様式Ｃ－１９

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２５年　５月２９日現在

機関番号：35302
研究種目：基盤研究（C）
研究期間：2010～2012
課題番号：22500265
研究課題名（和文）大規模・高次元データの発見的情報表現と効率的情報縮約およびその計算環境の研究
研究課題名（英文）Heuristic representation and effective reduction for large scaled and high dimensional information and its computational environments
研究代表者
森　裕一（Mori, Yuichi）
岡山理科大学・総合情報学部・教授
研究者番号：80230085

研究成果の概要（和文）：個体数や項目数が膨大であるデータに対して，次元縮約などでの発見的な考察を可能とする可視化手法を提案し，効率的な計算アルゴリズムや計算環境の開発を行った。具体的には，タッチパネルでの直接操作や色を表現手段に追加したグラフアプリとアソシエーション分析の大量ルールから有用な情報を見つけるインタラクティブツールを開発した。また，非計量主成分分析の計算の加速化に関して，計算コストが高い変数選択問題への適用や複数の手法を組み合わせた新加速化手法を提案し，十分な成果が得られることを確認した。

研究成果の概要（英文）：We discussed heuristic representation methods and effective computational algorithms for large scaled and high dimensional data in information visualization, data reduction, and variable selection problems. We developed a web-based graphics application which has an interactive interface using finger gestures on the touch-screen, a colored face graph which can represent a part of multiple variables by color, and an interactive tool to find useful rules in association analysis. We also studied acceleration techniques in non-linear principal component analysis by applying an acceleration algorithm to qualitative variable selection problem which needs high computation cost and proposed a two-stage acceleration algorithm to get estimations more speedy. We confirmed that all tools work well for heuristic and intuitive observations and that our proposed acceleration algorithms provide good performances in numerical experiments.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２０１０年度	1,300,000	390,000	1,690,000
２０１１年度	1,200,000	360,000	1,560,000
２０１２年度	800,000	240,000	1,040,000
年度			
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：多変量解析
大規模データ，可視化，次元縮約，加速化計算，インタラクティブ・インタフェース，変数選択，非計量主成分分析

１．研究開始当初の背景

Web，ゲノム，マーケティングなどの分野において，観測対象とする個体の数や観測観点である変数の数が膨大であるデータを解析しなければならない場面が増えてきた。

これら大規模データに対しては，データの

様相をいかに的確にとらえるかが重要となってくる。ここに，データを視覚的にとらえる技術の必要性が存在し，ブラッシングやグランドツアーといったダイナミックグラフィックスの手法が考案されてきた。しかし，大規模なデータほど，試行錯誤しながらの考察が有効になるが，従来の手法ではカバーできない大きさのデータや量的変数と質的変数を一緒に処理する場面では，データ入力やグラフ出力において，よりインタラクティブなインタフェースが望まれていた。その手段も，マウス操作だけでなく，iPhone に見られるようなタッチパネル上の指の動きに反応するインタフェースがあれば，考察の手段も広がり，データの可視化や解析結果の考察に新たな可能性が示唆されていた。

また，データ自体も複雑になってきており，観測対象がさまざまな属性で区分されている場合や，観測変数に量的なものと質的なものが混在している場合，あるいは，分析のための条件が膨大になるなど，設定された複雑さを考慮した分析が必要になっている。データの複雑さに対しては，制約つき主成分分析やあらゆる尺度に対応した主成分分析などが提案されていた。しかし，データのもつすべての性質は，これらの手法だけで分析し切れないことから，先行研究の成果を基にしながら，データの複雑さに応じた次元縮約手法をその場面ごとに提案していく必要が出てきていた。また，アソシエーション分析などでは，膨大なルールから有用なものを効率的に見つけていくために，より洗練された情報発見手段が求められていた。

以上の可視化のための計算や次元縮約のための計算は，データが大規模であるがゆえに，その計算量が大きい。質的な変数があると，計算方法によっては反復計算が入るため，さらに計算時間を要する。また，変数選択のように組み合わせの数だけ計算を行う必要がある場合も計算量は膨大になる。これらに対して，研究代表者らは計算の加速化の研究を進めており，その成果を適用するなどして，新しい加速化アルゴリズムを提案することが重要となっていた。

