DEWS2004 I-3-01

クラスの意味に基づく分類階層の構成

葛西　正裕†　　古川　哲也††

† 九州大学大学院経済学府　　〒 812–8581 福岡市東区箱崎 6–19–1
†† 九州大学大学院経済学研究院　〒 812–8581 福岡市東区箱崎 6–19–1
E-mail: {kuzu,furukawa}@en.kyushu-u.ac.jp

あらまし　収集したデータを利用するには、データを分析し整理する必要があり、階層的な分類はデータ整理の方法として有用である．データの分類は、それに用いる性質ごとに階層的に木構造で行われることを仮定し、分類によって生成されたデータのクラスをグループとする．データが多様であれば、階層における複数のグループの意味を持つデータを含み、またデータの概念の粒度も異なる．本稿では、データが複数のグループに属することを許した分類階層を提案し、グループ間における集合演算を検討する．粒度が異なるデータに関しては、グループの意味が問題となる．グループは分類の概念を表すものと分類の範囲を表すものがあり、両方の意味で利用できる分類階層の構成法を提案する．このような分類階層の構成により、多様なデータを整理し、様々な利用に供することができる．
キーワード　分類階層, 情報検索, データウェアハウス

Constructing Classification Hierarchies based on Semantics of Classes

Masahiro KUZUNISHI† and Tetsuya FURUKAWA††

† Graduate School of Economics, Kyushu University
Hakozaki 6–19–1, Higashi-ku, Fukuoka, 812–8581 JAPAN
†† Faculty of Economics, Kyushu University
Hakozaki 6–19–1, Higashi-ku, Fukuoka, 812–8581 JAPAN
E-mail: {kuzu,furukawa}@en.kyushu-u.ac.jp

Abstract　Collected data have to be analized and arranged in order to utilize the data. Hierarchical classfication is useful for arranging data. Data is assumed to be classified in a tree structure for each characteristics of classification, and a class of classified data is called a group. There can be data with semantics of several groups and the guranuality of concepts of data is not uniform. This paper proposes a classification hierarchy which allows data to belong to two or more groups, and discusses set operations between groups. For different granuality data, groups have two semantics, the concepts for classification and the ranges of classification. A classification hierarchy to use classes for both meaning is given. This hierarchy allows us to arrange and use various data.
Key words　Classification Hierarchies, Information Retrival, Data Warehouse

1. はじめに

近年におけるコンピュータの高性能化、小型化、廉価化は著しいものがあり、ネットワーク技術の進歩と環境の整備によって、流通し蓄えられるデータ量は増加している．また、コンピュータはあらゆる分野に利用されているため、蓄えられるデータは質的にも数値、文章、画像、音声等多岐にわたる．そのためデータを蓄え、それを利用するための研究が行われている．

データベースのデータモデルは、データ間の関連を表現するための枠組みであると考えることができ、どのようなデータ（値）と関連があるのかを条件としてデータを検索する方法を与えている．明確に構造化できないデータに対しては、部分的に構造化してデータ操作を可能にする半構造データに関する研究がある．非構造データにおいては、テキストデータや画像データに対するパタンマッチによる検索や分類の自動化が図られている [1][2]．

これらはいずれも、データ自身の形式的な情報である値を利用しており、値の対応関係に基づくデータ利用の枠を出ていない．本研究は、非構造データを含むデータの集合に対し、利用者がその内容から索引をつける、つまりデータをあるクラスに分類することでデータを整理すると同時に、参照に必要とする

データの抽出といった検索だけでなく、データ全体の分析を可能にするためのクラスの構成法を与えることを目的としている．

データの整理の際には、それに用いる性質に従ってクラスを生成し、クラス集合が階層構造となる分類階層が構成される．階層的に分類することは、クラス間の概念的な関係を体系化でき、また利用者の求める概念のレベルでの検索やデータの分布による分析などが可能となる．例えば、実験で得られた膨大な化学反応式を階層的に分類することによる共通パターンの発見や、階層を巡回することによる反応メカニズムの解明といった分析を可能にしている [3]．

