{L}_1 \preceq_{RU} \boldsymbol{L}_2$ であるようなラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対して，$L_2 \in \boldsymbol{L}_2$，$L_2' \notin \boldsymbol{L}_2$，$L_2 \preceq L_2'$ であるようなラベルを L_2 と L_2' とする．$\boldsymbol{L}_1$ 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 $\boldsymbol{L}_1 \preceq_{RU} \boldsymbol{L}_2 \cup \{L_2'\}$ である．よって，$\overline{\boldsymbol{L}_2}^{RU}$ と $\overline{\boldsymbol{L}_2'}^{RU}$ は同じオブジェクト集合である（**図 6**）．すなわち，$\preceq_{RU}$ では，ラベル集合 $\boldsymbol{L}$ は $\boldsymbol{L}$ 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 $\boldsymbol{L}$ の部分集合と同じオブジェクト集合を記述する．

排他でないラベル集合 $\boldsymbol{L}$ について，$\boldsymbol{L}$ 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 $\boldsymbol{L}$ の部分集合を $\boldsymbol{L}$ の下限とし，$l(\boldsymbol{L})$ で表す．

$$l(\boldsymbol{L}) = \{L \mid L \in \boldsymbol{L}, \forall L' \in \boldsymbol{L}\ s.t.\ L' \neq L, L' \not\prec L\}$$

補題 1　ラベル集合 $\boldsymbol{L}$ に対し，$\boldsymbol{L} \approx_{RU} l(\boldsymbol{L})$ である．

（証明）　$l(\boldsymbol{L})$ におけるすべてのラベルは，$\boldsymbol{L}$ におけるあるラベルに対してそのラベル以下なので，$\forall L \in \boldsymbol{L}$，$\exists L' \in l(\boldsymbol{L})$，$L' \preceq L$ であり，$\preceq_{RU}$ の定義より $l(\boldsymbol{L}) \preceq_{RU} \boldsymbol{L}$ である．また，$l(\boldsymbol{L}) \subseteq \boldsymbol{L}$ なので，$\forall L \in l(\boldsymbol{L})$，$\exists L' \in \boldsymbol{L}$，$L = L'$ である．よって，$\boldsymbol{L} \preceq_{RU} l(\boldsymbol{L})$ なので $\boldsymbol{L} \approx_{RU} l(\boldsymbol{L})$ である．　（証明終）

$\preceq_{RU}$ では，排他でないラベル集合 $\boldsymbol{L}$ で記述されるオブジェクト集合は，$\boldsymbol{L}$ の下限で記述されるオブジェクト集合に等しい．

定理 4　ラベル集合 $\boldsymbol{L}$ に対し，$\overline{\boldsymbol{L}}^{RU} = \overline{l(\boldsymbol{L})}^{RU}$ である．

（証明）　$\overline{\boldsymbol{L}}^{RU}$ におけるオブジェクト o に対して，補題 1 より $\widetilde{o} \preceq_{RU} \boldsymbol{L} \approx_{RU} l(\boldsymbol{L})$ なので，o は $\overline{l(\boldsymbol{L})}^{RU}$ に含まれる．よって，$\overline{\boldsymbol{L}}^{RU} \subseteq \overline{l(\boldsymbol{L})}^{RU}$ である．同様に，$\overline{l(\boldsymbol{L})}^{RU}$ におけるオブジェクト o に対して，$o \in \overline{\boldsymbol{L}}^{RU}$ であり，$\overline{l(\boldsymbol{L})}^{RU} \subseteq \overline{\boldsymbol{L}}^{RU}$ である．よって，$\overline{\boldsymbol{L}}^{RU} = \overline{l(\boldsymbol{L})}^{RU}$ である．　（証明終）

ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$（$\boldsymbol{L}_1 \neq \boldsymbol{L}_2$）に対して，$l(\boldsymbol{L}_1) = l(\boldsymbol{L}_2)$ ならば，定理 4 より $\overline{\boldsymbol{L}_1}^{RU} = \overline{\boldsymbol{L}_2}^{RU}$ なので，$\overline{\boldsymbol{L}_1}^{RU} = \overline{\boldsymbol{L}_2}^{RU}$ である．

