

研究終了報告書

「深層学習の高速化にむけた適応ネットワークの数学的発見と学習法開発」

研究期間：2018年10月～2022年3月

研究者：今泉 允聡

1. 研究のねらい

本研究は深層学習の高速化のため、データに即して変化するモデルである適応ネットワークを開発する。深層学習を用いるシステムは学習に膨大なコストがかかるため、実用上のシステム調整を迅速・柔軟に行うことが難しい。本研究は、適応ネットワークのための数学的理論を構築し、それに基づく効率的な学習アルゴリズムを開発することで、解析コストの問題を削減し深層学習の普及を促進する。

データ分析が社会に浸透する中で、深層学習と呼ばれる技術が中核的役割を果たしている。しかし、深層学習は分析に必要な解析コスト(計算時間・計算資源)が膨大化しやすく、それが実用化の障害になることが指摘されている。この高解析コストにより、深層学習を用いるプロダクトは(1)迅速なシステムの更新ができない、(2)細かい状況に応じた個別の調整が難しい、といった問題を内包している。深層学習の解析コストを削減するには、パラメータ数の少ない小規模なネットワークを用いる必要がある。しかし、深層学習の原理には不明な点が多く、小規模なネットワークを用いた場合の分析精度の低下が予測できないため、性能を確保する観点から必要以上に巨大なネットワークを用いることが多い。その結果として解析が高コスト化し、深層学習システムの迅速・柔軟な運用を阻んでいる。

本研究の最終的な到達地点は、上記の過剰な解析コスト問題を是正し、深層学習を用いたシステムの柔軟性・即時性を向上させることである。具体的な目的として、以下の三目的を達成する。(i)適切な構造を持つネットワーク(適応ネットワーク)の理論開発:与えられたデータに対して適切な大きさ・構造を持つネットワーク(適応ネットワーク)を考え、その理論を構築する。具体的には、分析対象となるデータの情報論的・幾何学的な性質に適したコンパクトなネットワークの構造を明らかにし、過大なネットワークを用いない深層学習を実現するための、数学的な手がかりを与える。(ii)適応ネットワークを学習するアルゴリズムの開発:明らかにした理論的性質を用いて、所与のデータから適応ネットワークを学習・選択するアルゴリズムの開発を行う。(iii)アルゴリズムの普及および実応用による効果の実証:開発したアルゴリズムの発表・普及の推進、および応用研究・事業化を行う。まず、画像解析の分野でその性能および解析コストの比較を行う。

2. 研究成果

(1) 概要

本研究は、理論フェイズと手法フェイズに二分される(図 1)。理論フェイズにおいては、深層学習が取り扱うさまざまなデータの構造に対して、誤差解析を通してその原理を明らかにしつつ、必要なニューラルネットワーク構造の特定を進めた。

具体的に取り扱った構造としては、(i)表現する関数の滑らかさ・特異性に関する構造、(ii)データに内在する並進・回転不変性といった幾何学的構造、(iii)過適合を回避する非凸・過剰パラメータ化を扱った。

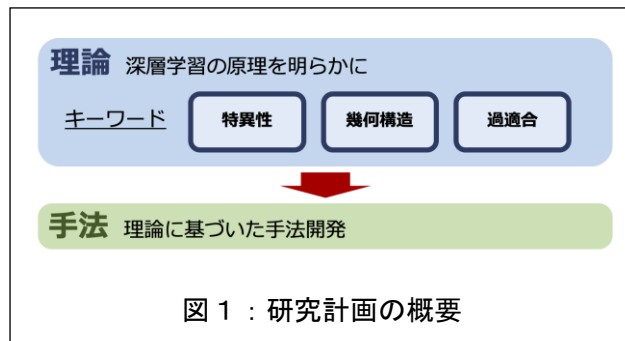
(i)の表現する関数の構造に関する解析では、深層学習が特異性を持つデータのもとで深層学習が従来法に優越することを発見し、さらに深層学習が優越するデータ空間の相図を開発した。また、データに内在する固有次元に注目し、深層学習が次元の呪いを緩和できることを示すと

