

ラベル集合の順序を用いた ランク付集合ラベルのデータの記述

葛西 正裕

- 1 はじめに
 - 2 ランクの導入と性質
 - 3 ラベル集合の順序とランクに応じたデータの記述
 - 4 順序の組合せによるデータの記述
 - 5 ランクの組合せによるデータの記述
 - 6 分析に有用な記述
 - 7 おわりに
- 参考文献

【要旨】

経済分析において、情報通信技術の発展により収集できるようになった多様なデータを利用することは有用である。データを階層的に分類しておくことで一元的に分析に供することが可能になるが、データの多様性は分類時にデータの不均一性となって現れる。すなわち、データには複数のカテゴリのラベル（集合ラベル）が付され、各カテゴリとデータの関連性の強さが異なる。関連の強さを示すランクをラベルに対して導入することで、分析の対象となるデータを関連性の強弱を反映させて指定できるようになるが、どのようにデータを記述すればよいか問題となる。そこで、本稿は、指定に用いるラベル集合とデータのランク付集合ラベルの関係に対してラベル集合の順序を適用し、考え得るデータの記述を導出したのち分析に有用な記述に集約することで、順序によるデータの記述に関する理論的な基盤を構築する。本稿の議論によって、順序とランクに応じて考え得るすべての組合せから同一のデータを記述するものを除いて43通りの記述を導出した。そして、43通りの記述は、データ分析時におけるラベル集合内外との関連の種類や程度を考慮した28通りの指定に対応することが判明した。同時に、残りの15通りの記述は単独では有用な意味をなさないことを示した。

【キーワード】 ラベル集合の順序、ラベルのランク、データの記述、データベースモデル、経済分析手法

1 はじめに

今日における情報通信技術の進展によって、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が大量に収集され分析に利用されている^{[2] [18] [19]}。多様かつ膨大なデータは、一般にビッグデータと呼ばれており、ビッグデータを対象とした分析手法が提案されている^{[7] [11]}。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することで従来とは異なる分析が可能になり、その分析手法の確立が求められる。

多様なデータに対してそれらの種類を問わず一元的に分析に供するためには、データを階層的に分類しておくことが有用である^{[13] [17]}。データの分類は業種や地域といった属性ごとに行われ、データには分類階層で該当するカテゴリのラベルが付される^[1]。

データの多様性は分類時にデータの不均一性となって現れる。均一なデータの分類であれば、1件のデータは1つのカテゴリに分類されることで単一ラベルが付されて、カテゴリとの関連の強さも同程度になる。一方、多様なデータでは複数のカテゴリのラベル（集合ラベル）が付され^{[4] [8] [14]}、かつ各カテゴリとデータの関連性の強さも異なることが多い。

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、経済データを業種を考慮して分析する際、主として自動車に関するデータであるが電気機器についても副次的に事業展開をしているようなデータに対して、単純に自動車と電気機器という2つのラベルを付すと、主として電気機器に関するデータとの区別ができない。一方、企業の財務情報などを参考にして、主として自動車に関する業種であるが電気機器についても関連した企業のデータと自動車と電気機器を同程度に展開する企業のデータを区別しておくことで、例えば両者の売上高や営業利益の差を比較するといった分析が可能になる。

カテゴリに対する関連性の強弱を反映させた分析を行うためには、関連の強さを示すランクをラベルに対して導入する必要がある。ランクを与えた集合ラベル（ランク付集合ラベル）をデータに付すことにより、分析の対象となるデータを関連性の強弱を反映させて指定することが可能になるが、どのようにデータを記述し分析に供すればよいか問題となる。例えば、自動車と電気機器に関連するデータを考えたとき、両業種に関連しているが自動車のみを主としているデータを分析の対象に含めるかどうかや、両業種以外に主たる関連があるが自動車にも副次的に関連しているデータを含めるかどうかといったように分析する際の観点は複数ある。すなわち、ランクを考慮しつつラベル集合に応じたデータの記述体

系が必要になる。

ランクを有しない集合ラベルのデータに対しては、論文 [5] でラベル集合の順序を用いることで分析に供することができることを明らかにしている。そこで、本稿は、ランク付集合ラベルのデータに対してラベル集合の順序を用いることによって、分析に有用なデータの記述を明らかにする。

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されている。それらは、検索結果の順位付けを目的としたものが中心であってランク付けの正確性や効率性を評価するものが多い^{[6] [15] [16]}。よって、関連研究の多くは分析という視点からの検討が行われていないため、関連研究の応用ではランクが複数であることや関連性の強弱を用いた精緻な分析が行えない。一方、論文 [10] は、ランク付ラベル集合を与えて分析の対象となるデータの指定と性質を明らかにし、研究 [12] において、ランクの強弱を意識したデータの指定法をより精緻に議論した。しかし、両者は指定されるデータを記述していく過程で結果的にラベル集合の順序を導出しているのに留まっており、ラベル集合の順序を用いたデータの記述について十分な議論を尽くしているとは言えない。それに対して、本稿は、指定に用いるラベル集合とデータの集合ラベルの関係において、当初よりラベル集合の順序を用いことでデータを記述する。その上で、順序とランクに応じて考え得るすべての組合せによるデータの記述を導出して、データの分析に有用かどうかという観点からデータの記述を整理する。本稿における網羅的な議論により、順序によるランク付集合ラベルのデータの記述に関する理論的な基盤を構築する。

本稿は以下のように構成される。2章は、データの集合ラベルにランクを導入し、ランクを有するデータの性質について述べる。3章は、データの記述にラベル集合の順序を用いて、ランクに応じたデータの記述を検討する。4章は、ランクを固定した上で、異なる順序で記述されるデータの組合せを検討し、新たな記述があるかどうかについて検討する。5章は、順序に加えてランクの組合せについても検証し、新たな記述があるかどうかについて検証する。6章は、これまでに導出した記述を分析に有用な記述に集約して整理する。7章はまとめである。

2 ランクの導入と性質

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、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。自動車産業における東海地域の現状といった業種と地域という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^[3]。

複数の属性に分類されるデータの分析に関しては、論文 [9] 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。同時に、分類されるデータの概念は分類階層の最下層のカテゴリに対応するようなデータとし、データは分類階層の最下層のカテゴリに分類されるものとする。

1件のデータを d 、データの分類後に付されるラベルを L 、ラベル集合を $\mathbf{L} = \{L_1, \dots, L_n\}$ とする。 d は予め与えられた分類階層において d 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。例えば、業種という属性でデータを分類したとき、自動車と電気機器という2つの事業を展開している企業の財務データあれば、 $\mathbf{L} = \{\text{自動車}, \text{電気機器}\}$ というラベル集合が付される。すなわち、 d 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 d に付される。ゆえに、 d のラベル集合は d の属性に対する意味を表している。また、 d に付されたラベル集合を $L(d)$ で表し、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、 $L_1 = \text{自動車}$ と $L_2 = \text{輸送機器}$ 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。