ラベル集合の下限と同様に，排他でないラベル集合 $\boldsymbol{L}$ 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 $\boldsymbol{L}$ の部分集合を $\boldsymbol{L}$ の上限とし，$u(\boldsymbol{L})$ で表す．

$$u(\boldsymbol{L}) = \{L \mid L \in \boldsymbol{L}, \forall L' \in \boldsymbol{L}\ s.t.\ L' \neq L, L \not\prec L'\}$$

$\prec_{RL}$ や $\preceq_{RN}$ では，$\prec_{RU}$ と同様の議論で，$\overline{\boldsymbol{L}}^{RL} = \overline{u(\boldsymbol{L})}^{RL}$，$\overline{\boldsymbol{L}}^{RN} = \overline{u(\boldsymbol{L})}^{RN}$ となる．$\overline{\boldsymbol{L}}^{RB}$ についても，$\boldsymbol{L}_1 \approx_{RB} \boldsymbol{L}_2$ となるようなラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ が存在する．

例 8　ラベル集合 $\boldsymbol{L}_1 =$ { アジア, 日本, 東京 }，$\boldsymbol{L}_2 =$ { アジア, 東京 } で，$\boldsymbol{L}_1 \preceq_{RB} \boldsymbol{L}_2$ かつ $\boldsymbol{L}_2 \preceq_{RB} \boldsymbol{L}_1$ なので，$\boldsymbol{L}_1 \approx_{RB} \boldsymbol{L}_2$ である（**図 7**）．

$ul(\boldsymbol{L})$ を $u(\boldsymbol{L}) \cup l(\boldsymbol{L})$ とする．図 7 で示した $\boldsymbol{L}_2$ は，$u(\boldsymbol{L}_2) \cup l(\boldsymbol{L}_2)$，すなわち $ul(\boldsymbol{L}_2)$ である $\boldsymbol{L}_1$ に縮約できる．

定理 5　ラベル集合 $\boldsymbol{L}$ に対し，$\overline{\boldsymbol{L}}^{RB} = \overline{ul(\boldsymbol{L})}^{RB}$ である．

（証明）　$ul(\boldsymbol{L})$ は $u(\boldsymbol{L})$ を含んでいるので，$\boldsymbol{L}$ 中のすべての L に対して，$L \preceq L'$ であるような $ul(\boldsymbol{L})$ 中のラベル L' が存在し，$\forall L \in \boldsymbol{L}$，$\exists L' \in ul(\boldsymbol{L})$，$L \preceq L'$ である．$ul(\boldsymbol{L})$ は $\boldsymbol{L}$ の部分集合なので，$\forall L \in ul(\boldsymbol{L})$，$\exists L' \in \boldsymbol{L}$，$L' \preceq L$ である．よって，$\preceq_{RB}$ の定義より $\boldsymbol{L} \preceq_{RB} ul(\boldsymbol{L})$ であり，$\overline{\boldsymbol{L}}^{RB}$ のオブジェクト o は $\overline{ul(\boldsymbol{L})}^{RB}$ のオブジェクトでもあるので，$\overline{\boldsymbol{L}}^{RB} \subseteq \overline{ul(\boldsymbol{L})}^{RB}$ である．

同様に，$ul(\boldsymbol{L})$ は $l(\boldsymbol{L})$ を含んでいるので，$\boldsymbol{L}$ 中のすべての L に対して，$L' \preceq L$ であるような $ul(\boldsymbol{L})$ 中の L' が存在し，$\forall L \in \boldsymbol{L}$，$\exists L' \in ul(\boldsymbol{L})$，$L' \preceq L$ である．$ul(\boldsymbol{L})$ は $\boldsymbol{L}$ の部分集合なので，$\forall L \in ul(\boldsymbol{L})$，$\exists L' \in \boldsymbol{L}$，$L \preceq L'$ である．よって，$ul(\boldsymbol{L}) \preceq_{RB} \boldsymbol{L}$ であり，$\overline{ul(\boldsymbol{L})}^{RB}$ のオブジェクト o は $\overline{\boldsymbol{L}}^{RB}$ のオブジェクトでもあるので，$\overline{ul(\boldsymbol{L})}^{RB} \subseteq \overline{\boldsymbol{L}}^{RB}$ である．　（証明終）

