

平成30年6月14日現在

機関番号：82626

研究種目：新学術領域研究（研究領域提案型）

研究期間：2013～2017

課題番号：25120011

研究課題名（和文）カーネル法による高次元データの非線形スパースモデリング

研究課題名（英文）Nonlinear sparse modeling for high dimensional data by kernel methods

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・研究グループ長

研究者番号：40356340

交付決定額（研究期間全体）：（直接経費） 41,400,000円

研究成果の概要（和文）：複数情報源からの情報抽出法に関して、高速計算可能な相互情報量を基準として用いたカーネル法を開発し、非線形で複雑な関係性の抽出法に成功した。また、複数タスクを統合する転移学習のために、情報幾何学の枠組みから研究を進め、e-混合と呼ばれる統合手法のノンパラメトリックなアルゴリズムを開発した。
ダイナミクスの解析では、神経ネットワークやシェアデータといった多変量時系列の相互関係を抽出する手法を開発した。また、時空間モデルを応用し、スロースリップイベントなどの解析を行った。
構造的データについては連想記憶モデルの相転移現象をデータ駆動で抽出することに成功した。

研究成果の概要（英文）：We developed a kernel-based method that extracts common features from multi-modal data based on mutual information. We also developed a transfer learning algorithm that is based on information geometrical e-mixture.
We developed a method to extract mutual relation among multivariate time series such as neuron networks and share-sales data. We applied spatio-temporal data analysis to slow slip event.
We succeeded to extract phase transition phenomena of associative memory neural network model in a data-driven manner.

研究分野：機械学習

キーワード：機械学習 多変量時系列 時空間モデル

1. 研究開始当初の背景

近年の計測技術の発展によって大量の高次元データが得られるようになってきている。そうしたデータのモデリングには従来単純な線形モデルが広く用いられてきたが、現実のデータは複雑で非線形なプロセスを経て観測されるため、線形モデルでは本質を捉えきれないことが多い(図1)。一方、非線形性

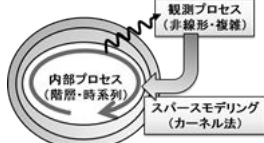


図1: 複雑・非線形な観測のモデリングの枠組み

とスパース性を扱うためのデータ解析手法としてカーネル法が提案され、様々な分野で応用されつつある。しかしながら、現状では観測データがもつ複雑性を十分反映するものにはなっていない。

2. 研究の目的

本研究課題では、カーネル法によるスパースモデリングを拡張することにより、マルチモーダルデータや時系列データ、さらに階層的・論理的な構造が内在するデータなどから複雑な構造を抽出することのできる新たなスパースモデリング手法の確立を目的とする。

3. 研究の方法

本計画研究では、カーネル法を軸として、データに内在する非線形性・階層性など複雑な構造を抽出する手法を開発し、高次元データ駆動科学の基盤となる枠組みの構築を目指す。具体的に、今後5年間で生物・地学データの解析において重要と考えられる三つの課題に着手する(図2)。

【課題1】マルチモーダルスパースモデリング: 複数のモダリティを持つ計測データに共通して含まれる情報をカーネル法によって非線形に抽出するための手法を確立する。

【課題2】ダイナミカルスパースモデリング: 時系列データに内在する非線形な潜在構造をカーネル法によって抽出する手法を開発する。

【課題3】構造的スパースモデリング: 階層構造や論理構造として表現される構造的な

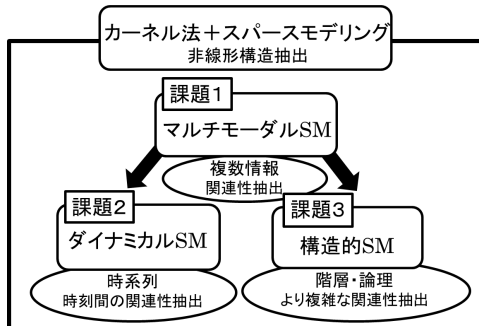


図2: 課題間の関係図

スパース性を抽出するための全く新しいカーネル法を開発する。

4. 研究成果

(1) 複数情報源から共通情報抽出法に対し、カーネル正準相関分析法を改良した手法を開発した。具体的には、成分の共通性を相関係数から相互情報量に拡張した。さらに、非線形性で自由度が向上した分、入力情報との関係性が小さくなるような成分が抽出されることを見出し、入力情報と抽出情報の間の相互情報量を加えることで有用な知識の抽出に成功した(図3)。また、相互情報量は計算上の困難を伴うが、quantile に基づくノンパラメトリックな推定法を採用することによってその問題を解決した。また、最適化計算にはMCMCを用いて局所解からの脱出を図る方法を提案した。

