ランク付集合ラベルデータの分析のための理論的枠組み

葛西　正裕†　　古川　哲也††

† 愛知学院大学経済学部　〒 462–8739 愛知県名古屋市北区名城 3-1-1
†† 九州大学大学院経済学研究院　〒 812–8581 福岡県福岡市東区箱崎 6–19–1
E-mail: †kuzunisi@dpc.agu.ac.jp, ††furukawa@econ.kyushu-u.ac.jp

あらまし　データの性質などをラベルとして表すモデルでは，該当する複数のカテゴリのラベル（集合ラベル）が付される．データに対するラベルの適合度は異なり，ラベルにランクを導入することで詳細なデータ分析ができる．本稿は，多様なデータを様々な目的で分析でき，ランクがないデータに対しても分析を可能にするための理論的枠組みを提案する．分析対象となるラベル集合とデータの集合ラベルとの関連の強さを評価する基準は，存在性，網羅性，排他性の要素条件からなる．要素条件の比較で評価基準間の強さの比較が可能かを判断できるので，評価基準の強さの順序（系列）が導ける．系列により，関連の強さを一元的かつ段階的に評価でき，異なる強さのデータ集合間の比較によって分析対象に関する有用な情報を得られる．さらに，ランクの区分を k 個に拡張することでより詳細な分析が可能になる．同時に，ランクが付けられないようなデータに対しては，最下位ランクよりも弱い $k+1$ 個目のランクの不明ラベルを設けることで対応できる．

キーワード　集合ラベル，ランク付ラベル，評価基準，データ分析，データモデル，データベース理論

1. はじめに

数値，テキスト，画像，音声，動画といった多様なデータが分析されている [6] [9] [10]．データの種類を問わず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．データの性質などをデータのラベルとして表すモデルでは，複数のラベル（集合ラベル）が付される [1] [2] [5]．データに対するラベルの適合度は異なることが多く，ラベルにランクを導入することで詳細なデータ分析ができる．例えば，企業の事業展開地域と業績との関連を分析する際，日本と米国の両地域に展開する企業の業績は全企業と比較して高くはないが，重点的に米国に展開している日本企業に限れば好業績であるというような知見が得られるようになる．さらに，両地域に展開はしているが両地域以外を主としている企業の業績についても分析が可能になる．分析対象を日本，米国といったラベル集合で与え，ラベル集合との関連の強さが異なるデータ集合間を比較することで有用な情報が得られる．

ランクの区分が多いほど詳細な情報を得られるが，議論を簡単にするため主ラベルと副ラベルの 2 つの区分で考えると，ランク付集合ラベルには次の性質が想定される．

性質 1　：少なくとも 1 つは主ラベルである．
性質 2　：主ラベルまたは副ラベルではないラベルは存在しない．
性質 3　：主ラベルかつ副ラベルであるラベルは存在しない．

これまでの研究 [4] は，これらの性質を満たすデータに対し分析の理論的基盤を提案した．より多様なデータを様々な目的で分析する際には，性質 1 を満たさない場合がある．例えば，バーゲンセール対象商品の売行きを分析するために POS で得られたバスケットデータに対し，バーゲンセールといったイベントの有無で主ラベルと副ラベルを区分すると，いずれの商品もバーゲンセールの対象でなければ主ラベルは存在しない．また，商品に対する評価が “きれい”，“使いやすい” といった感性に関するラベルで与えられたとき，いずれに対しても強くはない場合も主ラベルは存在しない．

ランクの判断が未着手な場合やランクを付けられないと判断した場合，ラベルにランクがないので性質 2 を満たさない．性質 3 については，主ラベルと副ラベルを区分する条件を同時に満たすことはなく，ラベルは主ラベル，副ラベルのどちらにかになるので常に満たす．本稿は，多様なデータを様々な目的で分析でき，主ラベルやランクがないデータに対しても分析を可能にする理論的枠組みを提案することで，汎用的かつ柔軟なデータ分析を可能にする．

ランクに関連する研究は主に情報検索の分野で行われており，個人の嗜好を数値化して評価した結果をランクとして与えることで，情報検索の高度化を図る研究がある [7] [8]．それに対し，本稿はランク付集合ラベルデータを対象にした分析を目的としている．研究 [3] は集合ラベルデータの分析を目的としているが，ランクを用いていない．

本稿は以下のように構成される．2 章は，性質 1 を満たさないランク付集合ラベルデータと分析対象となるラベル集合との関連の強さを議論し，強さを評価する基準を導く．3 章は，評価基準の強さについて述べ，その関係を明らかにする．4 章は，評価基準を組合わせた基準も評価基準になるので，関連の強さを評価するために必要なすべての評価基準を導く．5 章は，評価基準間の強さの有無について明らかにし，評価基準の強さの順序を系列として提案する．系列により，関連の強さを一元的かつ段階的に評価できる．6 章は，これまでの 2 ランクのデータから k（≥ 2）ランクのデータに拡張する．7 章は，性質 2 を

仮定しないデータについて述べ，ランクを付けられないデータには不明ラベルを $k+1$ 個目のランクとして設けることで対応できることを示す．8 章はまとめである．

2. ランク付集合ラベルデータと評価基準

データを分析に供する際，一般に概念の階層などに基づいてラベル集合が付される．ラベル集合の要素間には属性ごとに階層的な関係があることが多く [2] [3] [5]，例えば，地域であれば，日本，東海，愛知といった上下関係がある．本稿では，議論を簡単にするために属性は単一とする．また，ラベルは階層における最下層のものとし，データにはラベルが少なくとも 1 つは付されるものとする．

1 件のデータをオブジェクト o，ラベルを l，ラベル集合を $L=\{l_1,l_2,\cdots,l_n\}$ とする．o に付されたラベル集合を集合ラベルと呼び $L(o)$（$\neq\phi$）で表す．ラベルの概念の上下関係を $\prec$ とする．ラベル l_1，l_2 に対し，l_1 が l_2 の下位概念のラベルならば，l_1 は l_2 の下位（l_2 は l_1 の上位）であり，l_1 が l_2 の下位または l_1 と l_2 が等しい（$l_1 \preceq l_2$）とき，l_1 は l_2 に関連するという．また，ラベル l がラベル集合 L 中のいずれかのラベルに関連するとき，l は L に関連するという．

ラベル集合 L，L' に対し，L' に関連する L 中のラベルの集合を $Ring_{L'}(L)=\{l \mid l\in L, \exists l'\in L', l\preceq l'\}$，$L$ のラベルが関連する L' 中のラベルの集合を $Red_L(L')=\{l' \mid l'\in L', \exists l\in L, l\preceq l'\}$ で表す．$Ring_{L'}(L)\neq\phi$ であることは $Red_L(L')\neq\phi$ と等価である．

分析対象をラベル集合 $\mathcal{L}$（$\neq\phi$）で与え，$\mathcal{L}$ とオブジェクト o との関連の強さを検討する．$\mathcal{L}$ に関連するラベルが $L(o)$ にあるとき，すなわち，$Ring_{\mathcal{L}}(L(o))\neq\phi$ のとき，o は $\mathcal{L}$ に関連するという．また，$\mathcal{L}$ に関連するオブジェクトの集合を $\overline{\mathcal{L}}$ で表す．

オブジェクト o の集合ラベルの各ラベルに，o との関連の強さを表すランクを与える．o に強く関連する $L(o)$ 中のラベルを主ラベル（Primary Label），主ラベルほどではないが o に関連する $L(o)$ 中のラベルを副ラベル（Secondary Label）とし，主ラベルの集合を $P(o)$，副ラベルの集合を $S(o)$ で表す．ラベルがランクで区分された集合ラベルをランク付集合ラベルという．$L(o)$ には次の性質が想定される [4].

