

概念構造における欠落情報の推定と生成の研究

1 S-2

安江虹 石井哲子 藤原譲
筑波大学 電子・情報工学系

1 はじめに

意味処理を行なうために概念間の意味関係を記述する概念構造の構築が必要である。われわれは原情報の自己組織化による概念構造の構築手法の研究を行ってきた。即ち、原情報から意味関係を記述する断片的な情報を構造化情報として抽出し、それらを統合してある領域の知識をカバーできる概念構造を構築する。ところが、膨大な量の原情報があるにも関わらず、重要な情報が欠落している場合がよくある。

本研究では、構築された概念構造を用いて欠落情報の推定と生成方法について二つの処理手法が提案される。それは、情報空間の構成単位と構成規則が明確な場合、欠落した概念および概念間の関係を網羅的に生成する手法と、類推、帰納推論などによる推定する手法である。

2 概念構造およびその構築方法

概念構造とは概念間の同値関係、上・下位関係などを表す意味構造で、知識表現手法の一つとみなされる [1, 2]。知識表現についてはルールに基づく表現手法 (Production Rule System) は知識の多様性を表現し切れない欠点があり、事例に基づく手法 (Case-Based System) は事例間の関係が見えない上に記述が重複がちである。

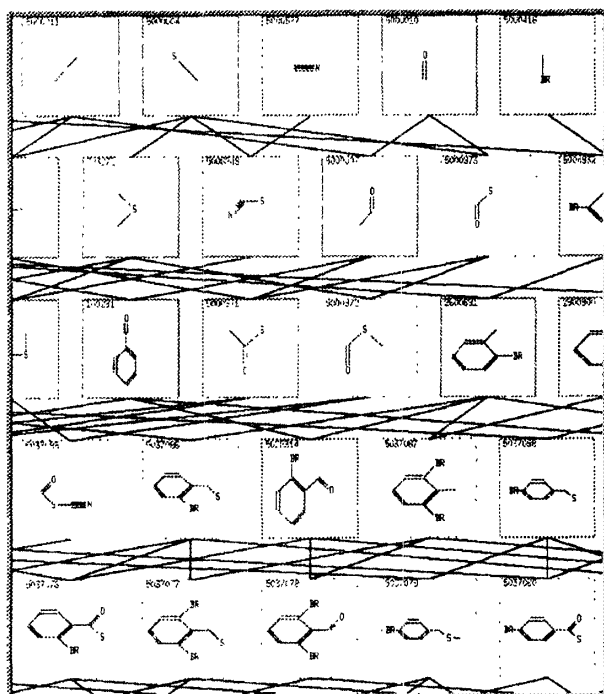


図-1 化合物概念構造の一部 (その1)

そのために、必要な個別事例を網羅的に集めて概念構造を構築するのは一つの解決方法である。更に、科学技術情報のような大量かつ増え続ける情報の概念構造は情報の自己組織化手法を用いて構築しなければならない。情報の自己組織化とは、情報資源に内在する意味的な関係を構造化情報として抽出し、それを用いて情報資源全体を自動的に組織化するということを指す [2, 3]。

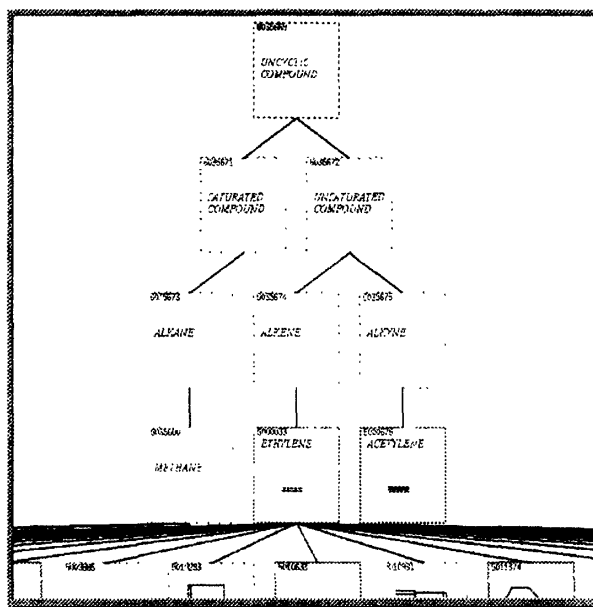


図-2 化合物概念構造の一部 (その2)

本研究では典型的な科学技術情報の例である化学情報を対象にした。具体的には有機反応の類推を行なえる研究支援システムの構築を目指している。有機反応の要素として、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。

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化合物の類似性は分子構造或いは原子結合の類似性に依存する。つまり、化合物が似た構造を持てば似た性質をもちうる。従って、本研究ではまず化合物の分子構造間の包含関係 (グラフ間のサブグラフ関係) によって概念構造を構築した。その概念構造の一部を図-1で示す。

この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。ラベルには分子構造間の違いが明記されている。この違いの単位は原子だけではなく、基本構造 (炭化水素、基本複数環系、特性基に相当する構造など) である場合もある。

ところが、分子構造における包含関係だけにに基づく方法は限界がある。それは次のことに原因がある：

1) 化学における概念はすべて個別な分子構造と対応するわけではなく総称表現がよく使われる；例えば、"ハロゲン化合物"とはI, Cl, Br, Fなどを含む化合物のことである。

