様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　６ 月 １７ 日現在

機関番号：８２６２６
研究種目：研究活動スタート支援
研究期間：2017～2018
課題番号：１７Ｈ０７３９０

研究課題名（和文）階層モデルの幾何学に基づく深層学習の理論構築と制御

研究課題名（英文）Theoretical construction and control of deep learning based on the geometry of hierarchical models

研究代表者

唐木田　亮（Karakida, Ryo）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・研究員

研究者番号：３０８０３９０２

交付決定額（研究期間全体）：（直接経費）　　2,300,000円

研究成果の概要（和文）：本研究課題では，幾何学の視点に基づいたアルゴリズムの提案と理論解析をとおして，制御が容易な深層学習の数理基盤の構築と開発を行った．具体的には，パラメータ空間の幾何学を定めるFisher情報行列の解析とそれに基づく勾配法の開発を行った．また，素子の入れ替え対称性は学習が停滞する特異領域をもたらすが，この特異領域を回避する条件を明らかにした．さらに，深層学習の応用上重要なWasserstein距離を情報幾何学的に調べるとともに，より自然なコスト関数の設計指針を与えた．

研究成果の学術的意義や社会的意義
深層学習は実用ベースの開発が進み，収束や解の性質が数理的に保証されていないヒューリスティクスへの依存が大きく，恣意性が多く学習の制御が難しい問題がある．本研究はこの問題に数理的な解決の基盤を与えている点で意義深い．今後，誰にでも使いやすい深層学習の開発につながることが期待できる．また，学習の挙動に性能保証を与えることは，安全性や信頼性が必要とされる実応用に深層学習を広げていくための土台となることも期待できる．

研究成果の概要（英文）：This research project analyzed and developed learning algorithms in deep learning from the perspective of geometry. It led to a mathematical foundation and practical applications for deep learning without uncontrollable heuristics. In more details, we analyzed the Fisher information matrix of deep neural networks, which determines the geometry of the parameter space, and applied it to propose some gradient methods. Besides, we investigated singular regions of the parameter space caused by permutation symmetry of units. We revealed a condition where such singular region and the slowing down of training are avoidable. Moreover, we gave information geometric insight into Wasserstein distance with the entropy constraint and proposed a novel and more natural cost function based on it.

研究分野：ソフトコンピューティング

キーワード：ニューラルネットワーク　機械学習　深層学習　数理工学　情報幾何　統計力学

１．研究開始当初の背景

神経回路を模したニューラルネットワークが再注目を集めており，素子を層状に並べた多層モデルの学習は深層学習と呼ばれる．深層学習は，画像識別や音声認識を筆頭に既存の機械学習を上回る性能を発揮している．しかし，大規模なデータとモデルを用いる深層学習では，数理的な裏付けがとれた学習手法が必ずしも利用できない．したがって性能重視の開発が進み，理論的な背景がはっきりしないヒューリスティックな手法が多く開発されてきた．特に，深層学習はモデルが非線形で理論的な考察が難しいこともあり，勾配法に限っても非常に多くのヒューリスティクスに依存している．また，理論が不足しているため，性能を向上するためのパラメータ設定を手作業で決める恣意性も残っており，制御が容易な学習手法の開発が問題として残ったままである．

２．研究の目的

幾何学の視点に基づいたアルゴリズムの提案と理論解析をとおして，制御が容易な深層学習法の確立を目指す．まず，パラメータ空間の幾何学構造に注目することで，勾配法の効率的な学習率の設定を実現する．これは収束の高速化のみならず，空間の歪みを反映しながら学習率を自動調整するアルゴリズムとみなすことができ，手動での学習係数の調整を排除できる利点がある．さらに，深層モデルのような階層モデルでは，素子の対称性に起因するパラメータ空間の幾何学的な歪みが生じることに注目する．この対称性が，学習の誤差局面に鞍点構造を生む．この鞍点が学習ダイナミクスを停滞させるメカニズムを深層学習の文脈で力学的に解析し，明らかにする．

