

## M. 構造化知識に基づく意味理解および仮説生成の機構

1. テーマ 構造化知識に基づく意味理解および仮説生成の機構  
The Mechanisms of Semantic Understanding and Abduction Based on Organized Knowledge
2. 研究代表者 情報科学科 藤原 謙
3. 共同研究者 (学内) 情報科学科 後藤 智徳、中山 克  
(学外) 井上 佳久 (ミドリ十字)、清水 恵次郎 (出光石油化学)  
菅原 駿吾 (NTT)、福柁 喜孝 (吉富製薬)、松井 孝司 (三共)、  
松浦 育敏 (中外製薬)
4. 研究期間 1997年4月1日～1998年3月31日
5. 研究の目的

国際的情報化の進展により生産、提供、流通、利用される情報は量的にも質的にも多種、多様である。しかし豊富なデジタルコンテンツも高度活用のための機能が未整備なため十分な活用からは程遠いのが現状である。そのためには計算機が学習した知識に基づき情報内容を理解、評価、選択することが必要である。本来情報は大量であればあるほどそれを適切に選択・活用することにより、新しい情報を生成できるところに真の価値がある。そのために新しい環境に則した情報資源の構築法、利用機能、とくに科学技術における新理論の確立に重要な仮説生成の機構およびその実証が本研究の直接の目的である。

また新しいモデルに基づく高度な情報資源の構築によりこれからの計算機、通信の理論や技術の展開の方向を示すことはこれまで行って来た研究と共通の目的である。

[キーワード] 多様化情報、情報解析、品質評価、内容理解、仮説生成、思考支援  
diversified information, information analysis, quality evaluation,  
content understanding, abduction and thinking support

### 6. 研究概要

#### 6. 1 はじめに

情報基盤の整備が進み電子図書館、電子博物館をはじめとしてインターネット上に、形式的にも内容的にも多様で多量の情報が地球規模で分散して集積されている。これまでのデータベースは、情報を集積してその中から必要なものをキーワードを手がかりとして取り出して利用する方式が主であった。さらにインターネットによりデータの収集が容易になり情報倉庫 (Information Warehouse) としてその感じが一層強くなっている。しかしそれらは、検索のみでなく各種の目的に利用されることになり、共通する基本的課題は豊富な情報資源を有効に活用するための思考支援など高度な処理機能である。それには意味の理解・処理が必要であり、そのためには、多様かつ豊富な意味関係を含む概念構造の自己組織化が有効な手段である。従来のいわゆる「意味処理」は主として機械翻訳や情報の検索などの応用に留まり、意味関係としてはシソーラスによる階層関係、関連関係、同義関係と翻訳用の用語辞書があるが、いずれも広範な意味を詳細に記述していない。また演繹推論に用いられている二値論理は閉じた世界に対応し、問題解決や発想などの情報生成を伴う開いた世界の問題に対応できないので多値論理導入が必要である [1-5]。すなわち新しい情報環境に対しては新しい情報資源と、新しい利用機能が必要である。さらにこのことは逆に新しい情報向きの計算機、ソフト、通信などの環境を展開させる指針を示すことにも

なる。

本研究で述べる自己組織化型情報ベースシステムでは、情報の蓄積、管理、検索だけでなく、類推、帰納推論、仮説推論、発想、連想などの意味処理も対象としている[10-16]。ここでは多様な目的に応用できる概念構造の構築を目的として、概念構造は十分な意味関係を含み、その記述項目と表現形式及び概念記憶構造の自動的な構築方法についても述べる。また構造化された情報に基づく意味処理、とくに思考支援で最も重要であるが従来の方式では直接処理ができなかった類似関係や関連関係の意味や程度を扱う方式についても概要を示す。

なお意味関係の抽出の際、表層的な多様性、曖昧性を適切に扱うには複数言語の表現を用いるのが有効であるので多言語情報を対象とする。このことはまた情報流通の面からも多数の言語にわたる情報をどの言語からでも横断的にアクセスできるので非常に好都合である[11, 12]。

## 6. 2 学習型情報資源化モデル

概念構造は概念間の意味関係に対応して階層関係の他、部分的重なり、多項関係、再帰構造、内部構造、相対性、動的関係などを含み、グラフの構造では対応できない[3-5]。そこで、ハイパーグラフ[8, 9]の多項関係や双対性を更に拡張した相対性(概念—関係、概念—属性など)、その他の関係に対応できる概念記憶構造である均質化2部グラフモデル(Homogenized Bipartite Model: HBM)に基づき、記述する[4, 5]。

$$E = 2V \quad (1)$$

$$V = V \cup E \quad (2)$$

$$E = E \cup V \quad (3)$$

$$\sigma : L \rightarrow V \cup E \quad (4)$$

ここで  $V, E, L$  はそれぞれノード、リンク、ラベルの集合(有限とは限らない)である。

このモデルでは関係は、ノード2つだけを結ぶのではなく任意の数の集合即ち多項関係に対応し、式(1)に示すように冪乗集合であることを示し、これだけであればハイパーグラフになる。式(2)と(3)でそれぞれ再帰構造と内部構造が許される。また(2)と(3)の両者を併せてノード  $V$ (概念)とリンク  $E$ (関係)は基本的に均質化されるが、ある時点では構造の最下部にはノードとしての役割を持つもの、最上部にはリンクの役割だけのものが存在する。一方ノードが整礎的集合(well founded set)ではなくなる。