計算環境については，統計解析環境である R を用い，その関数などを R パッケージとして広く公開することが主流となっていた。

２．研究の目的

観測個体数と項目数の一方あるいは両方が膨大で，既存の手法による分析には限界があるデータ，および分析のための考察対象が組み合わせ的に膨大となる場合に対して，データや条件の発見的な考察を可能とする可視化手法やデータがもつさまざまな性質を的確に反映できる次元縮約手法を提案し，データの様相をより正確にとらえられるようにする。また，次元縮約手法での変数選択場面など，より膨大な計算量や反復回数が必要な問題に対して，迅速で効率的に結果を得ることができる計算アルゴリズムについて，検討・開発を行うことを本研究の目的とする。

３．研究の方法

研究は，次の 5 つに分けて行った。

(1) 先行研究等の情報収集と分析・整理

大規模データの可視化と高次元データの次元縮約について，先行研究や公開されているソフトウェアなどの分析・整理を行う。特に，高次元データと複雑な構造をもつデータの次元縮約について最新の研究成果を集める。これらを通して，インタラクティブ手法の動向と，データサイズによらないインタラクティブな可視化手法の可能性と限界を明確にしていく。

(2) 可視化手法と次元縮約手法および変数選択手法の検討

(1)と同様，大規模データについてはインタラクティブな可視化のための技法，次元縮約については複雑さの程度を加味した情報処理の方法を整理する。また，これまでになかった複数の変数（尺度混在データ）を同時処理する多変量手法について検討を行う。

(3) インタラクティブシステムの構築

提案する各手法に指ジャスチャーを含んだインタラクティブ機能を導入し，発見的，試行錯誤的な考察ができるようにする。具体的には，可視化ではデータおよびデータグループの選択，可視化と次元縮約ではパラメータ指定，アソシエーション分析では情報の取捨選択に対話的操作を導入していく。同時に，可視化と次元縮約の融合も考えていく。

(4) 計算効率の検討

これまでの加速化の研究を再検討するとともに，未検討の加速化アルゴリズムを試し，さらには，複数の加速化アルゴリズムを組み合わせ，データの大きさや反復計算の頻出による膨大な計算に対する新しい加速化手法を，実用性を考慮して，提案していく。

(5) 計算環境の提供

確立した手法のツール化（R 関数，Excel アドイン）を行い，研究用に公開していく。

４．研究成果

３の(1)(2)は先行研究の整理であるので，(3)〜(5)の成果について，以下に報告する。

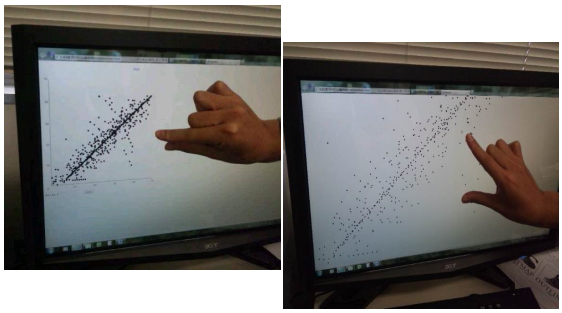
(1) インタラクティブ・インタフェース（タッチパネルによるインタラクティブグラフ）

SVG と Java Script を使用し，タッチパネル上でグラフを直観的に操作することができ，拡大しても線や点のサイズは一定のまま相対位置関係を広げることのできるアプリケーションを Web 上に開発した。

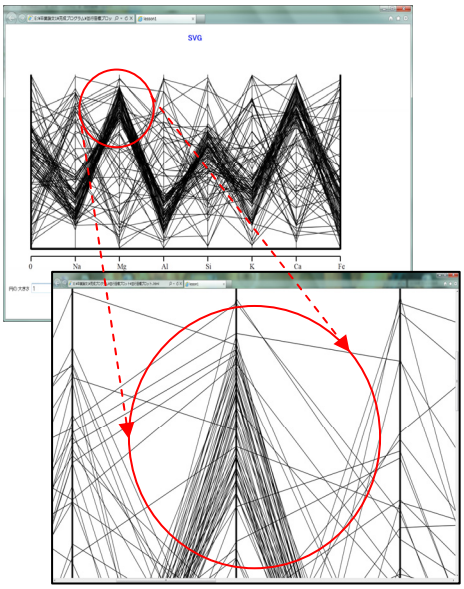
拡大方法は，タッチパネルの場合は，2 本

の指を遠ざける方向にスライドさせる（図 1）。縮小させる場合は，2 本の指を近付ける方向にスライドさせる。点の大きさだけを変える場合は，マウスホイールを上に回すと拡大され，下に回すと縮小される。

