マトリックス・データ解析法による土木工学科学生の進路分析

米 長 　 泰*・今 村 眞 明**・殿 守 育 子**

Reasoning study on the civil engineering student courses after leaving the colleges by matrics data analytical method

Yasushi YONENAGA, Masaaki IMAMURA and Ikuko TONOMORI

(平成元年 9 月18日受理)

Matrics data analytical method is known as a tool by which we anderstand the peculiarity of the phenomenon theoretically. We now attempted to apply this tool to reasoning study on the civil engineering student courses after leaving the colleges. Then we could catch some indications concerning students' consiousness and each college policy.

1. 本文の趣旨

本年度は学科主任を命ぜられたが，本校に赴任してまだ日が浅く，学内のことも学生のことも，まして他校のことも理解が不十分であることを痛感している。基本的な事実を知るにはいろいろな方法があるが，ふとマトリックス・データ解析法の活用を思いついた。同手法に関してはこれまで基礎研究と産業界での適用事例を積上げてきたが，[1],[2]経験的に学生進路分析にも十分活用可能と判断した。

同手法はインプット・データを収集するときに，対象事象の特性をマクロに把握していないと成功しないし，アウトプットされた散布図の解読も事象を深く理解していないと結論を誤ってしまう。その代り日頃気がつかない新事実を発見できる楽しみがある。丁度良い機会だと考え，首記のテーマに取組むこととした。

2. マトリックス・データ解析法とは

産業界には作業現場関係のデータを効率よく整理するための「ＱＣ七つ道具」と，デスクワーク効率化のための「新ＱＣ七つ道具」が存在する。双方は日本が豊かな国に成長する過程で非常に重宝がられたツールである。後者は七つの単一手法の集合体であるが，いずれも現存する手法の中から拾い出している。少し複雑で難解なものを平易に活用し易いものへと改良し，広く全国に普及されたところに意義がある。文献として[3]等がある。七手法とその出生となる手法を表１に示す。

そのうちマトリックス・データ解析法とは，多変量解析の主成分分析法にあやかったものである。相関行列等の高度な数学理論はすべてパソコン・ソフトに任せ，人間が注力するのは最初のインプット・データ作成と最後の散布図解読である。インプット・データは対象物の種類と特性という二軸において，相互の構成要素をすべて数値によって表現する。例えば物体の美しさ，人間の感性等も度合を数値化するか，YES，Noを（０，１）で表わす必要がある。いずれの場合でも出発点では必ず二元マトリックスの数値表が登場し，その解析結果を云々することになるので，マトリックス・データ解析法と命名された。解析の主役は主成分分析法に他ならない。

結果の解読では各主成分ごとに構成データの特長をキーワードに表わし，それらを総合して主成分を意味づける。即ち「データが何を言おうとしているか」を無言のパソコンに代って弁明し，目的を達成することになる。いくつかのキーワードを包含した新しいキーワードを創案するところがポイントで，手法としては表１の親和図法に似ているといえる。

解読の仕方としてもっともシンプルな例を図１に示す。22人の学生の身長と体重を x_1 軸と x_2 軸にとっ

＊秋田工業高等専門学校土木工学科
＊＊三菱重工業株式会社技術本部

表１　新ＱＣ七つ道具の構成

七　手　法	出生となる手法
1. 連関図法	因果関係整理図
2. 親和図法	川喜多二郎 KJ法のＡ形図解
3. 系統図法	価値工学の機能系統図
4. マトリックス図法	製品展開の品質分析表
5. マトリックス・データ解析法	主成分分析法の散布図
6. PDPC法	進展予測の過程決定計画図
7. アロー・ダイヤグラム	PERT　日程計画図

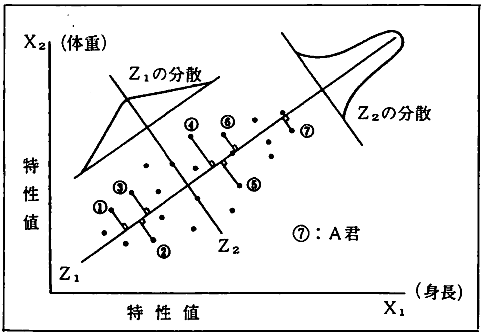

図１　身長と体重の相関

てプロットしたものである。明らかに身長が大きいと体重も大きい傾向にある。従って両者には強い正の相関があるといえる。ところで22個のデータの分散が最大となるような斜面Z_1を考えるとき，それは何を意味するのであろうか？

