

類推機能を持つ自己組織型情報ベースシステムの研究

張 曉冬* 宇陀 則彦* 藤原 譲**

*筑波大学工学研究科 **筑波大学電子・情報工学系

本論文は、自己組織型情報ベースシステムのモデルと設計について報告する。その目標は、(1)柔軟性のある情報モデルの導入、(2)動的な、自己組織型の情報構造化、(3)構造化情報に基づいての類推機能、である。その応用として、研究開発に高度な支援を提供するシステムの試作について報告する。

Research on the Self-Organizing Information Base System and its Analogical Reasoning Function

Xiaodong Zhang* Norihiko Uda* Yuzuru Fujiwara**

*Program in Engineering Sciences

**Institute of Information Sciences and Electronics

University of Tsukuba

We present in this paper, the theory of Self-Organizing Information Base system with a high flexibility information model, dynamic, self-organizing structuralization of information and analogical reasoning function based on the structured information. As an application of the theory, a prototype system that presents intelligent supporting to research and development is reported.

1 研究の背景と目的

近年、データベース技術の発展につれ、情報を蓄積、管理、検索する機能だけではなく、知識ベース、エキスパートシステムなど、演繹、推論などの機能までを含む、大量データプラス推論機能のシステムが数多く研究、開発されてきた。そこには以下にあげるような問題が存在する。

1. 情報の表現方法に柔軟性が欠けているため、大量な、変化している情報への対応能力が乏しい。それゆえ、システムの維持も非常に困難となる。
2. 情報に対する意味処理が充分に行なわれていないため、情報空間全体に対する把握能力が乏しい。結局、本当の意味での学習能力が足りない。
3. プロダクションルールなどのアプローチを取る場合が多く、クローズワールド的なシステムになってしまう。

これらの問題点に対処するため、われわれは次のような解決策を考えている。

- a. 表現モデルの柔軟性を強化するため、拡張ハイパーグラフによる新しい情報モデル。
- b. 情報の構造化。シソーラス、タキソノミーなど構造化用ツールの自動メンテナンスなどの技術の開発。情報の変化に対応し、動的な、情報自身に類する情報の分類、整理。すなわち、自己組織的な、情報の構造化メカニズムの確立。柔軟性のある情報モデルと合わせ、大量な、変化している情報にも対応できる仕組みの提案。

- c. 学習機能だけではなく、仮説生成、類推のような機能まで持つシステムの構築。

上記の解決策を実践するため、われわれは研究開発の支援を提供する自己組織型情報ベースシステムの研究を行なっている^[1]。ここでは、非線形光学材料の設計支援システムと有機合成反応設計支援システムを実例として取り上げる。

2 自己組織的な情報構造化

情報の構造化には、様々なものが考えられるが、本研究では次の三つの構造化を考える。

- ・ 物理構造化
情報を記録している位置関係を指す。一次情報におけるページ番号やインデックスの位置情報、書誌事項などの二次情報、構造化された情報における用語や記述の位置関係などである。
- ・ 概念構造化
概念間の様々な関係を指す。上下関係、同値関係、類似関係などである。概念構造化は、シソーラスとして記述する。
- ・ 論理構造化
文脈や状況に応じた概念間の関係を指す。因果関係はその一部である。論理構造化は、タキソノミーとして記述する。

この三つの側面から情報を構造化するための情報には、専門家が示唆する深い知識、専門分野を体系的に説明した事典的情報、専門用語を説明した辞書的情報、専門知識の背景となる基礎的情報、さらに、基礎となる一般常識などが考えられる。情報を構造化するには、これら全ての情報を利用するが、構造化する際の情報の重点のおきかたが三つの構造化によって異なる。この節では、題材として非線形光学情報を構造化

化するが、事典的情報と辞書的情報の情報源としてハンドブックを利用する。物理構造情報は、ハンドブックの目次及びインデックスから抽出する。概念構造情報は、目次、インデックス及びテキストから抽出する。論理構造情報は、目次、テキストから抽出する。

情報を構造化する手段も様々なものが考えられるがその大きなものに自然言語処理があげられる。形態素解析時の専門用語の抽出には、専門家が重要な概念をインデックス及び目次としてあげているのでそれを利用する。意味解析は、格文法に基づいた解析を行なう。自然言語処理の主な流れは以下の通りである。

