2次合成変量のアルファ係数の和を最大化する 多集合データの階層的主成分分析 II

村 上 隆

(名古屋大学)

Hierarchical component analysis of multiset data
maximizing the sum of alpha coefficients of second order composites II

TaKashi MURAKAMI
(Nagoya University)

Murakami(1987a) developed the hierarchical component analysis maximizing the sum of alpha coefficients of second order composites as a method for analyzing multiset data. The present paper offers some additional information on his method. First, three different types of hierarchical component analysis methods are formulated and compared with the present method. The main features of the method, that is, the scale free property and the interpretability of the second order composites are stressed. Second, three data sets, artificial and real, are analyzed by the method. The results showed that the fidelity and the bandwidth of the first order composites can be balanced according to expectation through changing the number of second order composites.

本研究は、村上(1987a)において定式化された、2次合成変量の α 係数の和の最大化を最適化基準とする階層的主成分分析について、幾つかの検討を行うことを目的とする。

1. 他のモデルとの関係

この節では、本研究で提案したモデルと、村上(1986)の階層的主成分分析のモデル、その他との関係を明らかにしたい。

まず、通常的主成分分析に関する幾つかの定式化とその間の関係について、考えてみよう。これには、大体以下の4つの方式がある。なお、 $\mathbf{Z}$ は $n \times N$ のデータ行列で、各行ごとに平均0、分散1に標準化されているものとする。したがって、 $\mathbf{R} = \frac{1}{N} \mathbf{Z} \mathbf{Z}'$ は変数間相関行列である。また、 $q (\leq \min(n, N))$ を合成変数の数とする。

定式化 I

$\mathbf{W}$ を $n \times q$ の重み行列とし、合成変量を $\mathbf{G} = \mathbf{W}' \mathbf{Z}$ で定義する。制約条件を $\mathbf{W}' \mathbf{W} = \mathbf{I}$ とする。最適化基

準は、

$$\text{tr} \frac{1}{N} \mathbf{G} \mathbf{G}' = \text{tr} \mathbf{W}' \mathbf{R} \mathbf{W} \quad (1-1)$$

の最大化、すなわち、合成変量の分散の和の最大化である。

この方式は、通常の変数解析のテキストで採用されていることが多い。(ただし、合成変量は、1つずつ順に定義される形をとることが多く、素データは、分散1に標準化されていないことが多い。)

定式化 II

定式化 I と同じく、 $\mathbf{F} = \mathbf{V}' \mathbf{Z}$ で合成変量を定義する。ただし制約条件は、 $\frac{1}{N} \mathbf{F} \mathbf{F}' = \mathbf{V}' \mathbf{R} \mathbf{V} = \mathbf{I}$ とする。最適化基準は、

$$\text{tr} \mathbf{V}' \mathbf{V} \quad (1-2)$$

の最小化、すなわち、合成変量の α 係数の和の最大化である。

この方式は、村上(1987a)の第2節で論じたものである。

定式化Ⅲ

A を $n \times q$ の因子負荷行列, F を $N \times q$ の因子得点行列とし, 制約条件は, $\frac{1}{N} FF' = I$ とする。最適化基準は,

$$\text{tr}(Z - AF)(Z - AF)' \quad (1-3)$$

の最小化。

この方式は, Keller (1962) による。これは, 通常の(共通)因子分析モデル,

$$Z = AF + E \quad (1-4)$$

において, 行列 EE' を対角行列とするという制約条件を外したものに相当し, 広い意味で因子分析と呼ばれていることもある(例えば, プログラム・パッケージ SPSS における PA1 等)。

定式化Ⅳ

定式化Ⅲの式 (1-4) から出発し, 制約条件 $\frac{1}{N} FE' = 0$ を加える。ここから,

$$R = AA' + \frac{1}{N} EE' \quad (1-5)$$

これにもとづいて, 最適化条件は,

$$\text{tr}\left(\frac{1}{N} EE'\right)^2 = \text{tr}(R - AA')^2 \quad (1-6)$$

の最小化とする。

これは, 結果的には, 相関行列に最小2乗法の意味で最もよく当てはまる因子負荷行列 A を求めることに相当する。ここで, 相関行列の対角要素を1でなく, 適当な共通性(行列 AA' の対角要素)に置き換えれば, 主因子法による因子分析そのものとなる。

ここでは, 導出過程の詳細は省略せざるを得ないが, これら4つの定式化は, 相関行列 R の固有値, 固有ベクトルを求める問題に帰着する, という意味で, すべて同等である。すなわち, A を R の最大 q 番目までの固有値を要素とする対角行列, Γ をそれに対応する正規直交固有ベクトルの行列とする ($R\Gamma = \Gamma A$, $\Gamma'\Gamma = I$)。このとき, 各定式化における重みあるいは因子負荷は,