［性質 1］ オブジェクト o に対し，$P(o)\neq\phi$ である．

［性質 2］ オブジェクト o に対し，$L(o)=P(o)\cup S(o)$ である．

［性質 3］ オブジェクト o に対し，$P(o)\cap S(o)=\phi$ である．

$L(o)$ のラベルで最も強く関連しているラベルを主ラベルと考えれば，主ラベルは必ず存在し性質 1 が成り立つ．また，$L(o)$ の各ラベルは主ラベルもしくは副ラベルであり，性質 2 と性質 3 が成り立つ．

ラベルのランクは，オブジェクトの性質や分析目的に応じて判断される．例えば，業種をラベルとする企業データでは，売上高などの財務情報で閾値を設定し判断される場合がある．閾値によっては，すべてのラベルが主ラベルと判断される閾値を超えない場合があり，そのようなオブジェクトは主ラベルが存在せず性質 1 を満たさない．また，事実の有無といった名義尺度や感性の程度といった順序尺度のような定性的データでも主ラベルが存在しないことがある．ランクの判断基準やデータの性質は様々なので，本稿では性質 1 を仮定しない．

ラベル集合とオブジェクトとの関連の強さを評価できる基準があれば，関連の強さの差異に基づく分析ができる．オブジェクト o_1，o_2 に対し，o_2 の方が o_1 よりもラベル集合 $\mathcal{L}$ との関連が強いことを $o_1 <_{\mathcal{L}} o_2$，$\mathcal{L}$ が明確であれば $o_1 < o_2$ で表す．

［定義 1］ オブジェクト o_1，o_2 と条件 cnd に対し，o_2 は cnd を満たし o_1 は cnd を満たさないとき $o_1 <_{\mathcal{L}} o_2$ ならば，cnd はラベル集合 $\mathcal{L}$ とオブジェクトとの関連の強さの評価基準である．

ラベル集合 $\mathcal{L}$ とオブジェクト o との関連の強さは，$\mathcal{L}$ と $L(o)$ との関係で考えられ，次の側面がある．

存在性 ：o は $\mathcal{L}$ のいずれかのラベルに関連する．

網羅性 ：o は $\mathcal{L}$ のすべてのラベルに関連する．

排他性 ：o は $\mathcal{L}$ と無関係なラベルに関連しない．

関連の強さの側面は，次の評価基準になる．

LE ：$\mathcal{L}$ のいずれかのラベルに関連するラベルが存在（Exist）する．すなわち，$Ring_{\mathcal{L}}(L(o))\neq\phi$（$Red_{L(o)}(\mathcal{L})\neq\phi$）．

LA ：$\mathcal{L}$ のすべて (All) のラベルに対し関連するラベルが存在する．すなわち，$Red_{L(o)}(\mathcal{L})=\mathcal{L}$.

LN ：$\mathcal{L}$ と無関係なラベル L（$L\not\preceq L', L'\not\preceq L, \forall L'\in\mathcal{L}$）が存在しない（Not Exist）．すなわち，$Ring_{\mathcal{L}}(L(o))=L(o)$.

これらの側面は主ラベルで考えても評価基準となる．

PE ：$\mathcal{L}$ のいずれかのラベルに関連する主ラベルが存在する（o は $\mathcal{L}$ のいずれかのラベルに強く関連する）．すなわち，$Ring_{\mathcal{L}}(P(o))\neq\phi$（$Red_{P(o)}(\mathcal{L})\neq\phi$）．

PA ：$\mathcal{L}$ のすべてのラベルに対し関連する主ラベルが存在する（o は $\mathcal{L}$ のすべてのラベルに強く関連する）．すなわち，$Red_{P(o)}(\mathcal{L})=\mathcal{L}$.

PN ：$\mathcal{L}$ と無関係な主ラベルが存在しない（o は $\mathcal{L}$ と無関係なラベルに強く関連しない）．すなわち，$Ring_{\mathcal{L}}(P(o))=P(o)$.

オブジェクト o の主ラベルは，o が $\mathcal{L}$ に強く関連することを表しており，o と $\mathcal{L}$ との関連の強さでは主ラベルと $\mathcal{L}$ との関連が優先される．よって，主ラベルがあれば副ラベルは関連の強さに影響しない．主ラベルがなければ，副ラベルに対する存在性，網羅性，排他性の評価基準はそれぞれ LE，LA，LN に一致する．したがって，副ラベルに基づく評価基準は必要ない．

評価基準 PN を満たすオブジェクトには，ラベル集合 $\mathcal{L}$ に関連しないオブジェクトが存在する．

［例 1］ ラベル集合 $\mathcal{L}$ = { 輸送機器，電気機械 } とし，$P(o_1)$ = { 四輪自動車 }, $S(o_1)$ = { 金融 } であるオブジェクト o_1，$P(o_2)=\phi, S(o_2)=$ { 四輪自動車，金融 } であるオブジェクト o_2，$P(o_3)=\phi, S(o_3)=$ { 金融 } であるオブジェクト o_3 について $\mathcal{L}$ との関連の強さを考える．o_1，o_2，o_3 は $\mathcal{L}$ と無関係な主ラベルを含まないので PN を満たす．一方，o_1 は主ラベル，o_2 は副ラベルに $\mathcal{L}$ の “輸送機器” に関連する “四輪自動車” があるので $\mathcal{L}$ に関連するが，o_3 は $\mathcal{L}$ に関連しない．

$\mathcal{L}$ は分析対象を表しているので，$\mathcal{L}$ に関連しないオブジェクトも満たすような PN は評価基準ではないとも考え得る．一方，PN を用いることでオブジェクトが $\mathcal{L}$ に関連するかどうかを問わずに分析対象の排他性を評価でき，分析対象ではない他

の概念と強いつながりがあるかが分かる．分析時に PN を用いるかは分析目的に応じて判断すればよいので，本稿は PN を評価基準として扱う．

3. 評価基準の強さの関係

評価基準の強さの関係が分かれば，ラベル集合とオブジェクトとの関連の強さを段階的かつ一元的に評価できるようになる．本章では，評価基準の強さの関係は評価基準の含意で判断できること [4] を用いて，評価基準の強さの関係を明らかにする．

