

令和 4 年 6 月 9 日現在

機関番号：34310

研究種目：若手研究(B)

研究期間：2017～2021

課題番号：17K12797

研究課題名（和文）ヒト生体情報データに対する3相3次元縮約解析法

研究課題名（英文）Three-way three-mode dimensional data analysis for biometric data

研究代表者

谷岡 健資 (Tanioka, Kensuke)

同志社大学・生命医科学部・助教

研究者番号：40782818

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：3相3元データとよばれる対象×変数×条件に対するデータから対象の分類構造および関連する特徴を同時に把握するためのデータ解析手法の開発を行った。実際に脳機能データに対して制約条件を課した3相3次元縮約クラスタリング法を適用し知見を獲得する研究や、ノイズを含んでいるようなデータに対しても条件間に差がある領域を特定することが可能な次元縮約クラスタリング法の開発、およびパス解析に基づくスパースな次元縮約クラスタリング法の開発を行った。本研究の成果として7本の論文が掲載済みであり、現時点1本は投稿中、1本が投稿準備段階である。

研究成果の学術的意義や社会的意義

情報技術の発達に伴うデータの大規模複雑化により、解析者はより複雑なデータを処理する必要に迫られている。3相3元データはそのような大規模複雑データのひとつであり、対象×変数×条件の3組間の構造を表現したデータであり、各固有の特徴を解釈することは困難である。今回開発した次元縮約クラスタリング法を用いることで、変数や条件固有の特徴を把握することが可能となる。また、パス解析の考えやスパース推定も加えた方法によって解析者の仮説をデータから検証し、解釈すべき項目数を削減することができることから、より解析者が容易に当該構造を解釈することが可能となる。

研究成果の概要（英文）：We have developed a data analysis method to simultaneously grasp the classification structure of subjects and related features from the tuple (subject, variable, condition), which is called Three-way three-mode data. We also developed a dimensional reduction clustering method that can identify regions of difference between conditions even in noisy data. We also developed a sparse dimensionality reduction clustering method based on path analysis. As a result of this research, seven papers have been published, one is in submission and one is in preparation for submission.

研究分野：多変量データ解析

キーワード：クラスタリング法 スパース推定 次元縮約

様式 C-19、F-19-1、Z-19（共通）

1. 研究開始当初の背景

実験等により得られた複雑データからある特徴を持つクラスターを検知し、その特徴と他の情報との関連を明らかにしたい場合を想定する。通常は(対象×変量)の値が記述された多変量データを扱うことが多いが、近年情報技術の発達により様々なデータが観測されるようになり、3相3元データ(対象×変量×条件)を扱う機会も増加してきている。3相3元データは、実験等で異なる複数の条件下で、同一対象、同一観測項目に対してデータを観測した際に得ることが可能である。例えば、fMRI データ等を観測するような実験の場合、被験者×変量数(脳の部位、脳の部位間の関係性)×実験条件のようなデータが3相3元データに該当する。

3相3元データは通常の多変量データと比較して、対象と条件の関係、変量と条件の関係に関する情報を保持しており、そのようなデータから対象のクラスターを検知しても変量数や条件数が多い場合、そのクラスターと関連する変量や条件の特徴を解釈することが困難な場合がある。そのような問題に対策する方法の一つとして Tandem Clustering (Arabie and Hubert, 1994)とよばれる2段階アプローチが応用上使用される場合がある。Tandem Clustering では最初のステップで主成分分析等の次元縮約のための多変量解析をデータへ適用し、低次元上の対象の座標行列を計算し、第2段階目で低次元上の座標行列に対して既存のクラスタリング法を適用するというアプローチである。Tandem Clustering は適用が容易という利点があるが、De Soete and Carroll (1994)や Vichi and Kiers (2001)によってクラスタリング結果に問題が生じるという点が指摘されている。そのような問題を克服する方法として多変量データに対する次元縮約クラスタリング法が提案されており (De Soete and Carroll, 1994; Vichi and Kiers, 2001), クラスタリング結果とクラスターを識別するための低次元空間を同時に推定する方法が提案されている。3相3元データに対しても Tucker model に基づく次元縮約クラスタリング法 (e.g. Rocci and Vichi, 2005; Vichi et al., 2007) が提案されているが、3相3元データに対する次元縮約クラスタリング法の必要性は高まっているにもかかわらず当該研究は依然少ない。