1. 平仮名から漢字、記号、アルファベット、片仮名、句読点に変わるところを文節として切る。
2. 文節ごとに右方向最長一致法により動詞を切り出す。(辞書があると仮定)
3. 文節ごとに右方向最長一致法により助詞と助動詞を切り出す。(辞書があると仮定)
4. 用言の結合価^[4]により名詞の属性を決定する。(動詞の格フレーム辞書があると仮定)
5. 助詞を手がかりにして名詞と用言の関係を意味として記述する。
6. 接続詞、接続助詞、代名詞を手がかりにして文間の関係を記述する^{[5][6]}。

以下に、情報を構造化する手順を光非線形デバイスを例にとって説明する。

まず、目次を見てみると、光非線形デバイスという節が非線形光学材料の章の中にある。さらに、光非線形デバイスの節は小節にわかれている(図1参照)。この目次のページ番号の位置情報より光非線形デバイスの物理構造情報が抽出できる。さらに、光非線形デバイス検討の

第3章 非線形光学材料

第1節 光非線形デバイス	357
1. 基本的考え方	357
2. 非線形光学材料から応用技術までの整理	357
3. 光非線形デバイスの適用検討の概要	358
4. 光非線形デバイス検討の具体例	363
4.1 光スイッチ	363
4.2 光変調器	369
4.3 光増幅素子	372
4.4 波長変換素子	381
4.5 光論理・光メモリ素子	390
4.6 光ソリトン素子	394
5. 今後の課題と動向	396

図1 目次の一部

MQW	17, 358, 365
Multi-Quantum Well	365
多重量子井戸 (MQW)	357, 365
光スイッチ	357, 358, 362, 363, 391, 396
光不安定	252, 358
光非線形現象	357, 358
光非線形デバイス	358, 360
光メモリ	357, 358, 397, 767

図2 インデックスの一部

具体例の節に光非線形デバイスの下位概念が記述されているとみなすことにより光非線形デバイスの上下関係という概念構造情報を抽出できる。

次に、インデックスを見てみると、光非線形デバイスというタームがある(図2参照)。このタームが指すページ番号の位置情報より光非線形デバイスの物理構造情報が抽出できる。目次からの位置情報とインデックスからの位置情報の違いは、指定範囲が異なることである。すなわち、目次に出ている場合は、その概念が中心的役割を果たしている位置を示しており、イ

1. バルク、ファイバ等の形態を伴った非線形材料に光電磁界が入力されると、材料内の非線形光学効果によって光入力と光出力が線形に対応しない光非線形現象（光波長変換、光双安定、等）を生じる。
2. この表では、光非線形現象を図-1と同様に"光の電磁界入力に対して物質の応答あるいは物質からの出力が線形でなくなる現象"と広く定義している。
3. 図-11は、多重量子井戸 (Multi-Quantum Well:MQW) 構造を有する半導体材料の電界屈折率効果 (正確には、量子閉じ込めシュタルク効果:Quantum Confined Stark Effect:QCSE) による屈折率変化を利用して、方向性結合器型の光スイッチを構成した例である。
4. そこで、この方向性結合器の導波路に図のように電界をかけると、導波路材料の有するEO効果によって導波路の屈折率が変化し、この結果結合係数が変化して、導波路(1)から導波路(1')から導波路(2')への結合状態が変わることとなる。

図3 テキストの例

ンデックスに出ている場合は、その概念が中心的役割を果たしている位置と補助的役割を果たしている位置を示している。また、目次の位置情報が示す範囲にあるインデックスタームは、目次の示す概念と関係があるといえるので、インデックスタームを目次の構成に沿って分類することができる。それにより、荒いものではあるが、専門家があげた重要概念を構造化できる。

次に、テキストを見ていくことにする(図3参照)。1の文でインデックスから参照されている概念は、バルク、非線形材料、光非線形現象である。ここで、光非線形現象というタームに注目してみる。この文から光非線形現象の説明及びその具体例として光波長変換、光双安定が記述されている。これより、光非線形現象の概念構造情報を抽出できる。また、インデックスのページ情報により光非線形現象について2の文が参照できる。これより光非線形現象の別の説明を知ることができる。その他、3の文では、多重量子井戸、Multi-Quantum Well、MWQ、量子閉じ込めシュタルク効果、QCSEがインデックスから参照されているが、この文から同値関係という概念構造情報を得ることができる。また、4の文を自然言語処理によって