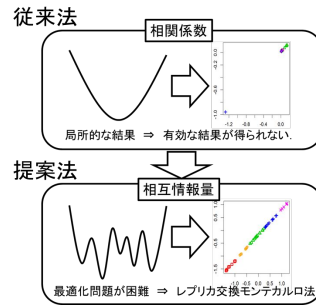


図3: 複数情報源からの情報抽出

(2) 複数のタスク間を統合してデータ不足の問題を解決する手段として、転移学習が注目されているが、ノンパラメトリックな確率分布を情報幾何学的な概念である e-混合と呼ばれる統合の方法を開発した(図4)。連続空間中の経験分布は陽に指数型分布族の元として書くことが困難であったが、e-混合を特徴づける方程式を用いて暗に定義することによって定式化を行った。また、そのために座標系を用いた射影を書き下すことも困難であったが、quantile に基づくダイバージェンスの推定法を援用することにより、ピタゴラスの定理を利用した座標値が不要な学習アルゴリズムを構成することに成功した。

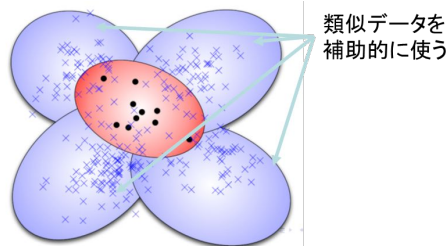
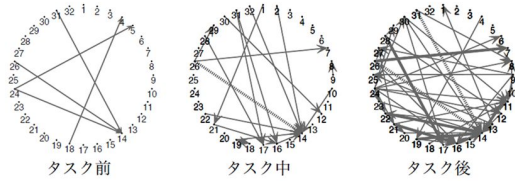


図4: e-混合による転移学習

(3) 脳活動の多点電極測定のような、高次元の多変量時系列の変量間の依存関係を抽出する手法を開発した。このような計測では、

直接の依存関係がないにも関わらず時間を経るに従って情報が伝播して統計的な相関が現れる．これを擬相関と呼ぶが、直接の結合関係を推定する際には障害となる．また、すべての細胞活動を計測しているわけではないので、計測外細胞からの影響もある．提案手法ではこれらの妨害因子を除去する統計量を考案し、人工データや実データでの関係性抽出に成功した（図 5）．



タスク後のグラフはタスク中のグラフと共通するエッジが多い。

図 5：ラットの脳活動からの関係性抽出

(4) 同じく多変量時系列ではあるが、シェアデータのようにある程度長い期間にわたる統計データであって、しかも比率データが与えられたとき、シェアの時系列から相互関係を抽出する方法を開発した．この場合、各シェアはマイクロなタイムスケールの推移行列の定常分布列だとみなし、その推移行列の推定問題として定式化した．それだけでは条件が足りないので、推移行列の変化は小さいと改訂したうえで、fused lasso と同様の差分の L1 正則化によってシェアデータの分析が可能になった．具体的には自動車メーカーの関係性やリコール等のイベントの影響が明らかになった（図 6）．

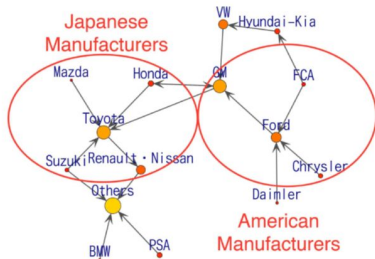


Fig 5. The estimated directed graph for "2007-1Q".
doi:10.1371/journal.pone.0169981.g005

図 6：自動車シェアからの推移行列抽出

(5) 地球科学においてスロースリップイベント (SSE) と呼ばれる通常の地震よりもゆっくりと起きる事象が注目されている．我々は、スイッチング時空間モデル解析により、SSE の開始点と終了点を精度よく推定することに成功した（図 7）．

