

令和 3 年 5 月 30 日現在

機関番号：17104

研究種目：基盤研究(C)（一般）

研究期間：2018～2020

課題番号：18K11472

研究課題名（和文）Tensor SOM Network：複雑な大規模複合データの包括的解析

研究課題名（英文）Tensor SOM Network: Comprehensive analysis method for large-scale complex data

研究代表者

古川 徹生（Furukawa, Tetsuo）

九州工業大学・大学院生命体工学研究科・教授

研究者番号：50219101

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究の目的は、複雑に複合した大規模データを包括的に可視化・解析し、知識発見するための汎用的手法の開発である。この目的のため、われわれは以下の4点について取り組んだ。(1) 複合関係データを可視化・解析するための手法開発。(2) 階層構造を持つ複合関係データを可視化・解析するための手法開発。(3) 実データに応用し、開発手法の有用性の実証。(4) 開発手法の基本構成要素となる生成的多様体モデリングの理論研究。われわれは生成的多様体モデルのネットワークにより実現した。これは汎用的なビジュアル・アナリティクスの方法論として位置づけられる。

研究成果の学術的意義や社会的意義

本研究では、解析者とシステムが視覚的インタフェースを介して双方向的に情報をやりとりすることを通して知識発見する、human-centeredな視覚的解析システム、すなわちビジュアルアナリティクス（VA）を実現した。とりわけ、多様体モデルネットワークという概念により、複合関係データに対する汎用的なVAシステムの構築法を提供できた。一方学術的な意義としては、マルチモード・マルチビュー・マルチレイヤーなデータの多様体モデリングの学習手法の開発を行った。とりわけ、階層的な多様体モデリングは単純な尤度最大化等の原理では実現できないことを見出し、新たな研究への手がかりを得た。

研究成果の概要（英文）：The purpose of this research is to develop a general-purpose method for comprehensive visualization, analysis, and knowledge discovery of complex large-scale data. For this purpose, we addressed the following four points. (1) Development of methods for visualization and analysis of complex relational data. (2) Development of a method for visualization and analysis of complex relational data with a hierarchical structure. (3) To demonstrate the usefulness of the developed method by applying it to real data. (4) Theoretical research on generative manifold modeling, which is the element component of the development method. We have achieved the method as a network of generative manifold models. This can be regarded as a general-purpose methodology for visual analytics.

研究分野：機械学習

キーワード：多様体モデリング ビジュアルアナリティクス 関係データ 埋め込み テンソルデータ

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

ビッグデータ解析における困難は、データサイズのみではない。複雑な構造を持つ複合データはデータ全貌の解析や理解を困難にする。図1は本研究者が自動車メーカーとの共同研究で用いた複合データの実例であるが、複数の調査データが複合したデータ構造を持つ。また調査ごとのサンプル不一致やデータ欠損も多い。このような複合データの全貌を包括的に解析する手法の必要性が高まっており、複合テンソル解析として研究されている。しかし現段階では線形テンソル分解にもとづく手法が主であり、また学習後の分析手法も確立されていない。さらにデータ構造に応じて個別にモデリング・解析システムを設計する必要があるため、システム開発のコストも要する。本研究では、このような複合データを包括的に理解するためのモデリングおよび視覚的解析システムの方法論を開発することをめざす。

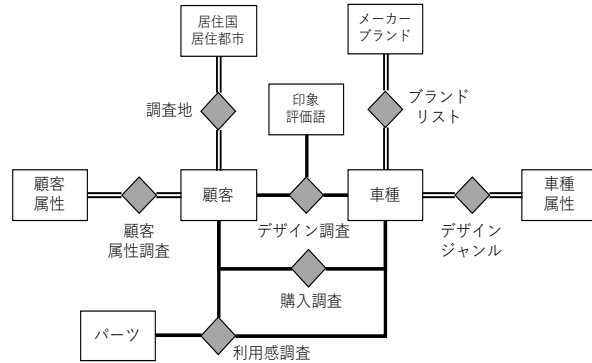


図1: 複合データの例

2. 研究の目的

本研究の目的は、複数の多様体モデルを相互的・階層的に結合したネットワークを用いてモデル化し、さらに可視化を通して包括的な知識発見を行うための汎用的解析手法の開発である。すなわち「多様体モデルネットワーク」がキーアイデアである。また多様体モデルネットワークを構成する基本的な構成要素がテンソル型自己組織化マップ (Tensor SOM) である。本研究のキーアイデアとなる「多様体モデルネットワーク」の例を図2に示す。

