様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　３年　６月１７日現在

機関番号：３５３０２
研究種目：基盤研究(C)（一般）
研究期間：2017～2020
課題番号：１７Ｋ０００６３

研究課題名（和文）尺度混在・複雑化大規模データの特徴抽出と対話的情報把握の研究

研究課題名（英文）Information extraction and its interactive system for large-scale mixed and complex data

研究代表者
森　裕一（Mori, Yuichi）

岡山理科大学・経営学部・教授

研究者番号：８０２３００８５

交付決定額（研究期間全体）：（直接経費）　3,600,000円

研究成果の概要（和文）：尺度混在や複雑性などを考慮し，情報縮約と分類の同時推定および変数処理により，隠れた構造や特徴を取り出せる手法の提案と，その結果を効率的に得る環境を提供することを目的として研究を行った。その結果，Reduced k-means法やPartial Least Squaresでの質的データの処理，変数選択手法の非計量主成分分析などへの拡張，交互最小二乗法を用いる反復計算の加速化の非計量主成分分析，非計量因子分析，Fuzzy c-meansクラスタリングへの適用，およびこれらの結果を用いた対話的な情報把握が可能となった。

研究成果の学術的意義や社会的意義
一般的なデータサイズや質的データに特化された既存の手法に対して，尺度の統一的な扱いを適用し，大規模性を次元縮約と変数選択により軽減させ，情報縮約と分類の同時推定により，対象の考察のための有用な情報を得ようとしたところに，学術的な特色がある。また，数値解析の分野で提案されている加速化法を取り入れ，計算時間の面からも大規模性を克服しようとしたところに独創性がある。これらの手法により，これまでの手法では観察しづらかった知見が手早く得られるようになり，マーケティング分野やデータマイニングの適用場面で強力なツールになりうる点に社会的な意義が見出せる。

研究成果の概要（英文）：For mixed measurement levels and complexity in data, we proposed methods to extract the latent structures and features in the data, and to efficiently obtain the computational results, using simultaneous estimation of information reduction and clustering and variable selection. The methods enable us to deal with qualitative data in reduced k-means clustering and partial least squares method, to select a reasonable subset in non-linear principal component analysis, and to obtain the results quickly in non-linear principal component analysis, non-linear factor analysis and fuzzy c-means clustering.

研究分野：計算機統計学

キーワード：クラスタリング　次元縮約　非計量多変量解析　数量化　計算効率　加速化アルゴリズム

１．研究開始当初の背景

計画的に集められたデータと異なり，一般に，ビッグデータは，冗長な情報や雑音など，無駄な情報が混在する。また，観測対象がさまざまな属性でグループ化されていたり，変数の観測尺度も質的変数と量的変数が混在していたりするなど，解析者の意図しない複雑な構造をもっていることが多い。これらに加え，計算時間の問題は常についてまわる。

これらに対して，従来，グループ化された観測値には制約つき主成分分析や Lasso／Ridge，質的データには交互最小二乗法による最適尺度法などが利用されてきたが，これらが提案された時期からも，データの複雑さや尺度混在を前提としたものとはなっていなかった。計算時間においても，並列処理システムのようなハードウェア的な解決策は講じられるが，適用する手法は従来のものであり，計算そのものを効率化したり，冗長性・複雑性に対応した処理を試行錯誤的に把握したりする工夫が積極的に指向されていたわけではなかった。

一方で，たとえば，マーケティングの分野では，膨大な顧客データから購買動向を効率的に把握し，顧客を分類して，群ごとあるいは群間の特徴を考察して販売戦略を立てることの重要性が増大してきていた。レビューサイト，SNS，人の行動を記録したデータなど，意図せず収集されたデータも容易に手に入るようになり，消費者の行動等の特徴をつかみ，それらへ素早く対応していくことが必要になっていた。すなわち，分類・クラスタリングを分析・考察のキーとして，単純な大規模データ処理にとどまれず，データの冗長性や複雑性への対応や試行錯誤的なデータ観察手法が求められていたといえる。