$\preceq_{RU}$ でラベル集合 $\boldsymbol{L}$ が記述するオブジェクトは，$\boldsymbol{L}$ の下限が記述するオブジェクトであり，$\preceq_{RL}$ や $\preceq_{RN}$ でラベル集合 $\boldsymbol{L}$ が記述するオブジェクトは，$\boldsymbol{L}$ の上限が記述するオブジェクトである．$\preceq_{RB}$ でラベル集合 $\boldsymbol{L}$ が記述するオブジェクトは，$\boldsymbol{L}$ 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，ラベル集合が排他でない場合には，ラベル集合をその上限または下限に縮約して考えなければならない．

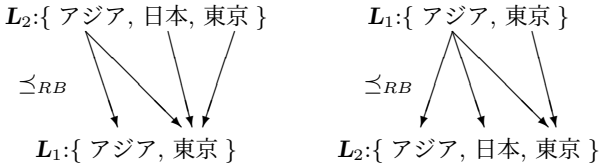

図 7　$\preceq_{RB}$ での等位性
Fig. 7　Equivalence by $\preceq_{RB}$.

5. 順序の健全性

4 章では $\overline{\boldsymbol{L}_1}^{x} = \overline{\boldsymbol{L}_2}^{x}$ となるためにラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ の比較による分析ができない原因をラベル集合の縮約で対処できることを示した．本章では，$\overline{\boldsymbol{L}_1}^{x} \neq \overline{\boldsymbol{L}_2}^{x}$ であっても $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ の比較による分析ができない場合について検討する．

ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対し，$\overline{\boldsymbol{L}_1}^{x} - \overline{\boldsymbol{L}_2}^{x} \neq \phi$ かつ $\overline{\boldsymbol{L}_2}^{x} - \overline{\boldsymbol{L}_1}^{x} \neq \phi$（$x \in \{RN, RU, RL, RB\}$）のとき，これらの差なども分析対象となりうるが，$\overline{\boldsymbol{L}_1}^{x}$ と $\overline{\boldsymbol{L}_2}^{x}$ の集約結果を比較してもその差が何のために生じたものなのか判断できないため，$\overline{\boldsymbol{L}_1}^{x}$ と $\overline{\boldsymbol{L}_2}^{x}$ を直接比較できない．$\overline{\boldsymbol{L}_1}^{x} \subseteq \overline{\boldsymbol{L}_2}^{x}$ であれば，$\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ の違いは $\overline{\boldsymbol{L}_1}^{x} - \overline{\boldsymbol{L}_2}^{x}$ によって分析できる．

単一のラベルでの分類では，ラベルの順序は分類階層におけるカテゴリの順序と一致する．すなわち，ラベル L_1 がラベル L_2 よりも下位ならば，分類階層における L_1 のカテゴリは L_2 のカテゴリよりも下位である．分類階層は概念の順序を示しているため，ラベルの順序と概念の順序は一致しているので，$L_1 \preceq L_2$ ならば，$\overline{L_1} \subseteq \overline{L_2}$ である．オブジェクトが集合ラベルを持つとき，単一ラベルでの分類と同様に，ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対して $\boldsymbol{L}_1$ が $\boldsymbol{L}_2$ 以下ならば $\overline{\boldsymbol{L}_1}$ におけるオブジェクトは $\overline{\boldsymbol{L}_2}$ の要素であれば，ラベル集合の順序は階層構造を規定する．

定義 4　ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対し，$\boldsymbol{L}_1 \preceq_x \boldsymbol{L}_2$（$x \in \{RN, RU, RL, RB\}$）であることが $\overline{\boldsymbol{L}_1} \subseteq \overline{\boldsymbol{L}_2}$ であることの必要十分条件であるとき，$\preceq_x$ は健全である．

$\preceq_{RN}$ は健全ではない．$\boldsymbol{L}_1 \preceq_{RN} \boldsymbol{L}_2$ であるようなラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ に対して，$\boldsymbol{L}_2$ のいずれのラベルに対しても順序にないラベル L_1 が $\boldsymbol{L}_1$ 中にある場合がある．L_1 以下であるようなラベルのオブジェクトは $\overline{\boldsymbol{L}_1}^{RN}$ の要素である．しかし，そのようなオブジェクトのラベルにはラベル集合 $\boldsymbol{L}_2$ のいずれかのラベルに対して順序があるラベルを含んでいるとは限らない．そのような場合にはそのオブジェクトは $\overline{\boldsymbol{L}_2}^{RN}$ の要素ではない．すなわち，$\overline{\boldsymbol{L}_1}^{RN}$

