

階層的分類における複数の意味を持つデータの利用

葛西 正裕[†] 古川 哲也^{††}

大量で多様なデータを効率的に利用するためにはデータを適切に構成する必要があり、データの意味に基づく階層的な分類はデータの構成法として有用である。収集されたデータは多様であるため分類に用いる属性に対して複数の意味を持つ場合が多い。データ（オブジェクト）が複数の子クラスに分類されるとオブジェクトは終端クラスで重複するため、クラスオブジェクトを終端クラスから求める際に重複を除く必要がある。また、クラス集合の意味に一致するオブジェクトを取り出すためには、オブジェクトがクラス集合の終端クラス以外の終端クラスに分類されていないかを調べる必要がある。本論文では、複数の子クラスに分類されたオブジェクトのうち1つで子クラス集合のオブジェクトを代表させることによってこれらの問題を解決できることを示す。

Utilization of Multiple Semantics Data with Hierarchical Classification

MASAHIRO KUZUNISHI[†] and TETSUYA FURUKAWA^{††}

A lot of heterogeneous data must be organized properly to be utilized efficiently, and hierarchical classification based on semantics of data is useful approach to organize data. The semantics of data for the classification attribute is often multiple. When data (objects) are classified to more than one child class, they are duplicated in the terminal classes, that requires the elimination of the duplications when the objects of a class are produced from the terminal classes. When the objects whose semantics is the semantics of a set of classes are required, we have to check that objects are classified to the other terminal classes. This paper shows these problems are handled by selecting one of the objects in the child classes to represent the object in the child classes.

1. はじめに

情報技術の発展とネットワーク環境の整備によって利用可能なデータは著しく増加し、同時にデータは数値、テキスト、画像、音声といった多様なものとなっている。データ量の増大やデータの多様性が増すとともにデータをいかにして一元的に管理するかが重要になってくる。データの種類を問わずに効率的に管理するデータの構成法として、データの意味に応じて階層的に分類しておくことが有用である^{2),8),11)}。たとえば、情報検索で用いられるカテゴリ索引などは、収集したデータをその意味に応じて階層的に分類することで管理し、必要なデータを効率的に取り出すことを可能にしている。

階層的分類を用いたデータの構成に関する研究の多くは、データが1つの終端クラスに属する分類階層を

想定している^{2),7),14)}。分類に用いる属性に対して意味が1つであるようなデータに対する分類階層は比較的簡単なものであり、クラス間の階層関係と終端クラスのデータを記憶しておいて、非終端クラスのデータはそのクラスの終端クラスにおけるデータの直和で求めるなどの方法が用いられる。しかし、収集されたデータは多様であるため、ある属性で分類を行うと複数の意味を持つことが多く、たとえば、世界経済を扱ったデータを地域で分類する場合、日本と中国に関する経済比較といった文献は{日本, 中国}という意味を持つ。複数の意味を持つデータに対する分類階層は、データが複数の子クラスに分類されることになり、次のような問題が生じる。

- (1) 複数のクラスに分類されるためデータが重複する。
- (2) クラスにはクラスの意味とは関係のない意味を持つデータが分類されている。

データが重複しているためにクラスの終端クラスからそのクラスに分類されるデータを取り出す際に重複を除く必要がある⁸⁾。データは複数の意味を持ち複

[†] 九州大学大学院経済学府
Graduate School of Economics, Kyushu University

^{††} 九州大学大学院経済学研究院
Faculty of Economics, Kyushu University

数の子クラスに分類されているので、クラスの積集合による問合せが行われる。しかし、クラスにはそのクラスの意味に対して同位または下位でない意味を含む意味のデータが分類されるため、集合演算によって得られたデータには演算に用いたクラス集合のいずれの意味に対しても同位または下位でない意味を含む意味のデータが含まれる。たとえば、クラス *JAPAN* と *CHINA* の積集合には、日本と中国のいずれに対しても同位または下位ではないパリの意味を含む { 東京, 北京, パリ } のデータが含まれる。データが複数の子クラスに分類される分類階層において特定の意味集合のみに関するデータを取り出すためには、データが他の終端クラスに分類されていないことを調べなければならない。すなわち、そのようなデータを求めるためにはすべてのデータを調べる必要がある。本論文は、子クラスに分類されたデータのうちから1つを代表に選び、代表に選ばれなかったクラスの深さや代表に選ばれなかった情報を用いることでこれらの問題が解決できることを示す。

近年の分類に関する研究は、主に大量のデータに対する分類の自動化を図るものである。統計学や機械学習の手法が分類の自動化に用いられており、それらはデータを事前に定義されたクラスへ自動的に分類するものである^{9),15)}。また、データマイニングにおけるクラスタリング¹⁶⁾ や決定木、回帰木を用いたクラシフィケーション¹⁾ は分類を用いて大量のデータから自動的に有益な情報を抽出する研究である。これらの研究は主に分類の自動化における正確さや高速化に関するもので、分類階層の構造に関しては議論が及んでいない。一方で、意味に基づく階層的な分類の研究として、オントロジに基づく分類階層が知られているが¹²⁾、研究の目的はクラス間の意味的な記述を行うことで分類の自動化を実現するための分類階層をあらかじめ構成しておくことにある¹⁰⁾。また、文献3),11)などはWebやテキストといった多様なデータを複数の子クラスに分類しているが、分類階層へのデータの自動的な分類を目的としており、データが複数の子クラスに分類されることによる問題については触れられていない。そのような問題に対して文献8)では、子クラス集合におけるデータのうちの1つで子クラス集合のデータを代表させることによりデータの重複に対応している。また、最初にデータが重複したクラスの深さを利用することでクラスの意味に一致するデータを取り出せることを示しているが、その情報を利用してもクラスの集合演算には対応できない。本論文は、代表に選ばれなかった情報を用いることで、クラスの集合演算で得

られたデータのうちクラス集合の意味に一致するデータを他の終端クラスを調べることなく取り出せることを示す。

本論文は、以下のように構成される。2章で基本的な分類階層の性質を述べる。3章では複数の意味を持つデータに対する分類階層について議論し、データが複数のクラスに分類されることによって生じる2つの問題を明らかにする。4章において、終端クラスにおけるデータの重複によって生じる問題は、子クラス集合におけるデータのうちの1つで子クラス集合のデータを代表させることによって解決できることを示す。5章では、代表に選ばれなかった情報を用いることでクラス集合の意味に一致するデータをそのクラス集合の終端クラスのみから求めることができることを示す。6章は結論と今後の課題である。