2) 共通性質を持つ化合物は必ずしも共通構造を持つとは限らない。例えば、"毒物"にはいろいろな種類があつて、さまざまな分子構造が持っている。

そのために、化合物の名称、定義の解析によって概念構造の構築も必要である。構築された構造の一部が図-2で示される。尚、化学反応間の意味関係を表す概念構造についてはここで省略する。

3 欠落情報の生成

構築された概念構造を利用して情報の生成ができる。生成する情報の特徴によって下記二種類にわけることができる：

1) 欠落情報の生成：情報空間の構成単位と構成規則が明確にした場合、欠落した概念および概念間の関係を補足する；

2) 候補解の生成：原因 \leftrightarrow 結果のような構成単位および構成規則が明確でない場合、類推、帰納推論、仮説生成など方法が必要である。

有機合成の目標は新しい化合物の合成である。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。

化合物の構成単位（原子、官能基など）および構成規則（原子間の結合）が相対的に明確であるため、概念構造に基づいて未知（仮想）化合物の生成が可能である。図-3ではその生成過程（ボトム アップ）が示されている。それによって、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予測できる。

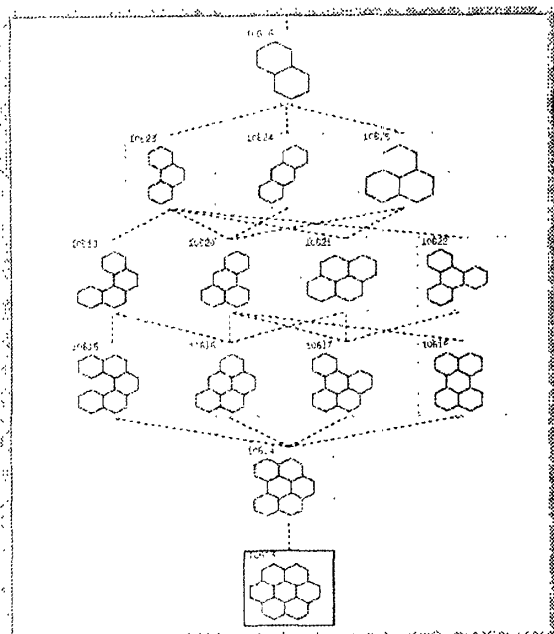


図-3 欠落化合物の生成

化学者は新しい反応及び反応経路の設計に次の知識を用いる：

- 1) 構造と物性の知識；
- 2) 有機合成の知識；
- 3) 既知の反応事例。

計算機システムにおいて、3) については記憶容量と正確性で専門家の能力を越えられる。問題は既知の反応事例を見る方法と分析能力にある。これが1) 及び2) と関係する。即ち、化学の知識をシステムに獲得させることとこの知識に基づいて事例を利用することである。

これが本研究において”構造化”の主な目標である。即ち、構造化によってこれらの知識を獲得し、また概念構造と論理構造によってこれらの知識をシステムにとって処理しやすい形で表現する。つまり、化合物の概念構造が1) に対応し、反応の概念構造と論理構造が2) と3) に対応する。

反応については、論理構造の拡張によって解空間を拡大しておくことで対応する。つまり、化合物と反応の概念構造および反応の論理構造を利用して候補解を作り出すことによって解空間を拡大する。図-4では未知反応の推定結果を示す。

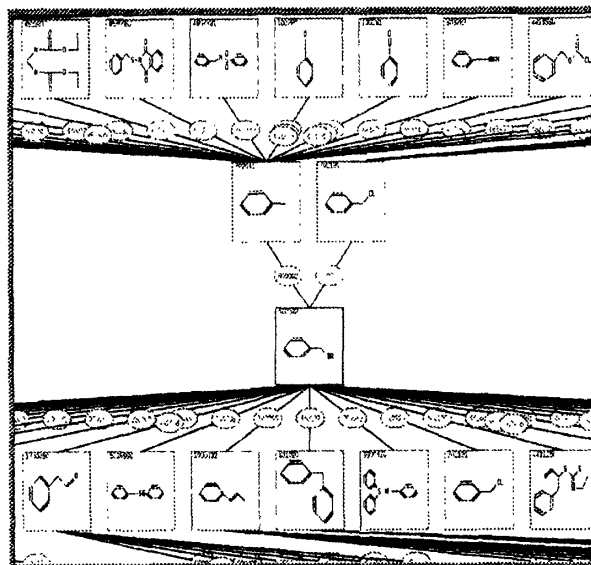


図-4 候補反応条件の推定

概念構造、論理構造の拡張は無限に可能なので、実際には制限する必要がある。拡張の制限と精度を考えて2種類の拡張を行なう：1) 問い合わせをする度に必要最小限の拡張を行なう；2) 十分に根拠のある拡張だけを行なう。

4 むすび

概念構造に基づいて欠落情報の推定と生成について述べた。それは思考支援システムをはじめ Open World に対応できるシステムにおける基本不可欠な機能である。概念構造は概念間の意味関係を合理的に記述できる意味構造である。また、科学技術情報のような大量かつ増え続ける情報の構造化は自己組織的に行なわなければならない。

参考文献

- [1] J. F. Sowa: Conceptual Structures. Addison-Wesley. 1984.
- [2] Yuzuru Fujiwara: *The Model for Self-structured Semantic Relationships of Information and Its Advanced Utilization*, International Forum on Information and Documentation, Vol. 19, No. 2. 8-10, 1994.
- [3] Jianghong An and Yuzuru Fujiwara: *Similarity of Compounds and Reactions Based on Self Organized Conceptual Structures of Organic Synthesis Information*, Journal of Japan Society of Information and Knowledge, Vol.6, 1996.