

■ 論文

ランク付ラベル集合による
分析を目的としたデータの指定と性質

葛西正裕

目次

- 1 はじめに
- 2 ランクの導入
- 3 ランク付ラベルに対応するオブジェクト
- 4 ランク付ラベル集合に対応するオブジェクト
- 5 分析に必要な性質の検証
- 6 おわりに

参考文献

▶ 概要

迅速かつ高度な経済分析を行う必要性が増す中、膨大に蓄積される多様なデータを利用した分析手法の確立が求められている。多様化したデータを一元的に分析に供するには、データを階層的に分類しておくことが有用であり、データには分類階層における複数のカテゴリのラベルが付される。その際、各カテゴリに対するデータの関連性の強さは一般に不均一なため、関連性の強弱を反映させた精緻な分析を行うためには、関連性の強弱を示すランクを有するラベルをデータに付しランク付ラベル集合を用いて分析対象となるデータを指定する理論的枠組みが必要である。そこで、本稿では、ランク付ラベル集合を有するデータに対する分析対象の考え方を述べた上で、ランク付ラベル集合による分析対象となるデータの指定法を体系化した。さらに、分析対象の考え方とランクの強弱に起因する分析対象となるデータ集合の包含関係を明らかにした。また、データを分析するために必要な妥当性（指定に用いるラベル集合が異なれば指定されるデータ集合も異なる）と健全性（指定に用いるラベル集合の上下関係と指定されるデータ集合の包含関係の一致）についても検証した。これらにより、分析対象となるデータ集合の拡張や限定といったことが精緻に行えるのと同時に、分析対象の増減といった変化を利用した分析が可能になる。

▶ キーワード

階層的分類、ランク付ラベル、データの指定、データの多様化、データベースモデル、データウェアハウス、経済分析手法

1 はじめに

複雑さを増し変化の激しい経済事象を分析していくには、迅速かつ高度な経済分析手法が必要である。従来の経済分析手法は、時間・費用・技術的側面の制約を受けるものが多い。一方、近年における情報通信技術の発達により、数値データのみならず、テキスト、画像、音声、動画等の多様なデータを収集できるようになっており分析に利用されている¹⁵⁾。経済分析においても、このような膨大に蓄積される多様なデータを利用した分析手法の確立が求められる。

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^{2) 5)}。その際、データの構成法として、データを階層的に分類しておくことが有用である^{3) 19)}。データの分類は、分類に用いる地域や業種といった属性ごとに行われ、データには予め与えられた属性ごとの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^{1) 4)}。

1件のデータには、対応する分類階層における複数のカテゴリのラベルが付される場合が多いが^{10) 16)}、各カテゴリに対するデータの関連性の強さは一般に不均一である。関連性の強さを考慮せずにラベルを付すと、データのカテゴリに対する関連性の強弱という不均一性を反映させた分析ができない。例えば、経済データを地域で分析する際、主に愛知に関するデータであるが、福岡についても若干関わりがあるようなデータに対して、愛知と福岡という単純に2つのラベルを付すと主に福岡に関するデータとの区別ができない。ゆえに、両方のデータを分析対象とした場合、福岡に対する関連性の強さを考慮しないために精緻な分析ができない。

データが分類の属性に対して持つ意味と分類階層のカテゴリとの関連性の強さを反映させた分析を行うためには、関連性の強弱を示すランクを有するラベルをデータに付した後、複数のランク付ラベル（ランク付ラベル集合）が付されたデータを分析対象に指定するための理論的枠組みが必要である。そこで、本稿では、ランク付ラベル集合のデータに対する分析対象の考え方を述べると同時に、ランク付ラベル集合による分析対象となるデータの指定法について議論する。その上で、分析対象の考え方とランクの強弱に応じて指定されるデータを整理する。さらに、データを分析するために必要な性質についても検証する。

データの分類におけるランクに関する関連研究は、主に情報検索の分野で研究がなされているが、検索結果の順位付けを目的としたものであり、ランク付けの正確性や効率性を評価するものが多い^{11) 17) 18)}。すなわち、関連研究の多くは、情報検索に重点が置かれており、分析という視点からランクを付した後のデータに対する利用についての議論は少ないため、ランク付ラベル集合が付されたデータに対する精緻な分析が行えない。また、これまでの既存研究⁷⁾では、データが分類階層のカテゴリに属する強度をランクとして与え、ランクを有するデータの構成法と利用法を提案してきたが、それらのデータを分析することにまで言及していない。

本稿は以下のように構成される。2節では、ランク付ラベルを導入しそれらを有するデータ

の性質を述べる。3節では、ランクを有する単数ラベル（ランク付ラベル）が指定する分析対象のデータについて述べる。4節では、3節の議論を複数のラベル集合（ランク付ラベル集合）に拡張し、ランク付ラベル集合によって指定される分析対象のデータを整理する。5節では、分析を行うために分析対象となるデータを指定する際に必要な性質として、妥当性（指定に用いるラベル集合が異なれば指定されるデータ集合も異なる）と健全性（指定に用いるラベル集合の上下関係と指定されるデータ集合の包含関係の一致）について検証する。6節はまとめである。

2 ランクの導入

データの分類は、分類に用いる属性に基づいて階層的に行われる。例えば、地域という属性では、東アジア、日本、東海、愛知といった階層的な分類が行われる。東海地域における自動車産業といった地域と業種のような複数の属性に関する分類では、属性ごとに分類した複数の分類階層を組み合わせる⁶⁾。複数の属性に分類されるデータの分析に関しては、既存研究¹⁴⁾で述べており、本稿では、議論を簡単にするために単一属性における分類を想定する。

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。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に応じて、 o は予め与えられた分類階層において o 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。すなわち、 o の意味に応じて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）が o に付される。ゆえに、 o のラベル集合は o の属性に対する意味を表している。ここで、ラベル集合を $L = \{L_1, \dots, L_n\}$ で表し、オブジェクト o のラベルを $L(o) = L$ とし、 o のラベルを集合ラベルと呼ぶ。

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。例えば、 $L_1 = \text{愛知}$ と $L_2 = \text{東海}$ に対して、 $L_1 \prec L_2$ となる。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半順序関係を示している。

オブジェクトに複数のカテゴリのラベルを付す際、オブジェクトが持つ意味とカテゴリとの関連性の強さをランクとして与えることで、分類後に精緻な分析が可能になる。本稿では、オブジェクトのカテゴリに対する関連性の強さを示すものをランクとして導入し、関連性の強さを区分する。ランクの区分が多いほど分類後により精緻な分析が可能になるが、本稿の議論はランクの区分を増やした場合にも同様に成り立つので、本稿では、ランクの区分を強弱の2つとし、主 (Main) と副 (Sub) に設定する。ラベル L が主の性質を持つ場合には主ラベルと呼び L^M で表し、 L が副の性質を持つ場合には副ラベルと呼び L^S で表す。