オブジェクト o_1 と o_2 のラベル集合 $\mathcal{L}$ との関連の強さが，評価基準 cnd_1 を満たすかどうかよりも評価基準 cnd_2 を満たすかどうかで決まるならば，cnd_2 は cnd_1 よりも強い $\mathcal{L}$ の評価基準であるとする．すなわち，o_1 や o_2 が cnd_1 を満たすかどうかに関わらず o_2 が cnd_2 を満たし o_1 が cnd_2 を満たさないとき $o_1 < o_2$ ならば，cnd_2 は cnd_1 よりも強い $\mathcal{L}$ の評価基準である．これは，o_1 が cnd_1 を満たしていても, すなわち，o_1 が cnd_1 の評価基準で o_2 よりも強い関連がある可能性があったとしても，cnd_2 を満たさないことで関連の強さが決まることによる．

［定義 2］ ラベル集合 $\mathcal{L}$ と評価基準 cnd_1，cnd_2 に対し，オブジェクト o_2 は cnd_2 を満たし，cnd_1 を満たすオブジェクト o_1 が cnd_2 を満たさないとき $o_1 <_{\mathcal{L}} o_2$ であるならば，cnd_2 は cnd_1 よりも強い $\mathcal{L}$ の評価基準であるといい，$cnd_1 <_{\mathcal{L}} cnd_2$ で，$\mathcal{L}$ が明確なときは $cnd_1 < cnd_2$ で表す．

ラベル集合 $\mathcal{L}$ に関連するオブジェクト集合 $\overline{\mathcal{L}}$ で評価基準 cnd を満足するオブジェクトの集合を $\overline{\mathcal{L}}^{cnd}$(=$\{o \mid o \in \overline{\mathcal{L}}, o$ は cnd を満たす $\}$) で表す．$\overline{\mathcal{L}}$ は $\mathcal{L}$ に関連するオブジェクトの集合なので，$\overline{\mathcal{L}} = \overline{\mathcal{L}}^{LE}$ である．

評価基準を満たすオブジェクト集合の包含関係は，評価基準に含意があるどうかで判断できる．評価基準の強さの関係は評価基準を満たすオブジェクト集合の包含関係で決まるため，評価基準の強さの関係は評価基準の含意で判断できる [4].

［定理 1］ 評価基準 cnd_1，cnd_2 に対し，$cnd_1 < cnd_2$ と $cnd_2 \Rightarrow cnd_1$ は等価である．

定理 1 より，評価基準の強さの関係はその含意で考えることができる．含意は推移律を満たすので，評価基準の強さの関係も推移律を満たす．

評価基準 LE と LA はランク付集合ラベル全体を対象としており，PE と PA は主ラベルの集合を対象とする同じ性質である．主ラベルの集合はランク付集合ラベルの部分集合なのでこれらの評価基準には含意があり，強さの関係がある．また，評価基準 LN を満たすオブジェクトは，ラベル集合 $\mathcal{L}$ と無関係なラベルに関連しないので，$\mathcal{L}$ に無関係な主ラベルはなく PN も満たす．

［補題 1］ 評価基準 LE と PE，LA と PA，PN と LN に対し，$LE < PE$，$LA < PA$，$PN < LN$ である．

（証明） PE を満たすオブジェクト o では $Ring_{\mathcal{L}}(P(o)) \neq \phi$ である．$P(o) \subseteqq L(o)$ なので $Ring_{\mathcal{L}}(L(o)) \neq \phi$ であり，o は LE を満たす．よって，$PE \Rightarrow LE$ なので $LE < PE$ である．同様に，$Red_{P(o)}(\mathcal{L}) = \mathcal{L}$ ならば $Red_{L(o)}(\mathcal{L}) = \mathcal{L}$ であり，$PA \Rightarrow LA$ が成り立つので $LA < PA$ である．

LN を満たすオブジェクト o では $Ring_{\mathcal{L}}(L(o)) = L(o)$ なので，$Ring_{\mathcal{L}}(P(o)) = P(o)$ を満たす．よって，$LN \Rightarrow PN$ が成り立つので $PN < LN$ である． （証明終）

評価基準 LA を満たすオブジェクト o は，ラベル集合 $\mathcal{L}$ のすべてのラベルに関連するので LE も満たす．PE と PA についても同様である．

［補題 2］ 評価基準 LE と LA，PE と PA に対し，$LE < LA$，$PE < PA$ である．

（証明） LA を満たすオブジェクト o は，$Red_{L(o)}(\mathcal{L}) = \mathcal{L}$ であり，$\mathcal{L} \neq \phi$ より $Red_{L(o)}(\mathcal{L}) \neq \phi$ なので LE を満たす．よって，$LA \Rightarrow LE$ なので，$LE < LA$ である．$PE < PA$ についても同様である． （証明終）

評価基準 LN を満たすオブジェクト o は，ラベル集合 $\mathcal{L}$ と無関係なラベルに関連しないので，$L(o)$ は $\mathcal{L}$ に関連するラベルが必ず存在するため LE も満たす．

［補題 3］ 評価基準 LE と LN に対し，$LE < LN$ である．

（証明） LN を満たすオブジェクト o では $Ring_{\mathcal{L}}(L(o)) = L(o)$ であり，$\mathcal{L}$ と無関係なラベルに関連するラベルを $L(o)$ に含まない．$L(o) \neq \phi$ なので，$L(o)$ 中に $\mathcal{L}$ に関連するラベルが必ず存在するため $Ring_{\mathcal{L}}(L(o)) \neq \phi$ であり，o は LE を満たす． （証明終）

補題 1 から補題 3 で得られた評価基準の強さの関係は，節点を評価基準，枝を強さの関係とするグラフで表すと図 1 となる．例えば，評価基準 LE から PE への枝は $LE < PE$ を示している．

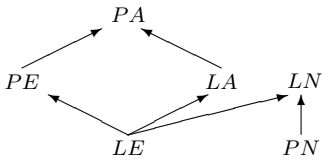

図 1 評価基準の強さの関係

4. 基底から得られる評価基準

本章は，これまでに明らかにした評価基準と評価基準の強さの関係をもとにして，評価基準を組合わせた基準を検討することで，関連の強さを評価するために必要なすべての評価基準を導く．

評価基準 cnd_i と cnd_j の積 $cnd_i \cdot cnd_j$ は評価基準である．評価基準の集合を $\boldsymbol{C} = \{cnd_1, cnd_2, \cdots, cnd_n\}$ とし，$\boldsymbol{C}$ から導出される評価基準集合，すなわち，$\boldsymbol{C}$ の要素の積によって得られる評価基準集合を $\boldsymbol{C}^+ = \{cnd_1, cnd_2, \cdots, cnd_n, cnd_1 \cdot cnd_2, \cdots, cnd_1 \cdot cnd_2 \cdots cnd_{n-1} \cdot cnd_n\}$ とする．評価基準集合 $\boldsymbol{C}_1$ と $\boldsymbol{C}_2$ は，$\boldsymbol{C}_1^+ = \boldsymbol{C}_2^+$ のとき等価であるという．また，評価基準集合 $\boldsymbol{C}$ は，等価な真部分集合が存在しないとき基底であるという．