さらに、解析者が持つ仮説を検証するために、当該データをデータ解析の結果に反映することは重要である。3相3元データを含む大規模複雑データに対するパス解析に基づく新たな次元縮約クラスタリング法について必要性があるにも関わらず、当該解析手法の整備は十分とは云えない。

2. 研究の目的

本研究では実験等により得られた3相3元データを含む複雑データからある特徴を持つクラスターを検知し、クラスターと関連する変量や条件の特徴を容易に把握することができるような次元縮約クラスタリング法を開発することを目的とする。具体的には次の2つの研究を実施する。

- (1) 3相3元データに対する仮定を緩めたノイズに強い次元縮約クラスタリング法の開発
- (2) パス解析とスパース推定に基づく次元縮約クラスタリング法の開発
- (3) (1)(2)の研究に関する付随研究

3. 研究の方法

前述の目的を達成するため、本研究では次の項目ごとに研究方法について説明していく。

- (1) 3相3元データに対する仮定を緩めたノイズに強い次元縮約クラスタリング法の開発
- (2) パス解析とスパース推定に基づく次元縮約クラスタリング法
- (3) (1)(2)の研究に関する付随研究

3相3元データに対する次元縮約クラスタリング法として主に2つの次元縮約クラスタリング法を開発する。特に認知科学実験や心理学実験等で観測される実際のfMRIデータ(被験者×脳領域×条件)に対するノイズに強く仮定が緩い新たな次元縮約クラスタリング法を想定した2つの次元縮約クラスタリング法を開発する。

(1-1) 相関行列の差の内積に対する次元縮約クラスタリング法について

暗算課題を被験者に実施し、被験者×脳領域×(条件1:高負荷な課題 or 条件2:低負荷な課題)に関するfMRIデータから条件間で顕著な差を持つ脳領域および脳領域のクラスターを検知するための新たな次元縮約クラスタリング法を新たに開発する。次に具体的な説明を行う。

まず、3相3元データを条件ごとのデータに分割し、分割されたデータごとに脳領域間の相関行列を計算し、2つの相関行列の差(高負荷な課題 - 低負荷な課題)を表現した行列を得る。この相関行列の差のクラスタリング構造に興味があるが、相関係数は-1から1の範囲をとり、

