

# 意味理解のための情報構造モデルとその応用

## The Information Model for Semantic Understanding and Its Applications

藤原 譲 (Yuzuru Fujiwara)      ○ 安 江虹 (Jianghong An)

### 概要

Thinking support functions are required for domain specialists to use information necessary for research and development. The thinking support systems should consist of semantic processing functions such as analogical reasoning, induction, abduction and so on. However, conventional models of databases and representation languages of knowledge bases have very strict constraints to describe semantic relationships.

The representation and storage of information based on the homogenized bipartite model is presented. Semantic relationships among concepts such as nested, many to many, dynamic and relative ones can be represented precisely. Based on the homogenized bipartite model an organic synthesis information model is constructed and the information is organized. Moreover, semantic processing functions such as analogical reasoning on structured information space are mentioned.

### 1 はじめに

研究開発に必要な情報を研究者、技術者が直接利用するには、検索、数値計算や演繹推論など従来型の処理のみではなく、意味処理を含む各種の思考支援機能を実現しなければならない。ところが従来のデータベースや知識ベースシステムでは表現形式や管理方式の制約が厳しく意味関係が十分に記述、表現できない [1, 2]。

そのために、より柔軟な情報モデルが必要である。モデルの役割は情報の意味を理解するための道具と、その意味関係を正確に仕様化して計算機で処理できる形まで表現する基準である。情報の意味を表現、記述するモデルについての研究が以前からなされてきた。実世界をデータベース化する立場からは関係モデル [3, 4]、E-Rモデル [5]、関数型モデル [6]、意味データモデル (SDM) [7]、オブジェクト指向データモデル [8] などが提案された。一方、人工知能における知識表現の立場からは述語論理、意味ネットワーク [9, 10]、フレーム [11] など表現方法が提案された。

ところが、それらの表現方法およびモデルでは集合、或いはグラフに基づくものであるため、情報の構造または意味を十分には表現できない。情報の意味は概念および概念間の関係によって記述されるので、まず概念と概念関係の特徴を解析し、その特性を考慮して意味表現モデル均質化2部グラフモデル [2] に基づく情報の記述、表現方法を示した。具体的に、有機合成情報を例にしてそのモデルの応用について述べる。

## 2 概念および概念関係の特徴

概念および概念関係の特徴としてハイパーグラフを拡張した均質化2部グラフ型モデルでは概念構造を中心に下記の点を取り扱える [2, 12] :

1) 階層構造：概念間は階層構造を成している。これによって概念の差分的な記述および概念間の属性値の継承ができる。

2) 内部構造：概念が内部構造を持つ。概念が抽象化または詳細化によって内部構造の複雑さが変わる。

3) 重なり：概念間には意味的な重なりが存在する。このような重なりは類似など意味関係の本質であるので類似度の測度の手がかりにもなる。

4) 多項関係：概念間には2項関係だけではなく多項関係も存在する。

5) 動的関係：概念間の関係は動的であり、視点によるものが多い。例として類似、関連など関係が挙げられる。

6) 相対性：概念と属性、または、概念と関係は相対的である。場合によって属性を概念（或いはその逆）とみなすことがあり、または、関係を概念（或いはその逆）とみなすことがある。

上述した概念および概念関係の六つの特徴を統合できるモデルとしてハイパーグラフを拡張した均質化2部グラフモデルを報告した。均質化2部グラフはハイパーグラフ [13] の2部グラフ表現を均質化することによって内部構造、相対性を含めて全ての関係を記述できるモデルである。