$$W = \Gamma \quad (1-7)$$

$$V = \Gamma A^{-\frac{1}{2}} \quad (1-8)$$

$$A = \Gamma A^{\frac{1}{2}} \quad (1-9)$$

のように, それぞれ, Γ の各列に比列した形で得られる。また, 合成変量については, 定式化Ⅱ~Ⅳの F はすべて同一であり, 定式化Ⅰの G との間には,

$$G = FA^{\frac{1}{2}}$$

図1 定式化Ⅰの主成分分析における主成分の回転における問題。

回転前(実線)では, 2つの主成分は独立。しかし, 回転後(点線)では, 2つの主成分は相関をもつ。

の関係がある。

なお, これらの定式化はすべて, 直交回転に関する不定性を有する。ここで, 定式化Ⅱ~Ⅳの各方式では, 合成変量が, 分散1で直交するという性質をもつから, この性質自体は回転しても変化しない。一方, 定式化Ⅰでは, 合成変量の分散が, 一般に相互に異なるので, 回転後の合成変量は, 相互に相関を有することになる(図1)。この点が, 通常の定式化Ⅰの記述が, 合成変量を順次求める形でなされ, 回転を導入しない形になっていることが多い理由であろう(制約条件として, $W'RW$ が対角行列であることを加えれば, 回転は不可能になる)。

以上を要約すれば, 定式化Ⅰの合成変量の単位を除き, 主成分分析の4つの定式化はすべて本質的に同一の解を生み出す, ということである。

階層的モデルに関しても, 通常の場合と平行的に4つの定式化を考えることができる。記号その他については, 3~4節のそれにしたがうが, 素データ行列 Z_k は, 各列が, 平均0, 分散1に基準化されているものとする。

定式化Ⅰ'

1次合成変量を, $F_k = V_k' Z_k$, 2次合成変量を, $G = \sum W_k^* F_k$ によって, 定義する。制約条件は,

$\frac{1}{N} F_k F_k' = V_k' R_{kk} V_k = I$, および, $\sum W_k^* V_k V_k' W_k^* = I$ とし, 最適化基準は,

$$\text{tr} \frac{1}{N} GG' = \text{tr} \sum_{k=1}^p \sum_{l=1}^p W_k^* V_k' R_{kl} V_l W_l^* \quad (1-10)$$

の最大化とする。このやり方は, 2次合成変量について, 通常的主成分分析における定式化Ⅰと同様, 分散の最大

化を目指すものである。この方法に関しては村上(1987b)に詳細に論じられている。

定式化Ⅱ'

一次合成変量が $F = V_k' Z_k$ 、二次合成変量が、 $G = \sum W_k V_k' Z_k$ によって定義され、制約条件の1つが、 $V_k' R_{kl} V_k = I$ である点は、定式化Ⅰと同じであるが、第2の制約条件は、 $\sum \sum W_k' V_k' R_{kl} V_l W_l = I$ とし、最適化基準は、

$$\text{tr} \sum W_k' V_k' V_k W_k \quad (1-11)$$

の最小化とする。

この定式化が、(素データの基準化を除き、) 村上(1987a)で展開した、 α 係数の和を最大化する階層的な主成分分析である。定式化Ⅰ'とⅡ'は、制約条件と最適化基準が入れ替わり、一種の双対(dual)となっているが、この点では、通常の場合の定式化ⅠとⅡも同様である。

定式化Ⅲ'

A_k を $n_k \times q_k$ の一次因子負荷行列、 C_k を $q_k \times Q$ の二次因子負荷行列、 F を $Q \times N$ の因子得点行列として、

$$Z_k = A_k C_k F + E_k \quad k = 1, \dots, p \quad (1-12)$$

なる形の基本モデルを考える。制約条件は、 $\frac{1}{N} F F' = I$ 、 $C_k C_k' = I$ とする。最適化基準は、

$$\text{tr} \sum E_k E_k' = \text{tr} \sum (Z_k - A_k C_k F) (Z_k - A_k C_k F)' \quad (1-13)$$

の最小化とする。

これは、村上(1986)における階層的な主成分分析である。

定式化Ⅳ'

基本モデル(1-10)から出発するが、制約条件、 $F_k E_k' = 0$ を追加する。これによって、

$$R_{kl} = A_k C_k C_l' A_l' + \frac{1}{N} E_k E_l' \quad k = 1, \dots, p \\ l = 1, \dots, p \quad (1-14)$$

という基本方程式が得られる。最適化条件は、

$$\sum_{k=1}^p \sum_{l=1}^p \text{tr} (R_{kl} - A_k C_k C_l' A_l') (R_{kl} - A_k C_k C_l' A_l') \quad (1-15)$$

の最小化である。

この形のモデルは、村上(1983, p.158)において、(3相データのみを対象とする特殊な形で)触れられている。

さて、このとき、これらの4種の定式化のうち3つまでは、実質的に異なった解に帰着する。すなわち、解はいずれも、2つの方程式を同時に満足しなければならない。すなわち、 Γ_k を $q_k \times Q$ の正規直交行列、 r_k を $n_k \times q_k$ の正規直交行列、 Θ を $Q \times Q$ の正定符号の対角行列として、方程式の一方は、

$$\sum_{l=1}^p r_k' R_{kl} r_l \Gamma_l = \Gamma_k \Theta \quad k = 1, \dots, p \quad (1-16)$$

である(これは、村上(1987a)の(4-22)そのもの)が、他方は、 A_k を $q_k \times q_k$ の正定符号の対角行列として、

定式化Ⅰ' 及びⅢ'

$$\{(\sum_{l=1}^p R_{kl} r_l \Gamma_l) \Theta^{-1} (\sum_{l=1}^p \Gamma_l' r_l' R_{lk})\} r_k = r_k A_k \\ k = 1, \dots, p \quad (1-17)$$

