

N315

任意の有機化合物の溶解度を対象とした定量的構造物性相関

(東大工) (学)右田 啓哉・(正)船津 公人*

1. 緒言

化学プロセスにおいて、有機溶媒は化学反応場や分離操作の溶剤などとして重要な役割を担っている。化合物の様々な溶媒に対する溶解度は、適切な溶媒を選択するための重要な指標であるため、なるべく正確な値を使用することが望ましい。しかし、殊にプロセス設計の初期段階において、関与する多数の有機化合物についての実測値は入手し難く、溶解度の値を迅速かつ正確に推算するための手法が必要となる。

定量的構造物性相関 (QSPR) は、推算速度の速さに特徴がある物性推算のための方法論であり、物性データと化学構造パラメータの関係を統計的に解析し、定量的な数値モデルを構築する。しかし溶解度に関する従来の QSPR は溶媒ごとの解析にほぼ限定されており、汎用性に難があった。そこで本研究では任意の有機化合物および有機溶媒の溶解度を推算するモデルの新規な構築手法を提案し、2成分系の常温における溶解度データを用いて検討を行う。

2. 手法

本研究では SVR^[1] (Support Vector Regression) に基づく溶解度予測モデルを構築する。SVR は回帰分析手法の一種であり、カーネル関数を利用し柔軟な非線形モデリングが行える。カーネル関数とは、簡単に言えばデータ間の類似度を単一の数値で表現するための明示的な関数であり、今回はガウス型を用いる。

任意の溶解度を扱うためには任意の溶質・溶媒ペアについての類似度が必要であり、本研究では溶質の類似度 K_A と溶媒の類似度 K_B の積として定義する。すると溶解度モデルのパラメータとしては両者の構造パラメータのみを用いればよいことになり、相互作用についての複雑な関係式を含まずとも、柔軟にモデルを構築することができる。SVR の学習アルゴリズムから各サンプルに対する重み a と切片 b が得られ、訓練サンプル数 n 、2つのデータ x と x' のカーネル関数を K とすると、予測モデル $f(x)$ は以下の式となる。

$$f(x) = \sum_{i=1}^n a_i K_A(x, x_i) \cdot K_B(x, x_i) + b$$

3. 結果と考察

今回は溶解度データとして CrossFire Beilstein^[2] から抽出・整理した結果、代表的な 18 種の有機溶媒と分子量 1000 以下の 2343 個の溶質を対象とした。溶解度の値は 20~25 度の室温付近で測定されており、単位は g/L で常用対数をとった。構造パラメータとしては、溶質は DRAGON^[3] で計算される分子構造記述

子 77 個を、溶媒は極性、モル分子容量、誘電率、水素結合のアクセプター数とドナー数の 5 つを用いた。

構築されたモデル

は、平均絶対誤差 (MAE) が 0.18g/L

で、実測値と予測値の相関係数が 0.97 であり、良好な計算精度である。

図 1 に示す溶媒別の MAE をみると、極性溶媒についての予測精度が比較的良く、炭化水素系の溶媒については比較的低い傾向が

みられた。非極性溶媒のデータ数及び構造パラメータをさらに充実させる必要があると考えられる。

次にモデルの要因解析の結果として、予測値の変動に対する溶媒パラメータの寄与を図 2 に示す。総和としてみれば水素結合の寄与が最も大きく、極性やモル分子容量が後に続く。これは溶解における自由エネルギー変化に影響する性質の物理化学的な理解とほぼ一致する。つまり本手法によって溶解現象の因子をも定量化できることを表し、モデルを理論的側面から解釈できる可能性を示唆している。

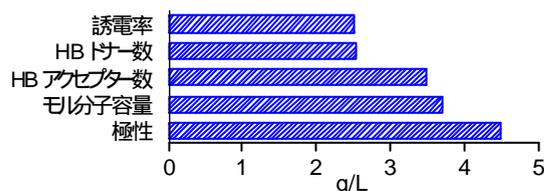


図 2 : 溶媒パラメータの寄与

4. 結論と展望

提案手法により、汎用的な溶解度モデルを構築し、良好な予測精度と理論的な解釈の可能性を示した。今後は非極性溶媒のデータやパラメータの補足、及び溶解度の温度依存性や混合溶媒への対応を行う。

参考文献

- [1] V. Vapnik, The Nature of Statistical Learning Theory, Springer, NY (1995).
- [2] <http://www.crossfirebeilstein.com/>
- [3] A. Mauri, et al. MATCH Commun. Math. Comput. Chem., 56, 237-248 (2006)

*Tel: 03-5841-7751, e-mail: funatsu@chemsys.t.u-tokyo.ac.jp