JP33 Chemish への原子団寄与法の組み込みに関する研究

(豊橋技科学大)○山田吉朗
(東大院工)荒川正幹、船津公人*
(三井化学(株))西村竜一

1. はじめに

1980 年代後半から、コンピュータが化学の分野で使われるようになり、大量の化学データの獲得とデータの処理が可能になってきた。それにともなって、大量のデータを数学的、統計学的に処理するための計量化学(Chemometrics)の分野が発展し、様々な手法が開発されてきた。計量化学の分野では、PLS 法や主成分分析法などの手法を用い、化学情報をモデル化する研究が行なわれている。

船津研究室では豊橋技術科学大学および東京大学において長年の間、PLS 法やニューラルネットワークなど、各種解析機能を備えたケモメトリクスソフトウェア Chemish[1]の開発を行っている。Chemish はケモメトリックスに興味を持っている研究者が手軽に利用することができるように、GUI による操作の簡便性、データ処理のしやすさ、結果の表示の見やすさなどに配慮して開発されたケモメトリックス総合システムである。動作環境は Microsoft Windows であり、C++言語によって記述されている。

本研究では、Chemish の利便性向上を目的として、化学構造ハンドリング機能、原子団寄与法によるデータモデリング機能、逆解析による新規構造提案機能などの機能を組み込んだ。以下に各機能の詳細と、Chemish を用いた原子団寄与法の解析例を示す。

2. 化学構造ハンドリング機能

化学データを読み込み、モデリングを円滑に進めることができるように、SDFile 形式のデータを Chemish に取り込むことを可能とするインターフェースの開発を行った。SDFile は MDL 社が定めた化学データファイルフォーマットで、複数の構造式と物性値などの分子特性情報を一つのファイルとして扱うことが可能である。また、化学データベースなどに広く利用されているファイルフォーマットである。SDFile を読み込む機能を Chemish に実装することで、本ソフトウェアと他の化学ソフトウェアとのデータのやり取りを容易に行うことが可能となり、本ソフトウェアの利便性向上が期待できる。また、メイン画面や散布図から対応する化学構造を表示する機能を開発し、Chemish による化学データの取り扱いがより効率的に行えるようになった。図 1 に Chemish の実行画面を示す。

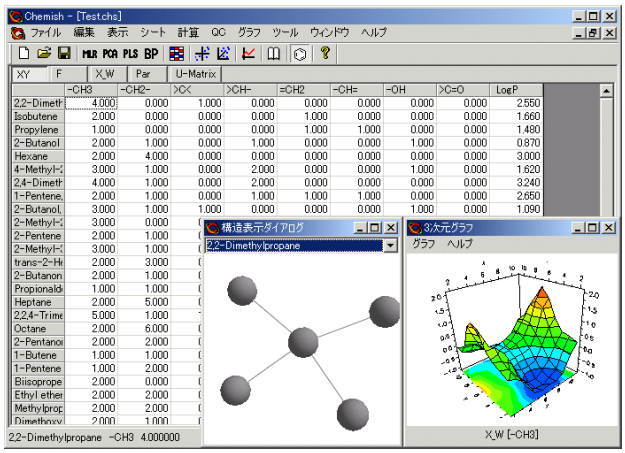

図 1. Chemish の実行画面

3. 原子団寄与法によるデータモデリング機能

原子団寄与法とは、構造中に含まれる各フラグメントの数量と物性値の相関関係をモデル化する手法である。本手法による物性の予測については様々な研究が報告されており、その有用性が明らかにされている。また、比較的単純なモデル式から物性値を予測することができるため、計算の高速性という観点においても有効な手法である。

原子団寄与法に基づくモデリングを実行するには、はじめに解析対象となる化学構造データの読み込みを行う。次にフラグメントライブラリを読み込み、解析対象構造に含まれる各フラグメント数を自動的に数えるプログラムにより、Chemish で取り扱うことのできるデータ形式に変換を行う。さらに Chemish のデータモデリング機能(MLR, PLS, Back Propagation, etc.)によってモデリングを行う。

4. 逆解析による新規候補構造提案機能

逆解析とは、モデル式の説明変数に値を入れて目的変数(予測値)を得る通常の操作とは逆に、目的変数に値を入れてそれを満たす説明変数の組(候補)を創出する操作である[2]。Chemish では全空間探索により逆解析を行うことが可能である。全空間探索とはデータ範囲などの制約条件を満たす説明変数の空間内の点を、設定した間隔にしたがって順次モデル