ともに、扱えるデータの潜在的な低次元構造の種類が広いことを示した。

(ii)の不変性に関する解析では、点群データを含む広いクラスの幾何的構造に着目し、深層学習がこのもとで非常に高い性能を示すことを理論的に示した。これを示すため、本研究では幾何的な構造を再現するニューラルネットワークの特定と、それに付随する概念の開発を行った。

(iii)の過適合に関する研究では、深層学習が多くのパラメータを持ちながら学習データに過適合しない原理を調べた。二重降下現象を対象とした研究では、本来は線形モデルでしか示されていない二重降下現象を、特定の深層ニューラルネットワークモデルも起こることを理論的に示した。非凸性に関する研究では、深層学習の学習アルゴリズムが一定の形状を持つ非凸な損失関数の元で、過適合を回避できることを示し、それをもたらすニューラルネットワーク構造の重要性を明らかにした。

手法フェイズにおいては、理論フェイズにおいて特定したニューラルネットワークを補正として用いる、効率化を目的とした学習アルゴリズムを開発した。これは、補助関数を用いて推定した関数および、データの構造に基づいて事前に特定したニューラルネットワークを用いて、ネットワーク構造の選択を簡略化しつつパラメータ学習を効率した汎用的な枠組みを構築した。



(2) 詳細

1 手法フェイズ: 深層学習の原理解明と適応ネットワークの特定

[関数の非滑らか性]

データ生成過程が十分滑らか(微分可能)な場合、深層学習ではない手法(浅い層を用いたニューラルネットワーク、カーネル法など)が理論的な最適精度を達成することが知られている。本研究はその理論の限界を克服するため設定を拡張し、非連続なデータ生成過程から得られるデータの元(図2)で、深層学習やその他の手法がどういった精度を達成するかを分析した。その結果、あるネットワーク構造の元で、深層学習が広い他クラスの手法に優越することを明らかにした。

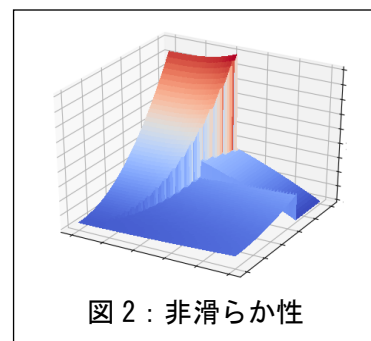
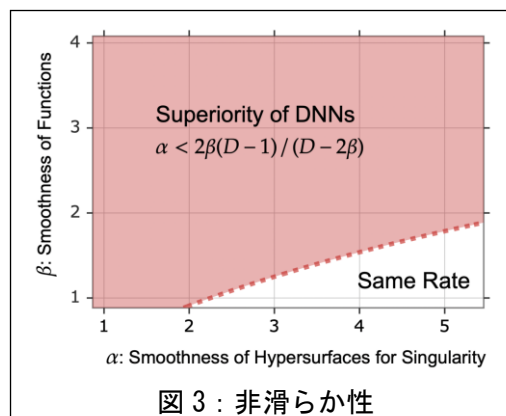


図 2 : 非滑らか性

[特異性と相転移]

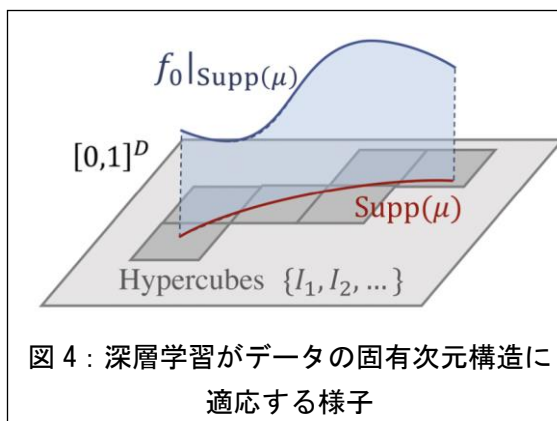
データの非滑らかさをより詳述した概念として、データを生成する関数の特異性について分析した。ここで特異性とは、データ関数が非連続になる台上の点を指し、その特異性の形状についてここで注目した。特異性を解消する方法としては、調和解析の分野で多くの手法が知られている。本研究ではそれらの手法と深層学習を比較し、深層学習の利点を発見することを行った。結果として、特異性の形状がある高次



