

令和 4 年 6 月 15 日現在

機関番号：82626

研究種目：基盤研究(B) (一般)

研究期間：2017～2021

課題番号：17H01793

研究課題名(和文) 情報幾何学に基づく分布データに対する機械学習手法の開発

研究課題名(英文) Development of machine learning method for distribution data based on information geometry

研究代表者

赤穂 昭太郎 (Akaho, Shotaro)

国立研究開発法人産業技術総合研究所・情報・人間工学領域・上級主任研究員

研究者番号：40356340

交付決定額(研究期間全体)：(直接経費) 14,340,000円

研究成果の概要(和文)：大量のデータを効率的に機械学習で扱う有効な方法として、データを確率分布のパラメータに縮約することが考えられる。本課題では、そのようなデータに対する機械学習の開発に取り組んできた。もともと、代表者が開発してきた分布データに対する主成分分析の拡張という研究があったが、それをより柔軟性のあるノンパラメトリックな枠組みへと拡張するというのが本研究課題の学術的な貢献であり、ガウス過程回帰の情報幾何などで達成できた。また、神経科学や地球物理学などへの応用についても情報幾何学的な行列分解の適用などを通じて実践できた。

研究成果の学術的意義や社会的意義

機械学習では、通常実数値ベクトルを入力として学習が行われることが前提であるが、大量のデータを集約した分布のパラメータなどは実数値として扱うことが必ずしも適切ではない。本研究課題では、そのような分布データに関する機械学習の手法を理論・応用の両面から深化させ、さまざまな応用課題(神経科学や地球科学をはじめとする多くの分野)に適用するための方法論を確立したことにある。

研究成果の概要(英文)：An effective way to handle large amounts of data by machine learning is to reduce the data to the parameters of a probability distribution. In this project, we have been working on the development of machine learning for such data. Originally, there was a study of extending principal component analysis to distributional data, which had been developed by a project member. The significant contribution of this project was to extend it to a more flexible nonparametric framework, which was achieved through information geometry of Gaussian process regression and other methods. We have also been able to apply information geometry to neuroscience and geophysics through the application of matrix factorization.

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 情報幾何 学習アルゴリズム

1. 研究開始当初の背景

本研究課題では、個々のデータが確率分布で与えられる状況でのデータ解析手法を考える。例えば、地球科学における研究で、岩石を分析してそこに含まれる元素量を計測したデータは絶対量は意味を持たず、相対的な比率のみが意味を持つ確率値データとみなすことができる。情報科学においても、人間の購買データやニュース記事の単語頻度など、多方面の分野において確率分布で表現されるデータを解析する必要性が高まっている。しかしながら、線形主成分分析などの従来の多変量解析は、データが従う分布のガウス性を暗に仮定しているため、確率データに適用すると誤った結論を導くことがある。本研究課題で用いる情報幾何学は、確率分布の空間に対するデータ解析に関して数学的な枠組みを提供し、機械学習の手法を統一的な観点から解釈することで問題点を明確化できるという利点をもつ。ただし、これまでの情報幾何学は主にパラメトリックな空間に対して構築されてきたため、時空間データ構造を持つ地球科学や脳科学といった自然科学分野で得られるデータに対しては必ずしも適用できる枠組みが整っていないという現状がある。

2. 研究の目的

個々のデータが確率分布で与えられる場合に適用可能な機械学習のアルゴリズムを情報幾何の観点から研究する。特に本研究課題では、時系列構造をもつ地球科学データや脳科学データなど、点過程としての性質をもつデータに適用可能なアルゴリズムの開発に取り組む。そのために、主にパラメトリックな空間で構築されてきた従来の情報幾何学をノンパラメトリックな枠組みに拡張したノンパラメトリック情報幾何学の確立を目指す。また、申請者がこれまで手掛けてきた地球科学・脳科学などを対象とした現実のデータ解析に適用することにより、単なる手法の解釈にとどまらず、データ駆動科学に新たな方法論を提供することを目的とする。

3. 研究の方法

本研究課題では、【課題1】確率データ解析の情報幾何的な統一的理解、【課題2】ノンパラメトリック確率データ解析法の開発、【課題3】地球科学・脳科学データなど実データへの適用、の3課題に取り組む。【課題1】では、情報幾何学の枠組みで既存の機械学習手法を整理・改良し、性質の理論解析に取り組む。【課題2】では、ノンパラメトリック情報幾何学の確立を通じて点過程データに適用可能な手法を開発する。また、【課題3】として、岩石の組成データや地球観測データ、顔認知課題などにおける脳活動のスパイクデータの解析に取り組む。さらに、協力研究者と議論しながら手法の改善や理論解析にフィードバックする。