Z_1軸の右上方ではx_1とx_2がともに大きな値をとるから，初期データからのキーワードはそれぞれ,「身長が大きい」「体重が大きい」となる。これらを包含して「図体が大きい」という新しいキーワードを設定すると，事象を１個の変数で説明できることになる。それを図２に示す。一般に主成分の意味あいは，こんな要領で解読することになる。

このようにマトリックス・データ解析法は二元の数値データが揃えば，さまざまな事象の特長を科学的な裏付けのもとに把握することができる。筆者はこれ迄主として新製品を開発するときの彼我の特長分析と新仕様の検証等で活用してきたが，それを学生進路分析に適用してみようというわけである。

3. 主成分分析の論理

主成分分析の論理(4) は知らなくてもパソコン・ソフトが処理してくれるが，要点だけをとりまとめておきたい。いまＰ個の特性値があるケースを考える。

(1) 相関係数行列

$$\begin{bmatrix} r_{11} & r_{12} \cdots\cdots & r_{1P} \\ r_{21} & r_{22} \cdots\cdots & r_{3P} \\ \vdots & \vdots \quad \ddots & \vdots \\ r_{P1} & r_{P2} \cdots\cdots & r_{PP} \end{bmatrix}$$

特性値間の相関係数を行列に並べたもの。r_{ij}とは特性値x_iと特性値x_jの相関係数である。対角要素が１の値をとる対称行列である。

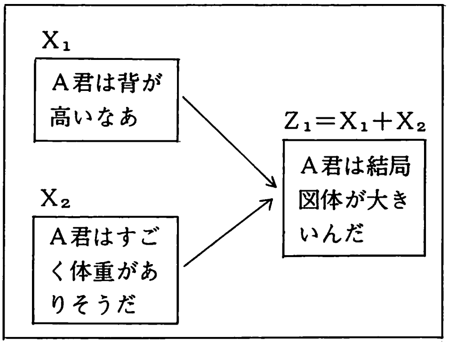

図２　キーワードの集約

(2) 固有値と固有ベクトル

下に示す行列方程式を解くことにより求められるものである。固有値 λ_K は主成分 Z_Kの分散に等しい。また固有値 λ_K（K＝１，２，…，Ｐ）の合計はPになる。Ｐ組の固有ベクトルが存在する。

$$\begin{bmatrix} r_{11} & r_{12} \cdots\cdots & r_{1P} \\ r_{21} & r_{22} \cdots\cdots & r_{2P} \\ \vdots & \vdots \quad \ddots & \vdots \\ r_{P1} & r_{P2} \cdots\cdots & r_{PP} \end{bmatrix} \begin{bmatrix} l_{K1} \\ l_{K2} \\ \vdots \\ l_{KP} \end{bmatrix} = \underset{\substack{\uparrow \\ \text{固有値}}}{\lambda_K} \begin{bmatrix} l_{K1} \\ l_{K2} \\ \vdots \\ l_{KP} \end{bmatrix} \text{固有ベクトル}$$

（条件$l_{K1}^2 + l_{K2}^2 + \cdots\cdots + l_{KP}^2 = 1$）

λ_Kを第Ｋ主成分の固有値と呼び，値が大きいほど重要な意味あいをもつ。値が１より小さな主成分は無視する。 $\lambda_1 \geqq \lambda_2 \geqq \lambda_3 \geqq \cdots$。

(3) 累積寄与率

$\frac{\lambda_1}{P}$, $\frac{\lambda_1+\lambda_2}{P}$, $\frac{\lambda_1+\lambda_2+\lambda_3}{P}$…を累積寄与率と呼び，事象全体の何％ぐらいが説明できているかを示す尺度である。通常90％前後に達すれば十分と考えられる。

(4) 主成分

固有ベクトル ℓ_{Ki} と特性値 X_i の積算和を第K主成分 Z_K という。 第m主成分迄求めるとき次式で与えられる。これらを順次第１主成分，第２主成分，……と呼ぶ。総合特性値である。上述のように $\lambda_m>1$ の範囲にとどめるのがよい。

$$\left.\begin{aligned} z_1 &= l_{11}x_1 + l_{12}x_2 \cdots\cdots + l_{3P}x_P \\ z_2 &= l_{21}x_1 + l_{22}x_2 \cdots\cdots + l_{2P}x_P \\ &\vdots \\ z_K &= l_{K1}x_1 + l_{K2}x_2 \cdots\cdots + l_{KP}x_P \\ &\vdots \\ z_m &= l_{m1}x_1 + l_{m2}x_2 \cdots\cdots + l_{mP}x_P \end{aligned}\right\}$$

