

情報の特性解析による意味構造モデル

○宇陀則彦†

藤原 譲†

The Semantic Structure Model based on Analysis of Information

Norihiko Uda†

Yuzuru Fujiwara†

Meaning is a serious problem to be solved in many fields such as databases, natural language understanding, artificial intelligence etc. It is necessary to analyze property of information for semantic processing. This paper describes an information model called the semantic structure model which is based on property of information. The semantic structure model focuses on internal structure of concepts, overlapping among concepts and relativity of concepts. The semantic structure model is formulated as extended hypergraphs, and can handle properties of information mentioned above.

1 はじめに

情報は、物質、エネルギーとともに現実世界を構成する重要な要素である。しかしながら、物質、エネルギーの構造や性質の解明に比べ、情報の構造や性質は、未だ基本的な部分において十分に解明されているとは言い難い。

情報自体に関する研究は、これまで人文科学の分野を中心に行なわれてきたが、近年における計算機の開発によって、これまで様々な媒体によって形を変えていた情報が計算機内部においてある程度統合して格納されるようになり、これまで困難であった情報の観測、実験が記号处理的に可能となった。これにより、情報の研究は、計算機という観測、実験装置を用いた自然科学として成立するようになった。しかしながら、これまでの情報科学においては、情報それ自体を対象とするより、道具としての計算機を対象とした処理方式を中心にする傾向にあった。そのため、情報に関する研究において本来最も重要であるはずの情報の意味を十分に考慮しない結果となった。

現在、データベース、ハイパーテキスト、情報検索、自然言語理解、人工知能、マルチメディアなどの研究において意味の記述精度の向上や処理機能の高度化が急速に議論されるようになってきた。データベースの分野では、リレーショナルモデルは、複雑な意味関係を扱うには柔軟性が不十分であるという認識によりオブジェクト指向モデルが注目を集めている。ハイパーテキストの分野では、フルテキストに対して、意味的に関連する情報を結び付け、ナビゲーションなどの機能により必要な情報をたどれるということが関心事の一つとなっている。情報検索の分野では、検索のためのシソーラス構築という観点から一次情報の主題に関する検索語の切りだしと体系化の研究が行なわれている。自然言語

†筑波大学 電子・情報工学系

†Institute of Information Science and Electronics, University of Tsukuba

理解の分野では、機械翻訳を中心に意味辞書の構築が行なわれている。人工知能の分野では、類推、帰納推論、学習など主に処理方式を中心に研究が行なわれている。

このように、意味の問題に関して様々な研究が行なわれているが、これらを含む意味に関する理論的枠組は確立されていない。そこで、本研究では、データベース、知識ベースを統合、拡張した自己組織型情報ベースシステムの開発の一環として、その理論的枠組である意味構造モデルと呼ばれる情報モデルの構築を行なった。

情報ベースシステムに格納される情報としては、一次情報、二次情報、意味関係情報、構造化情報などがあり、これまでデータベースや知識ベースに部分的、潜在的に格納されていた情報および、それらに収録できなかった情報を併せて自己組織的に統合し、意味処理に必要な情報を網羅的に顕在化して蓄積している。これらの情報に対して、検索や演繹推論だけでなく、類推、帰納推論、仮説推論などの高度な処理を行なう。対象は、実際の研究開発に使用されている情報であり、本稿では、具体的な例として非線形光学材料を用いる。情報ベースシステムによって情報を資源として活用することが可能になる。

本稿は、情報の解析により、情報の基本単位である概念の内部構造、相対性、概念間の重なりなどの特性について述べ、これらの特性によって生じる問題点と、これらの問題に対処するための意味構造モデルについて述べる。

2 情報の特性

情報の性質は、非常に複雑であり、その全てを列挙するのは困難であるが、従来から情報の取り扱いにおいて問題となっている基本的性質について述べる。

現実世界の情報空間の特徴は、概念の非可算性である。すなわち、プリミティブな概念は存在せず、概念は、一般に内部構造を持ち、複数の概念から構成されている。構成要素の各概念もまた内部構造を持ち、さらに複数の概念から構成されている。また、概念は、独立して存在するのではなく、様々な意味関係により関連づけられ、相互に重なりを持ちながら存在する。さらに、概念の構造、関係は相対的であり、一つの概念が、実体、属性、関係のいずれにもなりうる。つまり、概念は静的に存在するのではなく、目的や視点にしたがって、動的に個々の性質や全体の構造を変化させる。概念は主に用語によって表現される。

