

I-16 GMDHを用いた知識ベース構築法とその応用

Construction of Knowledge Database by utilizing GMDH and its Application

古田 秀博*

Hidehiro Furuta

【抄録】 GMDH(Group Method of Data Handling)を用いた機械学習のための演算法とその適用例について詳細に記述している。GMDHは、これまでデータ構造の近似や予測に用いられてきているが、一方で論理演算、データパターンの学習、および知識の蓄積能力等に優れた機能を発揮することが最近になって注目されている。本論文では、初めにGMDHの特徴について説明し、実際に論理演算の代表的な例である排他論理和(XOR)問題に適用し、具体的な計算手順とその性能を示す。最後に、この手法を実際の土木技術に関する学習問題に適用するとともに、その結果であるデータ構造について分析する。

【Abstract】 Automatic Learning System utilizing GMDH is presented in this paper. GMDH has been used for approximation of data structure and prediction so far. On the other hand, the ability which makes the logical arithmetic calculation, learns the data pattern and reserves knowledge is recently taken noticed. This paper explains about the details, features of GMDH and shows the actual procedure of calculation by solving the XOR as representative example. Finally, efficiency of the method is shown by applying this method to practical civil engineering problems. And the data structure of these examples is analysed

【キーワード】 GMDH, 機械学習, XOR問題, のり面崩壊, 発見的自己組織化

【Keywords】 GMDH, Machine Learning, XOR problem, Collapse of Cutt-off Slope, Heuristic Self-Organization

1. はじめに

1980年代は、エキスパートシステム(以下ESと呼ぶ)が最盛期を迎えた時代であった。しかしながら、その後期待された程の成果を上げることができず、現状、一時の勢いを全く失ってしまっている。また、新たに構築された実用的なシステムの例もほとんど聞かれない状況にある。この原因は、①専門家の持っている知識を客観的な事実としてうまく抽出できない。②専門技術の絶え間無い変化や発展に対応できない。③以前に作成さ

れた知識と新たに追加される知識の間の矛盾の解消作業が非常に煩雑であり、また、これを支援する確立された手法が見当たらない等の問題点にある。これらの問題の解決の糸口として佐野は、GMDH(Group Method of Data Handling)が有効であると文献1)で述べている。具体的には、GMDHを単なる構造同定手法としてではなく人工ニューラルネットの一種として捉え、これの論理演算機能と知識獲得機能について、誤差逆伝播学習法を用いた人工ニューラルネットなどに対する優位性を強調している。

* (株) 横河技術情報CAD・CAE事業部 システム部CAE課 (〒273 千葉県船橋市山野町27, TEL 0474(35)6120)

GMDHは元々、多入力1出力系の構造不明な多変数非線形システムを、数少ないデータを使って同定するための演算法である。そして、その方法も発見的自己組織化(Heuristic Self-Organization)の原理に基づいており、従来の方法では不可能であった数多くのシステムの同定が可能である²⁾⁻⁴⁾。これまで、GMDHの基礎的な研究や具体的な応用研究は多数行われてきているが、現状のGMDHに関する論文や事例はすべて多変数非線形関数の同定問題に関するものであり、文献1)にあるようにこれをニューラルネットワークと捉えて、知識の獲得や蓄積演算に具体的に適用しようとする試みはまだ始まったばかりである。そして、現実の土木工学の問題にGMDHの知識獲得・蓄積・判定処理演算を具体的に用いた例は、現状、文献5)以外見当たらない。

以下、本論文では始めに、GMDHの特長的な機能と計算手順について詳細に説明する。そして、これを典型的なパターン認識問題であるXOR問題に適用し、その結果について考察する。次に、土木分野での具体例として西等⁶⁾による調査結果について学習演算を行い、GMDHの知識獲得およびそのデータ構造の分析に関する有用性を示す。

なお、本研究では、GMDH手法の知識ベース構築のための具体的計算手順を整理し明示するため、以下の2つの項目について詳細に検討した。①ここで扱わねばならない問題が整数問題である上に、作成される連立方程式がほとんどの場合非正方行列を持つことになるため、これに適合した特殊な計算手段の構築と詳細な手順の提示を行う。②計算結果である2次多項式に基づいた調査・判定結果のデータ構造の分析方法(評価関数の相関関係、判定に影響する有効ビットの抽出等)を示す。

