

論文内容の要旨

博士論文題目 Analysis of Regularization and Optimization for Deep Learning
(深層学習の正則化と最適化の解析)

氏名 古庄 泰隆

(論文内容の要旨)

深層学習は多くの分野に適用されその優れた性能が報告されている。近年の深層学習の目覚ましい成果は正則化と最適化アルゴリズムに基づいている。例えば多層パーセプトロン(MLP)や ResNet などの深いニューラルネットワーク(DNN)は高い表現能力を持つ一方、データセットに含まれるノイズに過剰適合したり勾配降下法による学習の収束が遅いという問題がある。多様な正則化手法は DNN が訓練データに含まれるノイズに過剰適合するのを防ぎ、バッチ正規化(BN)や stochastic depth などの最適化アルゴリズムは ResNet の学習を加速させる。これらのメカニズムを明らかにするために多くの研究がなされてきた。正則化が MLP の汎化性能に及ぼす影響を解析するために Rademacher 複雑度などのモデルの複雑度が解析されてきた。しかしながら、最近の実験によりこれらの指標は上記の現象を説明するのに十分でないことが示された。したがって別の視点から正則化手法を解析する必要がある。また MLP の場合には BN による勾配降下法の高速度化について解析されている。この研究では BN が極小解付近の損失関数の形状を滑らかに、特に Fisher 情報行列の最大固有値を小さくすることで、勾配降下法で大きな学習率が使え学習が高速化することを示している。しかしながら ResNet の Fisher 情報行列の最大固有値や BN と stochastic depth がその固有値に及ぼす影響については明らかでなかった。そこで我々は正則化による過剰適合防止や BN と stochastic depth による学習高速化のメカニズム理解に向けて次の理論解析を行った。まず荷重減衰と L1 正則化、パス正則化が MLP のフーリエ係数に及ぼす影響を解析し、これらの正則化が MLP が訓練データの高周波成分を学習するのを防ぐことを示した。高周波成分の大部分が訓練データに含まれるノイズで構成されていると仮定すると、この結果は正則化により MLP がノイズを学習するのを防ぐことを示唆している。また ResNet の Fisher 情報行列の最大固有値と BN と stochastic depth がその固有値に及ぼす影響を解析し次の結果を示した。まず ResNet の最大固有値は層の数に対して指数的に増加する。そして stochastic depth はこの指数部を小さくし、BN は指数的な増加を線形な増加へと抑える。この固有値の抑制により勾配降下法で大きな学習率が使え学習が高速化する。

| | |
|----|-------|
| 氏名 | 古庄 泰隆 |
|----|-------|

(論文審査結果の要旨)

現代の AI の主流である深層学習は驚くべき性能を見せ、多くの分野に応用され AI 社会とも呼ばれる世界を作り出しつつある。しかしその一方で、その性能を生み出す仕組み、いわゆる理論の研究はほとんど進んでいない。実際、現代の d 深層神経回路網 (deep neural networks, DNN) は、skip-connection を導入した ResNet や入力を正規化する batch normalization (BN)、あるいは学習を加速化する dropout や stochastic depth などの多くのヒューリスティックを導入することでその収束性を向上させているが、これらがなぜ、どれほど性能に影響を与えるのかは全くわかっていない。

本論文はこの問題を解決するため、これらのヒューリスティックが DNN の性能に与える影響を理論的に導出したものである。DNN が勾配降下法で学習されることから、その特性を評価するために Fisher 情報行列の最大固有値を評価し、BN や stochastic depth などの効果を定量的に導出した。また、DNN のフーリエ係数に正則化が与える影響を評価し、学習の周波数特性を明らかにした。これらの成果は、今後の DNN 利用に設計指針を与えるものである。

以上をまとめると、本論文は DNN の各種技術の効果を理論的に評価したものであり、将来の DNN による AI の普及に大いに資すると考えられる。よって、博士 (工学) の学位に値するものと認められる。