この目的を達成するため、以下の目標を設定した。

- (1-1) 相互結合による多様体モデルネットワークの学習アルゴリズムを開発する。
これはマルチモード・マルチビューデータの学習法の開発である。co-training 法などを用いて、ビューごとに適切なメトリック推定を行う。
- (1-2) 階層結合による多様体モデルネットワークの学習アルゴリズムを開発する。
これは多様体モデルの階層学習法の開発である。また多様体モデルのマルチタスク・マルチレベル学習としても位置づけられる。情報幾何学的に見て2種類の多様体モデルの階層学習を実現する。
- (2) 多様体モデルネットワークを用いた複合データの可視化・解析法を開発する。
これは汎用的ビジュアルアナリティクス手法の開発として位置づけられる。
- (3) 実際の大規模複合データへ応用し、現実課題での有用性を検証する
- (4) さらなる発展に向け、Tensor SOMに代わるネットワーク基本構成を開発する。
とりわけ、確率空間内における多様体モデリングや、連続潜在変数の推定法を開発する。また、多様体モデルネットワークに関する基礎的学習理論についても取り組む。

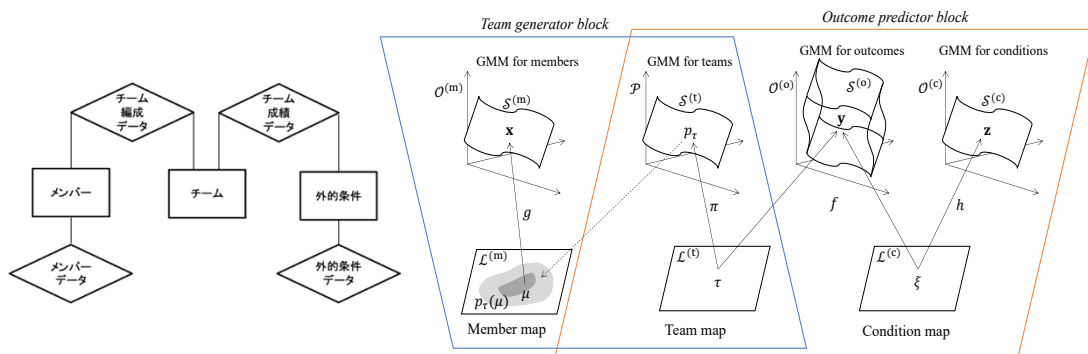


図2: チームデータ解析・チーム編成支援のための多様体モデルネットワーク。

左図: データ構造, 右図: 多様体モデルネットワーク。

3. 研究の方法

(1-1) Co-training 法によるマルチビュー・マルチモードデータの学習法

マルチモードデータのマルチビュー学習法を確立するため、最尤推定アプローチ、ベイズ推定アプローチおよび co-training アプローチによる学習法を実装するとともに、マルチモードデータのメトリック推定法を確立する。

(1-2) 階層学習法は、情報幾何学における m 混合、 e 混合に相当する 2 種の手法を開発する。M 型では全タスクを単一の多様体でモデリングするとともに、各タスクを多様体上の確率分布として m 混合によりモデル化する。一方 e 型では個々のタスクを多様体でモデル化するとともに、多様体モデル集合を高次の多様体でモデル化し、全体をファイバー束でモデル化する。また各データセットが小数サンプルしか持たない場合にも学習可能なマルチタスク学習法を開発する。