滑らかさを持つ場合に、深層学習とそれ以外の手法で精度の差が出るのが明らかになった。これを用いて、深層学習の優位性を示すデータの位相図を構築した(図 3)。[\[代表論文 1\]](#)

[固有次元]

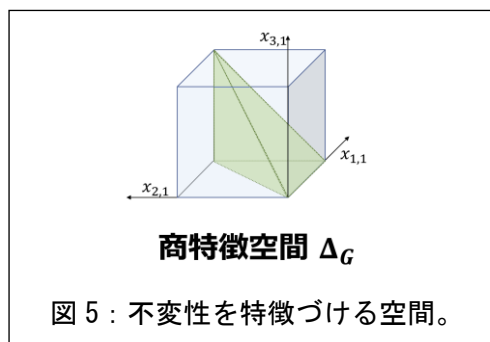
実データは見かけが高次元でも、その本質的な次元(固有次元)は低次元であることが知られている。例えば、機械学習で頻繁に用いられる画像データ(CIFAR-10)は見かけ次元は3072だが、固有次元は約 30 である。本研究は、深層ニューラルネットワークがデータの低次元性に適応し、その汎化誤差が速く収束することを証明した。これは、部分ネットワークによって構成される超立方体



が、標本空間中にあるデータの低次元構造を被覆するように分布することで達成される。(図 4)。また、他手法に対して、深層学習がフラクタルなどの複雑な形状の固有次元に適応することを示した。[\[代表論文 3\]](#) 本研究は、Rutgers 大学(アメリカ)との共同研究として行われた。

[不変性]

点群データ解析などに用いられるニューラルネットワークは、有限群に対する不変性を持つことが知られている。本研究では、このようなニューラルネットワークのもとでの汎化誤差が、不変性の度合い(例えば点の数)に対して指数的に減少することを示した。これは、部分ニューラルネットワークによって標本空間の引き戻して構成される商空間が、入力の不変構造のパラメータにつ



いて指数的に減衰することに由来する(図 2)。この関係性を表現するため、入力空間の体積に依存するエントロピー評価方法を開発した。これらの結果により、モデルに対して不変性の構造を導入するいくつかのニューラルネットワークが高い精度を達成することを理論的に示した。画像データなどは、回転や少量の平行移動を行っても予測結果が変わらないなどの、不変性を持っていることが知られている。本研究では、この不変性と深層学習の性能の関係を調べた。結果として、データが持っている不変性から標本の商空間という概念を定義し(図 5)、その商空間の大きさが深層学習の汎化誤差を記述できることを明らかにした。また、ReLU 活性化関数を持つニューラルネットワークが、この商特徴空間の構築を通して精度を向上させられることを示した。[代表論文 2]

[二重降下]

いくつかの統計モデルでパラメータの数を無限大にすると、汎化誤差の分散が減少することが知られている(例: 二重降下減少)。しかし、この現象が理論的に示されるのはランダム特徴量モデルなどの線形モデルに限られていた。本研究では、深層モデルを含む広いクラスのニューラルネットワークが、一定の正則条件のもとでこの現象に従うことを示した。具体的には、並列化ニューラルネットワークや ResNet などの特有の構造が、大パラメータ極限での誤差の減少を保証することを明らかにした。本研究は、Rutgers 大学(アメリカ)との共同研究として行われた。

[非凸損失関数と過適合]