データに複数のカテゴリのラベルを付す際、データ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、データ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、本稿では、議論を簡単にするためにランクの区分を強弱の2つとし、主 (Primary) と副 (Secondary) に設定する。ラベル L がカテゴリに対して主の性質を持つ場合には主ラベルと呼び L^P で表し、 L がカテゴリに対して副の性質を持つ場合には副ラベルと呼び L^S で表す。

ランクが付いたラベル集合を $\mathbf{L} = \{L_1^{R_1}, \dots, L_n^{R_n}\}$ ($R_1, \dots, R_n \in \{P, S\}$) とし、データ d の集合ラベル $L(d)$ はランク付集合ラベルとする。 $L(d)$ に対して、 $L(d)$ の主ラベルで構成されるランク付集合ラベルを $P(d) = \{L_i^{R_i} | L_i^{R_i} \in L(d), R_i = P\}$ で表し、 $L(d)$ の副ラベルで構成されるランク付集合ラベルを $S(d) = \{L_i^{R_i} | L_i^{R_i} \in L(d), R_i = S\}$ で表す。

[例1] 企業財務に関するデータセットに対して、企業ごとに事業展開する業種で分類を行う場合を考える。1件の企業のデータ d_1 が、売上高や営業利益といった財務情報をもとに自動車と電気機器を同程度に展開していると判断された場合

には両業種の主ラベルが付される。よって、 d_1 のランク付集合ラベルは $L(d_1) = \{\text{自動車}^P, \text{電気機器}^P\}$ となる。一方、自動車に関する事業の売上高や営業利益が電気機器と比較して十分に高いような企業のデータ d_2 であれば、 d_2 には“自動車”の主ラベルと“電気機器”の副ラベルが付されるので、 $L(d_2) = \{\text{自動車}^P, \text{電気機器}^S\}$ となる。

データ d に対して分類が行われ、 d にランク付ラベル集合が付された場合、 d 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] データ d に対して、 $L(d) = P(d) \cup S(d)$ である。

データ d の $L(d)$ のラベルに対してランクがついているので性質 1 が成り立つ。

[性質 2] データ d に対して、 $P(d) \cap S(d) = \phi$ である。

データ d の $L(d)$ のラベルとカテゴリとの関連の強さは、主副のいずれかになるので性質 2 が成り立つ。

[性質 3] データ d に対して、 $P(d) \neq \phi$ である。

データ d は特定の分類階層において分類されているということは、 d はいずれかのカテゴリの意味を主として有していることになり性質 3 が成り立つ。

[例 2] 企業財務に関するデータセットの分類を想定し、分類が完了しているものとする。データセット中の各企業のデータのラベルは主ラベルもしくは副ラベルなので、データのランク付集合ラベルは主ラベルの集合と副ラベルの集合の和となるため性質 1 が成り立つ。1 件の企業が自動車に主として関連していれば、同時に副次的にも関連するといったことは考えられないので、同一概念のラベルで主ラベルと副ラベルがともに存在することはない。よって、各企業が事業展開する業種のラベルは財務情報をもとに主ラベルもしくは副ラベルのいずれかになるので性質 2 が成り立つ。最後に、ある企業が業種という分類階層において分類できたということは、いずれかのカテゴリに主として関連することを意味しているため、ランク付集合ラベルに主ラベルが存在しないことはあり得ないので性質 3 が成り立つ。

3 ラベル集合の順序とランクに応じたデータの記述

本章は、最初にラベル集合の順序を導入した上で、データの記述に順序を適用し、ランクに応じたデータの記述を検討する。

分析に供するデータを指定する際、一般にはラベル集合を与えて、それに対応するデータを考える。ランクを有しない集合ラベルのデータに対しては、論文 [5]

により、指定されるデータはラベル集合の順序を用いて記述できることを明らかにしている。さらに、論文 [10] 及び研究 [12] では、ランク付集合ラベルのデータを対象として、ラベル集合の順序を用いた記述を議論している。これまでの研究によるラベル集合の順序は、定義 1 の通りである。

[定義 1] ラベル集合 L_1 と L_2 に対して、

$$L_1 \preceq_{RN} L_2 \iff \exists L_2 \in L_2, \exists L_1 \in L_1, L_1 \preceq L_2,$$

$$L_1 \preceq_{RU} L_2 \iff \forall L_2 \in L_2, \exists L_1 \in L_1, L_1 \preceq L_2,$$

$$L_1 \preceq_{RL} L_2 \iff \forall L_1 \in L_1, \exists L_2 \in L_2, L_1 \preceq L_2,$$

$$L_1 \preceq_{RB} L_2 \iff \forall L_2 \in L_2, \exists L_1 \in L_1, L_1 \preceq L_2$$

$$\wedge \forall L_1 \in L_1, \exists L_2 \in L_2, L_1 \preceq L_2$$

とし、順序 \preceq_{RN} 、 \preceq_{RU} 、 \preceq_{RL} 、 \preceq_{RB} をそれぞれ順序 RN 、 RU 、 RL 、 RB と表記する。□

ラベル集合 L_1 と L_2 に対して、 $L_1 \preceq_X L_2$ ($X \in \{RN, RU, RL, RB\}$) とする。順序 RN は、 L_2 のいずれかのラベルに対して L_1 のいずれかのラベルで順序があればラベル集合間に順序があると考ええる。次に、順序 RU は、 L_2 のすべてのラベルに対して L_1 のいずれかのラベルで順序があればラベル集合間に順序があると考ええるものであり、上位となるラベル集合に対してすべてのラベルで順序があるという制限を設けている。反対に、順序 RL は、 L_1 のすべてのラベルに対して L_2 のいずれかのラベルで順序があればラベル集合間に順序があると考ええるものであり、下位となるラベル集合に対してすべてのラベルで順序があるという制限を設けている。最後に、順序 RB は、順序 RU と RL の両方の考え方を組合せたものであり、上位と下位の両方のラベル集合に対してすべてのラベルで順序があるという制限を設けている。図 1 は、定義 1 を図で表したものである。