2. 分類階層の性質

データの分類はそれに用いる属性ごとに階層的に行われるものとする。たとえば、地域という属性で、国、都市といった分類が行われる。複数の属性に対する分類階層は文献4),11)などで議論されており、本論文では特定の属性に対する分類階層を対象とする。

分類の対象となるデータをオブジェクト o 、オブジェクトの分類によって作られるオブジェクト集合をクラス C とし、クラス C の要素を $m(C) = \{o_1, o_2, \dots, o_m\}$ で表す。クラス集合 C に対して、 $m(C) = \bigcup_{C \in C} m(C)$ とする。 $C = \{C_1, C_2, \dots, C_n\}$ が C の分類によるクラス集合であるとき、 C は C_i の親クラスであり、 C_i は C の子クラスである ($1 \leq i \leq n$) という。子クラスのオブジェクトをさらに分類することで、クラス集合は階層構造となり、分類によって生成される階層的なクラス集合を分類階層とよぶ。

オブジェクトがクラスの要素であるかどうかはオブジェクトとクラスの意味によって決定される。オブジェクト o とクラス C の意味をそれぞれ $S(o)$ と $S(C)$ で表し、 $S(C) = \{S(C) \mid C \in C\}$ とする。2つの意味 S_i と S_j に対して、 $S_i \prec S_j$ で S_j は S_i の上位概念 (S_i は S_j の下位概念) であることを示す。クラス C とオブジェクト o に対して、 $S(o) \preceq S(C)$ ならば、 o は C の要素である。よって $S(C_i) \prec S(C_j)$ ならば $m(C_i) \subseteq m(C_j)$ であり、分類階層において C_j は C_i の先祖 (C_i は C_j の子孫) である。分類階層で、子クラスを持たないクラスを終端クラスという。

分類階層に関する研究は一般に以下の性質が仮定される。

- 充足性：親クラスのオブジェクトはいずれかの子クラスに属する。
- 排他性：親クラスのオブジェクトは2つ以上の子クラスに属することはない。

$S(C)$ を $S(C)$ の下位概念で $S(o) (o \in m(C))$ の上位概念であるような意味の集合 $S = \{S_1, S_2, \dots, S_n\}$ に分割したときの C の子クラス集合を $C = \{C_1, C_2, \dots, C_n\} (S(C_i) = S_i (1 \leq i \leq n))$ とする。 $S(C)$ が $S(C)$ を被覆するとき、 C の任意のオブジェクト o に対して、 $S(o) \leq S(C_i)$ となるような子クラス C_i が必ず存在するので充足性は満たされる。分類階層が充足性を満たすとき、クラス間の階層関係と終端クラスに属するオブジェクトのみを記憶しておけば、非終端クラスのオブジェクトはそのクラスの終端クラスから求めることができる。充足性を満たせば非終端クラスはそのクラスのオブジェクトを記憶する必要がなく、オブジェクトの挿入や削除といった更新が効率的に行える。

定義 1 クラス C の子孫の終端クラスを C の基本クラスといい、その集合を $B(C)$ で表す。 □

補題 1 分類階層が充足性を満たせば、クラス C に対して、 $m(C) = m(B(C))$ である。 □

(証明) 分類階層が充足性を満たせばクラス C のオブジェクトはいずれかの子クラスに属するので、 C のオブジェクトは子孫のいずれかの終端クラスに属し、 $m(C) \subseteq m(B(C))$ である。また、 C の基本クラスは C の子孫なので、 $m(B(C)) \subseteq m(C)$ であり、 $m(C) = m(B(C))$ である。 (証明終)

意味 S_i と $S_j (i \neq j)$ に対して、 $S \prec S_i$ かつ $S \prec S_j$ となるような意味 S が存在しないとき、 S_i と S_j は意味的に排他である。分類階層における各クラスの子クラス集合 C の任意のクラス C_i と $C_j (C_i, C_j \in C, i \neq j)$ に対して、それらの意味 $S(C_i)$ と $S(C_j)$ が意味的に排他であれば、分類階層は排他性を満たし、そのような階層は木になる。

補題 2 分類階層が排他性を満たせば、クラス C の任意の基本クラス $C_i, C_j (C_i, C_j \in B(C), i \neq j)$ に対して $m(C_i) \cap m(C_j) = \phi$ である。 □

(証明) クラス C と C のオブジェクト o に対して、分類階層が排他性を満たせば、 o は C の1つの子クラスのみ分類される。したがって、 o が C の基本クラスで複数の子クラスに属することはない。 (証明終)

補題 2 より充足性を満たす分類階層が排他性も満たすとき、クラス C のオブジェクト $m(C)$ は、 C の基本クラスのオブジェクトの直和であり、和集合で重

複なく求めることができる。本論文は、複数の意味を持つオブジェクトに対する排他性を満たさない分類階層について議論する。

3. 複数の意味を持つオブジェクトの分類

オブジェクトの分類は、一般に分類に用いる属性に対する代表的な1つの意味を用いて排他性を満たすように分類が行われる。複数の意味を持つオブジェクトはそれぞれの意味に対して分類を行うことで、それぞれの意味でそのオブジェクトを求めることができる^{3),11)}。たとえば、地域という属性に対して {東京, 北京} の意味のオブジェクトは、クラス *JAPAN* と *CHINA* の両方に分類することで、日本と中国のどちらの意味で求めてもそのオブジェクトを得ることができる。このような分類ではオブジェクトは複数の子クラスに分類されるので、分類階層は排他性を満たさない。本章は複数の意味のオブジェクトが複数の子クラスに分類されるときに生じる問題点を明らかにする。

オブジェクトをある属性で分類する際、オブジェクトを複数のクラスに分類するとき、オブジェクトの意味は複数になるので $S(o)$ は意味の集合となる。たとえば、東京と北京に関する文献を地域という属性で分類する場合、そのオブジェクトの意味は {東京, 北京} である。複数の意味のオブジェクトに対する分類階層では、オブジェクトの複数の意味における各々の意味がクラスの意味に対して同位または下位であれば、オブジェクトはそのクラスに分類される。

定義 2 オブジェクト o とクラス C に対して、 $S_i \leq S(C)$ となる S_i が $S(o)$ に存在するとき、 o は C の要素である。 □

複数の意味を持つオブジェクトが複数の子クラスに分類されるとき、要求する意味集合のオブジェクトはクラスの和集合や積集合によって求めることになるが、次の問題が生じる。