３．研究の方法

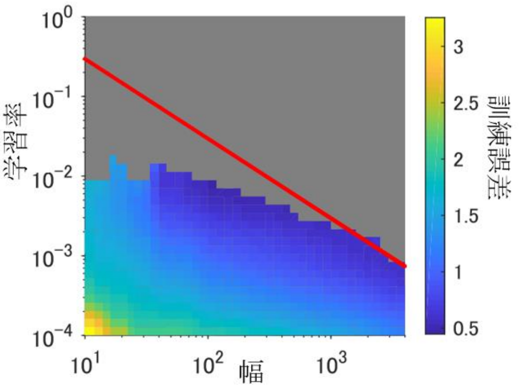
幾何学的なアプローチをとおして深層学習の数理的基礎を構築するために，おもに４つの解析を行った．（1）Fisher 情報行列の巨視的理論，(2)自然勾配法の開発，では，統計力学的極限操作を行い，固有値の統計性について理論解析を行った．また，(1)では網羅的な数値実験により，収束を速める学習率を推定した．(3)特異領域における学習の力学系解析，では，まず素朴な解析計算により特異領域を減少させる条件に当たりをつけた．次に,オンライン学習の統計力学的解析を利用して，その条件下での学習ダイナミクスを少数変数の常微分方程式に縮約し，より定量的な解析を行った．最後に，当初の目的からやや視点を広げ，(4)情報幾何学的な立場から見た Wasserstein 距離を使ったコスト関数の設計，を行った．これは，Wasserstein 距離が深層学習で有効であることが様々な文脈で報告されており，深層学習の理論と制御をテーマに研究を進めるうえで欠かせない要素となったためである，情報幾何学的な解析に基づき，エントロピー正則化付き Wasserstein 距離の潜在的な構造を数理的に明らかにした．

４．研究成果

(1)Fisher 情報行列の巨視的理論

ランダム結合を持つ全結合深層ニューラルネットワークにおいて，Fisher 情報行列の固有値を解析した．Fisher 情報行列はパラメータ空間の Riemannian 計量として働き，また，ある条件化で局所的な誤差局面の形状を決めるため，勾配法に基づく学習において本質的な役割を持つ．深層モデルの理論解析は，おもに非線形変換の繰り返しのため困難であるが，ランダム結合を持つモデルの幅が十分に大きいという極限をとることで，いくつかの統計量が解析的に評価できる．本研究では，Fisher 情報行列の固有値分布の平均，2 次モーメントや最大固有値が，4種類の秩序変数を使って表現できることを様々な設定の深層モデルで明らかにした．4種類の秩序変数は深層モデルの平均場理論として近年注目を集めているものと同じであり，これらの量が大規模深層モデルの解析にとって重要な量であることを示唆している．固有値の統計性については，固有値平均が幅無限の極限で漸近的にゼロに近づく一方，最大固有値は漸近的に無限に発散するオーダーが現れることが明らかになった．これは，幅が大きいネットワークでは固有値分布の歪みが強く現れることを示している．さらに，幾何学的には，パラメータ空間は局所的に，多くの方向に平坦であるが，その他の少数の方向には大きく歪んでいることを意味している．これが，大規模な深層モデルにおける普遍的な現象と考えられる．

特に，最大固有値は，最急勾配法の収束に必要な学習率の設定と関係している．モデルの幅が大きくなるにつれて最大固有値は大きくなることから，モデルサイズに応じて適切な学習率を小さくとる必要がある．本研究では，複数種類の深層モデルにおいて benchmark dataset での実験を行い，秩序変数から計算された最大固



