

GMDHによる生活環境評価関数の推定

徳島大学

正 員 青山吉隆

○ 徳島大学

学 員 瀬尾陽彦

§1 はじめに 従来、生活環境に対する住民の意識構造を分析するための方法としては数量化理論2類分析、AID分析などが用いられてきたが、これらは単に分析という段階にとどまることを待たないもので、生活環境評価を予測するモデルを作成するときには重回帰分析が適している場合が多い。しかしながら、重回帰分析では様々な問題点も多く、特に入手可能なデータが少ない場合に作成されたモデルは不安定なものとなってしまうことが多いため、予測に利用できるモデルを作成することは非常に困難である。そこで、本文ではGMDHを用いて、生活環境評価関数の推定について考察することにする。

§2 生活環境評価モデルとGMDH GMDHの原理及び構造についてはここではふれず、GMDHが重回帰分析に比べて生活環境評価モデルに有利であると考えられる理由を以下に列挙すると、

- 1) 多層構造で有用な変数を自己選択してゆくので、あらかじめ変数を選択しておく必要がない。
 - 2) 各層を通過してゆくに従って予測式の複雑さが増加するので、あらかじめ構造を決定する必要がない。
 - 3) 部分表現式を仲介して複雑な予測式を構成するので、非常に少ないデータでもモデリングできる。
 - 4) trainingデータとcheckingデータに分割してモデリングするので、構造的に安定なモデルの自己選択が可能。
- などがあるが、GMDHはすべてを自動的にモデリングを行えるわけではなく各種のheuristicsを含んでいるので、これらの最適化の必要がある。

§3 GMDHの適用 データには昭和53年と55年に行われた「徳島市の行政に関する市民の意識調査」アンケートの中の表1に示す26項目に対する満足及び不満の評価を行政地区(23地区)ごとに比率として集計したものを2回分46データとして使い、各環境項目に対する満足と不満の評価(x_i)から、住まいのまわり全体についての評価(満足度または不満度; y)を推定する。ここで、 $0 \leq y \leq 1$ の比率であることが望ましいからロジットモデルを適用して

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

$$\phi = f(x_1, x_2, \dots, x_{26}) \quad (2)$$

表1 住まいのまわりの環境項目

1 日あたりや風とおし	14 日常の買物
2 工場などからのにおい	15 通勤や通学の交通機関
3 工場、車などの騒音や振動	16 地震などの不慮の災害時の避難
4 川や池の水	17 散歩やいこいの場所
5 ほこりやばい煙、排気ガス	18 大雨が降ったときの水はけ
6 病院などの医療施設	19 かやはえなど
7 集会施設	20 のら犬やのら猫など
8 身近にできるスポーツ施設	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 街路樹などのまわりの緑	25 ごみを勝手に捨ててある
13 中心商店街への交通機関	26 現在の住まいについて

として ϕ をGMDHによって推定することにする。このため、heuristicsの最適化に改良を加えたGMDH²⁾を用いず、各種のheuristicsの最適化を行なう。

§4 heuristicsの最適化 GMDHのアルゴリズムには、1.データの分割方法の決定 2.部分表現式の決定 3.入力変数と中間変数の選択数の決定 4.中間変数の選択基準(regularity criterion)の決定 といったheuristicsを含んでいるので、1.については53年のデータをtrainingにする方法と分散の大きいものから半数をtrainingにする方法の2とおり、2.については ϕ がKolmogorov-Gaborの多項式で表わせると仮定した場合の部分表現式

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

と ϕ が1次式で表わされると仮定した場合の部分表現式

$$Z_k = a_0 + a_1 x_i + a_2 x_j \quad (4)$$

についてはい、4.では checking データによる (i) 相関係数、(ii) 2乗誤差 の他に、ロジットモデルを適用したため次のような *regularity criterion* を考えて最適化を行なった。

- (iii) training データによる相関係数と checking データによる相関係数の平均値
- (iv) training データによる (1) 式での変換後の相関係数と checking データによるものの平均値
- (v) checking データによる (1) 式での変換後の相関係数