(5) 因子負荷量

主成分 Z_K と元の特性値 X_i を基準化した X_i との相関係数のことである。 a_{iK} と表わせば

$$a_{iK} = \sqrt{\lambda_K}\ \ell_{Ki}$$

により求められる。もちろん $-1 \leqq a_{iK} \leqq +1$ であるが，絶対値が 0.6 より大きな a_{iK} だけに着目し，その符号を加味しながら第K主成分の意味あいを解読することになる。

(6) 主成分スコア

特性値 X_i がn個の個体から採取されたとき，最初のインプット・データはP×nの二元マトリックスとなる。第K主成分のスコアは固有ベクトル ℓ_{Ki} と基準化された特性値 X_i の積算和として求められる。即ち

$$Z'_K = l_{K1}x_1' + l_{K2}x_2' + \cdots\cdots + l_{KP}x_P'$$

固有ベクトル

実際には Z'_K が個体に応じてn個存在している。

(7) 散布図

２つの主成分スコアを直交座標上にプロットしたもの。これまでの数値解析結果がビジュアルになりわかり易い。一例を図３に示す。各個体にデータ番号を記すと見やすくなる。データは中心部になるほど平均的になり，左右上下に離散するほど何か特長を有することになる。実際にはさきの因子負荷量と散布図とを見比べながら，試行錯誤のうえ結論を得るのである。なお第３主成分迄取上げる場合，第１ vs 第２と第１ vs 第３の散布図でよいと思いがちであるが，念のため第２ vs 第３をブロットすると意外な新事実を発見することがある。これらは経験によって得られるノウハウである。

以上の道順をとりまとめると図４のようになる。

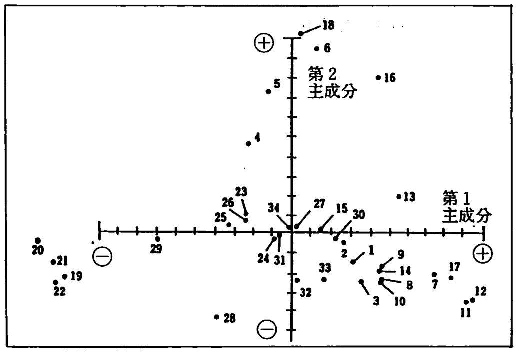

図３　散布図(第１主成分ＶＳ第２主成分)の一例

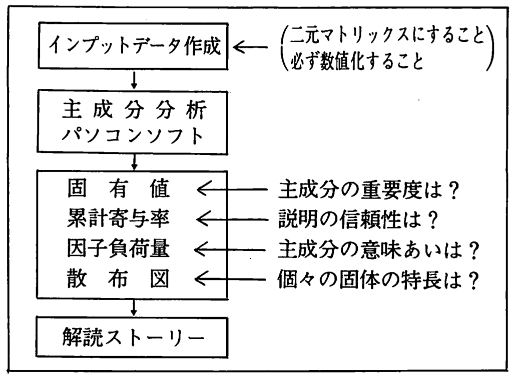

図４　マトリックス・データ解析法の活用要領

4. 学生進路分析

昭和63年３月卒業の全国高専土木工学科28校の学生進学・就職資料が手元に届いたので，これを基本入力として分析してみることにした。ただ「結果がこうでした」というものではなく，分析途上での筆者の考え方を明確にお伝えする。本法には正解というものがないし，時には誤解もありうる。何か見解の相異があったらご指摘いただきたい。

4・1　インプットデータの矞正

基本資料によれば卒業生の絶対数は最大49名，最小30名とバラツキが大きい。主成分分析の特長として，例えば企業体質を分析してみると，第１主成分には「図体の大きさ」，つまり売上が大きいとか社員が多いといった特長がプロットされる傾向にある。本題でもうっかりすると学生数の多少が第１主成分となり，肝心の分析目的が第２主成分以下となって重要度がぼけてしまう恐れがある。量よりも質について論じたい所存なので，インプットデータは各校とも卒業生が100名となるように調整し入力した。それを表２に示す。ここに公務員１はⅡ種，公務員