- (1) オブジェクトの重複：オブジェクトがクラス C の子孫の兄弟クラスで複数のクラスに分類されたとき、 $B(C)$ に $m(C_i) \cap m(C_j) \neq \phi$ となるクラス C_i, C_j が存在する。すなわち、 C のオブジェクトが $B(C)$ で重複する。
- (2) 拡張されたクラスの要素：クラス C には $S_i \not\leq S(C)$ となる S_i を含む意味のオブジェクトが含まれる。よって、クラスの和集合、積集合には、それらのクラス集合のいずれの意味に対しても同位または下位でない意味を含む意味のオブジェクトが含まれる。

例 1 図 1 は分類階層とオブジェクトの例である。オ

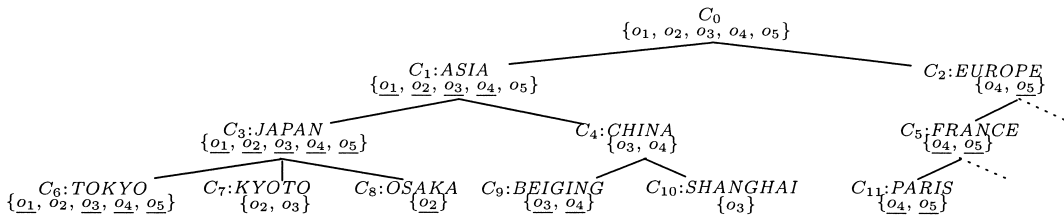


図 1 複数の意味を持つオブジェクトの分類

Fig. 1 Classification for multiple semantics objects.

ブジェクト集合 $\{o_1, o_2, o_3, o_4, o_5\}$ において, $S(o_1) = \{\text{東京}\}$, $S(o_2) = \{\text{東京, 京都, 大阪}\}$, $S(o_3) = \{\text{東京, 京都, 北京, 上海}\}$, $S(o_4) = \{\text{東京, 北京, パリ}\}$, $S(o_5) = \{\text{東京, パリ}\}$ であり, 図の $\{\}$ はクラスに分類されているオブジェクトを表している. 分類階層で実際にオブジェクトが分類されて記憶されるのは終端クラス $C_6, C_7, C_8, C_9, C_{10}, C_{11}$ のオブジェクトである. o_2, o_3, o_4, o_5 は複数の意味を持っており, 複数の子クラスに分類されるために次のような問題が生じる.

- (1) クラス *ASIA* のオブジェクトを求める際, *ASIA* の基本クラスである $C_6, C_7, C_8, C_9, C_{10}$ には o_2, o_3, o_4 が重複して存在する.
- (2) クラス *JAPAN* のオブジェクトには, 日本の同位または下位の意味ではない北京やパリを含む $\{\text{東京, 京都, 北京, 上海}\}$ の o_3 , $\{\text{東京, 北京, パリ}\}$ の o_4 , $\{\text{東京, パリ}\}$ の o_5 が含まれる. よって, クラス *JAPAN* と *CHINA* の和集合には, 日本と中国のいずれに対しても同位または下位ではないパリの意味を含む $\{\text{東京, 北京, パリ}\}$ の o_4 や $\{\text{東京, パリ}\}$ の o_5 が含まれる. *JAPAN* と *CHINA* の積集合にも $\{\text{東京, 北京, パリ}\}$ の o_4 が含まれる. □

クラスの意味が意味集合となるように分類階層を拡張すれば, 複数の意味を持つオブジェクトに対しても排他性を満たすことは可能であるが, n 個の意味に対してクラス数は $O(2^n)$ となり現実的ではない.

オブジェクトは複数の意味を持つので, 複数の意味による問合せが必要になる. 複数の意味による問合せには, 意味集合のいずれかの意味に対する同位または下位の意味を含む意味のオブジェクトを要求するものや意味集合すべての同位または下位の意味を持つオブジェクトを要求するものがある. 例 1 において, 日本または中国のいずれかに関するオブジェクトを求める場合や日本と中国の両方に関するオブジェクトを求める場合である. 分類階層においてクラスは意味を表しているの, 複数の意味による問合せに

対するオブジェクト集合はクラスの和集合や積集合によって容易に求められる. クラス集合 C に対して, $\exists S_1 \in S(o), \exists S_2 \in S(C), S_1 \preceq S_2$ となる o の集合は $\bigcup_{C \in \mathcal{C}} m(C)$ で, $\forall S_2 \in S(C), \exists S_1 \in S(o), S_1 \preceq S_2$ となる o の集合は $\bigcap_{C \in \mathcal{C}} m(C)$ で求まる.

クラスの集合演算で求められたオブジェクトには, 問合せには無関係な意味を含むものもある. 複数の意味での問合せでは, 求める意味集合以外の意味を含まないオブジェクトを要求する場合がある. たとえば, 日本と中国の両国間における関係の重要性を調べるために日本と中国のみの意味のオブジェクトを取り出し, その件数やアジアのオブジェクトの中での割合を調べる場合など, 分類結果からデータ分析を行うといったデータ利用では, 特定の意味のみに関するオブジェクトを取り出すことが必要になる. そのような特定の意味を求める問合せに対応するオブジェクトの集合をクラス集合に対するオブジェクト集合として議論する.

意味集合 S_1 と 1 つの意味 S_2 に対して, $\forall S_1 \in S_1, S_1 \preceq S_2$ ならば $S_1 \preceq S_2$ であり, 1 つの意味 S_1 と意味集合 S_2 に対して, $\forall S_2 \in S_2, S_1 \preceq S_2$ ならば $S_1 \preceq S_2$ である. これらを包含するような意味集合間の上下関係は, S_1, S_2 に対して, $\forall S_1 \in S_1, \exists S_2 \in S_2, S_1 \preceq S_2$ かつ $\forall S_2 \in S_2, \exists S_1 \in S_1, S_1 \preceq S_2$ ならば $S_1 \preceq S_2$ となる.

定義 3 クラス集合 C とオブジェクト o に対して, $\forall S_1 \in S(o), \exists S_2 \in S(C), S_1 \preceq S_2$ のとき, o は C の配属オブジェクトであり, C の配属オブジェクト集合を $m_A(C)$ で表す. また, $S(o) \preceq S(C)$ のとき, o は C の通常オブジェクトであり, C の通常オブジェクト集合を $m_N(C)$ で表す. □

クラス集合 C の配属オブジェクトは, オブジェクトの意味に C の意味に関係のない意味を含まないオブジェクトである. C の通常オブジェクト o は, 配属オブジェクトの条件に加え, $\forall S_1 \in S(C), \exists S_2 \in S(o), S_2 \preceq S_1$ が条件となっており, C の配属オブジェクトで C のすべての意味に対して同位または下位の意味を持つオブジェクトである.

