

徳島大学 正員 青山 吉隆
徳島大学 学生員 ○壁谷 康峰

§1. はじめに

生活環境に対する住民の意識構造の分析手法としては、数量化理論Ⅱ類、因子分析、重回帰分析などが用いられてきた。しかし、生活環境を構成する、多くの変数は、一般的に、相互に影響し合っていると考えられ、サンプル数が少ない場合には、不安定なものになるなどの問題があり、予測の可能なモデルの作成は困難である。本研究は、これらの問題を考慮して、生活環境評価関数を推定する手法として、GMDH (Group Method of Data Handling) を用いることを提案する。

§2. GMDHの適用

データとして、徳島市における、S53、S55の市民意識調査アンケートより、表-1に示す26項目に対する満足および不満足と評価した度数を、23の地区ごとに比率として集計したもの(x_i)を用いる。これらと、す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、 $0 \leq y \leq 1$ の制約条件があるため、ロジットモデル、式(1)を適用し、 ϕ を GMDH によって推定することにする。

§3. GMDHモデル

GMDHは、図-1のアルゴリズムに示されるように、部分表現式の重ね合わせ、変数の自己選択により、「定式化の困難さ」をなくし、変数として中間変数を用いることにより、少数データによる推定が可能である。また原データを、パラメータ推定に用いるtrainingデータと、自己選択に用いるcheckingデータに分割することにより、安定した構造のモデルが推定される。これらより、GMDHは、複雑なモデル推定に有用であると考えられる。

§4. heuristicsの最適化

GMDHのアルゴリズム中には、H1:部分表現式の決定、H2:データの分割方法、H3:入力変数・中間変数の選択数、H4:中間変数の選択基準、といったheuristicsを含んでいる。本研究では以下のようにこれらの最適化をおこなった。

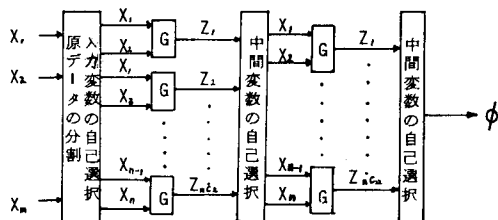
H1については、 ϕ をKolmogorov-Gaborの多項式で表わせると仮定したときの部分表現式一式(3)と、 ϕ を1次式と仮定した場合の部分表現式一式(4)を用いる。

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

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

$$\phi = f(x_1, x_2, \dots, x_{52}) \quad (2)$$



G: 部分表現式の発生器 Z: 中間変数
図-1 GMDH のアルゴリズム

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

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

H2については、データを年度により分割し、S55データをcheckingデータにする方法と、分散の小さい半数をcheckingデータにする方法の2通りを行った。

H4については、表-2に示すように、GMDHで一般に用いられている、(i)、(ii)の他に、ロジットモデルを適用していることを考え、(iii)~(V)も用いた。ここで、(iii)、(V)で両データによる相関係数の平均を用いたのは、H1において、分散による分割を行った場合に、checkingデータに対する適合が、trainingデータに対するものより、非常に高い値を示すため、データ全体に対する適合のよいものを中間変数に選択した方がよいと考えたためである。また(iv)、(V)では、式(1)に変換した後の相関係数を用いたのは、変換前後で部分表現式の相関係数の大小関係が変化するためである。

H3については、入力変数・中間変数の選択数を変化させて、checkingデータの相関係数との関係を表わした図-2より、選択数を多くした方が良い結果がえられることがわかる。また、この図より、第3層以降では、regularity criterionの改善が少なく、また式(3)を部分表現式とした場合には、 ϕ は8次式となること、式(4)を用いた場合、パラメータの符号条件を満足しない項が多くなることより、H5として、部分表現式の重ね合わせは3回とし、第3層において最良のものを完全表現式とした。

以上のheuristicsを各々変えて得た結果では、式(3)を用いて、H1として年度分割をおこない、H4として(V)を採用したとき、全データによるモデルの相関係数は満足度で0.9433、不満度で0.9150と最良であった。この満足度推定のモデル構造を図-3、パラメータを表-3に示す。

§5. 入力変数の許容変動域

推定されたモデルが、入力変数の変動にどの程度対応できるかを分析するため、1変数を $0 \leq x \leq 1$ の範囲で変化させ、他の変数は平均値に固定し、推定値の変化を調べた。式(4)を用いたモデルは、常識に合った変動を示す変数が多いが、式(3)を用いたモデルでは、データの存在範囲を越えると、不自然な変動をする変数がみられる。これは、重ね合わせの過程で満足を表わす変数と、不満の変数との積となる項が出現したためと考えられる。このようなモデルは、大きな変動には対応できないが、モデリングに使用した存在範囲での推定には有効であると考えられる。

§6. まとめ

本研究で用いたGMDHは、基本的なものであり、種々のheuristicsを含んでいる。それでも複雑なモデルをかなり安定した構造として推定できることが明らかとなった。部分表現式の合理的な決定方法や、regularity、criterionの改善などによって、さらに利用価値の高いシステムになると考えられる。

〔参考文献〕 池田三郎・榎木義一「GMDHと複雑な系の同定・予測」計測と制御 第14巻-第2号

表-2 中間変数の選択基準

i	$R_{\phi c}$
ii	checking data の2乗誤差
iii	$(R_{\phi t} + R_{\phi c}) / 2$
iv	$R_{y c}$
v	$(R_{y t} + R_{y c}) / 2$

R_{ϕ} : ϕ との相関係数
 R_y : y との相関係数
 c : checking data
 t : training data

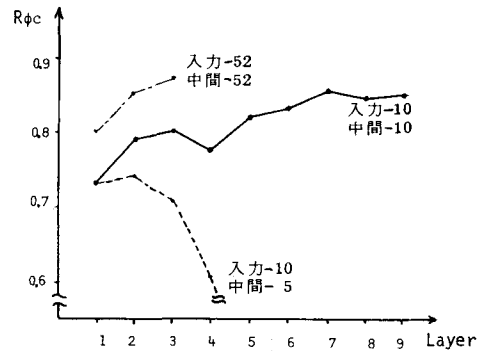


図-2 各層におけるchecking dataによる最良の相関係数

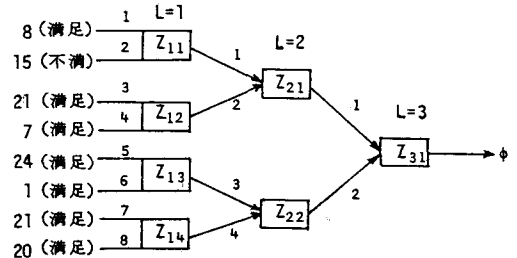


図-3 総合評価モデルの内部構造 (満足度・部分表現式-式3)

表-3 総合評価モデルのパラメータ

L	i	j	a_0	a_1	a_2	a_3	a_4	a_5
1	1	2	-0.685	0.915	-5.150	18.34	-2.746	2.930
	3	4	-3.639	10.28	1.639	26.95	-21.06	-15.35
	5	6	-11.55	8.152	21.25	-13.85	6.729	-9.439
	7	8	-4.192	5.775	17.29	-14.96	-0.129	-24.32
2	1	2	-0.389	0.494	-0.491	-0.220	0.063	-0.262
	3	4	0.019	0.946	-0.038	-0.078	0.114	-0.104
3	1	2	0.460	1.401	0.357	-0.145	0.435	-0.080