図：学習率の見積もり．秩序変数による学習率の見積もり（赤線）と実際のモデルの訓練結果（カラーマップ）．灰色は誤差が発散した領域．

有値を使って，勾配の発散を防ぐ学習率を見積もれることを明らかにした（図参照）．また，最も速く収束する学習率も見積もることができ，これまで恣意性が大きいハイパーパラメータとして制御が困難であった学習率に対して，ある程度，理論的な知見を与えられたことは実用上も意義深い．なお，当初の計画では，確率的最適化も個別に取り扱う予定であったが，ある程度大きいバッチサイズの確率的最適化では，バッチ学習を仮定した上記の理論を使って学習率を見積もれることが実験的には示された．この枠組みを拡張することで，今後さらに学習率の設定に関する知見が増え，深層学習の制御に貢献できることが期待される．以上の研究成果は，発表論文①に基づく．

(2) 自然勾配法の開発

自然勾配法は Fisher 情報行列を使って通常の最急勾配を補正する手法で，効率的にパラメータ空間を探索することができる．研究成果(1)では，巨視的な秩序変数を用いて Fisher 情報行列の固有値の性質を定量的に評価することに成功した．この枠組みを応用することで，Fisher 情報行列を近似的に計算し，自然勾配法に利用する研究を行った．通常の自然勾配法では Fisher 情報行列の逆行列の計算量が大きいため，実用が困難であるが，平均場近似に基づいた Fisher 情報行列は対角ブロック行列による表現が可能であり，さらに各ブロックは低ランクに支配されているため逆行列計算が容易であることを示した．これにより，今後，大規模な深層モデルで利用可能な自然勾配法の開発がより一層進み，学習率のヒューリスティックな制御を排除した効率的な勾配法の発展につながると期待できる．以上の研究成果は，発表論文②に基づく．

(3) 特異領域における学習の力学系解析

通常,学習アルゴリズムに対するパラメータの逐次変化を微分方程式に定式化し，解の性質や収束までの時間を求めることは難しい．しかし幸いなことに,3層パーセプトロンの一部では，学習力学を解析的に導出できることが報告されている．本研究ではこうした力学系解析を使うことで，特異領域を抜ける時間の定量化を行った．パラメータ空間における特異領域は勾配法が停滞しやすい領域として知られてきたが，学習の力学的解析をこの領域で実施したところ,モデルの出力素子数を増やす条件下で特異領域を減らせることが明らかとなった．この研究は出力素子(すなわちデータのラベル数)が1つではない多くのデータセットでプラトー現象を抑えられる可能性を明らかにしており，学習ダイナミクスの理解，アルゴリズムの開発を進めるうえで重要な知見といえる．以上の研究成果は，発表論文④と⑥に基づく．特に，発表論文④では，統計力学的なアプローチで学習のダイナミクスを少数の常微分方程式に縮約することに成功している．これにより，モデルの出力素子数が増えるにつれて，学習の停滞が避けられる様子が，より定量的に示された．今後，学習の挙動をより複雑な階層モデルで数理的に説明するうえで基礎になる知見と期待でき意義深い.

(4)情報幾何学的な立場から見た Wasserstein 距離を使ったコスト関数の設計

深層学習の基本的な設計部位であるコスト関数に関して研究を行った．パラメータ空間の幾何学はコストの設計に依存するため，勾配法の開発の前に，コスト関数の幾何学を調べることが有用だろう．通常使われる KL 情報量に基づくコスト関数は，自然画像のような高次元データでは十分な性能の学習結果が実現しない例が報告されている．この問題に対し有望と考えられる Wasserstein 距離を研究した．その結果，エントロピー罰則付き Wasserstein 距離の背後に双対微分幾何に基づく情報幾何の構造があることが明らかになった．以上の研究成果は，発表論文⑤と⑦に基づく．この幾何学の視点は，数理的に自然な学習アルゴリズムの開発につながると期待される．実際，発表論文③では，得られた視点に基づき，エントロピー正則付き Wasserstein 距離が正しい距離を構成するための補正法を提案し，その有効性を簡単な画像の重心問題で実験的に検証した．この補正法により，数理的に自然で扱いやすいアルゴリズムの開発がより一層期待できるとともに，画像生成の応用において，よりシャープで現実に近い画像の生成が期待できる.