動作例として，散布図（図 1）と並行座標プロット（図 2）を示す。いずれも拡大すると，点や線の大きさや太さが変わることなく，間隔が疎になるので，細かい分布の様子が鮮明になる。指での操作ができることで，直接的，直観的な操作も可能となっている。



（図 1）散布図（指ジャスチャーによる拡大）



（図 2）並行座標プロット（拡大の様子）

(2) 情報の発見的・効率的表現 1 （カラー顔グラフ）

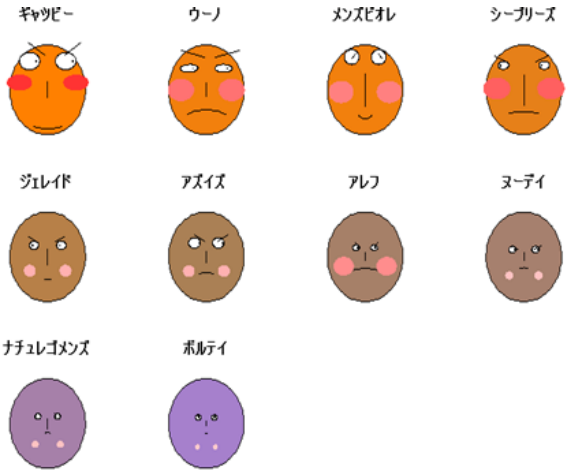
多変量データを一目で認識できる顔グラフの情報表現力を発展させ，従来，直線，孤，楕円などの図形のみであった表現手段に，「塗り分け」を加え，より多くの情報伝達を可能にしようとした。すなわち，図形によるモノクロ表現から，顔色などを加えた「カラー顔グラフ」の作成を行うものである。

開発では，R 上で動作する青木氏の顔グラフ関数（青木，http://aoki2.si.gunma-u.ac.jp/R/face.html）を改変し，顔色，頬の色を追加した。顔色は，数値が大きいほど赤に，低いほど青に，頬の色は，値が大きくなれば赤に，低くなれば白に近づくようにし，数値が高いほど活気ある印象を与える仕様にした。

男性用化粧品の商品イメージデータに対して，「知っている」を「顔色」に，「過去使ったことがある」を「頬の色」に対応させ，認知度を色で認識できるようにした（他の変数の割り当ては表 1 の通り）。図 3 がこの割り当てによるカラー顔グラフである。大きく 3 つのクラスターが観察でき，このうち，ギャツビー，ウーノ，メンズビオレ，シーブリーズは，血色が良く，元気な印象を受けることより，認知度の高さなどが色を含めた表情から考察できる効果が得られている。

（表 1）データの割り当て

番号	変　数	顔の部分への割り当て
x[1]	「知っている」	OP の長さ
X[2]	「知っている」	X 軸と OP の角度
x[3]	「知っている」	顔色
x[4]	「過去使ったことがある」	頬の色
x[5]	「安心感あり」	口の曲率の一部
x[6]	「効果高そう」	鼻の長さ
x[7]	「豊富」	口の位置
x[8]	「安心感あり」	口の曲率
x[9]	「安心感あり」	口の幅
x[10]	「価格手ごろ」	目の位置
x[11]	「使っている」	目の中心の離れ具合
x[12]	「パッケイジデザインよい」	目の傾き
X[13]	「パッケイジデザインよい」	目の楕円の離心率
X[14]	「価格手ごろ」	目の幅の半分
x[15]	「独自性あり」	ひとみの位置
x[16]	「宣伝積極的」	目から眉の位置
x[17]	「宣伝積極的」	眉の傾き
x[18]	「若者向け」	眉の長さ