4. 研究成果

- (1) モード回帰の情報幾何アルゴリズム的解釈：確率値の条件付モードを回帰するためのモード回帰に対して、ノンパラメトリックな手法であるカーネル密度推定に基づく従来手法が情報幾何学的に解釈できることを明らかにし、ロバスト性についても理論解析を行った。
- (2) ガウス過程回帰の次元削減の情報幾何学的アルゴリズムの開発：無限次元の自由度を持つガウス過程回帰に対し、問題が有限次元の問題に帰着できることを理論的に導出し、それに基づいて高速な次元削減アルゴリズムを開発した。これはマルチタスク学習や転移学習への展開が期待できる成果である。
- (3) 非負値行列分解の情報幾何的アルゴリズムと応用への展開：非負値行列分解の問題が確率値行列の分解に帰着できることを利用し、情報幾何の基本的性質である拡張ピタゴラスの定理に基づいた安定的なアルゴリズムを開発した。また、マウスの睡眠時脳活動カルシウムイメージングデータに対する応用研究も行い、睡眠時と覚醒時でセルアセンブリと呼ばれる神経細胞のグループ化のパターンが異なることが示唆された。
- (4) 神経ネットワークの推定法の開発：神経細胞の多点電極計測データから、神経同士の機能的なネットワークを推定する手法を開発した。この特徴は、観測されていないデータからの影響を除去できる点、細胞の結合の方向性を含めた推定が可能となる点、複数ステップでの偽相関を除去できる点である。また、神経細胞には興奮性と抑制性の2種類の細胞があるため、それを潜在変数として、情報幾何学的なアルゴリズムを開発した。
- (5) スパースアイソコン法の開発：岩石の分析などでは、データが質量や体積の相対的な比率である組成データで与えられるため、物質の移動量を絶対値として評価することが不可能である。そのため、地球科学分野では、不動元素というほとんど物質移動をしない元素を仮定することで、

絶対量の推定を行ってきたが、アイソコングラフというプロットから人為的に選ばれることが多かった。そのため、スパース性を評価基準に入れたスパースアイソコン法を開発し、客観的に不動元素を特定する技術を開発した。

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 8件/うち国際共著 0件/うちオープンアクセス 5件）

1. 著者名 Goto Shin-itiro, Hino Hideitsu	4. 巻 11712
2. 論文標題 Expectation Variables on a Para-Contact Metric Manifold Exactly Derived from Master Equations	5. 発行年 2019年
3. 雑誌名 Geometric Science of Information (LNCS)	6. 最初と最後の頁 239 ~ 247
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-26980-7_25	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 Goto Shin-itiro, Hino Hideitsu	4. 巻 95
2. 論文標題 Information and contact geometric description of expectation variables exactly derived from master equations	5. 発行年 2019年
3. 雑誌名 Physica Scripta	6. 最初と最後の頁 015207 ~ 015207
掲載論文のDOI (デジタルオブジェクト識別子) 10.1088/1402-4896/ab4295	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Akaho Shotaro, Hino Hideitsu, Murata Noboru	4. 巻 1143
2. 論文標題 On a Convergence Property of a Geometrical Algorithm for Statistical Manifolds	5. 発行年 2019年
3. 雑誌名 ICONIP 2019: Neural Information Processing (CCIS)	6. 最初と最後の頁 262 ~ 272
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-36802-9_29	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 Iwasaki Taishi, Hino Hideitsu, Tatsuno Masami, Akaho Shotaro, Murata Noboru	4. 巻 108
2. 論文標題 Estimation of neural connections from partially observed neural spikes	5. 発行年 2018年
3. 雑誌名 Neural Networks	6. 最初と最後の頁 172 ~ 191
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.neunet.2018.07.019	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 Akaho Shotaro, Hino Hideitsu, Nara Neneka, Murata Noboru	4. 巻 11303
2. 論文標題 Geometrical Formulation of the Nonnegative Matrix Factorization	5. 発行年 2018年
3. 雑誌名 International Conference on Neural Information Processing	6. 最初と最後の頁 525 ~ 534
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-04182-3_46	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Sando Keishi, Akaho Shotaro, Murata Noboru, Hino Hideitsu	4. 巻 11303
2. 論文標題 Information Geometric Perspective of Modal Linear Regression	5. 発行年 2018年
3. 雑誌名 International Conference on Neural Information Processing	6. 最初と最後の頁 535 ~ 545
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-04182-3_47	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yoshida Kenta, Kuwatani Tatsu, Hirajima Takao, Iwamori Hikaru, Akaho Shotaro	4. 巻 36
2. 論文標題 Progressive evolution of whole-rock composition during metamorphism revealed by multivariate statistical analyses	5. 発行年 2017年
3. 雑誌名 Journal of Metamorphic Geology	6. 最初と最後の頁 41 ~ 54
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/jmg.12282	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 H. Hino	4. 巻 9
2. 論文標題 Identifier: Intrinsic Dimension Estimation with R	5. 発行年 2017年
3. 雑誌名 R Journal	6. 最初と最後の頁 329-341
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 有
オープンアクセス オープンアクセスとしている (また、その予定である)	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 2件）

