【基盤研究（S）】

Fisher情報行列と記述長最小原理に基づく深層学習の理論と実践

研究代表者	九州大学・システム情報科学研究院・教授 竹内　純一（たけうち　じゅんいち）	研究者番号:80432871
研究課題情報	課題番号：23H05492 キーワード：深層学習、ニューラルネットワーク、人工知能、記述長最小原理、データサイエンス	研究期間：2023年度～2027年度

なぜこの研究を行おうと思ったのか（研究の背景・目的）

●研究の全体像

急速に普及が進む人工知能の基礎を固めるための研究を行う。現在、人工知能と呼ばれる技術のほとんどは、機械学習の一種である深層学習に基づいている。深層学習の性能は、他の機械学習技術に比べて飛び抜けて優れているが、その性質には不明な点が多く、経験に頼った開発が続けられている。そのため、現在人間の様々な知的活動を置き換えつつある人工知能の信頼性の保証は難しい。実際、画像分類システムで予想外の分類結果を返してしまう脆弱性の存在など、多くの問題点が指摘されている。本研究では、深層学習の信頼性を担保することを目標に、下記に述べる過学習の問題に関する謎に焦点を当てる。

深層学習は脳の構造を模した数学的モデルであるニューラルネットワーク（神経回路）を利用した機械学習手法である。動物の脳や神経機構には学習能力があり、経験を積むことで様々な機能を獲得する。ニューラルネットワークはこれにヒントを得て考案された機械学習のモデルである。特に、深層学習で用いる深層ニューラルネットワークは層の数が多く、とても複雑なことに特徴がある（図1）。機械学習にはニューラルネットワーク以外にも多くのモデルを用いたものがあり、特に1980年代以降、様々な研究がなされた。

そうした研究の中で、過学習の抑制は理論的課題の中心を占めてきた。過学習とは、訓練データを学習しすぎた結果、その特殊性に引きずられて新しい状況に対応できない状態を指す。これは、訓練データの量に比してモデルの近似能力（様々な機能を実現する潜在的能力であり、複雑なモデルほど能力が高い）が高すぎると起こりやすい。しかし、近似能力が低いと、訓練データに含まれる情報を十分に利用できない。したがって、適度な複雑さのモデルを選ぶこと（統計的モデル選択（図2））が必要となる。ここに、モデルの複雑さは、通常モデルのパラメータ数で表される。

この問題に対して初めて定量的な解を与えたのは、統計学における赤池情報量規準（AIC、1973年）であり、その後の統計学と機械学習の研究に多大な影響を与えた。本研究の柱である記述長最小原理（MDL原理）もその影響下で発展した情報量規準の理論である。これらの理論は、オッカムの剃刀という思考法の指針を定量的に実装したものと考えられる。これは「同等の説明能力ならば、より単純な法則が優れている」というもので、自然科学の指針でもある。

しかし、機械学習のこの常識は2015年ごろに起きた深層学習の台頭によって覆された。すなわち、深層ニューラルネットワークは、情報量規準の観点からすると考えられないほど複雑なものであり、深層学習ではむしろ、複雑なモデルほど良いという考え方が浸透している。このことは、機械学習の理論に対して、一体なぜこれほど複雑なモデルが高い性能を示すのかという、大きな未解決問題をつきつけている。本研究の第一の目標は、この謎を解くことにある。

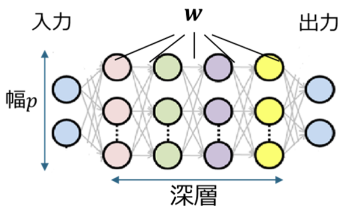

図1　深層ニューラルネットワークの模式図

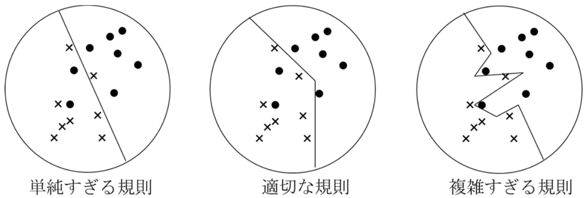

図2　統計的モデル選択のイメージ。折れ線で●と×を分類する問題の例。

[1] Y. Takeishi, M. Iida, and J. Takeuchi, “Approximate spectral decomposition of Fisher information matrix for simple ReLU networks,” *Neural Networks*, Volume 164, July 2023, Pages 691-706.

