

Auto-Modeller : 高精度物性推算式の自動構築機能

○田島澄恵¹、井上靖雄¹、Matthew Segall²¹株式会社ヒューリンクス (〒103-0015 東京都中央区日本橋箱崎町 5-14)² Optibrium Ltd. (7226 Cambridge Research Park, Beach Drive, Cambridge, CB25 9TL, UK)

【緒言】

新規物質材料の開発やプラント設計などにおいて、物性推算は非常に重要な役割を果たす。沸点・密度・発火点などの基本物性については、各種物性推算式が提案されている。しかし、希望物性に対する推算式がない場合や、希望する予測精度を満たす推算式がない場合には、各研究者がオリジナル物性推算式を構築しなければならない。

物性推算式の構築には、どのデータセットを利用するか、どの記述子を利用するか、どのアルゴリズムを利用するかなどの選択が必要となる。各々について最適なものを選択するためには、専門知識やノウハウが必要となる。しかし、StarDrop (<http://www.optibrium.com/stardrop/>) の Auto-Modeller 機能を利用することにより、誰もが簡単に高精度推算式を構築することが可能となる。さらに、StarDrop の Glowing Molecule 機能との連動が可能となり、物性値を上げる/下げるために、分子構造のどの箇所を修正すべきかを視覚的に確認することが可能となる。

【方法】

推算式構築用データベースには、US Environmental Protection Agency (EPA) の Toxicity Evaluation Software Tool (T.E.S.T.) toolkit¹⁾で利用されたものを用いた。T.E.S.T.には、沸点・密度・発火点・熱伝導率・粘度・表面張力・水溶性 (-logS) 物性推算式が含まれている。これらは、表1に示す教育用/テスト用データ数を用いてマニュアル構築されたものである。

Auto-Modeller 機能では、構造多様性や物性値を加味し、データベースを教育用/評価用/テスト用データセットに自動分割することが可能であるが、今回は、T.E.S.T.との精度比較のために、T.E.S.T.における教育用/テスト用データセットを利用した。Auto-Modeller 機能における記述子は、321 個の SMARTS 形式による 2D 構造情報および 9 個の分子物性値 (logP, TPSA, 分子量など) を自動認識した後、データベースにおいて利用率の低い記述子、および、記述子間の相関が高いものを省いて利用される。アルゴリズムには、Partial Least Square/Radial Basis Function fitting/Gaussian Processes: Fixed など 9 種類が搭載されており、推算式構築後、最適なアルゴリズムモデルが自動判別される。

表 1 各物性推算式構築に利用した教育用データ数およびテストデータ数

物性	教育用データ数	テスト用データ数	物性	教育用データ数	テスト用データ数
沸点(°C)	3003	751	粘度(cP)	346	87
密度(g/cm ³)	6885	1722	表面張力(dyn/cm)	1136	285
発火点(°C)	6480	1620	水溶性 -logS(mol/L)	4064	1015
熱伝導率(mW/mK)	275	71			

【結果】

Auto-Modeller 機能を用いて、各物性推算式を自動構築した。Auto-Modeller 機能により最も精度が高いと評価された新規推算式と、既存式 (T.E.S.T.) との精度比較を表2に示す。各物性を比較すると、熱伝導率と表面張力において、既存式のほうが良い結果となった。しかし、全般的には、Auto-Modeller 機能による自動構築推算式は、マニュアル構築による既存式とほぼ同精度といえる。つまり、Auto-Modeller 機能により、誰もが簡単に、高精度推算式を構築することが可能であることが分かった。さらに、Auto-Modeller 機能を用いて構築した物性推算式は、Glowing Molecule 機能と連動しており、物性値を上げる/下げる箇所を視覚的に確認することが可能である。表面張力推算式における Glowing Molecule 機能利用例を、図1に示す。物性値を上げる箇所は赤色、物性値を下げる箇所は青色で表示される。左側は、表面張力値を上げる寄与度の大きい塩素原子箇所が、赤色表示されている。右側では、塩素原子がフッ素原子に置換されており、赤色表示が無くなり、表面張力値が下がっている。

これらの結果から、StarDrop の Auto-Modeller 機能は、非常に有効な物性推算式構築ツールであることが分かる。

表2 Auto-Modeller 機能による新規推算式と既存式(T.E.S.T.)との精度比較

物性	Best T.E.S.T. Model						
	手法	R ²	RMSE	R ² _{pearson}	$\frac{R^2_{\text{newton}} - R^2_{\text{t.est}}}{R^2_{\text{newton}}}$	k	MAE
沸点(°C)	Consensus	/	15.73	0.96	0.00	1.00	10.00
密度(g/cm ³)	Consensus		0.09	0.94	0.01	0.99	0.04
発火点(°C)	Consensus		28.90	0.88	0.01	0.96	18.00
熱伝導率(mW/mK)	Consensus		13.19	0.85	0.10	0.96	8.39
粘度(cP)	Consensus		0.20	0.86	0.00	0.86	0.12
表面張力(dyn/cm)	Consensus		1.82	0.92	0.02	1.00	1.23
水溶性 -logS(mol/L)	Consensus		0.86	0.87	0.02	0.93	0.59
物性	Best StarDrop Model						
	手法	R ²	RMSE	R ² _{pearson}	$\frac{R^2_{\text{newton}} - R^2_{\text{t.est}}}{R^2_{\text{newton}}}$	k	MAE
沸点(°C)	GPFixed	0.97	14.40	0.97	0.00	1.00	8.32
密度(g/cm ³)	RBF	0.95	0.08	0.95	0.00	0.99	0.04
発火点(°C)	RBF	0.90	26.80	0.90	0.01	0.97	16.10
熱伝導率(mW/mK)	GPRFVS	0.83	13.11	0.84	0.03	0.97	9.86
粘度(cP)	GA RBF	0.94	0.13	0.95	0.00	0.92	0.09
表面張力(dyn/cm)	RBF	0.87	2.26	0.87	0.01	1.00	1.07
水溶性 -logS(mol/L)	RBF	0.88	0.88	0.88	0.01	0.93	0.52

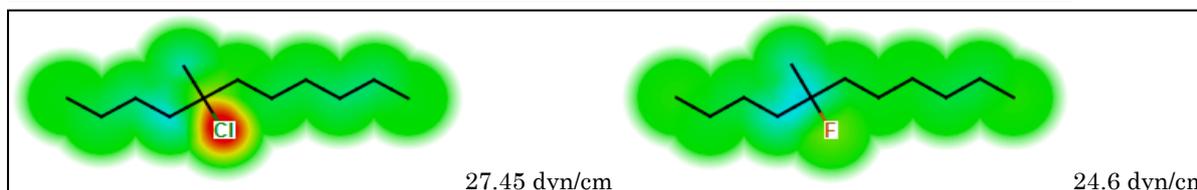


図1 Glowing Molecule 機能の利用例

参考文献

- 1) US Environmental Protection Agency. Quantitative Structure Activity Relationship. Environmental Protection Agency. [Online] 2011. [Cited: May 20, 2011.] <http://www.epa.gov/nrmrl/std/cppb/qsar/#TEST>.