

# 情報の構造化 - 学習・思考機能の実現に向けて -

2P-9

森本 貴之<sup>†</sup>, 真栄城 哲也<sup>‡</sup>, 藤原 譲<sup>†</sup>  
 神奈川大学 理学部<sup>†</sup>  
 ATR 人間情報通信研究所<sup>‡</sup>

## 1 はじめに

最近の計算機の高速化、大容量化と低価格化は目を見張るものがある。またそれにともない、インターネットの普及による情報化も加速度的に進んでおり、情報の発信や利用は急速に多種、多様、広域化している。しかしながら、現在普及している計算機の根底にあるものは数値計算やキーワード検索、演繹推論であり、豊富な情報や知識の内容を効率よく活用できるとは言難い。実際、情報の意味に関する高度な機能に対する要求も強く認識されるようになってきており、その代表的な機能としては学習・思考機能等が挙げられる。

人間が持っている高度な学習・思考機能を実現するためには、単に膨大な情報を記憶させるだけでなく、その意味を理解させる必要がある。そしてそのためには意味関係が表現できる構造化が要求される。しかしながら、情報の本質である特性、意味関係、構造などを解析し、それらの原理やモデルを体系化する手法は整備されていない。そこで、本論文では学習・思考機能実現に向けた情報の構造化に関する研究について述べる。

## 2 情報の構造化

前章でも述べたように情報・知識を有効に活用するためには符合や数値だけでなく、意味などを含めた面から理解する必要がある。そしてそのためにはまず以下に示す3点を実現しなければならない。

1. 情報の特性とくに意味関係の解析
2. 属性、特徴、意味、構造に関する基礎理論の確立、利用技術、手法の開発：体系化
3. 各分野の情報への具体的な応用のためのアルゴリズム、システムの整備

また、情報の意味内容は媒体を通して表現された文字や記号等を解釈するといった間接的な方法をとらざるを得ない。科学や技術の分野においては、用語、特に専門用語は抽象概念を表現する最も便利かつ強力な媒体である。そこで、概念を表現する最小単位として用

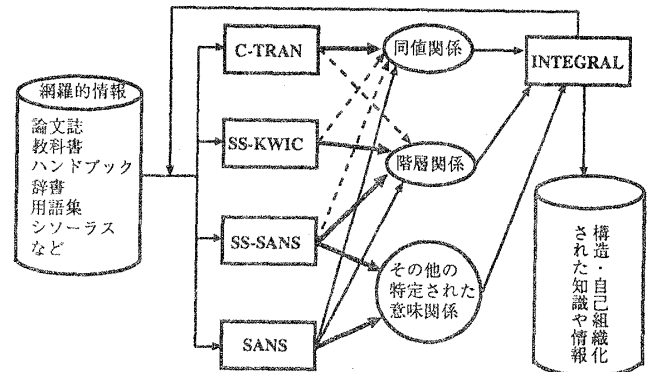


図 1: 情報・知識の自己組織化システム

語を取り上げ、この用語の体系化を行なう。

このような用語の体系化において、目的に対応した構造化を行なうためには多項関係が取り扱えなければならない。さらには入れ子構造や様相性、相対性等も表現可能でなければならないことが用語の意味関係に関する解析から得られており、ハイパグラフですら十分とはいえない。そこで、ハイパグラフに拡張を施し、思考機能に対応できる柔軟で意味関係を記述可能なモデルとして均質化2分グラフモデル (Homogenized Bipartite Model: HBM) を提案している。[1] このHBMを用いた類推、機能推論、仮説推論のメカニズムはすでに文献[2]で示している。

図1に現在開発を進めている用語を基にした概念間の各種関係を自動的に統合、調節するためのシステムの概略を示す。また、図中の各機構はそれぞれ以下に示す処理を行なう。

- C-TRAN 法 (Constrained Transitive Clousure): 同値関係 (同義語) および階層関係 (上位語、下位語) の抽出 [3]
- SS-KWIC 法 (Semantically Structured Key Word Element Index in Terminological Context): 階層関係および関連関係の抽出 [3]
- SS-SANS 法 (Semantically Specified Syntactic Analysis of Sentences): 各種意味関係の抽出 [4]
- SANS 法 (Semantically Analysis of Sentences): 意味解析
- INTEGRAL 法: 全体の構造・統合化

このシステムでは網羅的な情報を入力とし、上述の各機構が協調動作しながら、抽出した各種意味関係に基づ

Construction of Structuralized Information - for Realization of Machine Learning and Thinking -

Takayuki Morimoto<sup>†</sup>, Tetsuya Maeshiro<sup>‡</sup>, Yuzuru Fujiwara<sup>†</sup>

