

意味集合による複数の意味のデータの記述

葛西正裕[†] 古川哲也[‡]

[†]九州大学大学院経済学府

[‡]九州大学大学院経済学研究院

収集した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しておく必要があり、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。ある属性で分類を行ったとき、データは複数の意味を持つことが多い。複数の意味のデータは意味集合によって記述することになり、意味集合が表すデータ集合とは何かを明らかにしておく必要がある。本稿では、意味集合が表すデータ集合を意味集合の上下関係を用いることで明らかにし、その性質について議論する。

Specifying Multiple Semantics Data by Sets of Semantics

Masahiro Kuzunishi[†] and Tetsuya Furukawa[‡]

[†]Graduate School of Economics, Kyushu University

[‡]Faculty of Economics, Kyushu University

Collected data must be organized to be utilized efficiently, and hierarchical classification based on semantics of data is efficient approach to organize data. Semantics of data for a classification attribute is usually not single but multiple. Since multiple semantics data are specified by a set of semantics, we have to reveal what the set of the multiple semantics data represented by a set of semantics is. This paper discusses hierarchies of semantics sets for clarifying data set represented by a set of semantics, and shows the properties of the hierarchies.

1 はじめに

今日、利用可能なデータは著しく増加し、同時にデータはテキスト、画像、音声といった多様なものとなっている。データの多様性が増していることに伴い、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いる。データの種類を問わず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 [1][7][9]。

データの分類はオントロジのような意味を体系化したものに基づいて行われ、データは同じ属性で分類したときにでも複数の意味を持つことがある。例えば、ある企業に関するデータを業種という属性で分類すると、多角化経営をしている企業であれば情報技術と金融といった複数の業種に分類されること

がある。また、複数の地域で経営を展開している企業を地域で分類すると、日本と中国に拠点を置く企業であれば分類の属性に対するデータの意味は日本と中国になる。複数の意味を持つデータの分類では分類に用いる属性に対する代表的な1つの意味を用いてデータの分類を行われることがあるが [1][6][11]、複数の意味を分類に反映させることで複数の意味でデータを利用できる。

複数の意味を持つデータは意味集合を用いてデータを記述することになる。意味集合を $S = \{S_1, S_2, \dots\}$ とし、 S を用いてデータを記述するときの S の解釈は一般には2通りがある。

α S の意味のいずれかを持つ： $S_1 \vee S_2 \vee \dots$

β S の意味のすべてを持つ： $S_1 \wedge S_2 \wedge \dots$

例えば、 α は日本と中国のいずれかについて知りたい場合であり、 β は日本と中国の両方を扱っているものについて知りたい場合に用いられる。意味集合 S の2通りの解釈に対して、複数の意味を持つデータの意味のうちいずれかの意味が満たしているデータでよいのかすべての意味が満たすデータなのかで2通りある。よって、その組合せで4通りのデータの取り出し方がある。

- $\alpha 1$ データの意味のうち α を満たすものが存在する
- $\beta 1$ データの意味のうち β を満たすものが存在する
- $\alpha 2$ データの意味のすべてが α を満たす
- $\beta 2$ データの意味のすべてが β を満たす

意味集合を用いた複数の意味のデータの記述を意味集合の解釈とそれに対応するデータの取り出し方と考えると4通りになる。しかし、意味集合のすべての意味を持つデータを記述しようとする、 $\beta 1$ や $\beta 2$ ではその一部しか記述することができない。

[例 1] 図1は意味集合によるデータの記述について、 $\alpha 1$ から $\beta 2$ までの4通り示している。図2で示したような意味集合 {アジア, 環太平洋} に対してそのすべての意味を持つデータである {ソウル, シドニー, パリ} のデータは $\beta 1$ や $\beta 2$ では記述できない。すなわち、 $\beta 1$ や $\beta 2$ はデータの意味のうち1つで意味集合 S のすべての意味に対応することになるので、{ソウル, シドニー, パリ} のデータは該当しない。 □