階層的にデータを分類することに関しては用語学の分野で発展したシソーラスがあり、それは集めた用語を整理し、統制された用語間で概念体系を構成したものである．また、一般的な概念の階層をコンピュータシステムで利用可能にした MRD (Machine Readable Dictionary) などの研究がある [4]．オントロジーに関する研究領域では、対象の概念を明確化し体系付けることで、対象をモデル化するガイドラインを与えることを目的とした研究がなされている [5]．これらは、用語や対象の概念を意味的に問題にしており、その機能や実世界での位置づけ等の観点から構造化を図っている．本研究では、これに対し、クラス間における概念の意味的な関係は、階層を構成する利用者が考慮することであると考え、クラスの意味からデータを整理する際のデータ構成を問題にしている [6]．

オントロジーは、クラス間の概念や対象そのものとクラスの関係に重点が置かれ、それを意味的に解釈しているため、人工知能や分類の自動化の分野に応用されている．オントロジーを基に階層構造を自動的に構築したり、特定領域に特化させるように階層の修正を行うシステムが提案されている [7][8]．また、自動化に関しては、データマイニング技術のクラスタリングとの連携により、手作業での整理の負担が軽減されている [9]．本研究は分類の自動化を目的にしているわけでなく、データの整理を階層的な分類の構造によって行うものであり、その構成に関してクラスの意味から考察をしているものである．

分類では、１つのデータは２つの異なるクラスには属しないとする排他性と上位概念に属すデータは下位概念のいずれかに属するとする充足性が一般に仮定される．このような階層は汎化による階層 [10] と同様の考えに基づくものである．しかし、データが多様であれば、複数の概念を扱っているため複数のクラスに属するものやデータの扱う概念の大きさ、すなわち粒度が異なるために下位概念に属さないものも存在する．このようなデータを含むデータ集合に対しては、充足性と排他性を満たす単純な分類階層を構成することはできない．

データの粒度が異なるとき、粒度の違いによる分類におけるクラスの意味の違いを考える必要がある．分類階層では、分類のクラスは一般にそのクラスに分類されたデータ集合を表す．このような解釈では、そのクラスに分類されたデータの粒度の違いを区別することができない．本稿では、これらの問題を解決するための分類階層の構成法を提案し、その利用について検討する．

本稿は以下のように構成される．２章で、粒度が等しく１つの概念のみに分類されるデータ集合に対する充足性と排他性を満たす分類階層の構成を示す．３章では、複数の概念を扱っていて複数クラスに分類されるデータの性質と扱いについて検討する．４章では、分類階層を利用するためにクラス間におけるデータ集合の演算について議論する．５章では、異なる粒度のデータが存在する場合の分類階層の構成法を提案する．６章は全体のまとめである．

2. オブジェクトの分類階層

データの分類はそれに用いる性質ごとに階層的に木構造で行うものとする．その特定した性質のことを属性と呼び、例えば地理的に県、市などのようにデータを階層的に分類した場合、その属性は地域であると考える．したがって同じデータであっても地域以外に業種や規模といった属性で複数の階層を構成することが可能である．データを属性ごとに階層を構成して複数の分類階層を用意することは有効であり [11]、その実現方法は、複数の階層を利用者の要求に応じて合成するビューとともに提案されている [12]．本稿ではある性質に限定した特定の属性における階層構造について議論する．

対象となるデータをオブジェクト o、分類によって生成されたオブジェクトのクラスをグループ g とする．グループ g を構成するオブジェクトの集合を $m(g) = \{o_1, o_2, \cdots, o_n\}$ で表す．グループ集合 $\boldsymbol{g}$ に対して、$m(\boldsymbol{g}) = \bigcup_{g \in \boldsymbol{g}} m(g)$ とする．$\boldsymbol{g} = \{g_1, g_2, \cdots, g_m\}$ が g の分類によるグループ集合であるとき、g を g_i $(\in \boldsymbol{g}, 1 \leqq i \leqq m)$ の親グループ、g_i を g の子グループという．また、階層のグループ g_i, g_j において、g_i は g_j の先祖である（g_j は g_i の子孫である）とき、$g_i \succ g_j$ で表す．g_i は階層の属性における g_j の上位概念であり、$m(g_i) \supseteqq m(g_j)$ である．