これらのことから，大規模データの処理において，より積極的に，i) データの複雑性の処理，ii) 情報損失を最小限に抑えた大規模さの軽減 を行い，その上で，iii) データの観測手順（具体的には，分類やクラスタリング）の提供 を行っていく必要性がみてとれた。さらに，隠れた情報を発見的にとらえたり，さまざまな結果をすぐに考察できるように，iv) 対話的な可視化ツール や v) 効率的な計算アルゴリズム の開発が求められていた。

２．研究の目的

１に述べたような背景から，情報の損失を最大限に抑えて，冗長性，複雑性，大規模性をわれわれが容易に把握できる規模にまで縮約した上で，分類とクラスタリングを中心に，既存の手法の適用も含めた分析が実行できるデータ処理の環境を実現することを目的とする。具体的には，最適尺度法の多変量解析への適用と，次元縮約法および変数選択法により複雑性と大規模性を軽減し，構造に隠れた情報の抽出を行い，計算には，加速化アルゴリズムを積極的に適用して，大規模データの分析結果を迅速に得る工夫をすることを考えていく。

３．研究の方法

１の i)～v)の研究を次の年次計画で行う。i)の複雑性を処理するための既存手法の評価とその拡張の検討は 1 年目から行う。ii)の次元縮約と変数選択は，それらの手法の整理を 1 年目から行い，2 年目に，iii)の情報縮約と分類の融合手法に関する各種手法の評価の進行にともない，両者を加味した新しい手法開発に取りかかる。iv)の対話的情報表現については，2 年目の計算環境の構築とともに取り組んでいく。v)の加速化については，2 年目の手法開発と iv)の構築と並行して行う。内容としては，次の 6 つに分けて，研究を行う。

(1) 先行研究の情報収集と分析・整理

各種多変量手法における質的変数の処理手法や情報縮約と分類の同時推定手法およびソフトウェアに関する情報を入手し，分析・整理するとともに，各手法の性能とソフトウェアの解析を行い，本研究への利用可能性と改良点を明らかにする。また，対象となる場面や事例を収集し，求められている処理等を整理する。

(2) 尺度混在データの処理方法の検討

交互最小二乗法（ALS，Alternating Least Squares）を用いた最適尺度法（Young et al., 1978）による質的データの数量化を尺度混在データ中の質的変数に適用して全変数を量的データに変換後に各種多変量手法を実行する手順は，すでに，われわれの研究で確立できている。この手順の評価を主成分分析において行うとともに，本研究の対象とする大規模データやクラスタリングでも尺度混在データが扱えるようにする（(3)と(4)で実現）。

(3) 多変数の処理

多変数に対しては 2 つのアプローチをとる。1 つは，元のデータ（質的変数を含む）がもつ情報をできるだけ保持した形で変数の数を減らす変数選択，もう 1 つは，変数の数が多いことを容認した統計的手法で質的データを扱えるようにすることである。前者は，非計量拡張主成分分析を利用した変数選択はすでに開発されているので，その性能を評価することと，新たな選択手法として，テスト理論で用いられる項目反応理論（IRT，Item Response Theory）を意識調査などの質的変数に適用し，その潜在能力を維持する変数選択手法を提案する。後者は，個体数よりも変数の数が多い場合でも分析できる部分的最小二乗法（PLS，Partial Least Squares）を対象として，

最適尺度法を適用して，質的データを扱えるようにする。

(4) 次元縮約手法とクラスタリングの同時推定の検討

変数が多い場合でもその情報をできるだけ保ちながらクラスタリングを行える Reduced k-means Clustering（RKM）で質的データを扱えるようにする。RKM の次元縮約は主成分分析の文脈で行われており，計算そのものにも ALS の手順が用いられているので，非計量主成分分析で用いる数量化の手法を RKM に適用して，尺度混在データの RKM 手法を提案する。

(5) 計算効率の検討

データの大きさや反復計算の頻出による膨大な計算量の軽減を図るため，vector ε アルゴリズム（Wynn，1962）を利用した加速化を非計量主成分分析，非計量拡張主成分分析，非計量因子分析に適用してきた。ここでは，同様の手順を他の多変量手法，具体的には，ALS の手順が用いられている Fazzy c-means Clustering（FCM）に適用し，その性能を確認する。

(6) 対話性の導入

上記の各手順において，発見的，試行錯誤的な考察ができるように，対話的操作ができるインタフェースを考える。計算エンジンとして R を用いるので，R パッケージの Shiny を利用し，データ，パラメータ，出力の表示範囲などを対話的に指定するインタフェースの作成を試みる。

Young, F.W., Takane, Y., de Leeuw, J. (1978). Principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika*, 43, 279-281.