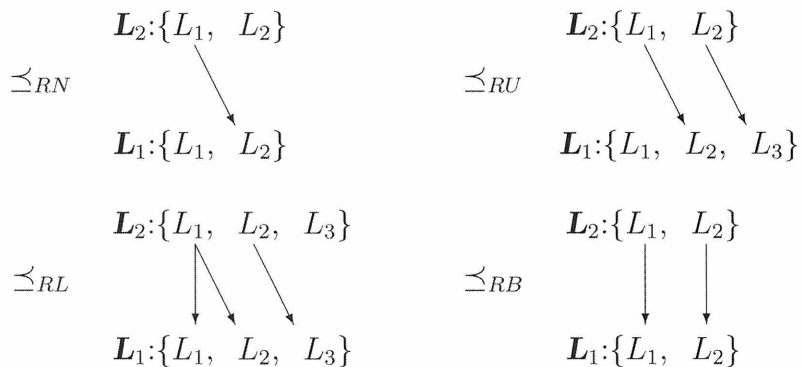


図 1 ラベル集合の順序

ラベル集合 L とラベル集合の順序によって記述されるデータ d に対して、まずは $L(d)$ のランクを考慮せずに L と $L(d)$ の関係を考える。 L と関連があるデータを指定する場合、対象となるデータが最も広範囲になるものとして、 L のいずれかのラベルに対して関連があるデータを指定する考え方がある。この場合、 L のいずれかのラベルに対して $L(d)$ のいずれかのラベルで関連があるという関係が L と $L(d)$ に成り立つデータが指定されればよいことになる。よって、指定されるデータの記述には順序 RN が対応する。一方、 L のすべてのラベルに対して関連があるデータを指定したい場合、 L のすべてのラベルに対して $L(d)$ のいずれかのラベルで関連があるという関係が L と $L(d)$ に成り立てばよいので、順序 RU が対応する。次に、ラベル集合 L 以外のラベルとの関連について検討すると、 L 以外のラベルに関連がないデータを指定したい場合、 $L(d)$ に L と関連のないラベルが含まれてはならない。よって、 $L(d)$ のすべてのラベルに対して L のいずれかのラベルで関連があるという関係が L と $L(d)$ に成り立てばよいので、順序 RL が対応する。最後に、 $L(d)$ に L と関連のないラベルが含まれず、かつ L のすべてのラベルに対して関連があるデータを指定したい場合は、順序 RL と RU の両方が成り立つ必要があるので、順序 RB が対応する。

これまでデータ d の $L(d)$ のランクを考慮せずに議論してきたがランクを考慮して議論していく。ラベル集合 L に対して、強い関連であれば $P(d)$ と L 、弱い関連であれば $S(d)$ と L 、強弱を意識しない関連であれば $L(d)$ と L の関係にラベル集合の順序を用いることで、関連性の種類に応じたデータの記述ができる、すなわち、関連の程度を考慮する指定や考慮しない指定が行える。

これより、データ d の $L(d)$ 、 $P(d)$ 、 $S(d)$ に順序 RN 、 RL 、 RU 、 RB を適用する。

[定義 2] ラベル集合 L とデータ d に対して、 $L(d)$ 、 $P(d)$ 、 $S(d)$ に順序 X ($X \in \{RN, RU, RL, RB\}$) を適用したときの L に対するデータを、 $\bar{L}^{X/L(d)} = \{d | L(d) \preceq_X L\}$ 、 $\bar{L}^{X/P(d)} = \{d | P(d) \preceq_X L\}$ 、 $\bar{L}^{X/S(d)} = \{d | S(d) \preceq_X L\}$ とする。また、 L と d が明確なときは、 $X/L (= \bar{L}^{X/L(d)})$ 、 $X/P (= \bar{L}^{X/P(d)})$ 、 $X/S (= \bar{L}^{X/S(d)})$ と表記する。□

定義 2 について、例えば、データ d の $L(d)$ に順序 RN を適用したときのラベル集合 L に対するデータは RN/L となる。

データ d に対して順序を適用する。ラベル集合 L に対する X/Y ($X \in \{RN, RU, RL, RB\}$ 、 $Y \in \{L, P, S\}$) で記述されるデータは、4 通りのラベル集合の順序があって、さらに順序を適用する対象が $L(d)$ 全体、主ラベルの集合 ($P(d)$)、

副ラベルの集合 $(S(d))$ の区分があるので、12通りで記述されるデータはそれぞれ異なる。

4 順序の組合せによるデータの記述

本章は、3章の4通りの順序で記述されるデータに対して、ランクは固定した上で、異なる順序で記述されるデータの組合せを検討する。本章では順序を適用するデータのランクを固定して考えるが、5章で異なるランクに拡張する。

ラベル集合 L とデータ d における、 L と $L(d)$ 、 L と $P(d)$ 、 L と $S(d)$ に対して、異なる2種類の順序で記述されるデータの組合せを $X/Z \cap Y/Z = \{d | d \in X/Z, d \in Y/Z\}$ ($X, Y \in \{RN, RU, RL, RB\}, Z \in \{L, P, S\}$) とする。なお、対称律が成り立ち、 $X/Z \cap Y/Z = Y/Z \cap X/Z$ である。

まず、ラベル集合 L とデータ d の $L(d)$ に対して、ラベル集合の順序を適用して記述されるデータの組合せを考える。4通りの順序から2つ順序を用いて記述される L に対するデータの組合せは6通り (${}_4C_2$) である。これらで記述されるデータについて以下の補題が成り立つ。

[補題1] ラベル集合 L とデータ d に対して、

$$RN/L \cap RL/L = RL/L \quad (1)$$

$$RN/L \cap RU/L = RU/L \quad (2)$$

$$RN/L \cap RB/L = RB/L \quad (3)$$

$$RL/L \cap RU/L = RB/L \quad (4)$$

$$RL/L \cap RB/L = RB/L \quad (5)$$

$$RU/L \cap RB/L = RB/L \quad (6)$$

である。□

(証明) (1) について、定義1のラベル集合の順序の定義より、順序 RL の方が順序 RN よりも厳しい条件となる。すなわち、順序 RL は順序 RN を含意しているので、 $RL/L \subseteq RN/L$ であり $RN/L \cap RL/L = RL/L$ である。同様に、(2) についても、順序 RU は順序 RN を含意しているので $RN/L \cap RU/L = RU/L$ である。また、(3) は、順序 RB は順序 RL と RU の両方の条件を満たす、すなわち順序 RB は順序 RL や RU を含意しているため順序 RN も含意する。よって、 $RN/L \cap RB/L = RB/L$ である。(4) は、順序 RB の定義より明らかである。また、(5) や (6) についても、順序 RB は順序 RL や RU を含意していることから明らかである。 (証明終)

補題 1 より、ラベル集合 L とデータ d の $L(d)$ に対して、ラベル集合の順序を適用して記述されるデータの組合せを考えても新しい意味を持つ記述はない。同様に、 $P(d)$ について検討する。

[補題 2] ラベル集合 L とデータ d に対して、

$$RN/P \cap RL/P = RL/P \quad (7)$$

$$RN/P \cap RU/P = RU/P \quad (8)$$

$$RN/P \cap RB/P = RB/P \quad (9)$$

$$RL/P \cap RU/P = RB/P \quad (10)$$

$$RL/P \cap RB/P = RB/P \quad (11)$$

$$RU/P \cap RB/P = RB/P \quad (12)$$

である。 □

(証明) (7) について、性質 3 より $P(d) \neq \phi$ なので、 RL/P に含まれる d の $P(d)$ のラベルは L のラベルと順序を持つ。よって、補題 1 の証明と同様に順序 RL は順序 RN を含意し、 $RL/P \subseteq RN/P$ であり $RN/P \cap RL/P = RL/P$ である。(8) から (12) についても性質 3 及び補題 1 の証明より明らかである。