ランク付ラベル集合を $L^R = \{L_1^{R_1}, \dots, L_n^{R_n}\} (R_1, \dots, R_n \in \{M, S\})$ と表す。ランク付ラベル集合 L^R に対して、 L^R 中の主ラベルで構成されるラベル集合を $M(L^R) = \{L_i^{R_i} | L_i^{R_i} \in L^R, R_i = M\}$

で表し, L^R 中の副ラベルで構成されるラベル集合を $S(L^R) = \{L_i^{R_i} | L_i^{R_i} \in L^R, R_i = S\}$ で表す。オブジェクト o に付されるランク付ラベル集合をランク付集合ラベルと呼ぶ。

〔例1〕 地域経済に関する文献を地域という属性で分類する場合を考える。自動車産業における愛知と福岡を比較する文献 o_1 であれば, 地域という属性で分類をすると, 両地域を同等に扱っているオブジェクトとして両地域の主ラベルが付される。よって, o_1 のランク付集合ラベルは $L(o_1) = \{\text{愛知}^M, \text{福岡}^M\}$ となる。一方, 愛知を中心に扱っている中で福岡について触れている文献 o_2 であれば, o_2 には愛知の主ラベルと福岡の副ラベルが付されるので, $L(o_2) = \{\text{愛知}^M, \text{福岡}^S\}$ となる。

オブジェクト o に対して分類が行われ, o にランク付ラベル集合が付された場合, o のランク付集合ラベルには以下の性質が成り立つ。

〔性質1〕 ランク付集合ラベルのオブジェクト o に対して, $L(o) = M(L(o)) \cup S(L(o))$ である。

〔性質2〕 ランク付集合ラベルのオブジェクト o に対して, L_i^M かつ L_i^S ($L_i^M, L_i^S \in L(o)$) であるような $L(o)$ の要素は存在しない。

〔性質3〕 ランク付集合ラベルのオブジェクト o に対して, $M(L(o)) \neq \phi$ である。

性質1と2は, ランク付ラベル集合の定義より明らかな性質であり, ランク付ラベル集合をオブジェクトに付す際においても満たすべき性質である。性質1については, オブジェクトの分類時にランクが未定である場合や確定できない場合を除いて, ラベルは主ないしは副のランクを有する。また, 性質2については, 例えば, 愛知に関連している1件のオブジェクトが愛知という地域に対して主と副の両方の意味を持つことはないので, そのオブジェクトに愛知の主ラベルと愛知の副ラベルが同時に付されることはあり得ない。

性質3は, オブジェクトの分類時においてランク付ラベル集合を付す際に生じる性質である。オブジェクトを特定の属性において分類を行う際, オブジェクトがその属性下で分類できる, すなわち, いずれかのラベルの意味をオブジェクトが有しているので, ランク付集合ラベルのいずれかの要素は主ラベルになる。

3 ランク付ラベルに対応するオブジェクト

本節では, 分析対象を指定するランク付ラベル集合の要素が単数の場合を考え, それが指定

するランク付集合ラベルのオブジェクト集合について述べる。本節の議論を踏まえて、4節において指定に用いるランク付ラベル集合の要素を複数に拡張して議論する。

ランク付ラベル集合 L^R のうち要素が単数であるような $L^R = \{L_i^{R_1}\}$ ($R_1 = \{M, S\}$)を L^R と記述し、ランク付ラベルと呼ぶ。ランク付ラベル L^R で分析対象に指定されるランク付集合ラベルのオブジェクト o は、 o と L^R の関係で決まり、 o と L^R の関係は以下の2通りである。

(A) 分析対象となるランク付集合ラベルのオブジェクト o は、ランク付ラベル L^R 以外のラベルと関連していてもよい。

(B) 分析対象となるランク付集合ラベルのオブジェクト o は、ランク付ラベル L^R 以外のラベルと関連しない。

関連の有無をオブジェクトのランク付集合ラベルの要素と指定に用いるランク付ラベル間の順序の有無で考えると、(A)と(B)の関係は以下のように言い換えられる。

(A) 分析対象となるランク付集合ラベルのオブジェクト o におけるランク付集合ラベルのいずれかの要素が、ランク付ラベル L^R に対して同位または下位である。

(B) 分析対象となるランク付集合ラベルのオブジェクト o におけるランク付集合ラベルのすべての要素が、ランク付ラベル L^R に対して同位または下位である。

(A)と(B)の関係で、どのようなランク付集合ラベルのオブジェクトがランク付ラベルによって指定されるのかを決めるには、オブジェクトのランク付集合ラベルとランク付ラベルの順序が必要になるので、まず、ランク付ラベル同士の順序について以下のように定義する。

[定義1] $L_i \preceq L_j$ であるようなラベル L_i と L_j に対して、 L_i と L_j がランク付ラベルであるとき、 $L_i^M \preceq L_j^M$, $L_i^S \preceq L_j^S$, $L_i^S \preceq L_j^M$ とする。

概念の上下関係があるラベル同士について、同一ランクであれば順序があると考え、次に、概念の上下関係があるラベル同士でランクが異なる場合、上位のラベルが副ラベルで下位のラベルが主ラベルの際、主ラベルであれば副ラベルの性質を有すると考えて順序があると考え。逆の場合、すなわち、上位のラベルが主ラベルで下位のラベルが副ラベルの際、副ラベルには主ラベル程の性質を有さないことから順序がないと考える。よって、ランク付ラベル同士の順序では、概念が上位にあるラベルが下位にあるラベルに対してランクが弱い場合にも順序があるので、概念の上位下位の関係とランクの強弱が逆になる場合がある。

ランク付集合ラベルのオブジェクト o がランク付ラベル L^R で指定されるかは、 $L(o)$ と L^R の順序で決まる。ゆえに、ランク付ラベルとランク付ラベル集合の順序を定義する。

[定義2] ランク付ラベル集合 L^R とランク付ラベル L^R に対して、 $\exists L_i^{R_i} \in L^R, L_i^{R_i} \preceq L^R \iff L^R \preceq_U L^R$ であり、 $\forall L_i^{R_i} \in L^R, L_i^{R_i} \preceq L^R \iff L^R \preceq_I L^R$ とする。

ランク付ラベルとランク付ラベル集合の順序によってランク付ラベル L^R により記述されるオブジェクト集合を $\overline{L^R}$ で表す。ランク付集合ラベルのオブジェクト o が $\overline{L^R}$ の要素であることは、 L^R と $L(o)$ におけるランク付ラベルとランク付ラベル集合の順序によって決まり、 $L(o) \preceq L^R$ ならば o は $\overline{L^R}$ の要素となる。すなわち、 $\overline{L^R} = \{o \mid L(o) \preceq L^R\}$ である。

ランク付ラベル L^R が、(A) の関係で指定するランク付集合ラベルのオブジェクト集合を $\overline{L^R}^{(A)}$ とし、(B) の関係で指定するランク付集合ラベルのオブジェクト集合を $\overline{L^R}^{(B)}$ とする。