機械学習手法の過適合を説明する理論的ツールとして、一様収束という概念が知られている。しかし一様収束は最悪の場合の誤差を評価するため、深層学習のような見かけの自由度が大きいモデルでは適切でないことが知られている。本研究は、深層学習が直面する非凸損失関数に着目し、この損失関数にある多くの局所最適値近傍の形状(図 6)が、深層学習の自由度を下げて過適合を防止する理論を構築した。この結論の元では、より細かい最適値近傍集合

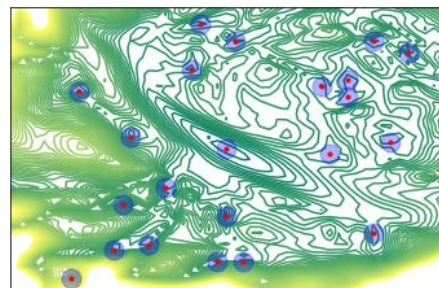


図 6 : 深層学習の損失関数と、その上にある局所解とその近傍。

を多く持つ損失関数や、それを実現するニューラルネットワークが良い性能を持つことが示唆される。本研究は、Leiden 大学(オランダ)との共同研究として行われた。

2. 手法フェイズ: 適応ネットワークに基づく効率的学習法の開発

理論的な知見をもとに、高速学習アルゴリズムの開発を行った。名前を「モチーフ模倣法 (Motif Mimicry)」とし、事前に学習したシンプルな手法を深層ネットワークに転移することで簡単な学習を実現する。提案法は、事前にサポートベクター回帰などの凸最適化で学習できるモデルをデータにフィットさせ、それを解析計算で深層ネットワークに転移させる(図 7)。ここで非凸最適化は必要でないため、通常の SGD などによる学習法よりも少ない計算リソースで高速な学習が可能になった(図 8)。

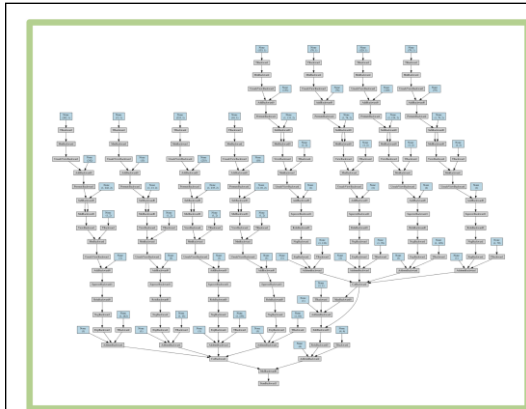


図 7：模倣でもとまったネットワーク

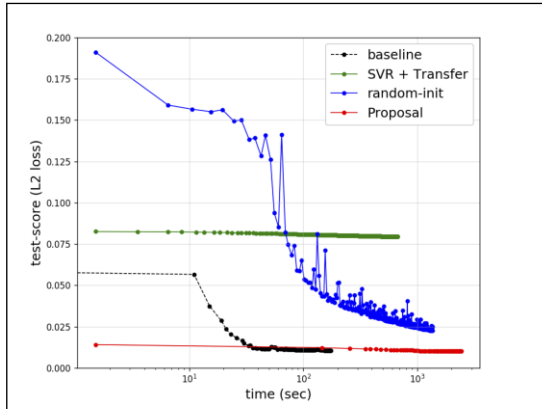


図 8：提案法の高速な学習

3. 今後の展開

本研究の展開は、理論と手法の両面で進められる。理論面については、未だ多くの解明すべき点が残されており、それらの点をより深く研究することが求められる。特に、本研究の理論フェイズの後半で扱った過剰パラメータや非凸損失関数のトピックについては、未だ理論には強い制約や必要であるなど、未解決点が多い。特に既存の数学的道具では根本的に解析が難しい点が多いため、より基礎的な側面から研究を進める必要がある。

手法面については、今回の研究で培った知見を生かし、より応用的なトピックを推進する。例えば、不変性の研究のなかでニューラルネットワーク機構に対する知見が溜まったため、それを用いたより実践的な点群データの解析などを行う。実際、さきがけ内部でそれらを用いた生物系分野におけるデータ解析のプロジェクトが進んでいる。このプロジェクトが進めば、その期間の3年以内で手法の開発が進められる。