（図 3）化粧品イメージデータのカラー顔グラフ

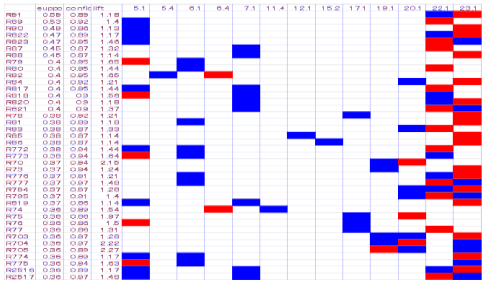
(3) 情報の発見的・効率的表現 2 （アソシエーションルールの可視化）

アソシエーション分析とは，たとえば，消費者がある商品（アイテム）a を買った場合に別の商品 b を買う傾向がどれだけあるかを調べるものであり，アイテム間の何らかの組み合わせの規則（アソシエーションルール）の中から有益な情報を見つけ出すものである。ルール「a ならば b」を「a⇒b」と書くとき，「⇒」の左辺を条件部，右辺を結論部とよぶ。アイテムの数を n とすると，$\sum_{i=1}^{n} {}_{n}\mathrm{P}_{i}$ だけのルールを調べる必要がある。

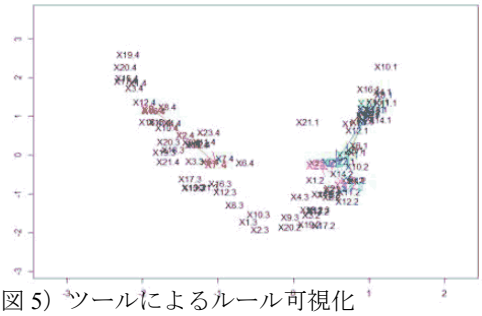
このルールの可視化ツールを R, MS Excel, RExcel を用いて開発した。

まず，Excel 上のデータに対して，アソシエーション分析を行い，色の異なるブロックを用いたルールの可視化を行う。図 4 は，行がルール，列がアイテムを示すテーブルで，ある頻度以上で出現する条件部のアイテムを青，結論部のアイテムを赤で示すようになっている。たとえば，抽出されたルールをアソシエーション分析の指標の 1 つである支持度の高い順に並べ替えて，この色分けを行うと，上位ルールの中で，条件部に頻出するアイテム，結論部に頻出するアイテムが発見でき，有用なルールの絞り込みが視覚的に行える。また，色のついた部分の対応を見ることで，どのアイテム間でルールが構成されやすいかを効率的に考察できる。

図 5 は，カテゴリスコア散布図上でルールの可視化を行うもので，テーブルによって見つけたルールの関係を対話的に図示できる。



（図 4）テーブルによるルールの可視化



（図 5）ツールによるルール可視化

(4) 計算の加速化

研究代表者らは，交互最小二乗法（Alternating Least Squares，ALS）を用いた非計量主成分分析の計算の加速化を，vector ε 法（vε 法）を用いて行ってきた。本研究では，この加速化を計算パワーがより必要となる場面に適用することや，より速い収束を目指した新たな加速化手法を提案することを行った。

① PRINCIPALS によるパラメータ推定

n 個×p 変数の標準化されたデータ行列 $\mathbf{X}$ は，$r(\le p)$ 個の主成分の得点行列 $\mathbf{Z}$ $(n\times r)$ と $\mathbf{X}$ の分散共分散行列の固有ベクトル行列 $\mathbf{A}$ $(p\times r)$ を用いて，$\hat{\mathbf{X}}=\mathbf{Z}\mathbf{A}^{\mathsf{T}}$ と表される。$\mathbf{X}$ の最適変換されたデータを $\mathbf{X}^*$ で表すと，r 個の主成分によって最もよく表現される $\mathbf{X}^*$ を求めることは，$\theta=\mathrm{tr}(\mathbf{X}^*-\hat{\mathbf{X}})^{\mathsf{T}}(\mathbf{X}^*-\hat{\mathbf{X}})$ を最小化するデータパラメータ $\mathbf{X}^*$ とモデルパラメータ $\mathbf{Z}$，$\mathbf{A}$ を求める最小二乗推定問題となる。

非計量データに適用される PRINCIPALS は，初期値 $\mathbf{X}^{*(0)}$ が与えられたもとで，データの最適変換 $\mathbf{X}^*$ と，これに最もよく一致するモデルパラメータ $\mathbf{Z}$，$\mathbf{A}$ を交互に推定する。すなわち，次の 2 つのステップを交互に繰り返す。

