

生活環境評価関数への改良形GMDHの適用

徳島大学 正員 青山 吉隆
○ 徳島大学 学生員 壁谷 康峰

1. 緒言

都市の生活環境は非常に多くの要因から構成されており、生活環境を評価し、環境計画を策定するためには、これらの多くの要因がどのような構造によって総合的な環境を形成しているかを明らかにすることが重要である。この生活環境の各個別要因と全体評価との構造を評価関数と呼び、関数モデルの推定には住民意識調査データを用いるのが最も一般的であり、また推定手法としては数量化理論、判別関数、重回帰分析などの多変量解析法やあるいはデルファイ法などの直観的方法も用いられている。

評価関数が線形構造の場合には、いわゆる“重み”の推定手法の優劣を比較することになる。しかし住民意識の構造はあくまで本質的にはブラック・ボックスの内部にあり、線形式として特定化することに問題が残る。この構造は単純な要因の一次結合ではなく、要因が相互に結合した複雑なシステムとみなすのが一般的であると考えられる。ところが複雑な構造式として特定化してそのパラメータを従来の統計的手法によって推定しようとすると、推定時の自由度が減少し、サンプル数の不足あるいはモデルの不安定性等の問題が発生し、十分な精度を持ち、かつrobustなモデル同定が困難となる。

GMDHは、こうした複雑な構造をもつモデルを少数のサンプルから推定する有効な手法であり、本研究はすでに発表した基本形GMDH⁽¹⁾の改良形を開発し、徳島市において実証的に研究したものである。

住民意識調査は昭和53、55年に行なわれたアンケートで、サンプル数は3608である。生活環境要因としては表-1に示す26要因、ゾーン数は23地区である。

2. GMDHモデル

モデルの説明変数と被説明変数は次のようである。

$\{x_{ik}\}$: ゾーン i の要因 k の満足率 or 不満率
 $\{y_i\}$: ゾーン i の総合的な満足率 or 不満率

明らかに、 $0 \leq y_i \leq 1$ であるから、モデル式として、式(1)を仮定し、 ϕ をGMDHによって推定する。

GMDH(Group Method of Data Handling)は、発見的自己組織化の原理にもとづくモデリングを行なう方法論である。その基本的なアルゴリズムを図-1に示す。ここでは、GMDH全体については、紙面の都合上述べず、その大きな特徴である2つの概念、「部分表現と中間変数」・「データ分割と自己選択」についてのみふれることにする。

部分表現とは、式(3)・式(4)に示すように、モデル中の一部の变数と出力変数 y との関係を表わしたものであり、この式を部分表現式という。これより求められた Z を中間変数と言い、次の層ではこの中間変数

表-1 住まいのまわりの環境項目

1	日あたり、風とおし	14	日常の買物
2	工場等からのにおい	15	通勤、通学の交通機関
3	工場、車等の騒音、振動	16	地震等の災害時の避難
4	川、池の水	17	散歩やいこいの場所
5	ほこり、ばい煙、排気ガス	18	大雨のときの水はけ
6	病院等の医療施設	19	か、はえ等
7	集会施設	20	のら犬、のらねこ等
8	身近のスポーツ施設	21	風紀について
9	子供の遊び場	22	防犯について
10	通園、通学路	23	夜道の照明
11	道路の舗装	24	空地の雑草
12	街路樹等のまわりの緑	25	ごみが捨ててある
13	中心商店街への交通機関	26	現在のすまいについて