●Fisher情報行列の近似固有値分解

この問題に取り組むため、我々は深層ニューラルネットワークの最小単位である2層ニューラルネットワーク（図3）のFisher情報行列について詳しく調べた。Fisher情報行列の固有値分解を行えれば、パラメータの空間の中でどの方向が学習にとって重要なのかを知ることができる。我々は、2層ニューラルネットワークの場合の、近似固有値分解の一般式を求めることに成功し、その結果、ごく限られた方向だけ（図3のニューラルネットワークの場合、$O(d^2)$個）が学習において重要であること（図4）を見出した。[1]

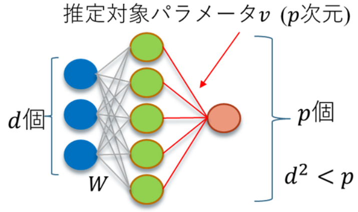

図3　2層ニューラルネットワーク

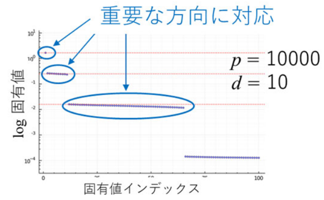

図4　Fisher情報行列の近似固有値分解

この研究によって何をどこまで明らかにしようとしているのか

●解明への道筋（図5）

近似固有値分解が示すようにd^2個の方向だけを学習すればよいのであれば、pがいくら大きかったとしても、このニューラルネットワークは、見かけよりもずっと単純である。このことに着目して2層ニューラルネットワークの性能保証を行うことを最初の目標とする（課題1）。

2層ニューラルネットワークの場合、Fisher情報行列が定数であることが解析を容易にしているが、一般の場合はどうであろうか。実は、2018年に発見された神経接核の理論によると、一般の場合にも、これは近似的に成立する。これを手掛かりとして、2層ニューラルネットワークでの結果を拡張することで、一般の深層ニューラルネットワークの性能保証と性能限界の見極めを行うこと（課題4）が本研究の方針である。これに加え、図6に示す個別課題に取り組む。

（課題1）2層ニューラルネットワークの解析
（課題2）二重降下現象（パラドックス）の直観的理解
近似固有値分解をヒントに解明する
（課題3）自然勾配法との関係の整理
FIM固有値の偏りを情報幾何学の立場で解釈
（課題4）一般の場合への拡張
（課題5）学習法設計
ニューラルネット構造設計，勾配法加速手法等．
（課題6）深層学習の実応用
- サイバーセキュリティにおけるデータ解析
- 磁気共鳴画像法（MRI）の画像再構成手法
- 誤り訂正符号の復号アルゴリズム設計

図6　個別課題の一覧

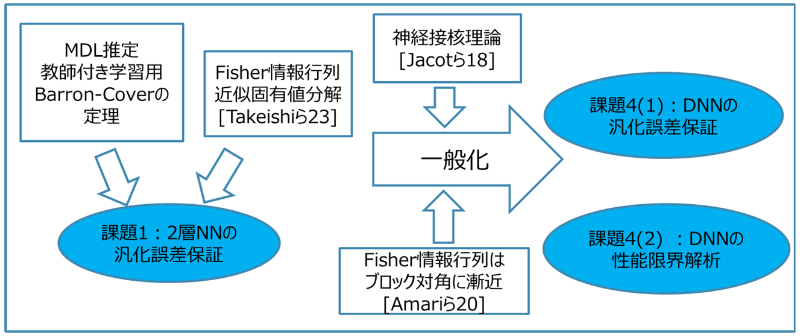

図5　深層ニューラルネットワークの未解決問題解明への道筋

汎化誤差：訓練データとは異なる新たなデータに対する誤差。
Barron-Coverの定理：MDL原理を採用して学習するときの汎化誤差への保証を与える定理。
二重降下現象：モデルを複雑にしていくと、一旦増加した汎化誤差が再度減少していくというパラドックス。

ホームページ等	http://www.me.inf.kyushu-u.ac.jp/%7etak/homepage.ja.html