(2) 多様体モデルネットワークを用いた可視化法を開発する。キーアイデアは (i) 多様体モデルネットワークをベイジアンネットワークとみなし、確率伝播を行うことでネットワークの離れた 2 点間の関係を可視化する方法を実現する。(ii) 積多様体モデルを非負テンソル分解し、トピックを可視化する方法を開発する方法を行う。さらに多様なデータ構造に柔軟に対応可能な汎用ビジュアルアナリティクスとしてまとめる。

(3) 上記の手法を実データに応用することで、開発した手法の有用性を検証するとともに、多様体モデルネットワークによるビジュアルアナリティクスツールとして完成させる。具体的には以下の実データに応用する。

- 風景画像や工業デザイン、食品嗜好等に対する感性データ解析を行う。とりわけ、被験者の主観評価データと物理特性等の客観評価データの関連を明らかにする課題に応用する。
- 文書データから文書・単語・トピックを可視化し、双方向的な検索法を実現する。ここでは文書と単語の多対多の相互関係を可視化することが課題となる。
- E コマースサイトの閲覧・購買データを同時モデリングする。ここでは、ユーザーが持つ複数の役割の使い分けに応じた解析が課題となる。
- ファッション系 SNS のスナップ写真をタグとスタイルの双方からモデル化し、検索するシステムを開発する。ここでは、コーディネートスタイル情報と付随するタグ情報の間の関連性を可視化することが課題となる。
- スポーツチームの編成支援のためのビジュアルアナリティクスを開発し、有用性を検証する。これは本研究課題において成果を典型的に示す課題として実施する。

(4) 理論・アルゴリズム開発においては、以下のアプローチで研究をすすめる。

(4-a) 離散化潜在変数を用いる Tensor SOM から連続潜在変数モデルへの拡張を行う。特に教師なしカーネル回帰法を用いた関係データ解析法を実現する。潜在変数を用いたノンパラメトリックなモデルを用いることで、写像と潜在変数の同時推定問題を、潜在変数のみの推定問題に帰着する。またスパース近似を導入し、大規模データに対してスケーラブルなモデリング手法を実現する。

(4-b) 積多様体モデリングを確率空間へ拡張する。とりわけ、共起確率データやグラフデータの積多様体モデリングの学習法を開発するとともに、その学習理論を明らかにする。とりわけ、潜在変数間の相互情報量に注目する。

(4-c) 階層的な多様体モデリングの学習原理を明らかにする。階層的な多様体モデリングは最尤推定やベイズ推定等のアプローチから導出することができない。これらの古典的な方法の代わりに、他のダイバージェンスや距離を用いて学習原理を明らかにする。また小数サンプルのマルチタスク課題にも拡張し、多様体モデリングにおけるマルチレベル・マルチタスク学習の学習原理を情報幾何学的な立場から明らかにする。

4. 研究成果

(1-1) マルチモード・マルチビューデータの学習法では、Co-training 法を用いたメトリック学習を行うことでビュー間のコンセンサスを見つけ、ビュー間に共通する潜在変数の表現と、多様体によるデータ分布のモデル化ができることを示した (図 3)。

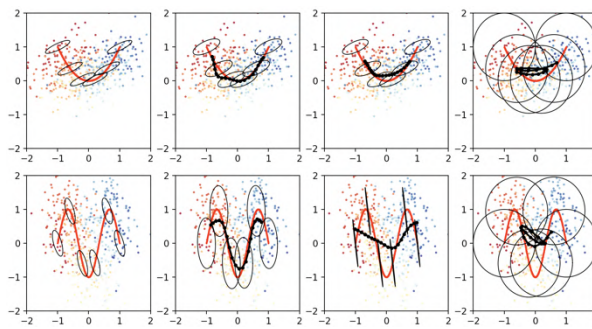


図 3: マルチビュー学習の実験結果。左から Ground truth, Co-training Learning, Ensemble Learning, Ordinary Learning.