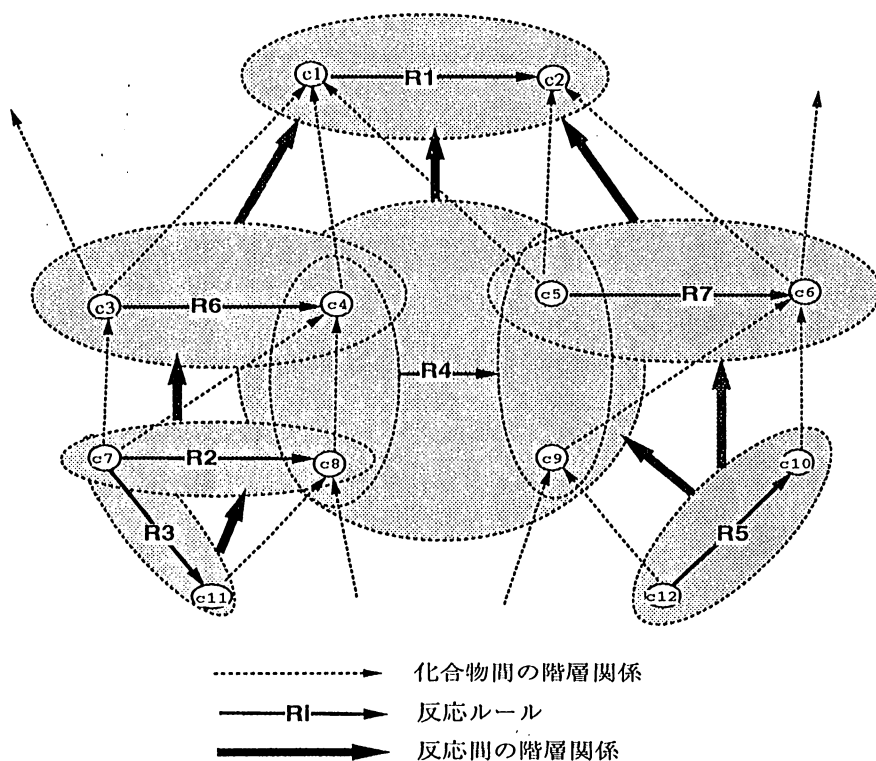


図-1：情報の構造の例

ここでは、例を使って概念関係の特徴およびそのモデル化について述べる。図-1は有機合

成情報の例である。この例において、まず、化合物間、または反応間には階層関係が存在することが分かる。次に、反応という概念は出発物、生成物など内部構造を持つことが示されている。また、反応の出発物と生成物は一つずつとは限らないので多項対多項関係である。最後に、反応は化合物間の関係を表すと同時に概念でもあるので相対性的である。

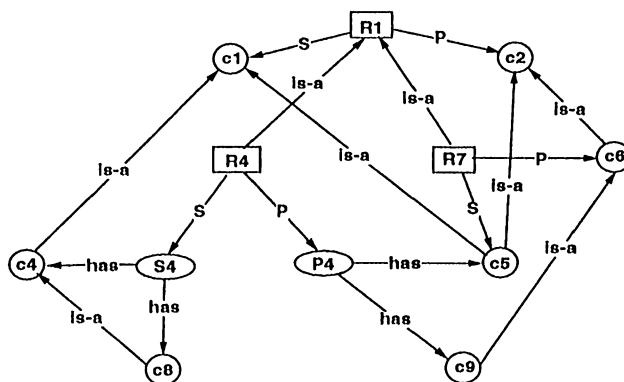
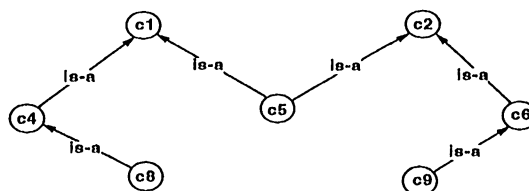
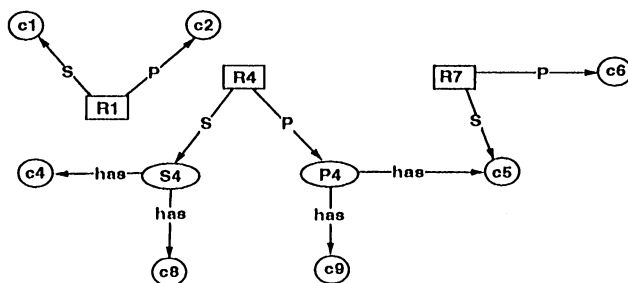


図-2-1：均質化2部グラフの表現例



1) 化合物の概念構造



2) 論理構造

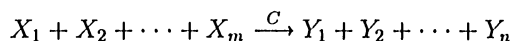
図-2-2：概念構造と論理構造

図-1の有機反応情報の一部を均質化2部グラフで表現した結果を図-2-1で示す。ここに、 $c_i$ は化合物、 $R_i$ は反応を代表するノードである。楕円で囲まれたノードは化合物のセットを意味する。SとPはそれぞれ反応の出発物と生成物を指す。

