

平成30年6月20日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2013～2017

課題番号：25730013

研究課題名(和文)大規模データにおける構造的推定手法の理論と応用

研究課題名(英文) Theories of structured estimation methods for large scale data and their applications

研究代表者

鈴木 大慈 (SUZUKI, Taiji)

東京大学・大学院情報理工学系研究科・准教授

研究者番号：60551372

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：近年あらゆる分野において扱うデータは急速に大規模化している。また同時に、巨大なデータは内在的に多様な構造を有するようになってきている。そのような多量かつ多様なデータを扱うため、構造的スパース性に着目し、それを利用した推定方法を体系的に提案・分析した。そのため、構造的正則化学習と呼ばれる学習方法を高速に解くための確率的最適化手法として、確率的交互方向乗数法を提案した。また、テンソルモデリングと呼ばれているデータ間の関係性を記述するモデルを考察し、ミニマックス最適性と呼ばれる統計的な最適性を満たす推定量を構築した。以上の研究トピックを通して理論から応用まで包括的な研究を行った。

研究成果の概要(英文)：Recently, the size of dataset is getting larger and larger in several areas. Moreover, such data often contains various structures. To deal with such a complicated and large data, we have focused on structured sparsity and developed new estimation methods and computational methods in a comprehensive manner. Specifically, we have proposed a new stochastic optimization method called stochastic alternating direction method of multipliers that work efficiently for structured regularization methods. We also studied so called tensor modeling and proposed some estimators that satisfy mini-max optimality. Through the above mentioned problems, we have studied theories and applications in a comprehensive manner.

研究分野：機械学習

キーワード：構造的正則化 テンソルモデリング ベイズ推定 再生核ヒルベルト空間 ガウシアンプロセス 確率的最適化 高次元統計 統計的学習理論

1. 研究開始当初の背景

計算機の発達やデータの観測技術の発展とともに、データ解析において我々が扱うデータの規模は増大の一途をたどっている。ここで、データの規模といったとき、それには高次元性と多量性の二種類の意味がある。まず、高次元性は一つの観測値の次元が非常に高いことを意味する。例えば DNA シークエンサからは一つの塩基配列につき数万次元というデータが得られる。一方、多量性とは観測の量が大量であることを意味する。例えば、コンピュータビジョンにおいては、数百万枚といった画像データを扱うことが珍しくない。このような二つの意味での大規模なデータをいかに扱うかはデータに関係する諸科学に共通な問題となっている。産業的な応用でいえば、コンピュータビジョン・音声認識・自然言語処理などが、理学的な応用では生物学・医学・物理学など、広範囲の応用が挙げられる。さらに、データの規模に加えて、様々な局面で多様なデータ構造が扱うニーズがある。例えば、ベクトルを推定するのではなく行列を推定する問題や、説明変数にグループ構造を仮定した推定問題などが重要になりつつある。そうした大規模かつ多様なデータからいかにして有用な情報を引き出すかは、現代社会において大きな課題であり、データサイエンスとしての新たな学問領域を開きつつある。

高次元データ解析の標準的なアプローチは、スパース性を利用する方法である。これは、高次元データには無駄な情報も多く含まれているため、それらをなるべく排除しよう（スパースにしよう）という考え方である。スパース推定の最も単純な方法として L1 正則化を用いた方法が重点的に研究されてきた。しかし、データの多様化にともない、データの構造を利用したデータ解析を行うため、L1 正則化以外の構造的なスパース性がより重要度を増している。実際、構造的スパース性は様々な文脈で現れ、例えば、低ランク行列推定やグループスパース推定、グラフ型正則化を伴った推定などが様々な場面で必要とされてきている。構造的正則化学習については、二つのアプローチが考えられる。一つは正則化付きリスク最小化、もう一つはベイズ推定である。正則化付きリスク最小化は、スパース推定の標準的な方法として広く研究が進められてきた。しかし、ベイズ推定を用いたスパース推定は、いくつかの例で理論的に（正則化付きリスク最小化と比べて）より少ない条件で好ましい性能を示すことが示されている。しかし、これら二つの方法は、どちらがどのような場面で有効であるかはまだ未解明な部分が多く、理論・実験の両面から比較し、その良し悪しを明確にすることは重要な課題である。