例 2 例 1 の分類階層におけるクラス集合 $\{JAPAN, CHINA\}$ に対して, o_1, o_2, o_3 は, オブジェクトのいずれの意味も日本, 中国の下位の意味なので配属オブジェクトである. o_4 と o_5 は, 日本, 中国の下位の意味ではないバリを含むので配属オブジェクトではない. 配属オブジェクトの中で, o_3 は日本と中国の両方について下位の意味を含むので $\{JAPAN, CHINA\}$ の通常オブジェクトである. 一方, o_1 と o_2 は中国に対する下位の意味を持たないので通常オブジェクトではない. □

クラス集合 C に対する $m_A(C), m_N(C)$ は, C 中のクラス C に $S(o) \not\leq S(C)$ であるようなオブジェクト o が含まれるため, C の和集合, 積集合のみでは求めることができない. C に分類されているオブジェクト o のうち $S(o) \leq S(C)$ であるもの, すなわち $m_N(\{C\})$ は, o が重複している最初のクラスの深さを用いることで求められる⁸⁾. しかし, クラス集合 C の $m_A(C)$ や $m_N(C)$ は, $m_N(\{C\})$ ($C \in C$) の和集合や積集合では求めることはできない. 例 1 において, o_3 の意味はクラス $JAPAN$ と $CHINA$ の下位のもののみからなり, $m_A(\{JAPAN, CHINA\})$ や $m_N(\{JAPAN, CHINA\})$ に含まれるが, $m_N(\{JAPAN\})$ や $m_N(\{CHINA\})$ には含まれない.

クラス集合 C に対する $S(o) \not\leq S(C)$ であるようなオブジェクト o は C の終端クラス以外の終端クラスに分類されているので, $m_A(C)$ や $m_N(C)$ を求めるためには一般にはすべての終端クラスを調べる必要がある. 例 1 において, クラス $JAPAN$ と $CHINA$ の和集合によって得られたオブジェクト集合と $FRANCE$ の差集合でフランスの同位または下位の意味を含まないオブジェクトが求まる. 一般に, $m_A(C)$ や $m_N(C)$ は C の和集合や積集合で得られた結果から C の終端クラス以外のすべての終端クラスとの差集合によって求めることになり, すべてのオブジェクトを調べるために多大なコストを要する.

基本クラスにオブジェクトの重複が存在する問題に関しては, 子クラスに分類されたオブジェクトの中から 1 つを代表に選ぶことで解決でき⁸⁾, クラスのオブジェクトは重複なく求まることについて 4 章で述べる. 5 章では, クラス集合 C の $m_A(C)$ や $m_N(C)$ は, 代表に選ばれなかった情報を用いることで C の終端クラスのみから求められることを示す.

4. 重複に対する代表オブジェクト

オブジェクトは終端クラスに分類されるので, オブ

ジェクトが複数の意味を持つ場合でも補題 1 は成り立つ. しかし, 複数の意味のオブジェクトは複数の子クラスに分類され, 分類階層は排他性を満たさないので, 補題 2 は成り立たない. すなわち, クラス C のオブジェクトを求める際, C の基本クラスである終端クラスにはオブジェクトの重複が存在する. 本章は, 子クラス集合における重複したオブジェクトを 1 つのオブジェクトで代表させることで, $m(C)$ は C の基本クラスにおけるオブジェクトから重複なく求まることを示す.

分類階層における根クラスを C_0 で, C_0 から C までの経路の長さであるクラス C の深さを $d(C)$ で表す. C のオブジェクトは C の 1 つ以上の子クラスに分類されている. その中から 1 つを代表オブジェクトとして選び, 子クラス集合における同一のオブジェクトを代表させる. クラス C のオブジェクト o に対して, o が代表オブジェクトとはならなかったクラスの深さを C における o の L 値といい, $l(o, C)$ で表す.

- 根クラス C_0 のオブジェクト o に対し $l(o, C_0) = 0$.
- クラス C のオブジェクト o と C の子クラス C_i に対して, C_i の o が C の o を代表するとき $l(o, C_i) = l(o, C)$, そうでないとき $l(o, C_i) = d(C_i)$.
- オブジェクト o が分類されていないクラス C ($o \notin m(C)$) に対して, $l(o, C) = null$.

クラス C の子孫のクラスにおいて, 代表オブジェクトに選ばれなかったオブジェクトの L 値は C の子孫の深さとなるので $d(C)$ よりも大きい. したがって, $B(C)$ ではオブジェクトは重複するが, L 値が $d(C)$ 以下のものはただ 1 つである. すなわち, $l(o, C)$ は深さが $l(o, C)$ である C の先祖から C までのクラスで o が代表オブジェクトとなっていることを表している.

クラス集合 C に分類されているオブジェクト o のうち, クラス C の o に対する代表オブジェクトは次のように定義できる.

定義 4 クラス C とクラス集合 C について, C に対する C の代表オブジェクト集合は, $rep(C, C) = \{o \mid o \in m(C), l(o, C') \leq d(C), C' \in C\}$ である. □

また, クラス C のオブジェクトは, C の基本クラスにおける代表オブジェクトから求まる.

補題 3 クラス C に対して, $m(C) = rep(C, B(C))$ である. □

(証明) オブジェクトが子クラスに分類される時

	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}
o_1 L	0	0	null	0	null	null	0	null	null	null	null	null
R	{0}	{0}	ϕ	{0}	ϕ	ϕ	{0}	ϕ	ϕ	ϕ	ϕ	ϕ
o_2 L	0	0	null	0	null	null	3	3	0	null	null	null
R	{0}	{0}	ϕ	{0}	ϕ	ϕ	ϕ	ϕ	{0, 3, 3}	ϕ	ϕ	ϕ
o_3 L	0	0	null	0	2	null	0	3	null	2	3	null
R	{0}	{0}	ϕ	{0, 2}	ϕ	ϕ	{0, 2, 3}	ϕ	ϕ	{3}	ϕ	ϕ
o_4 L	0	0	1	0	2	1	0	null	null	2	null	1
R	{0}	{0, 1}	ϕ	{0, 1, 2}	ϕ	ϕ	{0, 1, 2}	ϕ	ϕ	ϕ	ϕ	ϕ
o_5 L	0	1	0	1	null	0	1	null	null	null	null	0
R	{0}	ϕ	{0, 1}	ϕ	ϕ	{0, 1}	ϕ	ϕ	ϕ	ϕ	ϕ	{0, 1}

図 2 オブジェクトの L 値と R 値

Fig. 2 L and R values of objects.