［補題 4］ 評価基準集合 $\boldsymbol{C}_B = \{LE, LA, LN, PE, PA, PN\}$ は基底である．

（証明） 任意の評価基準 cnd（$\in \boldsymbol{C}_B$）に対し，$cnd \notin (\boldsymbol{C}_B - \{cnd\})^+$ なので，$\boldsymbol{C}_B$ の真部分集合は $\boldsymbol{C}_B$ と等価ではない．よって，$\boldsymbol{C}_B$ は基底である． （証明終）

評価基準 cnd_i と cnd_j が同値（$cnd_i \Leftrightarrow cnd_j$）ならば cnd_i と cnd_j の強さは同じであり，$cnd_i \equiv cnd_j$ と表す．また，$cnd_i \equiv cnd_j$ または $cnd_i < cnd_j$ であることを $cnd_i \leqq cnd_j$ と表す．cnd_i と cnd_j が同値ならば，cnd_i を満たすオブジェクト集合と cnd_j を満たすオブジェクト集合は一致する（$\overline{\mathcal{L}}^{cnd_i} = \overline{\mathcal{L}}^{cnd_j}$）．

基底を構成する評価基準の積で得られる評価基準に対し，評価基準の強さの関係を用いて新たな評価基準であるかを検討する．含意の関係にある評価基準の積は新たな評価基準ではない．例えば，$LA \Rightarrow LE$ なので $LE \cdot LA \equiv LA$ であり $LE \cdot LA$ は新たな評価基準ではない．また，評価基準 cnd 自身の積は cnd であり新たな評価基準ではない．一方，LA と含意の関係にない PE との積による評価基準 $PE \cdot LA$ は LA と同値ではなく，$LA < PE \cdot LA$ という強さの関係にあるので新たな評価基準である．図 2 において，下線を記した評価基準がこれまでの評価基準と同値ではなく新たな評価基準である．

	LE	PE	LA	PA	LN	PN
LE	–	PE	LA	PA	LN	$\underline{LE \cdot PN}$
PE	–	–	$\underline{PE \cdot LA}$	PA	$\underline{PE \cdot LN}$	$\underline{PE \cdot PN}$
LA	–	–	–	PA	$\underline{LA \cdot LN}$	$\underline{LA \cdot PN}$
PA	–	–	–	–	$\underline{PA \cdot LN}$	$\underline{PA \cdot PN}$
LN	–	–	–	–	–	LN
PN	–	–	–	–	–	–

図 2 基底の積で得られる評価基準

図 3 は，図 2 の新たに得られた評価基準と基底を構成する評価基準の積で得られる新たな評価基準を示している．評価基準 PA は $Red_{P(o)}(\mathcal{L}) = \mathcal{L}$ なので，$Red_{L(o)}(\mathcal{L}) = \mathcal{L}$ と $Red_{P(o)}(\mathcal{L}) \neq \phi$ がともに成り立つ．よって，$PA \Rightarrow PE \cdot LA$ であり $PE \cdot LA < PA$ なので，PA と $PE \cdot LA$ の積は新たな評価基準ではない．

図 3 で得られた評価基準 $PE \cdot LA \cdot LN$ について，PE，LA，LN 以外で基底を構成する LE，PA，PN との積で得られる評価基準について検討する．$PE \Rightarrow LE$，$LN \Rightarrow PN$ なので LE，PN との積は新たな評価基準ではなく，PA との積も $PA \cdot LN$ と同値となり新たな評価基準ではない．同様に，評価基準 $PE \cdot LA \cdot PN$ と LE，PA，LN との積で得られる評価基準は，$PE \Rightarrow LE$ なので LE との積は新たな評価基準ではなく，PA との積は $PA \cdot PN$，LN との積は $PE \cdot LA \cdot LN$ と同値となり，いずれも新たな評価基準ではない．

基底の積で得られるすべての評価基準を検証したので，これら以外に新たな評価基準は存在しない．よって，ラベル集合とオブジェクトとの関連の強さを評価するための基準は，基底（LE, PE, LA, PA, LN, PN）及び基底の積で得られる

	$PE \cdot LA$	$PE \cdot LN$	$LA \cdot LN$	$PA \cdot LN$
LE	$PE \cdot LA$	$PE \cdot LN$	$LA \cdot LN$	$PA \cdot LN$
PE	–	–	$\underline{PE \cdot LA \cdot LN}$	$PA \cdot LN$
LA	–	$PE \cdot LA \cdot LN$	–	$PA \cdot LN$
PA	PA	$PA \cdot LN$	$PA \cdot LN$	–
LN	$PE \cdot LA \cdot LN$	–	–	–
PN	$\underline{PE \cdot LA \cdot PN}$	$PE \cdot LN$	$LA \cdot LN$	$PA \cdot LN$

(a)

	$LE \cdot PN$	$PE \cdot PN$	$LA \cdot PN$	$PA \cdot PN$
LE	–	$PE \cdot PN$	$LA \cdot PN$	$PA \cdot PN$
PE	$PE \cdot PN$	–	$PE \cdot LA \cdot PN$	$PA \cdot PN$
LA	$LA \cdot PN$	$PE \cdot LA \cdot PN$	–	$PA \cdot PN$
PA	$PA \cdot PN$	$PA \cdot PN$	$PA \cdot PN$	–
LN	LN	$PE \cdot LN$	$LA \cdot LN$	$PA \cdot LN$
PN	–	–	–	–

(b)

図 3 新たな評価基準と基底との積で得られる評価基準

$PE \cdot LA, PE \cdot LN, LA \cdot LN, PA \cdot LN, LE \cdot PN, PE \cdot PN, LA \cdot PN, PA \cdot PN, PE \cdot LA \cdot LN, PE \cdot LA \cdot PN$ である．

5. 評価基準間の強さ

評価基準を満たすかどうかで，ラベル集合とオブジェクトとの関連の強さを評価することができる．複数の評価基準を用いれば，評価基準の強さの順序によって関連の強さを段階的かつ一元的に評価できるようになる．

評価基準 cnd は基底から得られるものなので，関連の強さを評価する側面，すなわち存在性，網羅性，排他性の条件で表せる．すべてのオブジェクトが満たす条件を $true$ とすれば，cnd は，$cnd_E (\in \{true, LE, PE\})$，$cnd_A (\in \{true, LA, PA\})$，$cnd_N (\in \{true, LN, PN\})$ を用いて $cnd_E \cdot cnd_A \cdot cnd_N$ で表せる．cnd に対し，$cnd \Leftrightarrow cnd_E \cdot cnd_A \cdot cnd_N$ のとき，(cnd_E, cnd_A, cnd_N) を cnd の要素条件という．例えば，LE では $(LE, true, true)$，$PE \cdot LA$ では $(PE, LA, true)$，$PE \cdot LA \cdot LN$ では (PE, LA, LN) が要素条件である．

評価基準の要素条件は，網羅性や排他性の条件が存在性の条件を含意する場合は 1 つに定まらない．例えば，$LA \Rightarrow LE$ なので，$(true, LA, true)$ と $(LE, LA, true)$ はともに LA の要素条件である．各側面の条件を個別に用いるとき，最も強い条件で考える必要がある．