2. GMDHによる非線形関数の同定・予測と学習

入出力データに基づく線形系の同定や予測問題に関して、回帰分析などのシステム同定手法は有用であるが、対象のシステムが複雑で大規模な構造である場合、推定すべき係数は莫大なものとなり、モデルを事前に決めることが困難である。このような場合、GMDHはその威力を発揮する。

いま、入出力変数($X_1, X_2, X_3, \dots, X_n$)と y が与えられているものとする。GMDHでは、入出力関数が次の

のような従属関係 f を満足するものとする。

$$y=f(X_1, X_2, X_3, \dots, X_n) \quad (1)$$

f の推定値 η を多項式と仮定すると、出力 η は以下のような関数式により表すことができる。

$$\begin{aligned} \eta = & \zeta(X_1, X_2, X_3, \dots, X_n) \\ = & C_0 + \sum_{k_1} C_{k_1} X_{k_1} + \sum_{k_1, k_2} C_{k_1 k_2} X_{k_1} X_{k_2} \\ & + \sum_{k_1, k_2, k_3} C_{k_1 k_2 k_3} X_{k_1} X_{k_2} X_{k_3} + \dots \end{aligned} \quad (2)$$

ただし、 C_{k_i} は係数であり、 $k_1, k_2, \dots, k_i$ はすべて1, 2, ..., n の値を持つ。

GMDHでは推定モデル η を得るために、まず、2変数2次式などの部分表現式により中間変数を求める。次に、回帰分析により、推定値と出力データとの差の2乗和が最小となるように部分表現式の係数を決定する。この操作を繰り返し、得られた部分表現式を階層的に組み合わせることにより最終的な推定モデル η を得る。すなわち、GMDHは部分表現式を階層的に適用するので、少ない入出力データで複雑な非線形系を同定し予測することが可能である。

以上は、数値予測のための演算処理であるが、この手法が論理演算の代表的な例であるXOR(排他論理和)問題やマルチプレクサー問題に対してすべて正しい判定結果を与えることが知られている¹⁾⁻⁵⁾。よくこの種の問題に適用されるニューラルネットワークの一種であるバックプロパゲーション方式は、XOR問題に対してすべて正しい判定結果を与えることができるが、マルチプレクサー問題については、大量のトレーニングをこなした後でも、すべて正しい判定結果を与えることができない。さらに、各ネットワークに割り付けられたウェイトからは、どのような算術演算が施されているかについての説明はできない。

しかし、GMDHはこの問題に対してもすべて正しい判定結果を与えることができるとともに、帰納的学習の結果、データに内在するデータ構造自体を学習し、その代数演算的な説明をも加えることができる¹⁾。さらに、佐野は文献1)のなかで、「バックプロパゲーションは、しきい値関数(シグモイド関数)を持ち込むことにより、論理演算であるXOR問題まで含めて正しい判定結果を示せるようになったが、入力パターンとウェイトを掛け合わせた一次式と非線形関数であるしきい値関数の混合

使用であるため、学習結果の代数演算的・論理演算的説明ができない。」と述べている。

3. GMDH を用いた知識獲得・判定計算法

GMDH を用いた知識獲得および判定処理のための具体的計算法(図-1参照)について以下に示す。

①表-1に示されるように、縦方向に判定処理対象を、横方向に判定項目を並べる。ここに並べる判定パラメータは、どのような表現でも許すことができる。たとえば、実数値であっても良いし、「VS: 非常に小さい,S: 小さい,,,VL: 非常に大きい」のような言語値表現でも良い。これに整数値(0,1,,,N)を割り当てる。なお、ここでは、GMDH の計算上の簡便性から、表-2にあるように数値をすべて2進数表現として扱う。

②右辺項に当たる判定値は10進数であっても2進数であっても良い。

③式(2)に判定表の各値を代入し連立方程式を作成する。ただし、この連立方程式は普通の正方行列ではなく、行と列の大きさが異なるものとなる。

④③の式を解く。ただし、この式は、ほとんどの場合正方行列とならないため、以下のような特殊な演算法を採用する必要がある。先ず、表-2の表を基に各行について1の値の数を数える。そして、すべての行の1の数の合計を求める。

⑤各行の中で最も1の数の小さい行を見出し、この行の最左列の1を0とする。そして、すべての行について、この0とされた列に対応する列の1を0とする。さらに、1から0にされた行についてすべて右辺項から1を引く。

