

令和元年6月20日現在

機関番号：62603

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K16108

研究課題名(和文)疎表現とノンパラメトリックモデルの融合によるデータ駆動型推論に関する研究

研究課題名(英文)Study on the data-driven fusion of nonparametric and sparse modeling

研究代表者

日野 英逸(Hino, Hideitsu)

統計数理研究所・モデリング研究系・准教授

研究者番号：10580079

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：研究期間を通して、ノンパラメトリック手法とそれを支える要素技術として確率分布の混合、一般化線形モデルに基づくデータの潜在次元の推定、スパースモデリングとそれを支える技術としては二重スパース性という概念を用いた画像超解像技術、潜在グラフ構造の推定などを開発した。本研究課題において重要である要素技術の応用においては、スロースリップ地震の解析、マグマテクトニクス場の分類といった当初の予定通りの地球科学分野の研究者との連携による成果に加え、X線分光顕微鏡計測の高速化、電気生理学的実験による神経細胞の結合推定など、様々な分野の研究者と連携しつつ各領域における重要な問題の解決に取り組み成果を上げた。

研究成果の学術的意義や社会的意義

古典的なデータ解析・統計手法が開発された時点とは質・量ともに異なるデータを扱うための新たなデータ解析手法の開発は、工学にとっても科学にとっても喫緊の課題である。データに特定の分布を仮定しないノンパラメトリックモデルは計算・メモリコストが高い。大規模高次元データに適用可能な疎表現アプローチはモデルとして柔軟さに欠ける。本研究課題では、ノンパラメトリックモデルとスパースモデルの融合を目指し、ノンパラメトリック手法、スパースモデリング手法それぞれで新たな手法を開発し、地球科学、材料科学、情報科学の諸問題に適用して学術的知見の発見、計測技術の向上を実現した。

研究成果の概要(英文)：Throughout the research period, we developed several component technologies to realize the concept of the fusion of nonparametric and sparse modeling. We developed a non-parametric mixture distribution estimation method in the framework of information geometry, an intrinsic dimension estimation method based on the generalized linear modeling. We have developed an image super-resolution method and an estimation method of latent graph structure. We also applied conventional and tailored methods to several fields of science. Particularly we collaborated with geoscientists as originally planned to analyse slow slip earthquakes and classification of magma tectonic fields. We also collaborated with material scientists and developed a method for accelerating the X-ray spectroscopy microscope measurement. Also, through the cooperation with researchers in electrophysiological measurements, we proposed a method for estimation of nerve cell connections based on partially observed signals.

研究分野：機械学習

キーワード：機械学習 ノンパラメトリックモデル スパースモデリング

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

古典的なデータ解析・統計手法が開発された時点とは質・量ともに異なるデータを扱うための新たなデータ解析手法の開発は、工学にとっても科学にとっても喫緊の課題である。こうしたデータの多くは、その生成分布が複雑であり、特定の生成モデルを仮定するパラメトリックモデルではデータの本質的な特徴を捉えることが出来ない。データに特定の分布を仮定しないノンパラメトリックモデルと呼ばれる統計的手法の研究は古くから行われている。ノンパラメトリックモデルは柔軟かつ汎用的であり、特定の分布の仮定が出来ない状況において検定や推定、あるいは潜在構造の推定に広く用いられている。一方で、種々の意思決定において過去の全データを保持しておく必要があり、大規模データに適用した際は大量のメモリと計算コストが必要になるという問題がある。一方、大規模データ解析のアプローチとして、近年、スパースモデリング(疎表現)が注目を集めている。疎表現は、データの本質的な部分はごく一部であるという仮定を、疎事前分布という形でモデルに取り込むことで、大規模データから結果に寄与する本質的な情報を抽出することを目的としている。一方で、疎表現に基づく既存モデルの多くはモデルとして柔軟さに欠け、重要な情報を見逃してしまう、あるいは逆に必要ない情報も取り込んでしまうという使いにくさがある。

2. 研究の目的

ノンパラメトリックモデリングの技法による疎性モデルの一般化と、疎性モデリングの技法による大規模データ向けノンパラメトリック潜在構造推定・情報縮約という、両方向からのアプローチにより、疎表現とノンパラメトリックモデルの融合の枠組み構築を目的とする。具体的には、ノンパラメトリックモデルを援用することで疎性モデルの適用範囲・柔軟性を拡張する方法論の研究と、データが有する潜在的構造推定のためのノンパラメトリック手法を、疎性モデルを援用することで効率化する方法論の研究を行う。両方法論に関して、具体的なアルゴリズムを導出し、人工データ及び実データを用いた評価によりそれらの性能と適用限界を明らかにする。その上で、個々の具体的なアルゴリズムの開発にとどまらず、数理モデリングの専門家と、地震・余効すべり解析及び画像解析の専門家と協業し、疎性モデリングとノンパラメトリックモデリングの融合という方法論の確立を目指す。