[補題 1] ランク付ラベル L^R に対して、 $\overline{L^R}^{(A)} = \{o \mid L(o) \preceq_U L^R\}$ であり、 $\overline{L^R}^{(B)} = \{o \mid L(o) \preceq_I L^R\}$ である。

(証明) ランク付ラベル L^R に対して、 $\overline{L^R}^{(A)}$ におけるランク付集合ラベルのオブジェクト o は、 o のランク付集合ラベルのいずれかの要素が、 L^R に対して同位または下位である。ゆえに、 L^R と $L(o)$ に対して、定義 2 のランク付ラベルとランク付ラベル集合の順序 \preceq_U の定義より $L(o) \preceq_U L^R$ が成り立つので、 $\overline{L^R}^{(A)} = \{o \mid L(o) \preceq_U L^R\}$ である。同様に、定義 2 のランク付ラベルとランク付ラベル集合の順序 \preceq_I の定義より $\overline{L^R}^{(B)} = \{o \mid L(o) \preceq_I L^R\}$ である。 (証明終)

これより、ランク付ラベルによって指定されるランク付集合ラベルのオブジェクト集合を、ラベルが各順序で記述するオブジェクト集合を明らかにした上で整理する。ランク付ラベルが分析対象に指定するランク付集合ラベルのオブジェクトは、例 2 の通りである。

[例 2] ランク付ラベル $\{\text{日本}^M\}$ 、 $\{\text{日本}^S\}$ が (A) と (B) の 2 通りの関係で指定される分析対象となるランク付集合ラベルのオブジェクト、すなわち、ランク付ラベルとランク付ラベル集合の順序 \preceq_U と \preceq_I で記述されるオブジェクトについて示したのが図 1 である。図 1 における各行は、1 件のオブジェクトが 2 通りの関係 (順序) と各ランク付ラベルで指定 (記述) されるかの有無を示している。図 1 の先頭行であれば、 $L(o) = \{\text{愛知}^M\}$ であるようなランク付集合ラベルのオブジェクト o は、(A) の関係 (順序 \preceq_U) においてランク付ラベル $L^R = \{\text{日本}^M\}$ で指定 (記述) される。同様に、(A) の関係 (順序 \preceq_U) においてランク付ラベル $L^R = \{\text{日本}^S\}$ で、(B)

行: $L(o)$ 列: L^R	$\overline{L^R}^{(A)}$ {日本 ^M }	$\overline{L^R}^{(A)}$ {日本 ^S }	$\overline{L^R}^{(B)}$ {日本 ^M }	$\overline{L^R}^{(B)}$ {日本 ^S }
{愛知 ^M }	○	○	○	○
{愛知 ^M , 福岡 ^M }	○	○	○	○
{愛知 ^M , 福岡 ^S }	○	○	×	○
{愛知 ^M , 北京 ^M }	○	○	×	×
{愛知 ^M , 北京 ^S }	○	○	×	×
{愛知 ^S , 北京 ^M }	×	○	×	×

図 1 ランク付ラベルが順序 \preceq_U と \preceq_I で記述するオブジェクト

の関係 (順序 \preceq_I) においてランク付ラベル $L^R = \{\text{日本}^M\}$ で、(B) の関係 (順序 \preceq_I) においてランク付ラベル $L^R = \{\text{日本}^S\}$ で指定 (記述) される。先頭行以外の行についても同様である。

ランク付ラベルが指定する分析対象となるランク付集合ラベルのオブジェクトの種類は、例 2 の図 1 で網羅されている。それは、以下の理由からである。まず、分析対象の指定に用いるランク付ラベルは、主と副の 2 つのランクであり、かつ (A) と (B) の 2 通りの関係で考えればよいので、その組合せである 4 通りについて検討すればよい。次に、ランク付ラベルで指定されるオブジェクトの種類については以下の通りである。性質 1 より、ランク付集合ラベルのオブジェクト o のランク付集合ラベルの要素は、主ラベルもしくは副ラベルかのいずれかなので、ランクの付されていない要素を含むランク付集合ラベルのオブジェクトは存在しない。性質 2 より、 $L(o)$ の要素は主ラベルないしは副ラベルのいずれかであるため、 $L(o) = \{\text{愛知}^M, \text{愛知}^S\}$ といったランク付集合ラベルのオブジェクトは存在しない。性質 3 より、 $L(o)$ のいずれかの要素は主ラベルになるので、 $L(o) = \{\text{愛知}^S\}$ や $L(o) = \{\text{愛知}^S, \text{福岡}^S\}$ といったランク付集合ラベルのオブジェクトは存在しない。これらの性質により、ランク付ラベルによって分析対象となるランク付集合ラベルのオブジェクトは、図 1 で示した 6 種類のオブジェクトで代表されている。

図 1 を基に、ランク付ラベルによって指定されるランク付集合ラベルのオブジェクト集合の包含関係についてまとめる。図 2 は、主ラベルないしは副ラベルで 2 通りの関係 (順序) が指定 (記述) するランク付集合ラベルのオブジェクト集合に対する包含関係を示している。

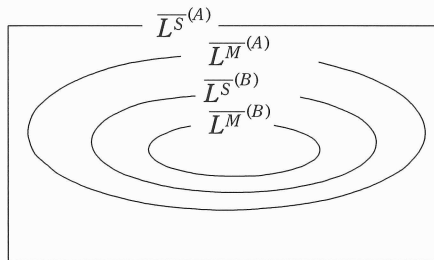


図 2 ランク付ラベルによって指定されるオブジェクト集合の包含関係

図 2 より、ランク付ラベルにおける主と副の 2 つのランク及び (A) と (B) の 2 通りの関係の組合せ 4 通りで分析対象に指定されるランク付集合ラベルのオブジェクト集合には包含関係があるので、分析対象を拡張させたり限定させることが的確にできる。

また、(A) の関係で指定する場合、すなわち、ランク付ラベルとランク付ラベル集合の順序 \preceq_U で記述する場合、ランク付ラベルに関連の無いラベルを含むランク付集合ラベルのオブジェクトも記述されることになる。この場合、関連が無いラベルの解釈は以下の 2 通りに分類できる。

1. ランク付ラベルに対して、概念の上下関係がないために関連の無いラベル
2. ランク付ラベルに対して、概念の上下関係はあるがランクが異なるので順序がないために関連の無いラベル

