

数理・情報のフロンティア  
2019 年度採択研究代表者

2020 年度 年次報告書
------------------

唐木田 亮

産業技術総合研究所 人工知能研究センター  
研究員

大自由度ニューラルネットワークの学習に潜む幾何学的構造の解析と  
信頼性評価への展開

## § 1. 研究成果の概要

本年度は, NTK に基づいた学習ダイナミクス解析の対象として, 自然勾配法に着目し研究を遂行した. 深層学習では学習の高速化のために自然勾配法の開発が進んできたが, 厳密な自然勾配は計算コストが大きく利用が難しい. そこで, ヒューリスティックな近似自然勾配法の開発が進められてきた. この近似自然勾配法に対して, 幅が十分に大きい深層モデルの NTK regime において訓練ダイナミクスの解析を行った. その結果, 深層学習で経験的に使われてきた近似自然勾配が, ステップ数に対して, 厳密な自然勾配と同じ速さで収束できることが明らかとなった. 具体的には, 勾配を補正している Fisher 情報行列の対角ブロック近似の中で, layer-wise 近似, unit-wise 近似, もっとも広く使われている K-FAC 近似で高速な収束が実現可能である. いずれの場合も, 関数空間での勾配が等方的になる条件を満たしており, これが NTK regime において最も早い収束を実現するために重要である. 一方で, 各近似手法の違いが汎化性能において現れる点や, より粗い近似において等方性が崩れる場合を数値実験を援用しながら明らかにした.

本年度はさらに, 非常に層数が多い深層モデルにおいて, 逆誤差伝播の発散・消失を防ぎながら訓練するためのアーキテクチャ・パラメータ初期値について, dynamical isometry の観点から研究を遂行した. 具体的には, 訓練サンプルに対して条件づけられた Fisher 情報行列の固有値分布をランダム行列理論によって解析し, 適切なスケールの直交行列初期値において固有値の集中がみられること, この固有値に対して適切な大きさをスケールした学習率で, 超深層のモデルの訓練を初期値から高速に進められることを明らかにした.

### 【代表的な原著論文情報】

[1] Ryo Karakida & Kazuki Osawa, "Understanding Approximate Fisher Information for Fast Convergence of Natural Gradient Descent in Wide Neural Networks", In Proceedings of Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 10891–10901, 2020.