Wynn, P. (1962). Acceleration techniques for iterated vector and matrix problems. *Mathematics of Computation*, 16, 301-322.

４．研究成果

３の 6 項目のうち，(1)(2)は先行研究の整理および既存の成果の検討であるので，(3)～(6)について，報告する。

４.１　項目反応理論を利用した変数選択

試験の問ごとに，正解に 1，間違いに 0 を与え，問の難易度や識別力，回答者の潜在能力を推定する IRT は，肯定的意識を 1，否定的意識を 0 とすれば，意識調査などに適用できる。このとき，質問項目を減らしても，回答者 i の推定した潜在特性値 θ_i が大きく変わらないような項目が特定できれば，質的なデータに対する変数選択が行える。そこで，変数を落とす前と落とした後の θ_i の差が最も小さくなる項目群を見つけるために差の二乗和を指標（d_q）とする方法と，落とす前と後の θ_i の順位が保たれている項目群を見つけるために順位相関係数（ρ_q）を指標とする方法の 2 つで変数選択を試みた。学科のイメージ調査データに適用した結果が表 1，2 である。

これにより，2 値で得られるデータに対して，潜在的な情報を基にした変数の精選ができるようになった。

表 1 変数選択結果（d_q：最小二乗法，変数減少法）

q						$\mathbf{Y}_1$ ｜	$\mathbf{Y}_2$						d_q
12	1	2	3	4	5	6	7	8	9	10	11	12	-
11	1	2	3	4	5	6	7	8	10	11	12	9	1.6077
10	1	2	3	4	5	6	7	8	10	12	9	11	3.8940
9	2	3	4	5	6	7	8	10	12	1	9	11	7.1607
8	2	3	4	5	6	7	10	12	1	8	9	11	10.6146
7	2	3	4	5	6	7	12	1	8	9	10	11	14.9702
6	3	4	5	6	7	12	1	2	8	9	10	11	19.4241
5	3	4	5	6	12	1	2	7	8	9	10	11	25.3764
4	3	4	5	6	1	2	7	8	9	10	11	12	35.0666
3	3	5	6	1	2	4	7	8	9	10	11	12	46.9073
2	3	6	1	2	4	5	7	8	9	10	11	12	68.0698

表 2 変数選択結果（ρ_q：順位相関係数，変数減少法）

q						$\mathbf{Y}_1$ ｜	$\mathbf{Y}_2$						ρ_q
12	1	2	3	4	5	6	7	8	9	10	11	12	-
11	1	2	3	4	5	6	7	8	10	11	12	9	0.9954
10	1	2	3	4	5	6	7	8	10	12	9	11	0.9883
9	1	2	3	4	5	6	7	10	12	8	9	11	0.9773
8	2	3	4	5	6	7	10	12	1	8	9	11	0.9659
7	2	3	4	5	6	7	12	1	8	9	10	11	0.9541
6	2	3	5	6	7	12	1	4	8	9	10	11	0.9361
5	2	3	5	6	12	1	4	7	8	9	10	11	0.9109
4	2	3	6	12	1	4	5	7	8	9	10	11	0.8736
3	3	6	12	1	2	4	5	7	8	9	10	11	0.8239
2	3	6	1	2	4	5	7	8	9	10	11	12	0.7706

４.２　部分的最小二乗法における質的データの扱い

PLS において，説明変数に質的データが含まれている場合の手法を提案する。ここでは，PLS の各種手法のうち，PLS 回帰（PLS-R）に着目する。

ALS による最適尺度法による数量化を PLS に統合することを行うが，数量化の対象とする行列によって，2 つのアプローチが考えられる。1 つは，説明変数行列に対して数量化を行ってから PLS-R を適用する Type1，もう 1 つは，説明変数と目的変数の両方を合わせた行列に数量化を行ってから PLS-R を適用する Type2 である。説明変数内の質的変数の数量化に目的変数の情報を用いるかどうかが 2 つの Type の違いである。