以下では、概念の内部構造、部分的重なり、相対性という情報の特性に焦点をしばって述べる。

まず、概念の内部は、次のような構造を持つものとして考える。

1. 属性: 概念は、その性質である属性的概念を持つ。
2. 抽象概念: 概念は、関連する概念を抽象した概念として持つ。
3. 階層: 概念は、上記の概念を階層構造として持つ。

次に、概念間の重なりとは、複数の概念が、内部構造を構成する概念を互いに共有することである。概念の類似性は、共有される構成要素の概念によって評価できる。

次に、概念の相対性とは、個々の概念の内部構造や全体の構造が視点や目的に応じて変化することを意味する。

1. 実体、属性、関係の相対性

ひとつの概念に対して、ある視点で実体として扱っていても、別の視点では、属性あるいは関係として扱うことがある。

2. 階層の相対性

ある目的で階層構造を形成している情報空間も利用目的や情報空間の大きさの変化に応じてその階層構造も変化する。たとえば、最下位の階層にあったものがさらに下位の階層を持ち、従って、新しい最下位から見ると上位の階層になる。

3. 限定による相対性

概念は、内部構造を持つが、常に内部構造全体を処理するわけではない。視点や目的に応じて限定の仕方が変わり、処理する内部構造が変化する。

3 意味構造モデル

意味構造モデルは、概念の内部構造、概念間の重なり、概念の相対性という情報の特性を考慮したモデルである。

意味構造モデルは、意味構造グラフと呼ばれる、拡張ハイパーグラフを基にしたネットワーク型のモデルである。

基本のハイパーグラフに対して、ラベルづけ、有向性、入れ子構造という点で拡張を行ない、意味構造グラフを定義する。ノードおよびリンクを入れ子にすることによってノードとリンクの区別はなくなり、これらを統一して扱うことができる。そこで、これらを一括してハイパーリンクと呼ぶ。

意味構造グラフは、リンクとノードの区別はなく、すべてハイパーリンクである。概念は、意味構造グラフ上で部分グラフとして定義され、構造体として処理される。また、各ハイパーリンクは、ラベルと状態を持つ。ラベルは、同義語集合であり、状態は、基点、内部、外部のいずれかの状態をとる。基点状態のハイパーリンクは、実体的概念に相当し、内部状態のハイパーリンクは、属性的概念に相当し、外部状態のハイパーリンクは、関係的概念に相当する。これらの状態は、視点や目的に応じて変化する。

以下に定義を示す。

定義 ハイパーグラフ

$X = \{x_1, x_2, \dots, x_n\}$ を有限集合とする。

X 上のハイパーグラフとは以下の (1),(2) の条件を満たす

X 上の部分集合族 $H = (E_1, E_2, \dots, E_m)$ である。

$$(1) E_i \neq \emptyset \quad (i = 1, 2, \dots, m)$$

$$(2) \bigcup_{i=1}^m E_i = X.$$

ここで x_1, x_2, \dots, x_n をノードと呼び、 E_1, E_2, \dots, E_m をリンクと呼ぶ。

定義 ハイパーリンク

X を有限集合とする。

ハイパーリンク hl は以下のように帰納的に定義される。

1. $x \in X$ は hl である。
2. $\{m \mid m \text{ は } hl \}$ は hl である。
3. 1、2 だけが hl である。

定義 意味構造グラフ

$$SSG = (HL, \psi)$$

ここで、

HL : ハイパーリンクの集合

$$\psi: HL \rightarrow HL$$

である。

ここで、ハイパーリンク hl の像 $\psi(hl)$ をハイパーノードと呼ぶ。

定義 ハイパーリンクの構成

$$\phi: HL \rightarrow N \times L \times S$$

ここで、

N : ハイパーリンクナンバー

L : ラベル

S : 状態

である。

ラベルは、同義語集合、状態は、基点、内部、外部のいずれかの状態を持つ。

4 情報の構造化

意味構造グラフと他のアプローチの最も違う点は、リンク主体であること、すなわち概念の関係の記述を詳細に行なうという点である。処理する概念が増加するにつれて、概念の記述だけでなく概念間の関係を詳細に記述しなければ複雑な意味関係を扱うことはできない。そのために情報の構造化が必要となる。構造化を行なうことによって複雑な意味関係を明確な処理対象にすることができる。

