

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 に基づく学習ダイナミクスの解明の第一歩として、深層学習で広く使われている normalization 手法の解析を実施した。具体的には、幅が十分に大きい深層モデルにおいて、BN が NTK および Fisher 情報行列の最大固有値を抑える仕組みを数理的に明らかにした。初期値周辺においては、Batch normalization(BN)がロス関数の歪みを抑え、勾配法の学習率を大きくとれることが示唆され、これは BN の経験的な知見とも一致する。とくに、BN は最大固有値のネットワーク幅への依存性を軽減しており、どのような大きさ(幅)のモデルでも大きな学習率で高速な訓練の収束が期待できる。さらに、本研究では BN の幅への依存性を軽減するためなら、最終層に BN をかけるだけで十分であることを明らかにした。下図は実際に深層モデルを再急降下法で訓練した結果であり、左図が BN なし、右図が最終層にのみ BN の簡易版をかけた結果である。赤線は理論線である。図からわかるように、最終層での BN が幅に依存しない学習率でロス関数を大きく下げることが成功している。なお、この効果は BN に特有であり、類似手法である Layer normalization では実現できないことも数理的に明らかにした。以上のように、これまでヒューリスティックに開発が進んできた normalization 手法に、普遍的に成立するメカニズムの系統的な理解・定量化を与えることができたといえる。