この手法を全国学力・学習状況調査から作成した擬似データに適用する。全国学力・学習状況調査は，教科に関する学力調査と質問紙調査からなり，学力調査は，3 教科で出題された成績（正答数），質問紙調査は，児童や学校長が勉強や学校の様子について 4 件法で回答したものである。このデータで，質問紙の回答を説明変数，成績を目的変数として，100 人分のデータを既知のデータとして PLS-R を行い，正答数のわかっている新たな 10 人の質問紙の回答から，正答数を予測する。

表 3 が最適尺度化されたカテゴリースコア，表 4 が（a）質的変数のまま PLS-R を行った結果，（b）Type1 の結果，（c）Type2 の結果である。これより，質的変数のままで予測を行うよりも，数量化してから，予測をする方がよいことが示唆された。

表 3　最適尺度化されたカテゴリースコア

Type1

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	-0.5673079	-0.5917914	-0.9254558	-0.3292346	-0.8241982	-0.9696602	-0.5327743	-0.8869627	-0.6672534	-1.0231223
2	-0.5873348	-0.5452766	-0.5644464	1.1231075	-0.7612506	-0.6520364	-0.0540313	-0.6835263	-0.8617743	-0.6093932
3	1.1546426	-0.3550446	0.1470265	-0.7938729	1.2848478	0.4110621	1.4205811	0.2788497	0.1974356	0.4631943
4		1.4921125	1.3428757		0.3006009	1.2106344	-0.8337755	1.2916392	1.3315920	1.1693211

Type2

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	-0.5520169	-0.6165673	-1.0627928	-0.2232932	-0.8503634	-0.9780857	-1.1744627	-0.9088254	-1.1741107	-1.2089955
2	-0.6023184	-0.5515667	-0.6026956	1.0927710	-0.8803839	-0.6356417	-0.4858100	-0.6807804	-0.4849876	-0.4129017
3	1.1543353	-0.3196339	0.5874855	-0.8694778	0.9104462	0.3962242	0.7736854	0.3219651	0.7590250	0.6488272
4		1.4877678	1.0780029		0.8203012	1.2175033	0.8865873	1.2676407	0.9000734	0.9730701

表 4　PLS-R による予測結果

児童番号	score	予測結果			scoreとの差		
		(a)	(b)	(c)	(a)	(b)	(c)
101	58	47.40042	54.66702	55.71089	10.59958	3.33298	2.28911
102	35	52.59543	50.8996	50.38544	-17.59543	-15.89960	-15.38544
103	38	50.98628	50.45163	49.70157	-12.98628	-12.45163	-11.70157
104	52	49.08289	50.3378	51.91628	2.91711	1.66220	0.08372
105	46	50.16606	49.62298	46.55941	-4.16606	-3.62298	-0.55941
106	55	50.20145	50.29768	48.90605	4.79855	4.70232	6.09395
107	43	45.23465	47.20462	44.19467	-2.23465	-4.20462	-1.19467
108	52	53.38026	51.05355	51.18096	-1.38026	0.94645	0.81904
109	45	45.75614	44.44456	43.22948	-0.75614	0.55544	1.77052
110	52	54.06716	47.93806	54.49698	-2.06716	4.06194	-2.49698
scoreとの差の二乗和					651.2290665	492.3322989	427.8024174

4.3　尺度混在データに対する次元縮約とクラスタリング

次元縮約とクラスタリングを同時に行う RKM を尺度混在データに適用することを考える。RKM は基本的に量的データに対する手法であるため，RKM の次元縮約部分を非計量主成分分析で置き換えた手法の提案する。