図-2-1を意味関係によって分けて記述したのは図-2-2である。ここでは、化合物の概念構造と論理構造を示した。反応の概念構造（反応間の階層関係）については後で述べる。

### 3 有機反応情報への応用

均質化2部グラフモデルに基づいて有機合成情報モデルの構築について述べる。有機合成情報は主として化合物についての情報と化学反応についての情報からなる。化合物については分子構造情報、化学性質情報などがある。反応が出発物、生成物及び反応条件など要素から構成される。一つの反応に参加する出発物と生成物はそれぞれ複数個ある：



ここで、 $X_i (i = 1, m)$  が出発物、 $Y_i (i = 1, n)$  が生成物、 $C$  が反応条件である。出発物と生成物は結合表 (connection table) で化学構造が記述されている。反応条件には試薬、触媒、溶媒、反応温度、圧力、時間、収率などがある。

#### 有機反応情報のモデル

$A$  を原子の集合 (抽象原子と偽原子も含む) とする、

$$\kappa = \langle A, E, B \rangle$$

が化合物構造の空間である。

ここで、 $E \subseteq A \times 2^A$  は原子間の結合であり、

$B$  は結合の種類である。 $B$  の中に抽象結合も含む。

結合関係  $A \times 2^A$  は配位化合物 (錯体) の表現も可能になる。

有機反応の情報空間が  $\kappa$  に基づいて構築される。ここで、処理効率を考えて化合物の概念構造、反応の論理構造および反応の概念構造をわけて記述することになった。