解析することにより、因果関係という論理構造を抽出できる。

3 情報モデル

情報の構造化及びそれに基づいての類推を行なうため、情報モデルに以下の性質を持たせることが必要であるとわれわれが考えている[2][3]。

(1) 再帰的な表現構造

再帰的な表現ができれば、現実世界の複雑な事象およびその相互関係を柔軟に対応できるようになる。従って、情報モデルの表現能力が強化される。さらに、再帰的な表現構造ができれば、概念間の階層関係が明確となる。その結果として、問題のスケールに応じて、既存の情報を階層の適切な深さまで展開することが可能となり、問題の単純化および解決空間の縮小に便利を持たせる。

(2) ラベル

意味を取扱うため、現実世界のさまざまな概念の識別子として、エンティティを表すノードにつけるラベルが必要である。

(3) ハイパーリンク

類推を行なう際、関係を表すリンクもオブジェクトとして取扱いたいという場合がある。従って、リンクをノードと同じように取扱える構造が必要なる。ハイパーグラフを使えば、従来のリンクがノードの集合となっているため、リンクに関する情報は一つのノードにより表現されることができる。

これらの必要性を考え、われわれは次のように情報モデルの数学的な基礎となる拡張ハイパーグラフ及びその関連する概念を定義する。

まず二つの空間を定義する。

UL = Universe of Labels

UC = Universe of Constants

ここでいうラベルは現実世界のさまざまな概念あるいは属性に対応し、定数は数字でも良い、文字列でも良い、集合、テーブルないしは図形などでも良い。

[拡張ハイパーグラフ]

Let N is a set of nodes. A Extended-Hypergraph 機能、ナビゲート機能などとの併用により、さまざまな類推を効率よく行なうことができる。

機能、ナビゲート機能などとの併用により、さまざまな類推を効率よく行なうことができる。

(1) $HL_i \neq \emptyset (i = 1, \dots, m)$

(2)

$$\bigcup_{i=1}^m HL_i = N$$

さらにノードには二種類のものがある：シンプルノードと、コンプレックスノード。

[シンプルノード]

シンプルノードはわれわれが取扱う情報空間の最小単位であり、それ以上は分割できないアトム的な存在となっている。任意のシンプルノード sn に対し、下記の二つの関数 f_l と f_c が存在する：

$f_l(sn) \rightarrow UL ; f_c(sn) \rightarrow UC$

[コンプレックスノード]

シンプルノードに対し、コンプレックスノードは内部構造を持つノードである。従って、その内容もまた再帰的に拡張ハイパーグラフ EH により表される。なお、任意のコンプレックスノード cn に対し、下記の関数 f_l が存在する：

$f_l(cn) \rightarrow UL$

4 情報ベースシステム

(1) 情報構造化機能

一次情報から自然言語処理技術などを用いて問題空間の各エンティティ間のさまざまな関係を表す大量なリンクを自動的に生成する。また、シソーラスおよびタキソノミーの自動メンテナンス機能により、情報の変化にすばやく対応することができ、問題空間に関する知識を蓄積できる。

(2) 情報操作機能

類似性を計算するためにグラフの変形、特徴抽出、部分削除などの機能も用意している。これらの機能の選択的な組合せにより、異なる許容基準による類推（精度重視の類推、あるいは再現率重視の類推）が可能となる。さらに検索

基本操作

- ・ ノード及びリンクの検索
- ・ リンクのマニュアル編集
- ・ ナビゲート
- ・ サブグラフの置き換え
- ・ 特定パターンの検出
- ・ サブグラフの削除
- ・ グラフからの特徴抽出（サブグラフの抽出）

複合操作

- ・ 問題空間の縮小、抽象化、帰納
- ・ パターンマッチング（完全、近似）

5 類推

人間の持つ類推能力をコンピュータにも持たせるため、多くの研究が行なわれてきた。その先駆的な研究例として Gentner の Structure-Mapping Engine^[7]、Winston の類似性による学習と推論の研究^[8]などが挙げられる。しかし、これらの研究の多くは、類推に用いられる情報の構造化に関してはあまり触れておらず、構造化が行なわれたという仮定で類推を行なっている。本研究では、情報の構造化こそ類推に決定的な役割を果たしていると考え、構造化された情報に基づいた類推を提案する。