$$y = \frac{1}{1 + \exp(-\phi)} \quad (1)$$

$$\phi = f(x_1, x_2, \dots, x_n) \quad (2)$$

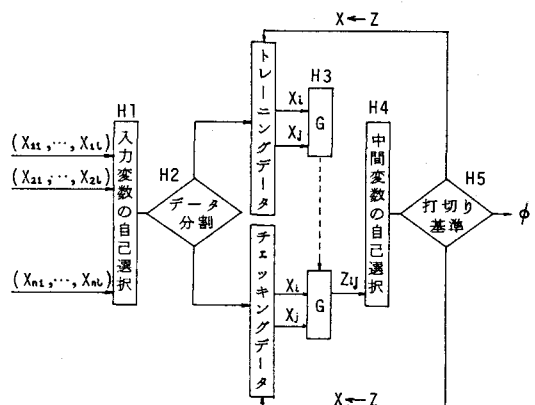


図-1 GMDHのアルゴリズム

を X として、部分表現式に代入し新たな中間変数を推定する。
この過程で、入力変数は層が増すにしたがい、相互に関係し合う
のでモデルは複雑な構造となる。

層が増すにしたがい出力変数と同程度に近似する中間変数が多
く出現してくる。また変数の組合せは、変数が n 個あれば、 nC_2
個ある。そこで、これらの中より、正則化を行なって、多くの組
合せの中より、モデルに必要な変数を選択する必要がある。ここ
で正則化とは、変数の独立性と微少変動に対する連続性を確保す
るためのアプローチである。GMDHでは、部分表現式のパラメ
ータ推定のためのデータ(Training Data)と、中間変数を選
択するためのデータ(Checking Data)とにデータを分割し、
これらを用いて、正則化を行ないモデルを構築してゆく。Checking
Dataを用いることにより「頑健な(robust)モデル」、すなわ
ちデータの微少変動によりパラメータが変化しない安定したモデ
ルが得られる。また層の打ち切り基準としては、heuristicsな閾値
的自己選択を行なう。

以上のような特徴により、モデルの構造に関する先験的情報を
必要とせず、少数データでも安定なモデルを推定することができ
る。しかし、GMDHでは表-2に示すheuristicsを含んでいる
ので、これらの最適基準を定めるためには、GMDHの繰り返し
計算を行わなくてはならない。さらにRSS(残差平方和)
を用いる基本形GMDHでは、①~③に示す問題が存在する¹⁾。

- ① H3で、部分表現式の型を決定すると、それにともない完全
表現式のモデルの構造が定まり、変数関係が一定のパターン
になる。また式(3)では多変数モデルは複雑すぎる。
- ② H2により、モデルのパラメータが変化する。また、Training
Dataの適合度がモデルに考慮されない。
- ③ H5を、かなり恣意的にきめなくてはならない。

3. 改良形GMDH

本研究では上記の問題点の改良をはかり、GMDHの最適化のため次の改良を行なった。

まず、部分表現式の型の選定にステップワイズ法を適用する。また、ステップワイズ法における選定基準・中
間変数の選択基準・層の打ち切り基準として、PSSまたはAICを用いて、部分表現式・完全表現式の最適化を
はかる。さらに、入力変数の自己選択についても、従来の出力変数と入力変数の相関係数による選択の他に、生
活環境評価関数であることを考慮する。

ここで用いる、PSSとは、 n 組のデータがあるとき、 j 番目のデータをのぞく($n-1$)組のデータよりモ
デルを推定し、このモデルに j 番目のデータを式(6)へ代入してえた推定値 $\hat{y}_j^*$ と y_j とのカイ離により、モデ
ルの評価を行なう基準である。また、AICは一般に式(7)で定義されるモデルの適合度とパラメータの増加へ
のペナルティを持つ評価基準である。GMDHの部分表現式の推定には、重回帰分析を適用する。この場合、A
ICは式(8)となる。ここで第3項は一定であるので、モデル式の相対的比較のためには式(9)を評価基準とし
て用いればよい。

2次の部分表現式

$$Z_k = a_0 + a_1 X_i + a_2 X_j + a_3 X_i X_j + a_4 X_i^2 + a_5 X_j^2 \quad (3)$$

双曲線の部分表現式

$$Z_k = a_0 + a_1 X_i + a_2 X_j + a_3 X_i X_j \quad (4)$$

表-2 heuristics について

	内 容
H1	入力変数の自己選択
H2	データ分割方法
H3	部分表現式の型の選択
H4	中間変数の選択基準
H5	層の積み重ね回数

$$PSS = \sum_{i=1}^n (y_i - \hat{y}_i^*)^2 \quad (5)$$

ただし

$$\hat{y}_i^* = a_0 i + \sum_{j=1}^m a_{ij} x_{ji} \quad (j = 1, \dots, m) \quad (6)$$

$$AIC = -2 \ln(\text{最大尤度}) + 2(\text{パラメータの数}) \quad (7)$$

正規分布の場合

$$AIC = n \ln \sigma^2 + 2K + n(\ln 2\pi + 1) \quad (8)$$

K : パラメータの数
n : サンプル数

$$AIC = n \ln \sigma^2 + 2K \quad (9)$$

(1). 入力変数の選択

入力変数として、1要因に対して満足率・不満率の2つの変数が存在するため、表-3、S1ケースのようにこの2変数をともに変数として計52変数を入力すると、評価関数を利用する際に、不適当な変数がモデルに組み込まれる場合がある。そこで、S2、S3の2ケースでは、1つの要因内の満足率・不満率のいずれかの1変数を入力とする。また、S4のケースでは、52変数を用いて因子分析を行ない、その因子得点を用いてクラスター分析を行なうことにより、グループングして、その中より出力変数との相関の高い変数を操作性を考慮して、9変数選択し、それを入力変数とする。