［モデルパラメータ推定ステップ］

固有値問題

$$\left[\frac{\mathbf{X}^{*(t)\mathsf{T}}\mathbf{X}^{*(t)}}{n}\right]\mathbf{A}=\mathbf{A}\mathbf{D}_r \qquad (1)$$

を解き，$\mathbf{A}^{(t)}$ を求める。$\mathbf{D}_r$ は対角要素が固有値の $r\times r$ 対角行列，$\mathbf{A}^{\mathsf{T}}\mathbf{A}=\mathbf{I}_r$ である。

［データパラメータ推定ステップ］

先のステップの $\mathbf{A}^{(t)}$ から $\hat{\mathbf{X}}^{(t)}(=\mathbf{Z}^{(t)}\mathbf{A}^{(t)\mathsf{T}})$ を求め，$\hat{\mathbf{X}}$ を固定し，最小二乗基準のもとで，$\mathbf{X}^{*(t+1)}=\arg\min_{\mathbf{X}^*}\mathrm{tr}(\mathbf{X}^*-\hat{\mathbf{X}}^{(t)})^{\mathsf{T}}(\mathbf{X}^*-\hat{\mathbf{X}}^{(t)})$

となる $\mathbf{X}^{*(t+1)}$ を求め，列ごとに基準化する。これを繰り返し，$\mathbf{X}^*$ の収束列 $\left\{\mathbf{X}^{*(t)}\right\}_{t\ge0}$ ができ，先の θ により収束したと判定された $\mathbf{X}^*$ によって得られるパラメータが解となる。なお，添字(t)は，その収束列の t 番目であることを示し，$t=\infty$ のとき，理論的に推定値となる。

② vε 法による ALS の加速化

vε 法は，1 次収束する反復法から生成されるベクトル列に対し，その収束を加速する方法で，ベクトル列を $\left\{\mathbf{Y}^{(t)}\right\}_{t\ge0}$ で表すとき，vε 法は，以下の式で加速列 $\left\{\dot{\mathbf{Y}}^{(t)}\right\}_{t\ge0}$ を生成する。

$$\dot{\mathbf{Y}}^{(t-1)}=\mathbf{Y}^{(t)}+\left[\left[\Delta\mathbf{Y}^{(t)}\right]^{-1}-\left[\Delta\mathbf{Y}^{(t-1)}\right]^{-1}\right]^{-1} \qquad (2)$$

ここで，$\Delta\mathbf{Y}^{(t)}=\mathbf{Y}^{(t+1)}-\mathbf{Y}^{(t)}$ で，ベクトルの逆行列は $\mathbf{Y}^{-1}=\mathbf{Y}/\langle\mathbf{Y},\mathbf{Y}\rangle$，$\langle\mathbf{Y},\mathbf{Y}\rangle$ はベクトルの内積を表す。

もとの列 $\left\{\mathbf{Y}^{(t)}\right\}_{t\ge0}$ に停留点 $\mathbf{Y}^{\infty}$ が存在するとき，得られる加速列は $\mathbf{Y}^{\infty}$ に収束し，$\left\{\mathbf{Y}^{(t)}\right\}_{t\ge0}$ の収束より速いという性質をもつ。

これを用いると，PRINCIPALS の［データパラメータ推定ステップ］で生成される $\left\{\mathbf{X}^{*(t)}\right\}_{t\ge0}$ を加速化でき，その手順は，次の 2 つの Step にまとめられ，この加速列 $\left\{\dot{\mathbf{X}}^{*(t)}\right\}_{t\ge0}$ の収束先が，$\mathbf{X}^*$ の推定値となる。

［Step1：PRINCIPALS ステップ］

$\mathbf{X}^{*(t)}$ から $\mathbf{Z}^{(t)}$ と $\mathbf{A}^{(t)}$ を計算し，$\mathbf{X}^{*(t+1)}$ を推定する。

［Step2：vε 加速ステップ］

部分列 $\left\{\mathbf{X}^{*(t-1)},\mathbf{X}^{*(t)},\mathbf{X}^{*(t+1)}\right\}$ から(2)式を用