(6) 高次元時系列データでは、時間方向と空間方向両方での次元縮約が必要となる．そこで、我々はベイズ推定の観点から確率的なモデル化を行い、動的因子分析モデルとして提案した．空間方向には基底関数展開、時間方向には AR モデルによる次元圧縮を行い、過学習を避けるために正則化項を導入した．地

表でのオゾン濃度データに適用し、オゾン分布の変化の時空間モデルを解析し、その要因を明らかにした（図 8）．

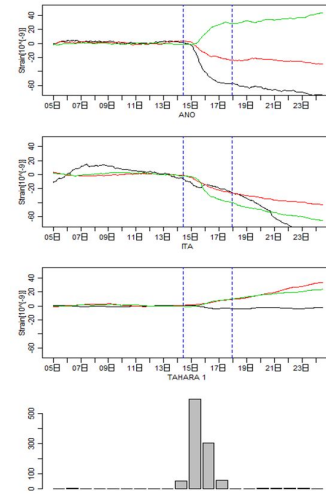


図 7：SSE の抽出結果

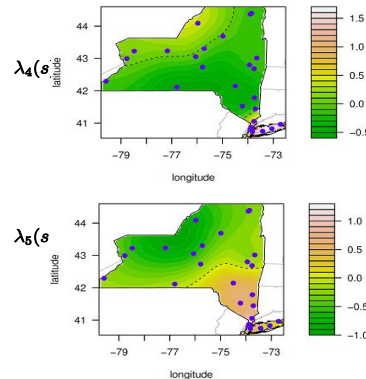


図 8：オゾンの時空間因子

(7) 連想記憶モデルと呼ばれる神経回路モデルは、温度パラメータによって記憶の想起が個別の事象から抽象化された概念まで階層上に変化することが知られている．その構造をデータ駆動で解析するために、モデルシミュレーションを行ったデータから、自動的に構造抽出するアルゴリズムを適用し、理論的な相転移と整合性が取れていることを確認した（図 9）．データから未知のデータ構造を抽出することは非常にチャレンジングな課題であり、今後のさらなる活用が期待できる．

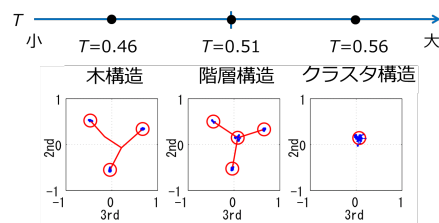


図 9：連想記憶想起の構造抽出

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 14 件)

麻生英樹, "多層ニューラルネットワークによる深層表現の学習", 人工知能学会誌, 28, 649-659, 2013

A. Noda, H. Hino, M. Tatsuno, S. Akaho, N. Murata, "Intrinsic Graph Structure Estimation Using Graph Laplacian", Neural Computation, 26, 1455-1483, 2014
10.1162/NECO_a_00603

K. Wakasugi, T. Kuwatani, K. Nagata, H. Asoh, M. Okada, "Verification of Effectiveness of a Probabilistic Algorithms for Latent Structure Extraction Using an Associative Memory Model", Journal of Physical Society of Japan, 83, 104801, 2014
10.7566/JPSJ.83.104801

T. Araki, K. Ikeda, S. Akaho, "An efficient sampling algorithm with adaptations for Bayesian variable selection", Neural Networks, 61, 22-31, 2015
10.1016/j.neunet.2014.09.010

K. Takano, H. Hino, S. Akaho, N. Murata, "Nonparametric e-Mixture Estimation", Neural Computation, 28, 2687-2725, 2016
10.1162/NECO_a_00888

T. Chiba, H. Hino, S. Akaho, N. Murata, "Time-Varying Transition Probability Matrix Estimation and Its Application to Brand Share Analysis", PLOS ONE, 12, e0169981, 2016
10.1371/journal.pone.0169981

日野英逸, "時間遷移のスパース性 マーケットシェアの遷移を捉えるスパースモデリング", 岩波データサイエンス, 5, 80-90, 2017

H. Hino, J. Fujiki, S. Akaho, N. Murata "Local Intrinsic Dimension Estimation by Generalized Linear Modeling" Neural Computation, 29(7), 1838-1878, 2017
10.1162/NECO_a_00969