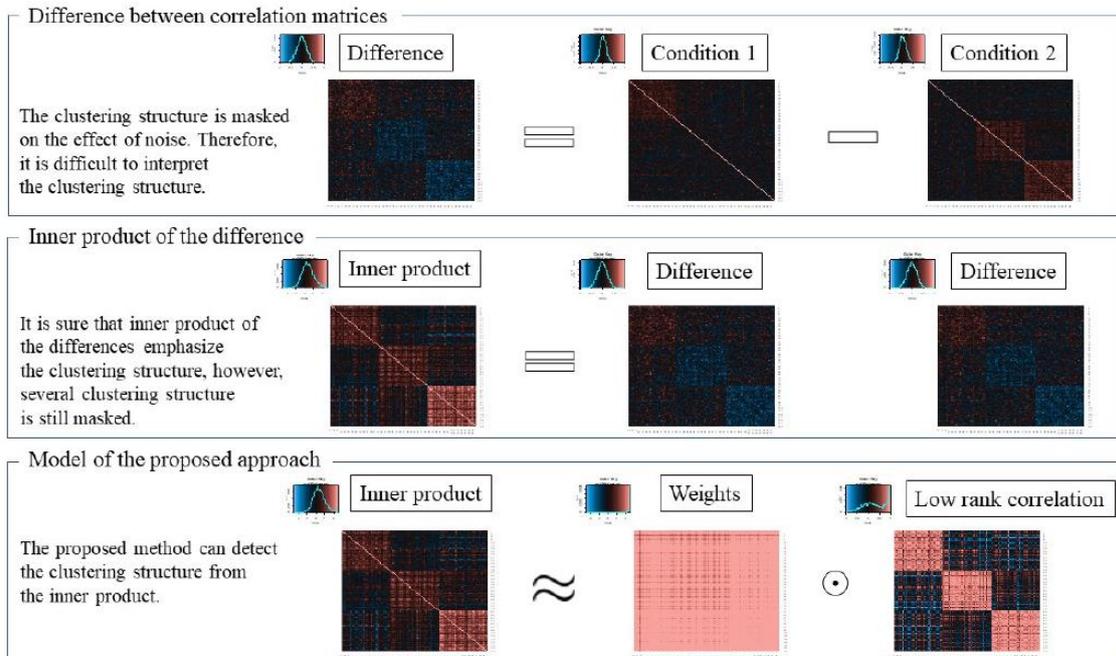


図 1：相関行列の差の内積に対するモデリングについて

その差は非常に小さな値となることが多く、クラスタリング結果をうまく検知することが困難になる場合がある。そこで、「相関行列の差」ではなく（図 1 上段）、「相関行列の差の内積」（図 1 中段）に対してモデリングすることでクラスタリング結果がよくなり、解釈が容易でかつノイズに影響を受けにくい新たな次元縮約クラスタリング法を開発することを試みる（図 1 下段）。

(1-2) Entropy 正則化に基づく複数の相関行列に対する低ランク相関行列の推定

相関行列が与えられた際に、大多数の相関行列が持つ共通の特徴とそうでない特徴に分類し、かつその特徴を容易に解釈したいという問題設定を考える。本研究に関連する研究として複数の相関行列から 1 つの低ランクな相関行列を推定するための Generalized low rank approximation of the correlation matrices (Duan et al., 2014) があるが、全ての相関行列が同じ構造を持っていることを仮定している。そのような問題を克服するため Entropy 正則化に基づく Fuzzy クラスタリング (Miyamoto and Mukaidono, 1997) で使用されている Entropy 正則化を用いた新たな次元縮約法を開発する。

Entropy 正則化された重みにより共通部分を持つ相関行列に高い重みを与え、そうではない相関行列に低い重みを与えることで相関行列を 2 つのクラスターに分類し、一つの低ランク相関行列を推定する方法となっている。本手法は 3 相 3 元データを 2 相 3 元相関行列に変換することで適用可能となる。

(2) Clustered Sparse Structural Equation Modeling for Heterogeneous Data

Structural Equation Modeling (SEM) (e.g. Bollen, 1989) は解析者が持つ仮説をデータで検証する際に有用なパス解析手法の一つである。しかし、データが大規模複雑化していることからデータが同質の構造を持っているとは必ずしも限らないことからクラスタリング法と SEM の同時分析法についても開発が行われている (e.g. Fordellone and Vichi, 2020)。しかし、クラスターごとにパス構造が異なる場合や変数数が多い場合、事前に正確なパス構造をデータに与えることは困難となる。このような問題点を克服するために、スパース推定に基づくパス構造の推定とクラスタリングを同時に実施する新たな次元縮約クラスタリング報を開発する。

本手法では 2 つのアプローチを開発しており、一つは Mixture に基づく方法であり、もう一つは k -means 法に基づく方法である。

(3) (1) (2) の研究に関する付随研究

ここでは 5 つの随研究について説明する。

(3-1) Thresholding Approach for Low-Rank Correlation Matrix Based on MM Algorithm

本研究は(1-2)の付随研究であり、(対象×変数×クラス)に関する多変量データが与えられた際にそこから計算した相関行列からスパースな低ランク相関行列を推定するための新たな次元縮