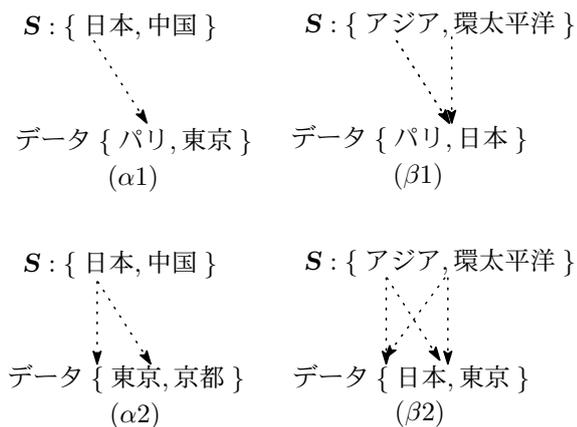


図1 意味集合の解釈とそれに対応する4通りのデータ

意味集合でデータを記述するためには、意味集合が表すデータはどのようなものがあるのかを明らかにしておく必要がある。本稿は意味集合とデータの

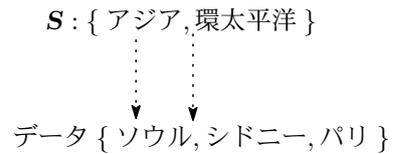


図2 意味集合のすべての意味を持つデータ

意味との間における意味集合の上下関係について検討することで意味集合が表すデータを明らかにし、意味集合の上下関係の性質について述べる。

複数の意味のデータに関する研究として、文献 [2][9] などは Web やテキストといった複数の意味のデータを対象としているが、複数のクラスへのデータの自動的な分類を目的としており、データの取り出し方にまでは議論が及んでいない。また、データの取り出し方についても意味集合による問合せに対してクラスの積集合や和集合の結果を返すといった利用に留まっている。一方で、オントロジに基づく分類階層が知られている。しかし、研究の多く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予め構成しているものであり [8][10]、意味集合が表すデータを明らかにすることで複数の意味に対するオントロジを提供するような研究は知られていない。

本稿は以下のように構成される。2節において、意味集合の上下関係は4種類あることを導く。3節では4種類であることの正しさを示すのと同時にデータを記述するには3種類でよいことを述べる。4節では、意味集合がデータを記述する際に必要な性質を明らかにして、3通りの意味集合の上下関係について検討し、5節では、別の性質を明らかにした上で、4節で示した性質を満たす上下関係について議論する。6節は2つの性質を満たす意味集合の上下関係を示し、7節では結論を述べる。

2 意味集合の上下関係を用いたオブジェクトの記述

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。例えば、地理という属性で、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は [3][9] などで議論されており、本稿ではある限定された属性でのデータの分類を対象とする。

1件のデータをオブジェクト o とし、分類で用いられる意味を S で表す。 S が表すオブジェクト集合を

\bar{S} で、分類に用いられる属性に対するオブジェクト o の意味を \bar{o} で表す。意味 S_1 と S_2 に対して、 S_2 が S_1 の上位 (S_1 が S_2 の下位) であることを $S_1 \prec S_2$ で、 S_1 と S_2 が同位であることを $S_1 \approx S_2$ で表す。 S_1 が S_2 の同位または下位であること、すなわち S_1 が S_2 以下であることを $S_1 \lesssim S_2$ で表す。オブジェクトが \bar{S} の要素であることは、 S とオブジェクトの意味によって決定され、 $\bar{o} \leq S$ ならば o は \bar{S} の要素となる。

オブジェクトをある属性で分類する際、オブジェクトの意味は複数になる場合があり、オブジェクトの意味は意味集合になる。そのようなオブジェクトは意味集合を用いて記述することになるので、意味集合が表すオブジェクト集合を明らかにしておく必要がある。意味集合 S が表すオブジェクト集合を \bar{S} で、意味集合の上下関係を \lesssim で表す。 \bar{S} を 1 つの意味が表すオブジェクトと同様に考えると \bar{S} は S 以下の意味集合のオブジェクトになる ($\bar{S} = \{o \mid \bar{o} \lesssim S\}$)。 \bar{S} を S 以下の意味集合のオブジェクトとして記述するには、 S とオブジェクトの意味集合に上下関係が必要になり、意味集合の上下関係を定義しなければならない。

本節では、下位となる意味集合の要素が 1 つとして意味集合の上下関係を考えた上で、要素を複数にして意味集合の上下関係を導く。意味集合 S_1 と S_2 に対して、 $|S_1| = 1, S_1 = \{S_1\}, |S_2| \geq 1$ のとき

A : S_2 のいずれかが S_1 以上であるとき S_2 は S_1 以上である、すなわち、 $\exists S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_A S_2$

B : S_2 のすべてが S_1 以上であるとき S_2 は S_1 以上である、すなわち、 $\forall S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_B S_2$

意味集合の上下関係は、 S_1 の要素を複数にしたときに、その意味集合のいずれかまたはすべてに対して上下関係があるかで 2 通りになる。

A1 : $\exists S_1 \in S_1, S_1 \lesssim_A S_2 \Rightarrow S_1 \lesssim_{A1} S_2$, すなわち、
 $\exists S_1 \in S_1, \exists S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{A1} S_2$

A2 : $\forall S_1 \in S_1, S_1 \lesssim_A S_2 \Rightarrow S_1 \lesssim_{A2} S_2$, すなわち、
 $\forall S_1 \in S_1, \exists S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{A2} S_2$