このような構造的スパース学習において、種々の学習方法の統計的性質の解明をはじめ、その理論的知見をもとにした新しい学習

方法の構築、さらにはその効率的計算手法の構築が望まれている。

2. 研究の目的

多量かつ多様なデータを扱うため、構造的スパース性に着目し、それを利用した推定方法を体系的に提案・分析する。本研究では、主として正則化付きリスク最小化とベイズ推定の枠組みを考察し、それらに統計的視点からの統一的な理論解析を与え最適性を議論すると同時に、大規模データに対してもスケラブルで効率的な推定アルゴリズムの開発を行う。理論からアルゴリズムまで体系的な研究を行うことで、包括的な視点を与えることが目的である。

3. 研究の方法

理論研究とアルゴリズムの開発を同時並行的または相補的に行ってゆく。理論研究においては、正則化付きリスク最小化とベイズ推定ともに対象とする問題を限定して研究を進め、得られた理論的知見をもとに各手法の長所短所を明確にしてゆく。研究の初めはアルゴリズム研究は正則化付きリスク最小化に絞りを絞る、なるべく汎用的で大規模データに適用可能な手法を開発してゆく。そこで得られた知見をもとに、理論研究・アルゴリズム開発とともに、関連する問題設定でより包括的な議論を進める。理論においては最適性の議論を通して、いかなる正則化関数および事前分布がある種の最適性を満たすかを解明する。アルゴリズム研究では、理論同様に一般的な設定で研究を進め、特にベイズ推定の効率的推定方法についても開発を進める。出来上がった提案手法は実問題を通してその有効性を検証してゆく。

4. 研究成果

(1) 大規模データにおける構造的な正則化の効率的計算手法

大量データの処理に有用な確率的最適化の枠組みで構造的な正則化を扱える手法を考察した [6]。提案手法は基本的に交互方向乗数法 (Alternating Direction Multiplier Method, ADMM) と呼ばれる手法に基づいており、それらのある種の確率的最適化への拡張と言える。提案手法は上に述べたような様々な応用に容易に適用可能であり、また実装の簡便さも重要な特徴である。確率的交互方向乗数法は主に分けて二つのタイプがある。それぞれオンライン型確率的最適化法 [25] と確率的双対座標降下法 [23] に対応し、それらの ADMM 型拡張が提案手法である。上記の二つの手法はともに大規模データにおける構造的な正則化学習問題を解くための一般的な枠組みを提供し、数多くの正則化項に適用可能であるだけでなく、実装が簡便なため、非専門家でも容易に扱うことができる。

オンライン型手法は従来の標準的な確率

的勾配降下法と比べ、適用範囲が広いにも関わらず同様の収束レートを達成することが理論的に示された [25] .

また、第二の確率的双対座標降下法の ADMM 拡張については、研究期間初期においては一次収束まで示したが、その収束速度を改善できることがわかった。すなわち、当初得られていた収束速度は問題の条件数によって決定されていたが、実はこの条件数への依存性を緩くすることができ、条件数そのものではなくその平方根が収束レートに現れることが示された [32] . これによって、悪条件の問題でもパラメータの設定を適切に行えば、通常の一次最適化法と比べて高速にできることが示されたことになる。示されたレートは一次最適化法の Nesterov の加速法と同じ収束レートになる。

さらに、この方法の欠点として、線形変換後のパラメータの次元が高くなりがちであるという点があるが、毎回の更新でサンプルの一部の観測と、パラメータの一部の更新で済む手法を考案した。この手法は毎回の更新がこれまでの方法よりも単純であるという点で特徴的である。さらに、理論的に収束レートを導出し、上記手法と比べて改善されることがわかった。

上記の確率的交互方向乗数法以外にもいくつかの確率的最適化手法を提案した。まずは、凸関数から凸関数を引いた形の目的関数を最小化する問題を確率的最適化によって効率的に解く手法を考案した [16] . 凸関数から凸関数を引いた形の目的関数を最小化する問題は DC 計画とよばれ、非凸関数の最適化問題になるが、応用上様々な場面で現れる最適化問題である。提案手法は単なる確率的勾配降下法よりも良い収束性能を示し、制限ボルツマンマシン等への応用によって他の学習手法よりも応用上良い性能を示すことが実験的にも示された。

さらに、確率的分散縮小勾配法と呼ばれる手法を改善し、ミニバッチサイズに対して従来法よりも計算量の増加が緩やかな手法を提案した [14] . これによって、各更新を並列計算することで少ない反復で最適解に収束させることができる。並列化の恩恵をより良く受けられるという点で大規模データ解析により適した手法である。これは、分散縮小勾配法と双対平均化法および Nesterov の加速法を組み合わせた手法である。

(2) 構造的スパース学習の学習理論

(i) 低ランクテンソル推定の解析

構造的正則化の中でも低ランクテンソル型正則化手法に着目し、その統計的性質を調べた。低ランクテンソル型正則化として、テンソルを各モードで行列化しそれぞれのトレースノルムを足し合わせた和型正則化と、

テンソルを各モードに対応したテンソルの和に分解し各成分のトレースノルムの和を用いる畳みこみ型正則化の二つを考えた。特に、今年度は畳みこみ型正則化を重点的に研究し、テンソルのランクがモードごとに偏っている場合には畳みこみ型正則化が有用であることが示された [24] .

