

研 究 報 告 書

「統合的統計モデリングの数理基盤と方法論」

研究タイプ: 通常型

研究期間: 平成26年10月～平成30年3月

研 究 者: 鈴木 大慈

1. 研究のねらい

本研究の目的は「モデリングのためのモデリング手法」を確立し、それに必要な数理基盤を整備することである。現在、ビッグデータ時代を迎え、利用可能なデータが爆発的に増加している中、急激な社会構造の変化および環境・エネルギー問題・顧客動向への迅速な対応といった数々の社会的問題をデータに裏付けられた定量的な意思決定によって解決するというニーズが高まっている。そのような社会的背景の中、いわゆるデータ科学が現在大きな注目を集め広く使われるようになってきており、それはまた同時に、データ解析においてその根幹をなすモデリング技術もその重要性を増していることを意味する。しかし、モデリングの技術は定まった方法論がなくデータごとに試行錯誤を繰り返して適切なモデルが選ばれていることが現状である。本研究課題では、データ解析現場において個別になされているモデリング作業に共通な、ある種の規則的構造を利用して、これまでは別個に扱われていた複雑な問題をまとめあげるための方法論を構築することが目的である。それによって、今まで人間が行ってきたモデリングの部分のある程度データから推論することを可能にさせる。この研究課題を通して、データ科学の応用におけるモデリングに数学的知見による新しいパラダイムをもたらすと同時に、数学諸分野へ理論的問題意識の還元を行う。

上記の理念を実現するために、具体的に以下の課題を軸にして研究を進める。研究課題は大きく分けて二つに分けられる: (1) 統合的統計モデリングを実現する各種モデルおよび学習方法に関する統計的学習理論の研究, (2) 統計的学習を効率的に実行するための効率的計算手法の研究。統計的学習理論の研究に関しては、さらに細分して「群対称性」のあるモデル、および「階層構造」のあるモデルを取り扱い、汎化誤差の評価および最適な学習方法の導出を行う。群対称性によって様々なデータ間に横たわる対称性を吸収し、階層構造によりデータ間に共通な因子をまとめる機構を有するモデリングを実現する。計算手法の研究については、確率的最適化を軸として、構造的正則化学習を高速に実行する最適化技法や並列計算技法を開発する。それによって、複雑な学習問題に対しても実装が簡単でかつ高速なアルゴリズムを提供する。

2. 研究成果

(1) 概要

統計的学習理論の研究においては、「群対称性」を有するモデルとして低ランクテンソルモデルを、「階層性」を有するモデルとして深層学習を中心に研究を進めた。低ランクテンソルモデルは様々なデータ間の関係を記述する非常に汎用性の高いモデルであるが、最適性を満たす手法の導出は不完全であり、そのような手法の導出と、最適な汎化誤差の導出が望まれていた。本研究ではそのような方法としてベイズ推定量を考察し、それが CP-分解が可能な

テンソルのクラスにおいてミニマックス最適レート達成を示した。さらに無限次元モデルにも理論を拡張し、ガウス過程事前分布を用いた推定法を考察し、それがやはりミニマックス最適レート達成を示した。また、ベイズ推定量の計算量を改善する形で、交互最適化法を考察し、ベイズ推定量よりも強い条件が必要になるが、良い初期解から始めることでミニマックス最適レート達成する局所最適解に線形収束することが示された。一連の研究は実データでの検証を通してその実用性を確かめた。

深層学習においては、データサイズに比してパラメータ数が大きくても良い汎化誤差を達成することが経験的に知られているが、その理由については理解が進んでおらず問題となっている。本研究では、その一つの解答として「自由度」を用いた汎化誤差の特徴づけを与えた。これにより複数データ・タスクに横たわる本質的な情報が圧縮可能であれば汎化性能を高めることができることがわかる。本理論は、真の関数を積分表現で表し、各中間層に対応するカーネル関数とそれに付随する再生核ヒルベルト空間を定義することによって、その「自由度」を定義し、各層の有限次元近似誤差を満たす横幅を導出することで成り立っている。これは言い換えれば、モデルが圧縮可能であれば高い汎化性能を有するとも言える。この観点から、モデル圧縮への応用も進め、新しい圧縮方法の提案と、実データにおける手法の有用性の実証を行った。