B1 : $\exists S_1 \in S_1, S_1 \lesssim_B S_2 \Rightarrow S_1 \lesssim_{B1} S_2$, すなわち、
 $\exists S_1 \in S_1, \forall S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{B1} S_2$

B2 : $\forall S_1 \in S_1, S_1 \lesssim_B S_2 \Rightarrow S_1 \lesssim_{B2} S_2$, すなわち、
 $\forall S_1 \in S_1, \forall S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{B2} S_2$

意味集合の上下関係を \lesssim_i ($i \in \{A1, A2, B1, B2\}$) とするときの意味集合 S が表すオブジェクト集合を \bar{S}^i で表す。1 節の例 1 における意味集合に対するオ

ブジェクトの 4 通りの取り出し方を意味集合の上下関係を用いて表すと以下の通りである。

- $\alpha 1 = \bar{S}^{A1}$
- $\beta 1 = \bar{S}^{B1}$
- $\alpha 2 = \bar{S}^{A2}$
- $\beta 2 = \bar{S}^{B2}$

意味集合の上下関係 $\lesssim_{A1}, \lesssim_{A2}, \lesssim_{B1}, \lesssim_{B2}$ は、意味集合における下位の意味集合 S_1 から上位の意味集合 S_2 を定義しているが、 S_2 から S_1 を定義すると以下のようなになる。

A1' : $\exists S_2 \in S_2, \exists S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{A1'} S_2$

A2' : $\exists S_2 \in S_2, \forall S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{A2'} S_2$

B1' : $\forall S_2 \in S_2, \exists S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{B1'} S_2$

B2' : $\forall S_2 \in S_2, \forall S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{B2'} S_2$

8 種類の意味集合が得られたが、 $\lesssim_{A1'}$ と $\lesssim_{B2'}$ はそれぞれ \lesssim_{A1} と \lesssim_{B2} に等しい。また、それ以外の上下関係 $\lesssim_{A2}, \lesssim_{A2'}, \lesssim_{B1}, \lesssim_{B1'}$ について、 \lesssim_{B1} と $\lesssim_{A2'}$ は、それぞれ $\lesssim_{B1'}$ と \lesssim_{A2} の定義を満たす定義なので、 $\lesssim_{B1'}$ と \lesssim_{A2} に含めると、8 通りの上下関係は以下に示す 4 通りの上下関係で考えればよい。

[定義 1] 意味集合 S_1 と S_2 に対して、

0 : S_1 のすべての意味が S_2 のすべての意味以下であるとき、 S_1 は S_2 以下である、すなわち

$\forall S_1 \in S_1, \forall S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_0 S_2$

I : S_1 のいずれかの意味が S_2 のいずれかの意味以下であるとき、 S_1 は S_2 以下である、すなわち

$\exists S_1 \in S_1, \exists S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_I S_2$

II : S_1 のすべての意味が S_2 のいずれかの意味以下であるとき、 S_1 は S_2 以下である、すなわち

$\forall S_1 \in S_1, \exists S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{II} S_2$

III : S_2 のすべての意味が S_1 のいずれかの意味以上であるとき、 S_2 は S_1 以上である、すなわち

$\forall S_2 \in S_2, \exists S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{III} S_2$

□

\lesssim_{A1} から $\lesssim_{B2'}$ までの意味集合の上下関係は定義 1 のいずれかの上下関係である。

- $\lesssim_{B2} = \lesssim_{B2'} = \lesssim_0$
- $\lesssim_{A1} = \lesssim_{A1'} = \lesssim_I$
- $\lesssim_{A2} = \lesssim_{II}$ ($\lesssim_{A2'}$ は \lesssim_{II} の条件を満たす)
- $\lesssim_{B1'} = \lesssim_{III}$ (\lesssim_{B1} は \lesssim_{III} の条件を満たす)

3 オブジェクトを記述するための意味集合の上下関係

2節は下位となる意味集合の要素を1つとして意味と意味集合の上下関係から意味集合の上下関係を導いた。本節では、上位となる意味集合から同様にして意味集合の上下関係を導いても意味集合の上下関係は4通りで考えればよいことを示す。

意味集合 S_1 と S_2 に対して、 $|S_1| \geq 1$, $|S_2| = 1$, $S_2 = \{S_2\}$ のとき

C : S_1 のいずれかが S_2 以下であるとき、 S_1 は S_2 以下である、すなわち、 $\exists S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_C S_2$

D : S_1 のすべてが S_2 以下であるとき、 S_1 は S_2 以下である、すなわち、 $\forall S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_D S_2$

意味集合の上下関係は、 S_2 の要素を複数にしたときに、その意味集合のいずれかまたはすべてに対して上下関係があるかで2通りになる。

$C1$: $\exists S_2 \in S_2, S_1 \lesssim_C S_2 \Rightarrow S_1 \lesssim_{C1} S_2$, すなわち、 $\exists S_2 \in S_2, \exists S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{C1} S_2$