ここで (iii), (iv) のように両データによる相関係数の平均と考えたりは、training データのあてはまりよりも checking データに対するあてはまりの方が非常によいという中間変数が多く存在したため、両データのあてはまりがよいものを選択するためなどの理由によるもので、(iv), (v) のように変換後の相関係数を用いたのは、変換前後でその大小関係が変化するためである。これら 1.~4.の *heuristics* の最適化の他、繰返し回数についても第3層で打切ることとした。これは、図1に示すように第3層以降 *regularity criterion* の改善があまり見られないうこと、また部分表現式に(3)式を用いた場合 ϕ はすてに8次式になつてきていること、(4)式の場合はパラメータの正負、その大きさなどからも第3層を超えてのモデリングにはあまり意味のないと判断したためである。

また、3.の入力及び中間変数の選択数の最適化によって、図1にも見られるように中間変数をできるだけ多く選択することが望ましいと考えられた。

以上の最適化により、2(3)式によるモデルの全データによる相関係数は満足度の場合の0.9033、不満足度の0.9150、また(4)式のときは0.8463と0.8483とかなり良い結果が得られた。これらはいずれも *regularity criterion* として (iv) を用いたものである。

§5 入力変数の許容変動域 前節で得られたモデルが入力変数の変動に対してどの程度対応できるかを見るために図2のように他のすべての入力変数は平均値に固定しておき、ひとつの変数のみを $0 \leq x_i \leq 1$ の範囲で変化させて推定値の変化を図化した。ここにあげた2つの曲線は、満足度のモデルで ϕ が8次式のものである。これを見ると a は非常に理想的であると考えられるものが得られているが、 b ではそのデータの存在する範囲にしか対応していないと考えられる。一般的には ϕ が8次式のものでは b のような曲線が a のように 1 次式のものとはほとんど説明変数に対して a のような曲線が得られている。したがって、2(3)式と部分表現式とした場合のモデリングでは予測式の自由度が大きいことから、実際の構造に対して複雑過ぎると思われる。このモデルでは大きな変動に対応できないがモデリングに使用したデータの存在範囲での推定にはこのままでもかなり有効であると考えられる。

§6 結論 GMDH によつて推定された生活環境評価関数は、推定が容易であるにもかかわらず、かなり安定したモデルが得られていることが明らかになり、今後予測式の複雑さを適当なものにするための改良を加える、すなわち部分表現式の決定にステップ・ワイズを用いるといったこととさらに利用価値の高いものとするこ

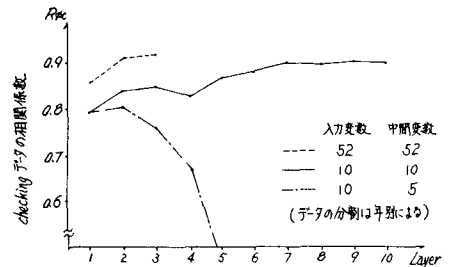


図1 各層での *regularity criterion* の最良値 (3式による満足度モデル)

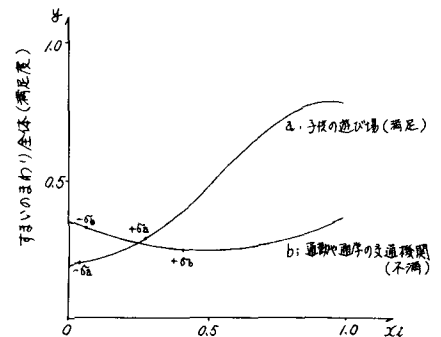


図2 入力変数の変動に対する推定値の変化 (ϕ が8次式の満足度モデル)

- 参考文献 1) 池田三郎・榎本義一著「GMDHと複雑な系の同定・予測」計測と制御 第14巻 第2号 (1975)
2) 田村坦之・近藤正 著「最近のGMDHの方法論と応用」オペレーションズ リサーチ2号 (1978)