計算理論に関しては、構造のある機械学習手法を解くための確率的最適化法を重点的に発展させた。構造的な正則化を解くための最適化手法として、既存手法より速い収束レートを達成する手法を提案し、グラフ正則化への応用を行った。また、DC計画の確率的最適化手法の提案や、分散縮小確率的勾配法の加速によるミニバッチサイズに対する計算効率を改善させた手法の提案などを実現した。

(2) 詳細

本研究ではデータおよびタスクの「群対称性」と「階層構造」に注目した高度なモデリング手法を支える統計的学習理論および、効率的計算のための「最適化理論」を柱に研究を進めた。

(1) 統計的学習理論:

「群対称性」を満たすモデルとしては特に低ランクテンソルモデルを、「階層構造」を持つモデルとしては特に深層学習を重点的に研究を行った。なお、これらの研究は独立ではなく、密接に関係しており、特に深層学習の研究においては低ランクテンソルモデルの学習理論で培った理論的道具が役に立っている。

・テーマ A: 「群対称性」を有するモデルの学習理論

群対称性を持つモデルとして特に低ランクテンソルの推定理論を重点的に展開した。多様なデータ間の高次の関係性を推定するための低ランクテンソルは重要なモデルである。関係データ解析や推薦システム、時空間解析、データの低次元ベクトル表現およびマルチタスク学習など様々な問題において現れるモデルである。低ランクテンソルモデルで表現される構造は、複数のデータおよびタスクに横たわる本質的に重要な情報源が少数であり、各データおよびタスクの関係はその少数の本質的な情報源を基に記述されるというものである。よって、見かけは多くのデータやタスクでもその少数の情報源を特定することによって少ないデータで効率的な推定が可能になる。しかし、情報源はデータの裏に隠れているため、情報源から観測データの間の関係は一意ではない。その一意性の喪失を、本研究では群対称性という言葉で

表している。本研究では、そのような低ランクテンソル推定について理論的に最適な推定法の導出及びその汎化誤差(予測誤差)の評価を行った。

まず、有限次元モデルを対象として、その最適な推定手法の導出を行った。本研究より前には凸正則化を用いた低ランクテンソル推定が盛んに研究されていた。そのような方法の推定誤差は理論的に導出されていたが、それらの方法はミニマックス最適性と呼ばれる最適性規準を満たさない。そこで、ミニマックス最適性を満たすための推定手法として、ベイズ推定法を考え、それがミニマックス最適性を満たすこと、およびその最適な汎化誤差の導出を行った。しかも、凸最適化による方法よりもかなり緩い条件でミニマックス最適性を満たすことが示された。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、実際に既存手法と比べて良い推定精度を有することが示された。当該研究成果は国際会議の ICML2015 に論文が採択された。

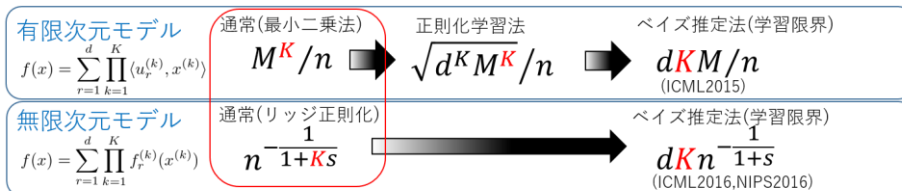
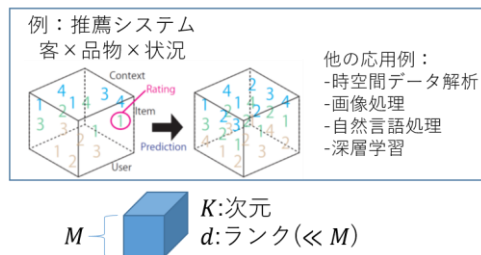
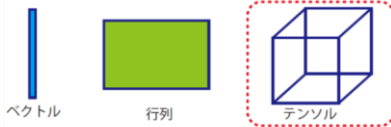
有限次元モデルを拡張して、無限次元モデルを考え、各データ間の非線形な関係も推定できる理論的枠組みを整備した。そのような無限次元モデルを推定するために、ガウス過程事前分布を用いたノンパラメトリックベイズ推定法を考察した。このノンパラメトリックベイズ推定法の汎化誤差を、ガウス過程事前分布の理論を用いて評価した。結果として、真の関数がガウス過程事前分布によって特徴づけられる再生核ヒルベルト空間に含まれていればミニマックス最適性を満たすことが示された。また、真の関数が再生核ヒルベルト空間に入っていない場合でも、補間空間の理論を用いてほぼミニマックス最適性を達成することが示された。当該研究成果は国際会議の ICML2016 に論文が採択された。