［定義 3］ 要素条件 (cnd_E, cnd_A, cnd_N) の cnd_E は，$cnd_E < cnd_{E'}$ である要素条件 $(cnd_{E'}, cnd_{A'}, cnd_{N'})$ が存在しないとき cnd の存在性上限であるという．同様に，cnd_A と cnd_N に対し，それぞれ $cnd_A < cnd_{A'}$，$cnd_N < cnd_{N'}$ である $(cnd_{E'}, cnd_{A'}, cnd_{N'})$ が存在しないとき cnd の網羅性上限，排他性上限という．評価基準 cnd の要素条件 (cnd_E, cnd_A, cnd_N) は，cnd_E が存在性上限，cnd_A が網羅性上限，cnd_N が排他性

上限のとき cnd の要素上限という．

評価基準 cnd は cnd の要素上限を構成する存在性上限，網羅性上限，排他性上限の積と同値である．

［補題 5］ 評価基準 cnd と cnd の要素上限 cnd_E，cnd_A，cnd_N に対し，$cnd \Leftrightarrow cnd_E \cdot cnd_A \cdot cnd_N$ である．

（証明） 評価基準 cnd は基底の積で得られる（$cnd \in \boldsymbol{C}_B^+$）．$cnd$ を構成する評価基準集合 $\boldsymbol{C}$（$\subseteqq \boldsymbol{C}_B$）の cnd_i と cnd_j に対し，$cnd_i < cnd_j$ ならば $cnd_i \cdot cnd_j \Leftrightarrow cnd_j$ なので，$\boldsymbol{C}$ から cnd_i を除いた積で得られる評価基準は cnd と同値である．$\boldsymbol{C}$ を構成する評価基準に対し，補題 1 を用いて存在性，網羅性，排他性で最も強いもの，該当する評価基準がなければ $true$ を cnd'_E，cnd'_A，cnd'_N とすると，$cnd \Leftrightarrow cnd'_E \cdot cnd'_A \cdot cnd'_N$ である．補題 2 と補題 3 より網羅性や排他性の評価基準は存在性の評価基準を含意する場合がある．cnd'_A に対し補題 2 を満たす存在性の評価基準 cnd''_E が $cnd'_E < cnd''_E$ のとき，$cnd''_E < cnd'_A$ なので $cnd \Leftrightarrow cnd''_E \cdot cnd'_A \cdot cnd'_N$ である．cnd'_N に対し補題 3 を満たす存在性の評価基準 cnd'''_E が $cnd'_E < cnd'''_E$ のとき，$cnd'''_E < cnd'_N$ なので $cnd \Leftrightarrow cnd'''_E \cdot cnd'_A \cdot cnd'_N$ である．補題 2 と補題 3 より $cnd'''_E \leqq cnd''_E$ が成り立つので，cnd''_E は存在性上限である．cnd''_E が存在しなければ cnd'''_E，cnd'''_E が存在しなければ cnd'_E が存在性上限である．一方，網羅性や排他性の評価基準は他の側面の評価基準によって含意されないので，cnd'_A は網羅性上限，cnd'_N は排他性上限である．よって，cnd は cnd の存在性上限 cnd_E，網羅性上限 cnd_A，排他性上限 cnd_N の積と同値である． （証明終）

補題 5 より，評価基準は要素上限の積で考えることができる．4 章で明らかにした評価基準は，含意される評価基準を省略したもの，すなわち基底における最も少ない評価基準で構成されたものである．

評価基準の要素条件を要素上限で表す．LA は LE を含意するので，LA の要素条件は $(true, LA, true)$ ではなく $(LE, LA, true)$ である．同様に，PA は $(PE, PA, true)$，LN は $(LE, true, LN)$ である．基底の積の評価基準についても同様であり，例えば，$LA \cdot PN$ は (LE, LA, PN) である．

［定理 2］ 要素条件が $(cnd_{1E}, cnd_{1A}, cnd_{1N})$，$(cnd_{2E}, cnd_{2A}, cnd_{2N})$ である評価基準 cnd_1，評価基準 cnd_2 に対し，$cnd_1 \leqq cnd_2$ と $cnd_{1E} \leqq cnd_{2E}$，$cnd_{1A} \leqq cnd_{2A}$，$cnd_{1N} \leqq cnd_{2N}$ は等価であり，$cnd_1 \equiv cnd_2$ は $cnd_{1E} = cnd_{2E}$，$cnd_{1A} = cnd_{2A}$，$cnd_{1N} = cnd_{2N}$ に限る．

（証明） 定理 1 より，$cnd_1 < cnd_2$ と $cnd_2 \Rightarrow cnd_1$ は等価である．$cnd_{1E} \leqq cnd_{2E}$，$cnd_{1A} \leqq cnd_{2A}$，$cnd_{1N} \leqq cnd_{2N}$ ならば，$cnd_{1E} \Leftrightarrow cnd_{2E}$，$cnd_{1A} \Leftrightarrow cnd_{2A}$，$cnd_{1N} \Leftrightarrow cnd_{2N}$ を除けば，$cnd_{2E} \Rightarrow cnd_{1E}$ または $cnd_{2A} \Rightarrow cnd_{1A}$ または $cnd_{2N} \Rightarrow cnd_{1N}$ なので $cnd_2 \Rightarrow cnd_1$ が成り立ち，$cnd_1 < cnd_2$ である．また，$cnd_{1E} \Leftrightarrow cnd_{2E}$，$cnd_{1A} \Leftrightarrow cnd_{2A}$，$cnd_{1N} \Leftrightarrow cnd_{2N}$ ならば，$cnd_1 \Leftrightarrow cnd_2$ が成り立つので $cnd_1 \equiv cnd_2$ である．よって，$cnd_{1E} \leqq cnd_{2E}$，$cnd_{1A} \leqq cnd_{2A}$，$cnd_{1N} \leqq cnd_{2N}$ ならば $cnd_1 \leqq cnd_2$ である．

補題 1 から補題 3 の評価基準の強さの関係より，存在性の評価基準は網羅性と排他性といった他の側面の評価基準によって含意される場合があるが，網羅性や排他性の評価基準は他の側面の評価基準に含意されない．まず，$cnd_{2A} < cnd_{1A}$ とすると，$cnd_{1A} \Rightarrow cnd_{2A}$ なので，cnd_2 を構成する cnd_{2E}，cnd_{2A}，cnd_{2N} のいずれも cnd_{1A} を含意しないため $cnd_2 \Rightarrow cnd_1$ が成り立たない．よって，$cnd_1 \leqq cnd_2$ が成り立たないので，$cnd_1 \leqq cnd_2$ ならば $cnd_{1A} \leqq cnd_{2A}$ である．cnd_{1N} と cnd_{2N} についても同様で，$cnd_1 \leqq cnd_2$ ならば $cnd_{1N} \leqq cnd_{2N}$ である．$(cnd_{1E}, cnd_{1A}, cnd_{1N})$ は要素上限なので，$cnd_{1E} < cnd_{1E'}$ となる $cnd_{1E'}$ が存在すれば cnd_{1A} や cnd_{1N} に関係なく，$(cnd_{1E'}, cnd_{1A}, cnd_{1N})$ の評価基準は cnd_1 よりも強い．$(cnd_{2E}, cnd_{2A}, cnd_{2N})$ も要素上限なので同様である．$cnd_{2E} < cnd_{1E}$ とすると，cnd_{2E}，cnd_{2A}，cnd_{2N} のいずれも cnd_{1E} を含意しないため $cnd_2 \Rightarrow cnd_1$ が成り立たないので，$cnd_1 \leqq cnd_2$ ならば $cnd_{1E} \leqq cnd_{2E}$ である．したがって，$cnd_1 \leqq cnd_2$ ならば $cnd_{1E} \leqq cnd_{2E}$，$cnd_{1A} \leqq cnd_{2A}$，$cnd_{1N} \leqq cnd_{2N}$ である．

