3R－5

概念構造に基づく類似性と類推の研究 ——化合物の名称・構造の類推を例にして

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

類推は問題解決システムにおいて必要かつ有効な機能の一つである。類推とは基底領域と目標領域のいくつかの対象間に類似性を検出し、その類似性を用いて基底領域の対象で成立する事実や知識を目標領域の対象に変換することにより、問題解決の手がかりを得たり未知の事実などを予測する推論方式である。しかし、類推に欠かせない類似性の扱いは概念間の意味関係に関わり処理であるため、意味関係の記述と処理が必要である[1]。

概念間の必要な意味関係は全て概念構造によって記述することにして、概念構造を構築する。次に、構築された概念構造に基づく概念間の類似度の測度手法を提案する。応用として有機化合物の名称・構造の類推について述べる。

2　有機化合物の構造と名称

化合物のグラフィカル表現として構造表現が使われている。構造表現では化合物の平面構造だけではなく立体異性、回転異性などを表現することもある。一方、文章、話の中に化合物を指す時には名称を使うのが便利である。化合物の命名（構造から名称を作ること）は3次元構造を線形文字列で表すことで、ユニークかつ曖昧さのない名称を作るには、複雑な命名規則が必要とする。

化合物の命名法が幾つかあるがその中に一番良く使われているのはIUPAC（International Union of Pure and Applied Chemistry）命名法[2]である。この命名法に基づいて一つの化合物の名称を作成には普通次の手順をとる：

1）化合物の本質によって、命名方法（置換、基官能、付加、減去、結合、代置）を決定する；

2）主基として用いる特性基があるなら、その種類を決定する；

3）母体構造（主鎖、母体環系など）を決める；

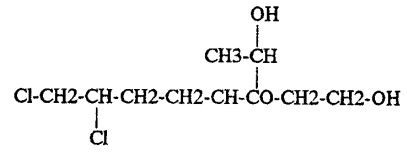
4）母体構造と主基を命名する；

5）接頭語、挿入語などを決めて命名する；

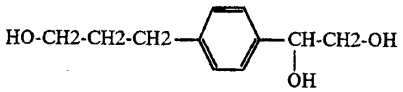
6）位置番号を完全に付ける；

7）離すことのできる接頭語をアルファベット順にならべ、部分名を集成して一つの完全な名称を作り上げる。

IUPACを使って命名した例を図－1で示す。



7,8-dichloro-1-hydroxy-4-(1-hydroxyethyl)-3-octanone



1-[p-(3-hydroxypropyl)phenyl]-1,2-ethanediol

図－1 構造と名称の例

IUPACをはじめ従来の命名法に基づく開発された構造・名称変換システムは、部分構造の辞書、命名規則およびプログラムから構成される。しかし、化合物の分子構造の複雑さと多様性によって辞書、規則およびプログラムが大変繁雑なものになる。その上、多様な化合物を対応するため、辞書の追加、規則の調整、時にはプログラムの修正も必要である。システムの維持コストがかなり高い。それは、従来の命名法はコンピュータ処理に向いていないことに原因がある。

一方、情報処理の立場から見ると、従来の命名法における一番の問題点はむりやりにユニークな名称を追求するところにある。それは単に命名規則を複雑化するだけではなく、化合物の視点による様々の特徴を表現すること（動的表現）もできなくなった。

本研究では新しい名称に基づいて構造・名称変換システムの開発を試みる。名称は基本的には部分構造の名称と結合関係によって命名する。部分構造の間は特に優先順位を付けない。即ち、命名者の視点に応じて名称が付けられる。それによって生み出された同義語はシステムが吸収する。