(2). 部分表現式の最適化

部分表現式を式(3)とすると、変数の選択の組み合わせの数は、 $\sum_{i=1}^5 C_i = 31$ 通りある。これに評価基準としてPSSあるいはAICを用いたステップワイズ法を適用し、その評価基準を最小とする部分表現式を、中間変数に採用する。このサブシステムにおいては、図-2に示すような特性を持つ部分表現式が推定される。つまり、G1タイプでは、次数の増加と変数の組み合わせが同時にでき、G2タイプでは変数の組み合わせのみが、G3タイプでは前層と同一構造の、中間変数ができる。このような部分表現式の最適化により、基本形GMDHに比較して変数の変動によく合う完全表現式となる。生活環境評価関数のような多変数で比較的低次の多項式と考えられるモデルに、この改良は有効と考えられる。

(3). 中間変数の最適化と打ち切り基準の最適化

部分表現式の最適化によってえた中間変数に対するPSSあるいはAICの値を比較し、小さい値を持つ中間変数を1個選択し、次層の変数とする。これにより、中間変数の最適化をはかる。AIC、PSSの両基準は、ともに最小値を持つので、最小値を示す層で打ち切りを行なえばよい。また、部分表現式は層の積み重ねにともないしだいにG3タイプになると考えられる。そこで全ての部分表現式がG3タイプとなり、モデル構造が変化しなくなった層で打ち切りを行なうことにする。この2つの基準を用いると、層の打ち切り基準における恣意性がなくなる。

またPSSとAICは、共にその性質より、データ分割を行なわなくてもよいので、問題②が解決される。

4. 適用例

徳島市に、改良形GMDHを適用した結果について述べる。

中間変数の選択数と各層における最良の評価基準の値の関係を図-3に示す。この図より、入力変数が一定であれば、中間変数の選択数が多いほど、完全表現式の適合度つまり最終層における適合度は良くなるように見える。しかし、それにともなって1つの層に対する計算量も増大することを考えると、中間変数の数をむやみに増加することは、効率的とは言えない。また、層の積み重ね過程において、いくつかの部分表現式の評価基準の値が類似してくる場合がある。このような部分表現は一つだけを考慮し、その他は中間変数として採用しなかった。この中間変数の評価基準と層の関係は、AIC・PSSともに同じような変動を示している。

次に各層で部分表現式の最適化を行なった結果について述べる。

表 -3 入力変数の選定

ケース	内 容
S1	52変数の中でy との相関が高い変数 (52)
S2	各要因での満足率・不満率のうち y との相関が高い変数 (26)
S3	各要因で、満足度推定には満足率 (26) # 不満足度推定には不満率
S4	因子分析・クラスター分析より (9)

() 内は最大入力変数の数

→[G1]→ 次数の増加と変数の組合せ

$$Z_k = a_0 + a_1 X_i + a_2 X_j + a_3 X_i X_j + a_4 X_i^2 + a_5 X_j^2$$

$$Z_k = a_0 + a_1 X_i + a_2 X_j + a_3 X_i X_j$$

→[G2]→ 変数の組合せ

$$Z_k = a_0 + a_1 X_i + a_2 X_j$$

→[G3]→ 前層と同一

$$Z_k = a_0 + a_1 X_i$$

図-2 部分表現式のタイプ

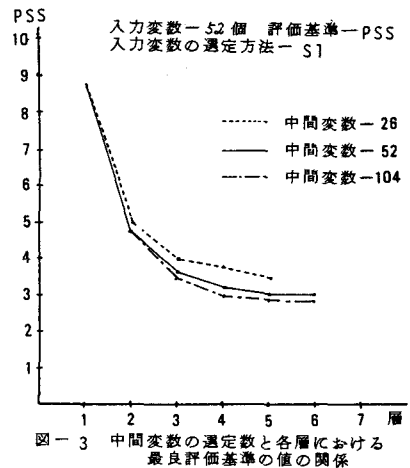


図-4に各層に出現する部分表現式をタイプ別に集計した頻度を示す。これより層の積み重ねによって部分表現式は次のような動向を示すことがわかる。層が積み重ねられるにしたがい、部分表現式のタイプはG1→G2→G3と変化する。AIC・PSSともに、層の積み重ねにより部分表現式はG3になることが実証された。