オブジェクトの概念の大きさをオブジェクトの粒度と呼ぶ．分類の属性において、オブジェクトの粒度が異なる場合がある．

［例 1］　地域経済の調査に関するデータであり、調査単位が九州である標本調査によって構成されるデータから福岡県に関することは調べることができない．このデータは、地域という属性で分類する場合、オブジェクトの概念の大きさ、即ち粒度は九州である．同様なデータであって調査単位が福岡県であれば、オブジェクトの粒度は福岡である．　□

また、分類の属性において、オブジェクトは１つのグループのみの意味を持つものと複数のグループの意味を持つものがある．前者を原子オブジェクト、後者を多義オブジェクトという．

［例 2］　福岡に関するデータ o_1、九州に関するデータ o_2 は、福岡や九州という分類のグループがあれば共に原子オブジェクトである．福岡と佐賀に関するデータ o_3 は、福岡・佐賀のグループが存在せず、福岡と佐賀のグループがある場合、２つのグループの意味を持つので、それは多義オブジェクトである．　□

一般にデータを階層的に分類するときには、親グループ g とその子グループ集合 $\boldsymbol{g}$ に対して２つの性質が仮定される．

（1） 充足性：親グループのオブジェクトは必ずいずれかの子グループに属する、すなわち $m(g) = m(\boldsymbol{g})$ である．

（2） 排他性：オブジェクトは子グループのいずれか 1 つのみに属する、すなわち $m(g_i) \cap m(g_j) = \phi\ (g_i, g_j \in \boldsymbol{g}, i \neq j)$ である．

これらの性質は、データの意味の粒度が同じであり、原子オブジェクトのみのオブジェクト集合を対象とした分類階層の場合に仮定されるものである．粒度が異なっているオブジェクト集合や多義オブジェクトを含むオブジェクト集合に対しては、これらの性質を満たす分類階層を構成するのは困難となる．

［例 3］ 親グループの意味が九州であり、子グループの意味が九州各県である分類階層を考える（図 1）．

- 九州全体に関するオブジェクト o_2 は親グループには属するが、いずれの子グループにも属さないために充足性を満たさない．
- 福岡と佐賀の両方に関するオブジェクト o_3 を、2 つの概念を扱っているために福岡、佐賀の両グループに属すとした場合、分類階層は排他性を満たさない． □

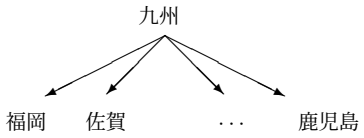

図 1 九州に関する分類階層

まず、粒度が等しい原子オブジェクトのみのデータに対する分類階層について議論する．そのようなデータを対象とした場合、原子オブジェクトであることから排他性は満たされる．また、適切なグループが存在しないオブジェクトについては、その他を意味するグループを設けることで子グループに含むことができ、充足性を満たす分類階層を構成できる．その他を意味するグループは、要素には親グループに属しているオブジェクト集合であるが子グループのいずれの意味にも該当しないオブジェクト集合である．例えば、九州の分類階層であれば、子グループとして福岡、佐賀、その他の県という子グループで構成される分類階層である．対象となるデータが分類階層の属性に対する粒度が等しい原子オブジェクト集合であれば明確な分類が可能である．

分類階層における全てのグループに対して、属するオブジェクトを記憶させておくことは、データ量やオブジェクトの挿入や削除を考えると適切ではない．分類階層は充足性を満たすので、子グループ集合の和集合で親グループのオブジェクトを得ることができる．分類階層では、子グループを再帰的に求めることになる．したがって、葉となるグループに対してのみオブジェクトを記憶させれば、充足性を利用して階層内全てのグループのオブジェクトを求めることができる．

3. 多義オブジェクトに対する分類階層

対象とするデータに多義オブジェクトが含まれる場合、分類階層における排他性が問題になる．多義オブジェクトへの対応は以下の 2 つが考えられる．

（1） 多義オブジェクトの持つ意味に対応するグループを作る．そのグループは、他のグループの意味を組合せた意味を持つ．

（2） オブジェクトが複数のグループに属することを許す．すなわち排他性を仮定しない．

意味の組合せのグループを用意することは、例 2 で考えれば、福岡・佐賀のデータは福岡と佐賀の 2 つのグループの意味を持つため多義オブジェクトであり、それを福岡・佐賀という組合せグループを置いた分類階層で対応する（図 2）．多義オブジェクトの種類が少なければ問題ないが、多い場合には問題である．親グループ g の意味を詳細化した基本となる意味が n 個あった場合に排他性を満たすべく子グループを準備するためにはその数は $O(2^n)$ となるため現実的ではない．利用時にも、1 つの意味に関するオブジェクト集合を求める際には、$O(2^{n-1})$ のグループの和集合により求めることになる．よってオブジェクトが複数のグループの要素になることを許すことで多義オブジェクトに対応する．