(証明終)

補題 2 より、ラベル集合 L とデータ d の $P(d)$ に対して、ラベル集合の順序を適用して記述されるデータの組合せを考えても新しい意味を持つ記述はない。最後に、 $S(d)$ について検討する。なお、新たな記述がある場合には先頭に * を付して明記する。

[補題 3] ラベル集合 L とデータ d に対して、

$$* RN/S \cap RL/S \quad (13)$$

$$RN/S \cap RU/S = RU/S \quad (14)$$

$$RN/S \cap RB/S = RB/S \quad (15)$$

$$RL/S \cap RU/S = RB/S \quad (16)$$

$$RL/S \cap RB/S = RB/S \quad (17)$$

$$RU/S \cap RB/S = RB/S \quad (18)$$

である。 □

(証明) (13) について、定義 1 のラベル集合の順序より、 $d \in RL/S$ となるような d のうち、 $P(d)$ のラベルが L のラベルと順序がなくかつ $S(d) = \phi$ の

とき、 $d \notin RN/S$ である。ゆえに、 RN/S と RL/S は包含関係になく、 $RN/S \cap RL/S$ はこれまでの記述に該当するものはない。

(14) については、 RU/S に含まれる d の $S(d)$ のすべてのラベルが L のラベルと順序を持つので $d \in RN/S$ である。よって、 $RU/S \subseteq RN/S$ なので $RN/S \cap RU/S = RU/S$ である。(15) から (18) については、順序 RB は順序 RN 、 RU 、 RL をそれぞれ含意することから明らかである。 (証明終)

補題3より、ラベル集合 L とデータ d の $S(d)$ に対して、ラベル集合の順序を適用して記述されるデータの組合せを考えたとき、 $RN/S \cap RL/S$ のみ新たな意味を持つ。 $RN/S \cap RL/S$ が記述するデータについて、 RN/S や RL/S を単独で用いた場合と比較をして検討する。まず、 RN/S は L のラベルに対して $S(d)$ のラベルが順序を持つデータを記述しており、 L のいずれかと副次的に関連をしているデータを指定する場合に用いる。一方、 RL/S は $S(d)$ のすべてのラベルが L のラベルと順序があるデータを記述しており、言い換えれば、 $S(d)$ には L と順序がないラベルが存在しない。よって、 L に対して副次的に関連しない意味を持たないデータを指定する場合に用いる。ただし、 $S(d) = \phi$ の場合には L と全く関連がないデータが該当することになる。例えば、 $L = \{\text{自動車, 電気機器}\}$ に対して、 $L(d) = \{\text{農業}^P\}$ といったデータも該当する。それに対して、 $RN/S \cap RL/S$ は、 L に関連しつつも L に対して副次的に関連しない意味を持たないデータを指定できるので、分析を考えた場合に有用な記述といえる。

5 ランクの組合せによるデータの記述

4章では、データのランクを固定して議論してきたが、本章は、ランクの組合せについても検証し、順序とランクに応じて考え得るすべての組合せによるデータの記述を明らかにする。

ラベル集合 L とデータ d に対して、ラベル集合の順序を適用する際に $L(d)$ 、 $P(d)$ 、 $S(d)$ の3通りがある。よって、まず3通りのうち2通りの組合せを議論し、最後に3通りすべての組合せを考える。なお、新たな記述がある場合には先頭に*を付して明記し、データの記述の種類を区別しやすいように(a)から(z)までの文字を表記する。

ラベル集合 L とデータ d に対して、 $L(d)$ と $P(d)$ の組合せを検討した結果は以下の補題の通りである。

[補題 4] ラベル集合 L とデータ d に対して、

$$RN/L \cap RN/P = RN/P \quad (19)$$

$$RN/L \cap RL/P = RL/P \quad (20)$$

$$RN/L \cap RU/P = RU/P \quad (21)$$

$$RN/L \cap RB/P = RB/P \quad (22)$$

$$RL/L \cap RN/P = RL/L \quad (23)$$

$$RL/L \cap RL/P = RL/L \quad (24)$$

$$* (a) : RL/L \cap RU/P \quad (25)$$

$$(a) : RL/L \cap RB/P = RL/L \cap RU/P \quad (26)$$

$$* (b) : RU/L \cap RN/P \quad (27)$$

$$* (c) : RU/L \cap RL/P \quad (28)$$

$$RU/L \cap RU/P = RU/P \quad (29)$$

$$RU/L \cap RB/P = RB/P \quad (30)$$

$$RB/L \cap RN/P = RB/L \quad (31)$$

$$RB/L \cap RL/P = RB/L \quad (32)$$

$$(a) : RB/L \cap RU/P = RL/L \cap RU/P \quad (33)$$

$$(a) : RB/L \cap RB/P = RL/L \cap RU/P \quad (34)$$

である。

□

(証明) (19) から (22) について、 $P(d) \subseteq L(d)$ かつ順序 RL 、 RU 、 RB は順序 RN を含意することから明らかである。(23) について、性質 3 より $P(d) \neq \phi$ なので、 RL/L に含まれる d の $P(d)$ のラベルは L のラベルと順序がある。よって、 d は RN/P に含まれるので $RL/L \cap RN/P = RL/L$ である。また、(24) では、 RL/L に含まれる d は L のラベルと順序がないラベルを $L(d)$ に持たない、かつ性質 1 から $P(d) \subseteq L(d)$ なので、 $RL/L \subseteq RL/P$ である。ゆえに、 $RL/L \cap RL/P = RL/L$ である。

一方、(25) は順序 RL と RU はお互いに含意しない上、順序を適用する対象も $L(d)$ と $P(d)$ で異なるため、 RL/L と RU/P に包含関係はなく、これまでの記述に該当するものはない。(26) は $RL/L \cap RB/P = RL/L \cap (RU/P \cap RL/P)$ であり、(24) の証明から $RL/L \cap RL/P = RL/L$ なので、 $RL/L \cap RU/P$ となり (a) に該当する。

(27) について、 RU/L に含まれる d のうち $P(d)$ のすべてのラベルが L のラベルと順序がないとき、 d は RN/P に含まれない。よって、 RU/L と RN/P は包含関係がなく、これまでの記述に該当するものはない。(28) は、(25) と同様に順序 RU と RL に含意しない上、順序を適用するのも $L(d)$ と $P(d)$ で異なっ

