

公開講演会要旨

# 化学情報の統計・ソフトコンピューティング 処理と実践的予測

佐藤 寛子<sup>†</sup>

(2003 年 11 月 5 日, 統計数理研究所 講堂)

## 1. はじめに

「化学は経験の学問である」といわれる。たしかに化学の歴史を振り返ってみると、化学者の頭の中で行われてきた化学情報と知識と経験が高度に統合化された情報処理により、新しい法則が発見され、それらの法則に適う実験結果や法則に合わない新現象が見出され、新たな情報や知識として蓄積されることの繰り返しによって、化学は発展してきたと捉えることができる。例えば周期表が作られ、その空白を埋める元素の性質が予測され、それらが発見されてきた歴史を見ても、化学が経験的な「情報」や「知識」を基盤として発展してきたといえる。

我々は、こうした従来の化学研究のかたちに加え、コンピュータを介した化学情報の有効な活用法を考えることで、化学研究のための新しい手段を開拓する研究を進めている(佐藤(2003))。人とコンピュータそれぞれに適した化学情報表現法の開発、新しい化学情報の提案、それらを利用した、化学研究の実践的問題を解決するための知的システムの開発などが主軸である。具体的には、立体的な特徴を含めた分子構造の表記法の開発、化学反応に関わる因子の表現法の開発、これらの特徴の可視化、さらに、化学反応予測や NMR(核磁気共鳴)スペクトル予測などの知的システムの開発を行っている。ここでは、化学における情報や統計・ソフトコンピューティング手法の活用について、我々の研究の概要を中心に紹介する。

## 2. 化学反応予測

化学反応予測とは、「任意の反応物の任意の反応条件下における化学反応によって得られる反応生成物とそれらの比を予測すること」である。化学反応は、種々の多様な因子の複雑な相互作用の結果として起こり、それぞれの因子の寄与の程度は反応によって異なる。この多様性と複雑さゆえ、化学反応を一般的に予測できる理論はいまだに確立されていない。

この問題を解くために、化学反応のコンピュータにおける新しい表現とモデル化の研究構想を描き、種々の研究を進めている。化学反応を支配する因子を多次元空間におけるデータであるとみなした場合、それらの複雑な相互作用の結果として起こる化学反応は、これら多次元データの非線形な繋がりであるとみなすことができる。この関係をうまく表現し、モデル化することができれば、複雑かつ多様な化学反応の表現と予測の糸口が掴めるのではないかと、この着想に基づいた研究構想である。ここでは、化学者が化学反応情報に共通する類似性を見出し、

<sup>†</sup> 国立情報学研究所 知能システム研究系：〒101-8430 東京都千代田区一ツ橋 2-1-2

詳細な解析によって規則性や理論を構築する過程を形式的に模倣し、データベースの化学反応の分類とモデル化を行う。化学反応分類においては、化学反応の方向を支配する種々の因子を数値化し、それらの類似性にもとづいた分類を行う。本研究において、非線形データを取り込んだ柔軟なモデル化の可能なニューラルネットワークなどのソフトコンピューティング手法は有効であると考えており、反応分類や予測モデル化の手法として利用している。

これまでに、シンプルな識別子による任意の化学反応の自動分類(Satoh et al. (1998))や、分子の立体的・静電的相互作用に基づく 3 次元的な分子の特性値化法 FRAU の開発(Satoh et al. (1999)), FRAU を用いた反応試薬の機能予測モデル化(Satoh et al. (2000a))など、種々の研究成果を得ている。

反応分類においては、データベースより任意に抽出した 131 件の化学反応について、反応部位の酸素原子の基本的な 6 種の物理化学パラメータの変化量によって化学反応を表現し、自己組織化ニューラルネットワーク: Kohonen neural network により自動分類を行った。その結果、これまでに化学者が経験的に分類してきた化学反応タイプと高い相関関係のあることが見出された。このことは、基本的なパラメータ表現だけでも、複雑な化学反応をある程度識別できることを示した興味深い結果であるといえる。

また、反応試薬の予測モデルにおいては、FRAU によって数値化された種々の試薬分子をもとに、カウンタプロパゲーションタイプの自己組織化ニューラルネットワークで分類・モデル化を行った。分類結果は、FRAU の特性値と化学反応試薬の機能との間に高い相関関係を示し、予測モデルは再現性と予測性を示した。さらに、化学合成実験による予測結果の検証と、実験結果のモデルへのフィードバックによる予測精度の向上についても確認している。

### 3. 核磁気共鳴(NMR)スペクトル予測

合成した化合物が望みの分子構造を有しているか、天然資源から採取された化合物がどのような分子構造を有しているかを調べる分子構造解析は、化学研究における不可欠な過程である。分子構造の解析のためには、核磁気共鳴(NMR)スペクトル、X 線構造解析、質量分析、赤外スペクトルなど、種々の機器分析手法が目的に応じて使い分けられている。中でも、NMR は現在、有機化学において最も有力かつ不可欠な分析手法として用いられている。