⑥1の数が1個となった行が現れたら、多項式中のある項に掛けられる係数の一つが決定される。

⑦以下④から⑥の計算を繰り返す。

⑧すべての行の合計が0になると計算を終了する。

⑨以上の計算が終了すると多項式の係数Cがすべて決定される。

⑩⑨で求めた係数Cを用いて式(2)に示される多項式を作成する。

⑪⑩で作成された多項式に評価項目に関する2進数を入力する。これにより判定結果のチェックを行う。

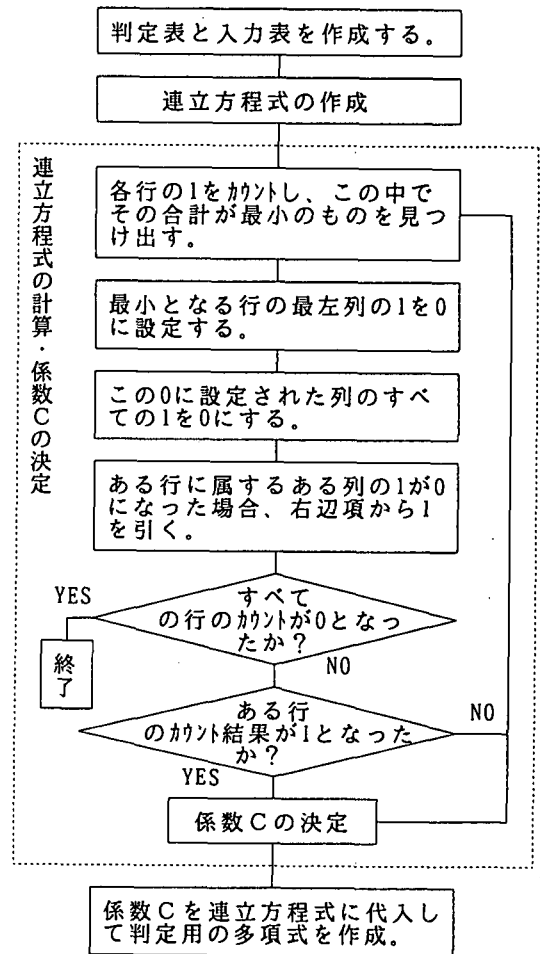


図-1 GMDH の計算処理の流れ

以上の計算手順を XOR 問題に適用してみよう。XOR 問題は排他論理和の問題であり、表-3に示すようなパターンを学習する問題である。ここで、式(3)に示すような2次の多項式に各条件の入出力値を代入すると、表-3にあるような連立方程式が得られる。

$$H(X_1, X_2) = a(X_1)^2 + b(X_2)^2 + cX_1 \cdot X_2 + dX_1 + eX_2 + f \quad (3)$$

表-3の中の2次の多項式 $H(X_1, X_2)$ による連立方程式を係数 abcdef について解くことにより、次に示すような多項式で XOR 論理関係を学習し表現できる。

$$H(X_1, X_2) = -2 \cdot X_1 \cdot X_2 + X_1 + X_2 \quad (4)$$

各入力値を式(4)に代入すれば、期待する教師信号として与えた出力値と完全に一致した結果が得られることは明かである。なお、この XOR 問題の場合、バックプロパゲーションはすべて正しい結果を与える。

表-1 判定表

	評価 1	評価 2		評価 M	判定値
対象 1	VS	S		L	SE
対象 2	M	L		VL	MM
⋮	⋮	⋮		⋮	⋮
対象 N	S	M		M	DD
(凡例) VS:Very Small,S:Small,M:Medium,L:Large VL:Very Large,SL:Slight,SE:Severe,MM:Moderate D:Destructive					

表-2 GMDH 法のための入力表

	評価 1	評価 2		評価 M	判定値
対象 1	000	001		011	010
対象 2	010	011		100	011
⋮	⋮	⋮		⋮	⋮
対象 N	001	010		010	100
(凡例) VS:000,S:001,M:010,L:011,VL:100,SL:001, SE:010,MM:011,DD:100					