また階層関係の属性継承関係を辿るナビゲーションは大部分の演繹推論に相当し、しかも単一化の計算量を著しく小さくすることが可能となる。さらに多値論理のうち実用性から最近注目されている Rough Set Theory [1, 2]における識別不能関係のみならず論理的帰結不能関係への拡張も可能となる。

## 6. 3 学習された知識に基づく意味理解

概念の意味内容を十分に記述することは外延型の関係DB、オブジェクト指向DB、単調論理型の知識ベースなどでは極めて困難である。そもそも必要な意味関係を網羅的に獲得することが問題である。

しかし概念の階層関係が構造化されていれば属性継承が可能で、しかも冗長度を最低限度に抑えられるので記憶容量のみならず管理上も好都合である。論理構造その他の関係も構造化により処理が簡単になる。特に演繹推論で問題となる単一化の計算量の大きさを気にしなくて済む。

情報の内容を十分理解するための概念記憶構造における意味関係の内容及び表現形式について述べる[4, 5]。概念構造の主たるものの一つは階層関係である。専門用語の階層関係には、事物所属の視点から形成する属階層関係と主題分野の視点から形成する分類階層関係があり、従来、それぞれの階層の関係によって互いに独立した概念体系が構成される。一方、意味関係を全面的かつ明確に表示するために、is-a や part-of など複数の階層関係を含む概念体系が必要である[4-7]。

また 命題論理、一階述語論理等において扱えない多値の問題も記述、表現の制約を受けなけ

れば対処可能となる。ただこのことは当然新しい推論方式を必要とするので次章で述べる。

その基礎となる類似関係、関連関係は問題解決で最も多く使われる類推や帰納推論に必須であるが、これらは十分な意味理解に基づく処理が必要で、それは網羅的な意味関係を含んでいる概念構造が利用できる。基本的には類似または関連の意味とその程度即ち定性的および定量的に処理を行うことになる。

このように、概念記憶構造が多重継承の階層関係を含み、意味内容の外延的記述の困難と冗長性を同時に解消できる。

検索用シソーラスでは階層関係以外の意味関係を全て単に関連関係として記述する。因果関係など論理関係の詳細な記述は類推、帰納推論、仮説推論などの応用にとって不可欠であるが、漠然とした関連関係では方向もなく、詳細な記述も無いので推論には使えない。このため、概念構造には関連特徴を記述するラベル付きの関連関係を構築する。過程—状態、実体—特徴などのラベル付け関連関係から、この中に概念構造の多項関係の表現例として、集合間の関連関係が示される。これらの関連関係を用いて論理関係の表現ができる。例えば、因果関係を原因と結果の関係として、例えば高分子の構造と物性の関係としても記述できる。

#### 6. 4 意味理解に基づく思考機能

演繹推論などで2値論理に基づく推論は記号処理として確立されているが、対象が閉じた世界の問題 (closed world problem) に限られるので思考支援に必要な開いた世界の問題 (open world problem) にはデータ、知識の生成を含めた新しい方式の確立が必要となる [1, 2, 4, 5, 11-16]。

これらは意味処理を含むので概念構造を用いることであれば類似関係、関連関係がその意味と程度を含めて機械処理が可能になる。

ここで基本の一つとなる類似関係は概念構造により処理できることは上で述べたが、もう一つのポイントであるところの新しい情報の生成もやはり概念構造が手がかりとなる。即ち概念の記述を基本としてさらにそれらを順次抽象化した階層関係が束 (lattice) 構造になることが利用できる。また概念構造の生成自体が類推を用いることもあり、手順が再帰的になる。

意味関係に対応して構造化された知識に基づく仮説生成の機構を下に示す。

$C=(V, E)$ , を知識世界とし,

$C_r=(V_r, E_r)$ ,  $C_s=(V_s, E_s)$ ,  $C_c=(V_c, E_c)$ , ここで  $r, s$ , および  $c$  はそれぞれ参照 reference、サンプル sample および共通 common 構造を示すものとする。