これまでに、NMR を用いた分子構造解析研究を支援し、また自動的に分子構造解析を行うことを目的とした、種々の NMR スペクトル予測システムや分子構造推定システムが開発されてきている。しかしいずれも立体化学の情報を的確に考慮できないために精度の高い予測が行えず、立体化学の考慮が必須である生物活性や物性などを対象とした実践的的化学研究に有効に利用することは不可能であった。

そこで我々は、こうしたシステムが根本的に抱えていたコンピュータにおける立体化学の規範的表記の問題を、独自の立体化学コード化法 CAST (CAnonical-representation of STereochemistry) を開発することで解決した(Satoh et al. (2000b), Satoh et al. (2001), Satoh et al. (2002))。CAST は分子内の 4 つの原子により構成される二面角にもとづくコード化手法であり、平面構造、絶対/相対配置、環構造の情報をそれぞれ独立に、あるいは組み合わせて表記し利用することができる。この CAST コードを分子構造記述の基盤として用い、立体化学を考慮することで  $^{13}\text{C}$ -NMR スペクトルを精密に予測するシステム CAST/CNMR を開発した(Satoh et al. (2003))。複雑な立体化学構造を有する種々の天然有機化合物や合成化合物に対して適用し、従来システムと比べて格段に精度の高い予測が可能であることを示し、立体化学の違いにより生じるスペクトルの違いよりも小さい誤差で予測できることを確認した。現在、実践的な活用に向けた種々の研究開発を進めている。

## 4. 最後に

近年のコンピュータやネットワーク環境の急速な進歩は、化学研究の方法論にも影響を与えつつある。たとえば、CPU能力の向上や並列・グリッドコンピューティング技術の飛躍的な進歩は、従来不可能とされた材料や生体分子などの大きな系への第一原理計算適用の可能性の道を開きつつある。また、本稿でも述べたように、統計的手法やニューラルネットワークなどのソフトコンピューティング手法は、多様な因子の複雑な相互作用によって引き起こされる様々な化学現象を解析するための有効な手段として活用され、機能性分子や材料の設計や予測モデル化等の研究における重要な情報技術基盤として位置づけられつつある。化学は実験科学であり、実験の重要性は今後も変わることはないであろう。しかし、目的を達成するための手段を拓げるものとしてコンピュータを捉えることで、情報とコンピュータの新しい有効利用のかたちが見えてくるのではないだろうか。こうした化学情報やコンピュータを基盤とした化学研究はまだ歴史は浅いが、無駄なものや有害なものをつくらないことの重要性が増すであろうこれからの化学において、予測を行った上で必要な実験のみを行う研究形態への変換などに大きく貢献できることが期待される。また、生命科学においても、分子の化学反応や動的な運動を基本とする生体内反応の解釈などにおいて分子と化学の視点は不可欠である。そのためにはバイオインフォマティクスのみならず、化学情報学の果たす役割も大きい。本分野への多くの研究者の参画を期待したい。

## 参 考 文 献

佐藤寛子 (2003) 『化学情報学—化学反応の系図と反応予測』, 丸善, 東京。

- Satoh, H., Sacher, O., Nakata, T., Chen, L., Gasteiger, J. and Funatsu, K. (1998). Classification of organic reactions: Similarity of reactions based on changes in the electronic features of oxygen atoms at the reaction sites, *Journal of Chemical Information and Computer Sciences*, **38**, 210–219.
- Satoh, H., Itono, S., Funatsu, K., Takano, K. and Nakata, T. (1999). A novel method for characterization of three-dimensional reaction fields based on electrostatic and steric interactions toward the goal of quantitative analysis and understanding of organic reactions, *Journal of Chemical Information and Computer Sciences*, **39**, 671–678.
- Satoh, H., Funatsu, K., Takano, K. and Nakata, T. (2000a). Classification and prediction of reagents' roles by FRAU system with self-organizing neural network model, *Bulletin of the Chemical Society of Japan*, **73**, 1955–1965.
- Satoh, H., Koshino, H., Funatsu, K. and Nakata, T. (2000b). A novel canonical coding method for representation of three-dimensional structures, *Journal of Chemical Information and Computer Sciences*, **40**, 622–630.
- Satoh, H., Koshino, H., Funatsu, K. and Nakata, T. (2001). Representation of configurations by CAST coding method, *Journal of Chemical Information and Computer Sciences*, **41**, 1106–1112.
- Satoh, H., Koshino, H. and Nakata, T. (2002). Extended CAST coding method for exact search of stereochemical structures, *Journal of Computer Aided Chemistry*, **3**, 48–55.
- Satoh, H., Koshino, H., Uzawa, J. and Nakata, T. (2003). CAST/CNMR: Highly accurate <sup>13</sup>C-NMR chemical shift prediction system considering stereochemistry, *Tetrahedron*, **59**, 4539–4547.