表2　各校進路内訳

単位：人

進路 ＼ 学校	1 公務員1	2 〃 2	3 〃 3	4 公団	5 建設業界	6 コンサルタント	7 基幹企業	8 ソフト	9 未定	10 進学	計
(A　1)	9	3	0	6	49	15	3	6	3	6	100
(B　2)	0	8	3	3	33	31	3	0	11	8	〃
(C　3)	0	0	13	7	40	0	7	0	20	13	〃
(D　4)	3	6	6	3	49	6	6	0	12	9	〃
(E　5)	0	0	24	9	39	3	6	0	6	12	〃
(F　6)	7	3	21	0	21	0	3	3	32	10	〃
(G　7)	6	0	3	0	30	9	6	0	25	21	〃
(H　8)	2	0	7	0	40	16	0	5	16	11	〃
(I　9)	0	12	12	0	37	12	0	6	6	15	〃
(J　10)	3	3	12	3	26	17	8	10	10	8	〃
(K　11)	6	0	17	6	30	3	8	0	6	24	〃
(L　12)	0	0	26	10	29	10	10	6	3	6	〃
(M　13)	24	0	14	5	23	3	5	3	8	15	〃
(N　14)	3	0	10	0	32	30	0	5	15	5	〃
(O　15)	3	0	19	0	34	19	3	6	3	13	〃
(P　16)	6	3	8	3	51	6	0	3	14	6	〃
(Q　17)	10	0	20	3	35	12	0	3	10	7	〃
(R　18)	6	0	3	0	58	18	3	0	3	9	〃
(S　19)	5	0	30	3	40	3	0	3	3	13	〃
(T　20)	0	0	22	0	53	9	0	0	0	16	〃
(U　21)	10	0	23	19	26	13	0	0	3	6	〃
(V　22)	2	2	10	2	47	21	0	7	7	2	〃
(W　23)	2	0	19	0	34	17	2	0	14	12	〃
(X　24)	3	0	13	3	60	9	3	3	6	0	〃
(Y　25)	0	0	6	3	38	13	9	9	13	9	〃
(Z　26)	3	0	17	8	24	14	6	0	17	11	〃
(a　27)	3	0	31	6	25	11	6	3	9	6	〃
(b　28)	0	0	0	0	65	6	0	0	10	19	100

2はⅢ種，公務員3は地方公務員の意である。なお主成分分析としての入力サイズは前述の記号でいうと

特性値の数　P＝10 ｝
個 体 の 数　n＝28 ｝（10×28マトリックス）

主成分の数　m＝4　として第4主成分までを取上げることとした。

4・2　各主成分の出力データと考察

表2入力データから得られた各主成分の出力データをとりまとめて表3に示す。これより以下のような事実を推察することができる。

(1)　総合評価

第5主成分以下は固有値が1以下となるので省略する。第1主成分の寄与率23.47％は一般に値が小さい。第4主成分迄の累積寄与率が66％となっているが，これは説明の信頼性が66％しかないということで，若干不都合である。通常第4主成分迄取上げれば90％近くになる筈であるが……。

ということは，各校とも就職・進学の状況に特に飛び抜けた特長，極端な差異というものがなく，指導方針，体制，学生の順応が大体似通っていることを意味している。おそらく第5主成分以下も固有値が0.9～0.8程度でしばらく続くものと思われる。

表3　出力データ概要

●印　要注意

主成分	固 有 値	累 積 寄 与 率	因有ベクトル	因 子 負 荷 量
第 1 主 成 分	$\lambda_1 = 2.347$	$\frac{\lambda_1}{P}$ 23.47 %	$\ell_{11} = 0.255$ $\ell_{12} = -0.312$ $\ell_{13} = 0.436$ $\ell_{14} = 0.420$ $\ell_{15} = -0.392$ $\ell_{16} = -0.388$ $\ell_{17} = 0.322$ $\ell_{18} = -0.184$ $\ell_{19} = 0.044$ $\ell_{1,10} = 0.163$	$a_{11} = 0.391$ $a_{12} = -0.478$ $a_{13} = 0.668$ ● $a_{14} = 0.643$ ● $a_{15} = -0.600$ ● $a_{16} = -0.594$ ● $a_{17} = 0.493$ $a_{18} = -0.282$ $a_{19} = 0.067$ $a_{1,10} = 0.250$
第 2 主 成 分	$\lambda_2 = 1.733$	$\frac{\lambda_1 + \lambda_2}{P}$ 40.80 %	$\ell_{21} = 0.003$ $\ell_{22} = 0.031$ $\ell_{23} = 0.245$ $\ell_{24} = 0.328$ $\ell_{25} = -0.212$ $\ell_{26} = 0.367$ $\ell_{27} = 0.049$ $\ell_{28} = 0.454$ $\ell_{29} = -0.331$ $\ell_{2,10} = -0.578$	$a_{21} = 0.004$ $a_{22} = 0.041$ $a_{23} = 0.323$ $a_{24} = 0.432$ $a_{25} = -0.279$ $a_{26} = 0.483$ $a_{27} = 0.065$ $a_{28} = 0.598$ ● $a_{29} = -0.436$ $a_{2,10} = -0.761$ ●
第 3 主 成 分	$\lambda_3 = 1.457$	$\frac{\lambda_1 + \lambda_2 + \lambda_3}{P}$ 55.37 %	$\ell_{31} = 0.040$ $\ell_{32} = -0.264$ $\ell_{33} = 0.163$ $\ell_{34} = 0.193$ $\ell_{35} = 0.515$ $\ell_{36} = -0.138$ $\ell_{37} = -0.370$ $\ell_{38} = -0.320$ $\ell_{39} = -0.581$ $\ell_{3,10} = -0.062$	$a_{31} = 0.048$ $a_{32} = -0.319$ $a_{33} = 0.197$ $a_{34} = 0.233$ $a_{35} = 0.622$ ● $a_{36} = -0.167$ $a_{37} = -0.447$ $a_{38} = -0.386$ $a_{39} = -0.701$ ● $a_{3,10} = -0.075$
第 4 主 成 分	$\lambda_4 = 1.063$	$\frac{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4}{P}$ 66.00 %	$\ell_{41} = -0.808$ $\ell_{42} = -0.007$ $\ell_{43} = 0.109$ $\ell_{44} = 0.071$ $\ell_{45} = 0.204$ $\ell_{46} = -0.104$ $\ell_{47} = 0.505$ $\ell_{48} = 0.016$ $\ell_{49} = -0.132$ $\ell_{4,10} = 0.072$	$a_{41} = -0.833$ ● $a_{42} = -0.007$ $a_{43} = 0.112$ $a_{44} = 0.073$ $a_{45} = 0.210$ $a_{46} = -0.107$ $a_{47} = 0.521$ ● $a_{48} = 0.016$ $a_{49} = -0.136$ $a_{4,10} = 0.074$