定式化Ⅱ'

$$\{(\sum_{l=1}^p R_{kl} r_l \Gamma_l) \Theta^{-3} (\sum_{l=1}^p \Gamma_l' r_l' R_{lk})\} r_k = r_k A_k \\ k = 1, \dots, p \quad (1-18)$$

定式化Ⅳ'

$$\{(\sum_{l=1}^p R_{kl} r_l \Gamma_l) (\sum_{l=1}^p \Gamma_l' r_l' R_{lk})\} r_k = r_k A_k \\ k = 1, \dots, p \quad (1-19)$$

となり、左辺中央の Θ にかかるべき指数が異なることになる。(それぞれの導出の詳細は省略。ただし、(1-17)は、村上(1987b)参照。式(1-18)は、村上(1987a)における(4-23)に(4-22)を変形して代入することによって得られる。ただし、そこでの A^{-1} を Θ とおく。)なお、(1-17)~(1-19)は、いずれも、行列の固有値、固有ベクトルを求める形式になっているが、(1-17)と(1-19)では、 A_k の要素を常に最大固有値に対応するようにとるべきことが示されるのに対し、(1-18)ではそうでないために、4節でのべたようなアルゴリズムを必要とすることになった。

これらの行列と本来のパラメータとの関係は、かなり複雑なので省略する。

いずれにしても、階層的な主成分分析では、通常の主成分分析における複数の定式化に対応する解が、必ずしも一致しないことがわかった。このことは、手法の選択と結果の解釈にあたって、多少面倒な問題を引き起こすことになる。すなわち、通常の主成分分析にあっては、定

式化 I を除き、どのような定式化をとっても同一の解を得ることになるから、特にどの定式化をとるかを事前に定める必要がないし、解釈にあたっては、複数の定式化の観点を適当に混在させることができるが、階層的モデルでは、あらかじめモデルを選ぶ必要があるとともに、解釈もその定式化の趣旨にもとづいて、なされねばならぬからである。

さて、(1-17)、(1-18)、(1-19) のアルゴリズムは、各集合の 1 次合成変数の数がすべて等しく、2 次合成変数の数もそれに等しい場合、すなわち、

$$q_1 = q_2 = \dots = q_p = Q \quad (1-20)$$

の場合に限り、同一の解に収束する。しかもそれは、村上 (1987b) に示したように、全集合の変数を全体として同時に主成分分析したものと同じになる。

一方、2 次合成変数の数が、1 次合成変数の数の総和に等しい、すなわち、

$$Q = \sum q_k \quad (1-21)$$

の場合、(1-17) は、集合ごとの個別主成分分析の結果と一致するが、(1-18) では一致しない。この点で、2 次合成変数の α 係数の和の最大化基準は、通常の主成分分析とのつながりがやや薄いということになる。

結局のところ、2 次合成変数の α 係数の和の最大化基準が、2 次合成変数の分散の和の最大化基準と比較して優れている点は、

- 1) 2 次合成変数を直交させたままの回転が可能であり、2 次合成変数そのものを解釈の対象とするのに便利である。
 - 2) 定式化全体が scale free である。すなわち、データの標準化に必然性がある。
- という 2 点になる。その一方で、前述の個別主成分分析との関係に関する問題点も生起する。

2. 人工データへの適用

ここでは、前節で述べたような、2 次合成変数の α 係

表 1 人工データ I の相関行列

		集合 1		集合 2	
		1	2	1	2
集合 1	1	1000	870	870	500
	2	870	1000	500	000
集合 2	1	870	500	1000	870
	2	500	000	870	1000

(小数点省略)

図 2 人工データ I の変数と 1 次合成変数
① $Q = 1$ ② $Q = 2$ ③ 個別主成分

数の和を最大化する階層的な主成分分析の特徴をもう少し明らかにするために、この方法を幾つかの人工データに適用して、その結果を検討してみよう。

まず、表 1 の相関行列を対象にする (これを人工データ I と呼ぶ)。ここで、それぞれの集合には 2 つずつの変数が含まれているが、それらは図 2 のように、完全に 2 次元空間の中におさまっている。このデータにおいて、1 次合成変数の数を 1 とし、2 次合成変数の数 Q を 1 とした場合と 2 とした場合における、1 次合成変数の負荷行列を、個別主成分分析の場合と並べて表 2 に、それぞれの場合における 1 次合成変数間相関行列を表 3 にそれぞれ示した。なお、表 2 には 1 次合成変数の α 係数を付記した。最大化されるべき、2 次合成変数の α 係数の和は、 $Q = 1$ のとき、約 0.87、 $Q = 2$ のとき、約 0.90 である。また図 2 には、それぞれの条件における 1 次合成変数を、ベクトルとして記入してある。