類推とは、ベースとターゲットそれぞれの既知な部分の間の類似性に基づいて対応関係を確立し、さらにその合理性を評価することによって行なう推論である。その手順は以下の通りである。

1. ベースとターゲットとの間の可能なマッピングに対しそれぞれ評価する。この対応関係確立—評価という過程の繰り返しにより、もっとも適切な対応を選出する。

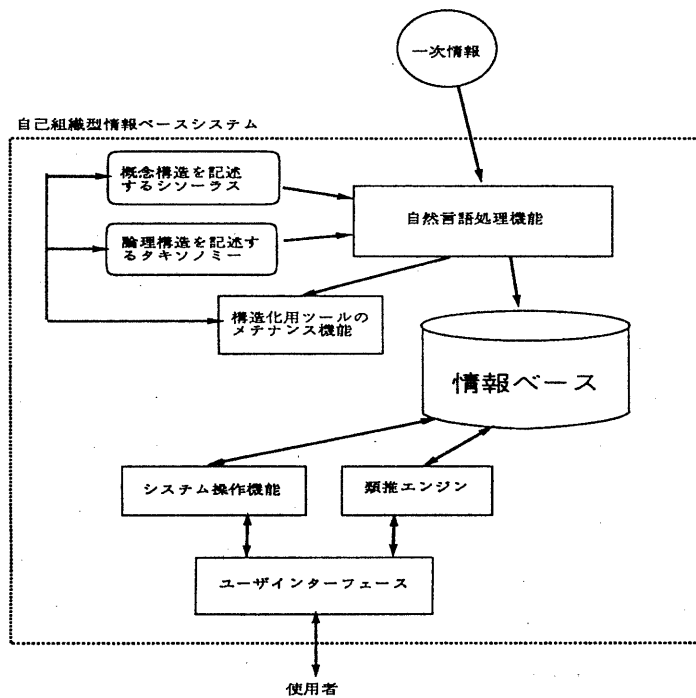


図4 システム構成図

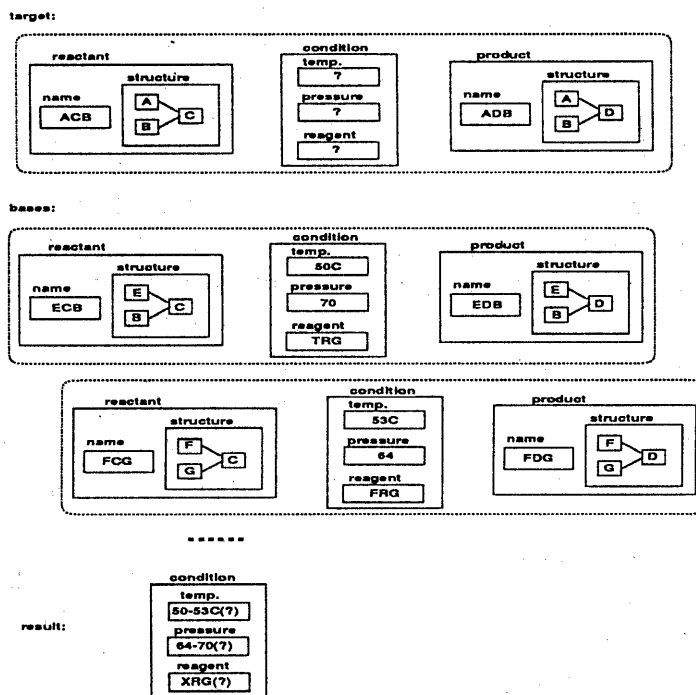


図5 類推の例示

2. 表現上の違いという考えを入れ、事実を表現したパターンの変形、意味解析辞書によるパターンの書き換えを行なってマッピングを評価する。
3. 対応関係を決めた後、ターゲットの中の未知な部分に対し、ベースの中の対応する部分を類推の結果としてターゲットに持ち込む。ベースとターゲット間の持ち越しは論理構造の間で行なう。

前述のように、われわれの研究では、構造化された情報が拡張ハイパーグラフ的な情報モデルにより記述されている。従って、類推の問題は結局拡張ハイパーグラフでの部分パターンマッチングという問題に帰結する。その具体的な手順は以下の通りである。