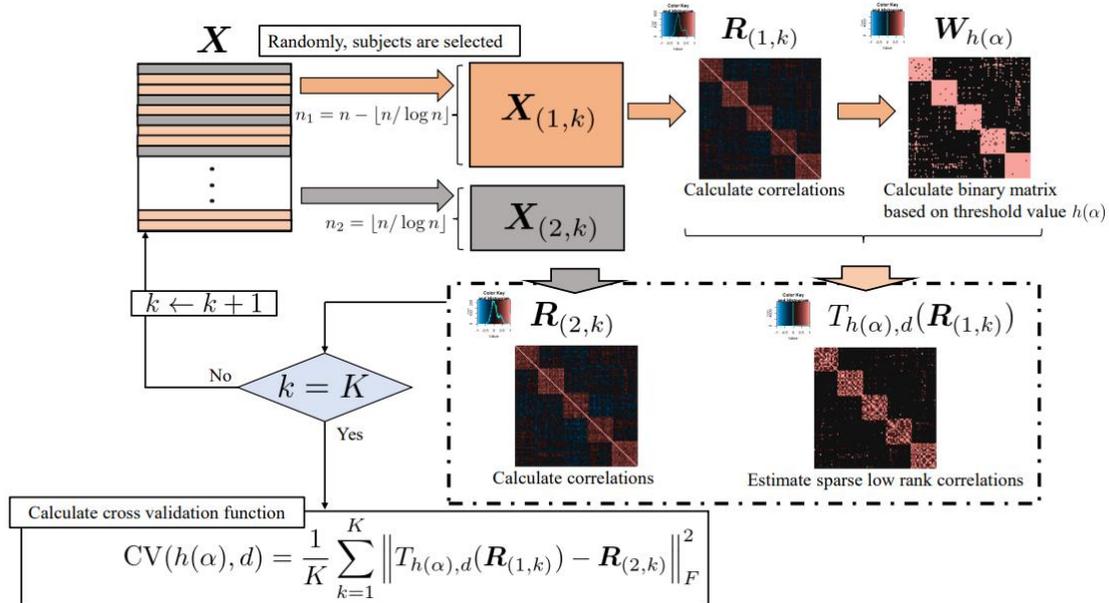


図 2: スパースな低ランク相関行列の推定に関するアルゴリズムの概要

約法を開発する。

相関行列から低ランクであるような相関行列を推定することで、変量間の差が強調され、heatmap と併用することで結果の解釈が容易となるが、本当は相関係数が 0 であるような箇所も誤って絶対値が大きくなるように推定してしまうという問題が生じる。さらに本当は相関係数の値が非ゼロだが 0 に近いような箇所についても絶対値が大きくなるように推定してしまい、誤った解釈を実施してしまうという問題も生じる。そのような問題を克服するような閾値の推定と低ランク相関行列の推定アルゴリズムを組み合わせた新たな低ランク相関行列を推定するための方法を開発する (図 2)。

(3-2) Simultaneous method of orthogonal non-metric non-negative matrix factorization and constrained non-hierarchical clustering

本研究は(1)の付随研究であり、質的多変量データに対する次元縮約クラスタリング法の開発を目的として実施する。質的多変量データはダミー変数行列と呼ばれる 1 と 0 のいずれかを取りうる 2 値データに変換して実施されることが多く、ダミー変数行列は非負値データとして捉えることができる。この非負値性に着目し、次元縮約する際に非負値行列因子分解を活用することで、正則化項を加えることなくスパース推定を実施することが可能となる。さらに、質的多変量データであることを考慮してクラスター重心を低次元空間の重心ではなく、質的多変量データのカテゴリと制約することで、各クラスターに所属する被験者が各変量でどのカテゴリを選択したかが分かるような方法として開発する。

(3-3) Constrained nonmetric principal component analysis

本研究は(1)の付随研究であり、質的多変量データが与えられた際に「対象の外部情報」や「変量の外部情報」それぞれの主成分得点や主成分負荷量を推定するための新たな解析手法を開発する。Optimal scaling (e.g. Young, 1981)を用いてカテゴリに目的関数の観点から最適な重みを推定することで、より近似がよい constrained nonmetric principal component analysis を開発する。