3. 研究の方法

疎性モデリングとノンパラメトリックモデリングの融合の枠組み構築という目標に向け、(1)疎性モデリングの柔軟性・適用範囲の拡大の枠組みの構築とアルゴリズム開発、(2)ノンパラメトリックモデリングによる潜在構造推定の効率化の枠組み構築とアルゴリズム開発、(3)実問題の解析を通じた評価と改善、という課題に順次取り組む。1年目は、(1)のアルゴリズム開発から着手し、基礎理論構築に向けた取り組みと、(3)の実問題解析を同時進行する。2年目以降は、(2)のアルゴリズム開発から着手し、(1)、(2)の基礎理論の構築と、(3)の実問題解析を通じた理論の検証を行う。

4. 研究成果

平成28年度には、(1)疎性モデリングの柔軟性・適用範囲の拡大の枠組みの構築とアルゴリズム開発、(2)実問題の解析を通じた評価と改善の2つに取り組んだ。(1)の課題については、変数の大部分が零であるという単純な疎性のみではなく、「変数の時間変化が疎である状況」や「グループ構造を持った疎性」をモデルに取り込み、疎性モデリングの適用範囲の拡大の枠組みを構築した。具体的には、変数の時間変化が疎であるという形の疎性を、消費者の商品の選好傾向の時間遷移に対して課すことで、ブランドのシェアの観測系列データから、ブランド間での消費者の移り変わりを推定する手法を開発し、査読付きの学術論文を発表した。また、開発手法の和文による解説を、書籍の1節として執筆した。「グループ構造を持った疎性」については、位置情報サービスを利用するユーザ群の移動パターンをクラスタリングしてから、各クラスタをグループ構造とみなしてグループ単位で疎性を仮定したテンソル分解を行うことで、各ユーザの存在位置を高精度に推定する位置情報攻撃手法を開発し、位置情報サービス利用における問題点を指摘した。本研究も、査読付き学術論文として発表している。また、ノンパラメトリック手法による柔軟な統計モデリングの基礎的な手法として、情報幾何学に基づく確率分布の混合モデリングの新たな方法を提案し、査読付き国際会議での発表及び査読付き学術論文として発表した。

平成29年度には、ノンパラメトリックな潜在構造の推定方法として、潜在的次元推定手法を開発した。従来の低次モーメントを用いた近似的なフラクタル次元の推定と比較して、高次のモーメントの情報を取り込んだ推定手法を開発した。高次モーメントを含めて潜在的次元推定の方法は自明ではなく、一般化線形モデルと多項式回帰、サンプリングといった手法を駆使して、高精度な推定方法を実現した。実現した手法の論文化に加えて、オープンソースソフトウェアとして提案手法と関連手法を統一的に実行できるプログラムを公開した。潜在次元の推定手法は様々な分野で必要とされる技術であり、本研究の遂行を通して得られた知見をもとに、生物データや宇宙科学データ分析における重要な前処理としての展開も実施している。当初の研究計画通り、スパースモデリングに基づくスロースリップ地震の分析を実施した。豊後水道付近のプレート境界におけるスロースリップ推定に取り組み、これまでの手法では検出できなかった、地殻構造を良く反映した滑りパターンの同定に成功した。本研究は Scientific Reports 誌に掲載され、共同研究機関と共同でプレスリリースも行った。その他、地球科学の様々な問題に、本研究課題のテーマであるスパースモデリング、ノンパラメトリックモデリングに基づく手法を適用し、火山灰の客観的分類、マグマテクトニクス場の特徴づけ等、これまで地球科学の諸分野で勤と経

験に基づき行われてきた分析を統計的な客観性を持った方法で実現することに成功している。さらに、材料科学分野の研究者との共同研究も行い、ノンパラメトリックモデリングによる材料計測時間の大幅な短縮を始めとして重要な成果をあげることができた。

研究期間を通して、ノンパラメトリックアプローチとそれを支える技術としては確率分布の混合、一般化線形モデルに基づくデータの潜在次元の推定、スパースモデリングとそれを支える技術としては二重スパース性という概念を用いた画像超解像技術、部分的な観測に基づくグラフ構造の推定などを開発した。本研究課題において重要である要素技術の応用においては、スロースリップ地震の解析、マグマテクトニクス場の分類といった当初の予定通りの地球科学分野の研究者との連携による成果に加え、X線分光顕微鏡計測の高速化、電気生理学的実験による神経細胞の結合推定など、様々な分野の研究者と連携しつつ各領域における重要な問題の解決に取り組み成果を上げた。