(1-2) 階層モデリングでは、m 型および e 型のアルゴリズムを開発し、それをマルチモードデータ（テンソルデータ）に拡張した。また開発したアルゴリズムをスポーツチームデータに応用した（図 4）。

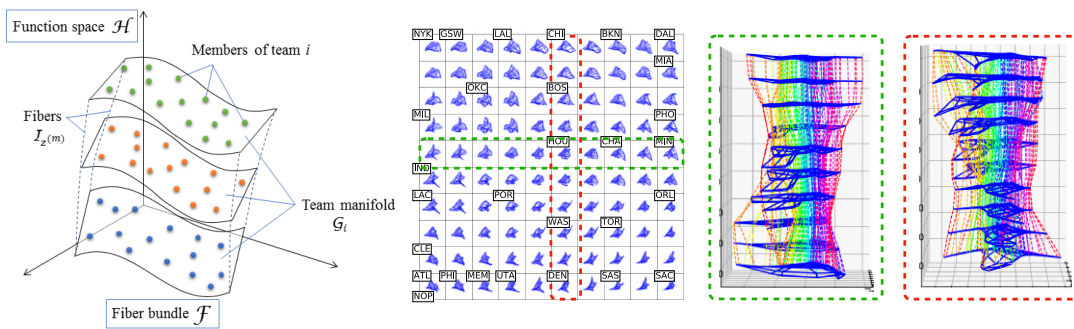


図 4: (左図) e 型階層モデリングではデータ分布をファイバー束でモデル化する。
(右図) e 型階層モデリングをスポーツチーム集合に応用した場合のファイバー束モデル。

(2) 複雑なデータ構造を持つデータから知識発見するための方法論を開発した。第 1 は多様体モデルにより得られた低次元表現を地図として可視化して解析者に提示するとともに、解析者の意図も地図の上に示すことでシステムと解析者間で視覚的なインタラクションを実現する手法としてコンセプトを明確化した。これは汎用ビジュアルアナリティクスの方法論として位置づけられる（図 5）。

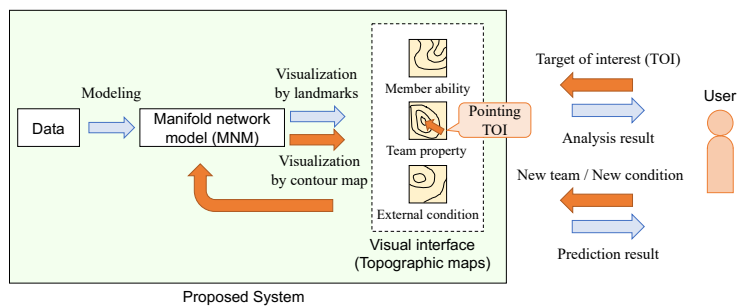


図 5: 多様体モデルネットワークを用いたビジュアルアナリティクスの概念図。多様体モデルによって得られた低次元表現を「地図」とみなし、データを地図上に可視化するだけでなく、解析者の意図 (Target of interest: TOI) も地図上で提示する人間中心的解析システムの枠組を実現した。

また具体的な可視化法として、多様体モデルをベイジアンネットワークとして情報伝搬する方法、および非負テンソル分解によりトピックを可視化する方法を開発した（図 6）。特に後者は、関係データのモード数が 3 以上の場合に有効な可視化法となる。

(3) 開発した多様体モデルを実データに応用して有効性を検証した。図 6 は文書データの可視化結果である。また図 7 は風景画像を用いた主観評価と客観評価の同時解析の例である。