ており、これまでの記述に該当するものはない。

(29) は、性質1から $P(d) \subseteq L(d)$ なので、 $RU/P \subseteq RU/L$ となり $RU/L \cap RU/P = RU/P$ である。(30) は、 $RU/L \cap RB/P = RU/L \cap (RU/P \cap RL/P)$ であり、(29) の証明から $RU/L \cap RU/P = RU/P$ なので、 $RU/P \cap RL/P = RB/P$ である。(31) について、 RB/L に含まれる d は、 $P(d) \neq \phi$ なので RN/P にも含まれるため $RB/L \cap RN/P = RB/L$ である。(32) について、 RB/L に含まれる d は、 \mathbf{L} のラベルと順序が無いラベルを $L(d)$ に含まないので $P(d)$ にも含まない。よって、 d は RL/P に含まれるので $RB/L \cap RL/P = RB/L$ である。

(33) は、 $RB/L \cap RU/P = (RL/L \cap RU/L) \cap RU/P$ であり、(29) の証明より $RU/P \cap RU/L = RU/P$ なので $RL/L \cap RU/P$ となり (a) に該当する。(34) は、 $RB/L \cap RB/P = (RL/L \cap RU/L) \cap (RL/P \cap RU/P)$ であり、(24) と (29) の証明より $RL/L \cap RL/P = RL/L$ かつ $RU/L \cap RU/P = RU/P$ なので、 $RL/L \cap RU/P$ となり (a) に該当する。 (証明終)

ラベル集合 \mathbf{L} とデータ d に対して、 $L(d)$ と $S(d)$ の組合せについて検討する。なお、補題3より $RN/S \cap RL/S$ が RN/S や RL/S とは異なるデータを記述することから $RN/S \cap RL/S$ ($RL/S \cap RN/S$) も組合せに加えて議論する。

[補題5] ラベル集合 \mathbf{L} とデータ d に対して、

$$RN/L \cap RN/S = RN/S \quad (35)$$

$$*(d) : RN/L \cap RL/S \quad (36)$$

$$(d) : RN/L \cap (RL/S \cap RN/S) = (RL/S \cap RN/S) \quad (37)$$

$$RN/L \cap RU/S = RU/S \quad (38)$$

$$RN/L \cap RB/S = RB/S \quad (39)$$

$$*(e) : RL/L \cap RN/S \quad (40)$$

$$RL/L \cap RL/S = RL/L \quad (41)$$

$$(e) : RL/L \cap (RL/S \cap RN/S) = RL/L \cap RN/S \quad (42)$$

$$*(f) : RL/L \cap RU/S \quad (43)$$

$$(f) : RL/L \cap RB/S = RL/L \cap RU/S \quad (44)$$

$$*(g) : RU/L \cap RN/S \quad (45)$$

$$*(h) : RU/L \cap RL/S \quad (46)$$

$$*(i) : RU/L \cap (RL/S \cap RN/S) \quad (47)$$

$$RU/L \cap RU/S = RU/S \quad (48)$$

$$RU/L \cap RB/S = RB/S \quad (49)$$

$$*(j) : RB/L \cap RN/S \quad (50)$$

$$RB/L \cap RL/S = RB/L \quad (51)$$

$$(j) : RB/L \cap (RL/S \cap RN/S) = RB/L \cap RN/S \quad (52)$$

$$(f) : RB/L \cap RU/S = RU/S \cap RL/L \quad (53)$$

$$(f) : RB/L \cap RB/S = RU/S \cap RL/L \quad (54)$$

である。 □

(証明) (35) は、性質1から $S(d) \subseteq L(d)$ であり、 $RN/S \subseteq RN/L$ なので $RN/L \cap RN/S = RN/S$ である。

(36) について、 RN/L に含まれる d のうち、 $S(d)$ に \mathbf{L} のラベルと順序がないラベルを含むような d は RL/S に含まれない。よって、 RN/L と RL/S は包含関係がなく、これまでの記述に該当するものはない。

(37) は、(35) の証明より $RN/L \cap RN/S = RN/S$ なので、 $RN/L \cap (RL/S \cap RN/S) = (RL/S \cap RN/S)$ となり新しい記述にはならない。(38) と (39) について、 RU/S や RB/S に含まれる d は、 \mathbf{L} のラベルと $S(d)$ のラベルが順序を持つので、 $S(d) \subseteq L(d)$ のため d は RN/L にも含まれる。よって、 $RN/L \cap RU/S = RU/S$ 、 $RN/L \cap RB/S = RB/S$ である。

(40) について、 RL/L に含まれる d のうち、 $S(d)$ のラベルが \mathbf{L} のラベルと順序がないような d は RN/S に含まれない。よって、 RL/L と RN/S は包含関係がなく、これまでの記述に該当するものはない。

(41) は補題4の(24)の証明と同様に考えれば明らかである。(42) は、(41)の証明より $RL/L \cap RL/S = RL/L$ なので $RL/L \cap RN/S$ となり (e) に該当する。

(43) は (41) の証明と同様に考えれば、明らかに RL/L と RU/S は包含関係がなく、これまでの記述に該当するものはない。

(44) は、 $RL/L \cap RB/S = RL/L \cap (RL/S \cap RU/S)$ となり、(41) の証明より $RL/L \cap RL/S = RL/L$ なので、 $RL/L \cap RU/S$ となり (f) に該当する。

(45) について、 RU/L に含まれる d のうち $S(d) = \phi$ のような d は RN/S に含まれないので、 RU/L と RN/S は包含関係がなく、これまでの記述に該当するものはない。(46) と (47) についても同様に考えれば明らかである。

(48) について、 $S(d) \subseteq L(d)$ であり $RU/S \subseteq RU/L$ なので $RU/L \cap RU/S = RU/S$ である。(49) は、 $RU/L \cap RB/S = RU/L \cap (RU/S \cap RL/S)$ であり、(48) の証明より $RU/L \cap RU/S = RU/S$ なので $RU/S \cap RL/S = RB/S$ である。

(50) について、 RB/L に含まれる d のうち $S(d) = \phi$ のような d は RN/S に含まれないので、 RB/L と RN/S は包含関係がなく、これまでの記述に該当するものはない。

(51) は、 $RB/L \cap RL/S = (RU/L \cap RL/L) \cap RL/S$ であり、(41) の証明より $RL/L \cap RL/S = RL/L$ なので $RU/L \cap RL/L = RB/L$ である。(52) は、(51) の証明より $RB/L \cap RL/S = RB/L$ なので $RB/L \cap RN/S$ であり (j) に該当する。(53) は、 $RB/L \cap RU/S = (RU/L \cap RL/L) \cap RU/S$ であり、(48) の証明より $RU/L \cap RU/S = RU/S$ なので $RU/S \cap RL/L$ であり (f) に該当する。(54) は、 $RB/L \cap RB/S = (RU/L \cap RL/L) \cap (RU/S \cap RL/S)$ であり、(48) の証明より $RU/L \cap RU/S = RU/S$ 、かつ (41) の証明より $RL/L \cap RL/S = RL/L$ なので、 $RU/S \cap RL/L$ であり (f) に該当する。 (証明終)