表-3 XOR 問題

入力値		変換	出力値	2次の多様体多項式
X_1	X_2	$\rightarrow$	$H(X_1, X_2)$	$H(X_1, X_2)=a(X_1)^2+b(X_2)^2+cX_1X_2+dX_1+eX_2+f$
1	1	$\rightarrow$	0	$H(1,1)=a(1)^2+b(1)^2+c(1)(1)+d(1)+e(1)+f=0$
1	0	$\rightarrow$	1	$H(1,0)=a(1)^2+b(0)^2+c(1)(0)+d(1)+e(0)+f=1$
0	1	$\rightarrow$	1	$H(0,1)=a(0)^2+b(1)^2+c(0)(1)+d(1)+e(0)+f=1$
0	0	$\rightarrow$	0	$H(0,0)=a(0)^2+b(0)^2+c(0)(0)+d(0)+e(0)+f=0$

4. のり面崩壊要因に関する学習計算例

西等によるのり面崩壊要因に関する調査・判定結果⁶⁾に対して、GMDHによる学習演算を適用してみる。この例は、表-4(共用時)と表-5(建設時)にあるように、多数ののり面の状況を観察し、その崩壊の可能性を判定したものである。ここでは、技術者の主観によりその要因(AからJ)に対して1から5のランク付けを行い、これらの結果を総合的に判断して、のり面崩壊の危険性をこれも技術者の主観により1から5のランク付けを行っている。この表-4と表-5に対してGMDHでは、表-6(共用時)と表-7(建設時)にあるような30ビットからなる学習演算用の表を用意する。ここに、要因AからJのランク1から5を3ビット2進数で表現している。なお、表-4と表-5の要因と判定に関する詳しい内容については、文献6)に詳しいのでここでは特に触れない。

また、建設時(のり面数44)と共用時(のり面数23)の2つの状況について検討した理由は、データ量(のり面数)の増加のデータ構造への影響を見るためである。

表-6と表-7に共用時、建設時の学習演算結果を示す。各表中の式Gはこれらのデータの構造を表現する式である。また、表中の[*]は式Gの中で使用されている項であり、[/]は使用されていない項である。表-6と表-7を見ると以下のようなことが分かる。

共用時の学習演算の結果を表現する式Gには、6, 9, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 29の各ビットが含まれていない。建設時は、5, 14, 15, 17, 18, 20, 21, 22,

24, 25, 27, 29の各ビットが含まれていない。このことから、これらのビットがのり面の崩壊判定に何等影響を及ぼしていないことが分かる。そうすると、各要因AからJに付けられている1から5のランクは、上記のビットが判定に関与しないことから、1から5のすべてのランクを考慮する必要が無くなる。たとえば、[*/*]の場合、1, 4, 5の3つのランクのみについて検討すれば良いことになる。このような考え方で表-6と表-7を整理すると表-8のようになる。念のため、表-8のデータを基に不要なデータをすべて0に設定し、同様のGMDHによる学習演算を行ったところ、データ構造を表す式Gはすべて同一の表現となり、当然のことであるが、すべて西等の判定結果と同一の結果となった。さらに、表-8から共用時と建設時の有効な評価ランクを比較すると、評価要因B, C, G, H, Iで多少の増減があるものの、両ケースとも同一の評価基準を用いていることから、データ量の増加による大きな差は認められない。

しかし、データ量(のり面数)が増加すると、互いに相関関係を持つ評価ランクの数が増加してくる。これは、データ量が増加するにつれて、独立性の高い評価を行うことが困難になることを示している。従って、数理化理論のような線形結合式で表現するような手法では、大量のデータを正確に表現することが困難であり、77%回帰モデルのような特殊な手法の併用が必要となる。

以上から、GMDHの学習結果について以下のことが言える。

(1)GMDHは、のり面崩壊に関する判定表のデータ構造

を明確にし、判定結果に影響を及ぼす調査項目として必要不可欠なものが何であることを示すことができる。例えば、供用時の要因 G,H に関する調査は、判定結果に何等影響を及ぼさないことが分かる。

(2) 各要因 A から J におけるランク付けについても、すべて 1 から 5 のランク付けを行う必要はない。例えば、供用時の要因 E,F,I や建設時の要因 B,E,F,G はすべてランク 4 である。また、供用時の要因 B,C と建設時の要因 H は 2,4、供用時の要因 J と建設時の要因 I,J は 1,4,5 のランク付けのみでよいことになる。