[例3] ランク付ラベル $L^R = \{\text{日本}^M\}$, $L(o_1) = \{\text{愛知}^M, \text{北京}^M\}$ となるようなオブジェクト o_1 , $L(o_2) = \{\text{愛知}^M, \text{福岡}^S\}$ となるようなオブジェクト o_2 について考える。ランク付ラベルとランク付ラベル集合の順序 \leq_U の場合, o_1 と o_2 はともに $\overline{L^R}^{(A)}$ の要素であるが, o_1 と o_2 のランク付集合ラベルには L^R とは関連のないラベルが含まれている。 o_1 については, $L(o_1) (= \{\text{愛知}^M, \text{北京}^M\})$ 中の “北京^M” は, $L^R (= \{\text{日本}^M\})$ の “日本^M” と概念の上下関係がないような無関連のラベルである。また, o_2 については, $L(o_2) (= \{\text{愛知}^M, \text{福岡}^S\})$ 中の “福岡^S” は, $L^R (= \{\text{日本}^M\})$ の “日本^M” と概念の上下関係はあるがランクが異なるため福岡^S $\not\leq$ 日本^M となるような無関連のラベルである。

指定に用いるランク付ラベルに対して、上記で述べたような関連の無いラベルを含むランク付集合ラベルのオブジェクトについても分析対象に含めたい場合には、オブジェクトの記述にはランク付ラベルとランク付ラベル集合の順序 \leq_U を用いて記述すればよい。逆に、そうしたオブジェクトを排除したい場合には、順序 \leq_I を用いて記述すればよい。

4 ランク付ラベル集合に対応するオブジェクト

3節で議論したランク付ラベルは、要素が単数のランク付ラベル集合なので、ラベル集合内の関係は1通りである。ゆえに、ランク付ラベル以外の関係、すなわち、ランク付ラベルに含まれない関係のみを議論した。本節では、分析対象を指定するラベルをランク付ラベル集合に拡張し、分析対象となるオブジェクトの考え方を、指定に用いるラベル集合に含まれるラベルとの関係と含まれないラベルとの関係の組合せで議論する¹²⁾¹³⁾。

ランク付ラベル集合 L^R に含まれないラベルとの関係は、3節の (A) と (B) の関係と同様に以下の通りになる。

(A) 分析対象となるランク付集合ラベルのオブジェクト o は、ランク付ラベル集合 L^R 以外のラベルと関連していてもよい。

(B) 分析対象となるランク付集合ラベルのオブジェクト o は、ランク付ラベル集合 L^R 以外のラベルと関連しない。

一方、ランク付ラベル集合 L^R に含まれるラベルとの関係でみると次のものがある。

(1) ランク付ラベル集合 L^R のいずれかのラベルに関連するランク付集合ラベルのオブジェクト o を分析対象とする。

(2) ランク付ラベル集合 L^R のすべてのラベルに関連するランク付集合ラベルのオブジェクト o

を分析対象とする。

よって、ランク付ラベル集合 L^R によって、分析対象に指定されるランク付集合ラベルのオブジェクト o の関係は、上記の関係を組合せたものになる。

(A-1) ランク付ラベル集合 L^R 以外のラベルと関連していてもよく、 L^R のいずれかのラベルに関連するランク付集合ラベルのオブジェクト o を分析対象とする。

(A-2) ランク付ラベル集合 L^R 以外のラベルと関連していてもよく、 L^R のすべてのラベルに関連するランク付集合ラベルのオブジェクト o を分析対象とする。

(B-1) ランク付ラベル集合 L^R 以外のラベルと関連せず、 L^R のいずれかのラベルに関連するランク付集合ラベルのオブジェクト o を分析対象とする。

(B-2) ランク付ラベル集合 L^R 以外のラベルと関連せず、 L^R のすべてのラベルに関連するランク付集合ラベルのオブジェクト o を分析対象とする。

ランク付ラベル集合間の順序によってランク付ラベル集合 L^R により記述されるランク付集合ラベルのオブジェクト集合を $\overline{L^R}$ で表す。ランク付集合ラベルのオブジェクト o に対して、 $L(o) \preceq L^R$ ならば、 o は $\overline{L^R}$ の要素となる。すなわち、 $\overline{L^R} = \{o \mid L(o) \preceq L^R\}$ である。

ランク付ラベル集合 L^R が、(A-1) の関係で指定するランク付集合ラベルのオブジェクト集合を $\overline{L^R}^{(A-1)}$ 、(A-2) の関係では $\overline{L^R}^{(A-2)}$ 、(B-1) の関係では $\overline{L^R}^{(B-1)}$ 、(B-2) の関係では $\overline{L^R}^{(B-2)}$ とする。これらに対応するランク付集合ラベルのオブジェクト集合を指定する際には、記述に用いるランク付ラベル集合とオブジェクトのランク付集合ラベルとの順序を定義する必要がある。ラベル集合間に対する順序は、文献⁸⁾で示した4通りの順序があり、これらをランク付集合ラベルに拡張した場合、以下の通りである。

[定義3] ランク付ラベル集合 L_1^R と L_2^R に対して、
 $\exists L_{i1}^{R_{i1}} \in L_1^R, \exists L_{i2}^{R_{i2}} \in L_2^R, L_{i1}^{R_{i1}} \preceq L_{i2}^{R_{i2}} \iff L_1^R \preceq_{RN} L_2^R,$
 $\forall L_{i2}^{R_{i2}} \in L_2^R, \exists L_{i1}^{R_{i1}} \in L_1^R, L_{i1}^{R_{i1}} \preceq L_{i2}^{R_{i2}} \iff L_1^R \preceq_{RU} L_2^R,$
 $\forall L_{i1}^{R_{i1}} \in L_1^R, \exists L_{i2}^{R_{i2}} \in L_2^R, L_{i1}^{R_{i1}} \preceq L_{i2}^{R_{i2}} \iff L_1^R \preceq_{RL} L_2^R,$
 $\forall L_{i2}^{R_{i2}} \in L_2^R, \exists L_{i1}^{R_{i1}} \in L_1^R, L_{i1}^{R_{i1}} \preceq L_{i2}^{R_{i2}} \text{ and } \forall L_{i1}^{R_{i1}} \in L_1^R, \exists L_{i2}^{R_{i2}} \in L_2^R, L_{i1}^{R_{i1}} \preceq L_{i2}^{R_{i2}} \iff$
 $L_1^R \preceq_{RB} L_2^R$ である。

ランク付ラベル集合の順序 \preceq_{RN} において、ランク付ラベル集合 L^R に対して同位または下位のランク付集合ラベルのオブジェクト集合を $\overline{L^R}^{RN}$ で示す。すなわち、 $\overline{L^R}^{RN} = \{o \mid L(o) \preceq_{RN} L^R\}$ である。同様に、ランク付ラベル集合の順序 \preceq_{RU} では $\overline{L^R}^{RU} = \{o \mid L(o) \preceq_{RU} L^R\}$ 、 \preceq_{RL} では $\overline{L^R}^{RL} = \{o \mid L(o) \preceq_{RL} L^R\}$ 、 \preceq_{RB} では $\overline{L^R}^{RB} = \{o \mid L(o) \preceq_{RB} L^R\}$ とする。

[定理1] ランク付ラベル集合 L^R に対して、 $\overline{L^R}^{(A-1)} = \overline{L^R}^{RN}$ 、 $\overline{L^R}^{(A-2)} = \overline{L^R}^{(RU)}$ 、