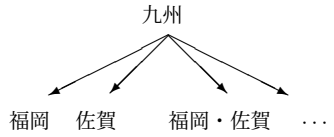

図 2 意味の組合せのグループを持つ分類階層

しかし、分類階層におけるオブジェクトとグループの関係を全て同等に扱うと利用時に不便なものとなる．

［例 4］ 地域経済に関するデータを地域という地理的な属性で分類する際、例 2 と同様に、九州という親グループ、福岡、佐賀といった子グループの分類階層を構成したとする．福岡の地域経済についての分析が中心で佐賀に関する記載がある多義オブジェクト o_4 と福岡と佐賀の経済比較といった対等に両者を分析した多義オブジェクト o_5 がある．2 つのオブジェクト o_4, o_5 を福岡、佐賀のグループに属するようにした場合、利用時に佐賀のグループを参照すると o_4, o_5 の区別が無く、佐賀を中心としたデータの要求に不都合である． □

オブジェクトが複数のグループに属することを許した分類階層では、オブジェクトとグループの関係には強弱があるので、分類階層はオブジェクトとグループの関係の強さを考慮したものとする．関係の強弱を区分することでグループにおけるオブジェクトの重要性を反映でき、利用目的に合ったデータの抽出が可能になる．関係の強さの段階が多いほどデータの利用時に高度な問合せが可能になる．しかし区分が多すぎると分類時における作業が煩雑で利便性に欠くことになり、区分の数が問題となる．本稿では、重要度の反映に最低限必要な 2 つの区分で検討する．2 つの区分における議論は 3 つ以上の区分にも容易に拡張できる．

［定義 1］　オブジェクト o がグループ g の意味を主として有しているときに、o は g の主オブジェクトであるという．オブジェクト o が g 以外のいずれかのグループの主オブジェクトであるが、g の意味を副次的に有しているときに、o は g の副オブジェクトであるという．　□

多義オブジェクトが分類されるべき複数のグループの中で、相対的に最も関係が強いグループにおいて、そのオブジェクトは主オブジェクトとなる．複数のグループに対して同等の関係の強さを持つとき、複数のグループで主オブジェクトとなる．したがって、全てのオブジェクトは少なくとも 1 つのグループの主オブジェクトとなる．

［例 5］　例 4 のオブジェクト o_4 は福岡が主となる内容であり、佐賀については福岡に関することとして副次的な記載があるだけなので、福岡のグループの主オブジェクト、佐賀のグループの副オブジェクトになる．o_5 は、福岡と佐賀の 2 つが主な内容であり対等に比較しているため、福岡と佐賀の両グループの主オブジェクトになる．　□

親グループと子グループ集合間の関係と主、副の 2 つの区分の性質について検討する．g の主オブジェクト集合、副オブジェクト集合をそれぞれ $m_M(g)$、$m_S(g)$ で表し、グループ集合 $\boldsymbol{g}$ に対して $m_M(\boldsymbol{g})$、$m_S(\boldsymbol{g})$ で表す．

グループに属する強さの区分の定義より、グループの主オブジェクトは同一グループの副オブジェクトにはなり得ない．

［性質 1］　グループ g のオブジェクト集合 $m(g)$ は $m_M(g) \cup m_S(g)$ であり、オブジェクトは主、副のいずれか、すなわち $m_M(g) \cap m_S(g) = \phi$ である．　□

子グループでの区分は親グループでの区分と関係する．親グループの主オブジェクトであることは、親グループの意味を主として有しているオブジェクトなので、少なくともいずれかの子グループの主オブジェクトである．親グループの副オブジェクトであることは、親グループの意味を副次的に有しているオブジェクトなので、いずれかの子グループの副オブジェクトである．すなわち、オブジェクトの親グループにおける区分は、少なくとも 1 つの子グループに継承される．また、オブジェクトの子グループにおける区分は、上位概念である親グループの区分より強くなることはない．例えば、他に主とするものがあり九州にも関連して副オブジェクトであるものは、福岡を主として扱っていることはない．