(2) 第1主成分の意味あい

a_{13}とa_{14}がプラス，a_{15}とa_{16}がマイナスで対比している。学生は地方公務員・公団を希望するか，それとも建設業界・コンサルタント会社に就職するか。いずれにしても建設現場の第一線に立って設計・製作・施工・監督等にタッチしたいとするグループが第一順位で指摘されている。

(3) 第2主成分の意味あい

a_{28}のプラスと$a_{2,10}$のマイナスが対比している。ソフト産業就職か進学かを選択するグループである。察するに保護者の経済事情により若干ゆとりがあれば進学を志ざし，ゆとりがなければ，あるいは早く実社会に出たいという希望があれば，頭脳勝負型のソフト産業にタッチしたいというものである。頭が良くパソコンも得意とするグループであろう。

(4) 第3主成分の意味あい

a_{35}のプラスとa_{39}のマイナスが負の相関を示している。建設業界の現場は文字どおり切った張ったの鉄火場である。ある意味では男らしく，ある意味ではハイテク化課題の宝庫だといえる。このような業界へ飛び込むか全く異種の方面に転身するか，ともかく公務員と進学の試験に手つかずの状態で，どこへ落付いたらよいか選択を迷っているグループだといえる。

(5) 第4主成分の意味あい

a_{41}のマイナスとa_{47}のプラスが相反関係にある。上級公務員を志望するか基幹産業への就職を採るか。志を大にし，一面都市部の一流大学学生に多く見られる傾向と一致している。勿論優秀な人材が集まるから昇進の道もけわしいが，国家を揺り動かすような大事業にタッチする機会に恵まれる。また別の意味で都市対抗野球に出場したいなど，体育文化活動の選手を狙うには基幹産業が適している。なお第3主成分迄で打切ろうとして，念のため第4主成分をアウトプットしたところ，貴重な事実を把握することができた。

4・3　散布図と各校の特長

各主成分の散布図を図5・6・7・8に示す。主成分スコアZ_Kの値は散布図から相対比較ができるので，数値表は省略する。散布図に対して若干の考察を試みる。

(1) 第1主成分 vs 第2主成分（図5）

- L校とU校は地方公務員・公団の比率が高く，建設業界の比率が低い。また進学率も低い。この点両校は非常によく似ている。
- b校は公務員関係がゼロであるが，3分の2が建設業界へ就職し，進学率も高い。L校・U校と全く対称的な方針をとっていることがわかる。
- K校は公務員の比率とともに進学率も高い。L校と半分似て半分違う。もっとも受験勉強に注力している学校ではないだろうか。
- v校は公務員と進学の実績が一番少ない。ただし民間に就職を奨励することは立派な方針と思う。

(2) 第1主成分 vs 第3主成分（図6）

- T校の学生は何のとまどいもなく建設業界の現場へ飛び込んでいる。U校もとまどいが少ない。反面F校は建設業界へ決心すべきかどうか一番悩みを抱えた学校のように思われる。
- 公務員比率ととまどいの度合ではU校とB校が全く対称的だといえる。

