

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 3 月 31 日現在

機関番号：14401

研究種目：基盤研究(C)

研究期間：2008～2010

課題番号：20500256

研究課題名（和文） 入出力 3 相データの主成分分析法に関する研究

研究課題名（英文） Studies on principal component analysis for three-way data of inputs and outputs

研究代表者

足立 浩平 (ADACHI KOHEI)

大阪大学・大学院人間科学研究科・教授

研究者番号：60299055

研究成果の概要（和文）：入力×出力×個体の三相データ配列を分析して、入力と出力に介在するコンポーネントを抽出する複数の主成分分析法に関する研究を行った。この研究によって、複数の主成分分析法の中から最適な分析法を選択する手法を完成して、選択されたモデルの解を有意義な単純解に変換する方法を開発した。

研究成果の概要（英文）：The principal component analysis (PCA) procedures were studied which can be applied to three-way data arrays of inputs by outputs by individuals for finding the components intermediating between inputs and outputs. These studies allowed us to find the procedure to find the optimal model among the PCA models considered and develop the method for optimally transforming the PCA solutions to useful ones of simple structures.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008 年度	1,200,000	360,000	1,560,000
2009 年度	1,100,000	330,000	1,430,000
2010 年度	1,000,000	300,000	1,300,000
2011 年度	0	0	0
年度			
総計	3,300,000	990,000	4,290,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：行動計量分析

1. 研究開始当初の背景

(1) 3 相データに適用できる PCA(主成分分析法)には PCA-SEP(SEP), PCA-SUP(SUP), TUCKER2(T2), TUCKER3(T3), CANDECOMP-PARAFAC (CP)があり、それらの間には $SEP \succ SUP \succ T2 \succ T3 \succ CP$ という階層関係がある。ここで、 $E \succ F$ は E 法が F 法より制約が弱いことを表す。

(2) 入出力 3 相データ $\{x_{ijk}; i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K\}$ とは、 I 種の入力(刺激)が個

体(被験体) k に惹起させる出力(反応)を J 種の指標で測定した結果であり、 x_{ijk} は入力 i に対する個体 k の出力 j (指標 j における反応)の大きさを表す。

(3) データへ最も高い適合を示す SEP のモデルは、 x_{ijk} を (i, j) 要素とするデータ行列を \mathbf{X}_k と表し、任意の P 次正則行列を \mathbf{S}_k と表すと、 $\mathbf{X}_k \equiv \mathbf{A}_k \mathbf{B}_k' = \mathbf{Y}_k \mathbf{S}_k \mathbf{S}_k^{-1} \mathbf{Z}_k'$ ($k = 1, \dots, K$) と書け、 \mathbf{S}_k を特定しなければ、個体間比較が可能な $(I \times P)$ の成分行列 $\mathbf{A}_k = \mathbf{Y}_k \mathbf{S}_k$ と $(J \times P)$ の \mathbf{B}_k

$= \mathbf{Z}_k \mathbf{S}_k^{-1}$ を特定できない致命的不定性がある。この問題を解決するため、研究代表者は、 $\phi = \alpha \sum_k \|\mathbf{A}_k - \mathbf{A}\|^2 + \beta \sum_k \|\mathbf{B}_k - \mathbf{B}\|^2 = \alpha \sum_k \|\mathbf{Y}_k \mathbf{S}_k - \mathbf{A}\|^2 + \beta \sum_k \|\mathbf{Z}_k \mathbf{S}_k^{-1} - \mathbf{B}\|^2$ を最小にする $\mathbf{A}, \mathbf{B}, \mathbf{S}_1, \dots, \mathbf{S}_K$ を求める同時プロクラステス分析(JPA)を開発している。ただし、 $\alpha = (\mathbf{IKP})^{-1}, \beta = (\mathbf{JKP})^{-1}$ である。

2. 研究の目的

(1) 入出力 3 相データの主成分分析法(PCA)P の統合プログラムを作成して、手法間の比較を行う。

(2) 解に正則変換の不定性のある Tucker2, Tucker3 の解の変換法を研究開発する。

(3) モデル選択規準の研究を通して、特定データに対して、最適な PCA モデルを選択する方法を完成させる。

3. 研究の方法

(1) PCA-SEP(SEP), PCA-SUP(SUP), TUCKER2(T2), TUCKER3(T3), CANDECOMP-PARAFAC(CP)のそれぞれについて、解全体の適合度だけでなく、各変数(入力・出力・個体)の分散説明率を表す指標を整備すると同時に、後者の短所に関わる解の安定性(stability)の指標を考えた上で、SEP > SUP > T2 > T3 > CP の階層関係を踏まえて、5 種の PCA それぞれをサブルーチンとして包含した単一のコンピュータ・プログラムを作成する。これを実データへ適用して、各手法の長短が現れる諸側面をみきわめる。