#### 化合物の概念構造：

$$\Gamma_1 = \langle \kappa, \Omega_1, L_1 \rangle$$

ここで、 $\Omega_1 \subseteq \kappa \times 2^\kappa$  は化合物間の階層関係で、

$L_1$  は化合物間の差を示すラベルの集合です。

#### 反応の論理構造：

$$\Gamma_2 = \langle \kappa, \Omega_2, L_2 \rangle$$

ここで、 $\Omega_2 \subseteq 2^\kappa \times 2^\kappa$  は化合物間の反応関係で、

$L_2$  は反応規則及び反応条件を示すラベルの集合です。

ここに、反応パターンなど総称的な反応の表現も含まれていることを注意して欲しい。

#### 反応の概念構造：

$$\Gamma_3 = \langle \Gamma_2, \Omega_3, L_3 \rangle$$

ここで、 $\Omega_3 \subseteq \Gamma_2 \times 2^{\Gamma_2}$  は反応間の階層関係で、

$L_3$  は階層関係も含む反応間の差を示すラベルの集合である。

反応の情報空間  $\Gamma$  を次のように定義する：

$$\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$$

$\Gamma$  は多種類のノード、関係及びラベルを持つハイブリッド (hybrid) モデルである。

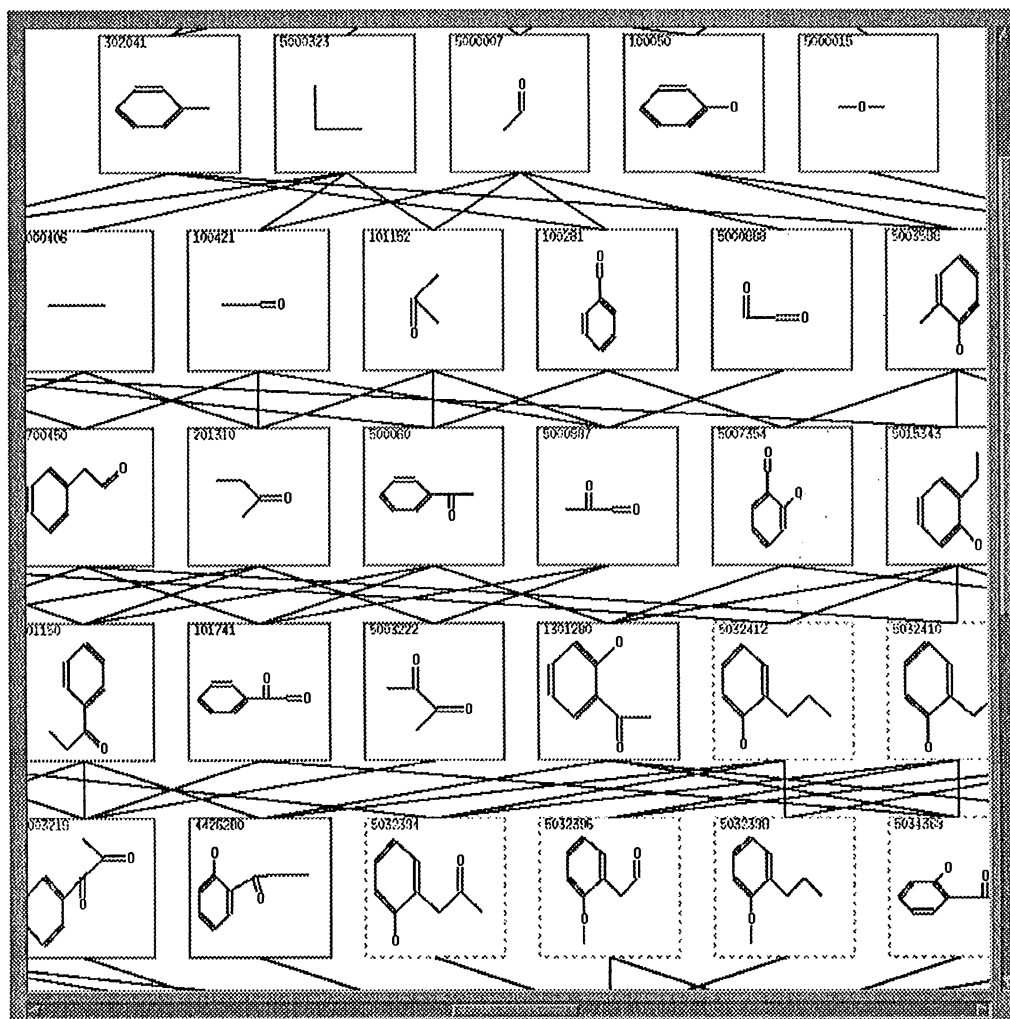


図-3：化合物概念構造の一部

#### 4 モデルに基づく情報の構造化

##### 有機化合物の概念構造の構築

化合物の概念構造は化合物間の概念としての同値、上下、包含など関係を表現する構造である。化合物の性質は基本的には分子構造によって決まるのでこの構造に基づいて構造化を行なう。化合物の概念構造が化合物を表現するノードとノード間の包含関係を表現するラベル付きリンクから構成される。構造間のラベルにはノード間分子構造の違いが明記されている。また、構造の中には反応事例に含まれる既存化合物の構造と生成された仮想化合物の構造がある。

反応の論理構造の構築

反応事例集合に基づいて生成した反応規則のネットワークを論理構造という。これは異なる反応に関係する化合物の構造の同型性判定によって反応規則を自動的に組織化したものである。

反応データベース CORES[14]に基づいて構築された化合物概念構造と反応論理構造の一部がそれぞれ図-3と図-4で示されている。

反応の概念構造の構築

反応の概念構造は反応間の概念としての同値、上下、包含など関係を表現する構造である。反応の要素として、出発物、生成物および反応条件がある。反応条件にはまた試薬、触媒、温度、圧力、収率などの要素が含まれている。特に、反応を見る時には出発物から生成物までの構造の変化が反応の本質を認識するにおいてとても重要であるのでここではこの構造の変化に基づいて反応の概念構造が構築された。図-5では反応の概念構造と論理構造の関係を示している。