$cnd_1 \equiv cnd_2$ は $cnd_1 \Leftrightarrow cnd_2$ と等価である．cnd_1 と cnd_2 の要素上限が異なると同値ではないので，$(cnd_{1E}, cnd_{1A}, cnd_{1N}) = (cnd_{2E}, cnd_{2A}, cnd_{2N})$ である．よって，$cnd_1 \equiv cnd_2$ は $cnd_{1E} = cnd_{2E}$，$cnd_{1A} = cnd_{2A}$，$cnd_{1N} = cnd_{2N}$ に限る． （証明終）

評価基準 cnd_1 と cnd_2 に対し，要素条件の cnd_E，cnd_A，cnd_N ごとに強さを比較することで，cnd_1 と cnd_2 に強さの関係があるかについて判断できる．例えば，評価基準 $PA \cdot PN$，$PE \cdot LA$ について，それらの要素条件は (PE, PA, PN)，$(PE, LA, true)$ なので，$PE = PE$，$LA < PA$，$true < PN$ となり $PE{\cdot}LA < PA{\cdot}PN$ である．一方，$PA{\cdot}PN$ と $PE{\cdot}LA{\cdot}LN$ では，$PE = PE$，$LA < PA$ となるが $PN < LN$ なので両基準には強さの関係がない．4 章で明らかにした評価基準に対し，その強さの関係は図 4 となる．

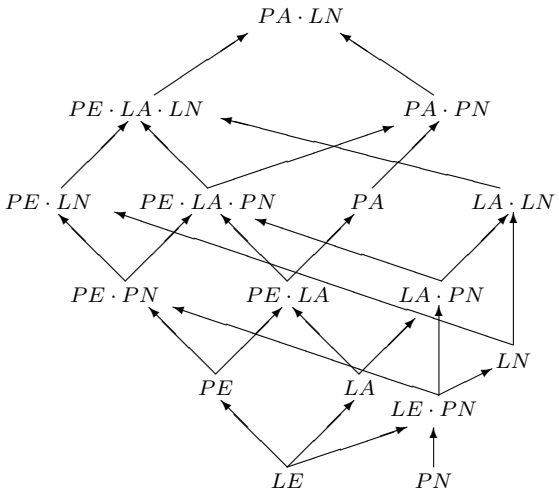

図 4 基底から得られる評価基準の強さの関係

［例 2］ 図 5 は，評価基準を満たすがそれよりも強い評価基準は満たさないランク付集合ラベルの例である．ラベル集合 $\mathcal{L}$ を

$\mathcal{L}=\{$ 輸送機器, 電気機械 $\}$ とし，$L(o)$ を $P(o), S(o)$ と記述している．下線のラベルは $\mathcal{L}$ と無関係なラベルである．ランク付集合ラベルが “$\{$ 四輪 $\}, \{$ 電気 $\}$” のオブジェクトは，評価基準 $PE\cdot LA\cdot LN$ まで満たすが，$\mathcal{L}$ の “電気機械” に関連する主ラベルを持たないので $PA\cdot LN$ は満たさない．

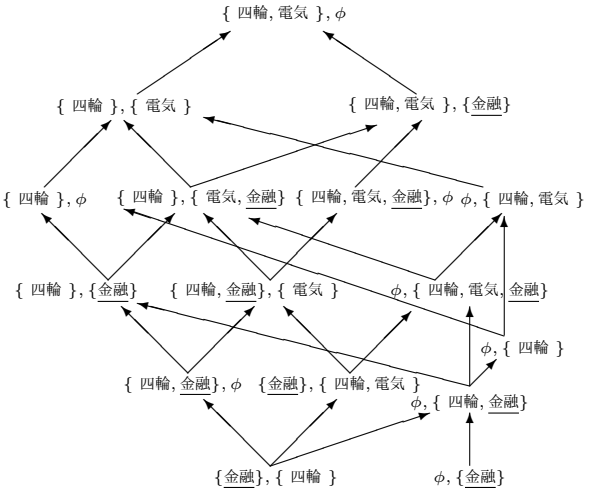

図 5　典型的なランク付集合ラベルの例

図 4 の LE と PN から $PA\cdot LN$ までの経路は，評価基準の強さの順序を示している．

［定義 4］ 評価基準集合 $\boldsymbol{C}$ に対し，$\forall cnd_i, cnd_j \in \boldsymbol{C}, cnd_i < cnd_j$ または $cnd_j < cnd_i$ であるとき，$\boldsymbol{C}$ は評価基準系列あるいは系列という．

評価基準 LE と PN から $PA\cdot LN$ までの経路に含まれる評価基準を要素とする評価基準集合及びその部分集合は系列である．系列は全順序集合であり，ラベル集合に対するオブジェクトの関連の強さを評価する 1 つの体系である．すなわち，オブジェクトが系列のどの評価基準まで満たすかによって関連の強さが段階的に決まり，オブジェクト間で強さを一元的に比較できる．定理 2 を用いることで，系列を構成する評価基準を容易に選択できる．例えば，評価基準 LA を満たすことが前提で，段階的に関連の強さを強くして分析したいのであれば，LA の要素条件は $(LE, LA, true)$ なので，存在性を強めるならば PE，網羅性を強めるならば PA，排他性を考慮するならば PN に要素条件を置き換えた $PE\cdot LA$，PA，$LA\cdot PN$ の評価基準を用いればよい．この場合，網羅性を強めると $(PE, PA, true)$ と存在性も強まる．先に存在性を強めた $PE\cdot LA$ を用いたのち，$(PE, LA, true)$ で PA に置き換えた PA を用いることで，段階的な分析が可能になる．PA の要素条件 $(PE, PA, true)$ は存在性と網羅性が上限なので，関連の強さを強くして分析するには排他性の強さを強くするのみである．

6. k ランクへの拡張

ランクの区分が多いほど詳細な情報を得られるので，これまでの 2 ランクの区分から制限のない k ランクに拡張する．k 個の区分を持つランクを強い順に $R_1, \cdots, R_i, \cdots, R_k$（$1 \leqq i \leqq k$）とする．これまでの議論は $k=2$ の場合であり，R_1 は主ラベル，R_2 は副ラベルである．

オブジェクト o の $L(o)$ におけるランク R_i のランク集合ラベルを $R_i(o)$ とし，R_i までのランク付集合ラベルを $R_i^*(o) = \bigcup_{j=1,i} R_j(o)$ とする．ランク付集合ラベルの性質 2 と 3 は，それぞれ次のようになる．