次に、各要因 (A ~ J) の相互関係について分析してみる。供用時と建設時に関するデータ構造を表現する多項式 G は、変数 ($X_1 \sim X_{30}$) を互いに掛け合わせた 2 次形式で表現されている。従って、これを整理することによって、各変数間の関係を知ることができる。例えば、 $X_i \cdot X_i$ ($i=1,2,\dots,n$) の場合、自己相関関係に、また、 $X_i \cdot X_j$ ($i=1,2,\dots,n; j=1,2,\dots,n; \text{ただし } i \neq j$) の場合、相互相関関係にあると言える。ただし、これは式 G を構成する各変数 ($X_1 \sim X_{30}$)、すなわちビット列にすぎないので、これを各要因に関連付けて考えなければならない。詳述すると以下ようになる。

ある項 $aX_i \cdot X_j$ (ただし、 a は任意のパラメータ) の変数 X_i (すなわち i 番目のビット) が要因 S に属し、変数 X_j (すなわち j 番目のビット) が要因 T に属しているとする、要因 S と T の間には何らかの相関性があると判断できる。この考え方で整理したものを表-9 と表-10 に示す。なお、これらの表中の左変数と右

変数は 2 次形式で表された、すなわち、これらの変数が掛け合わされてできた、式 G を構成する項の左側から数えて第 1 番目と第 2 番目の変数を表す。また、ビット列は、表-6 と表-7 で与えられている 3 つのビットを示しており、10 個の要因はすべてこれら 3 ビットによって表現されている。また、要因関係図は、式 G から想定される各要因間の相関関係図である。

表-9 をみると、要因 B,G,H を除いてすべての要因が自己相関関係にある。また、要因 A は、要因 B,C,D,J と相互相関関係にある。また、要因 B は要因 J とのみ相互相関関係を持っている。さらに、要因 G,H は自己相関および相互相関関係の両方の関係を持っておらず、この判定に関して何等影響を及ぼさないことが分かる。表-10 を見ると、要因 D を除いてすべての要因が自己相関関係にある。また、要因 A は、要因 C,D,E,F,G,H,I,J と相互相関関係にある。また、要因 B は J とのみ相互相関関係を持っている。

このような観測データを整理し知識ベース化する場合、各要因に関してそのデータの独立性が高いことが望ましい。ここで扱ったデータはかなり独立性の高い結果となったが、要因 A が建設時・供用時共に他のほとんどの要因と相関性を持っていることが明らかとなった。ただし、供用時はその数が少ない。従って、この要因 A と他の要因との相関性を小さくするようなランク付けや要因設定等を検討することが必要であろう。

表-4 供用時の判定表⁶⁾

のり 面 番 号	評価ランク										崩 壊 ラ ン ク	のり 面 番 号	評価ランク										崩 壊 ラ ン ク	
	地形・地質要因							土工要因					地形・地質要因							土工要因				
	A	B	C	D	E	F	G	H	I	J			A	B	C	D	E	F	G	H	I	J		
1	4	4	5	5	4	2	3	2	2	2	2	13	3	3	3	4	4	3	2	2	2	2	2	
2	4	4	5	5	4	2	3	1	2	2	2	14	4	4	4	4	4	4	1	2	2	1	2	
3	3	3	3	3	5	4	2	2	2	4	2	15	3	3	3	4	3	4	1	3	1	2	2	
4	3	3	3	5	4	2	2	1	2	4	2	16	3	3	3	4	2	4	4	2	3	2	2	
5	2	2	2	5	4	3	3	2	3	3	3	17	5	5	5	2	3	5	4	4	5	2	5	
6	2	2	2	4	1	1	2	2	1	2	1	18	4	4	4	2	2	4	3	3	4	2	3	
7	2	2	3	3	3	4	2	4	3	2	2	19	1	1	1	1	2	2	2	1	2	1	1	
8	3	3	3	3	4	3	5	4	3	2	2	20	3	3	4	3	3	3	4	2	2	2	3	
9	4	3	4	3	4	5	2	3	4	4	4	21	4	4	4	3	5	5	5	4	5	5	5	
10	4	4	3	4	4	5	2	5	4	3	5	22	2	2	2	3	3	1	1	1	2	2	1	
11	3	3	3	3	3	5	2	4	3	2	3	23	3	3	3	4	3	4	4	4	4	3	4	
12	3	3	3	3	3	5	1	3	2	3	2	-												
備考 A:地山地質,B:土質分類,C:節理等の状態,D:摂理の傾斜,E:被覆層・風化層の厚さ F:地下水・湧水の状況,G:降雨水の集中度,H:のり高さ,I:のり勾配,J:のり面保護工																								