$\overline{L^R}^{(B-1)} = \overline{L^R}^{(RL)}$ であり、 $\overline{L^R}^{(B-2)} = \overline{L^R}^{(RB)}$ である。

(証明) ランク付ラベル集合 L^R に対して、 $\overline{L^R}^{(A-1)}$ 中のランク付集合ラベルのオブジェクト o は、 L^R 以外のラベルと関連していてもよく、 L^R のいずれかのラベルに関連している。これは、 $L(o)$ のいずれかの要素が、 L^R のいずれかの要素に対して同位または下位であると解釈できる。よって、 $L(o)$ と L^R に対して、定義3におけるランク付ラベル集合の順序 \preceq_{RN} の定義より、 $L(o) \preceq_{RN} L^R$ が成り立つので $\overline{L^R}^{(A-1)} = \{o \mid L(o) \preceq_{RN} L^R\} = \overline{L^R}^{RN}$ である。同様に、定義3のランク付ラベル集合の順序 \preceq_{RU} 、 \preceq_{RL} 、 \preceq_{RB} の定義より、それぞれ $\overline{L^R}^{(A-2)} = \overline{L^R}^{(RU)}$ 、 $\overline{L^R}^{(B-1)} = \overline{L^R}^{(RL)}$ 、 $\overline{L^R}^{(B-2)} = \overline{L^R}^{(RB)}$ が成り立つ。(証明終)

これより、4通りの分析対象の考え方で、どのようなランク付集合ラベルのオブジェクトが分析対象になるかについて明らかにした上で、それらを整理する。4通りの分析対象の考え方について、まず、ランク付集合ラベルのオブジェクトの指定に用いるランク付ラベル集合の要素が1つの場合、すなわち、ランク付ラベル $L^R (= L^R)$ による指定に関しては、(A)の関係では、 $\overline{L^R}^{(A)} = \{o \mid L(o) \preceq_{RN} L^R\} = \{o \mid L(o) \preceq_{RU} L^R\}$ であり、(B)の関係では、 $\overline{L^R}^{(B)} = \{o \mid L(o) \preceq_{RL} L^R\} = \{o \mid L(o) \preceq_{RB} L^R\}$ である。これらは、3節で議論した通りなので、本節は、要素が2つ以上のランク付ラベル集合を考慮することとし、要素が2つの場合について、いずれも主、一方が主で他方が副、いずれも副、といったすべての場合に区分して考える。

まず、ランク付ラベル集合が順序 \preceq_{RN} 、 \preceq_{RU} 、 \preceq_{RL} 、 \preceq_{RB} で記述するランク付集合ラベルのオブジェクトを以下のように例示する。

[例4] ランク付ラベル集合 $\{\text{日本}^M, \text{中国}^M\}$ 、 $\{\text{日本}^M, \text{中国}^S\}$ 、 $\{\text{日本}^S, \text{中国}^M\}$ 、 $\{\text{日本}^S, \text{中国}^S\}$ が4通りのランク付ラベル集合の順序で記述される分析対象となるランク付集合ラベルのオブジェクト、すなわち、 \preceq_{RN} 、 \preceq_{RU} 、 \preceq_{RL} 、 \preceq_{RB} で記述されるオブジェクトは、それぞれ図3、図4、図5、図6の通りである。

図3、図4、図5、図6における各行は、1件のランク付集合ラベルのオブジェクトが各図が表すランク付ラベル集合の順序において各ランク付ラベル集合で記述されるかの有無を示している。図3の先頭行であれば、 $L(o) = \{\text{愛知}^M\}$ であるようなランク付集合ラベルのオブジェクト o は、ランク付ラベル集合の順序 \preceq_{RN} において、ランク付ラベル集合 $L^R = \{\text{日本}^M, \text{中国}^M\}$ で記述される。同様に、ランク付ラベル集合 $L^R = \{\text{日本}^M, \text{中国}^S\}$ 、ランク付ラベル集合 $L^R = \{\text{日本}^S, \text{中国}^M\}$ 、ランク付ラベル集合 $L^R = \{\text{日本}^S, \text{中国}^S\}$ についても記述されることを表している。先頭行以外の行についても同様であり、図4、図5、図6は、図3と同じランク付ラベル集合とランク付集合ラベルのオブジェクトに対して、記述の有無を例示したものである。

行: $L(o)$ 列: L^R	{日本 ^M , 中国 ^M }	{日本 ^M , 中国 ^S }	{日本 ^S , 中国 ^M }	{日本 ^S , 中国 ^S }
{愛知 ^M }	○	○	○	○
{北京 ^M }	○	○	○	○
{愛知 ^M , 福岡 ^M }	○	○	○	○
{愛知 ^M , 福岡 ^S }	○	○	○	○
{愛知 ^M , ソウル ^M }	○	○	○	○
{愛知 ^S , ソウル ^M }	×	×	○	○
{北京 ^M , ソウル ^M }	○	○	○	○
{北京 ^S , ソウル ^M }	×	○	×	○
{愛知 ^M , 北京 ^M }	○	○	○	○
{愛知 ^M , 北京 ^S }	○	○	○	○
{愛知 ^S , 北京 ^M }	○	○	○	○
{愛知 ^M , 北京 ^M , ソウル ^M }	○	○	○	○
{愛知 ^M , 北京 ^S , ソウル ^M }	○	○	○	○
{愛知 ^S , 北京 ^M , ソウル ^M }	○	○	○	○
{愛知 ^S , 北京 ^S , ソウル ^M }	×	○	○	○

図3 ランク付ラベル集合が順序 \triangleleft_{RN} で記述するオブジェクト

行: $L(o)$ 列: L^R	{日本 ^M , 中国 ^M }	{日本 ^M , 中国 ^S }	{日本 ^S , 中国 ^M }	{日本 ^S , 中国 ^S }
{愛知 ^M }	×	×	×	×
{北京 ^M }	×	×	×	×
{愛知 ^M , 福岡 ^M }	×	×	×	×
{愛知 ^M , 福岡 ^S }	×	×	×	×
{愛知 ^M , ソウル ^M }	×	×	×	×
{愛知 ^S , ソウル ^M }	×	×	×	×
{北京 ^M , ソウル ^M }	×	×	×	×
{北京 ^S , ソウル ^M }	×	×	×	×
{愛知 ^M , 北京 ^M }	○	○	○	○
{愛知 ^M , 北京 ^S }	×	○	×	○
{愛知 ^S , 北京 ^M }	×	×	○	○
{愛知 ^M , 北京 ^M , ソウル ^M }	○	○	○	○
{愛知 ^M , 北京 ^S , ソウル ^M }	×	○	×	○
{愛知 ^S , 北京 ^M , ソウル ^M }	×	×	○	○
{愛知 ^S , 北京 ^S , ソウル ^M }	×	×	×	○