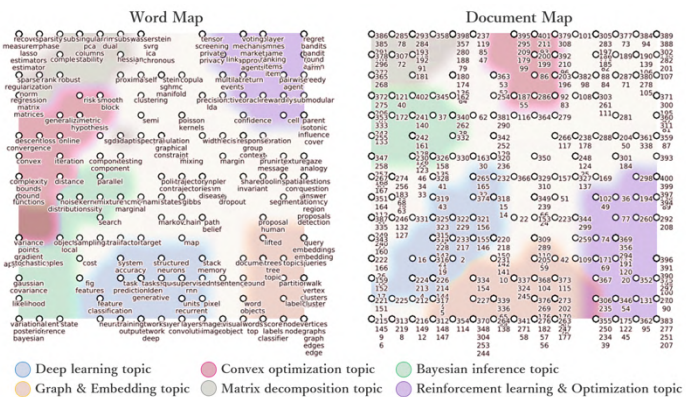


図 6: ドキュメントデータから、文書・単語およびトピックの三者を同時に可視化した結果。

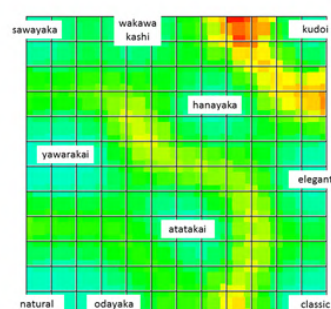
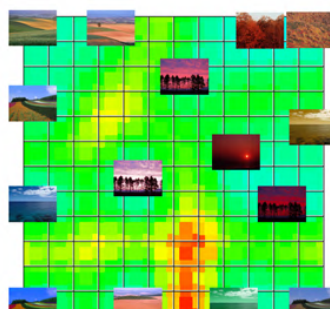
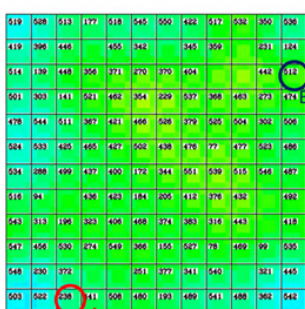


図 7: 風景画像に対する被験者の主観的な評価データおよび画像特徴量との関係を多様体モデリングにより解析した。その結果、被験者のタイプに応じて印象を左右する物理量に違いがあることが読み取れた。

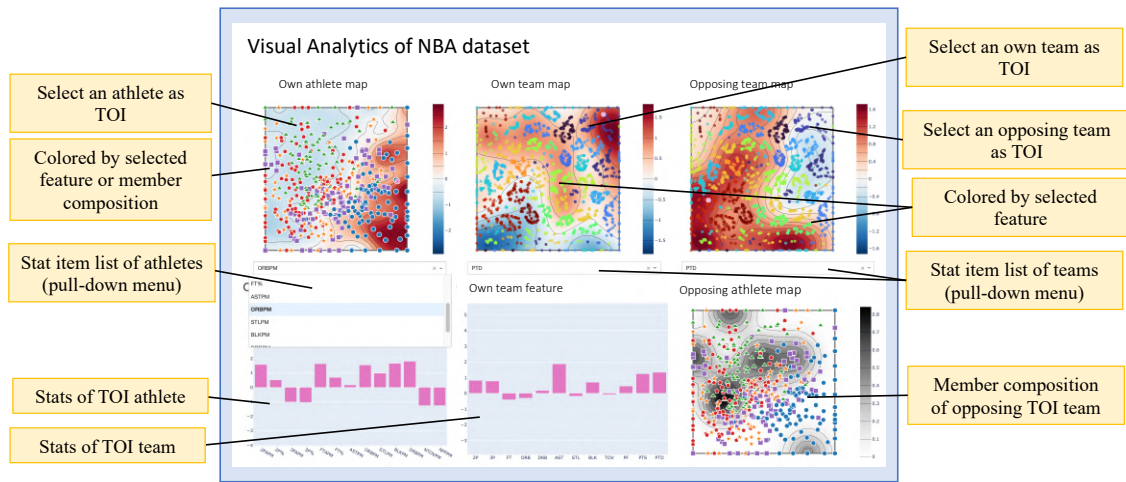


図 8: バスケットボールのチーム編成支援を行うためのビジュアルアナリティクスシステム。データは地図として可視化されるとともに、解析者の意図も地図上でシステムへフィードバックできる。また本システムは集合データの汎用ビジュアルアナリティクスシステムにもなっている。本システムでは (1-1) で開発したマルチモード・マルチビューデータのモデリング, (1-2) で開発した階層データモデリング, (2) で開発した対話的可視化法, (4) で開発した確率分布空間の連続化多様体モデリング法を用いており、本研究の集大成となっている。

