公開

研 究 終 了 報 告 書

「自由確率論による深層学習の研究」

研究期間: 2019 年 10 月～2022 年 3 月
研 究 者: 早瀬 友裕
加速フェーズ期間: 2022 年 4 月～2022 年 5 月

1. 研究のねらい

深層学習は画像処理・自然言語など広範囲の技術に性能向上とインパクトをもたらした. しかし, その理論的理解は不十分で, 適切なニューラルネットワークの学習のためには, 膨大な計算量が必要である. たとえば, ナイーブに活性化関数や初期値を選ぶと, 勾配の消失・発散が起き, 学習自体が難しくなる.

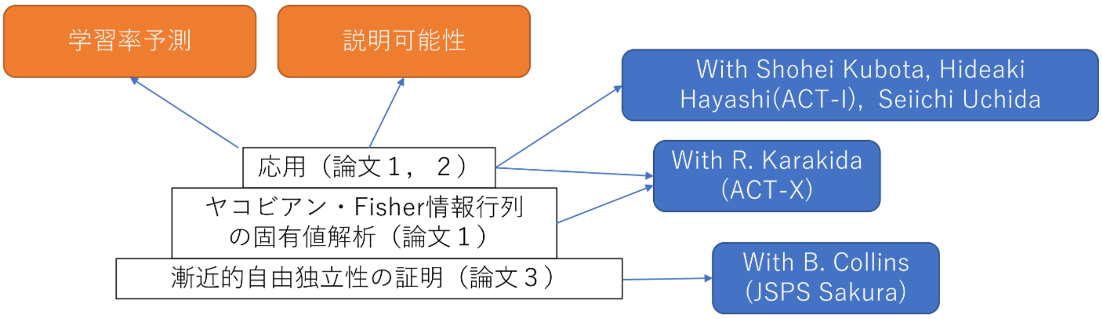
深層学習の理論には, 数理統計側からのアプローチ, 統計物理からのアプローチ(平均場・Neural Tangent Kernel)等がある. 今回の目標ではこれらの方法では足りない代数的な部分を補完・拡張するため, 自由確率論を利用する.

自由確率論はサイズの大きい行列の取り扱いに長けている. 特にランダム行列多項式の固有値解析を, 伝統的なランダム行列理論に比べ, 代数的に洗練された方法で行うことができる. ニューラルネットワークは多層構造になっているから, 特にその勾配や学習ダイナミクスを理解するためには, 行列の多項式を扱う必要が出てくる. 例えば, backpropagation を計算する際の Jacobian や, 学習ダイナミクスを理解するための Fisher 情報行列, Neural Tangent Kernel が含まれる.

今回の狙いは, 自由確率論を用いた固有値解析によって, これらの行列の固有値分布を明らかにし, ニューラルネットワークの設計指針や学習用ハイパーパラメータを理論的に導くことである. もちろん, 十分豊富な計算量があれば, 全てのハイパーパラメータは実験によってチューニングすればよいが, ハイパーパラメータ探査領域は指数的に増えていくから, 少しでも探査数を減らすことは有意義である. また, 体系的に深層学習の理論をまとめることは後代への知識継承として重要であると考える.

ただし, 本研究では理論に十分な拡張可能性を求めたい. 既存の理論は層数に制限があり, 非常にシンプルなモデルしか扱えない部分がある. したがって, 今回はトイモデルだけでなく, 画像処理等実問題で使用可能な複雑さを持つニューラルネットワークへの理論適用も模索する.

2. 研究成果



(1)概要

得られた研究成果は3つの方向性に分かれている.