図4 ランク付ラベル集合が順序 \triangleleft_{RU} で記述するオブジェクト

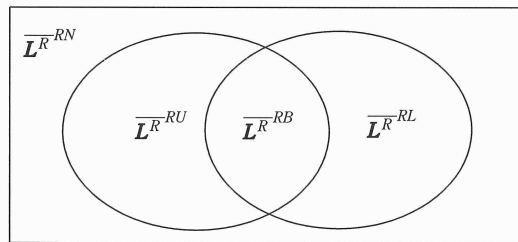
行: $L(o)$ 列: L^R	{日本 ^M , 中国 ^M }	{日本 ^M , 中国 ^S }	{日本 ^S , 中国 ^M }	{日本 ^S , 中国 ^S }
{愛知 ^M }	○	○	○	○
{北京 ^M }	○	○	○	○
{愛知 ^M , 福岡 ^M }	○	○	○	○
{愛知 ^M , 福岡 ^S }	×	×	○	○
{愛知 ^M , ソウル ^M }	×	×	×	×
{愛知 ^S , ソウル ^M }	×	×	×	×
{北京 ^M , ソウル ^M }	×	×	×	×
{北京 ^S , ソウル ^M }	×	×	×	×
{愛知 ^M , 北京 ^M }	○	○	○	○
{愛知 ^M , 北京 ^S }	×	○	×	○
{愛知 ^S , 北京 ^M }	×	×	○	○
{愛知 ^M , 北京 ^M , ソウル ^M }	×	×	×	×
{愛知 ^M , 北京 ^S , ソウル ^M }	×	×	×	×
{愛知 ^S , 北京 ^M , ソウル ^M }	×	×	×	×
{愛知 ^S , 北京 ^S , ソウル ^M }	×	×	×	×

図5 ランク付ラベル集合が順序 \triangleleft_{RL} で記述するオブジェクト

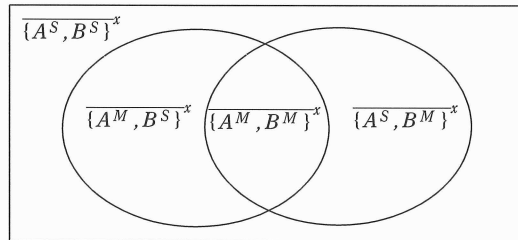
行: L^O 列: L^R	$\{\text{日本}^M, \text{中国}^M\}$	$\{\text{日本}^M, \text{中国}^S\}$	$\{\text{日本}^S, \text{中国}^M\}$	$\{\text{日本}^S, \text{中国}^S\}$
$\{\text{愛知}^M\}$	×	×	×	×
$\{\text{北京}^M\}$	×	×	×	×
$\{\text{愛知}^M, \text{福岡}^M\}$	×	×	×	×
$\{\text{愛知}^M, \text{福岡}^S\}$	×	×	×	×
$\{\text{愛知}^M, \text{ソウル}^M\}$	×	×	×	×
$\{\text{愛知}^S, \text{ソウル}^M\}$	×	×	×	×
$\{\text{北京}^M, \text{ソウル}^M\}$	×	×	×	×
$\{\text{北京}^S, \text{ソウル}^M\}$	×	×	×	×
$\{\text{愛知}^M, \text{北京}^M\}$	○	○	○	○
$\{\text{愛知}^M, \text{北京}^S\}$	×	○	×	○
$\{\text{愛知}^S, \text{北京}^M\}$	×	×	○	○
$\{\text{愛知}^M, \text{北京}^M, \text{ソウル}^M\}$	×	×	×	×
$\{\text{愛知}^M, \text{北京}^S, \text{ソウル}^M\}$	×	×	×	×
$\{\text{愛知}^S, \text{北京}^M, \text{ソウル}^M\}$	×	×	×	×
$\{\text{愛知}^S, \text{北京}^S, \text{ソウル}^M\}$	×	×	×	×

図6 ランク付ラベル集合が順序 \preceq_{RB} で記述するオブジェクト

ランク付ラベル集合によって指定されるランク付集合ラベルのオブジェクト集合について、図7 (a) はランク付ラベル集合の順序に起因するオブジェクト集合の包含関係を、図7 (b) は指定に用いるランク付ラベル集合の要素のランクに起因するオブジェクトの包含関係を示している。



(a)



$$x = \{RN, RU, RL, RB\}$$

(b)

図7 ランク付ラベル集合によって指定されるオブジェクト集合の包含関係

図7の通り、ランク付ラベル集合の順序と指定に用いるランク付ラベル集合の要素におけるランクの強弱に起因するランク付集合ラベルのオブジェクト集合の包含関係が明らかった。

これにより、順序やランクを変更することで分析対象となるオブジェクト集合を拡張させたり限定させたりすることができるので、的確に分析対象となるオブジェクトを指定することが可能になる。

5 分析に必要な性質の検証

分類したランク付集合ラベルのオブジェクト集合を分析する手法として、分析対象を指定するランク付ラベル集合を変えることで分析対象となるオブジェクト集合の差を比較することが有用である。本節では、分析を行うために分析対象となるランク付集合ラベルのオブジェクト集合を指定する際に必要な性質として、妥当性（ランク付ラベル集合が異なれば指定されるランク付集合ラベルのオブジェクト集合も異なる）と健全性（ランク付ラベル集合の上下関係と指定されるランク付集合ラベルのオブジェクト集合における包含関係の一致）の2つについて検証する。

5.1 ランク付ラベル集合の順序の妥当性

分析対象となるランク付集合ラベルのオブジェクト集合の変化を利用した分析を行う際には、ランク付ラベル集合の要素のランクや要素自体を変えることで分析対象を変える操作が必要になる。すなわち、ランク付ラベル集合 L_1^R と L_2^R が異なれば L_1^R と L_2^R が指定するオブジェクトも異なることが求められる。

[定義4] ランク付ラベル集合 L_1^R と L_2^R ($L_1^R \neq L_2^R$) に対して、 $\overline{L_1^R}^x \neq \overline{L_2^R}^x$ ($x \in \{RN, RU, RL, RB\}$) ならばランク付ラベル集合の順序は妥当である。

ランク付ラベル集合は、4節の図3, 図4, 図5, 図6で明らかにしたように、4通りの分析対象の考え方、すなわち、4通りのランク付ラベル集合の順序すべてにおいて、ランク付ラベル集合の要素におけるランクを変えることで異なるランク付集合ラベルのオブジェクト集合が指定される。また、ランク付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ンク付ラベル集合に限定した場合、ランク付ラベル集合中のラベルが異なればそれらによって指定されるランク付集合ラベルのオブジェクトも異なるので妥当である。ランク付ラベル集合中のラベル $L_i^{R_i}$ と $L_j^{R_j}$ が $L_i^{R_i} \preceq L_j^{R_j}$ ($L_i^{R_i} \neq L_j^{R_j}$) ではない場合、そのようなランク付ラベル集合を排他であると呼ぶことにする。ランク付ラベル集合のラベルを排他であると限定すれば妥当性が成り立つ。しかし、排他ではないラベルを含むランク付ラベル集合は、ランク付ラベル集合が異なっても同一のランク付集合ラベルのオブジェクト集合を指定する場合があるので、以下のようにランク付ラベル集合を縮約して考えることが有用である。