［性質 2］　親グループの主オブジェクトは、子グループ集合のいずれかのグループの主オブジェクトであり、他の子グループの主オブジェクト、副オブジェクトに成り得る．同様に、親グループの副オブジェクトは、子グループ集合のいずれかのグループの副オブジェクトであり、他の子グループの副オブジェクトに成り得る．　□

すべてが原子オブジェクトであるときと同様に分類階層における全てのグループに対して、2 つの区分別に属するオブジェクトを記憶しておくことは適切でない．性質 2 を用いれば、子グループ集合から親グループのオブジェクトを区分別に求めることができる．したがって、多義オブジェクトに対しても原子オブジェクトと同様に、葉のグループのオブジェクトから階層内のすべてのグループのオブジェクトを得ることができる．

［定理 1］　親グループ g と子グループ集合 $\boldsymbol{g} = \{g_1, g_2, \cdots, g_m\}$ に対して次の性質が成り立つ．

（1）　$m_M(g) = m_M(\boldsymbol{g})$

（2）　$m_S(g) = m_S(\boldsymbol{g}) - m_M(\boldsymbol{g})$　□

（証明）　親グループの主オブジェクトは 1 つ以上の子グループの主オブジェクトであり、子グループの主オブジェクトには、親グループで副オブジェクトになっているものはない．したがって、子グループ集合の主オブジェクトの和集合は親グループの主オブジェクト集合に一致する．親グループの副オブジェクトも 1 つ以上の子グループの副オブジェクトであるが、親グループの主オブジェクトで子グループでは副オブジェクトとなるものがある．逆に子グループの主オブジェクトで親グループでは副オブジェクトになるものはない．したがって、子グループ集合の副オブジェクトの和集合から子グループの主オブジェクトを除けば親グループの副オブジェクト集合となる．（証明終）

4. 多義オブジェクトの利用

葉となるグループのオブジェクトの和集合により、階層内のすべてのグループのオブジェクトを求めることができる．和集合演算で重複を除いたり差集合演算で同一オブジェクトを求めることは、多大なコストを要する．多義オブジェクトの性質を用いれば、親グループのオブジェクトを 2 つの区分で容易に求める方法を導入することができる．

親グループのオブジェクトは必ず子グループのオブジェクトとなっているので、その 1 つで親グループのオブジェクトであることを代表させる．そのようなオブジェクトを代表オブジェクトという．代表オブジェクトは、親グループと同じ区分になっているものとし、そのようなオブジェクトは必ず存在する．オブジェクト集合 $\boldsymbol{o}$ に対し、$\boldsymbol{o}$ 内の代表オブジェクトの集合を $r(\boldsymbol{o})$ とする．

［定理 2］　親グループ g と子グループ集合 $\boldsymbol{g} = \{g_1, g_2, \cdots, g_m\}$ に対して次の性質が成り立つ．

（1）　$m_M(g) = r(m_M(\boldsymbol{g}))$

（2）　$m_S(g) = r(m_S(\boldsymbol{g}))$　□

（証明）　親グループの各オブジェクトに対して、同じ区分の子グループの代表オブジェクトがただ 1 つ存在することから明らか．　（証明終）

代表オブジェクトは、子グループ集合のオブジェクトから親グループのオブジェクトを求めるためだけに利用するものであり、多義オブジェクトでも条件を満たすものであればどれでもよい．利用には無関係で利用者が意識することはないため、親

グループのオブジェクトの分類の際にシステムが自動的に指定することも可能である．

利用者はグループ g を指定することにより、主、副の 2 つに区分された g のオブジェクトを得ることができる．さらに、グループ間のオブジェクトの集合演算により、利用目的に合ったオブジェクトを得ることが可能になる．次に、集合演算に対して 2 つの区分をどのように扱えばよいかについて検討する．演算の対象となるグループを g_a, g_b とする．

積集合は最も基本的な演算であり、2 つのグループの意味を持っているオブジェクトの集合を求めるものである．演算結果の主オブジェクトを g_a と g_b において共に主であるオブジェクトとし、副オブジェクトは、g_a および g_b のオブジェクトで、積集合の主オブジェクト以外のものとする．