図 2 を見るとまず、個別主成分分析は当然のことながら、それぞれの集合の 2 つのベクトルのちょうど中間 (平均) のところに 1 次合成変数を生成する。このデータの場合には、2 つの変数の単純和を分散 1 に基準化したものでもある。2 つの集合の 1 次合成変数ベクトルの角度は 60° であり、この余弦が相関係数 0.5 に相当する。階層的な主成分分析では、 $Q = 1$ のとき、2 つの 1 次合成変数ベクトルの角度は約 55° となり、相互相関は、0.57 まで上昇する。 $Q = 2$ では、この角度は相関係数 0.38 に対応する 68° に広がる。

こうして見ると、2 次合成変数の α 係数の和の最大化基準は、一応想定したように機能していることがわかる。

表 2 人工データ I の 1 次合成変量の負荷行列

	$Q = 1$		$Q = 2$		個別主成分	
	I	h^2	I	h^2	I	h^2
1	977	954	948	899	967	935
2	956	913	982	964	967	935
2 乗和	1867	1867	1862	1862	1870	1870
α 係数	907		868		930	

(小数点省略)

表 3 人工データ I の 1 次合成変量間相関行列

	$Q = 1$		$Q = 2$		個別主成分	
	集合 1	集合 2	集合 1	集合 2	集合 1	集合 2
集合 1	1000	570	1000	379	1000	500
集合 2	570	1000	379	1000	500	1000

(小数点省略)

表 4 人工データ II の 1 次合成変量の負荷行列

	$Q = 1$		$Q = 2$		個別主成分	
	I	h^2	I	h^2	I	h^2
1	530	280	400	160	483	233
2	530	280	400	160	483	233
3	530	280	400	160	483	233
4	530	280	400	160	483	233
5	527	278	627	393	575	330
6	527	278	627	393	575	330
7	527	278	627	393	575	330
8	527	278	627	393	575	330
2 乗和	2233	2233	2209	2209	2254	2254
α 係数	626		614		636	

(小数点省略)

表 5 人工データ II の 1 次合成変量間相関行列

	$Q = 1$		$Q = 2$		個別主成分	
	集合 1	集合 2	集合 1	集合 2	集合 1	集合 2
集合 1	1000	462	1000	260	1000	394
集合 2	462	1000	260	1000	394	1000

(小数点省略)

すなわち、 $Q = 1$ の場合は、1 次合成変量間相関を高くするという意味において、より bandwidth を強調する 1 次合成変量が得られている。ただ、 $Q = 2$ の場合は fidelity を強調する解になっているかどうかは問題である。この場合、 α 係数で見た内的整合性の最も高い 1 次合成変量は、このデータでは個別主成分分析の場合に得られ、 $Q = 2$ の 1 次合成変量の α 係数は、むしろ $Q =$

1 の場合をすら下回っているからである。このデータに関する限り、 Q を増加させることは、1 次合成変量の内的整合性の向上よりも、1 次合成変量間の相関を低下させることにつながっている。

2 次合成変量は、 $Q = 1$ のとき、ちょうど 2 つの集合の中間に相当するところに来る。一方、 $Q = 2$ のときは、それぞれの集合の 2 番目の変数と一致する。このこと

は、2次合成変量の直交性の仮定からして直観的にも明らかであろう。ただ、このようにいわばデータの端の位置に2次合成変量を設定せざるを得なくなることが、 $Q = 2$ の場合において、1次合成変量の内的整合性が低下する一因である。

次に、村上(1987a)における表1の相関行列(人工データⅡと呼ぶ)に適用する。この場合も、各集合の1次合成変量の数は1とし、2次合成変量の数 Q を1または2として計算する。表4に1次合成変量間相関行列(2つの集合とも同じである)、表5に1次合成変量間相関行列をそれぞれ示した。

まず、負荷について見ると、 $Q = 1$ では集合間相関の高い変数1~4の負荷が、集合内相関の高い変数5~8の負荷を上回っている。 $Q = 2$ では、逆に変数5~8の負荷が変数1~4の負荷を上回っているが、先のデータと同様、その結果は個別主成分分析以上に変数5~8の側に偏っており、 α 係数は $Q = 1$ の場合よりもかえって低下している。

1次合成変量の集合間相関は、 $Q = 1$ 、個別主成分分析、 $Q = 2$ の順に低下している。値の範囲は、.26から.46の範囲にあり、これらは村上(1987a)で示した変数1~4、または5~8だけを選択して単純和で尺度構成した場合ほど極端な値(.00~.47)にはなっていない。

このようにどちらのデータでも、 Q を増加させると、予想通り集合間相関は低下するが、 α 係数を内的整合性の指標とする限り、内的整合性も低下してしまう。これは予期に反する結果であるが、1節で述べたこの方法と個別主成分分析との関連性の問題である。さらに、さまざまな形で人工データを生成して計算したが、 Q の増加とともに α 係数が増加するケースは(1次合成変量の数が1であるかぎり)かなり稀なようである。

3. 実際のデータへの適用

斎藤・村上・若林(1986)のデータの一部を分析する。このデータは、村上(1987b)でも分析されているが、愛知県下の大学一年生592名から得られたものである。

変数の集合は、

- 1) 職業志向性に関する10の質問項目(5段階評定)
- 2) ライフ・スタイルに関する10の質問項目(5段階評定)
- 3) 専攻(理科系を1, 文科系を2とコーディングした2値変数)
- 4) 性別(男子を1, 女子を2とコーディング)
- 5) 東海地方への将来の居住希望の程度(5段階評定)
- 6) 8つの企業の就職先としての魅力度(5段階評定)の6つである。したがって、 $p = 6$, $n_1 = n_2 = 10$,