排他ではないランク付ラベル集合 L^R について、 L^R 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^R の部分集合を L^R の下限とし, $l(L^R) = \{L_i^{R_i} \mid L_i^{R_i} \in L^R, \forall L_i^{R_i'} \in L^R \text{ s.t. } L_i^{R_i'} \neq L_i^{R_i}, L_i^{R_i'} \neq L_i^{R_i}\}$ で表す。

[命題1] ランク付ラベル集合 L^R に対して, $\overline{L^R}^{RU} = \overline{l(L^R)}^{RU}$ である。

(証明) ランク付ラベル集合 L^R に対して, $\overline{L^R}^{RU} = \overline{l(L^R)}^{RU}$ であることについての証明は, 文献⁹⁾による証明と同様である。 (証明終)

ランク付ラベル集合 $L_1^R = \{\text{東海}^M, \text{愛知}^M\}$, $L_2^R = \{\text{東海}^S, \text{愛知}^S\}$, $L_3^R = \{\text{東海}^S, \text{愛知}^M\}$ は, それぞれ $\overline{L_1^R}^{RU} = \overline{l(L_1^R)}^{RU}$, $\overline{L_2^R}^{RU} = \overline{l(L_2^R)}^{RU}$, $\overline{L_3^R}^{RU} = \overline{l(L_3^R)}^{RU}$ である。

一方, ランク付ラベル集合 $L_4^R = \{\text{東海}^M, \text{愛知}^S\}$ は, 2つの要素間には概念の上下関係はあるが, $\{\text{愛知}^S\} \not\leq_{RU} \{\text{東海}^M\}$ なので $l(L_4^R) \neq \{\text{愛知}^S\}$ であり, 縮約できない。例えば, $L(o) = \{\text{名古屋}^S, \text{福岡}^M\}$ のような, $\{\text{愛知}^S\}$ がランク付ラベル集合の順序 \leq_{RU} で記述するオブジェクト o は, $o \notin \overline{L_4^R}^{RU}$ である。

ランク付ラベル集合の下限と同様に, 排他ではないランク付ラベル集合 L^R 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^R の部分集合を L^R の上限とし, $u(L^R) = \{L_i^{R_i} \mid L_i^{R_i} \in L^R, \forall L_i^{R_i'} \in L^R \text{ s.t. } L_i^{R_i'} \neq L_i^{R_i}, L_i^{R_i} \neq L_i^{R_i'}\}$ で表す。

[命題2] ランク付ラベル集合 L^R に対して, $\overline{L^R}^{RL} = \overline{u(L^R)}^{RL}$, $\overline{L^R}^{RN} = \overline{u(L^R)}^{RN}$ である。

(証明) ランク付ラベル集合 L^R に対して, $\overline{L^R}^{RL} = \overline{u(L^R)}^{RL}$, $\overline{L^R}^{RN} = \overline{u(L^R)}^{RN}$ であることについての証明は, 文献⁹⁾による証明と同様である。 (証明終)

ランク付ラベル集合 $L_1^R = \{\text{東海}^M, \text{愛知}^M\}$, $L_2^R = \{\text{東海}^S, \text{愛知}^S\}$, $L_3^R = \{\text{東海}^S, \text{愛知}^M\}$ は, それぞれ $\overline{L_1^R}^{RL(RN)} = \overline{u(L_1^R)}^{RL(RN)}$, $\overline{L_2^R}^{RL(RN)} = \overline{u(L_2^R)}^{RL(RN)}$, $\overline{L_3^R}^{RL(RN)} = \overline{u(L_3^R)}^{RL(RN)}$ である。一方, ランク付ラベル集合 $L_4^R = \{\text{東海}^M, \text{愛知}^S\}$ は, 2つの要素間には概念の上下関係はあるが, $\{\text{愛知}^S\} \not\leq_{RL(RN)} \{\text{東海}^M\}$ なので $u(L_4^R) \neq \{\text{東海}^M\}$ であり, 縮約できない。例えば, $L(o) = \{\text{津}^M, \text{名古屋}^S\}$ のような, $\{\text{東海}^M\}$ がランク付ラベル集合の順序 \leq_{RL} で記述しないランク付集合ラベルのオブジェクト o は, $o \in \overline{L_4^R}^{RL}$ である。

ランク付ラベル集合の順序 \leq_{RB} に対して, $ul(L^R)$ を $u(L^R) \cup l(L^R)$ とすると, L^R は $ul(L^R)$ に縮約できる。

[命題3] ランク付ラベル集合 L^R に対して, $\overline{L^R}^{RB} = \overline{ul(L^R)}^{RB}$ である。

(証明) ランク付ラベル集合 L^R に対して、 $\overline{L}^{R^{RB}} = \overline{ul(L^R)^{RB}}$ であることについての証明は、文献⁹⁾による証明と同様である。 (証明終)

5.2 ランク付ラベル集合の順序の健全性

ランク付ラベル集合 L_1^R と L_2^R に対して、 $\overline{L}_1^{R^x} - \overline{L}_2^{R^x} \neq \phi$ かつ $\overline{L}_2^{R^x} - \overline{L}_1^{R^x} \neq \phi$ ($x \in \{RN, RU, RL, RB\}$)のとき、これらの差なども分析対象となりうるが、 \overline{L}_1^R と \overline{L}_2^R 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、 $\overline{L}_1^{R^x}$ と $\overline{L}_2^{R^x}$ を直接比較できない。 $\overline{L}_1^{R^x} \subseteq \overline{L}_2^{R^x}$ であれば、 L_1^R と L_2^R の違いは $\overline{L}_2^{R^x} - \overline{L}_1^{R^x}$ によって分析できる⁹⁾。

[定義5] ランク付ラベル集合 L_1^R と L_2^R に対して、 $L_1^R \preceq_x L_2^R$ ($x \in \{RN, RU, RL, RB\}$)であることと $\overline{L}_1^{R^x} \subseteq \overline{L}_2^{R^x}$ であることが等価であるならば、ランク付ラベル集合の順序 \preceq_x は健全である。

ランク付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] ランク付ラベル集合の順序が推移律を満たすことは、ランク付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) ラベル集合の順序が推移律を満たすことは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ことが文献⁹⁾で証明されているため、ランク付ラベル集合についても同様に成り立つ。 (証明終)

