

Transactions of JSCES, Paper No.

遺伝的プログラミングによる数理モデリング*

Mathematical Modeling by Genetic Programming

白神叶衣¹, 小野謙二²

Kanae SHIRAGAMI and Kenji ONO

¹ 九州大学大学院システム情報科学府 (〒819-0367 福岡市西区元岡 744)² 九州大学情報基盤研究開発センター (〒819-0367 福岡市西区元岡 744)

Genetic programming is one of evolutionary algorithm. It searches for optimal solutions through iterative evolution and selection of genes. It can be applied to equation discovery as a method for solving symbolic regression problems. To solve symbolic regression problems for various phenomena and to obtain concise and precise equation, we need to enhance exploration capability of GP. We try to achieve it by appropriately taking in elements of terms by referring to existing equations and terms on subject-area. In this paper, we construct GP which considers how to generate and calculate terms and how to choose useful ones. We report the effort to apply developed GP for modeling wind turbine wake.

Key Words: Equation Discovery, Genetic Programming, Multiple Regression, Wake Model, Mathematical Modeling

1. 緒 言

複雑な物理現象に対して簡潔で正確なモデリングを行うことで、現象の理解に役立てることができる。物理現象を適切に表現する方程式を発見するには、強化学習を用いる手法⁽¹⁾、あらかじめ項の候補を設計してスペース回帰を行う手法⁽²⁾など様々な手法がある。その中でも我々は遺伝的プログラミング(GP)によって記号回帰問題を解くことで方程式発見を行う手法に着目した。GPを用いた方程式発見の利点として、広大な探索空間をランダムに探索できる点、一度設計したGPのフレームワークを様々な現象に適用して方程式推定を行える点が挙げられる。我々の従来の研究⁽³⁾では、偏微分項を探索的に生成できるGPを構築した。構築したGPの能力を検証するために、流体の基礎方程式を用いて空間1次元のシミュレーションデータを作成し、利用した方程式をGPを用いて推定できることを確認した。本論文ではこのGPの適用範囲を広げ、後流モデルの推定に用いるための手法の紹介と後流モデルの推定実験を行う。

2. 偏微分項を探査するGP

* 原稿受付 2025 年 99 月 99 日, 改訂 2025 年 99 月 99 日, 発行 2025 年 99 月 99 日. ©2025 日本計算工学会.

Manuscript received, Dummy 99, 2025; final revision, Dummy 99, 2025; published, Dummy 99, 2025. Copyright ©2025 by the Japan Society for Computational Engineering and Science.

我々が従来の研究で構築したGPの概要、そのGPを方程式発見に適用する際の流れと得られた成果について説明する。

2.1 従来のGPの概要

遺伝子は関数記号と終端記号を組み合わせて表現される。遺伝子が計算式や項を表現するとき、関数記号はどのような演算を施すかを表す。終端記号は、関数記号が表す演算の引数となる。従来のGPでは、関数記号と終端記号を以下のように定義してきた。

- 関数記号：乗除 (mul, div)、微分 (d_dt, d_dx)
- 終端記号 : t, x, u(速度)

遺伝子のデータ構造として演算リストを用いる。

Fig.1に演算リストの例を示す。項の値を計算する際はリストの末尾の保持番号 (Fig.1の場合は 2) から終端記号を表す保持番号 (Fig.1の場合は 0) に辿り着くまで演算を解釈する。

{“0”:[“load”, “u”], “1”:[“d_dx”, 0], “2”:[“d_dx”, 1]}
Fig.1 Operation list representing $\frac{\partial^2 u}{\partial x^2}$.

2.2 従来のGPによる方程式発見の流れ

GPの計算手順を Fig.2 に示す。

まず、初期集団を生成する。

次に、遺伝的演算(複製、交叉、突然変異)を施して新たに N 個の項を生成する。

項を生成したのち、フィルタリングを施して Least

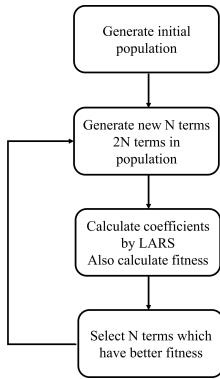


Fig.2 Each calculation step in GP.

