

化学における知的システム

神 沼 二 真

1. はじめに

知的システム (Intellective System) とは、人間の思惟 (Intellection) を支援したり、代行したりする計算機システムのことである。計算機技術の究極の目標のひとつは、このようなシステムを実現させることにある。現在、計算機があらゆる分野に普及し、浸透してゆくにつれ、計算機に期待される機能も次第に高次なものになってきている。つまり、比較的定型化され、定常化された仕事以外に、人間の専門家の判断や意思決定、創造活動を助けるような計算機への関心が急速に高まってきているのである。

このような試みの基礎となる技術のひとつは、思惟科学 (Intellective Science)、あるいは思惟工学 (Intellective Engineering) [12] とでも呼ぶべき、理論あるいはソフトウェアに関する知識である。

特に後者は、統計的決定理論、パターン認識、人工知能などという名のもとで、これまで研究が行われてきた。これはまた、理性的な行動計画技法であるORとも深い繋りがある。これらの学問はつまるところ、人間の思考のメカニズムをモデル化して、それを情報機械で実行させることを目的とする。しかし、これまで、そのような研究には、統一的な名称が与えられていなかった。筆

者は、インテレクトィクス (Intellectics, 思惟術) という用語を導入することによって、上記の諸分野を、ひとつの学問分野として統一したいと考えている。このような定義を前提とすれば、化学におけるインテレクトィクスの現状を紹介すること、および化学における知的 (インテレクトィク) なシステムを紹介することが、この論文の目的である。

2. 知的システムの分類

人間の思惟の基本パターンとしては、帰納、演繹、発想の3つがある。そして、人間の思惟はこれらの3要素を自在に構造化したものである。化学における思惟、すなわち、化学の研究や開発において、人間の思考を総合的に助けるようなシステムは、まだ存在していない。それゆえ、以下では、まず、帰納、演繹、発想 (創造) のそれぞれを個別に支援するシステムについて論じ、最後に、理性的な行動を支援する計画の技法について論ずることにする。

3. 帰納を支援するデータ解析システム

帰納とは、事実 (データ) を集め、そこから通則を導くことである。

化学においても、実験データの処理は、コンピュータ処理の大きな部分を占めており、その目的とするのは知識、またはその断片の生成にある。特に化学においては、Hansh解析 [7] に代表され

かみぬま つぐちか (財) 東京都臨床医学総合研究所

るような、分子の構造と、その属性（スペクトル特性、物性値、生物活性）との関係を明らかにすることが大きな課題である [4]。ここでの基本的なアプローチは、集めたデータから、それらの間の隠された関係（法則）を見出すという方法である。ところで、もちろん、これらのデータは多次元的なものであるから、データを眺めただけでは仮説に気づくのはむずかしい。しかも、データは記号や、数値だけとは限らない。スペクトルのような波形や、立体化学的な特性も扱える必要がある。

したがって、化学におけるデータ解析システムに要求される機能としては、

- (1) 数値や記号だけでなく、波形や立体構造データを扱えること、
- (2) 人間がデータに学び、データから発想することを助ける何かをもっていること、

があげられる。ところで、これまでデータ解析システムという点、BMDやSASのような統計パッケージが主流であった。統計パッケージは、これまでの統計学者の主たる関心事である仮説の検定手法は豊富に含んでいるものの、上記の2点への配慮は無きに等しかったと言えよう。

このギャップをうめるものが、パターン認識の手法である。しかし、統計的決定理論とパターン認識手法の本質的な違いは、双方の研究者の間でも正しく理解されていないことが多い。一言で言えば、パターン認識は統計学者が最近提唱し始めた、exploratory data analysis [18] の一種なのである。たとえば、上記のいわゆる構造活性相関の研究において、パターン認識手法 [8] が注目されているが、これは、仮説の検定を目的としたものでなく、むしろ、隠れた関係を明らかにするという、仮説づくりのほうが主たる目的なのである。