図 8 は本研究の集大成であり、バスケットボールチームの編成支援のためのビジュアルアナリティクスシステムである。本システムでは、個々の選手、個々のチームの編成データおよびチーム間の対戦成績による複合データを多様体モデルネットワーク (図 2) によりモデル化した。解析者は新規のシステム上でチーム編成を行って対戦シミュレーションをすることもできる。

(4) 潜在空間を連続化したパラメトリックな多様体モデリング法を開発した。また確率分布空間へ拡張し、文書や単語等の共起データのモデリングに使えることを示した。さらにその学習原理に、潜在空間における相互情報量最大化があることを示した。この手法は誘導点を用いたスパース近似を導入することが可能であり、 $O(N)$ のオーダーを実現できた。したがって大規模データに対するスケーラブルなモデリング法を提供できる。

マルチタスク多様体モデリングの学習原理についても研究を行った。図 9 は情報幾何学的見地から明らかにしたものである。またその背後には最適輸送距離による多様体推定があることを見出した (図 10)。これは今後の発展的研究テーマとなる。

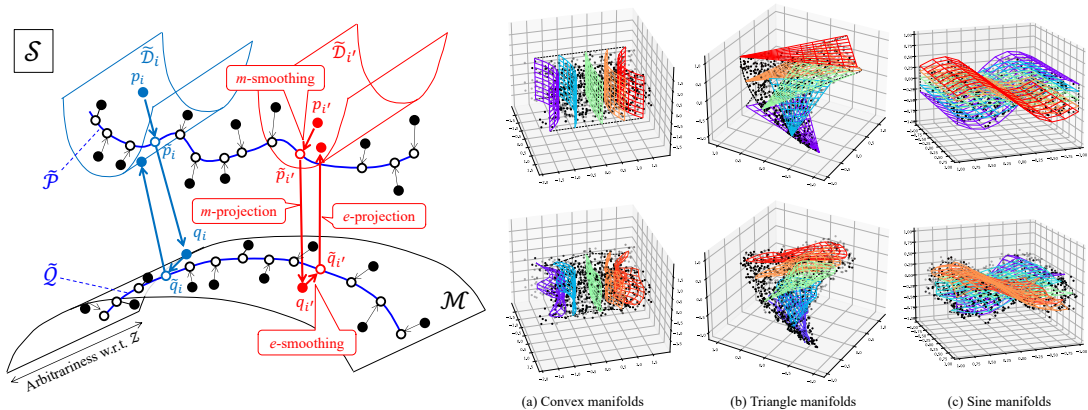


図 9: (左図) マルチタスク多様体モデリングの学習原理の情報幾何学的見地から描像した。(右図) 本学習原理を用いることで、小数サンプルデータからでもマルチタスク多様体モデリングが実現できる (上段: Ground truth, 下段: 推定結果)。