4. 自己評価

研究目的の達成状況について、特に理論面については多くの進捗を得られた。手法面については、理論的な研究との架橋および基礎的な部分については進捗を得られたが、この知見を活用した実応用については今後の進捗が求められる。それについては、上述のプロジェクトをはじめとした、複数の応用プロジェクトが共同研究先と開始している。

社会への波及効果という点について、手法面の応用は多くは進んでいないが、理論的な成果を直接伝えるという点で多くの進捗を得られた。具体的には、理論的な進展を直接説明する一般書を出版したほか、学会・研究会だけでなく一般向け公演などを数多く行い、新聞に掲載されるなどの大きな反響を得た。この点で、社会へ研究成果を伝えるという目的を遂行できている。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 12件

1. M.Imaizumi, K.Fukumizu. (2022) "Advantage of Deep Neural Networks for Estimating Functions with Singularity on Curves", Journal of Machine Learning Research, 23(111):1-54.

データ生成関数が特異性を持つ状況における、深層学習によるノンパラメトリック回帰の推定問題を解析した。深層学習の非深層手法に対する優位性を解析し、一定の特異性の条件のもとで、深層学習が理論的な最適性を達成するが、非深層法がそれを達成できない旨を証明した。その結果として、データ空間における深層優勢を達成する相関を導出した。また調和解析的な特異性解析手法との差別化を議論した。

2. A.Sannai, M.Imaizumi, M.Kawano, (2021), "Improved Generalization Bounds of Group Invariant / Equivariant Deep Networks via Quotient Feature Spaces", PMLR: Uncertainty on Artificial Intelligence 2021, in press.

データが有限群作用に対する不変性を持つ時のニューラルネットワークの汎化誤差を解析した。このような不変性を持つデータは、点群やグラフなどのデータ形式で頻出とされる。この研究では、不変性から導かれる商特徴空間を定義し、有限群の構成からこの空間を具体的に導出した。そしてそれを用いて汎化誤差の評価を行った。結論して、汎化誤差は不変性の程度に応じて指数的に誤差が減少することを明らかにした。

3. R.Nakada, M.Imaizumi (2020), "Adaptive Approximation and Generalization of Deep Neural Network with Intrinsic Dimensionality", Journal of Machine Learning Research, 21(174), 1-38.

見かけ上は高次元なデータが、ある低次元集合に集中している状況における、深層学習の汎化誤差を解析した。このようなデータの存在は、高解像度画像などでよく見られ、機械学習でも多様体仮説としてよく知られた事実である。本研究はこのようなデータの分布に適応するニューラルネットワークを具体的に構成して近似誤差を評価し、それに基づいて汎化誤差の収束速度がデータの固有次元のみで決まることを明らかにした。また、深層構造を用いることで、フラクタルのような複雑な低次元集合に適応できることを示した。

(2) 特許出願

研究期間全出願件数: 0 件 (特許公開前のものも含む)

(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)

[書籍] 今泉允聡 (2021) "深層学習の原理に迫る 数学の挑戦", 岩波科学ライブラリー, 岩波文庫, 2021 年 4 月.

[書籍] C. Aggarwal (著), 竹村彰通・今泉允聡 (監修) (2022) "ニューラルネットワークとディープラーニング", 岩波科学ライブラリー, 学術図書出版社, 2022 年 2 月 (予定).

[受賞] 日本統計学会 小川研究奨励賞, 2021 年 9 月.

[メディア掲載] 朝日新聞 GLOBE+ "人間が深層学習の AI を理解できないのには、理由がある", 2019 年 11 月.

[プレスリリース] 東京大学 ARTICLES, "人間が深層学習の AI を理解できないのには、理由がある", 2021 年 3 月. https://www.u-tokyo.ac.jp/focus/ja/articles/z0508_00102.html