表-5 建設時の判定表⁶⁾

のり 面番 号	評価ランク										崩 壊 ラ ン ク									
	地形・地質要因							土工要因												
A	B	C	D	E	F	G	H	I	J	A	B	C	D	E	F	G	H	I	J	
1	3	3	3	2	3	2	3	1	2	2	2	3	3	3	3	2	1	1	1	5
2	4	4	3	2	2	2	3	1	1	5	1	2	2	2	4	3	2	2	4	
3	5	5	5	3	4	2	1	4	5	2	5	5	5	5	4	4	3	2	4	
4	3	3	3	2	3	2	1	3	3	2	2	2	2	2	4	3	2	1	4	
5	3	2	3	4	3	2	1	1	2	2	2	3	3	3	4	3	2	1	4	
6	4	3	3	3	3	2	1	1	2	5	2	3	3	2	2	3	2	1	4	
7	2	2	1	2	1	1	1	1	1	3	1	2	2	2	2	4	4	3	2	
8	4	3	4	4	3	2	3	1	2	1	2	2	2	2	1	1	2	1	2	
9	4	3	4	3	3	2	3	1	1	5	2	2	2	2	2	3	3	2	2	
10	4	4	4	5	4	4	4	3	5	2	4	2	2	2	3	3	2	2	2	
11	3	3	4	4	4	2	1	2	3	1	3	3	3	3	5	5	3	5	4	
12	3	3	4	4	4	2	1	2	2	1	3	3	3	3	2	4	2	2	4	
13	3	3	3	4	4	3	2	1	1	5	2	5	4	5	3	3	2	1	2	
14	3	3	3	2	3	5	1	3	4	1	3	3	3	2	4	2	2	1	2	
15	3	3	3	2	3	2	1	2	3	2	2	2	2	2	4	3	2	1	2	
16	1	1	1	2	2	2	1	2	2	1	1	2	2	2	4	3	2	1	2	
17	3	3	3	3	3	2	1	1	2	2	2	2	3	3	3	3	2	1	2	
18	3	3	3	3	3	4	1	1	1	2	2	3	3	3	2	3	2	1	2	
19	3	3	3	3	3	2	1	1	1	4	2	4	3	4	2	3	2	1	4	
20	2	3	2	2	3	2	3	1	1	5	1	4	3	4	3	3	2	1	4	
21	3	3	3	5	5	3	1	4	4	3	5	3	3	2	1	1	1	2	2	
22	2	2	2	3	3	2	1	1	2	4	2	4	4	4	4	4	5	3	4	
備考 A:地山地質,B:土質分類,C:節理等の状態,D:摂理の傾斜,E:被覆層・風化層の厚さ F:地下水・湧水の状況,G:降雨水の集中度,H:のり高さ,I:のり勾配,J:のり面保護工																				

表-6 供用時の学習結果

のり面番号	評価ランク										崩壊ランク
	地形・地質要因							土工要因			
	A	B	C	D	E	F	G	H	I	J	
	,,*	*,*,/	*,*,/	*,*,*	*,*,/	*,*,/	/,/,/	/,/,/	*,*,/	*,*,*	
1	1,0,0	1,0,0	1,0,1	1,0,1	1,0,0	0,1,0	0,1,1	0,1,0	0,1,0	0,1,0	2
23	0,1,1	0,1,1	0,1,1	1,0,0	0,1,1	1,0,0	1,0,0	1,0,0	1,0,0	0,1,1	4
演算式	$G = -X_1 \cdot X_1 + X_2 \cdot X_2 + X_3 \cdot X_3 + X_7 \cdot X_7 + X_{13} \cdot X_{13} + X_{16} \cdot X_{16} + 3 \cdot X_{25} \cdot X_{25} - 2 \cdot X_{28} \cdot X_{28} + X_1 \cdot X_3 + 3 \cdot X_1 \cdot X_5 + X_1 \cdot X_8 - X_1 \cdot X_{11} - X_1 \cdot X_{12} - X_2 \cdot X_{30} - X_3 \cdot X_{10} + 4 \cdot X_4 \cdot X_{28} + 2 \cdot X_{10} \cdot X_{12}$										
判定	23個の評価対象すべてについて表-4の西等の判定と一致する結果を与えた。										
備考	A:地山地質,B:土質分類,C:節理等の状態,D:摂理の傾斜,E:被覆層・風化層の厚さ F:地下水・湧水の状況,G:降雨水の集中度,H:のり高さ,I:のり勾配,J:のり面保護工										