(3-4) Unfolding Models for Asymmetric Dissimilarity Data with External Information based on Path Structure

本研究は(2)の付随研究であり、多変量データおよび同一対象に対して非対称 (非) 類似性データ、および変量間のパス構造が与えられた際に非対称な関係性を表現可能な低次元空間を推定するための新たな非対称多次元尺度構成法を開発する。非対称多次元尺度構成法として様々な手法が開発されているが、本手法では Unfolding (Constantine and Gower, 1978) を基に当該目的を達成するための解析手法を開発する。

(3-5) Asymmetric MDS with Categorical External Information Based on Radius Model

本研究は(1)の付随研究であり、質的多変量データおよび同一対象に対して非対称 (非) 類似性

データが与えられた際に Radius model (Okada and Imaizumi, 1987)に基づく新たな非対称多次元尺度構成法を開発する。既存の非対称多次元尺度構成法では低次元空間での対象間の近さ遠さのみから対象間の関係性を解釈するための方法であったが、本手法を用いることで非対称な関係に加えて、各軸を主成分分析のように解釈することができることが可能となる。

4. 研究成果

(1-1) 相関行列の差の内積に対する次元縮約クラスタリング法について

提案手法の性能を検証することを目的として、数値シミュレーションおよび実際の暗算課題に関する fMRI データに対して適用し、実例を検討した。比較手法としては既存の次元縮約法やクラスタリング法を組み合わせた Tandem approach を 3 手法用意した。評価指標は Adjusted Rand Index (ARI) (Hubert and Arabie, 1985) で行い真のクラスター構造と推定したクラスター構造の類似度を評価して、値が高いか否かを検討した。数値シミュレーションの結果、提案手法の結果が比較手法と比較して結果が良好であることを確認した。さらに、実際の fMRI データに適用した結果、既存の Working Memory との解釈と矛盾がない結果を得ることができた。本研究結果は国際学術誌である applied sciences に既に掲載されている (Tanioka and Hiwa, 2021)。

(1-2) Entropy 正則化に基づく複数の相関行列に対する低ランク相関行列の推定

提案手法の性能を検証するために、既存手法である Duan, et al., (2014) や Simon and Abell (2010) を数値シミュレーションによって比較・検討した。評価方法は真の相関行列と推定した低ランク相関行列の差を採用し、比較した結果、提案手法の有用性を確認した。本研究結果については現在論文投稿の準備段階にある。

(2) Clustered Sparse Structural Equation Modeling for Heterogeneous Data

本研究についても、提案手法と既存の手法を組み合わせた Tandem approach を比較するために数値シミュレーションおよび適用例を実施し、提案手法の有用性を確認している。本研究結果については現在国際論文誌に投稿中で Under review 中である。

(3-1) Thresholding Approach for Low-Rank Correlation Matrix Based on MM Algorithm

本提案手法について数値シミュレーションおよび実際の遺伝子データに対して適用し、比較手法である Tandem approach と比較して良好な結果を得た。特に False positive rate および解釈容易性の指標であるスパースの度合いについては既存手法と比較して良好であることを確認した。本研究結果は国際学術誌である Entropy に掲載済である (Tanioka, et al., 2022)。

(3-2) Simultaneous method of orthogonal non-metric non-negative matrix factorization and constrained non-hierarchical clustering

本提案手法については効率的なアルゴリズムの導出および理論的性質について整理し、数値シミュレーションおよび実際の調査データに対して適用し、既存の質的多変量データに対する次元縮約クラスタリング法と比較して ARI が有用であることを確認している。また、本研究結果は国際学術誌である Journal of Classification に掲載済である (Tanioka and Yadohisa, 2019)。