［性質 4］ オブジェクト o に対し，$L(o) = R_k^*$ である．

［性質 5］ オブジェクト o に対し，$R_i(o) \cap R_j(o) = \phi$（$i \neq j$）である．

k ランクのランク付集合ラベルのオブジェクトについても，存在性，網羅性，排他性の側面から次の評価基準が得られる．

R_iE （$1 \leqq i \leqq k$）：o は R_i 以上のいずれかのランクで $\mathcal{L}$ のいずれかのラベルに関連する．すなわち，$Ring_{\mathcal{L}}(R_i^*(o)) \neq \phi$ $(Red_{R_i^*(o)}(\mathcal{L}) \neq \phi)$．

R_iA （$1 \leqq i \leqq k$）：o は R_i 以上のいずれかのランクで $\mathcal{L}$ のすべてのラベルに関連する．すなわち，$Red_{R_i^*(o)}(\mathcal{L}) = \mathcal{L}$.

R_iN （$1 \leqq i \leqq k$）：o は R_i 以上のランクでは $\mathcal{L}$ と無関係なラベルに関連しない．すなわち，$Ring_{\mathcal{L}}(R_i^*(o)) = R_i^*(o)$.

2 ランクのランク付集合ラベルのオブジェクトにおける補題 1 から補題 3 は，k ランクではそれぞれ補題 6 から補題 8 に拡張できる．

［補題 6］ 評価基準 R_iE，R_iA，R_iN（$2 \leqq i \leqq k$）に対し，$R_iE < R_{i-1}E$，$R_iA < R_{i-1}A$，$R_{i-1}N < R_iN$ である．

（証明） $R_{i-1}E$ を満たすオブジェクト o では $Ring_{\mathcal{L}}(R_{i-1}^*(o)) \neq \phi$ である．$R_{i-1}^*(o) \subseteqq R_i^*(o)$ なので $Ring_{\mathcal{L}}(R_i^*(o)) \neq \phi$ であり，o は R_iE を満たす．よって，$R_{i-1}E \Rightarrow R_iE$ なので $R_iE < R_{i-1}E$ である．同様に，$Red_{R_{i-1}^*(o)}(\mathcal{L}) = \mathcal{L}$ ならば $Red_{R_i^*(o)}(\mathcal{L}) = \mathcal{L}$ であり，$R_{i-1}A \Rightarrow R_iA$ が成り立つので $R_iA < R_{i-1}A$ である．R_iN を満たすオブジェクト o では $Ring_{\mathcal{L}}(R_i^*(o)) = R_i^*(o)$ なので，$Ring_{\mathcal{L}}(R_{i-1}^*(o)) = R_{i-1}^*(o)$ を満たす．よって，$R_iN \Rightarrow R_{i-1}N$ が成り立つので $R_{i-1}N(o) < R_iN$ である． （証明終）

［補題 7］ 評価基準 R_iE と R_iA（$1 \leqq i \leqq k$）に対し，$R_iE < R_iA$ である．

（証明） R_iA を満たすオブジェクト o は，$Red_{R_i^*(o)}(\mathcal{L}) = \mathcal{L}$ であり，$\mathcal{L} \neq \phi$ より $Red_{R_i^*(o)}(\mathcal{L}) \neq \phi$ なので R_iE を満たす．よって，$R_iA \Rightarrow R_iE$ なので，$R_iE < R_iA$ である．（証明終）

［補題 8］ 評価基準 R_kE と R_kN に対し，$R_kE < R_kN$ である．

（証明） R_kN を満たすオブジェクト o では $Ring_{\mathcal{L}}(R_k^*(o)) = R_k^*(o)$ であり，$\mathcal{L}$ と無関係なラベルに関連するラベルを $R_k^*(o)$ に含まない．$R_k^*(o) \neq \phi$ なので，$R_k^*(o)$ 中に $\mathcal{L}$ に関連するラベルが必ず存在するため $Ring_{\mathcal{L}}(R_k^*(o)) \neq \phi$ であり，o は R_kE を満たす． （証明終）

補題 6 から補題 8 より評価基準の強さの関係は図 6 となる．

［補題 9］ 評価基準集合 $\boldsymbol{C}_B = \bigcup_{i=1,k}\{R_iE, R_iA, R_iN\}$ は基底である．

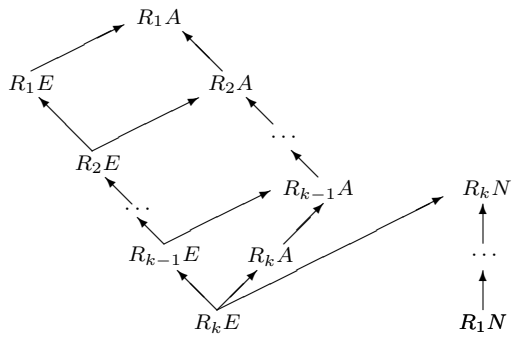

図 6　k ランクにおける評価基準の強さの関係

（証明）　任意の評価基準 cnd（$\in \boldsymbol{C}_B$）に対し，$cnd \notin (\boldsymbol{C}_B - \{cnd\})^+$ なので，$\boldsymbol{C}_B$ の真部分集合は $\boldsymbol{C}_B$ と等価ではない．よって，$\boldsymbol{C}_B$ は基底である．　（証明終）

評価基準の要素条件の cnd_E，cnd_A，cnd_N を以下のように定義すれば，補題 6 から補題 8 より，補題 5 が成り立つ．

$cnd_E \in \{true, R_kE, \cdots, R_1E\}$
$cnd_A \in \{true, R_kA, \cdots, R_1A\}$
$cnd_N \in \{true, R_kN, \cdots, R_1N\}$

ラベル集合と k ランクのランク付集合ラベルのオブジェクトとの関連の強さを評価するための基準は，基底から得られる評価基準であり補題 5 を満たす．評価基準の要素上限は，評価基準の強さの関係から導ける．k ランクのランク付集合ラベルのオブジェクトについても定理 2 が成り立つので，評価基準間に強さの関係があるのかは，要素条件の cnd_E，cnd_A，cnd_N ごとに判断でき，系列を構成する評価基準を容易に選択できる．

7. ランクがないデータへの対応

ランクの判断が未着手なオブジェクトやランクを付けられないと判断したオブジェクトがある場合，ラベルにランクがないオブジェクトがあるので性質 4 を満たさない．本章は，これまでの議論の拡張でこのようなデータにも対応できることを示す．

オブジェクト o の $L(o)$ のラベルに対しランクの判断が未着手な場合，ラベルにランクがないので性質 4 を満たさないが，そのようなオブジェクトを含むデータに対しても分析が行えることが望ましい．ランクを用いない分析の体系，すなわち，$L(o)$ 全体を対象とする評価基準 LE，LA，LN 及びそれらから導出される $LA \cdot LN$ を用いれば，ランクの処理が完了しても分析結果は変わらない．それらの評価基準の強さの関係は図 7 の通りで，LE から $LA \cdot LN$ までの 2 通りの系列となる．