(2) $\mathbf{X}_k \equiv \mathbf{A}_k \mathbf{B}_k' = \mathbf{Y}_k \mathbf{S}_k \mathbf{Z}_k^{-1}$ のモデルで表せる SEP は \mathbf{A}_k と \mathbf{B}_k を平等に扱うため、その解から、入力成分 \mathbf{A}_k の個体差と出力成分 \mathbf{B}_k の個体差の大小比較ができる。一方、T2 のモデルは、 $\mathbf{X}_k \equiv \mathbf{A} \mathbf{H}_k \mathbf{B}'$ ($k=1, \dots, K$) と表せ、T3 は \mathbf{H}_k に制約を加えた方法であるが、成分が個体間で共通の $\mathbf{A}(I \times P), \mathbf{B}(J \times Q)$ に制約され、 $\mathbf{H}_k(P \times Q)$ が個体差を説明する。この比較を T2, T3 の解から行うには、 \mathbf{H}_k を $\mathbf{H}_k = \mathbf{M}_k \mathbf{N}_k'$ と分解して、 $\mathbf{A}_k = \mathbf{A} \mathbf{M}_k, \mathbf{B}_k = \mathbf{B} \mathbf{N}_k'$ を求める必要があるが、 $\mathbf{A} \mathbf{H}_k \mathbf{B}' = \mathbf{A} \mathbf{M}_k \mathbf{N}_k' \mathbf{B}' = \mathbf{A} \mathbf{M}_k \mathbf{S}_k \mathbf{Z}_k^{-1} \mathbf{N}_k' \mathbf{B}'$ でわかるように、SEP の場合と同様の不定性が生じる。そこで、T2, T3 のための JPA(解の同時プロクラステス変換法)の開発を考える。

(3) 所与のデータに対する適合度と節約性を考慮した PCA のモデル選択法は、Ceulemans & Kiers(2006)によって提案されているが、SEP はこの選択法では考慮されていない。そこで、SEP も含めてモデル選択ができるように上記手法を修正する。さらに、他の選択規準も研究する。

4. 研究成果

(1) 同じ入出力 3 相データに対して、SEP, および、個別でない主成分分析(SUP), 三相主成分分析(T2, T3 と CP)のそれぞれを実行できるコンピュータ・プログラムを完成させ、手法間の比較を行い、以下の成果を得た。まず、 $I \times J \times K$ の三相配列(x_{ijk})をそれぞれ $JK \times I, KI \times J$ および $IJ \times K$ と二相化した行列の各列のデータ平方和が、モデル部の平方和と誤差平方和に分割できることを見出して、データ列平方和に対するモデル部平方和の比として、各入力・出力・個体の分散説明率を定義する指標を考案した。次に、折半されたデータに対する SEP の解どうし、および、データからのブートストラップ・サンプルに対する SEP の解の間の近接性を定義して、解の安定性(Stability)を求める方法を考案した。さらに、入出力 3 相の形をとる印象評定データにプログラムを適用した結果、SEP はデータへの適合が高く、分散説明率も高いが、安定性では劣ることが見出された。ただし、分析結果は妥当な解釈が可能であった。