*funatsu@chemsys.t.u-tokyo.ac.jp

式に入力していき、目的物性値を満たすものを候補とする操作である。

原子団寄与法に基づくデータモデリングを行った後、逆解析計算を行うことによって、目的の物性値を満たす各フラグメント数の組が算出される。逆解析によって得られた結果から、構造式を組み立てることが可能ならば、その構造の物性値は目的の値である可能性が高いと考えられる。本ソフトウェアでは、フラグメントとその数量の組み合わせから、結合スタック法を用いることで、可能なすべての構造を列挙し、目的の物性値を持つ可能性の高い候補構造を提案することができる。図 2 に本ソフトウェアの簡単な実行例を示す。

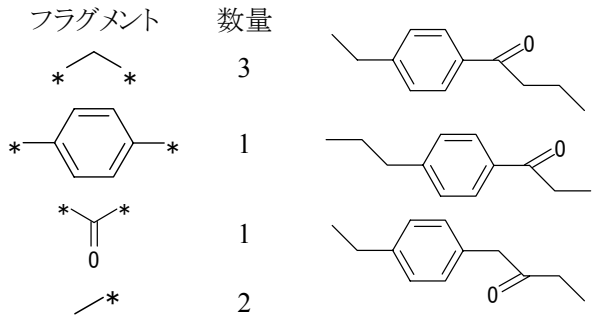

図 2. 構造生成プログラム実行例

5. フラグメントライブラリ作成支援機能

原子団寄与法によって予測性に優れたモデルを構築するためには、適切なフラグメントライブラリを用いることが重要である。解析の対象とする物性や、化合物群によって最適なフラグメントは異なるため、一般的なフラグメントライブラリ構築法は確立されておらず、個々の解析において最適化を行う必要がある。そのときの基準のひとつとなるのが、フラグメントのシェア率である。シェア率とは、各フラグメントが対象構造においてどの程度の割合を占めているかを数値化したものである。シェア率の低いフラグメントを用いることは、情報量の少ない説明変数を用いるということであり、モデルの精度や予測性を悪化させる要因になるものと考えられる。

そこで、Chemish を用いたフラグメントライブラリの作成を支援するために、各フラグメントのシェア率を算出する機能を追加した。これにより適切なフラグメントを選択することが容易となり、結果として精度のよい優れた相関モデルを構築することが可能となる。

6. 解析例

29 個の鎖式化合物の水／オクタノール分配係数 LogP をサンプルデータとして用い、原子団寄与法に基づくモデリングを行い、さらに逆解析による構造の提案を行った。PLS 法でモデリングを行った結果、Leave-one-out 法によるクロスバリデーションで求めた予測的相関係数の二乗値 Q^2 が、6 成分で 0.80 という良好なモデルが得られた。

この PLS モデルを用いて逆解析による新規構造の提案を行った。計算条件は要求特性(LogP の範囲)を 1.00～2.00 とし、各フラグメントの数の範囲を 0～1 とした。その結果、計 59 個の候補構造が提案された。解析に用いたフラグメントライブラリを図 3 に、提案された構造とモデル式より得られた LogP 値の一例を図 4 に示す。

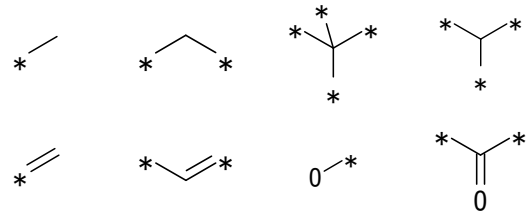

図 3. 解析に用いたフラグメント

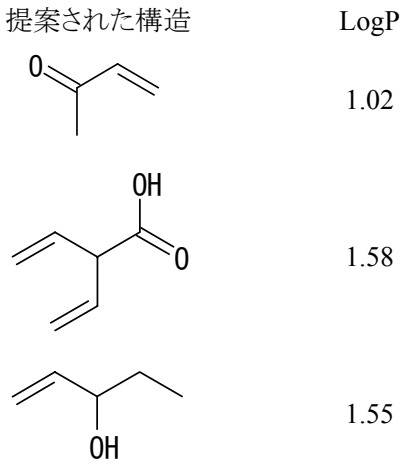

図 4. 提案された構造の一例

7. まとめ

本研究で開発した化学構造ハンドリング機能、原子団寄与法に基づくデータモデリング機能、逆解析による要求物性を満たす新規候補構造を提案する機能の追加により、化学データのモデリングにおける Chemish の利便性が向上した。また、フラグメントライブラリ作成支援機能の追加により、原子団寄与法に基づくデータモデリング機能の性能が向上した。

参考文献

[1] T. Tanada, et al., Journal of Computer Aided Chemistry, 1, 35-46 (2000).
[2] 橘大樹、豊橋技術科学大学修士論文(2003).