ランク付ラベル集合の順序 \preceq_{RN} は健全ではない。ランク付ラベル集合 L_1^R と L_2^R をそれぞれ $L_1^R = \{\text{東海}^M, \text{福岡}^S\}$ と $L_2^R = \{\text{九州}^S\}$ とする。ランク付集合ラベルが $L(o) = \{\text{愛知}^M\}$ であるオブジェクト o について、 $L(o)$ における“愛知^M”は L_1^R における“東海^M”以下なので、 o は \overline{L}_1^{RN} の要素であるが、 $L(o)$ における“愛知^M”は L_2^R における“九州^S”以下ではないので、 o は \overline{L}_2^{RN} の要素ではない。すなわち、 $L(o) \preceq_{RN} L_1^R$ 、 $L_1^R \preceq_{RN} L_2^R$ であるが $L(o) \not\preceq_{RN} L_2^R$ なので、 \preceq_{RN} は推移律を満たさない、よって \preceq_{RN} は健全ではない。それに対して、ランク付ラベル集合の順序 \preceq_{RU} 、 \preceq_{RL} 、 \preceq_{RB} は推移律を満たし、健全である。

[補題3] ランク付ラベル集合の順序 \preceq_{RU} 、 \preceq_{RL} 、 \preceq_{RB} は推移律を満たす。

(証明) ランク付ラベル集合の順序 \preceq_{RU} 、 \preceq_{RL} 、 \preceq_{RB} は推移律を満たすことは文献⁹⁾による証明と同様である。 (証明終)

[定理2] ランク付ラベル集合の順序 \preceq_x ($x \in \{RU, RL, RB\}$)は健全である。

(証明) 補題2, 3よりランク付ラベル集合の順序 \preceq_{RU} , \preceq_{RL} , \preceq_{RB} は健全である。(証明終)

6 おわりに

本稿では、オブジェクトと分類階層におけるカテゴリの関連性の強さを示すランクを導入し、ランク付集合ラベルのオブジェクトを分析に供する際に必要な分析対象の考え方を議論した。そして、分析対象の考え方に対応するランク付集合ラベルのオブジェクト集合の指定法をランク付ラベル集合間の順序を用いた記述によって体系化した。また、分析対象の考え方とランクの強弱に起因するランク付集合ラベルのオブジェクト集合の包含関係についても明らかにした。これにより、分析対象となるオブジェクト集合を拡張させたり限定させたりすることができるので、的確に分析対象となるオブジェクトを指定することができる。

本稿は、分析するために必要な性質についても順序の妥当性と健全性を議論することで検証した。ランク付ラベル集合の順序の妥当性を考えた場合、順序を変えたり指定に用いるラベル集合の要素のランクや要素自体を変えれば指定されるランク付集合ラベルのオブジェクトも異なるため、分析対象数の差をみるといった分析が可能である。ただし、ラベル間に上下関係があるようなランク付ラベル集合では、異なるラベル集合であっても同一の分析対象となり比較ができないため、ラベル集合を縮約して指定することが有用である。また、ランク付ラベル集合の順序 \preceq_{RU} , \preceq_{RL} , \preceq_{RB} は健全性を有するため、例えば、ランク付ラベル集合の上下関係に応じて変化する分析対象数の差を利用した分析が可能になる。本稿により、ランク付集合ラベルのオブジェクトに対するラベル集合による指定とその性質が体系化され、オブジェクトとカテゴリの関連性の強弱を加味した精緻な分析に必要な理論的枠組みが構築された。

本稿による研究成果は、ランク付集合ラベルのオブジェクトに対する分析手法に応用可能である。例えば、日本と中国の強弱関係（どちらの国の方が注目をされているか等）を分析する際に、 $L_1^R = \{\text{日本}^M, \text{中国}^S\}$ と $L_2^R = \{\text{日本}^S, \text{中国}^M\}$ に対して、順序 \preceq_{RB} で指定されるオブジェクト集合数の多寡を比較する場合を考える。この際、 $L_3^R = \{\text{日本}^M, \text{中国}^M\}$ として、 $\overline{L_1^R}^{RB} - \overline{L_3^R}^{RB}$ と $\overline{L_2^R}^{RB} - \overline{L_3^R}^{RB}$ の共通集合は空集合なので、それらを直接比較してオブジェクト集合の数が多いラベル側の方が強い関係にあることが分かる。仮に、 $\overline{L_1^R}^{RB} - \overline{L_3^R}^{RB}$ におけるオブジェクト集合の数が多ければ、日本を主として中国を副としているオブジェクトが中国を主として日本を副としているオブジェクトよりも多いので、日本の方が中国よりも注目度が高いという知見が得られる。今後は、本稿で明らかにしたランク付ラベル集合による指定と性質を利用して、ランク付集合ラベルのオブジェクトに対してどのような分析が可能かについてを詳細に検討したい。

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) Aho, A., Hopcroft, J., and Ullman, J.: *Data Structure and Algorithms*, Addison-Wesley Publishing Company (1987).
- 3) 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).
- 4) 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).
- 5) Elmasri, R. and Navathe, S.: *Fundamentals of Database Systems*, Addison-Wesley Publishing Company, Second Edition (1994).
- 6) 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).
- 7) Furukawa, T. and Kuzunishi, M.: Classification and Utilization of Data Belonging to Multiple Classes, *Proc. The 8th World Multiconference on Systemics, Cybernetics and Informatics*, Vol. 2, pp. 289-294 (2004).
- 8) 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).
- 9) 古川哲也, 葛西正裕: 集合ラベルを持つデータの集約範囲の記述, 情報処理学会論文誌: データベース, 情報処理学会, Vol.3, No.3, pp. 11-19 (2010).
- 10) Ghamrawi, N. and McMallum, A.: Collective Multi-Label Classification, *Proc. Int'l Conf. on Information and Knowledge Management (CIKM'05)*, pp. 195-200 (2005).
- 11) Kim, J. and Croft, B.: Ranking Using Multiple Document Types in Desktop Search, *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval (SIGIR'10)*, pp. 50-57 (2010).
- 12) 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, 情報処理学会九州支部, A-3-2 (2010).
- 13) Kuzunishi, M., Furukawa, T., and Lu, K.: Analyzing Multi-Labeled Data Based on the Roll of a Concept against a Semantic Range, *Proc. World Academy of Science, Engineering and Technology*, Vol. 69, pp. 498-504 (2010).
- 14) 葛西正裕: 多重属性を持つラベル集合を用いたデータの記述, 地域分析, 愛知学院大学産業研究所, 第49巻, 第1号, pp. 43-65 (2010).
- 15) Mikroyannidis, A. and Theodoulidis, B.: Ontology Management and Evolution for Business Intelligence, *International Journal of Information Management*, Vol. 30, pp. 559-566 (2010).
- 16) 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).
- 17) Veloso, A., Almeida, H., Goncalves, M., and Meira, W.: Learning to Rank at Query-Time Using Association Reles, *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval (SIGIR'08)*, pp. 267-274 (2008).
- 18) Wang, C., Zhang, M., Ru, L., and Ma, S.: Automatic Online News Topic Ranking Using Media Focus and User Attention Based on Aging Theory, *Proc. Int'l Conf. on Information and Knowledge Management (CIKM'08)*, pp. 1033-1042 (2008).
- 19) 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).