いて，vε 加速列 $\left\{\dot{\mathbf{X}}^{*(t)}\right\}_{t\geq 0}$ を生成する。δ を収束判定指標として，

$$\left\|\mathrm{vec}\left(\dot{\mathbf{X}}^{*(t-1)}-\dot{\mathbf{X}}^{*(t-2)}\right)\right\|^2<\delta$$

により，収束をチェックする。

この方法を vε-PRINCIPALS とする。

③拡張主成分分析による変数選択の加速化

vε 法による加速化を，組み合わせの数だけ計算が必要となる変数選択問題に適用する。ここでは，主成分分析における変数選択として，その計算過程に変数選択を自然に含む拡張主成分分析（Modified PCA，M.PCA）を用いる。M.PCA は，標準化された n 個体×p 変数のデータ行列 $\mathbf{X}$ を，$n\times q$ の部分行列 $\mathbf{X}_{V_1}$ と $n\times(q-1)$ の $\mathbf{X}_{V_2}$ の 2 つに分解したとき，$\mathbf{X}$ 全体を最もよく再現する r 個の主成分を $\mathbf{X}_{V_1}$ から抽出しようというものである。そのために，$\mathbf{X}=(\mathbf{X}_{V_1},\mathbf{X}_{V_2})$ の共分散行列を $\mathbf{S}=\begin{pmatrix}\mathbf{S}_{11} & \mathbf{S}_{12}\\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{pmatrix}$ として，一般化固有値問題

$$[(\mathbf{S}_{11}^2+\mathbf{S}_{12}\mathbf{S}_{21})-\mathbf{D}_r\mathbf{S}_{11}]\mathbf{A}=\mathbf{0} \qquad (3)$$

を解く。

この M.PCA による質的データの変数選択に対して，vε-PRINCIPALS が PRINCIPALS よりどれだけ速く収束するかを変数減少法（Back）と変数増加法（For）それぞれで，反復回数と計算時間により評価する。すべての計算は R で行い，δ を 10^{-8} とし，CPU 時間は **`proc.time`** 関数で計測した。なお，M.PCA を対象とした PRINCIPALS は，(1)式の固有値問題を(3)式に置き換えればよい。

表 2 は，100 個体，5 カテゴリーの 10 変数の人工データに対して，主成分数 r を 3 として，PRINCIPALS と vε-PRINCIPALS のそれぞれで，$q=10$ から $q=3$ での最適な q 変数を求める Back と For を実行した結果である。

（表 2）質的データの M.PCA における変数選択にかかった反復回数と CPU 時間

(a) 変数減少法（Back）

	PRINCIPALS		vε-PRINCIPALS		Speed-up	
q	反復回数	CPU 時間	反復回数	CPU 時間	反復回数	CPU 時間
10	141	1.70	48	0.68	2.94	2.49
9	1363	17.40	438	6.64	3.11	2.62
8	1620	20.19	400	5.98	4.05	3.37
7	1348	16.81	309	4.80	4.36	3.50
6	4542	53.72	869	11.26	5.23	4.77
5	13735	159.72	2949	35.70	4.66	4.47
4	41759	482.59	12521	148.13	3.34	3.26
3	124	1.98	44	1.06	2.82	1.86
総計	64491	752.40	17530	213.57	3.68	3.52

(b) 変数増加法（For）

	PRINCIPALS		vε-PRINCIPALS		Speed-up	
q	反復回数	CPU 時間	反復回数	CPU 時間	反復回数	CPU 時間
3	4382	67.11	1442	33.54	3.04	2.00
4	154743	1786.70	26091	308.33	5.93	5.79
5	13123	152.72	3198	38.61	4.10	3.96
6	3989	47.02	1143	14.24	3.49	3.30
7	1264	15.27	300	4.14	4.21	3.69
8	340	4.38	108	1.70	3.15	2.58
9	267	3.42	75	1.17	3.56	2.93
10	141	1.73	48	0.68	2.94	2.54
総計	178249	2078.33	32405	402.40	5.50	5.16

Speed-up の列は，PRINCIPALS の計測値を vε-PRINCIPALS の計測値で割ったものである。これより，vε-PRINCIPALS の方が，反復回数も計算時間も少なく，反復回数で 3 倍から 5 倍，計算時間で 2 倍から 5 倍の加速化ができていることがわかる。