表-7 建設時の学習結果

のり面番号	評価ランク										崩壊ランク
	地形・地質要因							土工要因			
	A	B	C	D	E	F	G	H	I	J	
	,,*	*,/,/	*,*,*	*,*,*	*,/,/	*,/,/	*,/,/	*,*,/	/,*,/	*,/,*	
1	0,1,1	0,1,1	0,1,1	0,1,0	0,1,1	0,1,0	0,1,1	0,0,1	0,1,0	0,1,0	2
44	1,0,0	1,0,0	1,0,0	1,0,0	1,0,0	1,0,1	0,1,1	0,1,1	1,0,1	0,1,1	4
演算式	$G=X_1 \cdot X_1+X_2 \cdot X_2-2 \cdot X_4 \cdot X_4+X_7 \cdot X_7+X_{13} \cdot X_{13}-X_{16} \cdot X_{16}+2 \cdot X_{19} \cdot X_{19}+5 \cdot X_{22} \cdot X_{22}+X_{26} \cdot X_{26}+X_{28} \cdot X_{28}+2 \cdot X_1 \cdot X_3+3 \cdot X_1 \cdot X_{10}-3 \cdot X_1 \cdot X_{13}+2 \cdot X_1 \cdot X_{16}-X_1 \cdot X_{26}-2 \cdot X_2 \cdot X_{19}-3 \cdot X_2 \cdot X_{22}+X_3 \cdot X_9-X_3 \cdot X_{11}+2 \cdot X_3 \cdot X_{16}-2 \cdot X_6 \cdot X_{30}$										
判定	44個の評価対象すべてについて表-5の西川等の判定と一致する結果を与えた。										
備考	A:地山地質,B:土質分類,C:節理等の状態,D:摂理の傾斜,E:被覆層・風化層の厚さ F:地下水・湧水の状況,G:降雨水の集中度,H:のり高さ,I:のり勾配,J:のり面保護工										

表-8 学習演算に有効なビット

評価時期	評価ランク									
	地形・地質要因							土工要因		
	A	B	C	D	E	F	G	H	I	J
供用時	12345	24	24	12345	4	4	-	-	4	145
建設時	12345	4	12345	12345	4	4	4	24	145	145
備考	A:地山地質, B:土質分類, C:節理等の状態, D:摂理の傾斜, E:被覆層・風化層の厚さ F:地下水・湧水の状況, G:降雨水の集中度, H:のり高さ, I:のり勾配, J:のり面保護工									

表.9 共用時の各変数間および各要因間の関連

式Gにおける各変数の関係		式Gから見いだされる要因関連図		
左の変数	右の変数	ビット列	要因関連図	ビット列
X_1	$(X_1, X_3, X_5, X_8, X_{11}, X_{12})$	1→3		1→3
X_2	(X_2, X_{30})	4→6		4→6
X_3	(X_3, X_{10})	7→9		7→9
X_4	(X_{28})	10→12		10→12
X_7	(X_7)	13→15		13→15
X_{10}	(X_{12})	16→18		16→18
X_{13}	(X_{13})	19→21		19→21
X_{16}	(X_{16})	22→24		22→24
X_{25}	(X_{25})	25→28		25→28
X_{28}	(X_{28})	28→30		28→30

表.10 建設時の各変数間および各要因間の関連

式Gにおける各変数の関係		式Gから見いだされる要因関連図		
左の変数	右の変数	ビット列	要因関連図	ビット列
X_1	$(X_1, X_3, X_{10}, X_{13}, X_{16}, X_{28})$	1→3		1→3
X_2	(X_2, X_{19}, X_{22})	4→6		4→6
X_3	(X_3, X_{11}, X_{16})	7→9		7→9
X_4	(X_4)	10→12		10→12
X_6	(X_{30})	13→15		13→15
X_7	(X_7)	16→18		16→18
X_{13}	(X_{13})	19→21		19→21
X_{16}	(X_{16})	22→24		22→24
X_{19}	(X_{19})	25→28		25→28
X_{22}	(X_{22})	28→30		28→30
X_{26}	(X_{26})			
X_{28}	(X_{28})			