T. Kato, H. Hino, N. Murata "Double sparsity for multi-frame super resolution" Neurocomputing, 240, 115-126, 2017
10.1016/j.neucom.2017.02.043

R. Nakata, H. Hino, T. Kuwatani, S. Yoshioka, M. Okada, T. Hori "Discontinuous boundaries of slow slip events beneath the Bungo Channel, southwest Japan" Scientific Reports, 7, 6129, 2017
10.1038/s41598-017-06185-0

H. Hino "Intrinsic Dimension Estimation with R" R Journal, 9(2), 329-341, 2017

T. Ueno, H. Hino, A. Hashimoto, Y. Takeichi, M. Sawada, K. Ono "Adaptive design of an X-ray magnetic circular dichroism spectroscopy experiment with Gaussian process modelling" npj Computational Materials, 4, 1-8, 2018
10.1038/s41524-017-0057-4

Takamitsu Araki, Tadafumi Ochi, Norio Matsumoto, Shotaro Akaho, "Simultaneous Estimation of the Spatio-Temporal Slip Distribution and Duration of the Slow Slip Event by a Switching Model", Journal of Signal Processing, 21(6), 297-308, 2017

Kenta Yoshida, Tatsu Kuwatani, Takao Hirashima, Hikaru Iwamori, Shotaro Akaho, "Progressive evolution of whole-rock composition during metamorphism revealed by multivariate statistical analyses", Journal of Metamorphic Geology, 36(1), 41-54 (2018)

[学会発表](計 10 件)

赤穂昭太郎 機械学習におけるコストをめぐる話題, 電子情報通信学会 IBISML研究会 (招待講演)

H. Hino, A. Noda, M. Tatsuno, S. Akaho, N. Murata, "An Algorithm for Directed Graph Structure Estimation", Lecture Notes in Computer Science (ICANN2014), 8681, 145-152, 2014
10.1007/978-3-319-11179-7_19

T. Araki, H. Hino, S. Akaho, "A kernel method to extract common features based on mutual information", Lecture Notes in Computer Science (ICONIP2014), 8835, 26-34, 2014
10.1007/978-3-319-12640-1_4

H. Asoh, I. Kobayashi, "Zero-Shot Learning of Language Models for Describing

Human Actions Based on Semantic Compositionality of Actions", Proc. of the 28th Pacific Asia Conference on Language, Information and Computing, 1, 12-14, 2014

H. Suetani, "Macroscopic reliability of chaos in recurrent neural networks: An approach based on canonical correlation analysis", Proc. of the 21th International Symposium on Artificial Life and Robotics, 1, 545-548, 2016

S. Akaho, S. Higuchi, T. Iwasaki, H. Hino, M. Tatsuno, N. Murata, "Graph structure modeling for mult-neural spike data", Journal of Physics: Conference Series, 699, 12012, 2016
10.1088/1742-6596/699/1/012012

H. Hino, K. Takano, S. Akaho, N. Murata, "Non-parametric e-mixture of Density Functions", Lecture Notes in Computer Science, 9948, 3-10, 2016
10.1007/978-3-319-46672-9_1

H. Hino, S. Akaho, N. Murata, "An Entropy Estimator Based on Polynomial Regression with Poisson Error Structure", Lecture Notes in Computer Science, 9948, 11-19, 2016
10.1007/978-3-319-46672-9_2

T. Kato, H. Hino, N. Murata
"Doubly Sparse Structure In Image Super Resolution"
Machine Learning for Signal Processing (MLSP) 2016 IEEE 26th International Workshop on, 1-6, 2016
10.1109/MLSP.2016.7738902

Hiromichi Suetani, "Assimilating Nonlinear Dynamics with Force-Learning : a Perspective from Chaotic Synchronization", Proceedings of The 2017 International Symposium on Nonlinear Theory and Its Applications (NOLTA 2017), pp.268-270 (2017)

6. 研究組織

(1) 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・研究グループ長
研究者番号：40356340

(2) 研究分担者

麻生 英樹 (ASOH, Hideki)
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・副研究センター長
研究者番号：10344194

(3) 研究分担者

日野 英逸 (HINO, Hideitsu)
筑波大学・システム情報系・准教授
研究者番号：10580079

(4) 研究分担者

末谷 大道 (SUETANI, Hiromichi)
大分大学・理工学部・教授
研究者番号：40507167