④ vε 法と GM 法による 2 段階加速法

GM 法は，以下の式で，ベクトル列 $\left\{\mathbf{Y}^{(t)}\right\}_{t\geq 0}$ の加速列 $\left\{\ddot{\mathbf{Y}}^{(t)}\right\}_{t\geq 0}$ を生成する。

$$\ddot{\mathbf{Y}}^{(t-1)}=\mathbf{Y}^{(t+1)}-\frac{\left\langle\Delta\mathbf{Y}^{(t)},\Delta\mathbf{Y}^{(t)}\right\rangle}{\left\langle\Delta\mathbf{Y}^{(t)},\Delta^2\mathbf{Y}^{(t)}\right\rangle}\Delta\mathbf{Y}^{(t+1)} \qquad (4)$$

ここで，$\Delta^2\mathbf{Y}^{(t)}=\Delta\mathbf{Y}^{(t)}-\Delta\mathbf{Y}^{(t-1)}$ である。

この GM 法を vε 法で生成した加速列 $\left\{\dot{\mathbf{X}}^{*(t)}\right\}_{t\geq 0}$ に対して適用する。すなわち，先の Step1，2 に続けて，次の Step3 を実行する。

［Step3：GM 加速ステップ］

vε 加速列によって生成される加速列の部分列 $\left\{\dot{\mathbf{X}}^{*(t-1)},\dot{\mathbf{X}}^{*(t)},\dot{\mathbf{X}}^{*(t+1)},\dot{\mathbf{X}}^{*(t+2)}\right\}$ から(4)式を用いて，GM 加速列 $\left\{\ddot{\mathbf{X}}^{*(t)}\right\}_{t\geq 0}$ を生成する。

この段階で，δ による収束判定を行う。

この 2 段階加速法を vεGM-PRINCIPALS とする。この vεGM-PRINCIPALS の Step3 で得られた $\left\{\ddot{\mathbf{X}}^{*(t)}\right\}_{t\geq 0}$ の最終値を $\mathbf{X}^*$ の推定値とし，この値を用いて $\mathbf{Z}$ と $\mathbf{A}$ の推定値を求める。

vεGM-PRINCIPALS を軽傷意識障害データ（87 個体×23 変数，4 または 2 カテゴリー）と授業アンケートデータ（56 個体×14 変数，5 カテゴリー）に適用し，その計測値を PRINCIPALS と vε 法による加速結果とともに表 3 に示す。データサイズや変数数およびカテゴリー数の違いによる詳細な検討は必要であるが，2 段階加速化の効果が見て取れる。

（表 3）vεGM-PRINCIPALS の結果（() 内は加速率）

データと主成分数		PRINCIPALS	
		反復回数	計算秒数
軽傷意識データ	$r=2$	53	1.05
	$r=3$	105	1.94
授業アンケートデータ	$r=2$	270	2.27
	$r=3$	489	3.98

データと主成分数		vε 法	
		反復回数	計算秒数
軽傷意識データ	$r=2$	15 (3.5)	0.43 (2.4)
	$r=3$	31 (3.4)	0.68 (2.9)
授業アンケートデータ	$r=2$	75 (3.6)	0.69 (3.3)
	$r=3$	189 (2.6)	1.62 (2.5)

データと主成分数		vεGM 法	
		反復回数	計算秒数
軽傷意識データ	$r=2$	12 (4.4)	0.42 (2.5)
	$r=3$	26 (4.0)	0.66 (2.9)
授業アンケートデータ	$r=2$	74 (3.6)	0.70 (3.2)
	$r=3$	191 (2.6)	1.64 (2.4)

(5) 計算ツールの開発

(1)でJava Scriptによる Webアプリケーション，(2)で R 関数，(3)で MS Excel のマクロ（R と RExcel が必要），(4)で R 関数を作成した。末尾のホームページで整理・公開していく。

(6) 考　察

個体数や項目数が膨大であるデータを念頭に，発見的な考察を可能とする可視化手法を検討・提案するとともに，そのための効率的な計算アルゴリズムの開発を行った。具体的には，タッチパネルを直接操作するインタフェースをもったグラフアプリケーション，多次元データのグラフ表現に「色」の要素を追加した可視化手法，アソシエーション分析で対象となる大量のルールから有用な情報を発見するための対話的可視化ツールを開発した。また，非計量主成分分析の計算の加速化を，計算回数が膨大となる変数選択問題へ適用するとともに，既存の加速化手法を組み合わせたより高速な加速化手法を提案し，数値実験により，十分な成果（3〜5 倍の加速）が得られることを確認した。