Angle Regression(LARS)⁽⁴⁾ に入力する項を選択する。フィルタリングは以下の手順で行い、LARS に入力する項の集合を作成する。

- 項の集合を目的変数との相関が高い順に並べておく
- 新たに項を追加する際、項集合の先頭から順に新たに追加する項と項集合内の項の相関係数を計算する
- 相関係数が 0.95 以上である場合、項の複雑度が低い方を項集合に残し、もう一方は項集合から削除する
- 相関係数が 0.95 未満である場合、項集合の次の要素に対して上の処理を行う

項の複雑度は、項に含まれる関数記号の数と終端記号の数の和で計算する。このような手順でフィルタリングを行うことで $(u * v) / u$ と v のように表記は異なるものの意味が全く同じ項があった場合に、表記が簡単な v を LARS に入力する項として残すことができる。また、多重共線性の回避が期待できる。

LARS により各項の係数を計算したのち、方程式を生成する。方程式への寄与が高い項を優先して式に用いるために、重要度という指標を新たに定義して各項の方程式への寄与を判断する。重要度は係数の絶対値の総和に対して各項の係数の絶対値が占める割合によって計算する。

適合度の計算の際、線形独立である項、方程式に用いられた項、方程式の中での寄与が高い項が残るよう考慮する。

2N 個の項の中から、適合度が高い N 個を選択する。以上の手順を繰り返す。

2.3 従来の GP の性能の検証

従来の研究⁽³⁾ で、流体の基礎方程式である線形移流方程式、拡散方程式、Burgers 方程式から空間 1 次元のシミュレーションデータを作成し、左辺 $\frac{\partial u}{\partial t}$ に適合する右辺の形式を探査した。いずれの場合についても、10000 個オーダーのシミュレーションデータから元の方程式の特徴を再現する方程式が得られることを確認した。また、シミュレーションデータにノイズがある

場合でも元の方程式の特徴を再現する方程式が得られることを確認した。

3. 風車後流の推定

大規模な風力発電を行うウインドファームでは、風車を通過した風(後流)には速度減少や乱れが生じるため風下の風車に発電効率の低下などの影響を与える。後流をモデリングすることで風下の風車が受けける風を分析し、風車の配置や設計に役立てて発電効率の改善に役立てることが期待できる。また、このような後流モデルを GP を用いて推定することができれば探索範囲の広さと探索のランダム性により、画期的なモデルを得られる可能性がある。

風車の後流をモデルリングしたものとして Bastankhah and Porté-Agel モデル (BP モデル)⁽⁵⁾ が挙げられる。BP モデルのような既存の後流モデルの式を参考にして GP の関数記号や終端記号などを十分に定義しておくことで、後流モデルの式を探索するための十分な探索範囲を設計することができる。しかし、BP モデルの式は複雑であるため GP で推定することが難しい。そこで、問題設定が行いやすく GP による推定がやりやすい CFD Porous Disk モデル (PD モデル)⁽⁶⁾ の推定を試みることにした。

PD モデル (Eq.(1)) は Navier-Stokes 方程式の外力項に空力抵抗を導入したものである。

$$\frac{\partial \bar{u}_i}{\partial t} + \bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} = - \frac{\partial \bar{p}}{\partial x_i} + \frac{1}{R_e} \frac{\partial}{\partial x_j} \frac{\partial \bar{u}_i}{\partial x_j} + \frac{\partial \tau_{ij}}{\partial x_j} - u_i |U| \phi C_{RC} \Omega \quad (1)$$

$$f_i = -u_i |U| \phi C_{RC} \Omega \quad (2)$$

$$\phi = e^{\frac{-r^2}{2\sigma^2}} = e^{\frac{-y^2}{2(\frac{0.5R}{3})^2}} \quad (3)$$

$$|U| = \sqrt{u_i u_i} \quad (4)$$

Ω は格子ごとの体積占有率、 ϕ (Eq.(3)) は y 軸方向の正規分布、 C_{RC} は係数である。速度、圧力のデータをもとに Eq.(1) を用いて外力項の値を計算し、その値に適合するような式を GP で探索し、PD モデルの外力項の定義 Eq.(2) と一致するか確かめる。

4. 風車後流推定のための GP

この章では、従来手法の GP からの変更点を説明する。

4.1 関数記号と終端記号の定義

終端記号として空間変数の x と y 、速度 u と v 、圧力 p 、また $|U|$ を用いる。関数記号の設計はモデリングを行いたい現象に関連する既存の方程式などを参考を行う。従来用いていた関数記号と BP モデルを比較し、BP モデルを表現するのに必要な e^{\wedge} を追加した。関数記号 e^{\wedge} については次節で詳細を説明する。