しかし、上記の凸最適化を用いたテンソル推定手法の予測誤差を理論的に評価するためにはデザイン行列にある種の強凸性を仮定する必要がある。そこで、ベイズ推定を用いて推定することにより、これらの制約を取り除くことができることを理論的に示すことに成功した [22,26] . さらにテンソル推定問題における予測誤差のミニマックスレートを導出し、ベイズ法による低ランクテンソル推定手法がそのミニマックスレートを達成することが示された。また、理論のみでなく、実際のデータにおいても計算機実験を行い、既存の凸最適化を用いた手法と比べて実験的にも良い推定精度を有することが示された。

上記の研究はパラメトリックなテンソルモデルであったが、それを拡張して、ノンパラメトリックなテンソル型関数推定も考察した。これにより、これまでモーダル間の線形な関係性までしか捉えられなかったことに対して、モーダル間の非線形な関係性までも捉えられるようになった。推定手法として、ノンパラメトリックテンソルをガウシアンプロセス事前分布を用いてベイズ推定する手法を提案した [18] . パラメトリックな方法と同様に統計的に非常に弱い条件のもと、ミニマックス最適性をほぼ達成することが理論的に証明された。さらに、別の推定方法として交互最適化法を考察し、その汎化誤差の理論評価をした [17] . 交互最適化法は各モードに対応する非線形関数を再生核ヒルベルト空間の中で交互に最適化する。理論的解析により、適切な初期値から始めることで、ミニマックス最適な解に線形収束することが示された。理論には、これまでの研究で調べてきたベイズ推定においては必要としなかった仮定が必要であったが、一方でアルゴリズムの線形収束といったより強い結果が得られている。また、大規模数値実験によって、従来法との比較をし、交互最適化法が実用的にはベイズ法と同程度の性能を示すことが確認された。

(ii) その他の構造的正則化学習

テンソル推定以外に、様々な正則化学習法の統計理論の研究を行った。高次元グラフィカルモデルの推定 [19,20] , 高次元グラフィカルモデルの変化点検出 [3,4,21] , 深層学習の汎化誤差解析 [11] , 多変量回帰のモデル選択 [2] , Multiple Kernel 学習の汎化誤差解析 [1] , 非凸正則化を用いたスパース推定手法の提案と理論 [13] といった研究成果

を挙げることができた。

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 27 件)

【他査読付き国際ジャーナル論文】

- [1] Taiji Suzuki: Fast Learning Rate of Non-Sparse Multiple Kernel Learning and Optimal Regularization Strategies. *Electronic Journal of Statistics*, accepted.
- [2] Yuichi Mori and Taiji Suzuki: Generalized ridge estimator and model selection criteria in multivariate linear regression. *Journal of Multivariate Analysis*, 165:243–261, 2018.
- [3] Song Liu, Taiji Suzuki, Relator Raissa, Jun Sese, Masashi Sugiyama, and Kenji Fukumizu: Support Consistency of Direct Sparse-Change Learning in Markov Networks. *The Annals of Statistics*, 45(3):959–990, 2017.
- [4] Song Liu, Kenji Fukumizu and Taiji Suzuki: Learning Sparse Structural Changes in High-dimensional Markov Networks: A Review on Methodologies and Theories. *Behaviormetrika*. 44(1):265–286, 2017.
- [5] Yoshito Hirata, Kai Morino, Taiji Suzuki, Qian Guo, Hiroshi Fukuhara, and Kazuyuki Aihara: System Identification and Parameter Estimation in Mathematical Medicine: Examples Demonstrated for Prostate Cancer. *Quantitative Biology*, 4(1):13–19, 2016
- [6] Taiji Suzuki: Stochastic Alternating Direction Method of Multipliers for Structured Regularization. *Journal of Japan Society of Computational Statistics*, 28:105–124, 2015.
- [7] Taiji Suzuki, and Kazuyuki Aihara: Nonlinear System Identification for Prostate Cancer and Optimality of Intermittent Androgen Suppression Therapy. *Mathematical Biosciences*, 245(1):40–48, 2013.
- [8] Taiji Suzuki, and Masashi Sugiyama: Fast learning rate of multiple kernel learning: trade-off between sparsity and smoothness. *The Annals of Statistics*, 41(3):1381–1405, 2013.
- [9] Taiji Suzuki: Improvement of Multiple Kernel Learning using Adaptively Weighted Regularization. *JSIAM Letters*, 5:49–52, 2013.
- [10] Masashi Sugiyama, Takafumi Kanamori, Taiji Suzuki, M. C. du Plessis, Song Liu, Ichiro Takeuchi: Density Difference