(3) 第2主成分 vs 第3主成分（図7）

- 進学と就職未定者の多少はU校とG校が対称的である。

(4) 第1主成分 vs 第4主成分（図8）

- 図5で似かよっていたL校とU校が第4主成分で全く別れてしまった。L校は基幹企業へ10%就職しているのにU校はゼロである。

(5) 平均校

- W校は図5～8においていずれも原点近くにある。諸事に関し全国の平均的存在だと推察できる。
- 本学はズバリD校である。平均校の中に入る。第2主成分スコアが若干マイナス寄りであるが，これはたまたまソフト産業への就職がゼロだったためである。

(6) 散布図の読図について

- 以上の指摘は少し訓練を積めば，散布図と因子負荷量との関連において誰にも可能なことである。
- 表2の数表を詳しく眺めれば，5学年担任教官なら上述程度のことは常識的に気付くことであろう。しかしそれらの事実を第1，第2，第3と順位づけ，しかも全体でなお検討すべき事項の何%が舌足らずなのか，パソコンは科学的に指摘してくれる。
- 筆者のような勤続の浅い教官にも事実が手に取るようにわかるし，またベテラン教官が居眠りして見過してしまっても，ドキュメントが忠実に記録に留めてくれる。
- マトリックス・データ解析法は結局たいへん便利な一面を備えていることが伺われる。