ラベル集合 \mathbf{L} とデータ d に対して、 $P(d)$ と $S(d)$ の組合せについて検討する。

[補題6] ラベル集合 \mathbf{L} とデータ d に対して、

$$* (k) : RN/P \cap RN/S \quad (55)$$

$$* (l) : RN/P \cap RL/S \quad (56)$$

$$* (m) : RN/P \cap (RL/S \cap RN/S) \quad (57)$$

$$* (n) : RN/P \cap RU/S \quad (58)$$

$$* (o) : RN/P \cap RB/S \quad (59)$$

$$* (p) : RL/P \cap RN/S \quad (60)$$

$$RL/P \cap RL/S = RL/L \quad (61)$$

$$(e) : RL/P \cap (RL/S \cap RN/S) = RL/L \cap RN/S \quad (62)$$

$$* (q) : RL/P \cap RU/S \quad (63)$$

$$(f) : RL/P \cap RB/S = RL/L \cap RU/S \quad (64)$$

$$* (r) : RU/P \cap RN/S \quad (65)$$

$$* (s) : RU/P \cap RL/S \quad (66)$$

$$* (t) : RU/P \cap (RL/S \cap RN/S) \quad (67)$$

$$* (u) : RU/P \cap RU/S \quad (68)$$

$$* (v) : RU/P \cap RB/S = RU/P \cap RU/S \cap RL/S \quad (69)$$

$$* (w) : RB/P \cap RN/S \quad (70)$$

$$(a) : RB/P \cap RL/S = RL/L \cap RU/P \quad (71)$$

$$* (x) : RB/P \cap (RL/S \cap RN/S) = RL/L \cap RU/P \cap RN/S \quad (72)$$

$$* (y) : RB/P \cap RU/S = RU/P \cap RU/S \cap RL/P \quad (73)$$

$$* (z) : RB/P \cap RB/S = RU/P \cap RU/S \cap RL/L \quad (74)$$

である。 □

(証明) (55) について、 RN/P に含まれる d のうち $S(d)$ のラベルが \mathbf{L} の

ラベルと順序がないような d は RN/S に含まれないので、 RN/P と RN/S は包含関係がなく、これまでの記述に該当するものはない。(56) は、 RN/P に含まれる d のうち $S(d)$ のラベルが L 以外のラベルと順序があるような d は RL/S に含まれないので、 RN/P と RL/S は包含関係がなく、これまでの記述に該当するものはない。(57) は、(55) と (56) の証明から明らかにこれまでの記述に該当するものはない。(58) について、 RN/P に含まれる d のうち $S(d) = \phi$ であるような d は RU/S に含まれないので、 RN/P と RU/S は包含関係がなく、これまでの記述に該当するものはない。(59)、(60)、(63) は (58) と同様に考えれば明らかである。

(61) は、 $RL/P \cap RL/S$ に含まれる d の $P(d)$ と $S(d)$ のいずれのラベルも L のラベルと順序がある。すなわち、 RL/L で記述されるデータに該当するので、 $RL/P \cap RL/S = RL/L$ である。(62) は、(61) の証明より $RL/P \cap RL/S = RL/L$ なので $RL/L \cap RN/S$ であり (e) に該当する。(64) は、 $RL/P \cap RB/S = RL/P \cap (RU/S \cap RL/S)$ であり、(61) の証明より $RL/P \cap RL/S = RL/L$ なので $RU/S \cap RL/L$ であり (f) に該当する。

(65) について、 RU/P に含まれる d のうち $S(d) = \phi$ であるような d は RN/S に含まれないので、 RU/P と RN/S は包含関係がなく、これまでの記述に該当するものはない。(66) から (69) までは同様に考えれば明らかである。また、(70) も $RB/P \subseteq RU/P$ が成り立つので同様に考えれば明らかである。

(71) は、 $RB/P \cap RL/S = (RU/P \cap RL/P) \cap RL/S$ であり、(61) の証明より $RL/P \cap RL/S = RL/L$ なので $RU/P \cap RL/L$ であり (a) に該当する。

(72) は、 $RB/P \cap (RL/S \cap RN/S) = (RU/P \cap RL/P) \cap (RL/S \cap RN/S)$ であり、(61) の証明より $RL/P \cap RL/S = RL/L$ なので $RL/L \cap RU/P \cap RN/S$ である。 $RL/L \cap RU/P \cap RN/S$ における RL/L 、 RU/P 、 RN/S は互いに包含せず、(72) はこれまでの記述に該当するものはない。(73) と (74) も同様である。 (証明終)

最後に、ラベル集合 L とデータ d に対して、 $L(d)$ 、 $P(d)$ 、 $S(d)$ のすべての組合せについて検討する。なお、補題 4、5、6 の通り $L(d)$ 、 $P(d)$ 、 $S(d)$ のうち 2 組を選択した場合はすでに議論してきた。その結果、組合せによっては新たな記述とはならないものが明らかになっており、補題 4、5、6 の中では補題 4 が示す (a)、(b)、(c) の 3 通りが最小である。よって、補題 4 に $S(d)$ を組合せて考えるのが最も効率的であり、以下の補題が成り立つ。

[補題7] ラベル集合 L とデータ d に対して、

$$(x) : RL/L \cap RU/P \cap RN/S \quad (75)$$

$$(a) : RL/L \cap RU/P \cap RL/S = RL/L \cap RU/P \quad (76)$$

$$(x) : RL/L \cap RU/P \cap (RL/S \cap RN/S) \\ = RL/L \cap RU/P \cap RN/S \quad (77)$$

$$(z) : RL/L \cap RU/P \cap RU/S \quad (78)$$

$$(z) : RL/L \cap RU/P \cap RB/S \\ = RL/L \cap RU/P \cap RU/S \quad (79)$$

$$*(bgk) : RU/L \cap RN/P \cap RN/S \quad (80)$$

$$*(bhl) : RU/L \cap RN/P \cap RL/S \quad (81)$$

$$*(bim) : RU/L \cap RN/P \cap (RL/S \cap RN/S) \quad (82)$$

$$(n) : RU/L \cap RN/P \cap RU/S = RN/P \cap RU/S \quad (83)$$

$$(o) : RU/L \cap RN/P \cap RB/S = RN/P \cap RB/S \quad (84)$$

$$*(cgp) : RU/L \cap RL/P \cap RN/S \quad (85)$$

$$RU/L \cap RL/P \cap RL/S = RB/L \quad (86)$$

$$(j) : RU/L \cap RL/P \cap (RL/S \cap RN/S) \\ = RB/L \cap RN/S \quad (87)$$

$$(q) : RU/L \cap RL/P \cap RU/S = RL/P \cap RU/S \quad (88)$$

$$(f) : RU/L \cap RL/P \cap RB/S \\ = RL/P \cap RB/S = RL/L \cap RU/S \quad (89)$$