<sup>†</sup>Department of Information Science, Kanagawa University

<sup>‡</sup>ATR Human Information Processing Research Laboratories

いた情報・知識の構造・自己組織化を自動的に行なう。以下の章では、階層・関連関係の抽出を行なう SS-KWIC 法について実際の抽出結果と合わせて述べる。

### 3 SS-KWIC 法

SS-KWIC 法は専門用語の構成規則に基づいて、複合用語を基本構成用語に分解し、相互の関係を解析することによって階層関係および関連関係を獲得する方法である。ここで注目する専門用語は以下のような特徴を持つ。また、曖昧さの排除や他の用語との区別を明確にするために、長い合成語になることが多い。

- ほとんどが名詞
- 後部分の語基の性質や状態を、前部分の語基が修飾または限定するなどの修飾関係が多い
- 用語は複数の語基を含むことが多い
- 同じ語基を持つ用語は、何らかの関係を持つことが多い

そこで、用語の構成規則は以下に示す合成語の構成規則に準ずるものと考えられる。

合成語 ::= 複合語 | 派生語

複合語 ::= 語基 + 語基 | 語基 + 連結要素 + 語基

派生語 ::= 接辞 + 語基 | 語基 + 接辞

語基 ::= 単純語基 | 複合語基

単純語基 ::= 単純語

複合語基 ::= 語基 + 語基

連結要素 ::= ・ | / | の | な

接辞 ::= 接頭語 | 接尾語 | 数詞 | 量詞

例えば、「意味処理」は語基である「意味」と「処理」から構成される複合語と考える。

以下に学術情報センターの「NACSIS テストセレクション」のデータを用いた結果を示す。このデータは多くの論文の題目と要旨から構成され、ここでは人工知能の分野のデータを用いる。階層・関連関係の抽出は以下の手順で行なう。

1. 日本語形態素解析システム “JUMAN” [5] を用いて構文解析
2. 用語の抽出
3. 階層・関連関係の抽出

図 2 に結果の一部を示す。各行が 1 用語に対応し、1 用語の中の空白は合成語の構成規則による区切りを、インデントは階層性を表す。したがって、図 2 は用語 “モデル” を共通の先祖とした階層・関連関係が抽出されていることを示す。

今回のデータからは約 18,000 個の用語の階層・関連関係が抽出されたが、現在の SS-KWIC 法の実装では、非常に単純な階層・関連関係しか抽出できない。実際には、今回の実装では抽出できない修飾関係がより複雑に入り組んだ用語も存在する。また、造語規則からだけでは見つけることのできない階層・関連関係も存在するが、C-TRAN 法や SS-SANS 法等と組み合わせ

モデル

知識 獲得 モデル

確率 モデル

学習 モデル

PAC 学習 モデル

反駁 PAC 学習 モデル

機械 学習 モデル

図 2: SS-KWIC 法による関係の抽出

ることで抽出することが可能になる。例えば “フーリエ変換” と “FFT” の関係は、“FFT” の対訳である “高速フーリエ変換” がわかっているならば抽出することができるが、これは同値関係 (C-TRAN 法) から導き出すことができる。

### 4 終りに

情報化が加速度的に進む現代において要求される計算機の新しい機能として学習・思考機能に着目し、その実現に向けての知識・情報の構造化に関する研究について述べてきた。本研究の特徴は意味関係に対応した情報の構造化されたものを知識資源とし、それによって類推や仮説推論といった思考機能の実現を目指す点であり、その中の一機能である用語の合成規則から階層・関連関係を抽出する方法について具体例を挙げて示している。また、情報の分野においては原理やモデルの一般的な体系化すら整備されておらず、本研究は一石を投じるものでもある。

現在、各種意味関係抽出を行なう機構のプロトタイプが完成している。今後は均質化 2 グラフモデルの実装を含めた各機構の統合と情報の構造化を行ない、その概念構造を基にした類推や仮説推論の実現を目指す予定である。

### 参考文献

- [1] Y. Fujiwara and Y. Liu, *The Homogenized Bipartite Model for Self Organization of Knowledge and Information*, IFID 2 (1), pp13-17, 1998.
- [2] 藤原 謙, 情報学基礎論の現状と展望 - 学習・思考機構と超脳計算機への応用 -, 情報知識学会誌, Vol.9, No.1, pp-13-29, 1999.
- [3] Y. Fujiwara and J. Lai, *An Information-Base System Based on the Self-Organization of Concepts Represented by Term*, Terminology, Vol.3(2), pp313-314, 1997.
- [4] H. Sano and Y. Fujiwara, *Syntactic and semantic structure analysis of article titles in analytical chemistry*, J. Inf. Sci. Principles and Practice 19, pp119-124, 1993.
- [5] <http://www.nagao.kuee.kyoto-u.ac.jp/nl-resource/juman.html>