この章では、意味関係の構造化を行なう方式と、構造化された意味構造上での操作機能について述べる。

情報の構造化は、次のような手順で行なう。

まず、意味構造グラフを構築する。

1. ハイパーリンクの生成
2. ハイパーリンクのラベルづけ
3. ハイパーリンクの結合
4. ハイパーリンクの状態設定

次に、意味構造グラフ上に概念を形成する。

1. 基点ハイパーリンクに対して、結合している全ての内部ハイパーリンクと基点ハイパーリンクを結合テーブルとして構成する。
2. これを全ての基点ハイパーリンクに対して行なう。

ここまでで、内部構造を持つ概念が、外部ハイパーリンクで関係づけられ、互いに重なりを持ちながら存在することになった。

構造化された意味構造グラフに対して、次のような構造操作が可能である。

1. 概念の検索
2. 概念の属性を知る
3. 概念間のナビゲーション

上記の構造操作を複合して高度な処理を行なえる。

構造化された意味構造グラフでは、概念間には、様々な関係があるわけだが、目的に応じて注目する関係は、関係のうちで一部分だけである。例えば、上下関係だけ、あるいは因果関係だけに注目してナビゲーションを行ないたいことは研究開発の過程でよくあることである。その際には、特定の関係だけに注目した結合テーブルを構成することによっ

て不要な関係を見ることなしにナビゲーションを行なえる。ナビゲーションの途中で、ある概念の属性を知ることも可能である。

構造間をナビゲーションすることは、単に関連する情報をたどることではない。ハイパーリンクの関係情報や結合に応じて、ナビゲーション自体が情報操作を意味する。たとえば、因果関係をたどることは、演繹推論または仮説推論を意味し、特定の関係情報および共有する内部構造に注目してナビゲーションを行なうことは、複数の概念がどのような意味でどのくらい類似しているかということ、ナビゲーションのパスの経路や距離で評価することを意味する。

また、ある視点で内部構造として持っていた概念を、別の視点で外部の概念として扱いたいこともよくある。その際には、ハイパーリンクの状態を内部から外部に変えることによって、視点の違いによる相対性を扱うことができる。

図1は、非線形光学材料を対象とした意味構造グラフの例である。

非線形光学効果と光学材料それぞれの階層関係を示している。ここで、四角で囲んだハイパーリンクは、基点状態のハイパーリンクであり、太線の丸で囲んだものは、外部状態のハイパーリンクで、細線の丸で囲んだものは、内部状態のハイパーリンクである。矢印は、ハイパーリンク間の意味的方向性を表す。有機高分子を概念とすると、その属性は、軟らかい、融点が低い、非対称性となる。また、有機高分子に非対称性の性質が生じると光カー効果が生じることを示している。さらに、有機高分子と有機材料の間には isa の関係があり、外部状態のハイパーリンクとなっているが、状態を内部にすると、有機材料は有機高分子の属性であるという見方ができる。

5 考察

この章では、意味構造モデルと他の分野との意味処理方式の比較を考察する。

データベースの分野におけるデータモデルは、現実世界の射影として位置付けられている。データの三大モデルは、階層型、ネットワーク型、リレーショナル型であるが、最近では、オブジェクト指向型データモデルの研究が行なわれている。

ここではまず柔軟性と構造化機能で注目されているオブジェクト指向型データモデルについて考察する。オブジェクト指向モデルは、特定のモデルを指すのではなく、様々なモデルの総称であるため、一概には定義できないが、だいたい以下のようなことが共通している。

現実世界の事象それぞれをすべてオブジェクトとみなし、事象とオブジェクトを一対一に対応させる。オブジェクトには、クラスオブジェクトとそのクラスに属するインスタンスオブジェクトがあり、クラスオブジェクトがインスタンスオブジェクトの性質を規定する。クラスオブジェクトの階層においては、下位のオブジェクトは、上位のオブジェクトの性質を継承する。各オブジェクトは、他のオブジェクトからのメッセージによって