オブジェクトの 1 つは代表オブジェクトとなり、その L 値は親の L 値と一致するので、クラス C の終端クラスで、 C の L 値と同じ L 値を持つオブジェクトが存在する。したがって、 $m(C)$ の任意のオブジェクト o に対し、L 値が $l(o, C)$ であるオブジェクト o が $B(C)$ の中に存在する、すなわち $m(C) \subseteq \text{rep}(C, B(C))$ である。定義 4 より $\text{rep}(C, C) \subseteq m(C)$ なので、 $\text{rep}(C, B(C)) \subseteq m(B(C))$ である。また、補題 1 より $m(B(C)) = m(C)$ なので、 $\text{rep}(C, B(C)) \subseteq m(C)$ であり、 $m(C) = \text{rep}(C, B(C))$ となる。(証明終)

クラス C の基本クラスにおける代表オブジェクトには重複がない。

補題 4 クラス C の基本クラス C_i と C_j ($C_i, C_j \in B(C), i \neq j$) に対して、 $\text{rep}(C, \{C_i\}) \cap \text{rep}(C, \{C_j\}) = \phi$ である。□

(証明) クラス C におけるオブジェクト o に対して、 $o \in \text{rep}(C, \{C_i\}) \cap \text{rep}(C, \{C_j\})$ とする。 C_i の o と C_j の o はともに C の o に対する代表オブジェクトなので、それらは C_i と C_j の共通の先祖のクラスと同じオブジェクトを代表することになり、子クラス集合の中で 1 つが代表に選ばれるという代表オブジェクトの定義に反する。よって、 $\text{rep}(C, \{C_i\}) \cap \text{rep}(C, \{C_j\}) = \phi$ である。(証明終)

クラス C のオブジェクトは、 C の子孫の終端クラスのオブジェクトの中で L 値が $d(C)$ 以下のものであり、それらには重複がない。

定理 1 クラス C に対して、 $m(C)$ は $\text{rep}(C, \{C'\})$ ($C' \in B(C)$) の直和である。□

(証明) 補題 3 より $m(C) = \text{rep}(C, B(C))$ であり、補題 4 より $\text{rep}(C, \{C_i\}) \cap \text{rep}(C, \{C_j\}) = \phi$ なので、 $m(C)$ は $\text{rep}(C, \{C_i\})$ の直和となる。

(証明終)

例 3 図 1 では、下線で代表オブジェクトの例を示している。図 2 で、オブジェクトの上段は各クラスにおけるオブジェクトの L 値である。null はそのクラスにオブジェクトが含まれていないことを、太字はオブジェクトが実際に存在することを表している。 C_0 のオブジェクト集合の分類において、 C_1 の o_1, o_2, o_3, o_4 と C_2 の o_5 が代表に選ばれているので、 C_1 の o_1, o_2, o_3, o_4 と C_2 の o_5 の L 値は、親クラス C_0 における L 値、すなわち 0 である。一方、代表に選ばれなかった C_1 の o_5 、 C_2 の o_4 の L 値は、それぞれ C_1 、 C_2 の深さ、すなわち 1 である。 C_1 のオブジェクトを C_1 の基本クラス $B(C_1) = \{C_6, C_7, C_8, C_9, C_{10}\}$ から求める。 $B(C_1)$ における代表オブジェクトは L 値が $d(C_1) = 1$ 以下である C_6 の o_1, o_3, o_4, o_5 と C_8 の o_2 であり、 $m(C_1)$ に対する $\text{rep}(C_1, \{C_k\})$ ($C_k \in B(C_1)$) にはオブジェクトの重複がない。□

5. クラス集合の意味に一致するオブジェクト

複数の意味を持つオブジェクトの分類では、クラス C の要素には $S' \not\leq S(C)$ であるような意味 S' を含む意味のオブジェクトが含まれるので、クラス集合 C の和集合、積集合には $S' \not\leq S(C)$ であるような意味 S' を含むオブジェクトが含まれている。よって、 $m_A(C)$ 、 $m_N(C)$ を求めるためには一般にすべての終端クラスを調べる必要がある。本章では、代表に選ばれなかった情報を用いることで、 $m_A(C)$ や $m_N(C)$ は C の終端クラスのみから求まることを示す。

クラス C_i, C_j ($S(C_i) \preceq S(C_j)$) に対して、 $m(C_i) \subseteq m(C_j)$ なので、 $m_A(\{C_i, C_j\}) = m_A(\{C_j\})$ であり、 $m_N(\{C_i, C_j\}) = m_N(\{C_i\})$ である。よって、問合せにおけるクラス集合 C は、 $S(C_i) \not\leq S(C_j)$ 、 $i \neq j, C_i, C_j \in C$ であるようなクラス間に上下関係

のないクラス集合とする。

クラス C で代表に選ばれたオブジェクト o に対して、深さが $l(o, C)$ の先祖から C までのクラスにおける代表に選ばれなかったオブジェクトの L 値の重複を許した集合を C における o の R 値といい、 $r(o, C)$ で表す。

- 根クラス C_0 のオブジェクト o に対し $r(o, C_0) = \{0\}$.
- クラス C のオブジェクト o と C の子クラス集合 C に対して、 $C_i (\in C)$ の o が C の o を代表するとき $r(o, C_i) = r(o, C) \cup \bigcup_{C_j \in C} \{l(o, C_j) \mid j \neq i, o \in m(C_j)\}$, そうでないとき $r(o, C_i) = \phi$.
- オブジェクト o が分類されていない C ($o \notin m(C)$) に対して、 $r(o, C) = \phi$.