1. 発表者名 永山瑞生・有竹俊光・日野英逸・上田壮志・宮崎峻弘・柳沢正史・赤穂昭太郎・村田 昇
2. 発表標題 非負値行列因子分解を用いたカルシウムイメージングデータからの睡眠状態解析
3. 学会等名 電子情報通信学会 IBISML 研究会
4. 発表年 2019年

1. 発表者名 Mizuo Nagayama, Toshimitsu Aritake, Hideitsu Hino, Takeshi Kanda, Toshihiro Miyazaki, Masashi Yanagisawa, Shotaro Akaho, Noboru Murata
2. 発表標題 Sleep State Analysis Using Calcium Imaging Data by Non-negative Matrix Factorization
3. 学会等名 International Conference on Artificial Neural Networks 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Tatsu Kuwatani, Kenta Yoshida, Kenta Ueki, Ryosuke Oyanagi, Masaki Uno, Shotaro Akaho
2. 発表標題 Estimation of Material Transfer from Compositional Dataset Using Sparse Modeling
3. 学会等名 Asia Oceania Geosciences Society 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 吉田健太, 桑谷 立, 上木賢太, 大柳良介, 赤穂 昭太郎
2. 発表標題 スパースモデリングによる沈み込み帯深部流体活動の定量的見積り
3. 学会等名 日本地質学会第126年学術大会
4. 発表年 2019年

1. 発表者名 桑谷 立, 赤穂昭太郎
2. 発表標題 スパースアイソコン法の資源探査分野への応
3. 学会等名 資源・素材 2 0 1 9
4. 発表年 2019年

1. 発表者名 石橋 英朗, 赤穂昭太郎
2. 発表標題 情報幾何に基づくガウス過程事後分布集合の次元削減法
3. 学会等名 電子情報通信学会 IBISML 研究会
4. 発表年 2019年

1. 発表者名 桑谷立, 赤穂昭太郎
2. 発表標題 全岩化学組成データセットからのプロセス推定: カコウ岩の風化を例に
3. 学会等名 日本地質学会
4. 発表年 2018年

1. 発表者名 桑谷立, 赤穂昭太郎
2. 発表標題 岩石風化プロセスのフォワードモデリングと逆解析
3. 学会等名 資源・素材学会秋季大会
4. 発表年 2018年

1. 発表者名 唐木田亮, 赤穂昭太郎, 甘利俊一
2. 発表標題 深層ニューラルネットワークにおけるFisher情報行列の普遍性
3. 学会等名 IBIS2018
4. 発表年 2018年

1. 発表者名 三戸圭史, 赤穂昭太郎, 村田 昇, 日野英逸
2. 発表標題 最頻値線形回帰の情報幾何
3. 学会等名 第32回情報論的学習理論と機械学習研究会 (IBISML)
4. 発表年 2018年

1. 発表者名 石橋英朗, 古川 徹生, 赤穂昭太郎
2. 発表標題 事後分布推定されたガウス過程間のKLダイバージェンスは有限次元の正規分布間のKLダイバージェンスで評価できる
3. 学会等名 情報論的学習理論ワークショップ (IBIS2017)
4. 発表年 2017年

1. 発表者名 江口 脩, 田中 勝, 藤木 淳, 栗田 多喜夫
2. 発表標題 深層学習による楽器音の音名識別
3. 学会等名 画像の認識・理解シンポジウム (MIRU2017)
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	藤木 淳 (Fujiki Jun) (10357907)	福岡大学・理学部・教授 (37111)	
研究分担者	日野 英逸 (Hino Hideitsu) (10580079)	統計数理研究所・モデリング研究系・教授 (62603)	
研究分担者	村田 昇 (Murata Noboru) (60242038)	早稲田大学・理工学術院・教授 (32689)	

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