5. 主な発表論文等

{雑誌論文}(計 10 件)

1. **Geochemical Discrimination and Characteristics of Magmatic Tectonic Settings: A Machine-Learning-Based Approach**, Kenta Ueki, [Hideitsu Hino](#), Tatsu Kuwatani, *Geochemistry, Geophysics, Geosystems*, vol.19, 2018 (査読あり)
2. **Estimation of neural connections from partially observed neural spikes**, Taishi Iwasaki, [Hideitsu Hino](#), Masami Tatsuno, Shotaro Akaho, Noboru Murata, vol.108, 2018 (査読あり)
3. **Adaptive design of an X-ray magnetic circular dichroism spectroscopy experiment with Gaussian process modelling**, Tetsuro Ueno, [Hideitsu Hino](#), Ai Hashimoto, Yasuo Takeichi, Masahiro Sawada, Kanta Ono, vol.4, 2018 (査読あり)
4. **Identifier: Intrinsic Dimension Estimation with R**, [Hideitsu Hino](#), vol.9, 2017 (査読あり)
5. **Discontinuous boundaries of slow slip events beneath the Bungo Channel, southwest Japan**, Ryoko Nakata, [Hideitsu Hino](#), Tatsu Kuwatani, Shoichi Yoshioka, Masato Okada, Takane Hori, vol.7, 2017 (査読あり)
6. **Local Intrinsic Dimension Estimation by Generalized Linear Modeling**, [Hideitsu Hino](#), Jun Fujiki, Shotaro Akaho, Noboru Murata, *Neural Computation*, vol.29, 2017 (査読あり)
7. **Time-Varying Transition Probability Matrix Estimation and Its Application to Brand Share Analysis**, Tomoaki Chiba, [Hideitsu Hino](#), Shotaro Akaho, Noboru Murata, *PLoS ONE*, vol.12, 2017 (査読あり)
8. **Double sparsity for multi-frame super resolution**, Toshiyuki Kato, [Hideitsu Hino](#), Noboru Murata, *Neurocomputing*, vol.240, 2017 (査読あり)
9. **Group Sparsity Tensor Factorization for Re-identification of Open Mobility Traces**, Takao Murakami, Atsunori Kanemura, [Hideitsu Hino](#), *IEEE Transactions on Information Forensic and Security*, vol.12, 2016 (査読あり)
10. **Non-parametric e-mixture estimation**, Ken Takano, [Hideitsu Hino](#), Shotaro Akaho, Noboru Murata, *Neural Computation*, vol.28, 2016 (査読あり)

{学会発表}(計 6 件)

1. **Geometrical Formulation of the Nonnegative Matrix Factorization**, Shotaro Akaho, [Hideitsu Hino](#), Neneka Nara, Noboru Murata, The 25th International Conference on Neural Information Processing (ICONIP2018), Siem Reap, Cambodia, December 13-16, 2018 (査読あり)
2. **Information Geometric Perspective of Modal Linear Regression**, Keishi Sando, Shotaro Akaho, Noboru Murata, [Hideitsu Hino](#), The 25th International Conference on Neural Information Processing (ICONIP2018), Siem Reap, Cambodia, December 13-16, 2018 (査読あり)
3. **An Entropy Estimator Based on Polynomial Regression with Poisson Error Structure**, [Hideitsu Hino](#), Shotaro Akaho, Noboru Murata, The 23th International Conference on Neural Information Processing (ICONIP2016), Kyoto, Japan, 16-21, October 2016 (査読あり)
4. **Non-parametric e-mixture of density functions**, [Hideitsu Hino](#), Ken Takano, Shotaro Akaho, Noboru Murata, The 23th International Conference on Neural Information Processing (ICONIP2016), Kyoto, Japan, 16-21, October 2016 (査読あり)
5. **Doubly Sparse Structure In Image Super Resolution**, Toshiyuki Kato, [Hideitsu Hino](#), Noboru Murata, The 2016 IEEE Machine Learning for Signal Processing Workshop

- (MLSP2016), Vietri sul Mare, Salerno, Italy, 13-16, September 2016 (査読あり)
6. **Change-Point Detection in a Sequence of Bags-of-Data**, Kensuke Koshijima, Hideitsu Hino, Noboru Murata, International Conference of Data Engineering (ICDE), TKDE poster track, Helsinki, Finland, 16-20, May, 2016 (査読あり)

〔図書〕(計 1 件)

池田思朗, 伊庭幸人, 麻生英樹, 井手剛, 本谷秀堅, 日野英逸, 尾崎隆, 中谷秀洋, 野間久史, 他, 岩波書店, 岩波データサイエンス Vol.5, 2017 年, p.80 - p.90, 時間遷移のスパース性(執筆分担)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

ホームページ等

6. 研究組織

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。