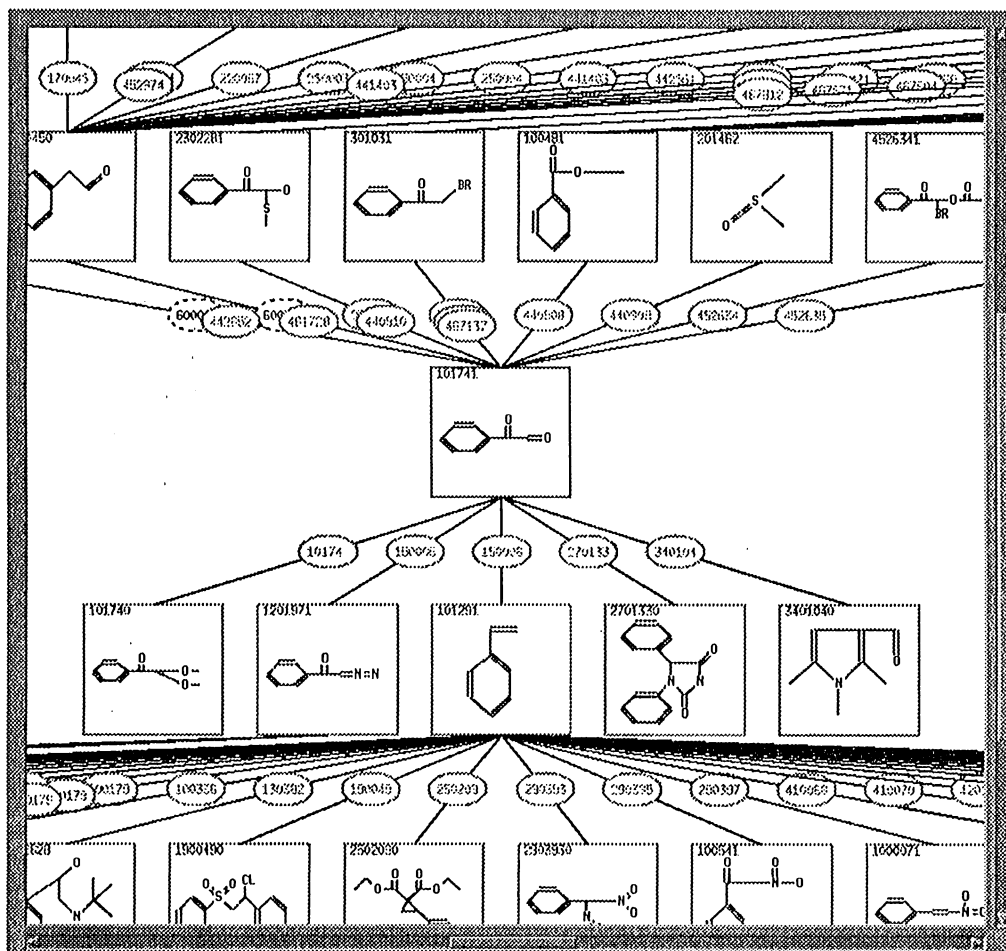


図-4：反応論理構造の一部

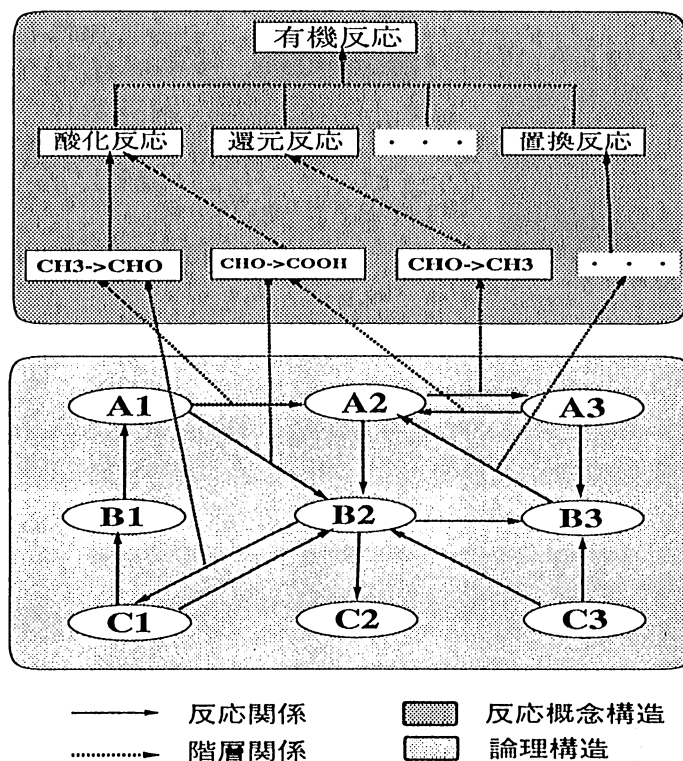


図-5：反応の概念構造と論理構造の関係

### 構造化された情報に基づく類推

新しい化合物の合成には未知化合物の取り扱いの問題を解決しなければならない。即ち、今まで知られていない化合物に対してその性質を予測する手段を提供しなければならない。これについては、化合物の概念構造の構築に当って仮想化合物生成で対処する。つまり、未知化合物の概念構造における位置を特定すればまわりの既存化合物によって性質が推測できる。反応については、概念構造と論理構造を利用して類推で解決する。