Estimation. *Neural Computation*, 25(10): 2734–2775, 2013.

【他査読付き国際会議論文】

- [11] Taiji Suzuki: Fast generalization error bound of deep learning from a kernel perspective. *AISTATS2018, Proceedings of Machine Learning Research*, 84:1397–1406, 2018.
- [12] Atsushi Nitanda and Taiji Suzuki: Gradient Layer: Enhancing the Convergence of Adversarial Training for Generative Models. *AISTATS2018, Proceedings of Machine Learning Research*, 84:1008–1016, 2018.
- [13] Masaaki Takada, Taiji Suzuki, and Hironori Fujisawa: Independently Interpretable Lasso: A New Regularizer for Sparse Regression with Uncorrelated Variables. *AISTATS2018, Proceedings of Machine Learning Research*, 84:454–463, 2018.
- [14] Tomoya Murata and Taiji Suzuki: Doubly Accelerated Stochastic Variance Reduced Dual Averaging Method for Regularized Empirical Risk Minimization. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pp. 608–617, 2017.
- [15] Song Liu, Akiko Takeda, Taiji Suzuki and Kenji Fukumizu: Trimmed Density Ratio Estimation. *Advances in Neural Information Processing Systems 30 (NIPS2017)*, pp.4521–4531, 2017.
- [16] Atsushi Nitanda and Taiji Suzuki: Stochastic Difference of Convex Algorithm and its Application to Training Deep Boltzmann Machines. *The 20th International Conference on Artificial Intelligence and Statistics (AISTATS2017), Proceedings of Machine Learning Research*, 54:470–478, 2017.
- [17] Taiji Suzuki, Heishiro Kanagawa, Hayato Kobayashi, Nobuyuki Shimizu, and Yukihiro Tagami: Minimax Optimal Alternating Minimization for Kernel Nonparametric Tensor Learning. *The 30th Annual Conference on Neural Information Processing Systems (NIPS2016)*, pp. 3783–3791, 2016.
- [18] Heishiro Kanagawa, Taiji Suzuki, Hayato Kobayashi, Nobuyuki Shimizu, and Yukihiro Tagami: Gaussian process nonparametric tensor estimator and its minimax optimality. *The 33rd International Conference on Machine Learning (ICML2016), Proceedings of Machine Learning Research* 48:pp. 1632–1641, 2016.
- [19] Song Liu, Taiji Suzuki, Masashi

Sugiyama, and Kenji Fukumizu: Structure Learning of Partitioned Markov Networks. International Conference on Machine Learning (ICML2016), Proceedings of Machine Learning Research, 48: pp.439-448, 2016.

[20] Satoshi Hara, Tetsuro Morimura, Toshihiro Takahashi, Hiroki Yanagisawa, Taiji Suzuki: A Consistent Method for Graph Based Anomaly Localization. The 18th International Conference on Artificial Intelligence and Statistics (AISTATS2015), JMLR Workshop and Conference Proceedings, 38: pp.333--341, 2015.

[21] Song Liu, Taiji Suzuki, and Masashi Sugiyama: Support Consistency of Direct Sparse-Change Learning in Markov Networks. The Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI2015), 2015.

[22] Taiji Suzuki: Convergence rate of Bayesian tensor estimator and its minimax optimality. The 32nd International Conference on Machine Learning (ICML2015), JMLR Workshop and Conference Proceedings 37:pp. 1273--1282, 2015.

[23] Taiji Suzuki: Stochastic Dual Coordinate Ascent with Alternating Direction Method of Multipliers. International Conference on Machine Learning (ICML2014), JMLR Workshop and Conference Proceedings 32(1):736--744, 2014.