情報を縮約するというのも、パターン認識の重要な応用のひとつである。これには、たとえば各種のスペクトルを、そのピークの位置と高さ、お

よび半幅など少数のパラメータに圧縮して記録したり、分子の3次元構造を、いくつかの特性パラメータで縮約して表現するなどの研究がある。

さらに新しい傾向としては、統計を専門とする研究者の道具ではなく、化学を専門とする研究者の道具となるデータ解析システムを開発する試みがある。たとえば、BBN社で開発されたRS/1は、後述する化学研究サポートシステム、PROPHET [1] のデータ解析部分のみを独立したシステムとしたものである。この種のシステムの特徴としては、

- (1) 処理はコマンドによって行なわれる
- (2) いわゆる関係型のデータベースをもつ
- (3) 結果の図表による、表示機能がすぐれている

ことがあげられる。すなわち、図や表によって人間の発想を刺激することを狙っているのである。

4. 演繹を支援する判断システム

帰納によって導出された知識を用いれば、推測や意思決定を、ある程度自動的に行なうことも可能である。一般に、ある判断規則（群）を計算機に入れておいて、新しいデータが与えられた時、それに応じて演繹的推論を行なうための計算機システムを、知的応答システムという。

このようなシステムは、推定、予測、およびコンサルテーションに使われる。たとえば、与えられたスペクトルデータから分子構造を推定 (elucidate) したり、既知の化合物の毒性や薬効のデータから、新しい（仮想的な）物質の毒性や薬効を予測するためのシステムは、推測や予測のためのシステムである。また、ある既知の、あるいは未知の化合物をいかなる経路をたどって合成すべきかについて助言するシステムは、コンサルテーションシステムと呼ばれる。

このような知的応答システムを構成するには、2通りのアプローチがある。第1は、基礎となるデータを集め、これを統計的決定論やパターン認

識手法によって解析することによって、判断規則を作成する方法である。よく知られているように、この時使われるのは、いわゆる Bayes の決定理論やパターン認識の識別理論である。これらの方法論では、ある種の最適性が保証されているものの、データが十分揃わなかったり、欠損している場合には信頼性が急落するという欠点がある。第2の方法論は、人間の専門家の判断定石を、計算機で実行可能な規則の形で書き下して、これを推論規則とするものである。このような方法論は、最初医学診療の分野で、従来の統計的手法やパターン認識手法に代わる方法論として、1970年代の始めに提唱された [9],[17] ものである。その後それは、「人工知能 (AI)」手法 [3] と俗称され、第五世代計算機の計画などと関連して注目を集めるにいたった。

現在、第1の方法論は、主に構造活性相関の研究に使われているが、これは他で解説されるので、ここでは省略する。

第2の方法論は、構造の推定と合成コンサルテーションシステムの構築に使われている。

構造推定とは、既知の化合物のスペクトルや構造に関する情報(知識)を用いて、未知の化合物の分子構造を推定しようとするもので、佐々木慎一らの CHEMICS [16]、Lederberg や Buchannann らの DENDRAL [13] がよく知られている。CHEMICS は可能な分子の種類と大きさに制限を課す代わりに、高い予測精度を狙い、他方 DENDRAL は、基礎となる知識の量が大きく、ユーザーとの対話が柔軟に行なえるのが特徴である。いずれのシステムも実用までもう一步というところにある。しかし、本当に実用的なシステムに脱皮するためには、両方とも、方法論的飛躍が必要であろう。

5. 化学合成における論理と発見

人間の専門知識を、コンピュータで実行可能な推論規則として書き下ろし、この規則にもとづい

てコンピュータにも推論を行なわせて、逆に人間の判断を助ける道具にしようというのが、コンサルテーションシステムの思想である。しかし、コンサルテーションシステムの本質は、あくまでも、判断機能にあり、それを“助け”とするか否かは人間の側の勝手な解釈の問題になる。したがって、その限りにおいてコンサルテーションシステムは、統計的決定理論やパターン認識にもとづくシステムと同じ範疇に属しているのである。