ひとつは理論の礎になる，漸近的自由独立性の証明である．自由確率論（FPT）は，Dynamical Isometry，フィッシャー情報行列，学習ダイナミクスなど，DNN に関連する研究に現れる，ランダム行列に起因する数学的困難を処理するための豊富な知識を提供する．しかし，各層のヤコビアンの漸近的自由独立性という重要な性質は，今まで完全には解かれていなかった．これを厳密に証明した.(論文3)

二つ目は，Multi-Layer Perceptron (MLP)の Fisher 情報行列の固有値解析である．厳密には，ひとつサンプルが与えられたときの Fisher 情報行列を考えている． フィッシャー情報量行列（FIM）は，パラメータ空間の局所的な計量を記述するため，DNN の学習可能性を理解する上で基本となる．Dynamical Isometry(DI)は，勾配消失・発散を防ぐためにネットワークが満たすための性質で，DI を満たしていれば勾配の信号強度（ノルム）は逆伝搬の途中で保存される．この研究では DI を満たしていても，Fisher 情報行列の固有値が層数に依存し，さらには最大固有値に集中する非自明な現象を示した．（論文1）

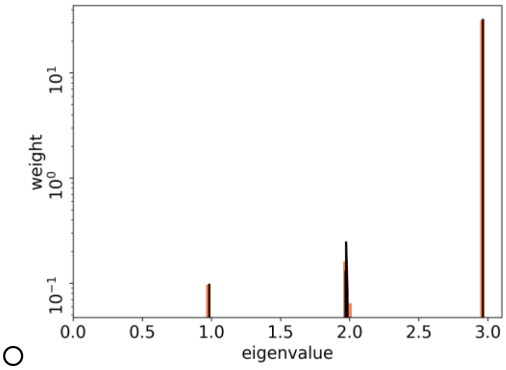
三つめは，第二の結果の応用にあたる．DIを満たしているネットワークのオンライン学習時における学習率が層数に反比例すればよいことを理論と実験で示した．別の応用先として，DNN の解釈可能性は，機械学習を用いた意思決定プロセスの透明性を高めるために，困難ではあるが必須のテーマである．DIを満たすためには Haar ランダム行列を使うのが慣例だが，よりシンプルな単位行列を使ってもできることを示し，さらにこれを分類モデルの分類結果が，どのパラメータに基づいているか(解釈可能性)の分析に応用した.(論文2)

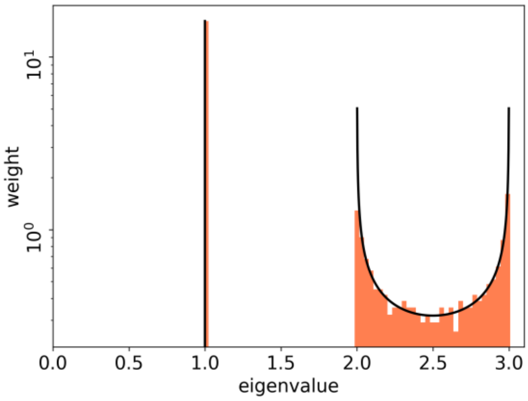
(2)詳細

研究テーマ1「複雑なモデルへの拡張：Jacobian, Fisher 情報行列の研究」

ニューラルネットワークの Jacobian の固有値解析は，勾配消失・発散問題を防ぐうえで重要である．フィッシャー情報量行列（FIM）は，パラメータ空間の局所的な計量を記述するため，DNN の学習可能性を理解する上で基本となる．しかし, FIM は Jacobian に比べてランダム行列モデルとして複雑で，固有値解析も難しい．今回，入力が one-sample である場合のFIMに着目し，DI を満たしているニューラルネットワークの FIM の漸近的固有値解析を行うことができた．ここ

で漸近的というのはパラメータ行列のサイズが十分大きいことを意味する．３層（隠れ層２層）の場合は，自由確率論を用いて，陽に漸近固有値分布を計算することができた:





○ …

一般の層数の場合，FIM の最大固有値は層数に比例し，しかもほとんどの固有値が最大固有値に集中することを示した．このことを利用して，L2ロス関数を用いたオンライン学習の場合，学習率を層数に反比例させると，学習初期段階においては良い学習率になることを実験的に示した：