以下 L 値と R 値に関する集合演算は重複を許すものとする。クラス C のオブジェクト o の $r(o, C)$ は、 o が C の先祖以外のクラスに属していることについての情報、すなわち o の重複についての情報である。

例 4 図 2 のオブジェクトの下段は、例 3 の分類階層におけるオブジェクトの R 値である。 C_0 のオブジェクト集合の分類において、 C_1 の o_1, o_2, o_3, o_4 と C_2 の o_5 が代表に選ばれているので、それらの R 値は C_0 における R 値に C_0 の子クラスで代表に選ばれなかったオブジェクトの L 値を加えたものである。すなわち、 C_1 の o_1, o_2, o_3 は C_2 に分類されていないので、その R 値は C_0 における R 値 $\{0\}$ のままであり、 C_1 の o_4 と C_2 の o_5 の R 値は、 C_0 における R 値 $\{0\}$ に、それぞれ C_2 の o_4, C_1 の o_5 の L 値 $\{1\}$ を加えた $\{0, 1\}$ である。一方、代表に選ばれなかった C_2 の o_4 と C_1 の o_5 の R 値は ϕ である。 □

クラス集合 C とオブジェクト o に対して、 C における o が分類されている C ($C \in C, o \in m(C)$) の o の L 値を要素とするような重複を許した集合を $l(o, C) = \{l(o, C) \mid C \in C, l(o, C) \neq null\}$ で表し、同様に C における R 値の重複を許した集合を $r(o, C) = \bigcup_{C \in C} r(o, C)$ で表す。

オブジェクト o に対し、クラス集合 C は、 C 中のクラスの先祖や子孫ではないクラスに o が含まれないとき、 o の被覆であるという。 C が o の被覆ではないとき、 o を含む C に属さない C' が存在し、 $S(C') \not\subseteq S(C)$ なので $S(o)$ は $S(C)$ のいずれの意味に対しても同位または下位ではない意味を含む。そのため、 o は $m_A(C), m_N(C)$ のいずれの要素でもない。 o が C' に分類されていないことは、 $l(o, C)$ と $r(o, C)$ が等しいことで判断できる。 o が複数の子

ラスに分類されたときに、代表に選ばれなかった o の L 値は代表に選ばれた o の R 値に含まれることになる。よって、 o の被覆である C に対する R 値には o が代表に選ばれなかったときの L 値が含まれているので、 $l(o, C)$ と $r(o, C)$ は等しい。

補題 5 クラス間に上下関係のないクラス集合 C がオブジェクト o の被覆であれば、 $l(o, C) = r(o, C)$ である。 □

(証明) 根クラス C_0 と o が分類されている C_0 の子クラスの集合 C_t に対して、 R 値の定義より、

$$\{l(o, C_0)\} \cup r(o, C_t) = r(o, C_0) \cup l(o, C_t) \quad (1)$$

である。 C_t 中のクラス C'_i に対して、 C_t を $C_t - \{C'_i\}$ と $\{C'_i\}$ に分割することで、

$$\begin{aligned} \{l(o, C_0)\} \cup r(o, C_t - \{C'_i\}) \cup r(o, C'_i) = \\ r(o, C_0) \cup l(o, C_t - \{C'_i\}) \cup \{l(o, C'_i)\} \end{aligned} \quad (2)$$

となる。 o が分類されている C'_i の子クラスの集合を C'_t とすると、式 (1) は C'_i に対しても成り立つので、

$$\{l(o, C'_i)\} \cup r(o, C'_t) = r(o, C'_i) \cup l(o, C'_t) \quad (3)$$

である。式 (2) と式 (3) の両辺を加えて整理すると、

$$\begin{aligned} \{l(o, C_0)\} \cup r(o, C_t - \{C'_i\}) \cup r(o, C'_i) = \\ r(o, C_0) \cup l(o, C_t - \{C'_i\}) \cup l(o, C'_i) \end{aligned} \quad (4)$$

となる。 $C_t - \{C'_i\} \cup C'_t$ を C_t に定義しなおすと式 (1) が成り立つ。 o の被覆である C 中の o を含むクラスの集合を C' とする。このクラスの展開を C_t 中で C に含まれないクラス C'_i に対して繰り返すことで、 $C_t = C'$ となる C_t で式 (1) が成り立ち、 $\{l(o, C_0)\} = \{0\}, r(o, C_0) = \{0\}$ なので $r(o, C') = l(o, C')$ である。 o を含まないクラスでの L 値と R 値はそれぞれ $null$ と ϕ なので、 $l(o, C') = l(o, C), r(o, C') = r(o, C)$ であり $l(o, C) = r(o, C)$ である。 (証明終)

クラス集合 C に分類されているオブジェクト o に対して、 $l(o, C) = r(o, C)$ ならば、 o は C の被覆ではないクラス C' に分類されていない。

補題 6 クラス間に上下関係のないクラス集合 C とオブジェクト o ($o \in m(C)$) に対して、 $l(o, C) = r(o, C)$ ならば、 C は o の被覆である。 □

(証明) オブジェクト o が重複していないとき o を含むクラス集合 C は o の被覆である。クラス間に上下関係のないクラス集合 C がオブジェクト o の被覆ではないとき、 $C \subseteq C_v$ となる o の被覆であるクラス間に上下関係のないクラス集合 C_v が存在し、 $C' = C_v - C$ とする。 C_v は o の被覆なので補題 5 より $l(o, C_v) = r(o, C_v)$ であり、 $C_v = C \cup C'$ ($C \cap C' = \phi$) なので $l(o, C) \cup l(o, C') = r(o, C) \cup r(o, C')$ である。 C_v 中に根クラスを代表す

るオブジェクトは1つだけ存在するので、 L 値が0であるオブジェクトは1つだけ存在し、 C, C' のうちそれを含むクラスを C_1 , 他方のクラスを C_2 とする。 $l(o, C) \neq 0$ のとき $r(o, C)$ の要素は、深さが $l(o, C)$ である C の先祖から C までの o についての重複の情報なので、 $l(o, C)$ よりも大きい。したがって、 C_2 には $l(o, C) = 0$ となる o を含まないので、 R 値の要素の最小値は L 値よりも大きく、 $l(o, C_2) \neq r(o, C_2)$ である。また、 $l(o, C_1) \cup l(o, C_2) = r(o, C_1) \cup r(o, C_2)$ であることから $l(o, C_1) \neq r(o, C_1)$ である。したがって、 C が C_1, C_2 のいずれであっても $l(o, C) \neq r(o, C)$ である。すなわち、オブジェクト o の被覆ではないクラス集合に対する L 値と R 値は一致しない。よって、 C が o の被覆でなければ $l(o, C) \neq r(o, C)$ なので、 $l(o, C) = r(o, C)$ ならば C は o の被覆である。

(証明終)

補題5と補題6より、クラス集合 C の和集合(積集合)に含まれるオブジェクト o で $l(o, C) = r(o, C)$ であるような o は C の配属オブジェクト(通常オブジェクト)であり、それらの集合は $m_A(C) (m_N(C))$ である。