部分表現式、式(3)、式(4)を比較する。入力変数の選定方法を表-3のS1とし、入力変数・中間変数を各々52個選定し、中間変数・部分表現式をAICにより自己選定する場合、式(3)で表わされるような二次式を部分表現式に用いると、完全表現式中に存在する入力変数は9変数あり、式(4)で表わされる双線形式を用いると11変数存在している。

また式(4)を用いた場合、入力変数の積で表わされる項の中で、最も変数が多い項でも4変数によって構成されたものであるのに対し、式(3)では12変数によって構成された項が存在している。このように式(4)の方が依次で多変数より構成されていると考えられる生活環境評価関数モデルに適していると考えられる。

入力変数の選定について比較する。まず表3のS2・S3を比較すると、モデルの構造・パラメータ値ともに、あまり差異がなく、類似したモデルとなった。これは、S2において、満足と不満の変数間で入れ換えが行なわれた要因は、完全表現式中には1変数しか入っていないためS3とほとんど差異のないモデルとなった。

S4は、RSS・評価基準値とともに低い、入力変数9個の内8個が完全表現式に採用されている。これは、入力情報が完全表現式に生かされており、入力変数の因子分析・クラスター分析による選定効果が現われている。

図-5に、改良形GMDHによって推定されたモデルの構造体系を示しておく。

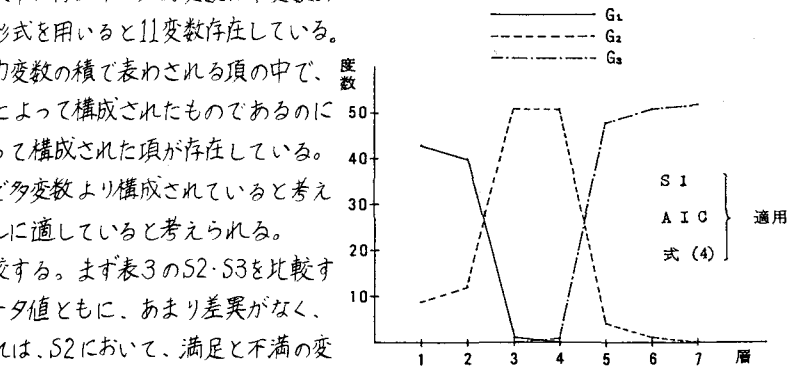


図-4 部分表現式のタイプ別出現頻度

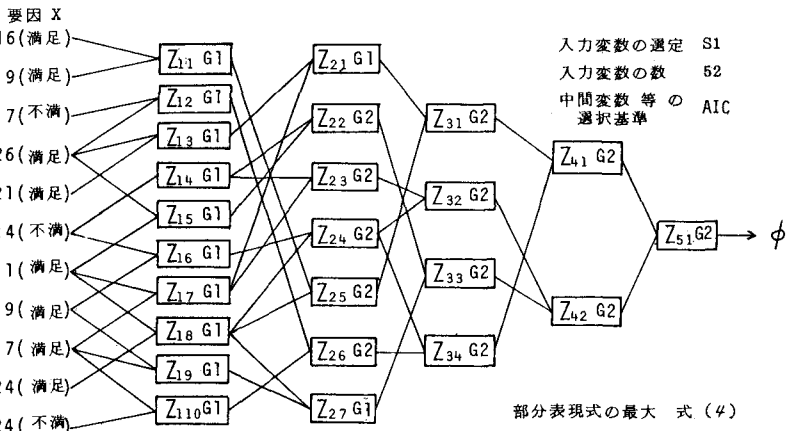


図-5 改良形GMDHによる総合満足度モデルの構造系

5. まとめ

生活環境評価関数は、要因が相互に関係・影響し合う複雑な構造を持つシステムと考えられる。そして、ここに述べたように、GMDHの各サブシステムに関して最適化をはかった改良形GMDHはこのシステムの同定手法として有効である。また、本手法は、

- ① 基本形GMDHで問題となった、各種のheuristicsの最適化がはかれる。
 - ② 少数データにもかかわらず、多変数が相互に関係し合うモデルの推定を、適度に複雑さをそなえたモデル構造で推定することができる。
- などの特徴を持っている。

参考文献

- 1) 青山・壁谷「生活環境の評価関数に対するGMDHの適用」第37回土木学会年次学術講演会概要集
- 2) 田村坦・近藤正「最近のGMDHの方法論と応用」オペレーションズ・リサーチ 1978 2月号