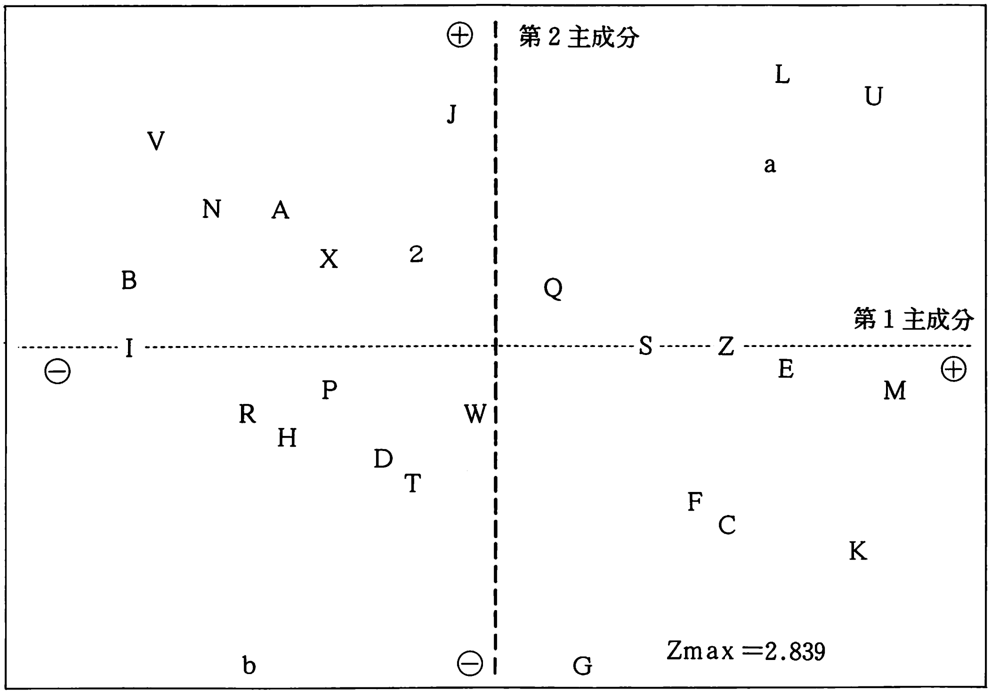

図５　第１主成分ＶＳ第２主成分

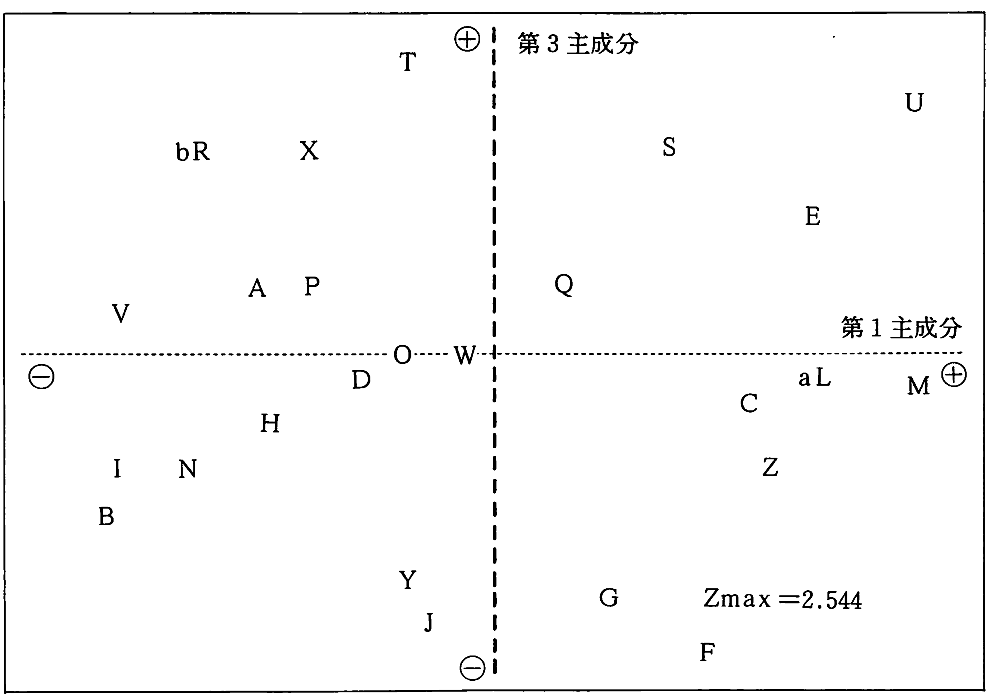

図６　第１主成分ＶＳ第３主成分

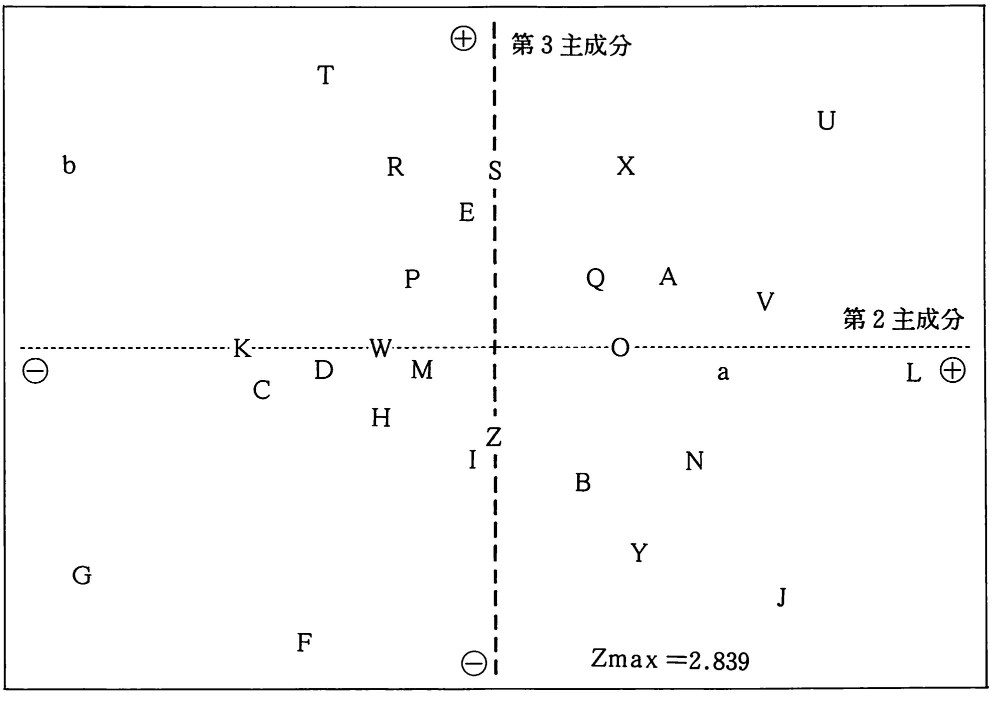

図7　第2主成分VS第3主成分

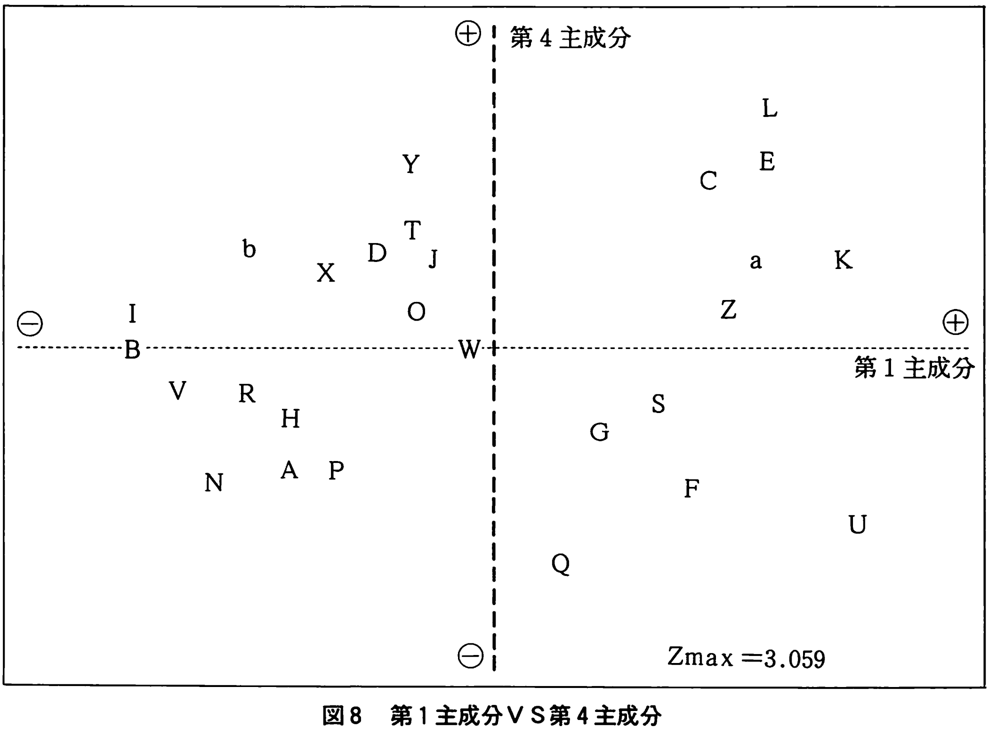

図8　第1主成分VS第4主成分

5. 残された課題

図5～8の散布図に従い特長ある学校を何校か指摘したのであるが，本学がすこしも登場して来ないなら，苦労の意味がない。ではどうすれば良いのか？　ひとつの方法として，25周年記念誌に要約された過去15年間の進路先資料にもとづいて，これ迄と全く同じ解析を試みるとよい。（実際に近い将来実施してみるつもりであるが……）

つまり表2の全国28校の代りに，本学15年間の各年度をそれぞれひとつの学校（ひとつの個体）とみなし，横軸の特性値はそっくり同じにして実行すればよい。そうすると15年間に同じ学校の同じ学科でどのような変遷を辿ったのか，鮮明に浮彫りされる。

昨年と今年と似ているのか異質なのか，その原因は何か。また業界の好不況や求人動向などの周辺条件を加味すれば，これ迄の進路指導の良かった点，至らなかった点が従来にも増して把握し易くなるだろう。また近未来の世間の動向に対しても，進路指導方針に適切に折込み易くなる筈である。