ただし、次の点は重要である。コンピュータを自動判断システムとするならば、判断の主体はコンピュータにあるから、その論理が人間にわかりやすい必要はない。すなわち、結果よければ、すべてよしである。これに対して、判断の主体は人間にあり、コンピュータはその補佐役にすぎないとすれば、コンピュータの用いる論理は、ステップ・バイ・ステップで進む人間の判断形態に見合った、人間にとってわかりやすいものである必要がある。それゆえ、この場合には統計学やパターン認識の理論にもとづくアルゴリズムよりは、専門家の経験知識を整理することによって作成したアルゴリズムのほうがすぐれているであろう [9]。しかし、後者においては、前者のような最適性や信頼性に関する保証がないことは、もちろんである。

経験則に頼らざるをえない、というもうひとつの理由として、問題の複雑さがあげられる。一般に、統計理論やパターン認識理論は、データを集めることが簡単な対象に関しては、一応有効である。しかし、たとえば、合成経路を探索(正しくは探検)するというような複雑な問題においては、たとえば統計学的な検討に値するだけのデータを集めること自身、非常に困難な仕事である。したがって、このような問題領域においては、専門家の経験知識から出発せざるをえないのである。

このような前提に立つと、次に問題となるのは、専門家の経験知識を引き出し、それをもとに

してコンピュータ用の判断規則を作成する仕事である。つまり、専門家の勘やヒラメキ、アイデアの生み出し方などを、具体的な論理や発見的な手続きとして表現することである。このようなプログラムを、有機合成の分野で推進したのが、E. J. Corey の LHASA (Logic and Heuristics Applied to Synthetic Analysis) プロジェクト [2] である。同じような試みは、他の幾人かの研究者によっても行なわれている [5]。けれども、これまでのところ、最も成功を収めているのは、LHASA プロジェクト、あるいはその流れをくむシステム [19] であると言えよう。これは、このような研究にたずさわっている有機化学者として、Corey が最もすぐれていることの裏返しでもある。

LHASA プロジェクトで開発されている、同じ名称のコンピュータシステム、LHASA は、有機合成に関する知識ベースをもち、専門家の経験則にしたがって判断 (後戻り推論) を行なう。この意味で LHASA は、現在話題になっている、いわゆる人工知能システムであり、また知識 (工学) システムであると言える。それにもかかわらず、人工知能や知識工学の研究者は、Stanford の Heuristic Programming Project (HPP) の DENDRAL [13] や MYCIN [17] は宣伝しても、LHASA にはまったく言及していない。

それは、この分野の研究者が HPP の責任者である Feigenbaum [3] の宣伝上手に乗せられて、インテリククスにおける方法論と、その実現におけるコンピュータプログラムの作成技法とを混同しているからに他ならない。たしかに、LHASA は主として FORTRAN で書かれた単独のシステムであるので、HPP 流の INTERLISP によって書かれたシステムほど、コンピュータプログラムとしては洗練されていない。しかし、その構成と秘められた有用性から言えば、いわゆる人工知能あるいは知識工学的なシステムとしては、今後最も発展性のあるシステムであるという

ことができる。

6. モデルと発想 (創造) の支援

人間の発想のメカニズムが不明な現在、これをコンピュータに代行させることは不可能である。しかしながら、新しい化合物を探索したり、化合物の働きを解明したり、新しい合成方法を考えたりする局面において、化学者を助けることを目的とするシステムは現に存在する。この時カギとなるのは、分子の立体的構造である [6]。

化学の究極の目標のひとつは、分子の構造 (形) とその機能 (働き) の関係を明らかにすることである。これはまた、生命科学の基礎学問である分子生物学の目標のひとつである。そして科学の原理から言えば、構造を知ることによって、機能は予測することができるはずである。したがって、化合物のアイデンティティそのものである、分子の構造を表現することは、化学において最も重要な課題なのである。ところで、分子式や WLN など、これまで化学者が慣用としてきた構造表現は二次元的であり、しかも静的であった。つまり、それらは、現実に行きわたる機能を発現する、“生きた” 分子像を伝えるものとしてはきわめて不十分なものであった。