表6 2次合成変量の数の関数としての α 係数の総和
($q_1 = q_2 = q_6 = 2$)

Q	A
2	1.526
3	2.072
4	2.534
5	2.944
6	3.136
7	3.112

$n_3 = n_4 = n_5 = 1$, $n_6 = 8$ である。また、1次合成変量の数は、 $q_1 = q_2 = q_6 = 2$, $q_3 = q_4 = q_5 = 1$ とする。集合3~5の1次合成変量は、変数の値そのものとなる。この条件の下で、2次合成変量の数 Q を変化させた。

表6は、 Q が2から7までの場合における、最大化された2次合成変量の α 係数の和を示したものである。最大の値は、 $Q = 6$ の場合において得られていることがわかる。もちろん、この値は1次合成変量の数の影響も受けるから、あまりとらわれるべきではない。以下においては、 $Q = 2, 4, 6$ の各場合をとりあげよう。

表7~9は、集合1, 2, 6の負荷行列である。表7の職業志向性では、 Q の増減はほとんど負荷に影響を与えておらず、2つの合成変量は、「挑戦志向」と「安定志向」と解される。しかしながら、第Ⅱの合成変量に高く負荷する項目1と3の負荷が、 Q が増加するほど低下していく傾向が認められ、このことはライフ・スタイルの第Ⅰの合成変量との相関をあわせて考えると興味深い。一方、表8のライフ・スタイルでは、 Q が増加するにつれて、各因子の寄与、および α 係数が増加している。しかも、 $Q = 4$ と $Q = 6$ の間で、2つの合成変量の意味が、「体制型」と「自己実現型」の組み合わせから、「会社・出世型」と「私的生活享楽型」の組み合わせに変化する。表9の就職先としての魅力度に関しては、 $Q = 2$ と $Q = 6$ の間に多少の違いが見られる。特に、「CBC」の負荷する合成変量が、ⅠからⅡに移動することが目立つ。

このデータでは、 Q の増加に伴う α 係数の低下は、職業志向性の第2合成変量を除き生起していない。

表10は、2次合成変量の負荷行列である。 $Q = 2$ の場合、複数の変数を含む3つの集合の1次合成変量は、2つの2次合成変量に一定の負荷を示すが、その構造は単純ではない。 $Q = 4$ の場合には、職業志向性とライフ・スタイルの2つの1次合成変量が、最初の2つの2次合成

表7 2次合成変量の数を変化させた場合の職業志向性の負荷行列の比較
($q_1 = q_2 = q_6 = 2$ の場合)

項 目	$Q = 2$			$Q = 4$			$Q = 6$		
	I	II	h^2	I	II	h^2	I	II	h^2
1. 安定した会社や勤め先	02	70	49	01	63	40	01	54	29
2. 困難な仕事に挑戦する機会	75	-23	61	74	-22	60	75	-22	61
3. 高い給与やボーナス	09	78	62	08	74	56	08	66	45
4. 仕事上の責任の重さ	74	-17	57	74	18	58	75	-22	60
5. 休日多い・勤務時間短い	01	74	54	01	77	59	01	79	63
6. 能力が試される機会	81	09	67	81	09	66	81	10	67
7. 仕事の気楽さ	-12	62	40	-12	70	50	-11	75	58
8. 仕事を通じ勉強・成長	71	12	51	72	14	53	71	16	52
9. 通勤の便利さ	01	64	41	01	65	43	01	68	46
10. 自力で成し遂げる機会	84	14	72	84	15	72	83	17	72
2 乗 和	298	256	555	298	258	556	298	253	551
α 係 数	73	67		74	68		74	66	

(小数点省略)

表8 2次合成変量の数を変化させた場合のライフ・スタイルの負荷行列の比較
($q_1 = q_2 = q_6 = 2$ の場合)

項 目	$Q = 2$			$Q = 4$			$Q = 6$		
	I	II	h^2	I	II	h^2	I	II	h^2
1. お金持ちになること	74	04	54	79	-03	62	85	-11	73
2. 仕事で成功すること	38	57	47	39	51	42	73	17	56
3. 自分にあった仕事	12	67	46	12	63	41	38	35	27
4. 親友をもつこと	25	64	47	17	73	56	17	73	56
5. 安定した仕事につく	77	18	63	72	29	60	53	51	54
6. 大きな組織で高い地位	75	10	58	75	12	58	75	16	58
7. 好きなことをする暇	41	18	20	45	13	22	27	19	11
8. よい結婚相手・幸福な家庭	47	43	40	40	55	46	24	66	50
9. 世の中の不平等なくす	00	64	41	-04	63	39	-06	69	48
10. 両親や親類のそばに住む	35	23	18	32	28	18	13	52	28
2 乗 和	244	190	434	236	207	443	241	219	460
α 係 数	48	28		48	38		54	47	

(小数点省略)

表9 2次合成変量の数を変化させた場合の魅力度の負荷行列の比較
($q_1 = q_2 = q_6 = 2$ の場合)