である。□

(証明) (75) について、補題4の (a) の $RL/L \cap RU/P$ に対して RN/S の組合せを考える。 $(RL/L \cap RU/P) \cap RN/S$ は補題6の (x) に該当する。(76) についても同様に、 $(RL/L \cap RU/P) \cap RL/S$ は、(41) の証明より $RL/L \cap RL/S = RL/L$ なので $RL/L \cap RU/P$ となり (a) に該当する。(76) の証明と同様に (77) は $RL/L \cap RU/P \cap (RL/S \cap RN/S) = RL/L \cap RU/P \cap RN/S$ になり (x) に該当する。(78) は、 $(RL/L \cap RU/P) \cap RU/S$ であり (z) に該当する。(79) は、 $(RL/L \cap RU/P) \cap RB/S = (RL/L \cap RU/P) \cap (RU/S \cap RL/S)$ となり、(41) の証明より $RL/L \cap RL/S = RL/L$ なので $RL/L \cap RU/P \cap RU/S$ となり同じく (z) に該当する。

(80) について、補題4の (b) の $RU/L \cap RN/P$ に対して RN/S の組合せを考える。 RU/L 、 RN/P 、 RN/S は互いに包含せず、(80) はこれまでの記述に該当するものはない。また、(b) に補題5の (g) の $RU/L \cap RN/S$ と補題6の (k) の $RN/P \cap RN/S$ を組合せたものと考えられるので (bgk) としている。

(81)、(82)、(85) の証明についても同様である。

(83) は、(48) の証明より $RU/L \cap RU/S = RU/S$ なので $RN/P \cap RU/S$ となり (n) に該当する。(84) は、(49) の証明より $RU/L \cap RB/S = RB/S$ なので $RN/P \cap RB/S$ となり (o) に該当する。(86) は、(61) の証明より $RL/P \cap RL/S = RL/L$ なので $(RU/L \cap RL/P) \cap RL/S = RU/L \cap RL/L$ となり RB/L になる。(87) について、 $RU/L \cap RL/P \cap (RL/S \cap RN/S)$ は、(61) の証明より $RL/P \cap RL/S = RL/L$ なので $RU/L \cap RL/L \cap RN/S$ となり $RB/L \cap RN/S$ となって (j) に該当する。(88) は、(48) の証明より $RU/L \cap RU/S = RU/S$ なので $RL/P \cap RU/S$ となり (q) に該当する。(89) は、(49) の証明より $RU/L \cap RB/S = RB/S$ なので $RL/P \cap RB/S$ となる。さらに、(64) の証明より $RL/P \cap RB/S$ は (f) に該当するので (89) も (f) に該当する。
(証明終)

これまでの議論により、順序とランクに応じて考え得るすべての組合せから同一のデータを記述するものを除いた記述は定理1にまとめられる。

[定理1] ラベル集合 L とデータ d に対して、 L と d の関係にラベル集合の順序を適用して考えられるデータの記述は、 $X/Y (X \in \{RN, RU, RL, RB\}, Y \in \{L, P, S\})$ 、 $RN/S \cap RL/S$ 、(a) から (z)、(bgk)、(bhl)、(bim)、(cgp) の記述である。□

(証明) これまでの議論と補題1から7より明らかである。(証明終)

6 分析に有用な記述

本章は、これまでの議論で導出したラベル集合の順序を適用して考えられるデータの記述に対して、分析に有用な記述に集約して整理をする。

定理1で示した43通りの記述に関して、データの分析に有用な指定かどうかという観点で整理をする。まず、ラベル集合 L とデータ d に対して、 d が L のいずれかに関連するのか、すべてに関連するのかで2通りの指定が考え得る。さらに、関連の種類に関しても、 L と主たる関連がある、副次的な関連がある、主たる関連もしくは副次的な関連がある (ランクを意識しない)、主たる関連かつ副次的な関連もあるという4種類の指定がある。すなわち、それらの指定は L のラベルに対して、 $P(d)$ のラベルで順序がある、 $S(d)$ のラベルで順序がある、 $P(d)$ もしくは $S(d)$ ($P(d) \cup S(d) = L(d)$) のラベルで順序がある、 $P(d)$ と $S(d)$ の両方 ($P(d) \cap S(d)$) のラベルで順序がある記述が対応する。一方、 L

以外のラベルとの関連を考えると、 L と関連がないラベルを含んでもよい指定と含まない指定に分かれる。含まない指定については、さらに、主たる関連がなければよい、副次的な関連がなければよい、どちらの関連もあってはならないという3通りの指定がある。それらの指定は、 L のラベルと順序がないラベルが、 $P(d)$ がない、 $S(d)$ がない、 $P(d)$ と $S(d)$ ($L(d)$) がないということであり、 $P(d)$ 、 $S(d)$ 、 $L(d)$ に制限をつけることを意味する。これまでに議論した指定法に対して、定理1の記述を対応させたものが図2である。

図2で示した通り、ラベル集合 L と L 以外に対して、 L のいずれかが P と関連しかつ P のみ制限する指定と L のいずれかが L と関連しかつ P のみ制限する指定はともに RL/P の記述、 L のいずれかが P と関連しかつ L を制限する指定と L のいずれかが L と関連しかつ L を制限する指定はともに RL/L の記述、 L のいずれかが S と関連しかつ P のみを制限する指定と L のいずれかが P と S で関連しかつ P のみ制限する指定はともに $RN/S \cap RL/P$ の記述、 L のいずれかが S と関連しかつ L を制限する指定と L のいずれか P と S で関連しかつ L を制限する指定はともに $RN/S \cap RL/L$ の記述になる。それらは、性質3の $P(d) \neq \phi$ によるところであり、4組が同一の記述になるので指定の種類も集約できる。これまでの議論をまとめると、図2に示したように3つの観点を組合せた32通りの指定になるが、4組の指定が同一のデータを記述するので最終的に28通りの指定となる。

L の いずれか	ラベル集合 L 以外のラベルとの関連			
	制限無	P を制限	S を制限	P と $S(L)$ を制限
P	RN/P	RL/P^*	$RN/P \cap RL/S(l)$	$RN/P \cap RL/L = RL/L^{**}$
S	RN/S	$RN/S \cap RL/P^{***}(p)$	$(RN/S \cap RL/S)$	$RN/S \cap RL/L^{****}(e)$
$P \cup S = L$	RN/L	$RN/L \cap RL/P = RL/P^*$	$RN/L \cap RL/S(d)$	RL/L^{**}
$P \cap S$	$RN/P \cap RN/S(k)$	$RN/S \cap RL/P^{***}(p)$	$RN/P \cap (RN/S \cap RL/S)(m)$	$RN/S \cap RL/L^{****}(e)$