(3-3) Constrained nonmetric principal component analysis

本提案手法では目的関数の分解や MM algorithm に基づくアルゴリズムの更新式について整備し、既存の Constrained principal component analysis と当てはまり度合いについて比較検証し、その有用性を確認した。本研究結果は国際学術誌である Behaviormetrika に掲載済である (Yamagishi, et al., 2019)。

(3-4) Unfolding Models for Asymmetric Dissimilarity Data with External Information based on Path Structure

実際に各パスに対応する係数および低次元空間の座標行列を同時に推定する方法を開発・実装し、実際の switching 行列と関連する多変量データから非対称な関係を視覚化することに成功している。本研究結果は国際学術誌である International Journal of Software Innovation に掲載済である (Tanioka and Yadohisa, 2018)。

(3-5) Asymmetric MDS with Categorical External Information Based on Radius Model

本提案手法では実際に非対称な関係と同時に解釈可能な軸を推定する新たな非対称多次元尺度法を開発し、実データを視覚化することに成功している。本研究結果は国際誌である Procedia Computer Science に掲載済である (Tanioka and Yadohisa, 2018)。

5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 7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 Tanioka Kensuke, Furotani Yuki, Hiwa Satoru	4. 巻 24
2. 論文標題 Thresholding Approach for Low-Rank Correlation Matrix Based on MM Algorithm	5. 発行年 2022年
3. 雑誌名 Entropy	6. 最初と最後の頁 579 ~ 579
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/e24050579	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 -
1. 著者名 Tanioka Kensuke, Hiwa Satoru	4. 巻 11
2. 論文標題 Low-Rank Approximation of Difference between Correlation Matrices Using Inner Product	5. 発行年 2021年
3. 雑誌名 Applied Sciences	6. 最初と最後の頁 4582 ~ 4582
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/app11104582	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 -
1. 著者名 Kensuke Tanioka, Hiroshi Yadohisa	4. 巻 36
2. 論文標題 Simultaneous method of orthogonal non-metric non-negative matrix factorization and constrained non-hierarchical clustering	5. 発行年 2019年
3. 雑誌名 Journal of Classification	6. 最初と最後の頁 73-93
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s00357-018-9284-8	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 -
1. 著者名 Yuki Yamagishi, Kensuke Tanioka, Hiroshi Yadohisa	4. 巻 46
2. 論文標題 Constrained nonmetric principal component analysis	5. 発行年 2019年
3. 雑誌名 Behaviormetrika	6. 最初と最後の頁 313-332
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s41237-019-00087-3	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 -

1. 著者名 Miyoshi Takuma, Tanioka Kensuke, Yamamoto Shoko, Yadohisa Hiroshi, Hiroyasu Tomoyuki, Hiwa Satoru	4. 巻 13
2. 論文標題 Revealing Changes in Brain Functional Networks Caused by Focused-Attention Meditation Using Tucker3 Clustering	5. 発行年 2020年
3. 雑誌名 Frontiers in Human Neuroscience	6. 最初と最後の頁 1-11
掲載論文のDOI (デジタルオブジェクト識別子) 10.3389/fnhum.2019.00473	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 -

1. 著者名 Tanioka Kensuke, Yadohisa Hiroshi	4. 巻 6
2. 論文標題 Unfolding Models for Asymmetric Dissimilarity Data With External Information Based on Path Structures	5. 発行年 2018年
3. 雑誌名 International Journal of Software Innovation	6. 最初と最後の頁 53 ~ 66
掲載論文のDOI (デジタルオブジェクト識別子) 10.4018/IJSI.2018070104	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 -

1. 著者名 Tanioka Kensuke, Yadohisa Hiroshi	4. 巻 140
2. 論文標題 Asymmetric MDS with Categorical External Information Based on Radius Model	5. 発行年 2018年
3. 雑誌名 Procedia Computer Science	6. 最初と最後の頁 284 ~ 291
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.procs.2018.10.318	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13件 (うち招待講演 1件 / うち国際学会 9件)

