

審 査 の 結 果 の 要 旨

氏 名 謝 沢 柯

本博士論文は深層学習における学習のメカニズムをランジュバン動力学を代表とする拡散理論の観点から解明するものである。博士論文は全部で7章から構成されており、1章から2章で深層学習における学習のダイナミクスについての既存研究の説明がある。3章から6章までは、謝氏の主な貢献としての拡散理論に基づく深層学習のダイナミクスについて記述されている。具体的には、深層学習で用いられる種々の学習アルゴリズムをランジュバン動力学として表現し、その拡散係数の振る舞いを解析する方法を提案している。

第3章では、深層学習で用いられる確率的勾配降下法(stochastic gradient descend, SGD)についてそのダイナミクスを解析している。未知のデータに対する深層学習モデルの予測性能の意味では、学習に用いられる経験的な損失関数に対して、平坦な局所解(flat minima)と呼ばれる局所解を見つけることが重要であることが知られている。第3章では、SGDを用いることで、ある条件の下ではこのような平坦な局所解が効率的に見つかることをKramer's escape rate problemを利用して解析している。具体的には、局所解に対する回避時間率(escaping time rate)を解析することでSGDが指数的オーダーで早く尖った局所解(sharp minima)から平坦な局所解に移動することを示した。

第4章では、SGDを改良したAdamという学習アルゴリズムに関して、そのダイナミクスを解析している。AdamはSGDに比べて学習データにおける経験的な損失関数を最小化するという意味で収束は早いですが、未知のデータに対する予測性能を表す汎化能力が劣るということが知られている。これはすなわち、十分時間をかけて学習すると、SGDは最終的にはAdamよりも高い汎化能力を発揮するということである。この現象に対して第4章では、第3章で開発した局所解における回避時間解析技術を用いることでAdamの解析を行っている。解析の結果解析、AdamはSGDに比べて経験損失の鞍点を効率的に避けることができる一方で、尖った局所解(sharp minima)の回避時間率が長いことを解明した。つまり、SGDに比べて鞍点を効率的に避けるが、尖った局所解に囚われやすいために、収束は早いですが汎化能力が劣ることを解明した。

第5章では、4章での理論解析結果を元にAdamの問題点を解決するアルゴリズムを提案している。Adamと同程度に鞍点を避け、sharp minimaを回避することができるアルゴリズムを提案している。ここではAdamとSGDの違いの一つである慣性項の有無に着目している。ランジュバン動力学の観点では拡散項とドリフト項に分けて考えることができる。慣性項はドリフト項に関係し、この慣性項によって鞍点からの効率的な回避を実現している。しかし、この慣性項は尖った局所解(sharp minima)の回避時間率に対しては悪影響を及ぼしていることを解明し、適応的な慣性項を導入することでこの問題を解決している。これによって、提案アルゴリズムはAdamと同程度に鞍点を効率的に避けることができ、SGDと同程度の尖った局所解(sharp minima)の回避時間率を有する学習手法となっている。実験では、現在現実的に良く使われている大規模な深層ネットワーク(ResNet-18, ResNet-34, ResNet-50, VGG-16, DenseNet-121, GoogLeNet)を用いて、大規模なベンチマークデータセット(CIFAR-10, CIFAR-100, ImageNet)に対する提案アルゴリズムの優位性を示している。

第6章では、神経科学における実験的に知られている事実を動機とした深層学習アルゴリズムの解析をしている。具体的には、神経活動は常に変動しているという神経変動性をモチベーションとして、損失関数のパラメータに対してノイズで平均をとる形で最適化する方法に関して、拡散理論の観点で分析を行っている。つまり、通常の機械学習では損失関数は固定されているのだが、損失関数そのものがノイズによって変動しているという定式化を行いその学習アルゴリズムの解析している。このような損失関数の確率的勾配降下法は、ノイズに対する損失関数のテイラー展開を考えるとヘッセ行列を拡散係数としたランジュバン動力学と見ることができ、これまで培った拡散理論を使った学習アルゴリズムの解析ができる。これは、単に勾配に対してノイズを加えたものとは異なる挙動をする。このような学習が深層学習で見られる破滅的忘却を防ぐ効果があること、ミスラベルによるノイズの影響を受けにくいことを発見した。

以上のように、謝氏は深層学習の最先端の学習アルゴリズムに対して拡散理論を基にした統一的な解析手法を開発したことは学位論文として十分に認められる内容として、審査員一同合格としました。

よって本論文は博士（科学）の学位請求論文として合格と認められる。

以上 2078 字