完全（近似）パターンマッチング

1. ユーザによるターゲット記述
2. システムの機能により、ターゲット記述をサブグラフ EH_i に変換する
3. EH_i を用い、グラフにより表現されている問題区間 EH で（近似）パターンマッチングを行ない、その結果を EH_j とする。

$$EH_r = \sum_{j=1}^k EH_j$$

ここで足し算の意味は、事実の出現回数により、その信用性が影響されるということになる。出現回数が多いほど、その信用度が高くなる。

4. システムの機能により、 EH_r を変換し、結果としてユーザに返す。

有機合成反応における類推の一つの例は図5に示されている。

6 考察

近年、情報量の急増に伴い、さまざまなシステムの維持が大きな問題となっている。ここでわれわれが提案する情報モデルおよび自己組織型の情報構造化により、維持性の良いシステムができる。さらにシソーラス、タキノミーなど構造化ツールの自動メンテナンスを通じ、情報構造に関する知識の累積機能が設けられる。従って、情報量の増加により、情報空間に対する把握の程度もより完全となってくる。さらに、類推機能の導入により研究開発技術者にとってはより強力なサポート役を果たし、人工頭脳型の情報ベースシステムの実現になる。

しかし、構造化も類推も、結局は情報の意味を扱うということになる。意味を取扱うことは、根本的には以下の二つの問題に帰結できる。

1. 意味の表現。現実世界の概念とその間の複雑な関係を、計算機が処理できる形式でどう表現するかの問題。
2. 意味の処理。計算機技術および数学理論、アルゴリズム理論のレベルで、表現された意味をどう操作するかの問題。

表現の観点からいうと、現実世界の意味構造は非常に複雑なものであるので、ある程度複雑な構造を持つ表現モデルでないと現実問題に対応しきれない。しかしその一方、処理の観点からいうと、表現方法の複雑さは意味の処理にも複雑さを持ち込むことになり、あまり複雑すぎる表現を導入すると取り扱いが非常に難しくなる。従って、両者の間にうまくバランスをとる必要がある。本研究の情報モデルでは、処理に関してまだ不十分である。また、類推に関しては計算コストの問題もある。処理系の充実および類推の実用化が今後の課題として残されている。

7 結論

本論文では、情報の構造化の方法とモデルについて述べた。また、情報ベースシステムの機能、特に類推機能について述べた。それにより、高度で複雑な情報を扱えるようになった。

参考文献

- [1] Yuzuru Fujiwara, Jihong He, Gyoto Chang, Nobuo Ohbo, Hiroyuki Kitagawa and Kazunori Yamaguchi: Self Organizing Information Systems for Material Design. Proc. of the 1st CAMSE. Tokyo. (Aug.,1990).
- [2] Yuzuru Fujiwara: The Self Organizing Information-Base Systems with Learning and Analogical Reasoning. Proc. of the 3rd Beijing Int. Symp. on Computer. Information Management. S3A-7. (Oct.14-18,1991).
- [3] Yuzuru Fujiwara, Nobuo Ohbo, Hiroyuki Kitagawa and Kazunori Yamaguchi: The Information Base Systems for Materials Research. CODATA Bull.24-1, pp.1-7(1992).
- [4] 水谷静夫, 石綿敏雄他: 文法と意味 I. 朝倉書店. (1985).
- [5] 高松 忍, 西田富士夫: 見出し情報を用いたテキスト解析と情報抽出. 情報処理学会論文誌. Vol.29,NO.8 pp.760-769(1988).
- [6] 福本淳一, 安原 宏: 日本語文章の構造化解析. 情報処理学会, 自然言語処理研究会,85-11. (1991).
- [7] D. Gentner: Structure-mapping:A theoretical framework for analogy,Cognitive Science, 7, pp.155-170 (1983).
- [8] Patrick.H.Winston: Learning and Reasoning by Analogy, Communications of the ACM, Vol. 23, No. 12, pp.689-703 (1980).
- [9] Harold Boley: Directed Recursive Labelnode Hypergraphs: A New Representation-Language, ARTIFICIAL INTELLIGENCE, Vol. 9, No. 1, pp.49-85(1977).
- [10] John F. Sowa: Conceptual Structure. Addison-Wesley(1984).
- [11] 田中穂積: 自然言語解析の基礎. 産業図書. (1989).
- [12] 高木 朗, 伊藤幸宏: 自然言語の処理. 丸善. (1987).
- [13] 電子情報通信学会: 日本語情報処理. 電子情報通信学会. (1987).