それゆえ、もしコンピュータによって分子を動的にかつ立体的に表現できれば、化学者はより豊かな発想を始めるに違いない。そしてそれは、構造の推定、構造活性相関予測、新しい合成法の探索、新しい化合物の設計などに非常に役に立つことになるであろう。

このような研究は、コンピュータによるドラッグ・デザインとの関連で、最近非常に注目されるようになった。その中心となるのは、与えられた (原子結合) 構造をもつ分子が実際にどうふるまうかを、動的にシミュレーションする、Molecular Shape Analysis [15] と呼ばれる研究と、その表示を行なうコンピュータグラフィックス技術である。

このような傾向は、化学における一種の思考革命ということができるだろう。ここにおいて、コンピュータ自身が発想したり、創造したりするわけではないけれども、コンピュータは発想や創造のためのなくてはならない道具となるのである。

現在このような路線で精力的なシステムづくりを行なっているのは、DuPontのPensakであり、彼のシステムである、TRIBBLE [14]はこの分野で最も進んだシステムである。正確に言えば、TRIBBLEはサブシステムとしてLHASAをもっている。もし、TRIBBLEに先に述べたデータ解析や構造推定の機能を加えれば、それは次に述べる研究開発のトータルシステムとなる。

7. 総合化

人間の思惟は、帰納、演繹、発想とバラバラに行なわれるものではない。それらはある目的に対して有機的に構造化されなければ意味がない。

ところで、化学における総合的研究サポートは次の機能を有する。

- (1) 人間にとって望ましい物質を探索し、
- (2) それを効率的に合成し、
- (3) 効果的に使用する。

しかし、これまで述べてきた中には、上記の(3)をサポートするシステムは含まれていない。実際、人類が現在直面している最大の課題のひとつが、実は(3)の問題なのであり、これはまた、ORの問題、正確には制御の問題そのものとも言えよう。

8. 化学とOR

上記の問題は結局のところ、環境における化学物質の流れとその制御の問題と、生体内における化学物質の流れとその制御の問題に帰着する。前者に関連するのは、環境汚染、それによる健康被害、毒性(安全性)試験、化学物質の使用規制などの問題であり、後者に関連するのは、生体内の物質輸送(キネティクス)、物質の代謝、作用(ダ

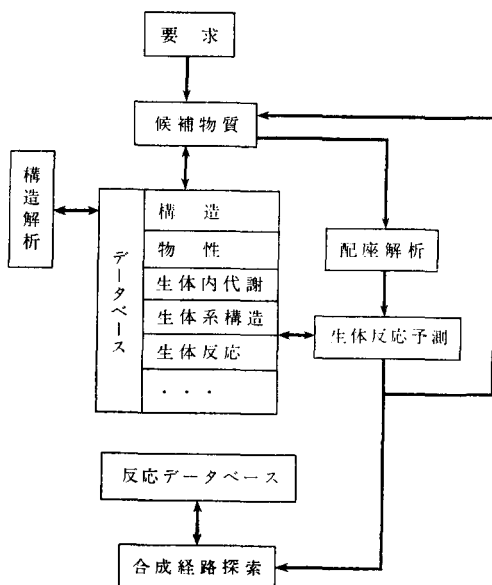


図1 化学における研究サポートシステム

イナミックス)、投与計画[10]などの問題である。

いずれの問題にも、基礎データを蓄積した、データ(知識)ベース[11]と、理性的な制御を計画するためのモデル作成とシミュレーション[20]が不可欠な道具となる。

さらに、いずれの場合も基礎となる知識は、化学物質と生体系、あるいは生態環境系との相互作用である。しかし、この知識は化学物質そのものに関する知識に比較して常に絶対的に不足している。したがって、われわれは絶えず、極度に貧弱な情報のもとで、ある特定の化学物質を使用したり、規制したり、禁止したりするという、意思決定問題に直面しているのである。