X を n 個体 p 変数の中心化されたデータ行列，k をクラスター数，r（一般に，$k \geq r+1$）を次元数とする。通常の RKM は，U を $n \times k$ のメンバーシップ行列，A を $p \times r$ の負荷行列，$\mathbf{Z} = \mathbf{XA}$ を $n \times r$ の主成分得点行列，F を $k \times r$ のクラスター中心行列とすると，最小化する目的関数は，$f_{\mathrm{RKM}}(\mathbf{U}, \mathbf{F}, \mathbf{A}) = \|\mathbf{X} - \mathbf{UFA}^{\mathrm{T}}\|^2$であり，この U と F を交互最小二乗法により求めるものである。ここに，質的データを最適尺度化して主成分分析を行う非計量主成分分析を統合する。すなわち，最適に尺度変換された元データ X を$\mathbf{X}^*$とすると，尺度混在データに対する RKM の目的関数は，$f_{\mathrm{RKM/NLPCA}}(\mathbf{U}, \mathbf{F}, \mathbf{A}) = \|\mathbf{X}^* - \mathbf{UFA}^{\mathrm{T}}\|^2$となる。具体的な手順は，次のようになる。

[Step 1] 初期化: クラスター数 k, 主成分数 r を決め，初期値として U に乱数を与える。

[Step 2] 数量化: 主成分分析における最適尺度法により $\mathbf{X}^{(t+1)}$ を数量化し，数量化行列 $\mathbf{X}^{*(t)}$ を求める。

[Step 3] クラスタリング: k平均法により，$f_{\mathrm{RKM/NLPCA}}$を最小にする $\mathbf{U}^{(t)}, \mathbf{F}^{(t)}$, $\mathbf{A}^{(t)}$を求める。

[Step 4] 終了判定: t番目と$t+1$番目の$f_{\mathrm{RKM/NLPCA}}$の値に差がなければ終了，さもなくば，$t = t+1$として Step2 へ戻る。

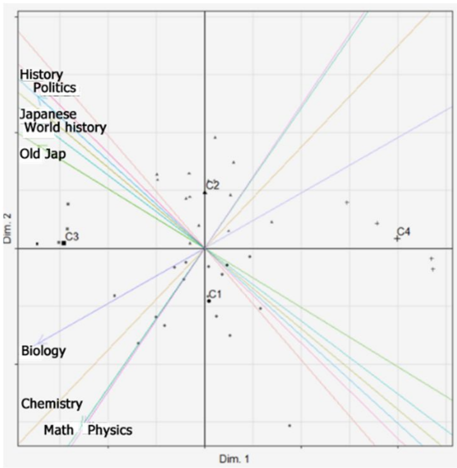

図 1　質的データの RKM 結果

この手順を 9 科目の成績を 5 段階評価したデータに本手法を適用したときの主成分空間におけるクラスター中心，主成分得点，負荷量を布置したのが図 1 である。元の素点データの RKM を基準としたとき，5 段階評価データを量的データとして RKM を実行した結果より付置の再現性などが高くなることがわかった。また，本手法は，GROUPALS (Van Buuren, Heiser, 1989) や Corresponding Clustering (van de Velden, 2017) と同一性をもつことも確認された。

Van Buuren, S. and Heiser, W. J. (1989). Clustering n objects in k groups under optimal scaling of variables, *Psychometrika*, 54, 699-706.

van de Velden M., Iodice D' Enza, A., Palumbo, F. (2017). Cluster correspondence analysis. *Psychometrika*, 82(1), 158-185

4.4　vector ε アルゴリズムによるファジィ c 平均法の計算の加速化

FCM の収束スピードを加速するために，vector ε アルゴリズム（Wynn, 1962）を組み込む。これを vε-FCM とする。メンバーシップ行列Uもしくはクラスター中心行列Aが既知ならば，もう

一方の推定が可能であるので，vε-FCM では，クラスター中心行列Aをベクトル化し，vector ε アルゴリズムによって収束を加速する。したがって，加速されるAをA_εと記すと，目的関数は，$F(U,A)=\|X-UA_\varepsilon\|^2$となり，アルゴリズムは，次のようになる。