反応条件を類推する手順は次の通りである：

- (1) 入力された出発物と生成物の分子構造の変化から反応パターンを認識する；
- (2) 反応の概念構造から類似反応パターンを持つ反応（類似反応候補）をアクセスする；
- (3) 上述のような計算方法で類似反応候補と問い合わせ反応の出発物同士、生成物同士の類似度を計算する；
- (4) 計算された類似度と閾値によって、類似反応の集合を決定する；
- (5) 類似反応の反応条件の内容を含めて詳細な類似性を検討する。

類似反応の反応条件を問い合わせ反応に適用するには類似反応と問い合わせ反応間の比較をするのみならず類似反応同士間の比較もしなければならない。

## 5 むすび

本論文は情報の構造、特に概念関係の特徴を解析し、均質化2部グラフモデルにおける情報の表現と記述について述べた。また、均質化2部グラフモデルに基づく有機合成情報モデルを構

築した。更に、構築された有機合成情報モデルに基づいて情報の構造化を行なった。最後に、構造化された情報空間における類推など意味処理について述べた。

## 参考文献

- [1] Y. Fujiwara, Z.Q. Wang et al: The Multicategorical Structures of Information for Inferences and Reasoning in the Self-organizing Information-Base System, Computer Aided Innovation of New Materials II, Elsevier Science Publishers B.V. ( M. Doyama ed. ), 1993, 27-32.
- [2] Y. Fujiwara : The Model for Self-structured Semantic Relationships of Information and Its Advanced Utilization, International Forum on Information and Documentation, Vol.19, No.2, 1994, 8-10.
- [3] E. F. Codd: A Relational Model of Data for Large Shared Data Banks, Communications of ACM, Vol.13, No.6, 1970, 377-387.
- [4] E. F. Codd: Extending the Database Relational Model to Capture More Meaning, ACM transactions on Database Systems, Vol.4, No.4, 1979, 397-434.
- [5] P. P. Chen: The Entity-Relationship Model: Toward a Unified View of Data, ACM transactions on Database Systems, Vol.1, No.1, 1977, 9-36.
- [6] D. W. Shipman: The Functional Data Model and the Data Language DAPLEX, ACM transactions on Database Systems, Vol.6, No.1, 1981, 140-173.
- [7] M. Hammer and D. McLeod: Database Description with SDM: A Semantic Database Model, ACM transactions on Database Systems, Vol.6, No.3, 1981, 351-386.
- [8] J. Banerjee, W. Kim, H. J. Kim and H. F. Korth: Semantics and Implementation of Schema Evolution in Object-Oriented Databases, Proceedings of ACM SIGMOD, 1987, 311-322.
- [9] R. M. Quilian: Semantic Memory, Semantic Information Processing, Minsky, M.A. (ed.) The MIT Press, Cambridge, MA, 1968.
- [10] J. F. Sowa: Conceptual Structures, Addison-Wesley, 1984.
- [11] M. A. Minsky: A Framework for Representing Knowledge, The Psychology of Vision, Winston, P. H. (ed.) McGraw-Hill, New York, 1975.
- [12] Y. Fujiwara and H. Gotoda : Representation Model for Relativity of Concepts, International Forum on Information and Documentation, Vol.20, No.1, 1995, 22-30.
- [13] C. Berge: Hypergraphs, North-Holland, 1989.
- [14] H. Ishizuka, Y. Inoue, K. Oda, K. Katoh, Y. Kawashima, K. Shimizu, Y. Naka, M. Shiroki, T. Sekiya, H. Takashima, T. Nakayama, T. Nishioka, N. Hara, S. Sugawara, I. Matsuura, T. Miyagishima, M. Miyamura, R. Kani, S. Yoshida, N. Ohbo, H. Kitagawa, K. Yamakuchi and Y. Fujiwara: Development of Computer/CD-ROM Aided Organic Synthesis Research System: CORES(In Japanese), 13th Symposium on Chemical Information and Computer Science, 1990, 13-16.