項 目	$Q = 2$			$Q = 4$			$Q = 6$		
	I	II	h^2	I	II	h^2	I	II	h^2
1. 愛知県庁	-05	81	66	-02	66	44	-02	66	44
2. 中部電力	57	13	34	50	30	34	50	29	34
3. C B C	43	17	21	16	45	23	16	45	23
4. 東海銀行	19	64	44	-01	74	55	00	73	54
5. トヨタ	64	15	43	71	06	50	72	05	52
6. 日本電装	68	-12	48	83	-11	70	84	-11	71
7. ブラザー	69	06	48	69	11	49	67	12	47
8. 松阪屋	12	65	44	12	67	46	11	67	47
2 乗 和	190	157	347	195	175	370	195	175	370
α 係 数	45	33		54	47		54	47	

(小数点省略)

表10 2次負荷行列
($q_1 = q_2 = q_6 = 2$ の場合)

		$Q = 2$		$Q = 4$				$Q = 6$					
		1	2	1	2	3	4	1	2	3	4	5	6
1. 職業志向性	I	-70	67	-15	96	10	-13	-03	-02	03	100	00	-02
2.	II	57	54	82	07	16	-10	09	06	-01	01	98	18
3. ライフ・	I	71	42	77	-10	02	23	08	-06	04	02	21	96
4. スタイル	II	-05	65	39	52	-19	18	00	96	-05	14	09	06
5. 理系(1)一文系(2)		29	05	23	-12	59	-22	64	02	-03	-06	10	06
6. 男(1)一女(2)		-11	04	-09	10	27	-09	23	-03	-04	15	02	-18
7. 東海地方居住希望		18	02	09	-07	-01	26	04	36	15	-13	-01	-07
8. 就職先として	I	07	22	-08	14	08	90	-09	13	96	02	00	07
9. の魅力度	II	42	06	15	-08	88	25	93	08	14	-05	01	07
2 乗 和		163	141	153	126	126	111	136	109	108	106	102	101
α 係 数		76	77	77	72	48	57	48	45	52	69	53	46

(小数点省略)

表11 1次合成変量間相関行列 ($Q = 2$)

		1	2	3	4	5	6	7	8	9
1. 職業志向性	I	100	00	-18	33	-09	09	-09	07	-13
2.	II	00	100	49	16	16	-01	04	04	18
3. ライフ・	I	-18	49	100	00	14	-06	13	11	23
4. スタイル	II	33	16	00	100	01	-03	02	08	-04
5. 理系(1)一文系(2)		-08	16	14	01	100	13	-05	-18	40
6. 男(1)一女(2)		09	-01	-06	-03	13	100	01	03	00
7. 東海地方居住希望		-09	04	13	02	-05	01	100	10	07
8. 就職先として	I	07	04	11	08	18	03	10	100	00
9. の魅力度	II	-13	18	23	-04	40	00	07	00	100

(小数点省略)

表12 1次合成変量間相関行列 ($Q = 4$)

		1	2	3	4	5	6	7	8	9
1. 職業志向性	I	100	00	-18	29	-09	09	-09	03	-07
2.	II	00	100	48	15	16	-02	03	01	14
3. ライフ・	I	-18	48	100	00	15	-06	11	09	19
4. スタイル	II	29	15	00	100	01	-04	05	09	00
5. 理系(1)一文系(2)		-09	16	15	01	100	13	-05	-23	37
6. 男(1)一女(2)		09	-02	-06	-04	13	100	01	-02	05
7. 東海地方居住希望		-09	03	11	05	-05	01	100	10	08
8. 就職先として	I	03	01	09	09	-23	-02	10	100	00
9. の魅力度	II	-07	14	19	00	37	05	08	00	100

(小数点省略)

表13 1次合成変量間相関行列 ($Q = 6$)

		1	2	3	4	5	6	7	8	9
1. 職業志向性	I	100	00	00	11	-09	09	-09	03	-07
2.	II	00	100	37	16	15	-02	01	-00	12
3. ライフ・	I	00	37	100	00	11	-07	02	08	13
4. スタイル	II	11	16	00	100	05	-03	16	10	07
5. 理系(1)一文系(2)		-09	15	11	05	100	13	-05	-23	37
6. 男(1)一女(2)		09	-02	-07	-03	13	100	01	-02	05
7. 東海地方居住希望		-09	01	02	16	-05	01	100	10	08
8. 就職先として	I	03	-00	08	10	-23	-02	10	100	00
9. の魅力度	II	-07	12	13	07	37	05	08	00	100

(小数点省略)

変量にそれぞれ1つずつ高い負荷を示し、魅力度の1次合成変量は、残りの2つの2次合成変量に高く負荷する形となり、ほぼ単純構造を呈するようになる。 $Q = 6$ では、6つの1次合成変量は、6つの2次合成変量にほぼ完全に別れて負荷する。

表11～13は、1次合成変量間の相関行列である。ここでは、2次合成変量の負荷に対応した形の相関パターンが得られている。 Q が小さい場合ほど、一般に相関は高い。

この実際のデータの分析では、2次合成変量の α 係数の和の最大化基準は、だいたい予期したような方向で機能していると言えよう。

また、同一のデータを、2次合成変量の分散の和を最大化する基準で分析した、村上(1987b)の結果と比較してみると、先に述べたように、 $Q = 2$ の場合にはまったく同一の解となっており、また $Q = 4$ の場合もほとんど同一の解を生み出している。このデータに関する限り、基準の違いはほとんど問題にならないと言ってよい。また、現在までのところ、この2つの方法の結果が決定的に異なるような実際のデータは見出されていない。