仮説生成の機構:

$$C_c \subseteq C_{s1} \cap C_{s2} \cap \dots \cap C_{sn} \cap C_r,$$

$$C_s = (V_s, E_s) = C(V_r, E_r),$$

$$\text{ここで } V_r = V_c + \delta V_r, E_r = E_c + \delta E_r,$$

$$V_s = V_c + \delta V_r, E_s = E_c + \delta E_r.$$

類推および帰納推論の機構:

$$C_c \subseteq C_s \cap C_r,$$

$$\text{ここで } V_r = V_c + \delta V_r, E_r = E_c + \delta E_r,$$

$$V_s = V_c + \delta V_r, E_s = E_c + \delta E_r.$$

知識構造の操作からみると、仮説生成と同様に情報生成を伴う類推や帰納推論の機構は共通の形式の処理となることが興味深い点である [17]。

自動抽出された断片的な意味関係の構造化による自己組織化概念構造システムの概要と応用例として以上の手法を組み合わせる高分子のNMRの情報をも自己組織的に概念構造を構築して、それを高分子のNMRスペクトルの解析に応用するためのシステムの例を示す。概念構造部分は主としてC-T R A Nで同値関係を抽出し、S S - K W I Cで階層関係と関連関係を抽出し[11,12]、SS-SANSで因果関係を抽出し[13,14]、高分子の複雑な構造に対応する非常に数多い類似概念多重継承を含む構造になっている。これを帰納推論に用いスペクトルの帰属に適用できる[2,16,17]。

## 6. 5 むすび

広範な情報源から得られる質的量的に多種多様な情報を有効に活用するために概念間の意味関係を解析し、多様な目的に対応できる概念構造の特性を明らかにし、そのモデルとして概念記憶構造に対応する均質化2部グラフモデルを示した。それに基づき概念を表現する用語間の意味関係から専門知識の自己組織的構築方法つまり情報の意味内容を計算機に学習される機構を示した。概念構造に、多様なラベル付き階層関係、関連関係、内部構造、相対性などの複雑な意味関係を含ませることでこれまで困難であった類似関係、関連関係を定性的にも定量的にも対処できることを示した。それに基づき、多値論理を含む仮説推論は類推、帰納推論と同様に、構造化知識から情報を生成するプロセスとして発想を実現し、それらを結合して問題解決、意思決定、評価などの思考機構を明らかにし、新世代型情報資源のあり方を示した。

このことはさらに新しい情報環境即ち新しい計算機機構や通信技術の展開の基本的方向付けにも資するものである。

## 参考文献

- [1] Z.Pawlak, "Rough Sets" Int. J. of Computer and Inf. Sti. Vol.11 pp341-54, 1982
- [2] Z.Pawlak and W.P.Ziarko "Rough Sets, Fuzzy Sets and Knowledge Discovery" Kluwer Academic Publ. ISBN 0-7923-1472-7
- [3] Fugmann, Robert: "Theoretische Grundlagen der Indexierungspraxis". 1992.
- [4] Yuzuru Fujiwara: "The Model for Self Structured Semantic Relationships of Information and Its Advanced Utilization". International Forum on Information and Documentation, vol. 19(2) PP 8-10, January 1994.
- [5] Yuzuru Fujiwara and Hironobu Gotoda: "Representation Model for Relativity of Concepts". International Forum on Information and Documentation, vol. 20, 1, PP 22-30, January 1995.
- [6] 藤原譲、劉野「電子図書館の高度利用のための機能」情報学基礎研究会(44-6)1996
- [7] 藤原譲「用語間の意味関係の自己組織化」脳と用語シンポジウム論文集(日本学術会議、1996-12-12)
- [8] C.Berge. "Hypergraphs". North-Holland, 1989
- [9] H.Boley. "Directed Recursive Labelnode Hypergraphs: A New Representation-Language". Artificial Intelligence, 9(1):49-85, 1997.
- [10] Patric H. Winston, "Learning and Reasoning by Analogy" Communication of the ACM, 23(12), pp689-703, 1980
- [11] J.Lai, H.Chen and Yuzuru Fujiwara: "Extraction of Semantic Relationships Among Term - SS - KWIC". Proceedings of The 47th FID Conference and Congress, Oct. 1994.
- [12] Jinjuan Lai, Hanxiong Chen and Yuzuru Fujiwara "An Information-Base System Based on the Self-Organization of Concepts Represented by Terms" Terminology Vol.3(2), 313-34, 1996
- [13] Hikomaro Sano and Yuzuru Fujiwara "Syntactic and Semantic Analysis of Article Titles in Analytical Chemistry" J. Inf. Sci. Principles and Practice 19, 119-124, 1993
- [14] Hikomaro Sano and Yuzuru Fujiwara "Automatic Assignment of Word Categories for Improved Facet Analysis of Titles and Indexes" J. Inf. Sci. Principles and Practice 20, 23-31, 1994

[15]Hironobu Gotoda, Jianghong An, and Yuzuru Fujiwara: ''Analogical Reasoning of Organic Reactions Based on the Structured Compound-Reaction Diagram''. International Conference on Forum and Applied Practical Reasoning, FAPR' 96, PP 276-290, Germany, June 1996.

[16] Yuzuru Fujiwara, Jianghong An and Jinjuan Lai "Materials Information Systems for Open World Problem Solving" Computerization and Networking of Materials Databases" Vol.5, ASTM-STP 1311, 273-278, 1997

[17] Yuzuru Fujiwara and Ye Liu, "The Homogenized Bipartite Model for Self Organization of Knowledge and Information", The Proceedings of the 48<sup>th</sup> FID Conference and Congress (23-25, Oct. 1996 in Graz), to be published in IFID 1998