$C2$: $\forall S_2 \in S_2, S_1 \lesssim_C S_2 \Rightarrow S_1 \lesssim_{C2} S_2$, すなわち、 $\forall S_2 \in S_2, \exists S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{C2} S_2$

$D1$: $\exists S_2 \in S_2, S_1 \lesssim_D S_2 \Rightarrow S_1 \lesssim_{D1} S_2$, すなわち、 $\exists S_2 \in S_2, \forall S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{D1} S_2$

$D2$: $\forall S_2 \in S_2, S_1 \lesssim_D S_2 \Rightarrow S_1 \lesssim_{D2} S_2$, すなわち、 $\forall S_2 \in S_2, \forall S_1 \in S_1, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{D2} S_2$

\lesssim_{C1} から \lesssim_{D1} までの意味集合の定義について、意味集合における下位の意味集合 S_1 と上位の意味集合 S_2 の順を逆にして定義すると以下の定義が得られる。

$C1'$: $\exists S_1 \in S_1, \exists S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{C1'} S_2$

$C2'$: $\exists S_1 \in S_1, \forall S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{C2'} S_2$

$D1'$: $\forall S_1 \in S_1, \exists S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{D1'} S_2$

$D2'$: $\forall S_1 \in S_1, \forall S_2 \in S_2, S_1 \lesssim S_2 \Rightarrow S_1 \lesssim_{D2'} S_2$

\lesssim_{C1} から $\lesssim_{D2'}$ までの上下関係は定義1の上下関係のいずれかである。

- $\lesssim_{D2} = \lesssim_{D2'} = \lesssim_{B2} = \lesssim_{B2'} = \lesssim_0$
- $\lesssim_{C1} = \lesssim_{C1'} = \lesssim_{A1} = \lesssim_{A1'} = \lesssim_1$
- $\lesssim_{D1'} = \lesssim_{A2} = \lesssim_{II}$ ($\lesssim_{D1} = \lesssim_{A2'}$ は \lesssim_{II} の条件を満たす)
- $\lesssim_{C2} = \lesssim_{B1'} = \lesssim_{III}$ ($\lesssim_{C2'} = \lesssim_{B1}$ は \lesssim_{III} の条件を満たす)

意味集合の上下関係を \lesssim_0 で定義して記述できる意味集合 S のオブジェクトは、 S の共通の意味を持

つ意味のうち最も下位となる意味1つに対して定義 \lesssim_{II} で定義して記述できるオブジェクトに等しい。例えば、意味集合 $S = \{\text{アジア}, \text{日本}\}$ を \lesssim_0 で定義したときに S で記述されるオブジェクトは、 $\{\text{東京}\}$ や $\{\text{関東}, \text{九州}\}$ といったオブジェクトになり、それは、 \lesssim_{II} で定義したときに S の共通の意味のうち最も下位となる意味である $\{\text{日本}\}$ のオブジェクトに等しくなる。 \lesssim_0 で定義して記述できる意味集合 S のオブジェクトは、定義 \lesssim_{II} で定義しても記述できるので \lesssim_0 を \lesssim_{II} に含めて考えることにする。

4, 5節において、意味集合の上下関係を考えるときに満たすべき2つの性質についてそれぞれ述べて、3通りの意味集合の上下関係について検討する。

4 健全性を満たす上下関係

3節ではオブジェクトを記述するための意味集合の上下関係には3通りがあることを示した。本節は、意味の上下関係で成り立つ性質から意味集合の上下関係においてもオブジェクトの記述するために必要な性質について明らかにし、3通りの意味集合の上下関係がその性質を有するかを議論する。

意味の上下関係は推移律を満たすので、意味 S_1 と S_2 に対して、 S_1 と S_2 の上下関係と S_1 のオブジェクト集合と S_2 のオブジェクト集合の包含関係は一致する。

[補題1] 意味 S_1 と S_2 に対して、 $S_1 \lesssim S_2 \Leftrightarrow \overline{S_1} \subseteq \overline{S_2}$ である。□

(証明) $\forall o \in \overline{S_1}$ について、 $\tilde{o} \lesssim S_1, S_1 \lesssim S_2$ より、 $\tilde{o} \lesssim S_2$ なので $o \in \overline{S_2}$ である。よって、 $S_1 \lesssim S_2$ ならば $\overline{S_1} \subseteq \overline{S_2}$ である。 $\tilde{o} \approx S_1$ であるようなオブジェクト o は、 $o \in \overline{S_1}$ であり、 $\overline{S_1} \subseteq \overline{S_2}$ ならば $o \in \overline{S_2}$ である。よって、 $\tilde{o} \lesssim S_2$ であり、 $\tilde{o} \approx S_1$ なので $S_1 \lesssim S_2$ である。(証明終)