1. 発表者名 谷岡健資, 日和悟
2. 発表標題 低ランク近似に基づく相関行列の差の推定について
3. 学会等名 日本計算機統計学会 第34回シンポジウム
4. 発表年 2020年

1. 発表者名 Kensuke Tanioka, Hiroshi Yadohisa
2. 発表標題 Dimensional reduction clustering with modified outcome method
3. 学会等名 16th Conference of the International Federation of Classification Societies (国際学会)
4. 発表年 2019年

1. 発表者名 Kensuke Tanioka, Satoru Hiwa, Tomoyuki Hiroyasu, Hiroshi Yadohisa
2. 発表標題 Joint analysis of the low rank correlation matrices and clustering based on majorization
3. 学会等名 25TH ANNUAL MEETING OF THE ORGANIZATION FOR HUMAN BRAIN MAPPING (国際学会)
4. 発表年 2019年

1. 発表者名 相本武瑠, 風呂谷侑希, 谷岡健資, 日和悟, 宿久洋, 廣安知之
2. 発表標題 ワーキングメモリ課題における課題負荷量が脳機能ネットワーク構造に及ぼす影響: 機能的結合行列の低ランク近似に基づく検討
3. 学会等名 第 16 回日本ワーキングメモリ学会大会
4. 発表年 2018年

1. 発表者名 谷岡健資, 日和悟, 廣安知之, 宿久洋
2. 発表標題 相関行列の差の内積に対するクラスタリングを伴う低ランク近似について
3. 学会等名 2018年度人工知能学会全国大会(第32回)
4. 発表年 2018年

1. 発表者名 谷岡健資, 日和悟, 廣安知之, 宿久洋
2. 発表標題 脳機能ネットワークに対する相関分析法について
3. 学会等名 2017年度人工知能学会全国大会 (第31回)
4. 発表年 2017年

1. 発表者名 Kensuke Tanioka, Hiroshi Yadohisa
2. 発表標題 Generalized Structured Component Analysis for dissimilarity data and multivariate data
3. 学会等名 Big Data, Cloud Computing, and Data Science Engineering (BCD 2017) (国際学会)
4. 発表年 2017年

1. 発表者名 Yuki Yamagishi, Kensuke Tanioka, Hiroshi Yadohisa
2. 発表標題 Constrained Principal Component Analysis for Nonmetric Data
3. 学会等名 61th World Statistics Congress (国際学会)
4. 発表年 2017年

1. 発表者名 Kensuke Tanioka
2. 発表標題 Dimension Reduction Clustering based on Constrained Centroids
3. 学会等名 International Federation of Classification Societies 2017 (国際学会)
4. 発表年 2017年

1. 発表者名 Kensuke Tanioka, Hiroshi Yadohisa
2. 発表標題 Cluster Difference Scaling for Asymmetric Dissimilarity Data based on Unfolding models
3. 学会等名 Hangzhou International statistical symposium (国際学会)
4. 発表年 2017年

1. 発表者名 Toshio Shimokawa, Kensuke Tanioka,
2. 発表標題 Two Stage Approach to Data-Driven Subgroup Identification in Clinial Trials
3. 学会等名 New Zealand Statistical Association and the International Association of Statistical Computing (Asian Regional Section) Joint Conference 2017 (国際学会)
4. 発表年 2017年

1. 発表者名 Ke Wan, Kensuke Tanioka, Kun Yang, Toshio Shimokawa
2. 発表標題 Adjusted Adaptive Index Model For Binary Response
3. 学会等名 New Zealand Statistical Association and the International Association of Statistical Computing (Asian Regional Section) Joint Conference 2017 (国際学会)
4. 発表年 2017年

1. 発表者名 Kensuke Tanioka, Hiroshi Yadohisa
2. 発表標題 Clusterwise Low-Rank Correlation Analysis Based on Majorization
3. 学会等名 New Zealand Statistical Association and the International Association of Statistical Computing (Asian Regional Section) Joint Conference 2017 (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
--	---------------------------	-----------------------	----

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
---------	---------