(2) x_{ijk} を (i,j) 要素とする個体 k の入出力データ行列を \mathbf{X}_k で表すと、TUCKER2(T2)の解は、 $\mathbf{X}_k \equiv \mathbf{A} \mathbf{H}_k \mathbf{B}' = \mathbf{A} \mathbf{S} \mathbf{S}^{-1} \mathbf{H}_k \mathbf{T}'^{-1} \mathbf{T}' \mathbf{B}' = \mathbf{A}^* \mathbf{H}_k^* \mathbf{B}'^*$ と表せ(\mathbf{S} と \mathbf{T} は任意の正則行列)、正則変換に関する不定性を持つが、成分行列 $\mathbf{A}^* = \mathbf{A} \mathbf{S}$ と $\mathbf{B}'^* = \mathbf{B} \mathbf{T}$ および核行列 $\mathbf{H}_k^* = \mathbf{S}^{-1} \mathbf{H}_k \mathbf{T}'^{-1}$ が単純構造を達成し、さらに、 $\mathbf{H}_k^* = \mathbf{S}^{-1} \mathbf{H}_k \mathbf{T}'^{-1}$ によって表現される個体差が最小になるような正則行列 \mathbf{S} と \mathbf{T} を求めるために、目的関数 $f = \alpha \|\mathbf{A} \mathbf{S} - \mathbf{A}_T\|^2 + \beta \|\mathbf{B} \mathbf{T} - \mathbf{B}_T\|^2 + \gamma \sum_k \|\mathbf{S}^{-1} \mathbf{H}_k \mathbf{T}'^{-1} - \mathbf{H}_k\|^2$ を最小化する $\mathbf{S}, \mathbf{T}, \mathbf{H}$ を求める交互最小二乗アルゴリズムを開発した。ここで、 α, β, γ は所与の正の定数、 \mathbf{A}_T と \mathbf{B}_T は単純構造を持つターゲット行列、そして、 \mathbf{H} は各個体の核行列 \mathbf{H}_k^* を集約したいわば個体間共通の核行列である。このアルゴリズムでは、ターゲット行列選定のために、バリマックス回転後の成分行列を強調化するプロマックス法を用い、 \mathbf{H} の推定には、一定数の 0 要素以外は未知要素の行列とする制約つきシンプリマックス法と呼べる方法を採用した。そして、正則行列 \mathbf{S} と \mathbf{T} の推定には、同時プロクラステス分析の特異値分解リパラメトリゼーション法を用いた。シミュレーションによって開発手法が真のパラメータをよく再現できることを確認した上で、実際の入出力 3 相データへの適用を通して、合理的な解釈ができる解が得られることを示した。

(3) Ceulemans & Kiers(2006)の数値凸包モデル選択法に基づいて、SEP, SUP, T2, T3, CP の中から最適なモデルを選択する方法を完成させた。なお、CP の解の退化の頻発は、2010

年度に見いだされ、これの原因解明と対処の追求が、2011年度への研究繰越の契機となったが、原因は、CPのモデルのパラメータ過少にあり、それへの対処法として、パラメータ行列の条件数を不等式制約する方法を開発して、その有用性を例証できた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計6件)

① Adachi, K., Three-way Tucker2 component analysis solutions of stimuli \times responses \times individuals data with simple structure and the fewest core differences. *Psychometrika*, 査読有, vol. 76, 2011, pp. 285-305.
DOI: 10.1007/S11336-011-9208-6

② 足立浩平. 最小二乗置換によるサイズ固定クラスタリング. データ分析の理論と応用, 査読有, 1巻, 2011, pp. 11-22.

③ Adachi, K., Constrained principal component analysis of standardized data for biplots with unit-length variable vectors. *Advances in Data Analysis and Classification*, 査読有, vol. 5, 2011, pp. 23-36.

DOI: 10.1007/s11634-010-0081-4

④ Adachi, K., Joint Procrustes Analysis for Simultaneous Nonsingular Transformation of Component Score and Loading Matrices. *Psychometrika*, 査読有, vol. 74, 2009, pp. 667-683.

DOI: 10.1007/S11336-009-9131-2

⑤ Adachi, K. Joint Procrustes Analysis: Minimizing a LS Function of a Nonsingular Matrix and Its Inverse Matrix with SVD Reparameterization. *The Proceedings of International Association of Statistical Computing (IASC) 2008 (CD-ROM版)*, 査読無, 2008, pp. 32-39.

⑥ Adachi, K. & Murakami, T., Correspondence Analysis as Reduced Rank Approximation of Quantified Indicator Matrices. *The Proceedings of International Association of Statistical Computing (IASC) 2008 (CD-ROM版)*, 査読無, 2008, pp. 40-49.

[学会発表] (計21件)

① Adachi, K. Nonsingular Transformation of Tucker2 Solutions for Representing Stimulus-

Response Relationships by Sparse Networks, *International Meeting of Psychometric Society*, 2011, 2011年7月19日, 香港教育学院.

② Adachi, K. & Murakami, T., Three Kinds of Hierarchical Relations among PCA, Nonmetric PCA, and Multiple Correspondence Analysis, *International Meeting of Psychometric Society*, 2011, 2011年7月19日, 香港教育学院.

③ Adachi, K. Singular Value Reparameterization with Its Applications to Least Squares Computations, *The 3rd International Conference on COMPUTING & STATISTICS*, 2010年12月11日, ロンドン大学.

④ Adachi, K. A Factor Analysis Procedure for Jointly Estimating Factors, Loadings, and Specificities, 第10回中国日本統計シンポジウム, 2010年10月16日, 西南财经大学.