意味集合の上下関係が、補題1と同様の性質を有するとき、その上下関係は健全である。

[定義2] 意味集合 S_1 と S_2 に対して、 $S_1 \lesssim S_2 \Leftrightarrow \overline{S_1} \subseteq \overline{S_2}$ ならば、意味集合の上下関係 \lesssim は健全な上下関係である。□

意味集合が健全でないとき、意味集合 S_1 と S_2 に対して、 $S_1 \lesssim S_2$ であっても S_1 の要素となるオブジェクトが S_2 の要素にはならない場合があり問題である。意味集合の上下関係 \lesssim が健全な上下関係であるかは \lesssim が推移律を満たすかを議論すればよい。

[補題 2] 意味集合の上下関係 \lesssim が健全な上下関係であることに対して、 \lesssim が推移律を満たすことは必要十分である。□

(証明) 意味集合の上下関係が \lesssim が推移律を満たせば、補題 1 の証明と同様に、 $S_1 \lesssim S_2 \Leftrightarrow \overline{S_1} \subseteq \overline{S_2}$ が成り立ち \lesssim は健全である。また、意味集合の上下関係が推移律を満たさないならば、意味 S_1 のオブジェクト o と S_1 より上位の意味 S_2 に対して、 $o \lesssim S_1$ かつ $S_1 \lesssim S_2$ であっても、 $o \lesssim S_2$ が成り立たない場合がある。よって、 o は S_2 のオブジェクトとは限らないことになり、オブジェクトの包含関係は成り立たないので、 \lesssim は健全な上下関係ではない。すなわち、 \lesssim が健全な上下関係であれば \lesssim は推移律を満たす。したがって、 \lesssim が健全な上下関係であることに対して、 \lesssim が推移律を満たすことは必要十分である。(証明終)

$\lesssim_I, \lesssim_{II}, \lesssim_{III}$ について、推移律が成り立つかを検討することで健全な上下関係であることを明らかにする。

意味集合の上下関係 \lesssim_I は推移律を満たさないので健全ではない。

[例 2] 図 3 は意味集合の上下関係を \lesssim_I としたときに推移律を満たさない例を表している。意味集合 $S_1 = \{ \text{日本, パリ} \}$ と $S_2 = \{ \text{フランス, 北京} \}$ とする。 $\{ \text{福岡, 中国} \}$ のオブジェクトを o とするとき、 o は S_1 の日本の下位である福岡の意味を持つので o は S_1 の要素であるが、 S_2 のフランスや北京の下位の意味を o は持たないので o は S_2 の要素ではない。よって、 \lesssim_I は推移律を満たさないので健全ではない。□

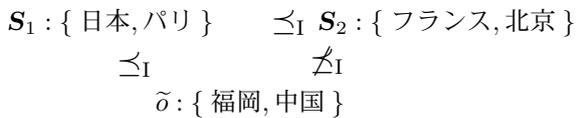


図 3 推移律を満たさない \lesssim_I

意味集合の上下関係 $\lesssim_{II}, \lesssim_{III}$ は推移律を満たすので健全な上下関係である。

[定理 1] 意味集合の上下関係 $\lesssim_{II}, \lesssim_{III}$ は健全な上下関係である。□

(証明) 意味集合 S_1, S_2, S_3 に対して、 $S_1 \lesssim_{II} S_2$ かつ $S_2 \lesssim_{II} S_3$ とし、推移律が成り立つことを証明する。 $\forall S_1 \in S_1, \exists S_2 \in S_2, S_1 \lesssim S_2$ かつ

$\forall S_2 \in S_2, \exists S_3 \in S_3, S_2 \lesssim S_3$ より $\forall S_1 \in S_1, \exists S_3 \in S_3, S_1 \lesssim S_3$ であるので、 $S_1 \lesssim_{II} S_3$ が成り立つ。よって、 $S_1 \lesssim_{II} S_2$ かつ $S_2 \lesssim_{II} S_3$ ならば $S_1 \lesssim_{II} S_3$ が成り立つので \lesssim_{II} は推移律を満たすので、補題 2 より \lesssim_{II} は健全な上下関係である。 \lesssim_{III} についても同様である。(証明終)

意味集合の上下関係 \lesssim_I が健全な上下関係でないのに対して、 \lesssim_{II} と \lesssim_{III} は健全な上下関係である。

5 意味集合の上限と下限

4 節では、 \lesssim_I は健全ではないが \lesssim_{II} と \lesssim_{III} は健全な上下関係であることを示した。本節は、意味集合でオブジェクトを記述するときの別の性質を明らかにした上で、 \lesssim_{II} と \lesssim_{III} に対してその性質について議論する。