補題7 クラス間に上下関係のないクラス集合 C に対して、 C の配属オブジェクト集合は $m_A(C) = \{o \mid o \in \bigcup_{C \in \mathcal{C}} m(C), l(o, C) = r(o, C)\}$ であり、 C の通常オブジェクト集合は $m_N(C) = \{o \mid o \in \bigcap_{C \in \mathcal{C}} m(C), l(o, C) = r(o, C)\}$ である。□

(証明) $\bigcup_{C \in \mathcal{C}} m(C)$ に含まれる o が C の配属オブジェクトならば、 o は C 中のクラスの先祖や子孫ではないクラスに分類されておらず、 C は o の被覆となり、補題5より $l(o, C) = r(o, C)$ である。また、 $l(o, C) = r(o, C)$ ならば補題6より C は o の被覆であり、 o は C 中のクラスの先祖や子孫ではないクラスに含まれないので、 C の配属オブジェクトである。よって、 $m_A(C) = \{o \mid o \in \bigcup_{C \in \mathcal{C}} m(C), l(o, C) = r(o, C)\}$ となる。 C の配属オブジェクトのうち C のすべての意味を持つオブジェクトは C の通常オブジェクトなので、クラス集合 C の積集合に含まれるオブジェクト o で $l(o, C) = r(o, C)$ であるような o が C の通常オブジェクトである。すなわち、 C の通常オブジェクト集合は $m_N(C) = \{o \mid o \in \bigcap_{C \in \mathcal{C}} m(C), l(o, C) = r(o, C)\}$ となる。(証明終)

クラス C と C' に分類されているオブジェクト o に対して、 $l(o, C)$ と $r(o, C)$ は C の o を代表する o が分類されている C の終端クラスにおける o の L 値と R 値から求まる。 C の o を代表する o が分類されている C の基本クラスを $B_r(C, o)$ で表すと、

$l(o, B_r(C, o)) = l(o, C)$ である。よって、クラス集合 C に対して $B_r(C, o) = \{B_r(C, o) \mid C \in \mathcal{C}\}$ とすると、 $l(o, C) = l(o, B_r(C, o))$ となる。

$r(o, B_r(C, o))$ には C の子孫での代表に選ばれなかったクラスの o の L 値が含まれるので、 $r(o, B_r(C, o))$ の要素で $d(C)$ 以下のものからなる集合が $r(o, C)$ に等しくなる。よって、 $r_v(o, B_r(C, o)) = \bigcup_{C \in \mathcal{C}} \{k \mid k \in r(o, B_r(C, o)), k \leq d(C)\}$ とすると、 $r(o, C) = r_v(o, B_r(C, o))$ となる。

これにより、クラス集合 C の配属オブジェクトや通常オブジェクトは、 C の終端クラスの代表オブジェクトとその L 値、 R 値から求めることができる。

定理2 クラス間に上下関係のないクラス集合 C に対して、 C の配属オブジェクト集合は $m_A(C) = \{o \mid o \in \bigcup_{C \in \mathcal{C}} \text{rep}(C, B_r(C, o)), l(o, B_r(C, o)) = r_v(o, B_r(C, o))\}$ であり、 C の通常オブジェクト集合は $m_N(C) = \{o \mid o \in \bigcap_{C \in \mathcal{C}} \text{rep}(C, B_r(C, o)), l(o, B_r(C, o)) = r_v(o, B_r(C, o))\}$ である。□

(証明) 補題3より $m(C) = \text{rep}(C, B_r(C, o))$ である。 $B_r(C, o)$ の o は C の o の代表オブジェクトなので、 $l(o, C) = l(o, B_r(C, o))$ であり、 $l(o, C) = l(o, B_r(C, o))$ となる。 $r(o, B_r(C, o))$ は、 $r(o, C)$ に C の子孫における L 値を加えたものなので、 $r(o, B_r(C, o))$ から $d(C)$ より大きい要素を除くと $r(o, C)$ に一致する。したがって、 $r(o, C) = r_v(o, B_r(C, o))$ である。補題7をこれらを用いて表すと定理2となる。(証明終)

クラス集合 C に対する配属オブジェクトや通常オブジェクトを求めるために、 C の各クラスのオブジェクトはそのクラスの基本クラスの代表オブジェクトから求められる。 C の和集合、積集合におけるオブジェクト o で $l(o, B_r(C, o)) = r_v(o, B_r(C, o))$ となるものがそれぞれ、 C の配属オブジェクト、通常オブジェクトである。

例5 例3の分類階層において、クラス集合 $\{JAPAN, CHINA\}$ の配属オブジェクトと通常オブジェクトは以下のように求められる。

- (1) $JAPAN$ と $CHINA$ のオブジェクトをそれぞれ $\text{rep}(C_3, B_r(C_3))$, $\text{rep}(C_4, B_r(C_4))$ から求める。
- (2) $JAPAN$ と $CHINA$ の和集合で得られるオブジェクト集合 $\{o_1, o_2, o_3, o_4, o_5\}$ と積集合で得られるオブジェクト集合 $\{o_3, o_4\}$ は、それぞれ配属オブジェクトと通常オブジェクトの候補である。
- (3) C_6, C_9 の o_3 はそれぞれ、 C_3, C_4 の o_3 の代表オブジェクトである。 $l(o_3, B_r(\{C_3, C_4\}, o_3))$,

すなわち $\{l(o_3, C_6), l(o_3, C_9)\} = \{0, 2\}$ は $r(o_3, C_6)$ の要素で $d(C_3)$ 以下のものからなる集合 $\{0, 2\}$ と $r(o_3, C_9)$ の要素で $d(C_4)$ 以下のものからなる集合 ϕ の重複を許した集合 $r_v(o_3, B_r(\{C_3, C_4\}, o_3)) = \{0, 2\}$ に等しいので, o_3 は $\{JAPAN, CHINA\}$ の配属オブジェクトであり通常オブジェクトである. それに対して o_4 は, $\{l(o_4, C_6), l(o_4, C_9)\} = \{0, 2\}$ が $r(o_4, C_6)$ の要素で $d(C_3)$ 以下のものからなる集合 $\{0, 1, 2\}$ と $r(o_4, C_9)$ の要素で $d(C_4)$ 以下のものからなる集合 ϕ の重複を許した集合に等しくないので, $\{JAPAN, CHINA\}$ の配属オブジェクトでも通常オブジェクトでもない. 同様に, o_1 と o_2 は $\{JAPAN, CHINA\}$ の配属オブジェクトであり, o_5 は配属オブジェクトではない. □

クラス集合に対する配属オブジェクトや通常オブジェクトを取り出すには, L 値と R 値を用いない場合, オブジェクトが他の終端クラスに分類されていないかを調べるため, 一般にはクラス集合の和集合や積集合に対してクラス集合以外の終端クラスとの差集合が必要になる. クラス集合の終端クラスにおけるオブジェクト数を n , 分類階層における全オブジェクト数を N とすると, 一般には処理するオブジェクト数は N であるのに対し, L 値と R 値を用いれれば処理するオブジェクト数は n で, 用いない場合に比べて n/N になる.