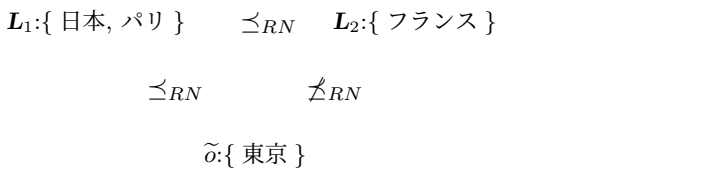

図 8　$\preceq_{RN}$ の健全性
Fig. 8　Soundness of $\preceq_{RN}$.

の要素であるが $\overline{\boldsymbol{L}_2}^{RN}$ の要素ではないようなオブジェクトが存在する．

例 9　ラベル集合 $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ をそれぞれ { 日本, パリ } と { フランス } とする．ラベルが { 東京 } であるオブジェクト o は，東京は日本以下のラベルなので $\overline{\boldsymbol{L}_1}^{RN}$ の要素であるが，東京はフランス以下のラベルではないので o は $\overline{\boldsymbol{L}_2}^{RN}$ の要素ではない（**図 8**）．

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 2　順序が推移律を満たすことは，順序が健全であることの必要十分条件である．

（証明）　順序 $\preceq_x$（$x \in \{RN, RU, RL, RB\}$）は推移律を満たすとする．ラベル集合 $\boldsymbol{L}_1$ とオブジェクト o に対して，$\widetilde{o} \preceq_x \boldsymbol{L}_1$ ならば，o は $\overline{\boldsymbol{L}_1}^x$ に含まれる．また，$\boldsymbol{L}_1 \preceq_x \boldsymbol{L}_2$ であるようなラベル集合 $\boldsymbol{L}_2$ に対して，$\widetilde{o} \preceq_x \boldsymbol{L}_1$ と $\boldsymbol{L}_1 \preceq_x \boldsymbol{L}_2$ から $\widetilde{o} \preceq_x \boldsymbol{L}_2$ なので，o は $\overline{\boldsymbol{L}_2}^x$ の要素である．$\overline{\boldsymbol{L}_1}^x$ のすべてのオブジェクトは $\overline{\boldsymbol{L}_2}^x$ のオブジェクトなので，$\overline{\boldsymbol{L}_1}^x \subseteq \overline{\boldsymbol{L}_2}^x$ である．一方，$\overline{\boldsymbol{L}_1}^x \subseteq \overline{\boldsymbol{L}_2}^x$ ならば，$\overline{\boldsymbol{L}_1}^x$ におけるオブジェクト o は $\overline{\boldsymbol{L}_2}^x$ に含まれる．また，$\widetilde{o} = \boldsymbol{L}_1$ であるとき $\widetilde{o} \preceq_x \boldsymbol{L}_2$ であり $\boldsymbol{L}_1 \preceq_x \boldsymbol{L}_2$ である．したがって，$\preceq_x$ が推移律を満たせば $\preceq_x$ は健全である．

$\boldsymbol{L}_1 \preceq_x \boldsymbol{L}_2$ かつ $\boldsymbol{L}_2 \preceq_x \boldsymbol{L}_3$ であるようなラベル集合 $\boldsymbol{L}_1$，$\boldsymbol{L}_2$，$\boldsymbol{L}_3$ に対して，$\preceq_x$ が健全ならば $\overline{\boldsymbol{L}_1}^x \subseteq \overline{\boldsymbol{L}_2}^x$ かつ $\overline{\boldsymbol{L}_2}^x \subseteq \overline{\boldsymbol{L}_3}^x$ である．$\overline{\boldsymbol{L}_1}^x \subseteq \overline{\boldsymbol{L}_3}^x$ なので，定義 4 より $\boldsymbol{L}_1 \preceq_x \boldsymbol{L}_3$ が成り立つため，$\preceq_x$ が健全ならば $\preceq_x$ は推移律を満たす．　（証明終）