意味集合 S_1 と S_2 に対して、 S_1 と S_2 が異なれば、 S_1 と S_2 によって記述されるそれぞれのオブジェクト集合もオブジェクトを記述するときには一般に異なる。

[定義 3] 意味集合 S_1 と S_2 に対して、 $S_1 \neq S_2$ ならば $\overline{S_1} \neq \overline{S_2}$ であるとき、意味集合の上下関係 \lesssim は妥当な上下関係である。□

オブジェクトの意味の中には、 $\{ \text{アジア, 日本} \}$ といった意味集合の意味に上下関係があるものがある。例えば、日本の経済状況をアジア全体の経済状況と比較するような文献の意味は $\{ \text{アジア, 日本} \}$ になる。全体とその一部について述べているようなオブジェクトは少なくないため、上下関係のある意味集合を考えることが必要になる。意味集合 S が $S_i \prec S_j$ ($S_i, S_j \in S$) である S_i と S_j が存在するような意味集合であるとき、 S は上下関係のある意味集合とする。意味集合の上下関係 \lesssim_{II} と \lesssim_{III} では、上下関係のある意味集合を考えたとき、 $S_1 \lesssim S_2$ かつ $S_2 \lesssim S_1$ となる意味集合 S_1 と S_2 が存在する。 $S_1 \lesssim S_2$ かつ $S_2 \lesssim S_1$ であるとき S_1 と S_2 は同位であるとよび、 $S_1 \approx S_2$ で表す。 \lesssim_{II} と \lesssim_{III} では $S_1 \approx_{II} S_2, S_1 \approx_{III} S_2$ である意味集合が存在する。

[例 3] 図 4 は意味集合の上下関係を \lesssim_{II} としたときに同位となる意味集合を示している。意味集合 $S_1 = \{ \text{アジア, 福岡} \}$ 、 $S_2 = \{ \text{アジア, 日本} \}$ としたとき、 S_1 のアジアと福岡は S_2 のアジア、日本の意味以下なので $S_1 \lesssim_{II} S_2$ である。また、 S_2 のアジアと日本は S_1 のアジアの意味以下なので $S_2 \lesssim_{II} S_1$ となるため、 $S_1 \approx_{II} S_2$ である。また、図 5 は意味集合の

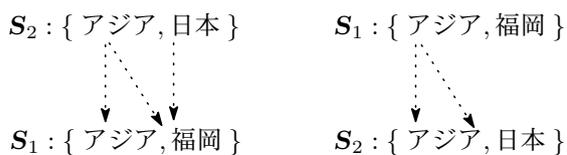


図4 \lesssim_{II} で同位となる意味集合

上下関係を \lesssim_{III} としたときに同位となる意味集合を示している. 意味集合 $S_1 = \{日本, 福岡\}$, $S_2 = \{アジア, 福岡\}$ としたとき, S_2 のアジアと福岡は S_2 の日本, 福岡の意味以上なので $S_1 \lesssim_{III} S_2$ である. また, S_1 の日本と福岡は S_1 の福岡の意味以上なので $S_2 \lesssim_{III} S_1$ となるため, $S_1 \approx_{III} S_2$ である. \square

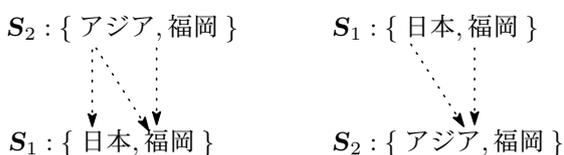


図5 \lesssim_{III} で同位となる意味集合

上下関係のある意味集合 S に対して, 上下関係のある意味集合の中で最も上位または下位の意味で構成される S の部分集合を以下のように定義する.

- 上下関係のある意味集合 S に対して, 上下関係のある意味集合の中で最も上位の意味で構成される S の部分集合を S の上限とし, $u(S) = \{S \mid S \in S, S \neq S' \text{ であるような } \forall S' \in S, S \not\prec S'\}$ で表す.
- 上下関係のある意味集合 S に対して, 上下関係のある意味集合の中で最も下位の意味で構成される S の部分集合を S の下限とし, $l(S) = \{S \mid S \in S, S \neq S' \text{ であるような } \forall S' \in S, S' \not\prec S\}$ で表す.

上下関係のある意味集合 S で記述されるオブジェクト集合は, \lesssim_{II} では S の上限で記述されるオブジェクト集合に等しく, \lesssim_{III} では S の下限で記述されるオブジェクト集合に等しい.