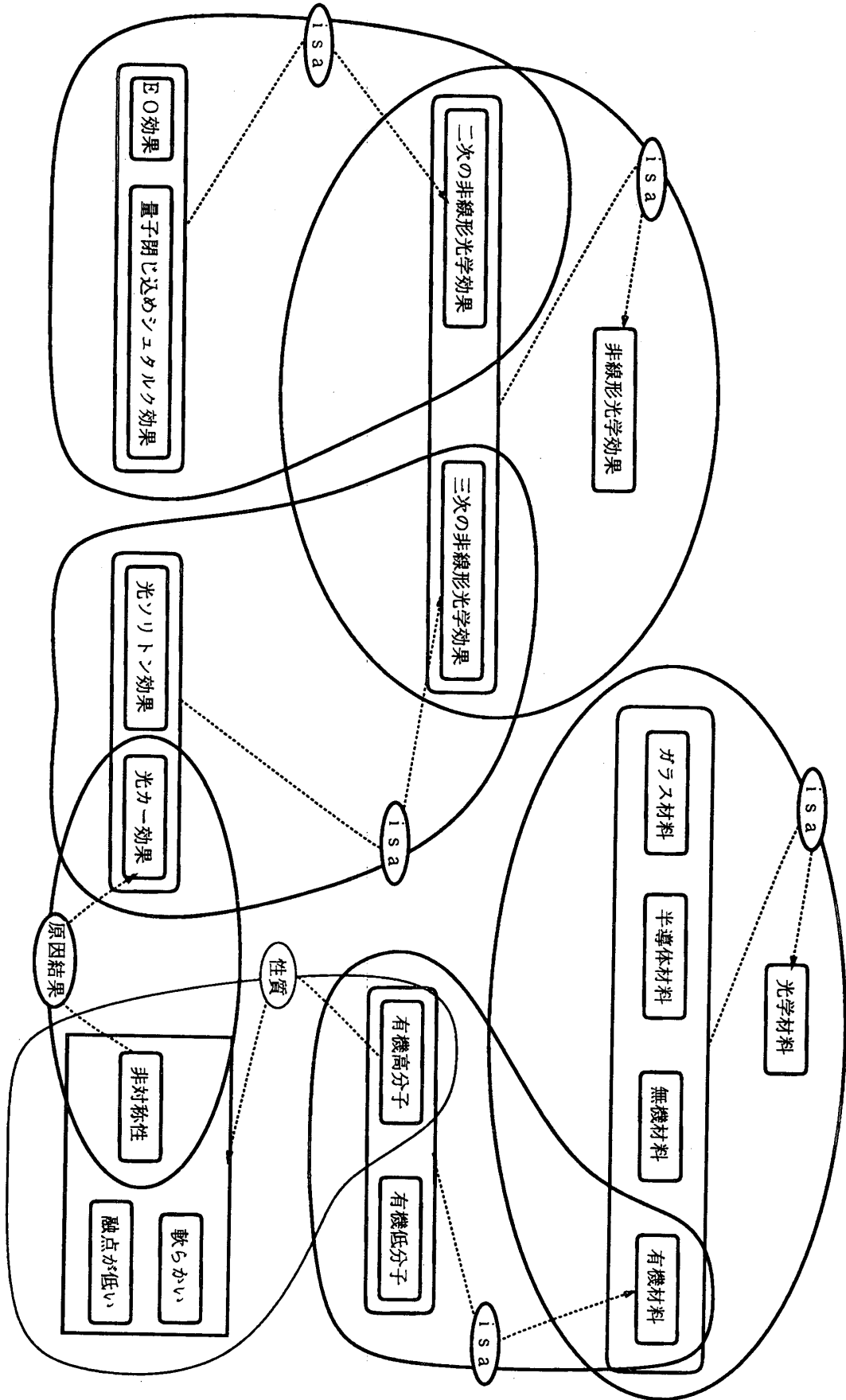


図1 意味構造グラフ

その振舞いを決定する。また、各オブジェクトはカプセル化されているため、他のオブジェクトの性質を直接に参照することができない。

オブジェクト指向モデルは、現実世界と一対一の対応がとれるはずであり、情報空間を直截にモデル化できると考えられているが、情報の特性を考えた場合には、次のような問題点がある。

1. 各オブジェクトは独立して存在するため、個々のオブジェクトの性質は記述できるが、オブジェクト間の関係を明示的に記述することは困難である。
2. 各オブジェクトはカプセル化されているため、複数のオブジェクトの重なりを直接参照することができない。
3. オブジェクトの性質は静的なものであるため、視点や目的による相対性は扱うことができない。
4. クラスとインスタンスにわかれているため、インスタンスを独立に生成することはできない。また、クラスとインスタンスの間の関係の処理はできない。
5. 継承においては、多重継承を行なう場合に衝突が起こる。多重継承を許さない場合では、オブジェクトの性質は限定されてしまう。
6. オブジェクトの識別が困難である。

これに対して、意味構造グラフにおける概念は、相互に関連して存在している。そのため、概念の属性、概念間の関係、類似概念を直接扱うことができる。また、意味構造グラフは様々な相対性を扱うことができる。

フルテキストデータベースと関連が深いハイパーテキストは、意味的に関連する情報を結び付け、ナビゲーションなどの機能により必要な情報をたどることができる。

ハイパーテキストには次のような問題点がある。

1. ハイパーテキストのネットワーク構造は、平坦であるため、単純な意味づけしか行なえない。
2. 関連する情報が複数リンクされている場合、どれをたどればよいか決定するのが困難である。