4.2 $e^{-\frac{r^2}{\sigma^2}}$ の導入

関数記号に e^{\wedge} を加える。BP モデルの外力項では e^{\wedge} は y 方向の正規分布 $e^{-\frac{r^2}{2\sigma^2}}$ として現れる。 e^{\wedge} の指数部分も含めて初めから推定しなければ完全なモデリングができたとはいえないが、現在の GP の設計では指数部分も推定するのは困難であるため今回は e^{\wedge} の指数を y 軸方向の正規分布の形式に固定して推定を行った。すなわち、関数記号 e^{\wedge} が項の中で用いられる際は必ず $e^{-\frac{r^2}{2\sigma^2}} = e^{-\frac{y^2}{2(\frac{0.5R}{3})^2}}$ として解釈される。Fig.3 のように、 e^{\wedge} を用いる保持番号では便宜上第二引数を指定するものの、実際にはその保持番号をたどることがない。

{"0":["load", "u"], "1":["d_dx", 0], "2":["e^", 1]}

Fig.3 Operation list representing $e^{-\frac{r^2}{2\sigma^2}}$.

今後、事前知識なしで指数部分も推定可能か調査する必要がある。

4.3 式の x 成分と y 成分を同時に推定する

これまで空間 1 次元のシミュレーションデータから式を推定していたが、今回は空間 2 次元のシミュレーションデータから式を推定する。外力項の x 成分と y 成分の推定を別々に行うと全く異なる意味を持つ式が生じるおそれがある。また、一方の成分の特徴をよく表す式がもう片方の成分の特徴をよく表すとは限らない。したがって、 x 成分と y 成分の式を同時に推定し、その中で x 成分と y 成分どちらの特徴も表現できるような式を得られるようにする必要がある。そこで、従来の手順に以下の処理を加える。

- 係数の計算後、どちらか一方の成分で著しく係数が小さくなる場合、その項は式に用いないようにする。
- 係数の計算後、どちらか一方の成分で著しく重要度が低くなる場合、その項は式に用いないようにする。
- 適合度は、 x 成分の式と目的変数で計算した最小二乗誤差と y 成分の式と目的変数で計算した最小二乗誤差の和として計算する。

5. 実験

5.1 問題設定

今回の実験では空間 2 次元の PD モデルのシミュレーションデータをもとに計算した f_i の値を GP へ入力し、この f_i と項の誤差をもとに適合度を計算して式の探索を行った。モデル内の各変数の値は以下の通りである。

Table 1 Model parameter of PD model simulation

R_e	Smagorinsky model constant	C_{RC}
290000.0	0.17	15.0

シミュレーションの条件は以下の通りである。

Table 2 Spatial Condition of PD model simulation

range of x	N_x	range of y	N_y	dx and dy
[-5.01, 19.01]	2403	[-7.51, 7.51]	1503	0.01

Table 3 Temporal Condition of PD model simulation

range of t	N_t	dt
[149., 150.]	100	0.0005

また、用いたデータの条件は以下の通りである。

Table 4 Spatial Condition of Input data

range of x	N_x	range of y	N_y	dx and dy
[-0.25, 0.25]	52	[-0.76, 0.76]	202	0.01

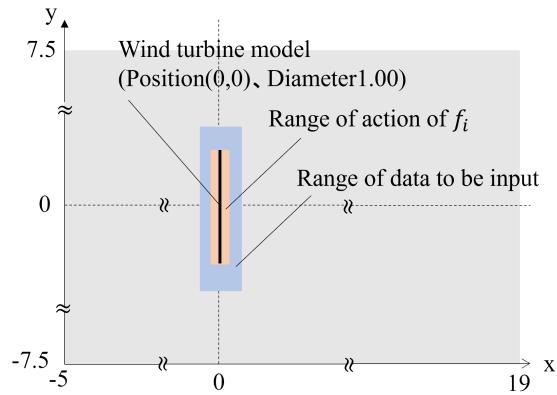


Fig.4 Simulation data.