[定理 2] 意味集合 S に対して, $\bar{S}^{II} = \overline{u(S)^{II}}$ であり $\bar{S}^{III} = \overline{l(S)^{III}}$ である. \square

(証明) $\forall o \in \bar{S}^{II}$ について, $\forall S \in \tilde{o}, \exists S' \in S, S \lesssim S'$ であり, $\forall S' \in S, \exists S'' \in u(S), S' = S''$ である. よって, $\forall S \in \tilde{o}, \exists S'' \in u(S)$ に対して, $S \lesssim S''$

なので $\tilde{o} \lesssim_{II} u(S)$ であり $o \in u(S)$ であるから, $\bar{S}^{II} \subseteq \overline{u(S)^{II}}$ が成り立つ. $\forall o \in \overline{u(S)^{II}}$ について, $\forall S \in \tilde{o}, \exists S' \in u(S), S \lesssim S'$ であり, $u(S) \subseteq S$ なので, $\forall S \in \tilde{o}, \exists S'' \in S, S \lesssim S''$ である. よって, $\tilde{o} \lesssim_{II} S$ であり $o \in S$ であるから, $\overline{u(S)^{II}} \subseteq \bar{S}^{II}$ が成り立つ. したがって, $\bar{S}^{II} = \overline{u(S)^{II}}$ である. $\bar{S}^{III} = \overline{l(S)^{III}}$ についても同様である. (証明終)

定理 2 より, $S_1 \neq S_2$ であるような意味集合 S_1 と S_2 に対して, $u(S_1) = u(S_2)$ ならば $\bar{S}_1^{II} = \bar{S}_2^{II}$ であり, $l(S_1) = l(S_2)$ ならば $\bar{S}_1^{III} = \bar{S}_2^{III}$ である. よって, \lesssim_{II} と \lesssim_{III} はともに妥当な上下関係ではない.

6 健全かつ妥当な上下関係

4.5 節において, \lesssim_{II} と \lesssim_{III} は健全ではあるが妥当な上下関係ではないことを示したのに対して, 本節では \lesssim_{II} と \lesssim_{III} の条件を両方満たす上下関係は健全であり妥当な上下関係になることを明らかにする.

意味集合の上下関係 $\lesssim_I, \lesssim_{II}, \lesssim_{III}$ の他にそれら 3 通りの組合せによる意味集合の上下関係を考えたとき, \lesssim_I と \lesssim_{III} の両方を満たす上下関係, \lesssim_I と \lesssim_{II} の両方を満たす上下関係はそれぞれ $\lesssim_{II}, \lesssim_{III}$ になるが, \lesssim_{II} かつ \lesssim_{III} は別の意味集合の上下関係の定義になる.

[定義 4] 意味集合 S_1 と S_2 に対して, $\forall S_1 \in S_1, \exists S_2 \in S_2, S_1 \lesssim S_2$ かつ $\forall S_2 \in S_2, \exists S_1 \in S_1, S_1 \lesssim S_2$ ならば S_1 は S_2 以下であるとする上下関係を $S_1 \lesssim_{IV} S_2$ で表す. \square

[例 4] 図 6 は意味集合の上下関係 \lesssim_{IV} において意味集合 S_1 は S_2 以下である意味集合の例を示している. 意味集合を $S_1 = \{福岡, 上海\}$, $S_2 = \{日本, 中国\}$ としたとき, S_1 におけるすべての意味である福岡と上海は S_2 における日本と中国の意味以下, かつ S_2 におけるすべての意味である日本, 中国は S_1 における福岡, 上海の意味以上なので $S_1 \lesssim_{IV} S_2$ である. \square



図6 意味集合の上下関係 \lesssim_{IV}

意味集合 S_1 と S_2 の上下関係には, 1 組の意味に上

下関係があればよいとする場合があり、 \lesssim_I が対応する (図 7). それに対して、意味集合のすべての意味に対して上下関係があるとき、意味集合に上下関係があるとする場合があり、下位の意味集合のすべてに対してか、上位の意味集合のすべてに対してかの 2通りがあって、下位の意味集合であれば \lesssim_{II} , 上位の意味集合であれば \lesssim_{III} が対応する. すなわち、 \lesssim_I は上位と下位の意味集合のどちらにも制限がないのに対して、 \lesssim_{II} は下位の意味集合に制限がある関係、 \lesssim_{III} は上位の意味集合に制限がある関係と考えることができる. 意味集合のすべての意味に対して、上位と下位の意味集合の両方である場合は \lesssim_{IV} が対応し、 \lesssim_{IV} は上位と下位の意味集合の両方に制限がある関係である.