特に、数量化理論は、複数の調査項目を線形結合式で評価するものであり、各調査項目間の独立性が保証されている場合に限って精度の高い判定値を示す。しかし、現実には、この独立性の高いデータを得ることは容易ではなく、必ず、これらの間には何らかの相関関係が混入してしまう。さらに、これらのデータには、データを採り判定する技術者の主観や曖昧な部分が混在しており、文献8)で実施されているような線形ファジィ回帰モデルのような特殊な手法を用いることが必要となってくる。これに対しGMDHは元々、データ構造を非線形なものと考えている上に、各項目間に相関関係があることを認めており、線形結合式による表現に比べてはるかに柔軟性に富んだ計算法と言える。

一方、人工ニューラルネットワークのもう一つの大きな特長である汎化機能についてここで触れておく。人工ニューラルネットワークの一種であるバックプロパゲーションでは、学習パターンに対する判定も汎化機能（未知パターンに対する判定）もすべて、収束計算によって求められた重み係数によって表現されている。しかし、この重み係数を眺めただけでは、学習パターンの持つデータ構造を理解することは困難である。これに対し、GMDHは表-3, 表-6, 表-7に示されているような2次多項式で、その学習パターンの持つデータ構造を表現している。従って、学習パターンと同様に未知パターンに関しても、すべてこの多項式の内挿と外挿の形式で明確に表現されていると言える。

5. まとめ

矛盾の無い客観的事実あるいは知識を抽出する作業が、ESを構築する上で大きな障害となっている。この問題がある程度解決されなければ、ESの土木技術における利用が急速に広がることは困難であると考えられる。

本論文では、構造同定手法として主に用いられているGMDHを上記の問題点の1つの有力な解決法と考え、これを知識ベース構築に採用することとした。そして、その基本的な考え方や具体的な知識獲得演算法の説明、土木分野における実際問題への適用、およびそれらの学習の結果から得られたデータ構造に関して分析を行った。その結果、以下に示すようなことが明らかとなった。

①判定結果を出力するに至った学習パターンのデータ構造を多項式形式で示すことができる。このため、データ構造の分析や必要とする調査項目を減少させることも可能である。

②調査項目に関して、そのデータ構造を一つの多項式として表現し、これを知識として蓄えるため、必要記憶容量が非常に小さくなる。また、判定演算も入力信号とパラメータの単なるかけ算と足し算だけであり、その計算量は非常に少ない。

③知識獲得処理が煩雑な辻褄合わせではなく、単純な数値演算の繰り返しとなるため、知識ベースのメンテナンスが容易である。

④1つの調査項目が追加される度に、単純な数値演算であるGMDHの知識獲得演算を行うため、新たな知識が追加される度に知識間の矛盾を検査することができる。⑤西等¹⁾によって実施された「のり面破壊に関する調査・判定」に対してGMDHを適用したところ、供用時と建設時の両方に関して、その判定結果がすべて教師信号と一致していることを確認できた。

⑥さらに、この調査項目に関するデータ構造を式Gによって表現できているため、各ビット（各変数）間の関係、すなわち各要因間の関係を明らかにし、また、必要不可欠なデータが何であるのかを示すことができる。

なお、今後の課題として、GMDHがどの程度の量の知識を記憶可能か、また、どのような性質の知識パターンに対して相性が良いかなどのGMDHの限界に関し、今後十分な検討を行なっていく必要がある。

【参考文献】

1. 佐野千遥：人工知能と人工生命，日刊工業新聞社，1994.2
2. 浅井喜代治編著：現代システム工学の基礎，オーム社，昭和54年11月
3. 林 勲：コンピュータの玉手箱，773 GMDH(その1)，bit，No.8，Vol.24，1992.8
4. 林 勲：コンピュータの玉手箱，773 GMDH(その2)，bit，No.2，Vol.25，1993.2
5. 古田秀博：GMDHによる知識ベース構築に関する考察，鋼構造年次論文報告集第3巻，pp.131-138，1995.11
6. 西邦正，古川浩平，中川浩二：773理論を用いたのり面の崩壊要因および崩壊可能性の評価について，土木学会論文集，No.445/Ⅲ-18，pp.109-118，1992.3