残念ながら、過去の事例を見るかぎり、この問題に関するわれわれの意思決定の仕方は、理性的というにはほど遠かった。一方で、コンピュータを駆使した理性的な思考術を発達させながら、他方で、理性的な思惟が真に必要な、人類の重要な意思決定問題は、ほとんど研究の対象とさえなっていない。

化学とORとの関係を考えると、「科学技術の進歩と調和した、科学技術を活用するための科学と技術の研究」の必要性を痛感する。それはまた

インテレクティクスの究極の目標でもある。もちろん、インテレクティクスが直接何かすぐれた解答を出すとは期待できなくとも、インテレクティクスが、人間のすぐれた知恵の発現の呼び水となることは、大いに期待できるであろう。

参 考 文 献

- [1] P. A. Castleman et al. : The Implementation of The PROPHET System, *AFIPS Conf. Proc.*, Vol. 43, 1974, 457
- [2] E. J. Corey and W. T. Wipke : Computer-Assisted Design of Complex Organic Synthesis, *Science*, 166, 1969, 178
- [3] E. A. Feigenbaum : The Art of Artificial Intelligence *AFIPS Conf. Proc.*, Vol. 47, 1978, 227
- [4] 藤田稔夫 : 構造活性相関の意義と役割 (薬物の構造活性相関, 構造活性相関懇話会編). 化学の領域増刊122号, 1979, 1
- [5] H. L. Gelenter et al. : Empirical Explorations of SYNCHEM, *Science*, Vol. 197, 1977, 1041
- [6] P. Gund : Molecular Geometry as an Indicator of Drug Activity, *Trends in Pharmacological Sciences*, Vol. 3, 1982, 56
- [7] C. Hansh and A. J. Leo : *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley, 1979
- [8] P. C. Jurs : Chemical Data Interpretation Using Pattern Recognition Techniques : in *Computer Representation and Manipulation of Chemical Information* (W. T. Wipke et al. eds.), Robert E. Kriger Pub. Co., 1981, 265
- [9] T. Kaminuma and K. Machii : Semiempirical Approach to Computer Diagnosis, *Proc. of 4th IJCP*, 1978, 889
- [10] 神沼二真 : 制御理論の治療への応用 (ライフサイエンスの進歩, 第5集, 日本医師会特別分科会編), 春秋社, 1978, 29
- [11] T. Kaminuma and A. Kurihara : A Data Base of Data Banks for Toxicological Information, *Analytica Chimica Acta*, Vol. 133, 1981, 707
- [12] 神沼二真 : 思惟工学よりみたアドヴァンスト・データベース・システム (「アドヴァンスト・データベース・システム」シンポジウム予稿集, 情報処理学会, 1981, 21
- [13] R. K. Lindsay et al. : Applications of Artificial Intelligence for Organic Chemistry : The Dendral Project, McGraw-Hill, 1980
- [14] D. A. Pensak, The TRIBBLE System, CBIワークショップニュース, No.5, 1982
- [15] R. Potenzzone, CAMSEQ-II, Ph. D Thesis, Case Western Reserve Univ., 1978
- [16] Shin-ichi Sasaki : Automated Chemical Structure Analysis of Organic Compounds -CHEMICS-F, (in *Informatio Chemistry*, S. Fujiwara and B. Mark Jr. eds.), Japan Scientific Societies Press, 1975, 227
- [17] E. H. Shortliffe : Computer-Based Medical Consultations : MYCIN, Elsevier, 1976 (日本語版, 神沼・倉科共訳, 診療コンピュータシステム, 文光堂, 1981)
- [18] J. Tukey : *Exploratory Data Analysis*, Wiley, 1977
- [19] W. T. Wipke et al. : SECS-Simulation and Evaluation of Chemical Synthesis : Strategy and Planning, (in *Computer-Assisted Organic Synthesis*, W. T. Wipke and W. J. Howe eds.), American Chemical Society, 1977, 97
- [20] T. Wipke et al. : Use of Computerized Methods to Predict Metabolic Pathways and Metabolites, *J. of Environmental Pathology and Toxicology*, 1978, 123