4. 残された問題

Cronbach (1970) のbandwidth-fidelity dilemmaは、心理学的な量を測定するため尺度を構成するための、すなわち、本研究の文脈で言えば、質問項目、あるいは問題項目に対する被験者の反応の1次結合としての合成変量の定義するための、単一の基準は存在しないことを意味している。本研究では、2つの視点を同時に考慮しながら、尺度構成をしていくための方法を提案し、それはある程度適切に機能することが示されたと思われる。

しかしながら、なお検討されなければならない問題も多い。この節では、それらの問題を幾つか取り上げて論じておこう。

まず第一に、この方法では、1つの変数集合から導かれる1次合成変量が、組み合わせられる他の変数集合が何であるかによって、異なったものになる点に若干奇異の念がもたれるかも知れない。しかしながら、これは通常の因子分析の性質をそのまま受け継いだものと見なすべきである。なぜなら、通常の因子分析において、変数の一部が削除されたり、別のものと入れ換えられた場合、あるいは、変数が追加された場合、他の部分のデータはまったく不変なままにとどまっているにもかかわらず、それらの変数の因子負荷は変化し、したがって、因子得点(あるいは主成分得点)は違ったものとなるからである。

因子分析を適切に利用するためには、出来る限り必然性のある変数の選択をすることが必要であろう。それと同様に、本研究の方法を用いる場合にも、分析の対象となる変数集合を、必然性のあるものに限定する必要があることになる。しかし、それには当然限界がある。問題としている心理学的な構成概念の性質がある程度明らかでない限り、他のいかなる構成概念との相関を検討すべきか、ということは明らかにならない。したがって、この方法の結果は多くの場合、研究の途上における過渡的なものとみなされるべきである。(この点は、尺度構成のために用いられる因子分析の結果と同様である。)

第二に、階層的主成分分析の方法が、定式化I'～IV'の4通り存在することになったが、そのうちのどれを選択すべきか、という問題がある。1節において述べたように、基本的なアルゴリズムにおいては、定式化I'とIII'は同一であり、定式化II', IV'は異なっている。しかしながら、3節の結果に見られるように、定式化I'とII'は、実際のデータでは比較的似た結果をもたらす。したがって、主要な選択は、データにモデルを当てはめる、定式化III'をとるか、合成変量を定義する形の定式化I'

またはⅡ'をとるか、ということになる。(定式化Ⅳ'は現在のところ十分な検討がなされていないのでここでは論じない。)

この点について、ある程度見当をつけるために、3節で分析したデータを、定式化Ⅲ'の形で分析した。ただし、 $Q=2$ の場合に限って結果を示す。表14~16は集合1, 2, 6の因子負荷行列(式(1-12)の A_k)、表17は単一項目からなる集合3~5の負荷行列、表18は、2次負荷行列(式(1-12)の C_k)、表19は、 $F_k = C_k F$ で定義される各集合の主成分得点間相関行列である。

これらの結果を、3節の結果と比較してみると、職業志向性の因子負荷については、ほぼ同一であり、ライフ・スタイルについては負荷の絶対値がかなり小さいもの

の、定性的にはほぼ同一の結果が得られている。しかし、就職先としての魅力度については、すべての値がきわめて小さく、ほとんど解釈不能な結果となっている。また、単一の変数からなる集合の負荷も専攻(文系一理系)以外、大変小さい。

表16 定式化Ⅲ'による就職先としての魅力度の負荷行列
($q_1 = q_2 = q_6 = Q = 2$ の場合)

項 目	I	II	h^2
1. 愛知県庁	04	36	13
2. 中部電力	13	06	02
3. C B C	12	07	02
4. 東海銀行	12	24	07
5. トヨタ	15	17	05
6. 日本電装	08	-01	01
7. ブラザー	16	10	04
8. 松阪屋	04	25	06
2 乗 和	30	10	40

(小数点省略)

表14 定式化Ⅲ'による職業志向性の負荷行列
($q_1 = q_2 = q_6 = Q = 2$ の場合)

項 目	I	II	h^2
1. 安定した会社や勤め先	-00	62	39
2. 困難な仕事に挑戦する機会	73	-22	57
3. 高い給与とボーナス	07	67	46
4. 仕事上の責任の重さ	69	-12	50
5. 休日多い・勤務時間短い	-00	53	28
6. 能力が試される機会	80	05	64
7. 仕事の気楽さ	-15	40	18
8. 仕事を通じ勉強・成長	68	07	47
9. 通勤の便利さ	-00	48	23
10. 自力で成し遂げる機会	82	07	68
2 乗 和	282	159	441

(小数点省略)

表17 定式化Ⅲ'による1項目だけの集合の負荷

	I	h^2
理系(1)一文系(2)	29	08
男(1)一女(2)	11	01
東海地方居住希望	18	03

(小数点省略)

表15 定式化Ⅲ'によるライフ・スタイルの負荷行列
($q_1 = q_2 = q_6 = Q = 2$ の場合)