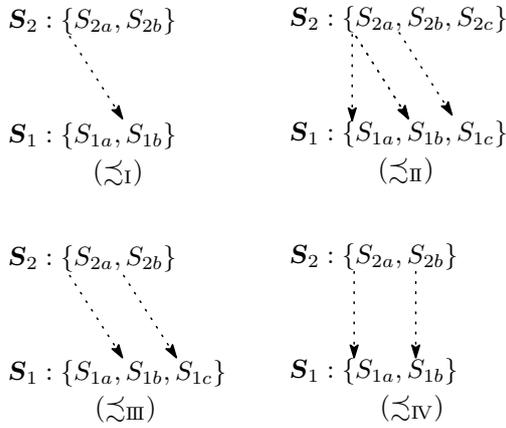


図 7 上位下位の意味集合の制限

意味集合の上位と下位の両方に制限がある意味集合の上下関係 \lesssim_{IV} は、オブジェクトを記述する上下関係として健全であり妥当な上下関係である.

[定理 3] 意味集合の上下関係 \lesssim_{IV} は健全な上下関係である. \square

(証明) 意味集合 S_1, S_2, S_3 に対して、 $S_1 \lesssim_{IV} S_2$ かつ $S_2 \lesssim_{IV} S_3$ とし、推移律が成り立つことを証明する. $\forall S_1 \in S_1, \exists S_2 \in S_2, S_1 \lesssim S_2$ かつ $\forall S_2 \in S_2, \exists S_3 \in S_3, S_2 \lesssim S_3$ より、 $\forall S_1 \in S_1, \exists S_3 \in S_3, S_1 \lesssim S_3$ が成り立つ. 同様に $\forall S_3 \in S_3, \exists S_1 \in S_1, S_1 \lesssim S_3$ が成り立つので $S_1 \lesssim_{IV} S_3$ である. よって、 $S_1 \lesssim_{II} S_2$ かつ $S_2 \lesssim_{II} S_3$ ならば $S_1 \lesssim_{II} S_3$ が成り立つので \lesssim_{IV} は推移律を満たすので、 \lesssim_{IV} は健全な上下関係である. (証明終)

[定理 4] 意味集合の上下関係 \lesssim_{IV} は妥当な上下関係である. \square

(証明) 意味集合が同位であることと意味集合が等しいことが一致すれば、意味集合が異なれば同位になることはなく意味集合のオブジェクト集合も異なるので、 \lesssim_{IV} は妥当な上下関係である. 意味集合 S_1 と S_2 に対して、 $S_1 \approx_{IV} S_2 \Leftrightarrow S_1 = S_2$ であることを証明する. S_1 と S_2 に対して、 $S_1 \approx_{IV} S_2$ ならば、 S_1 のすべてが S_2 のいずれか以下かつ S_2 のすべてが S_1 のいずれか以上であり、かつ S_2 のすべてが S_1 のいずれか以下、かつ S_1 のすべてが S_2 のいずれか以上である. S_1 のすべてが S_2 のいずれか以上かつ以下、すなわち S_2 と等しくかつ、 S_2 のすべてが S_1 のいずれか以上かつ以下、すなわち S_2 と等しいので、 $S_1 = S_2$ である. また、 $S_1 = S_2$ ならば \approx_{IV} の定義より $S_1 \approx_{IV} S_2$ である. よって、 $S_1 \approx_{IV} S_2 \Leftrightarrow S_1 = S_2$ が成り立つので、 \lesssim_{IV} は妥当な上下関係である. (証明終)

7 むすび

本稿では、意味集合が表すオブジェクト集合を意味集合の上下関係を用いることで記述し、その性質について議論した. 意味集合の上下関係が推移律を満たさないとき、意味集合の上下関係とその意味集合が表すオブジェクト集合の包含関係は一致しないのでオブジェクトの記述に用いる上下関係として健全ではない. また、意味集合が異なっても同位となる意味集合が存在するような上下関係では、異なる意味集合に対してもその意味集合が表すオブジェクト集合が一致する場合があるので、オブジェクトの記述に対して妥当な上下関係ではない. それに対して、意味集合の上位と下位の両方に制限がある意味集合の上下関係 \lesssim_{IV} は、オブジェクトを記述する上下関係として健全であり妥当な上下関係である.

参考文献

- [1] 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).
- [2] 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).
- [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] Furukawa, T. and Kuzunishi, M.: Hierarchical Classification of Heterogeneous Data, *Proc. IASTED Int'l Conf. on Databases and Applications (DBA2005)*, pp. 252–257 (2005).
- [6] Koller, D. and Sahami, M.: Hierarchically Classifying Documents Using Very Few Words, *Proc. Int'l Conf. on Machine Learning*, pp. 170–178 (1997).
- [7] 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).
- [8] Spyns, P., Meersman, R. and Jarrar, M.: Data Modelling Versus Ontology Engineering, *ACM SIGMOD Record*, Vol. 31, Nov. 1, pp. 12–17 (2002).
- [9] Sun, A. and Lim, E.: Hierarchical Text Classification and Evaluation, *Proc. IEEE Int'l Conf. on Data Mining (ICDM2001)*, pp. 521–528 (2001).
- [10] 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).
- [11] 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).