（1） $m_M(g_a \cap g_b) = m_M(g_a) \cap m_M(g_b)$

（2） $$\begin{aligned} m_S(g_a \cap g_b) &= (m(g_a) \cap m(g_a)) - (m_M(g_a) \cap m_M(g_b)) \\ &= (m_M(g_a) \cap m_S(g_b)) \\ &\quad \cup (m_S(g_a) \cap m_M(g_b)) \\ &\quad \cup (m_S(g_a) \cap m_S(g_b)) \end{aligned}$$

和集合の主オブジェクトは、少なくともいずれか一方のグループで主となるオブジェクトとする．副オブジェクトは、少なくとも一方のグループで副となり、和集合の主ではないオブジェクトである．

（1） $m_M(g_a \cup g_b) = m_M(g_a) \cup m_M(g_b)$

（2） $$\begin{aligned} m_S(g_a \cup g_b) &= (m_S(g_a) \cup m_S(g_b)) \\ &\quad - (m_M(g_a) \cup m_M(g_b)) \\ &= (m_S(g_a) - m_M(g_b)) \\ &\quad \cup (m_S(g_b) - m_M(g_a)) \end{aligned}$$

差集合は、ある性質以外のオブジェクトを求める際に用いる．差集合の主オブジェクトは、g_a の主オブジェクトで g_b ではないものとなる．“g_b でない” の考え方は様々あるが、ここでは g_b の主オブジェクトではないとし、g_a の主オブジェクトで g_b の副オブジェクトであるものも差集合の主オブジェクトであると考える．“g_b でない” を副オブジェクトでもないとした場合も軽微な修正で以下の議論は成り立つ．実現の際に差集合演算のオプションとすることで、様々な差集合演算が可能となる．副オブジェクトについても同様である．

（1） $m_M(g_a - g_b) = m_M(g_a) - m_M(g_b)$

（2） $m_S(g_a - g_b) = m_S(g_a) - m_M(g_b)$

これらの演算結果のオブジェクト集合はグループのオブジェクト集合と同じデータ構造であり、性質 1 を満たす．したがって、演算結果に対しても更なる集合演算を適用することが可能であり、演算の組合せは個々の演算を再帰的に適用することによって計算できる．

［例 6］　福岡に分類される文献のグループと佐賀に分類される文献のグループ間の集合演算を考える．

（1） 積集合は、福岡と佐賀の両方に関する文献のグループを生成する．その主オブジェクトは両方が中心的に述べられているもの、副オブジェクトは両方の記述があるものである．

（2） 和集合の演算結果の主オブジェクトは少なくとも一方が中心的に述べられているものであり、副オブジェクトは少なくとも一方の記述があるものである．

（3） 差集合で得られるものは福岡に分類される文献で佐賀には関係しないものである．

（4） (福岡 ∪ 佐賀) ∩ 北海道 という演算では、福岡か佐賀のどちらかと同時に北海道に関する文献が求められる．主オブジェクトは、福岡または佐賀と北海道の両者が中心的に述べられている文献である．副オブジェクトは、福岡または佐賀と北海道の両者についての記載があるものである．　□

5. 異なる粒度のデータへの対応

オブジェクトの粒度が異なるとき、粒度が大きいオブジェクトは下位概念では分類できないという問題が生じ、これまでの議論は適用できない．本章では、異なる粒度のオブジェクトの扱いについて議論する．

オブジェクトの粒度がグループの概念と一致するとき、グループの更なる分類では、そのオブジェクトを子グループに属させることができない．例 2 の o_2 は九州のオブジェクトであり、親グループの九州という概念の粒度と一致しているので、それは子グループの概念の粒度からするといずれのオブジェクトでもない．

充足性を満たす分類階層を構成するためには、最大粒度のオブジェクトに分類の粒度を一致させることになるので現実的でない．すなわち、オブジェクトの粒度が異なるとき、充足性を満たす分類階層を構成できない．

また、データの利用の際には、グループ g に属するオブジェクトの意味を、g の粒度のオブジェクトとして用いたい場合もある．例 2 で、o_2 は九州のグループに属するオブジェクトとして求めたいが、福岡である o_1 は含めたくない場合などである．すなわち、グループに属するオブジェクトの意味は、