意味構造グラフは、ハイパーテキストのリンクに比べて関係情報が豊富でかつ結合が直接意味に対応した構造になっているため、複雑な意味関係を処理することができる。

情報検索の分野においてシソーラスの構築は必須である。情報検索のためのシソーラスは、統制語という意味あい強いが、検索語の意味関係を記述しているとみなすことが

できる。シソーラスの基本的な記述項目は、同値関係、階層関係、関連関係であるが、このようなシソーラス構築の際には、情報の特性を考えた場合に次のような問題点がある。

1. 視点によって階層関係における用語の位置が異なるが、どこに記述するかを決定するのは困難である。
2. 階層関係には種類があるが、区別して記述することは困難である。
3. 関連関係には様々なものが考えられるが詳細に記述することは困難である。

検索用シソーラスは、概念の内容の関係ではなく表現の用語の関係の記述になっている。そのため、階層関係、関連関係の詳細な記述が困難であるが、意味構造グラフでは、これが可能であり、意味関係に対する様々な操作が可能である。

自然言語理解の分野における意味辞書は、初期には、動詞に対する格を付与したものであったが、現在では、その考えをさらに押し進め、意味素を各単語に付与することによって辞書を構築している。このような意味辞書の構築には、次のような問題点がある。

1. どのような意味素を付与するかを決定するのが困難である。
2. どのくらいの数の意味素を付与すればよいのかを決定するのが困難である。
3. 意味素も言葉である以上、意味素の表現に多義性、多様性が生じる。

意味辞書は、ひとつひとつの言葉を単位として意味や関係する言葉が記述されているが、それは意味についての断片的な記述にとどまっている。意味構造グラフでは、断片的な記述ではなく構造として意味を捉えているため、情報空間全体からみた意味を処理することができる。

人工知能の分野において、類推、帰納推論、学習の研究が盛んに行なわれている。ここでは、特に類推について考察する。

類推とは、類似性に基づく推論であり、最も重要な問題は、類似性の評価である。

人工知能の類推研究においては、次のような問題点がある。

1. 類似概念を検出するのが困難である。
2. 類似性の定量化を行なうのが困難である。
3. 類似性の意味規定を行なうのが困難である。

意味構造グラフにおいては、概念に内部構造をもたせているため類似性の評価が容易である。

以上、他の方式との比較を行ない、意味構造モデルによる意味処理方式の利点を述べた。最後に、意味構造モデルを基にした情報ベースシステムの自己組織化について述べる。

情報ベースシステムに蓄積されている情報は、複雑かつ大量である。そのため、構造化処理は自己組織的に行なう必要がある。すなわち、情報自身に基づいて構造化処理を自動的に行なう。大量の情報の構造化処理は、人間には極めて困難であり、計算機による処理が必須である。自己組織型情報ベースシステムは、大量の情報に対する構造化の自動処理が可能である。

6 結論

本稿は、概念の内部構造、概念の相対性、概念間の重なりなどの情報の特性を解析し、それに対応した意味構造モデルと情報の構造化について説明した。また、意味構造モデルによる意味の表現および処理方式とその応用の方向を示した。

参考文献

- [1] Y. Fujiwara, N. Uda, and X. Zhang. "Analogical Reasoning in Polymer Information-Base Systems". *the 13th CODATA*, October 1992.
- [2] 藤原 譲. "情報知識学試論". 情報知識学会誌, 1(1):3-10, 1990.
- [3] C. Berge. "*Hypergraphs*". North-Holland, 1989.
- [4] H. Boley. "Directed Recursive Labelnode Hypergraphs: A New Representation-Language". *Artificial Intelligence*, 9(1):49-85, 1977.
- [5] H. Yasukawa and K. Yokota. "Labeled Graphs as Semantics of Objects". Technical Report TR-600, ICOT, 1990.
- [6] S.B. Zdonik and D. Maier. "*Readings in Object Oriented Database Systems*". Morgan-Kauffman, 1990.
- [7] F.W. ランカスター著 松村多美子, 鈴木祐滋訳. "情報システムのためのシソーラスの構築と利用". 情報科学技術協会, 1989.
- [8] 長尾 眞. "自然言語における意味処理". 情報処理学会 情報学基礎研究会, 92-FI-28, 1992.
- [9] J.W. Shavlik and T.G. Dietterich. "*Readings in Machine Learning*". Morgan-Kauffman, 1990.