次に本稿で指摘したいくつかの特長校を訪問して，実情はどうなっているか調査してみることも有用と思われる。これ迄の解析結果から，どの学校のどの面を知ればよいか，的が絞られている。ただ具体的に記述することは控えたい。

6. 結　　語

マトリックス・データ解析法は主成分分析という高度な数学をパソコンの助けを借りて，我々教官・技術者・学生に活用の門戸を開いている。同手法は対象が学生であっても，会社人間でも，商品でも，企業や国家の比較であっても，何に対しても応用が効く，事象の特性を把握する手法である。ただしインプット・データの作成方法と因子負荷量の評価の仕方に熟達しないと実戦の武器とはならないことは，他のあらゆる手法と全く同じことで，言及の必要はないだろう。

表3に暗示された各主成分の意味あいをじっくり解読していると，若い学生諸君の心の動きがわかるような気がして，思わず親しさが倍旧するのを感じている。このような分析は数学的な興味からではなく，あくまで学生を指導し奉仕する信念で取組まなければ効果が生まれない。本稿が最終学年担任教官や受入れ企業人事部門長の一助となれば幸いである。

参 考 文 献

1) 米長粂他　新ＱＣ七つ道具活用の手引（三菱重工社内技術マニュアル）　1986. 3.
2) 米長粂　Ｎ７社内普及活動および製品開発におけるマトリックス・データ解析法の活用について　日科技連第10回Ｎ７シンポジウムテキスト　P. 161～166　1988. 3.
3) 納谷嘉信他　新ＱＣ七つ道具の企業への展開　日科技連
4) 今村真明，殿守育子他　統計的手法活用マニュアル（主成分分析）ＭＭ 06－861　三菱重工技術本部・非売品