上記のベイズ推定法は、統計的性質については優れているが、その計算には多量の計算量を必要とする。そこで、交互最適化法を考察し、その汎化誤差を評価した。交互最適化法は、テンソル分解した各成分を交互に最適化することによって推定をする方法である。これは、非凸な目的関数の上での座標降下法の形をなす。また、モデルとしては、上記のガウス過程事前分布に付随する再生核ヒルベルト空間を用いる。結果として、真の関数に近い場所からアルゴリズムを開始すれば、交互最適化法はミニマックス最適な誤差まで線形収束し、普通の教師ありカーネル学習とほとんど計算量が変わらないことが示された。これらの方法をいくつかの実データへ実際に適用し、その性能を評価したところ、確かにこれらの手法は既存手法に比べて高い性能を持つことが観察された。特に、Yahoo Japan との共同研究を進め、Yahoo shopping のデータを用いて顧客の購買予測に提案手法を適用した。当該研究成果は国際会議の NIPS2016 に論文が採択された。

また、国際シンポジウム The First Korea-Japan Machine Learning Symposium で招待講演をし、国際シンポジウム STM2016 でも発表をした。一連の研究と関連して、2016 年度統計関連学会連合大会において「統計・計算理論で広がる機械学習」というタイトルでチュートリアル講演を行い、また第 11 回日本統計学会研究業績賞を受賞した。

テンソルモデリング

複数のデータ間の関係を学習
→機械学習に頻繁に出現



次元の呪い

・テーマB:「階層構造」を有するモデルの学習理論

「階層構造」を持つ学習機構として深層学習を重点的に研究した。深層学習は現在、ワンショット学習や転移学習およびマルチモーダル学習など複数タスクをまたがる学習問題において威力を発揮しており、本研究課題の統合的統計モデリングにおいて重要な役割を果たす。実は、深層学習は非線形な関数を通した一般化テンソル分解と結び付けて考察することができる。この知見を活かして、テーマAの結果を援用する形で深層学習の汎化誤差とその特徴づけを行った。深層学習においては、パラメータの数がデータより大きくても良く汎化することが観察されており、その理由はいまだ謎とされている。また、訓練誤差が収束した後でも汎化誤差が下がり続けることがあり(汎化ギャップ)、これもまた深層学習の不思議な性質として世界的に議論が盛んになされている。これから述べる本研究の結果は、これら深層学習における本質的に重要な疑問へ対して一つの解答を与えるものである。そのため、深層学習の構造がいかに汎化誤差に影響を与えるかという問題を考える。最適な構造を決定することは応用上も非常に重要な問題である。深層学習の万能近似能力を示すために積分表現が重要であるが、この積分表現を用いると各層に対応する再生核ヒルベルト空間が定義できる。すると、この積分表現を通してカーネル法の学習理論が援用できる。特に、カーネル関数に対応した「自由度」を定義することにより、積分表現で表される真の関数を有限和で近似する際の近似誤差を見積もることができた。さらに、その有限次元近似モデルの中での推定誤差も見積もることで、推定誤差のバイアス-バリエーションの分解を与えることができる。このことより、バイアス-バリエーションのトレードオフを最小化する横幅の特徴づけが可能になった。これらの証明にはこれまで行ってきた無限次元テンソルモデルの理論も利用している。この理論はベイズ推定量および経験誤差最小化の両方に適用することができる。上記の「自由度」は各層に定義されたカーネル関数の固有値の振る舞いによって決まり、固有値が速く減少すればより自由度は小さくなる。この自由度は実質的なモデルの次元と呼べるもので、実際に学習されたネットワークの自由度を求めてみるとパラメータ数よりもはるかに小さくなることが多い。これより、深層学習は自由度を小さくすることにより過学習を防いでいると言うことができる。また、複数データ・タスクに横たわる本質的な情報が圧縮可能であれば汎化性能を高めることができる。そのような自由度を減少させる学習は、例えばDropoutやWeight decayおよび