［Step1］ チューニングパラメータ m と，クラスター数 K の初期値を与える。
［Step2］ Aの初期値$A^{(0)}$を定める。定めた初期値を代入したAを$\hat{A}^{(0)}=(\hat{c}_1{}^{(0)},\dots,\hat{c}_k{}^{(0)})$とする。
［Step3］ $\hat{A}^{(t)}$を固定し，$\hat{U}^{(t)}=\arg\min_{U\in M}F(U,\hat{A})^{(t)}$を求める。
［Step4］ $\hat{U}$を固定し，$\hat{A}^{(t+1)}=\arg\min_{A}F(\hat{U}^{(t)},A)$を求める。
［Step5］ vector ε アルゴリズムにより，$\{\hat{A}^{(t)},\ \hat{A}^{(t+1)},\ \hat{A}^{(t+2)}\}$を用いて，$\mathrm{vec}\ \hat{A}^{(t)}=A^{(t+1)}+\left[\left[\Delta A^{(t+1)}\right]^{-1}-\left[\Delta A^{(t)}\right]^{-1}\right]^{-1}$を計算し，vector ε 加速列$\left\{\mathrm{vec}\hat{A}_\varepsilon^{(t)}\right\}_{t\ge0}$を生成する。ここで，vec $\hat{A}$は，$\hat{A}$をベクトル化したものである。
［Step6］ $\left\|\mathrm{vec}\hat{A}_\varepsilon^{(t)}-\mathrm{vec}\ \hat{A}_\varepsilon^{(t-1)}\right\|^2<\delta$より収束判定をする。終了していなければ，Step3 へ戻り計算を続ける。

人工データによるシミュレーション（K=5 となる 500×10 の乱数データに対して，m=2 として実行）と実データを用いた数値実験（178 個体×13 変数のワインデータ，K=3（3 等級））の結果を示す。表 5 より，シミュレーションでは，反復回数，CPU 時間ともに 1.5 倍ほど速く収束していることがわかる。また，表 6 の通り，実データにおいても，反復回数，CPU 時間とも 1.2 倍から 2.3 倍の加速ができるという結果を得た。

表 5：乱数データから得られたファジィ c 平均法と v ε ファジィ c 平均法の結果の比較

Stats		Number of iteration		CPU time	
		ファジィc平均法	vεファジィc平均法	ファジィc平均法	vεファジィc平均法
Min.	:	95	56	1.78	0.95
1st Qu.	:	221.8	114.8	3.978	1.958
Median	:	414.5	194	7.315	3.46
Mean	:	842.3	541.1	15.255	9.399
3rd Qu.	:	810.2	401.2	15.275	6.902
Max.	:	8303	9296	157.35	160.02
[Speed Up]		-	1.56	-	1.62

表 6：ワインデータから得られたファジィ c 平均法と v ε ファジィ c 平均法の結果の比較

	ファジィc平均法		vεファジィc平均法		加速率	
Number of cluster	# of iteration	cpu_time	# of iteration	cpu_time	# of iteration	cpu_time
k=3	51	0.2	40	0.16	1.28	1.25
k=4	232	1.24	111	0.58	2.09	2.14
k=5	706	4.65	327	2.14	2.16	2.17
k=6	1016	8.16	458	3.53	2.22	2.31
k=7	410	3.81	229	2.12	1.79	1.80

4.5　対話性の導入

次元縮約とクラスタリングを同時に行う R のパッケージ clustrd の一連の分析過程を対話的に進めるための Shiny アプリケーションを試作した（図 2，3）。スライダーなどでパラメータや出力が制御できるので，考察の手段が増え，発見的・試行錯誤的に検討できる環境が実現できた。

図 2　RKM の Shiny インタフェース（計算結果とプロット）

以上より，質的データへの最適尺度法の適用による数量化を次元縮約と分類の同時推定などの既存の多変量手法に適用することで混在尺度を統一的に扱えるようにしたこと，反復計算をその計算過程にもつ手法の加速化の適用範囲を広げたこと，対話的インタフェースによる計算効率の向上が図れたこと，さらに，変数が多い場合の対処法を検討したことから，複雑性や大規模性を考慮した特徴抽出という本研究の目的はおおむね達成できた。