（1） そのグループに分類されたオブジェクト

（2） そのグループに関するオブジェクト

の 2 つがある．(1) はこれまでの議論で用いてきた解釈で、$g_i \succ g_j$ ならば $m(g_i) \supseteq m(g_j)$ を意味する．すなわち、グループは分類の範囲を表している．一方、(2) の解釈では、$g_i \neq g_j$ ならば $m(g_i) \cap m(g_j) = \phi$ となり、グループは概念自体を表していると考えることができる．オブジェクトの粒度が同じであるとき、グループの解釈は (1) で問題ないが、異なるときには、これら両方の解釈で分類階層を利用できるようにする必要がある．

これらの問題を解決するため、グループ g に、g に関するオブジェクトを要素とする子グループ g_o を加える．g_o は子グループを持たず、階層の葉となるグループである．g の子グループは、g よりも下位概念のオブジェクトを分類したグループと g_o から成る．

この階層は充足性を満足し、グループの解釈はそのグループに分類されたオブジェクトの集合である．グループに関するオブジェクトは g_o で求めることができる（図 3）．

粒度の異なるオブジェクトに対しても多義オブジェクトは存

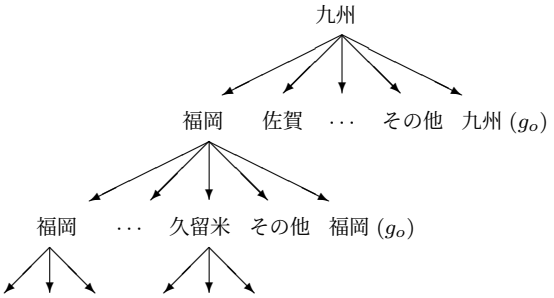

図 3　2 つの解釈で利用可能な分類階層

在する．3、4 章で議論したオブジェクトが複数のグループの要素となる分類階層は、粒度が異なるオブジェクトに対しても適用できる．したがって、分類の範囲を表すグループと分類の概念を表すグループ間や分類の概念を表すグループ間での集合演算が可能になる．

また、子孫と先祖のグループに対する多義オブジェクトも分類階層は対応する．上位概念を持つオブジェクトと下位概念を持つオブジェクト間の集合演算はこれまでの議論を適用できる．

［例 7］　福岡県の文献のグループと福岡市の文献のグループ間の集合演算を考える．

（1） 福岡県に関する文献のグループと福岡市に分類される文献のグループの積集合は、福岡県に関する記述と福岡市に分類される概念の記述の両方を含むものとなる．上下関係の概念にあるグループに対して、両方とも範囲を表すグループ、すなわち福岡県に分類される文献のグループと福岡市に分類される文献のグループの積集合は、福岡市に分類される文献のグループであり意味がない．

（2） 福岡県に関する文献のグループと福岡市に分類される文献のグループの和集合は、福岡県に分類された文献から福岡県に関する文献と福岡市に分類される文献以外の文献を除くことになる．福岡県に分類される文献のグループと福岡市に分類される文献のグループの和集合は、福岡県に分類される文献のグループであり意味がない．

（3） 福岡県に分類される文献のグループと福岡市に分類される文献のグループの差集合により、福岡市に分類される文献が除かれる．福岡市に関する文献のグループから、福岡県に分類される文献のグループの差集合は、空集合になり無意味である．　□

分類の粒度によっては、いずれのグループの概念にも一致しないオブジェクトも存在し得る．例えば、北部九州に関するオブジェクトの粒度は、九州の下位概念であるが、福岡、佐賀などには分類できない．このオブジェクトが、例えば統計データの集計値のように福岡や佐賀に直接関する部分を持たなければ、多義オブジェクトとして複数のグループに属させることは不適切である．このようなオブジェクトは、子グループのその他を意味する子グループに分類されることになる．したがって、データ利用時に、あるグループの上位概念に関するオブジェクトも求めたいときは、祖先のグループに関するオブジェクトだけでなく祖先のグループの分類におけるその他を意味するグループにもそのようなオブジェクトが存在し得ることに注意する必要がある．

データを分類する際は、グループが分類の概念を表すものとしてオブジェクトをグループに属させていくが、分類階層の構成は、グループが分類の範囲を表す階層になる．利用時には、親グループ g に関するオブジェクトを要素とする子グループ g_o により、グループを両方の意味で利用できる．