L2 正則化を利用することで実現できる。

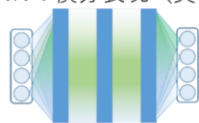
さらに、上記の理論をモデル圧縮に応用し、大きなサイズのネットワークで学習した結果を、汎化誤差を損なうことなく小さいモデルに圧縮するために、自由度を利用して各層ごとに適応的に幅を決める方法を提案した。さらに、その正当性として、経験分布から推定される自由度を訓練誤差一定のもと小さくすることでより良い汎化誤差を得ることができることを示し、モデル圧縮可能性と汎化性能を結び付けた理論を導出した。また、cifar10 等のデータセットで学習した VGG ネットワークおよび Cityscape データセットで学習した SegNet に提案手法を適用し、その有効性を確かめた。すると、パラメータサイズが約半分になっても精度がほとんど落ちないことが数値的に確かめられた。これは実際に学習されたネットワークには多くの小さな固有値が含まれていることによる。なお本研究の遂行にあたり、株式会社 NTT データ数理システムと共同研究を実施した。本研究は、AISTATS2018 に論文が採択された。また、French-Japanese Workshop on Deep Learning and Artificial Intelligence 2017 で招待講演を行った。

深層学習はなぜうまくいくのか？

深層学習の汎化誤差（学習性能）を評価

- 「自由度」という深層ネットワークの構造を表す量が汎化誤差を決めることを解明 → 構造決定の指針を与える：モデル圧縮への応用
- 汎化誤差を導出するための理論的枠組みを整理

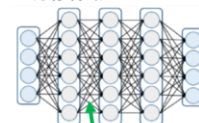
深層NNの積分表現（真の関数）



有限近似

求積法の理論

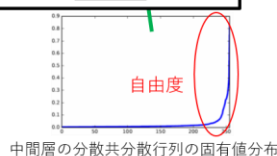
有限次元モデル



深層学習の汎化誤差を決める要素は何か？ → **自由度**が重要

自由度：中間層の独立な隠れ成分の数

$$N_{\ell}(\lambda) = \sum_{j=1}^{\infty} \frac{\mu_j^{(\ell)}}{\mu_j^{(\ell)} + \lambda}$$



(2) 計算理論

テーマ C: 機械学習における大規模最適化の理論:

上記の統計理論に加え、大規模データを用いた機械学習を効率的に実行するための計算方法の研究を進めた。特に、複数の問題を同時に行うためのマルチタスク学習や構造的な正則化を解くための確率的最適化法を重点的に発展させた。確率的最適化においては各更新でサンプルの一部（ミニバッチと呼ぶ）を用いて計算を進め、結果的に全体的に必要な計算の総量を少なくすることが可能である。構造的な正則化は複雑な正則化項をもつ推定問題であるが、この最適化を効率的に実行するための方法として交互方向乗数法がある。交互方向乗数法の確率的最適化手法が提案されていたが、本研究では、既存研究に必要な条件を緩め、さらにその収束レートの解析を改善し、ネステロフの加速法を用いた方法と同じ収束レートを達成することを示した。本手法の応用として、k-劣モジュール関数から誘導されるグラフ正則化を用いた正則化学習に関して効率的な確率的最適化手法を提案した。

また、非凸問題の最適化として凸関数から凸関数を引いた形の DC 計画問題を考察し、その確率最適化手法を提案した。論文は国際会議 AISTATS2017 に採択された。

有限サンプルの確率的最適化においては、分散縮小確率的勾配法が良く用いられているが、その加速法を提案した。これまでの方法は並列計算を用いない限りミニバッチサイズを増やしても理論的には計算効率性が上がるわけではなかったが、提案手法はミニバッチサイズに対して計算効率が向上する手法である。これによって直列計算の計算効率性に加え、並列計算の計算効率も向上させることができる。当該研究は国際会議 NIPS2017 に論文が採録された。

これらの研究成果に基づいて、2015 年には機械学習の国際サマースクールである MLSS2015 にて確率的最適化に関する講義を行い、また統計数理研究所において「統計・機械学習における確率的最適化」というタイトルで一般向けに確率的最適化に関するチュートリアル講演を行った。

3. 今後の展開

本さがけ研究で得られた深層学習の汎化誤差理論は今後も様々なオープンプロブレムに展開できる可能性がある。現在、機械学習界隈では深層学習の応用が急速に進む一方で、その理論研究の重要性も増している。例えば、(a) 訓練誤差が収束した後も汎化誤差が減少を続ける「汎化ギャップ」の現象の解明、(b) 深層学習の大域最適性の問題、(c) 敵対的入力に関する現象の理論的解明とその対処法、といった本質的問題の解決へ本研究の知見は有用な鍵を与えると期待される。また、確率的最適化を中心とした高速な最適化手法の開発を非凸関数まで広げて進めることで、ハードウェアの発展と相まってさらなる深層学習をはじめとする様々な機械学習手法の高速計算が実現できるものと期待される。さらに、さがけ研究で得られた理論的知見を社会実装に繋げるべく、企業との共同研究などを積極的に進め、理論と応用との包括的発展に寄与したいと考えている。