例 9 で示したように，$\widetilde{o} \preceq_{RN} \boldsymbol{L}_1$，$\boldsymbol{L}_1 \preceq_{RN} \boldsymbol{L}_2$ であるが，$\widetilde{o} \not\preceq_{RN} \boldsymbol{L}_2$ であり，$\preceq_{RN}$ は推移律を満たさない．それに対して，$\preceq_{RU}$，$\preceq_{RL}$，$\preceq_{RB}$ は推移律を満たす．

補題 3　順序 $\preceq_{RU}$，$\preceq_{RL}$，$\preceq_{RB}$ は推移律を満たす．

（証明）　$\boldsymbol{L}_1 \preceq_{RU} \boldsymbol{L}_2$ かつ $\boldsymbol{L}_2 \preceq_{RU} \boldsymbol{L}_3$ であるようなラベル集合 $\boldsymbol{L}_1$，$\boldsymbol{L}_2$，$\boldsymbol{L}_3$ に対して，$\forall L_2 \in \boldsymbol{L}_2$，$\exists L_1 \in \boldsymbol{L}_1$，$L_1 \preceq L_2$ かつ $\forall L_3 \in \boldsymbol{L}_3$，$\exists L_2 \in \boldsymbol{L}_2$，$L_2 \preceq L_3$ なので，$\forall L_3 \in \boldsymbol{L}_3$，$\exists L_1 \in \boldsymbol{L}_1$，$L_1 \preceq L_3$ である．よって，$\preceq_{RU}$ の定義より，$\boldsymbol{L}_1 \preceq_{RU} \boldsymbol{L}_3$ なので $\preceq_{RU}$ は推移律を満たす．$\preceq_{RL}$ と $\preceq_{RB}$ についても $\preceq_{RU}$ と同様である．　（証明終）

$\preceq_{RU}$，$\preceq_{RL}$，$\preceq_{RB}$ は推移律を満たすので健全である．

定理 6　順序 $\preceq_{RU}$，$\preceq_{RL}$，$\preceq_{RB}$ は健全である．

（証明）　補題 2，3 より $\preceq_{RU}$，$\preceq_{RL}$，$\preceq_{RB}$ は健全である．　（証明終）

$\boldsymbol{L}_1 \preceq_x \boldsymbol{L}_2$ となる $\overline{\boldsymbol{L}_1}^x$ と $\overline{\boldsymbol{L}_2}^x$ の比較は $x = RN$ では意味をなさないが，定理 6 より，$x \in \{RU, RL, RB\}$ では $\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ の違いを分析できる．

6. むすび

ラベル集合でオブジェクトの集約範囲を記述するための順序は，$\preceq_{RN}$，$\preceq_{RU}$，$\preceq_{RL}$，$\preceq_{RB}$ であることを示し，それらがどのようなオブジェクト集合を記述しているのかを明らかにした．集約範囲は，与えたラベル集合を和集合か積集合のどちらで解釈するのかとオブジェクトの集合ラベルが与えたラベル集合の範囲内であるかどうかで決定される．従来のラベル集合による記述は，ラベル集合の各ラベルが記述するオブジェクト集合の和集合と積集合で得られるオブジェクト集合であり，ラベル集合の範囲を考えないものであった．それに対して，$\preceq_{RL}$ と $\preceq_{RB}$ はラベル集合の範囲内であるオブジェクト集合を記述している．

本論文で明らかにした順序により集約範囲の考え方が整理され，様々な集約範囲による分析が明確になる．たとえば，ラベル集合 { 日本, 米国 } が $\preceq_{RU}$ で記述するオブジェクト集合と $\preceq_{RB}$ で記述するオブジェクト集合の集約値を比較することで，日本と米国の結び付きの開放性が分かるので，日本と米国による広域経済圏のグローバル化の程度などが分析できる．