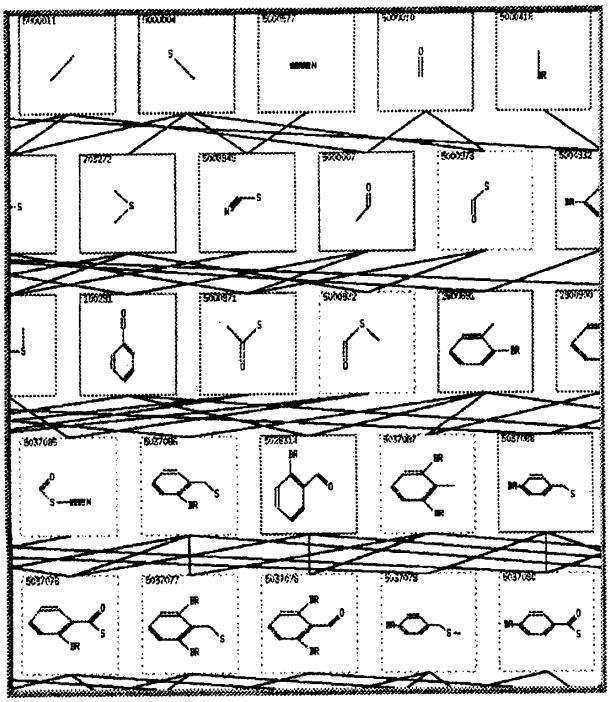
そのため、化合物の概念構造を構築し、化合物と化合物、化合物とその部分構造の関係を体系的に表現、記述し、それに基づいて化合物の命名、化合物間の類似度の計算および既に命名された化合物の名称と化合物間類似度を利用して名称・構造の類推を行なう。

Similarity and Analogical Reasoning in Conceptual Structures
Jianghong AN, Tetsuko ISHII, Yuzuru FUJIWARA
Institute of Electronics and Information Sciences
University of Tsukuba

3 概念構造の構築

概念構造とは概念間の同値関係、上・下位関係などを表す構造である。その構造は情報の自己組織化手法を用いて構築される [3]。情報の自己組織化とは、情報資源に内在する意味的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。

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持ち、その逆も言える。特に、名称の場合では、構造と直接関連するので分子構造に基づいて概念構造を構築するのは適切である。これで、本研究では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。構築された概念空間の一部を図−2で示す。



図− 2 化合物概念構造の一部

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ではなく、命名における基本構造（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。

化合物の類似度が分子構造間の重なりに依存するので化合物間の類似度は基本的には化合物の概念構造における距離（ノード間パスの長さ）と共通上位概念の数の関数とする。つまり、パスの長さが短いほど、共通上位概念が多いほど重なりが多く類似度が高い。

概念構造を構築する時、まず、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）に名称を付ける。それから、概念構造を利用して基本構造を含む構造にも名前を付けておく。また、利用に従って名称付きの構造が順々増えてゆく。

4 概念構造を用いる名称・構造の変換

構造から名称を作る手順：

（1）命名する構造を概念構造において同定する；

（2）同定できた場合対応する名称を返して終了する；

（3）同定できない場合は新しい構造としてまず概念構造に追加する；

（4）命名する構造の最近名称付き上位概念を取得する；最近というのは概念構造において最短パスを持つ名称付き上位概念である；

（5）その上位概念を語尾とし、パスに記述されたラベル（構造間の差）を修飾語として構造を命名する；最短パスを持つ名前付上位概念が複数個あった場合、複数の名称を作ることになる。

名称から構造を組み立てる手順：

（1）与えられる名称を概念構造において同定する；

（2）同定できた場合は対応する構造を返して終了する；

（3）同定できない場合はその名称を解析してそれに基づく新しい構造を作り出す；

（4）作り出した構造を概念構造に登録する。

5 むすび

概念構造に基づいて化合物の名称・構造の相互変換について述べた。概念構造を利用することによって名称・構造の相互変換は基本要素から変換するのではなく、できるだけ類似性の高い構造（名称）、即ち、主要構成要素の大部分を共有し、結合関係もほとんど同じ構造から出発することがポイントである。また、コンピュータ処理に適する名称を提供する。すなわち、ユニークな名称のみならず、視点の違いによって名称の多様性を認め、使用の便利さを図る。また、従来のシステムより維持が容易であることも示された。

参考文献

[1] Patric H. Winston: *Learning and Reasoning by Analogy*,Communications of the ACM, Vol.23, No.12, 1980, 689-703.

[2] International Union of Pure and Applied Chemistry: *Nomenclature of Organic Chemistry, Section A, B, C, D, E, F, and H*, Pergamon Press, Oxford, 1979.

[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, 1996.