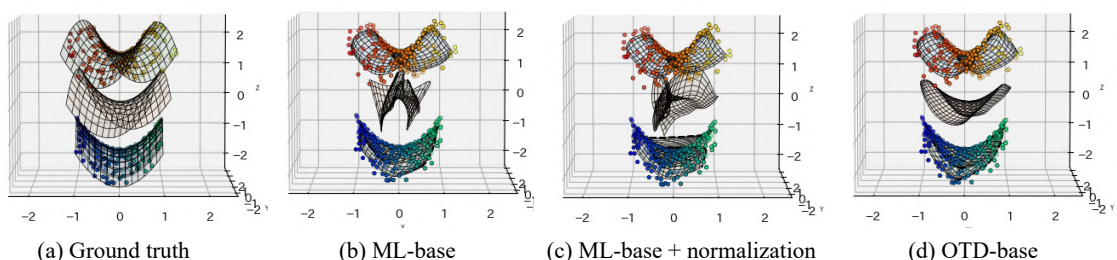


図 10: さまざまな距離基準に基づくマルチ多様体モデリング。最適輸送距離 (OTD) がもっとも望ましい結果を示した。

5. 主な発表論文等

〔雑誌論文〕 計1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 Ishibashi H., Furukawa T.	4. 巻 Vol. E101-A, No.11
2. 論文標題 Hierarchical Tensor SOM Network for Multilevel-Multigroup Analysis	5. 発行年 2018年
3. 雑誌名 Neural Processing Letters	6. 最初と最後の頁 pp.1745-1755
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s11063-017-9643-1	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 0件/うち国際学会 7件）

1. 発表者名 石田琢朗, 米田圭佑, 波田野創, 古川徹生
2. 発表標題 文書・単語の同時分布モデル化による両者の関係性可視化
3. 学会等名 電子情報通信学会ニューロコンピューティング研究会
4. 発表年 2020年

1. 発表者名 宮崎一希, 渡辺龍二, 古川徹生
2. 発表標題 関係データの直積空間への埋め込みによる可視化
3. 学会等名 電子情報通信学会ニューロコンピューティング研究会
4. 発表年 2020年

1. 発表者名 瀬野浦貫太, 石橋英朗, 古川徹生
2. 発表標題 Multi-Level SOMによるメンバー構成によるチームパフォーマンスの可視化
3. 学会等名 電子情報通信学会ニューロコンピューティング研究会
4. 発表年 2020年

1. 発表者名 Florian Tambon, Tetsuo Furukawa
2. 発表標題 Optimal transport based autoencoder for class and style disentanglement
3. 学会等名 電子情報通信学会ニューロコンピューティング研究会
4. 発表年 2020年

1. 発表者名 楠元啓介, 堀尾圭一, 古川徹生
2. 発表標題 Tensor SOMを用いたグループディスカッションにおける幼児間のインタラクション可視化
3. 学会等名 電子情報通信学会ニューロコンピューティング研究会
4. 発表年 2020年

1. 発表者名 Takuro Ishida, Hajime Hatano and Tetsuo Furukawa
2. 発表標題 Simultaneous Visualization of Documents and Words by using Tensor Self-Organizing Map
3. 学会等名 SCIS&ISIS2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Ryuji Watanabe, Hideaki Ishibashi, Tohru Iwasaki and Tetsuo Furukawa
2. 発表標題 Non-parametric Continuous Self-Organizing Map
3. 学会等名 SCIS&ISIS2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Kazushi Higa, Hideaki Ishibashi and Tetsuo Furukawa
2. 発表標題 Multi-task Learning for Self-Organizing Maps
3. 学会等名 SCIS&ISIS2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Kyouhei Itonaga, Kaori Yoshida and Tetsuo Furukawa
2. 発表標題 Tensor Self-Organizing Map for Kansei Analysis
3. 学会等名 SCIS&ISIS2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Kaido Iwamoto, Tohru Iwasaki and Tetsuo Furukawa
2. 発表標題 Visualization of User-item Rating Matrix by Hierarchical Tensor SOM Network
3. 学会等名 SCIS&ISIS2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Keisuke Yoneda, Kimihiro Nakano, Keiichi Horio, Tetsuo Furukawa
2. 発表標題 Simultaneous Analysis of Subjective and Objective Data Using Coupled Tensor Self-organizing Maps: Wine Aroma
3. 学会等名 ICONIP2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Hideaki IshibashiKazushi HigaTetsuo Furukawa
2. 発表標題 Multi-task Manifold Learning Using Hierarchical Modeling for Insufficient Samples
3. 学会等名 ICONIP2018 (国際学会)
4. 発表年 2018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