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 西山ちとせ，片山浩子，森　裕一	４．巻 3
２．論文標題 部分的最小二乗法における質的データの扱い	５．発行年 2021年
３．雑誌名 経営とデータサイエンス	６．最初と最後の頁 26-33
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Masahiro Kuroda, Yuichi Mori, Masaya Iizuka	４．巻 Chapter of Book
２．論文標題 Initial Value Selection for the Alternating Least Squares Algorithm	５．発行年 2020年
３．雑誌名 Advanced Studies in Classification and Data Science	６．最初と最後の頁 227-239
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-981-15-3311-2	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計12件（うち招待講演　2件／うち国際学会　10件）

１．発表者名 森　裕一，吉岡嵩紹，片山浩子，黒田正博
２．発表標題 尺度混在データに対する次元縮約とクラスタリング
３．学会等名 2019年度統計関連学会連合大会
４．発表年 2019年

１．発表者名 Yoshioka, T., Kuroda, M., Mori, Y.
２．発表標題 Reduced K-Means with nonlinear principal component analysis
３．学会等名 The 62nd International Statistical Institute 2019 World Statistics Congress（国際学会）
４．発表年 2019年

1．発表者名 Mori, Y., Katayama, H., Yoshioka, T., Kuroda, M., Iizuka, M.
2．発表標題 Some Applications in Multivariate Methods with Alternating Least Squares
3．学会等名 Data Science, Statistics & Visualisation 2019（国際学会）
4．発表年 2019年

1．発表者名 吉岡嵩紹，黒田正博，森　裕一
2．発表標題 vεアルゴリズムによるファジィc平均法の計算の加速化
3．学会等名 日本計算機統計学会第32回大会
4．発表年 2018年

1．発表者名 Kuorda, M., Mori, Y.
2．発表標題 Speed-up of bootstrap computation of the covariance matrix of MLEs from incomplete data
3．学会等名 The 2nd International Conference on Econometrics and Statistics（国際学会）
4．発表年 2018年

1．発表者名 Yoshioka, M., Kuroda, M., Mori, Y.
2．発表標題 Acceleration of computation for fuzzy c-means clustering
3．学会等名 The 23rd International Conference on Computational Statistics（国際学会）
4．発表年 2018年

１．発表者名 Kuroda, M., Mori, Y.
２．発表標題 Speed-up of bootstrap computation to incomplete data
３．学会等名 IASC-ARS 25th Anniversary Conference & CASC 2nd Annual Conference（国際学会）
４．発表年 2018年

１．発表者名 Yoshioka, M., Kuroda, M., Mori, Y.
２．発表標題 Computational efficiency for fuzzy clustering
３．学会等名 The 11th International Conference of the ERCIM WG on Computational and Methodological Statistics（招待講演）（国際学会）
４．発表年 2018年

１．発表者名 Mori, Y., Iizuka, M., Kuorda, M.
２．発表標題 Variable selection for mixed measurement level data in dimension reduction methods and its computation
３．学会等名 The 1st International Conference on Econometrics and Statistics (EcoSta 2017)（国際学会）
４．発表年 2017年

１．発表者名 Kuroda, M., Mori, Y., Iizuka, M.
２．発表標題 Initial value selection for the alternating least squares algorithm
３．学会等名 Conference of the International Federation of Classification Societies 2017（国際学会）
４．発表年 2017年

1．発表者名 Katayama, H., Mori, Y.
2．発表標題 Item selection for impression survey
3．学会等名 2017 Hangzhou International Statistical Symposium（国際学会）
4．発表年 2017年

1．発表者名 Kuroda, M., Mori, Y., Iizuka, M.
2．発表標題 Improvement of Computation for Nonlinear Multivariate Methods
3．学会等名 The 10th Conference of the IASC-ARS/68th Annual NZSA Conference（招待講演）（国際学会）
4．発表年 2017年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	黒田　正博 (Kuroda Masahiro) (90279042)	岡山理科大学・経営学部・教授 (35302)	
研究分担者	飯塚　誠也 (Iizuka Masaya) (60322236)	岡山大学・全学教育・学生支援機構・教授 (15301)	
研究分担者	久保田　貴文 (Kubota Takafumi) (30379705)	多摩大学・経営情報学部・准教授 (32695)	

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関