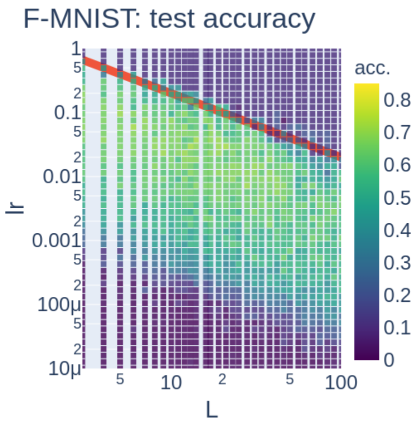


Figure Fashion-MNIST における学習初期でのテスト分類精度．赤線が学習が発散する理論線 2/L を表す．

研究テーマ2「信号処理からの応用」

Dynamical Isometry を満たしているときに，固有値分の１の周りの広がりがどの程度学習に影響を及ぼすか，その指針を作るのが目標であった．固有値分布から決まる逆伝搬の信号容量に基づいて指針を構成したが，この指針が最高値を示すのはすべての特異値が1になる場合であり，非自明な結論は得られなかった．実験的には，ある程度固有値が広がっていたほうが性能は良かった．実際，純伝搬と逆伝搬がともに信号強度を保存すると，ネットワークの非線形性が過度に弱まってしまうので，表現力を担保するためにはどちらかの保存を緩める必要があるとわかった．

研究テーマ3「応用」
学習率予測・解釈可能性への応用を行うことができた．学習率予測については上の通りである．
解釈可能性については，単位初期化というシンプルな構造の初期化を行うことにより，DI を達成して層を深くすることで，各層の初期値からの変化を微小にすることにより，初期値からの変動を見て解釈可能性を分析できるように工夫した．

[本 ACT-X 研究を通じて実現した，当 ACT-X 研究領域内外の研究者や産業界との連携]

- ACT-I 林さんとの連携 上で述べた解釈可能性への応用．
- ACT-X 唐木田さんとの連携．統計物理的アプローチの導入に至る．
- 福田 & Ion Nichita: JSPS Sakura Program Random Tensor and Matrix to Quantum information and Machine Learning. ランダムテンソルの機械学習への応用も視野に入るようになった．
- ACT-X 平木さん VR. Projection Mapping における Projector 配置問題に数理的アプローチを行った. IEEE Poster.

3． 今後の展開
論をランダムテンソルまで拡張し，大規模データセットで理論を実証し，学習率予測・学習ダイナミクスの理解を実用レベルのニューラルネットワークへ拡張する．
[1 年間で達成したい目標] 理論で扱えるほどシンプルかつ実用レベルの性能を持つモデルの理論解析の推進．
内訳：
- (順伝搬)モデルの Gaussian Process （順伝搬の共分散）の解析．それを用いた活性化関数などのハイパーパラメータ調整手法の提案．（自由確率論不要.）
- （逆伝搬）モデルの逆伝搬・Jacobian，Fisher 情報行列(one-sample)の固有値解析.（自由確率論を利用.） 及びそれを使った最適学習率予測．
- （実験）以上を踏まえた，大規模データセットでの実証．

[3～5年で達成したい目標] 現状の深層学習は成功しているが，多くの問題を抱えている．実験論文で動いていても，実用レベルにおいては結局データセットごと・環境ごとにハイパーパラメータ探索は必須である．
もちろん大規模な計算資源があれば，ハイパーパラメータ探索を行うことはできる．しかし，動作原理が分からない深層学習を利用することは，実用レベルの応用では不安が残る．したがって数学的・理論的理解をすることで，可能な限り動作原理を明らかにすることは重要になる．かと

いって，トイモデルでのみ証明された理論のみでは実モデルに援用するのは不安が残る．本研究では，できるだけ実運用レベルの深層学習モデルに対し理論を拡張することで，この不安を払拭したい．

本研究の目指す価値創造は，このような後ろ向きなものに限らない．たとえば，トイモデルと実モデル間に，理論的な違いが見つかり，科学的に興味深い結果が得られることも期待している．また本研究は線形性(逆伝搬)と非線形性（順伝搬）の境界にあり，非線形性をよりうまく扱える数理が生まれないかとも期待している．

4.　自己評価

基礎部分は予想以上の成果であった：Fisherの解析はできる可能性は低かったができた．しかしMLP 以外のネットワークに拡張するのは簡単ではなかった．社会状況により，旅費・謝金への使用が困難であった．計算資源は企業内のほうを使うほうが利便性が高かったため，想定以上に不要であった．

[研究成果の科学技術及び社会・経済への波及効果※]