[24] Ryota Tomioka, and Taiji Suzuki: Convex Tensor Decomposition via Structured Schatten Norm Regularization. Advances in Neural Information Processing Systems (NIPS2013), 1331--1339, 2013.

[25] Taiji Suzuki: Dual Averaging and Proximal Gradient Descent for Online Alternating Direction Multiplier Method. International Conference on Machine Learning (ICML2013), 2013, JMLR Workshop and Conference Proceedings 28(1): 392--400, 2013.

【日本語論文】

[26] 鈴木大慈: ベイズ法を用いた高次元テンソル推定. 『日本応用数理学会論文誌』, 27(3), 7-14, 2017.

[27] 鈴木大慈: 統計的学習理論概説. 『日本応用数理学会論文誌』, 23(3), 537-561, 2013.

〔学会発表〕(計 50 件)

[28] 鈴木大慈: 「機械学習・人工知能における数学の役割」 2018 年度数学教育学会春季年会, 総合講演 1. 2018 年 3 月 19 日. 東京大学駒場キャンパス.

[29] 鈴木大慈: 「人工知能・機械学習における課題, 数学の役割と期待について」 日

本数学会 2018 年度年会, 数学連携ワークショップ「Society 5.0 と数学---量子コンピュータと人工知能を題材に---」. 2018 年 3 月 19 日. 東京大学駒場キャンパス.

[30] 鈴木大慈: 「機械学習技術の進展とその数理基盤」 数理システムユーザーコンファレンス 2017, 基調講演. 2017 年 11 月 2 日. 東京コンファレンスセンター・品川.

[31] Taiji Suzuki: Generalization error analysis of deep learning and its application to network structure determination. French-Japanese Workshop on Deep Learning and Artificial Intelligence. October 25th 2017. Lycée Henri IV, Paris, France.

[32] 鈴木大慈: 「構造のある機械学習問題における最適化技法」 第 29 回 RAMP シンポジウム, 企画セッション「機械学習と最適化」. 2017 年 10 月 12 日~2017 年 10 月 13 日. 筑波大学春日講堂. (10/12 発表)

[33] 鈴木大慈: 低ランクテンソルの学習理論と計算理論. IBIS2016. 2016 年 11 月 16 日. 京都大学.

[34] 鈴木大慈: 統計・計算理論で広がる機械学習. 統計関連学会連合大会チュートリアル講演. 2016 年 9 月 4 日. 金沢大学. Taiji Suzuki: Stochastic Optimization. 機械学習サマースクール (MLSS2015, Kyoto), 2015. Kyoto, Japan. 23/Aug-4/Sep, 2015 (presented in 2-4/Sep/2015).

[35] 鈴木大慈: マルチプルカーネル学習とスパース推定の統計的性質. 日本数学会年会, 特別講演. 2014 年 3 月 17 日. 学習院大学, 口頭発表.

その他

〔図書〕(計 4 件)

[36] 金森敬文, 鈴木大慈, 竹内一郎, 佐藤一誠: 機械学習のための連続最適化 (機械学習プロフェッショナルシリーズ). 講談社, 2016 年 12 月 7 日.

[37] 鈴木大慈: 確率的最適化 (機械学習プロフェッショナルシリーズ). 講談社, 2015 年 8 月 8 日.

[38] 合原一幸(編著), 占部千由, 近江崇宏, 阿部力也, 長谷川幹雄, 鈴木大慈(著): 暮らしを変える驚きの数理工学 (ウェッジ選書 53). ウェッジ, 2015 年 5 月 25 日.

[39] Trevor Hastie, Robert Tibshirani, Jerome Friedman (原著), 杉山将, 井手剛, 神鳥敏弘, 栗田多喜夫, 前田英作(編), 井尻善久, 井手剛, 岩田具治, 金森敬文, 兼村厚範, 烏山昌幸, 河原吉伸, 木村昭悟, 小西嘉典, 酒井智弥, 鈴木大慈, 竹内一郎, 玉木徹, 出口大輔, 冨岡亮太, 波部斉, 前田新一, 持橋大地, 山田誠(訳): 統計的学習の基礎: データマイニング・推論・予測. 共立出版, 2014.

〔産業財産権〕

なし。

〔その他〕

ホームページ等

<http://ibis.t.u-tokyo.ac.jp/suzuki/>

6．研究組織

(1)研究代表者

鈴木 大慈 (SUZUKI, Taiji)

東京大学・大学院情報理工学系研究科・准教授

研究者番号：60551372