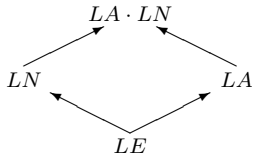

図 7　集合ラベル全体を対象とする評価基準の強さの関係

ランクの処理が進んでも，ラベルの適合度を表す値自体が分わからない場合，すなわちランクの判断の根拠となるものがないときにはランクを付けられない．そのようなオブジェクト o については，$L(o)$ 中のラベルでランクがないラベルを不明ラベル（Unranked Label）とする．また，$L(o)$ 中の不明ラベルの集合を $U(o)$ で表し，$L(o) = R_k^*(o) \cup U(o)$ とする．

ラベル集合とオブジェクトとの関連には，R_k 以上のいずれかのラベルで関連する場合と不明ラベルで関連する場合，すなわち明確な強さで関連する場合と不明確な強さで関連する場合がある．ラベル集合 $\mathcal{L}$ のいずれかのラベルに対し，評価基準 R_kE を満たすオブジェクトは R_k 以上のいずれかのラベルで関連し，評価基準 LE を満たし R_kE を満たさないオブジェクトは不明ラベルのみで関連する．明確な強さで $\mathcal{L}$ に関連するオブジェクトは，不明確な強さでしか $\mathcal{L}$ に関連しないオブジェクトよりも $\mathcal{L}$ との関連が強いと考え得るので，評価基準間の強さの関係に関する定義 2 より，$LE < R_kE$ である．同様に，$\mathcal{L}$ のすべてのラベルに対し，評価基準 R_kA を満たすオブジェクトは R_k 以上のいずれかのラベルで関連し，評価基準 LA を満たし R_kA を満たさないオブジェクトは不明ラベルでも関連する．$\mathcal{L}$ のすべてのラベルに明確な強さで関連するオブジェクトは，$\mathcal{L}$ のすべてのラベルに関連するが不明確な強さがあるオブジェクトよりも $\mathcal{L}$ との関連が強いと考え得るので，$LA < R_kA$ である．また，評価基準 LN を満たすオブジェクトは不明ラベルも含めて $\mathcal{L}$ と無関係なラベルに関連せず，評価基準 R_kN を満たし LN を満たさないオブジェクトは R_k 以上のラベルでは $\mathcal{L}$ と無関係なラベルに関連しない．$\mathcal{L}$ と無関係なラベルに関連しないオブジェクトは，明確な強さでは $\mathcal{L}$ と無関係なラベルに関連しないオブジェクトよりも $\mathcal{L}$ との関連が強いと考え得るので，$R_kN < LN$ である．よって，不明ラベルは最も関連が弱いランク R_k よりも関連が弱いランクのラベルと考えればよく，$k+1$ 個目のランク R_{k+1} を設けることで性質 4 を満たし，$R_{k+1}E < R_kE$，$R_{k+1}A < R_kA$，$R_kN < R_{k+1}N$ なので補題 6 が成り立つ．$R_{k+1}E < R_{k+1}A$ と $R_{k+1}E < R_{k+1}N$ は明らかなので補題 7 と補題 8 も成り立つ．また，不明ラベル以外，すなわち R_{k+1} 以外のランクのラベルであれば R_1 から R_k のいずれかのランクのラベルになるので性質 5 も満たし，これまでの分析体系が適用できる．最下位ランクよりも弱い $k+1$ 個目のランクの不明ラベルを設けることで，R_kE，R_kA，R_kN 以上の強さの評価基準を用いれば，不明確な強さでの関連を考慮しない分析ができる．$R_{k+1}E$，$R_{k+1}A$，$R_{k+1}N$ の評価基準を用いれば，不明確な強さでの関連を考慮した分析が可能であり，それらは LE，LA，LN を用いた分析体系である．

8. おわりに

本稿は，分析目的やデータ自体が多様で主ラベルやランクがないデータに対しても分析ができる理論的枠組みを提案した．分析対象となるラベル集合とデータの集合ラベルとの関連の強さを評価する基準には，存在性，網羅性，排他性の側面があり，評価基準は存在性上限，網羅性上限，排他性上限からなる要素上限で表せる．評価基準の要素上限を側面ごとに比較することで評価基準間の強さの関係を判断でき，評価基準間の強さの順序である系列を容易に導ける．系列により，関連の強さを一元

的かつ段階的に評価でき，異なる強さのデータ集合間を比較することで分析対象に関する有用な知見を得られる．

本稿では，ランクの区分を k 個に拡張することでより詳細な分析を可能にした．同時に，ランクが付けられないようなデータに対しては，最下位ランクよりも弱い $k+1$ 個目のランクの不明ラベルを設けることで対応できることを示した．

文　献

[1] Canuto, S., Goncalves, M., Santos, W., Rosa, T., and Martins, W.: An Efficient and Scalable MetaFeature-based Document Classification Approach based on Massively Parallel Computing, *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval* (SIGIR '15), pp. 333–342 (2015).

[2] Ding, B., Wang H., Jin, R., Han, J., and Wang, Z.: Optimizing Index for Taxonomy Keyword Search, *Proc. ACM SIGMOD Int'l Conf. on Management of Data* (SIGMOD '12), pp. 493–504 (2012).

[3] 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).

[4] Kuzunishi, M. and Furukawa, T.: Strength of Relationship Between Multi-labeled Data and Labels, *Proc. Information and Communication Technology - Third IFIP TC 5/8 Int'l Conf., ICT-EurAsia 2015, and 9th IFIP WG 8.9 Working Conf., CONFENIS 2015, Held as Part of WCC 2015*, pp. 99–108 (2015).

[5] Ren, Z., Peetz, M., Liang, S., Dolen, W., and Rijke, M.: Hierarchical Multi-Label Classification of Social Text Streams *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval* (SIGIR '14), pp. 213–222 (2014).

[6] Tang, B., Han, S., Yiu, M., Ding, R., and Zhan, D.: Extracting Top-K Insights from Multi-dimensional Data, *Proc. ACM SIGMOD Int'l Conf. on Management of Data* (SIGMOD '17), pp. 1509–1524 (2017).

[7] Wang, X., Bendersky, M., Metzler, D., and Najork, M.: Learning to Rank with Selection Bias in Personal Search, *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval* (SIGIR '14), pp. 115–124 (2014).

[8] Wang, H., He, X., Chang, M., Song, Y., White, R., and Chu, W.: Personalized Ranking Model Adaptation for Web Search, *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval* (SIGIR '13), pp. 323–332 (2013).

[9] Wasay, A., Wei, X., Dayan, N., and Idreos, S.: Data Canopy: Accelerating Exploratory Statistical Analysis, *Proc. ACM SIGMOD Int'l Conf. on Management of Data* (SIGMOD '17), pp. 557–572 (2017).

[10] Zhu, X., Song, S., Lian, X., Wang, J., and Zou, L.: Matching Heterogeneous Event Data, *Proc. ACM SIGMOD Int'l Conf. on Management of Data* (SIGMOD '14), pp. 1211–1222 (2014).