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 3 件）

① Kuroda, M., Mori, Y., Iizuka, M. and Sakakihara,M. Acceleration of convergence of the alternating least squares algorithm for nonlinear principal components analysis, Sanguansat, P. (Ed.), Principal Component Analysis, 査読有, 2012, 129-144, http://www.intechopen.com/books/principal-component-analysis/acceleration-of-convergence-of-the-alternating-least-squares-algorithm-for-nonlinear-principal-compo, InTech Publications.

② Kuroda, M., Izuka, M., Mori, Y. and Sakakihara, M. Principal Components Based on a Subset of Qualitative Variables and Its Accelerated Computational Algorithm, Proceedings of the 58th World Statistics Congress (ISI2011), 査読有, 2011, 全 9 ページ.

③ Kuroda, M., Mori, Y., Izuka, M. and Sakakihara, M. Acceleration of the alternating least squares algorithm for principal components analysis, Computational Statistics and Data Analysis, 査読有, 55(1), 2011, 143-153, DOI: 10.1016/j.csda.2010.06.001, Elsevier Science.

〔学会発表〕（計 7 件）

① 松居俊宏（飯塚誠也）．アソシエーションルールの可視化について，北海道大学情報基盤センター共同利用研究集会「大規模・高次元データの発見的情報表現と効率的計算」, 2013 年 2 月 9 日, 北海道大学（北海道）.

② 森 裕一（黒田正博・榊原道夫・飯塚誠也）．非計量主成分分析の加速化ー 実データへの適用，日本行動計量学会第 40 回大会 2012 年 9 月 13 日〜16 日, 新潟県立大学（新潟県）.

③ Mori, Y. (Kuroda, M., Iizuka, M. and Sakakihara, M.). Two-stage acceleration for non-linear PCA, COMPSTAT 2012, 2012 年 8 月 27-31 日, Amathus Beach Hotel（キプロス）.

④ 森 裕一（榊原道夫・黒田正博・飯塚誠也）. Graves-Morris による交互最小二乗法の加速化, 日本計算機統計学会第 26 回大会, 2012 年 05 月 12-13 日, 香川県社会福祉総合センター（香川県）.

⑤ Kuroda, M.(Mori, Y., Izuka, M. and Sakakihara, M.). Variable Selection in Principal Components Analysis of Qualitative Data Using the Accelerated ALS Algorithm, The 7th IASC-ARS Conference, 2011 年 12 月 16 日, Academia Sinica（台北市，台湾）.

⑥ Kuroda, M. (Mori, Y., Izuka, M. and Sakakihara, M.). Principal components based on a subset of qualitative variables and its accelerated computational algorithm, The 58th World Statistics Congress of the International Statistical Institute (ISI2011) , 招待, 2011 年 8 月 22 日, Convention Centre Dublin（ダブリン, アイルランド）.

⑦ Mori, Y. (Kuroda, M., Iizuka, M. and Sakakihara, M.). Improvement of acceleration of the ALS algorithm using the vector ε algorithm, The 19th International Conference on Computational Statistics, 2010 年 8 月 26 日, フランス国立工芸院（パリ，フランス）.

〔その他〕

ホームページ等

http://mo161.soci.ous.ac.jp/vasmm/

６．研究組織

(1) 研究代表者

森　裕一（Mori, Yuichi）

岡山理科大学・総合情報学部・教授

研究者番号：80230085

(2) 研究分担者

飯塚　誠也（Iizuka, Masaya）

岡山大学・環境学研究科・講師

研究者番号：60322236

黒田　正博（Kuroda, Masahiro）

岡山理科大学・総合情報学部・准教授

研究者番号：90279042

(3) 連携研究者

足立　浩平（Adachi, Kohei）

大阪大学・人間科学研究科・教授

研究者番号：60299055

中野　純司（Nakano, Junji）

統計数理研究所・統計計算開発センター・教授

研究者番号：60136281