6. むすび

多義オブジェクトを含み、粒度が異なるデータに対する分類階層の構成について議論した．多義オブジェクトに関しては、オブジェクトが複数のグループに属することを許した分類階層を提案し、グループ間における集合演算を検討した．粒度が異なるデータに関しては、グループの意味が問題となる．グループには、分類の概念を表すものと分類の範囲を表すものがある．親グループに関するオブジェクトを要素とする子グループと子グループのいずれにも属さない概念のオブジェクトを要素とするその他を意味するグループを置くことで対応する．これにより、分類階層の利用時に両方の意味でグループを利用することができる．

このような分類階層の構成により、多様なデータを整理し、様々な利用に供することができる．本論文で議論した階層は木構造を仮定しているが、概念の上下関係を考慮すると、木構造で表せない場合もある．子グループが親グループを複数持つような構成になるが、それに対しても本稿での議論は容易に適用可能である．しかし、そのような分類階層では、分類作業は複雑なものとなる．利便性を高めるようなユーザインタフェースをどのように構成するかが問題となる．

謝辞　本研究の一部は文部科学省科学研究費補助金基盤研究 (C) (2)（課題番号 15500072）の支援を受けている．

文　献

[1] Cheng Lu and Mark S. Drew, “Construction of a Hierarchical Classifier Schema Using a Combination of Text-Based and Image-Based Approaches,” *ACM SIGIR Conf.*, pp. 438–439, New Orleans, USA, Sept. 2001.

[2] Sharon McDonald, Ting-Sheng Lai, and John Tait, “Evaluating a Content Based Image Retrival System,” *ACM SIGIR Conf.*, pp. 232–240, New Orleans, USA, Sept. 2001.

[3] J. Royce Rose and Johann Gasteiger, “Hierarchical Classification as an Aid to Database and Hit-List Browsing,” *Proc. Conf. on Information and Knowledge Management* (CIKM ’94), pp. 408–414, Gaithersburg, USA, Nov. 1994.

[4] Robert Krovetz, “Sence-Linking in a Machine Readable Dictionary,” *Proc. 30th Annual Meeting of the Association for Computational Linguistics*, Newark, USA, pp. 330–332, 1992.

[5] Thomas R. Gruber, “Toward Principles for the Design of Ontologies Used for Knowledge Sharing?” *Int’l Journal of Human-Computer Studies*, Vol. 43, Issues 5–6, pp. 907–928, Nov. 1995.

[6] Tetsuya Furukawa and Masahiro Kuzunishi, “Classification

and Utilization of Data Belonging to Mutiple Classes," *The 8th World Multiconference on Systemics, Cybernetics and Informatics* (SCI '04), Orland, USA, July 2004 (to appear).

[7] Chizuru Aoki, Reiko Sekiuchi, Masaki Kurematsu, and Takahira Yamaguchi, "DODDLE: A Domain Ontology Rapid Development Enviroment," *Proc. 5th Pacific Rim Int'l Conf. on Artificial Intelligence*, pp. 194–204, Singapore Nov. 1998.

[8] Yusuke Itoh and Makoto Haraguchi, "Conceputual Classification Guided by a Concept," *Proc. 11th Int'l Conf. on Algorithmic Learning Theory*, pp. 166–78, Sydney, Australia, Dec. 2000.

[9] Enrico Giacoletto and Karl Aberer, "Automatic Expansion of Manual Email Classifications Based on Text Analysis," *CoopIS/DOA/ODBASE2003*, LNCS, Vol. 2888, pp. 785–802, Catania, Italy, Nov. 2003.

[10] John Miles Smith and Diane C. P. Smith, "Database Abstraction: Aggeregation and Generalization," *ACM Trans. Database Systems*, Vol. 2, No. 2, pp. 105–133, June 1977.

[11] Tetsuya Furukawa, "Multiple Classfication Hieratchies in Cooperative Databases," *Advanced Database Syst. for Integration of Media and User Environments '98*, Advanced Database Reseach and Development Ser., Vol. 9, pp. 309–314, 1998.

[12] 河野弘史, 王磊, 古川哲也, "進化する多重階層索引の実現方法," 情報基盤センター年報 2 号, 九州大学情報基盤センター, pp. 81–87, March 2002.