また，ラベル集合で記述された集約範囲を分析に用いる際の性質について検討した．ラベル集合が排他であれば異なるラベル集合は異なる集約範囲を記述し，排他でなければ縮約したラベル集合を用いる必要がある．$\preceq_{RU}$，$\preceq_{RL}$，$\preceq_{RB}$ は健全であり，ラベル集合 $\boldsymbol{L}_1$，$\boldsymbol{L}_2$ に対して，$\boldsymbol{L}_1 \preceq_x \boldsymbol{L}_2$ と $\overline{\boldsymbol{L}_1}^x \subseteq \overline{\boldsymbol{L}_2}^x$ は等価で，$\boldsymbol{L}_1$ と $\boldsymbol{L}_2$ の違いは $\overline{\boldsymbol{L}_2}^x - \overline{\boldsymbol{L}_1}^x$ として分析できる．

オブジェクトの利用を考えるとオブジェクトのラベルは排他ではないこともあるので，ラベル集合を排他に限定すべきではない．特に経済分析では，近年の経済活動のグローバル化にともなう経済現象の複雑化によって，分析の対象となる経済事例といったオブジェクトのラベルは排他ではないことが多い．たとえば，経済事例を地域という属性で分類した際には，広域経済圏，道州，県，市といった様々なレベルに複数関連するオブジェクトがある．そのような排他ではないオブジェクトを利用する際には，本論文で明らかにしたようにラベル集合を上限や下限に縮約することによって適切な記述が可能となる．

参 考 文 献

1) 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).
2) Bertino, E., Fan, J., Ferrari, E., Hachi, M. and Elamagarmid, A.: A Hierarchical Access Control Model for Video Database Systems, *ACM Trans. Inf. Syst.*, Vol.21, No.2, pp.151–191 (2003).
3) 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).
4) 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).
5) Chuang, S. and Chien, L.: Taxonomy Generation for Text Segments: A Practical Web-Based Approach, *ACM Trans. Inf. Syst.*, Vol.23, No.4, pp.363–396 (2005).
6) 崔　超遠，陳　漢雄，古瀬一隆，大保信夫：グローバル分析とローカル分析に基づく検索支援，情報処理学会論文誌：データベース，Vol.45, No.SIG14 (TOD24), pp.54–63 (2004).
7) 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).
8) 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).
9) 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).
10) 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).
11) Ghamrawi, N. and McMallum, A.: Collecticve Multi-Label Classification, *Proc. Int'l Conf. on Information and Knowledge Management* (*CIKM'05*), pp.195–200 (2005).
12) 葛西正裕，古川哲也：階層的分類における複数の意味を持つデータの利用，情報処理学会論文誌：データベース，Vol.47, No.SIG8 (TOD30), pp.1–10 (2006).
13) Sun, A. and Lim, E.: Hierarchical Text Classification and Evaluation, *Proc. IEEE Int'l Conf. on Data Mining* (*ICDM2001*), pp.521–528 (2001).
14) 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).
15) Toutanova, K., Chen, F., Popat K. and Hofmann, T.: Text Classification in a Hierarchical Mixture Model for Small Training Sets, *Proc. Int'l Conf. on Information and Knowledge Management* (*CIKM'01*), pp.105–112 (2001).
16) 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).
17) Wang, Y. and Oyama, K.: Web Page Classification Based on Surrounding Page Model Representing Connection Type and Directory Hierarchy, *IPSJ Trans. Databases*, Vol.2, No.2, pp.29–43 (2009).
18) Yang, B., Sun, J., Wang, T. and 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).

(平成 22 年 3 月 19 日受付)
(平成 22 年 7 月 6 日採録)

(担当編集委員　樋口 健)

古川　哲也（正会員）

昭和 58 年京都大学工学部情報工学科卒業．昭和 60 年同大学院工学研究科修士課程修了．昭和 63 年九州大学大学院工学研究科博士後期課程修了．工学博士．九州大学工学部助手，大型計算機センター講師，同助教授，経済学部助教授を経て，現在同大学院経済学研究院教授．この間，平成 7 年米国パデュ大学客員研究員．データベースの設計論，情報システムの研究に従事．電子情報通信学会，ACM，IEEE，日本データベース学会等各会員．

葛西　正裕（正会員）

平成 15 年九州大学経済学部経済工学科卒業．平成 17 年同大学院経済学府修士課程修了．平成 20 年同大学院経済学府博士後期課程単位取得退学．平成 19 年財団法人九州経済調査協会調査研究部研究員．平成 21 年愛知学院大学商学部専任講師．情報システムにおけるデータ利用や情報システムを用いた経済分析の研究に従事．日本経済政策学会会員．