

GA による GMDH モデル構築の最適化

3B-11

*吉原 郁夫 **佐藤 周一

*日立製作所 システム開発研究所 **日立システムテクノロジー

1. はじめに

GMDH(Group Method of Data Handling)は、簡単な非線形式を組み合わせて複雑な非線形モデルを構築する手法であり¹⁾、予測、システム同定、モデリングなどに幅広く応用されている。しかし、説明変数の組合せを決める際、人が介入するため、属人的であったり、時間がかかる等の問題がある。

我々は遺伝的アルゴリズム(GA)を用い、GMDH のモデル構築を自動化することにより上記問題点を解消しようと考えた。

2. GMDH

システムへの入力変数を x_1, x_2, \dots, x_n とし、出力を y とする。普通、簡単なモデル式としては2次式 $G(x_1, x_2) = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2$ を用い、その組み合わせで高次のモデルを構成する。まず、二変数ずつの組み合わせを考え、出力 y に対する近似度の良いもの p 組を選ぶ。次に、入力変数及び p 個のモデル式の出力 $u_{ij} = G(x_i, x_j)$ を改めて入力(説明変数)としてその組み合わせを考え、出力 y に対する近似度の良いもの q 組を選ぶ。以下同様にして高次モデルを構成してゆく^{2) 3)}。さらに高次にしても精度が向上しないとき「完全表現」と呼び、それを GMDH のモデル式とする。

3. GA によるモデル構築の最適化

一般に、高次のモデルは入力の変動に敏感で、特にノイズに弱い傾向があるため、ここではモデル式の次数があまり高くならないように抑えることを基本方針とする。

(1) GMDH モデルの二分木表現

入力変数の組み合わせは、入力変数を葉とし、各ノードを $G(\cdot, \cdot)$ とする二分木で表現できる。説明変数が n 個のとき、全ての説明変数を葉に持つ二分木の高さは、 $h = \log_2 n \uparrow$ (\uparrow は切り上げの意) 以上である。逆に、高さ h の木が表せるのは、説明変数が $n^* (= 2^h)$ 個までであり、 x_i の次数も n^* までである。

説明変数が n^* 個の場合、モデルは葉が全部のついた完全な二分木で表される。ここでは n 個の変数に $n^* - n$ 個のダミー変数 d を補い、常に n^* 個の変数を扱うようにする(図1)。

(2) GA による モデル構築 のアルゴリズム

n^* 個の変数の並びをそのまま染色体とする。主な遺伝的操作は次の通りである。

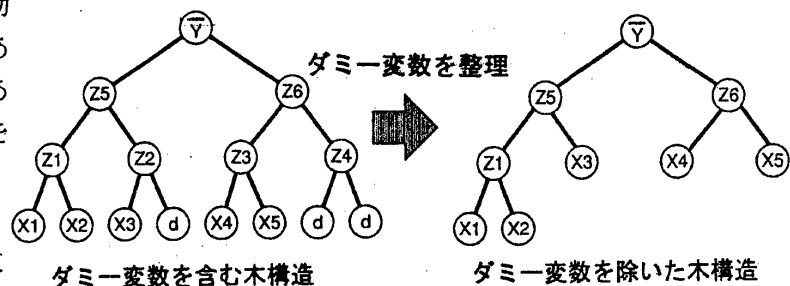


図1 GMDH モデルの二分木表現

- a) 交差……一点交差の位置 s をランダムに決定する。片方の親個体の染色体から s の左側の遺伝子列をとり、そのまま子個体に転写する。次にこの親個体から取らなかった遺伝子を、他の親個体から出現順に取り出し、子個体の s の右側に転写する。
- b) 突然変異……染色体の一点をランダムに選び、その右側と左側の遺伝子を入れ換える。ただし、最優良個体(エリート)に対しては、突然変異は施さない。
- c) 個体評価と自然選択……各個体は、評価データに対する GMHD モデルの出力と真値 y の最小自乗誤差で評価する。誤差が小さい個体ほど適応度が高いとし、順位選択する。

4. 検証実験

データ点数 500、説明変数 8 からなる実験データを 4 ケース用意し本手法の有効性を検討した。350 点をモデル決定に用い、75 点を評価に用いる。真の説明変数の他に余分な変数を 5 個加え、説明変数候補を 13 とし、変数の選択及びモデル構築を行ってみる。なお、変数選択のため前節の染色体を拡張しているが、その詳細は別途研究会にて報告する⁴⁾。

図 2 は個体数 48、世代数 300 での最優良個体の出力値である。ピーク位置などを正しく予測していることがわかる。

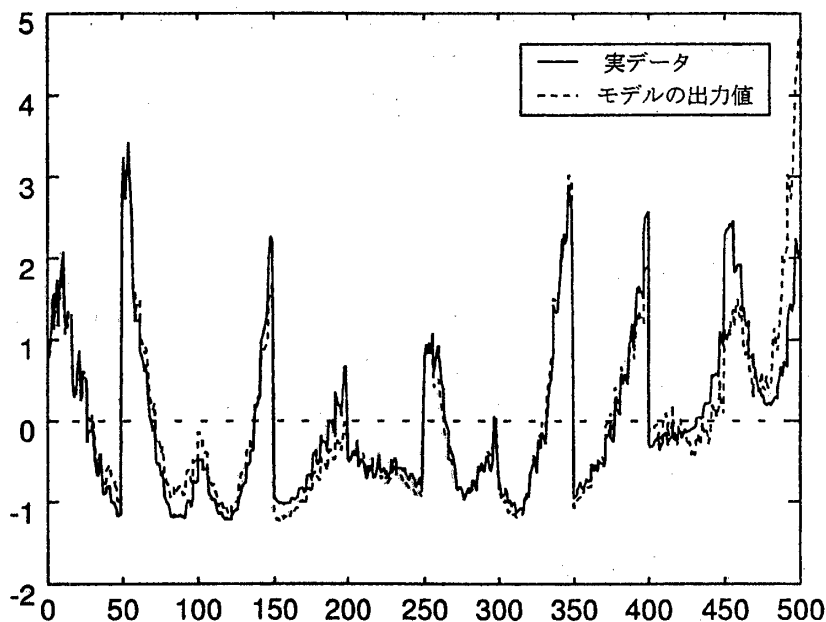


図 2 GMDH による推定結果

5. おわりに

非線形のモデル構築手法である GMDH に GA を組み込み、ヒューリスティクスに頼らないモデル構築を可能にした。我々の試みは、あまり高次でないモデルの方が好ましいとの経験判断を反映したものであり、遺伝的プログラミング手法を用いた従来手法とは異なる³⁾。

上記人工データのほか、41 項目の経済時系列実データの解析も行った。後者に関しては、主要な説明変数の抽出に有用との専門家の評価を得ている。

<参考文献>

- 1) Ivakhnenko, A.G.: The Group Method of Data Handling, A Rival of the Method of Stochastic Approximation: Soviet Automatic Control, Vol.1, pp.43-55, (1968)
- 2) 池田: GMDH(変数組合せ計算法)の基礎と応用: システムと制御, Vol.23, No.12, pp.710-717, (1979)
- 3) 伊庭: 遺伝的アルゴリズムの基礎 - GA の謎を解く - : オーム社, 1994
- 4) 吉原, 佐藤: GA を用いた非線形モデル構築の最適化 - GA と GMDH の融合 - : AI 研究会, AI-105, pp.1-6, (1996)