

シンボリック回帰を用いた化学における数理モデルの自動構築

(早大先進理工¹・早大理工総研²・(株)三菱ケミカルホールディングス³)○磯田 拓哉¹・中野 匡彦^{2,3}・清野 淳司^{1,2}

Automatic Construction of Mathematical Models in Chemistry Using Symbolic Regression

¹*School of Advanced Science and Engineering, Waseda Univ.*, ²*WISE, Waseda Univ.*, ³*Mitsubishi Chemical Corp.*) ○Takuya Isoda,¹ Masahiko Nakano,^{2,3} Junji Seino,^{1,2}

Many chemical laws and principles have been constructed inductively from experimental data so far. Recent developments of informatics can expect that computers automatically construct new and complex chemical laws and principles from huge amount of data obtained from calculations and experiments. In this presentation, we will show that several chemical laws can be reproduced using the symbolic regression, which can automatically construct an optimal mathematical model using chemical features, four arithmetic operations, and so on.

Keywords : Symbolic Regression; Chemical Law; Mathematical Model; Informatics

【緒言】これまで多くの有用な化学法則や原理が、実験データから帰納的に構築されてきた。近年のインフォマティクス技術の発展により、計算・実験を含む膨大なデータから新規で複雑な化学法則・原理を、コンピュータが自動的に構築できることが期待されている。本研究では、解釈可能な法則を導き出すインフォマティクス技術であるシンボリック回帰を用いて、これまで提案された化学法則を再現できるか検証した。

【方法】本研究ではシンボリック回帰として、物理変数と数学的演算を組み合わせるヒューリスティックに最適な数理モデルを構築する遺伝的プログラミング (GP)¹、算術演算子による特徴量生成と LASSO 回帰による特徴量選択を繰り返し、最終的な式の形を決定する再帰的 LASSO 型

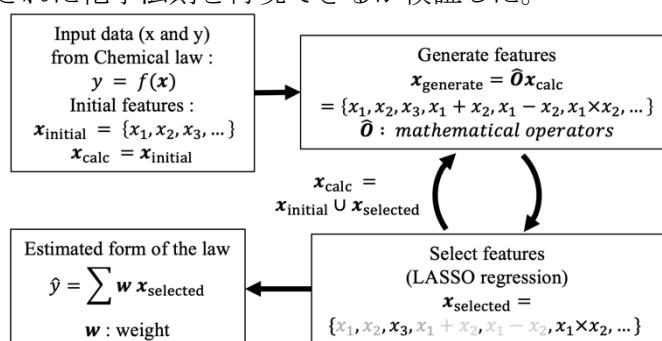


Fig. 1. Flowchart of the customized RLS method.

シンボリック回帰 (RLS)²、対称性・分割可能性・簡略化・次元解析を活用して、より複雑な関数から成る物理法則を発見する AI-Feynman³を用いた。GP と AI-Feynman は既存の Python ライブラリ (gplearn、aifeynman) を、RLS は先行研究²) の思想に基づき、独自にカスタマイズして実装した (Fig. 1)。

【結果】アトキンス物理化学に記載されている法則に対し、実際に取り得る範囲の数値を代入してデータセットを作成した。ここでは、原理上 RLS の特徴量生成に必要な最低繰り返し回数 (cycle) に応じて、法則の複雑さを分類した。Table 1 に各 cycle における代表的な式と、各手法で導出までに要した計算時間を示す。GP はあらかじめ定数を与える必要があること、AI-Feynman は GP や RLS よりも長い計算時間が必要であることがわかった。発表当日は手法の詳細と他の法則についての導出結果についても示す予定である。

Table 1. CPU times (second) for derivations of chemical laws.

Cycle	Name	Formula	GP	RLS	AI-Feynman
0	Henry's law	$\rho_B = K\chi_B$	2.6 ^{*1}	9.5	1243.8
1	Raoult's law	$p_A = \chi_A p_A^*$	2.6 ^{*1}	7.5	701.8
2	Rate equation of secondary reaction	$t = ([A]_0 - [A])/k_r[A]_0[A]$	N.C. ^{**2}	18.1	3599.6
3	Moseley's law	$\nu = A(Z - b)^2$	N.C. ^{**2}	113.0	4729.0

^{*1} Constants are given for derivation in advance. ^{**2} Not converged.

1) <https://gplearn.readthedocs.io/>. 2) Y. Iwasaki, M. Ishida, Probab. Math. Stat. 07, 16 (2021).

3) S. M. Udrescu, M. Tegmark, Sci. Adv. 6, 2631 (2020).