項 目	I	II	h^2
1. お金持ちになること	59	22	39
2. 仕事で成功すること	34	50	37
3. 自分にあった仕事	19	44	23
4. 親友をもつこと	31	43	28
5. 安定した仕事につく	67	19	49
6. 大きな組織で高い地位	55	28	38
7. 好きなことをする暇	37	22	19
8. よい結婚相手・幸福な家庭	44	34	32
9. 世の中の不平等なくす	11	37	15
10. 両親や親類のそばに住む	30	20	13
2 乗 和	178	114	292

(小数点省略)

表18 定式化Ⅲ'による2次負荷行列
($q_1 = q_2 = q_6 = Q = 2$ の場合)

		$Q = 2$	
		1	2
1. 職業志向性	I	-57	82
2.	II	82	57
3. ライフ・	I	100	10
4. スタイル	II	-10	100
5. 理系(1)一文系(2)		100	-02
6. 男(1)一女(2)		-88	48
7. 東海地方居住希望		100	-07
8. 就職先として	I	20	98
9. の魅力度	II	98	-20
2 乗 和		577	323

(小数点省略)

表19 定式化Ⅲ'による主成分間相関行列 ($Q = 2$)

		1	2	3	4	5	6	7	8	9
1. 職業志向性	I	100	00	-48	88	-58	89	-62	70	-72
2.	II	00	100	88	48	81	-45	78	72	70
3. ライフ・	I	-48	88	100	00	99	-83	99	29	96
4. スタイル	II	88	48	00	100	-12	56	-17	96	29
5. 理系(1)一文系(2)		-58	81	99	-12	100	-89	100	18	98
6. 男(1)一女(2)		89	-45	-83	56	-89	100	-91	30	-96
7. 東海地方居住希望		-62	78	99	-17	100	-91	100	13	99
8. 就職先として	I	70	72	29	96	18	30	13	100	00
9. の魅力度	II	-72	70	96	29	98	-96	99	00	100

(小数点省略)

一方、2次因子負荷はきわめて大きな値が目立ち、また構造は3節のものと比べてはるかに単純である。また因子得点間相関行列には、非常に大きな値が目立つ。このことは、村上(1987a)において指摘したように、この定式化Ⅲでは、集合ごとの因子得点が全集合の変数の1次結合として定義されているからである。特にこのデータでは、ほとんど負荷をもたない魅力度や、性別、東海地方居住希望のような集合の因子得点間の相関が著しく高いことははなはだ奇異であり、通常の解釈方法から考えれば、やはり誤った結論に導かれる恐れがある。したがって、このような場合、定式化Ⅲ'の結果を表示することは不適切であるとせざるを得ない。

しかしながら、データの説明力という点から見たとき、この結果は、少なくとも $Q = 2$ の場合に魅力度の変数はほとんど説明されていないということを示しており、この観点からすれば、定式化Ⅰ'、Ⅱ'の結果も、見掛けほど良いものではないことを物語っているとも考えられる。つまり、 $Q = 2$ の場合、魅力度の集合から導かれる合成変量は、余りにも bandwidth の方向に偏り過ぎた結果であると考えられることもできる。

以上の点から、結果の解釈、あるいは分析結果にもとづく尺度構成の指針としては、一応定式化Ⅰ'、Ⅱ'が適切であるが、定式化Ⅲの結果も適宜参考にされる必要がある、とこの段階では結論しておきたい。この点は今後の検討がなお必要である。

最後に、定式化Ⅱ'のアルゴリズムの問題について触れておこう。村上(1987a)で記述されたアルゴリズムは、村上(1987b)における定式化Ⅰ'のアルゴリズムに比べて収束がかなり遅く、また Q が大きい場合には、時に不適解を生ずることがある。1節において述べたように、(1-18)式において常に最大固有値に対応する固有ベクトルが解であれば、定式化Ⅰ'と同様な2つの固有値問

題の反復という形の解法が定式化できた。実際、すべての集合について、最適解が最大固有値に対応するものであることがわかっている場合について、その形のアルゴリズムを適用すれば、収束ははるかに速いことが確かめられている。しかし、一般には解がそのような性質をもつことは確認できないから、それとは違ったより頑健で高速なアルゴリズムを開発する余地は残されている。

文 献

- Cronbach, L. J. 1970 *Essentials of Psychological Testing, 3rd ed.* New York : Harper & Row.
- Keller, J. B. 1962 Factorization of matrices by least squares. *Biometrika*, **49**, 239-242.
- 村上 隆 1986 多集合データのための階層的な主成分分析 名古屋大学教育学部紀要—教育心理学科—, **33**, 35-48.
- 村上 隆 1987a 2次合成変量のアルファ係数の和を最大化する多集合データの階層的な主成分分析 経営行動科学, **2**, 37-47.
- 村上 隆 1987b 複数の変数集合の主成分と正準変量 名古屋大学教育学部紀要—教育心理学科—(印刷中)
- 斎藤和志・村上 隆・若林 満 1986 就職先としての魅力度と職業志向性及びライフ・スタイルとの関連 経営行動科学, **1**, 101-113.

(1987年10月2日受稿, 1987年10月9日受理)