L の すべて	ラベル集合 L 以外のラベルとの関連			
	制限無	P を制限	S を制限	P と $S(L)$ を制限
P	RU/P	$RU/P \cap RL/P = RB/P$	$RU/P \cap RL/S(s)$	$RU/P \cap RL/L(a)$
S	RU/S	$RU/S \cap RL/P(q)$	$RU/S \cap RL/S = RB/S$	$RU/S \cap RL/L(f)$
$P \cup S = L$	RU/L	$RU/L \cap RL/P(c)$	$RU/L \cap RL/S(h)$	$RU/L \cap RL/L = RB/L$
$P \cap S$	$RU/P \cap RU/S(u)$	$RU/P \cap RU/S \cap RL/P(y)$	$RU/P \cap RU/S \cap RL/S(v)$	$RU/P \cap RU/S \cap RL/L(z)$

注：*、**、***、****が付された記述は同一のデータであることを示している。

図2 データ分析に有用な指定と記述の対応

一方、図2に該当しない定理1の記述は以下の通りである。

1. RL/S
2. $(b) : RU/L \cap RN/P$
3. $(g) : RU/L \cap RN/S$
4. $(i) : RU/L \cap (RL/S \cap RN/S)$
5. $(j) : RB/L \cap RN/S$
6. $(n) : RN/P \cap RU/S$
7. $(o) : RN/P \cap RB/S$
8. $(r) : RU/P \cap RN/S$
9. $(t) : RU/P \cap (RL/S \cap RN/S)$
10. $(w) : RB/P \cap RN/S$
11. $(x) : RB/P \cap (RL/S \cap RN/S)$
12. $(bgk) : RU/L \cap RN/P \cap RN/S$
13. $(bhl) : RU/L \cap RN/P \cap RL/S$
14. $(bim) : RU/L \cap RN/P \cap (RL/S \cap RN/S)$
15. $(cgp) : RU/L \cap RL/P \cap RN/S$

1の RL/S は、4章で述べたように、ラベル集合 L のラベルと全く関連のないデータも記述されることになるので有用な意味を持つとは言えない。また、図2で示した記述は単独で分析に有用な意味を持つものに対して、残りの記述は単独で有用な意味を持つわけではなく、複数の指定を組合せた問合せである。例えば、2の**(b)** : $RU/L \cap RN/P$ について考えると、 RU/L により L のすべてのラベルに関連するデータを指定した上で、さらにその中で主たる関連があるものを選択する場合に RN/P を組合せて行う問合せになる。したがって、 RU/L と RN/P を個別に用意しておき、組合せて用いればよいので、単独で有用な意味を持つ記述として考える必要はない。その他残りの記述に関しても同様である。

7 おわりに

本稿は、ラベル集合とデータのランク付集合ラベルの関係にラベル集合の順序を適用し、順序とランクに応じて考え得るすべての組合せによるデータの記述を導出した。その上で、分析に有用な指定に対応する記述であるかどうかという観点で記述を整理して、分析に有用なデータの記述を明らかにした。

まず、4通りのラベル集合の順序があり、さらに順序を適用する対象がデータ d の $L(d)$ 、 $P(d)$ 、 $S(d)$ という3通りの区分があるので12通りのデータの記述がある。それらの記述はそれぞれ異なるデータを記述する。そして、ランクを固

定した上で4通りの順序から2つ順序を用いて記述されるデータの組合せは6通りであり、それらを3つの区分に対して適用すると18通りの記述になる。それらの記述では、 $RN/S \cap RL/S$ の1通りのみが新たな意味を持つ。さらに、ランクを組合せた71通りの記述では、(a) から (z) と (bgl)、(bhl)、(bim)、(cgp)の30通りが新たな意味を持つ記述になる。

次に、これまでの43 (=12 + 1 + 30)通りの記述のうち、分析に有用な記述かどうかに関してラベル集合 L とデータ d のランク付集合ラベルの関係に着目して考察を行った。 L 内のラベルとの関連という観点では、いずれかに関連、すべてに関連するのかで2通り、 L との関連の程度という観点では、主たる関連、副次的な関連、主たる関連もしくは副次的な関連 (ランクを意識しない)、主たる関連かつ副次的な関連があるという4通り、最後に、 L 以外のラベルとの関連という観点では、 L と関連がないラベルを含んでもよい、主たる関連がなければよい、副次的な関連がなければよい、どちらの関連もあってはならないという4通りがある。よって、3つの観点を組合せた32通りの指定になるが、4組の指定が同一のデータを記述するので最終的に28通りの指定となる。そして、28通りの指定は、43通りの記述のうちで対応することが判明したのと同時に、残り15通りの記述は単独では有用な意味をなさないことも示した。

参考文献

- [1] 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).
- [2] Fazzinga, B., Flesca, S., and Furfaro, F.: RFID-Data Compression for Supporting Aggregate Queries, *ACM Transaction on Database Systems*, Vol. 38, Issue 2, pp. 11–45 (2013).
- [3] 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).
- [4] 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).
- [5] 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)。

- [6] 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).
- [7] Kumar, A., Niu, F., and Re, C.: Hazy: Making It Easier to Build and Maintain Big-Data Analytics, *Communications of the ACM*, Vol. 56, No. 3, pp. 40-49 (2013).
- [8] 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、情報処理学会、Vol. 47、No. SIG8 (TOD30)、pp. 1-10 (2006)。
- [9] 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、愛知学院大学産業研究所所報「地域分析」、愛知学院大学産業研究所、Vol. 49、No. 1、pp. 43-65 (2010)。
- [10] 葛西正裕: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、愛知学院大学論叢「商学研究」、愛知学院大学商学会、Vol. 52、No. 3、pp. 1-17 (2012)。
- [11] 葛西正裕:企業におけるビッグデータの利活用とデータベース技術を用いた分析、愛知学院大学論叢「商学研究」、愛知学院大学商学会、Vol. 53、No. 1、pp. 67-90 (2012)。
- [12] 葛西正裕、古川哲也:データ分析におけるランク付集合ラベルの利用、第6回データ工学と情報マネジメントに関するフォーラム(第12回日本データベース学会年次大会)、A9-4 (2014)。
- [13] Silla, C. and Freitas, A.: A Survey of Hierarchical Classification Across Different Application Domains, *Data Mining and Knowledge Discovery*, Vol. 22, Issue 1-2, pp. 31-72 (2011).
- [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] Veloso, A., Almeida, H., Goncalves, M., and Meira, W.: Learning to Rank at Query-Time Using Association Rules, *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval (SIGIR'08)*, pp. 267-274 (2008).
- [16] 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).

- [17] 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).
- [18] Wang, Y., Yang, B., Qu, L., Spaniol, M., and Weikum, G.: Harvesting Facts from Textual Web Sources by Constrained Label Propagation, *Proc. Int'l Conf. on Information and Knowledge Management (CIKM'11)*, pp. 837–846 (2011).
- [19] Zhang, X., Yang, Y., Han, Z., Wang, H., and Gao, C.: Object Class Detection: A Survey, *ACM Computing Surveys*, Vol. 46, Issue 1, pp. 10–53 (2013).