4. 評価

(1) 自己評価

(研究者)

統合的統計モデリングを実現するという目的に沿って、統計的学習理論および計算理論を展開し、機械学習に現れる様々な問題について包括的に研究を進めることができた。群対称性のあるモデルに関しては低ランクテンソルモデルを中心に、有限次元モデルおよび無限次元モデルにおいてミニマックス最適な汎化誤差の下限および、その下限を達成する学習方法の導出ができた。階層構造のあるモデルに関しては、深層学習の理論として、再生核ヒルベルト空間およびそれに付随する自由度の概念を用いて、その汎化誤差の特徴づけを行った。この理論研究は、学習モデルの圧縮といった応用に繋げることができた。この理論はモデルの圧縮だけでなく、その他の様々な問題にも展開できる可能性を持っており、今後の展開の足掛かりを作ることができた。また、これらの理論研究の実用面を実証すべく、企業との共同研究も進め、実用への展開についても一定以上の成果を挙げられたと考えている。

また、最適化技法に関しても確率的最適化を中心として、構造的学習問題の高速解法や、並列化によって従来法より効率を改善できる方法などを提案し、当該領域に貢献できたもの

と考えている。

より実用に近い応用研究を期間内に進めたいところであったが、さがけ研究を通してアドバイザーや同領域研究者から多様な知見を提供していただき、また企業との共同研究の遂行など、研究の幅を大きく広げられたことから、未公開の成果とあわせて今後の発展に繋がる統合的統計モデリングの基礎となる研究が実現できたと考えている。

(2) 研究総括評価(本研究課題について、研究期間中に実施された、年2回の領域会議での評価フィードバックを踏まえつつ、以下の通り、事後評価を行った)。

(研究総括)

データ科学におけるモデリング技法を、従来の個別のデータごとの試行錯誤によるモデル選択から脱却し、より統合的なモデリングの数理基盤と方法論を構築することを目指し、低ランクテンソルモデルの推定理論と深層学習における汎化誤差に関する理論研究などにおいて大きな成果を挙げた。特にミニマックス最適レートを達成する低ランクテンソル推定手法のクラスを見出し、それを実現する手法の一つである交互最適化法の局所最適解への線形収束を証明した。そこで培われた方法を用いて、深層学習の汎化誤差に対して自由度の概念を導入することで、深層学習における基本的な問題への1つの解答を与えることに成功した。これらの成果はいずれも複数のトップレベル国際会議に採録され、また日本統計学会研究業績賞など4件の受賞や国際シンポジウム・サマースクールなどでの基調講演や連続講義などに招待されたことにもみられるとおり、国内外で高い評価を得ており、さがけ研究の成果として申し分のないものであるといえる。

これらにより開発された手法は既に実データに適用して良い効果が検証されており、現実の社会的な問題への本格的な応用実施についてはこれからであるが、今後の展開が大いに期待される。特に社会的課題を捉えて複数の企業と連携し、現場の視点も踏まえた共同研究が行われたことは、本さがけ領域の趣旨にも合致している。

さがけ期間中には「確率的最適化」などの著書の刊行や、さまざまなチュートリアル講義、800人以上の企業参加者への講演などの研究成果の普及活動もめざましく、またアウトリーチ活動や領域内外の研究交流にも積極的に取り組んで、領域の活動にも大きな貢献があった。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

1. Taiji Suzuki: Fast generalization error bound of deep learning from a kernel perspective. AISTATS2018, accepted.
2. 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.
3. 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.

4. 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.

5. 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.

(2)特許出願

研究期間累積件数:0件

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

【招待講演】

1. 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.
2. 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.

【受賞】

鈴木大慈: 第 11 回日本統計学会研究業績賞 (2017 年度). 2017 年 9 月 5 日.

【著書】

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.

【他査読付き雑誌論文】

Taiji Suzuki: Fast Learning Rate of Non-Sparse Multiple Kernel Learning and Optimal Regularization Strategies. Electronic Journal of Statistics, in press.