５．主な発表論文等

〔雑誌論文〕（計　7　件）

① Ryo Karakida, Shotaro Akaho & Shun-ichi Amari, Universal statistics of Fisher information in deep neural networks: mean field approach, International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 89:1032-1041, 2019 (査読あり).

②Shun-ichi Amari, Ryo Karakida & Masafumi Oizumi, Fisher information and natural gradient learning in random deep neural networks, International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 89:694-702, 2019 (査読あり).

③Shun-ichi Amari, Ryo Karakida, Masafumi Oizumi & Marco Cuturi, Information Geometry for Regularized Optimal Transport and Barycenters of Patterns, Neural Computation, vol.31, pp.827-848, 2019 (査読あり).

④Yuki Yoshida, Ryo Karakida, Masato Okada & Shun-ichi Amari, Statistical Mechanical Analysis of Learning Dynamics of Two-Layer Perceptron with Multiple Output Units, Journal of Physics A, vol.52, pp.184002(1)-(17), 2018（査読あり）.

⑤Shun-ichi Amari, Ryo Karakida & Masafumi Oizumi, Information Geometry Connecting Wasserstein Distance and Kullback-Leibler Divergence via the Entropy-Relaxed Transportation Problem, Information Geometry, vol.1, pp.13-37 2018（査読あり）.

⑥Shun-ichi Amari, Tomoko Ozeki, Ryo Karakida, Yuki Yoshida & Masato Okada, Dynamics of Learning in MLP: Natural Gradient and Singularity, Neural Computation, Vol.30(1), pp.1-33, 2018（査読あり）.

⑦Ryo Karakida & Shun-ichi Amari, Information geometry of Wasserstein divergence, Proceedings of Geometric Science of Information (GSI), pp.119-126, 2017（査読あり）.

〔学会発表〕（計　8　件）
①Ryo Karakida, Theoretical analysis of RBMs with Gaussian visible units -Dynamical analysis and Riemannian optimization-, American Institute of Mathematics (AIM) workshop "Boltzmann machines", 2018.

②唐木田 亮, 深層ニューラルネットワークにおける Fisher 情報行列の普遍性, 第 30 回 RAMP シンポジウム, 2018.

③唐木田 亮, 機械学習から見たニューラルネットワークの数理, 東京理科大学 脳学際研究部門 第 2 回公開シンポジウム, 2018.

④唐木田 亮, 深層ニューラルネットワークの数理: 平均場理論の視点, 産総研 AI セミナー, 2018.

⑤唐木田 亮, 赤穂昭太郎, 甘利俊一, 深層ニューラルネットワークにおける Fisher 情報行列の普遍性, IBIS2018, 2018.

⑥唐木田 亮, 赤穂昭太郎, 甘利俊一, ランダム深層ニューラルネットワークにおける Fisher 情報行列の巨視的理論, 日本物理学会年次大会, 2018.

⑦Ryo Karakida, Shun-ichi Amari, Information geometry of Wasserstein divergence, Geometric Science of Information, 2017.

⑧唐木田 亮, エントロピー正則化付き Wasserstein 距離の情報幾何, 第 64 回幾何学シンポジウム, 2017.

〔図書〕（計　1　件）
唐木田 亮, 麻生 英樹, 深層学習の数理, 8 ページ, 数理科学(サイエンス社), 2018.

〔産業財産権〕

○出願状況（計　0　件）
○取得状況（計　0　件）

〔その他〕
ホームページ等
https://sites.google.com/view/ryokarakida/

６．研究組織
(1)研究代表者
唐木田　亮　　(KARAKIDA, Ryo)
産業技術総合研究所・人工知能研究センター・研究員
研究者番号:30803902