Fig.4 の灰色の部分がもととなるシミュレーションの計算領域、橙色の部分が外力項の作用範囲、青色の部分が GP で用いるデータの範囲である。Schmit らの研究⁽⁷⁾を参考に交叉確率 75%、突然変異確率 1%に設定した。また、従来手法と同様にアイランドモデル⁽⁸⁾を用いて並列実行した。遺伝子数 32 の 8 個の島それぞれで計算を行い、100 世代ごとに島同士の個体の交換を行いながら計 300 世代の計算を行った。

5.2 実験結果

以上の設定で GP を動かした。理想的な出力は $f_x = -15.0u|U|e^{-\frac{y^2}{2.0\sigma^2}}$ 、 $f_y = -15.0v|U|e^{-\frac{y^2}{2.0\sigma^2}}$ であるが、プログラムの出力にこのような項は現れなかった。

6. 議論

従来の手法では LARS へ入力する項を相関係数によってフィルタリングしていた。しかし、今回の BP モデルの推定では生成された項同士の相関係数が高いものが多いという問題があった。

その一例として、 $e^{\frac{-r^2}{2\sigma^2}}$ と、 $e^{\frac{-r^2}{2\sigma^2}}$ に変数を乗じた形式の項の相関係数が非常に高くなっていた。 $e^{\frac{-r^2}{2\sigma^2}}$ は生成されやすいため、仮に $u|U|e^{\frac{-r^2}{2\sigma^2}}$ が生成されても $e^{\frac{-r^2}{2\sigma^2}}$

との衝突で、複雑度の高い $u|U|e^{\frac{-r^2}{2\sigma^2}}$ は項の集合から除外されやすい。よって、 $e^{\frac{-r^2}{2\sigma^2}}$ に変数を乗じた形式の項は式生成に用いられにくくなっていた。

表記が異なるが同じ意味を示すような項が同時に回帰に用いられないようにするためにフィルタリングを行っていたが、項同士の相関係数が高いものが多いことから項同士の相関係数が 0.95 を超える場合は単純な項のみを残すという条件では用いられる項はごくわずかになくなってしまう。したがって、回帰に用いる項を事前に減らすよりも回帰を行う過程で回帰式に用いる項を絞り込む方がより柔軟で質の良いモデリングを行うことができると考えられる。例えば、計算の過程で多重共線性を考慮できる回帰手法を導入することが必要であると考えられる。

7. おわりに

後流モデルを推定できるように遺伝的プログラミングの実装方法を改良した。現時点では PD モデルの外力項の形式をうまく推定することができない。多重共線性を考慮した回帰手法を用いることで得られる方程式の質を向上できる可能性があるので、引き続き PD モデルの外力項の推定を行っていく予定である。また、現時点では $|U|$ と $e^{-\frac{r^2}{2\sigma^2}}$ の形式が推定しやすいうように終端記号と関数記号を設定しているため、これらが初めから生成されるような実装に変更をしていきたい。

参考文献

- (1) Petersen, B. K., Landajuela, M., Mundhenk, T. N., Santiago, C. P., Kim, S. K., and Kim J. T., Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients, arXiv:1912.04871
- (2) Brunton, S. L., Proctor, J. L., and Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proceedings of the National Academy of Sciences*, 113(15), 2016, pp.3932–3937.
- (3) 小野謙二, 古賀壱成, 遺伝的プログラミングによる支配方程式の推定, 日本計算工学会論文集, 2020 卷 (2020)1 号, p.20201004.
- (4) Efron, B. and Hastie, T. and Johnstone, I. and Tibshirani, R., Least Angle Regression, *The Annals of Statistics*, Chapter 6, ed. A. Einstein, Vol.32, No.2, 2004, pp. 407–451.
- (5) Bastankhah, M. and Porté-Agel, F., A new analytical model for wind-turbine wakes, *Renewable Energy*, Volume 70, October 2014, pp. 116–123.
- (6) Uchida, T., Taniyama, Y., Fukatani, Y., Nakano, M., Zhiren Bai, Tadasuke Yoshida, T. and Inui, M., A New Wind Turbine CFD Modeling Method Based on a Porous Disk Approach for Practical Wind Farm Design, *Energies* 2020, 13(12), 3197.

- (7) Schmidt, M. D. and Lipson, H., Distilling free-form natural laws from experimental data, *Science*, Vol. 324, Issue. 5923, 2009, pp. 81–85.
- (8) F. Francisco, S. Giandomenico, T. Marco, and V. Leonardo, Parallel Genetic Programming, In *Parallel Metaheuristics*, 2005, pp. 127–153.