⑤ Adachi, K. A constrained condition-number LS algorithm with its applications to reverse component analysis and generalized oblique Procrustes rotation, *19th Symposium of IASC on Computational Statistics (COMPSAT 2010)*, 2010年8月26日, パリ国際会議場.

⑥ Adachi, K. A New Formulation of Factor Analysis with Fixed Common and Specific Factors, *The 75th Meeting of Psychometric Society*, 2010年7月9日, ジョージア大学.

⑦ Adachi, K. Singular Value Reparameterization with Its Applications to Rank Preserving Computation, Procrustes Transformation, and Factor Analysis, *The 19th International Workshop on Matrices and Statistics*, 2010年6月6日, 上海金融学院.

⑧ 足立浩平. Reduced Condition Number Regression and its Alternating Least Squares Algorithm with Singular Value Reparameterization, 科学研究費(基盤研究(B)18300094)シンポジウム「多変量データ解析法への計算機統計学・行動計量学的アプローチの新展開」, 2009年12月4日, 大阪大学.

⑨ 足立浩平. 入出力データの三相主成分分析 — Tucker2の成分と核行列の同時変換 —, *日本計算機統計学会第23回シンポジウム*, 2009年11月7日, 札幌学院大学.

⑩ 足立浩平. パーセプトロンとしての三相主成分分析 — 刺激・主成分・反応のネットワーク表現 —, 2009年度統計関連学会連合大会, 2009年9月8日, 同志社大学.

⑬ Adachi, K. Permutimin: Factor Rotation to Specified Simple Structure with Least Squares Permutation, International Meeting of Psychometric Society, 2009, 2009年7月22日, ケンブリッジ大学.

⑭ Adachi, K. Joint Procrustes Algorithm for Nonsingular and Compensatory Transformation of Component and Loading Matrices. The 1st Institute of Mathematical Statistics Asia Pacific Rim Meetings, 2009年6月29日, ソウル国立大学.

⑮ 足立浩平. 最小二乗パーミュテーションの因子回転とサイズ固定クラスタリングへの応用, 日本計算機統計学会第23回大会, 2009年5月17日, 福岡女子大学.

⑯ Adachi, K. Joint Procrustes Analysis: Minimizing a LS Function of a Nonsingular Matrix and Its Inverse Matrix with SVD Reparameterization, International Association of Statistical Computing, 2008, 2008年12月8日, パシフィコ横浜.

⑰ Adachi, K. & Murakami, T. Multiple Correspondence Analysis as Reduced Rank Approximation of Quantified Indicator Matrices, International Association of Statistical Computing, 2008, 2008年12月6日, パシフィコ横浜.

⑱ 足立浩平. 心理統計学と実験心理学のねじれた関係, 日本心理学会第72回大会, 2008年9月21日, 北海道大学.

⑲ 足立浩平. 負荷行列の列ではなく行のノルムを1と制約した主成分分析の性質 —アダマール積 vech 演算子を用いた考察—, 2008年度統計関連学会連合大会, 2008年9月8日, 慶応義塾大学.

⑳ 足立浩平. 多変量データ解析法のタクソノミー, 日本行動計量学会 第36回大会, 2008年9月4日, 成蹊大学.

㉑ Adachi, K. Joint Procrustes Component Analysis for Exploring the Perceptual and Semantic Structures in Three-Way Semantic Differential Data, International Congress of Psychology, 2008, 2008年7月24日, ICC Berlin.

㉒ Adachi, K. Joint Procrustes Analysis with Constrained Simplimax Rotation: Nonsingular Transformation of Component Score and

Loading Matrices Toward Simple Structure. The 32nd Annual Conference of the German Classification Society, 2008年7月16日, HELMUT-SCHMIDT-UNIVERSITÄT.

(21) Adachi, K. Joint Procrustes Analysis in the S-O-R Framework. International Meeting of Psychometric Society, 2008, 2008年7月1日, University of New Hampshire.

[図書] (計2件)

① 足立浩平・村上 隆. 非計量多変量解析法—主成分分析から多重対応分析へ—. 朝倉書店, 171.

② 足立浩平. 心理学フロンティア (I 章 15節: 多次元尺度法, 16節: 構造方程式モデリング) 新曜社, 8.

[その他]

ホームページ等

<http://www.dma.jim.osaka-u.ac.jp/view?1=ja&u=5163&k=%E8%B6%B3%E7%AB%8B%E6%B5%A9%E5%B9%B3&kc=1&sm=keyword&sl=ja&sp=1>

6. 研究組織

(1) 研究代表者

足立 浩平 (ADACHI KOHEI)

大阪大学・大学院人間科学研究科・教授

研究者番号: 60299055