6. おわりに

本論文は, データが複数の子クラスに分類されることによる問題は, 子クラスに分類されたデータのうち1つを代表に選び, 代表に選ばれなかったクラスの深さや選ばれなかった情報を用いることで解決できることを示した. L 値により, クラスのデータを重複なく取り出せ, L 値と R 値により, すべてのデータを調べることなく特定の意味のデータを取り出せる. 利用者は L 値や R 値を意識する必要がないため, どのクラスのデータを代表とするかはシステムが任意に決めることができる. また, L 値と R 値はオブジェクトが分類される際に親クラスの L 値や R 値と子クラスの深さから決まるので, L 値と R 値を生成するためのコストはほとんどない.

L 値や R 値は, セマンティックウェブやナレッジマネジメントのような収集した多様なデータを分類したうえで特定の意味集合のみのデータを取り出す必要がある分野に適用可能である. また, 分析目的で特定の

意味集合のみのデータを取り出す必要がある場合にも有用であるため, 分析に供するデータを供給するデータウェアハウスのデータモデルとしても応用が可能である.

本論文ではデータが複数のクラスに分類される問題に対して L 値や R 値が十分な情報であることについての議論に重点を置くため, データの意味は一様であるとし, データは終端クラスに分類されるものとした. しかし, データの多様性にはデータの意味のレベルが異なるものがあり, そのようなデータに対する分類階層は充足性を満たさない^{(3),(6)}. たとえば, 親クラスが国レベルで子クラスが都市レベルで構成されている分類階層では, 日本に関するデータは子クラスに分類できない. また, 分類が完了していないためにクラスのデータが子クラスに分類されていない場合もある. そのような場合では, 分類階層は充足性を満たさない⁽⁵⁾. 本論文で提案した L 値や R 値は, 基本クラスの定義を修正することで充足性を満たさない分類階層にまで拡張できる.

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)(課題番号 15500072)の支援を受けている.

参 考 文 献

- 1) Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J.: *Classification and Regression Trees*, Wadsworth (1984).
- 2) Chakrabarti, S., Dom, B., Agrawal, R. and Raghavan, P.: Using Taxonomy, Discriminants, and Signatures for Navigating in Text Databases, *Proc. Int'l Conf. on Very Large Data Bases (VLDB'97)*, pp.446-455 (1997).
- 3) Dumais, S. and Chen, H.: Hierarchical Classification of Web Content, *Proc. ACM Int'l Conf. on Research and Development in Information Retrieval*, pp.256-263 (2000).
- 4) 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).
- 5) Furukawa, T. and Kuzunishi, M.: Classification and Utilization of Data Belonging to Multiple Classes, *Proc. 8th World Multiconference on Systemics, Cybernetics and Informatics*, Vol.2, pp.289-294 (2004).
- 6) Furukawa, T. and Kuzunishi, M.: Hierarchical Classification of Heterogeneous Data, *Proc. IASTED Int'l Conf. on Databases and Appli-*

- cations (DBA2005), pp.252–257 (2005).
- 7) Koller, D. and Sahami, M.: Hierarchically Classifying Documents Using Very Few Words, *Proc. Int'l Conf. on Machine Learning*, pp.170–178 (1997).
- 8) Kuzunishi, M. and Furukawa, T.: Classification Hierarchies for Multiple Semantics Data, *Proc. World Multiconf. on Systemics, Cybernetics and Informatics*, Vol.3, pp.134–139 (2005).
- 9) Schutze, H., Hull, D. and Pedersen, J.O.: A Comparison of Classifiers and Document Representations for The Routing Problem, *Proc. Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp.229–237 (1995).
- 10) Spyns, P., Meersman, R. and Jarrar, M.: Data Modelling Versus Ontology Engineering, *ACM SIGMOD Record*, Vol.31, No.1, pp.12–17 (2002).
- 11) Sun, A. and Lim, E.: Hierarchical Text Classification and Evaluation, *Proc. IEEE Int'l Conf. on Data Mining (ICDM2001)*, pp.521–528 (2001).
- 12) Tiun, S., Abdullah, R. and Kong, T.E.: Automatic Topic Identification Using Ontology Hierarchy, *Proc. Int'l Conf. on Intelligent Text Processing and Computational Linguistics (CI-Cling 2001)*, pp.444–453 (2001).
- 13) Trech, M., Palmer, N. and Luniewski, A.: Type Classification of Semi-Structured Documents, *Proc. Int'l Conf. on Very Large Data Bases (VLDB'95)*, pp.263–274 (1995).
- 14) Wang, K., Zhou, S. and Liew, S.C.: Building Hierarchical Classifiers Using Class Proximity, *Proc. Int'l Conf. on Very Large Data Bases (VLDB'99)*, pp.363–374 (1999).
- 15) Yang, Y.: Expert Network: Effective and Efficient Learning from Human Decisions in Text Categorization, *Proc. Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp.13–22 (1994).
- 16) Zhang, T., Ramakrishnan, R. and Livny, M.: BIRCH: An Efficient Data Clustering Method for Very Large Databases, *Proc. ACM SIGMOD Int'l Conf. on Management of Data*, pp.103–114 (1996).

(平成 17 年 9 月 19 日受付)

(平成 18 年 3 月 13 日採録)

(担当編集委員 石川 博, 有次 正義, 片山 薫,
木俣 豊, 土田 正士)



葛西 正裕 (学生会員)

平成 15 年九州大学経済学部経済工学科卒業。平成 17 年同大学院経済学府修士課程修了。現在、同大学院経済学府博士後期課程に在学中。情報システムにおけるデータ利用の研究に従事。電子情報通信学会, 日本データベース学会等学生会員。



古川 哲也 (正会員)

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