学習率などのハイパーパラメータの決定について，一定の理論的理解を得られたため，最適化コストの削減が見込める．深層学習の挙動理解．安全に使えるようになることを期待したい．

[*研究者ネットワーク*]

Computer Vision，数学，物理の研究者と共同研究を行い，ネットワークを広げることができた．

[今後の見込み]

今回のモデルをランダムテンソルへ拡張することで，より広い範囲のニューラルネットワークかつ実用レベルのものに拡張できると期待している．

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：3件

1.　T. Hayase, Ryo Karakida, "The Spectrum of Fisher Information of Deep Networks Achieving Dynamical Isometry", Proceedings of The 24th International Conference on Artificial Intelligence and Statistics, PMLR 130:334-342
フィッシャー情報量行列（FIM）は，パラメータ空間の局所的な計量を記述するため，DNN の学習可能性を理解する上で基本となる．本研究では，Dynamical Isometry(DI)を達成しているMLP に着目して，条件付き FIM（1 つのサンプルを与えられたときの FIM）のスペクトル分布を調べた．その結果，条件付き FIM のスペクトルは最大値付近に集中し，その値は深度の増加に伴って直線的に増加することを明らかにした．このスペクトルを調べるために，ランダムな初期化を考慮して，自由確率論に基づく代数的な方法論を構築する．最後に，DNN のオンライン学習における適切な学習率は，条件付き FIM のスペクトルで決まる深さに反比例する

ことを実験で検証した.
2. S. Kubota, H. Hayashi, T. Hayase and S. Uchida, ″Layer-Wise Interpretation of Deep Neural Networks using Identity Initialization,″ ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 3945-3949, doi: 10.1109/ICASSP39728.2021.9414873.
DNN の解釈可能性は，機械学習を用いた意思決定プロセスの透明性を高めるために，困難ではあるが必須のテーマである．解釈可能性を欠く原因の一つは，ランダムな重みの初期化であり，入力が各層で異なる特徴空間にランダムに埋め込まれることである．本論文では，NN の最も一般的なアーキテクチャである深層多層パーセプトロンに対して，単位初期化（すなわち，単位行列を用いた初期化）に基づく解釈手法を提案する．提案手法により，各隠れ層における各ニューロンの分類への貢献度とクラス尤度を分析することができる.
3. Benoit Collins, Tomohiro Hayase, ″Asymptotic Freeness of Layerwise Jacobians Caused by Invariance of Multilayer Perceptron : The Haar Orthogonal Case″, to appear in Communications in Mathematical Physics, （arXiv:2103.13466）
層ごとのヤコビアンの漸近的自由独立性という重要な仮定は，これまで完全には証明されていなかった．漸近的自由性の仮定は，スペクトル分布を層を介して伝搬させる際に基本的な役割を果たす．動的アイソメトリーを実現するためには，ハール分布直交行列が不可欠である．本研究では，この場合の多層パーセプトロン（MLP）の層ごとのヤコビアンの漸近的自由独立性を証明する．証明の鍵となるのは, MLP の不変性である．各層の隠れユニットを固定する直交行列を考慮し，各層のパラメータ行列を自分自身に直交行列を掛けたものに置き換えると, MLP は変化しない．さらに，元の重みが Haar 直交行列であれば，この置き換えによってヤコビアンも変化しない．最後に，この重要な事実を利用して，各重みを活性化関数のヤコビアンから独立した Haar 直交ランダム行列に置き換えることで完了する.

（2）特許出願

なし

（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）

*-invited talk

- *2021.11 Workshop on “Non-commutative Probability and Related Fields 2021”
- *2021.3 Sakura Project RMTQIML Mathematics of Neural Networks, Random Matrix in Fisher Information of Deep Nets Achieving Dynamical Isometry
- *自由確率論による深層神経回路網の解析，情報系 WINTER FESTA Episode 5, 2019.12
- 2021, Takefumi Hiraki, Tomohiro Hayase, Yuichi Ike, Takashi Tsuboi, Michio Yoshiwaki, ″Viewpoint Planning of Projector Placement for Spatial Augmented Reality using Star-Kernel Decomposition″ accepted into IEEE VR 2021 Poster Track.