

審査の結果の要旨

氏 名 二反田 篤史

コンピュータの性能向上とデータ収集能力の向上に伴い、大規模データを用いたより複雑な機械学習モデルの学習に対する需要が高まっている。特に、データの大規模化により機械学習の計算コストが膨大になりつつあることから、学習の高速化は喫緊の課題である。さらに、学習方法とそれに用いる最適化手法は時に切り離して議論することが難しく、より良い学習法の設計には最適化手法も同時に論じる必要がある。本論文では、機械学習における最適化法を軸として (1) 大規模データを用いた機械学習における確率的最適化手法の研究、および (2) 関数勾配を用いた新しい機械学習手法の構築を行い、種々の問題設定に対して効率的な最適化手法および学習方法を提案すると共にその有効性を理論的、実験的に確認している。

本論文は「Efficient Machine Learning from Gradient Method Perspective in Finite and Infinite Dimensional Spaces」(有限・無限次元空間における勾配法の観点からの効率的機械学習)と題し9章からなる。論文は大きく分けて二つのパートに分かれており、それぞれ「大規模データを用いた機械学習における確率的最適化手法の研究」「関数勾配を用いた新しい機械学習手法の構築」に主眼を置いている。

第1章「Introduction」では、機械学習における確率的最適化および関数勾配法の研究背景と、本論文の構成および論文の主結果についての概要を与えている。

第1パート：大規模データを用いた機械学習における確率的最適化手法の研究

第2章「Stochastic Gradient Descent」では、第1パートの主眼である確率的最適化法およびその機械学習における利用方法について概要を与えている。

第3章「Accelerated Variance Reduced Stochastic Gradient Descent I」では、凸損失関数を用いた有限次元正則化付き経験損失最小化問題における確率的分散縮小勾配降下法の加速法を提案している。滑らかな目的関数を持つ凸最適化問題において、Nesterovの加速法は通常の勾配法より優れた収束率を持つが、大規模データによる経験損失最小化問題においては確率的勾配降下法に遅れを取る事が知られていた。本研究では確率的最適化において標準的な手法である確率的分散縮小勾配降下法にNesterovの加速法を組み込む手法を提案し、問題が強凸関数である場合に条件数のオーダーのミニバッチを用いることでNesterovの加速法と同じ収束率が実現される事を示している。

第4章「Accelerated Variance Reduced Stochastic Gradient Descent II」では、第3章で提案した手法にリスタート法と呼ばれる技法を組み込み、より実用的な手法を提案している。これにより強凸の場合だけでなく最適解を基点とした強凸性を持つ関数、あるいは一般の滑らかな凸関数に対しても理論が拡張される。

第5章「Stochastic Difference of Convex Algorithm」では、目的関数が凸関数の差分、即ち、Difference of Convex (DC) 関数の場合に有用な確率的最適化手法を提案している。提案手法は、理論的には確率的勾配降下法と少なくとも同等の収束性能を持ち、いくらかの追加条件下では収束性が改善される事が示されている。更に数値実験によって制限・深層ボルツマンマシンの両モデルにおいて通常確率的最適化手法を上回る性能を発揮することを確認している。

第2パート：関数勾配を用いた新しい機械学習手法の構築

第6章「Functional Gradient Descent」では、第2パートの主眼である無限次元関数空間における関数勾配法による機械学習手法の一般論および、第2パートの各章の概要および関係性を述べている。適切な問題設定における関数勾配法の収束性、および有限次元でのパラメトライズによって生じる局所解に陥らない事が示されている。

第7章「Stochastic Particle Gradient Descent for Infinite Ensembles」では、無限次元空間におけるL1正則化付きアンサンブル学習を考察し、アンサンブルに含まれる弱学習器をサンプリングするための確率測度を、関数勾配法を使って直接学習する方法を提案している。提案手法を用いることで停止時刻を注意深く設計せずとも過学習を防げることが理論と実験から示されている。

第8章「Enhancing the Convergence of Adversarial Training」では、Wasserstein GAN (WGAN) というニューラルネットワークを用いた生成モデルの学習法を関数勾配法によって改良する方法を提案している。実際に画像データセットを用いた数値実験において少ない更新数で効率的にインセプションスコア（生成画像群の質を測るスコア）を改善する事が確認されている。

第9章「Functional Gradient Boosting Based on Residual Network Perception」では、分類問題のためのResidual Networkの特性を持つ勾配ブースティング法の学習方法を提案している。元来、Residual Networkのネットワーク構造とデータマイニング分野で非常に強力とされる勾配ブースティング法との関連が指摘されていたが、本論文ではこの関係性を関数勾配法という観点からより直接的に見直し、新しい勾配ブースティング法を提案している。関数勾配の性質を活かした二種の汎化誤差上限を与え、数値実験では提案法が勾配ブースティング法の著名な実装であるLightGBMを含むいくつかの代表的な判別手法を上回る事を確認し、その有効性を示している。

以上を要するに、本論文では最適化の観点から機械学習における幅広い問題設定にお

いて計算量と汎化性能を改善する方法論を包括的に与えている。特に、確率的最適化においては分散縮小勾配降下法の加速法およびDC問題を効率的に解くための確率的最適化手法を提案し、また関数勾配法を用いた学習の方法論を提唱し、その概念をアンサンブル学習、ニューラルネットワークの生成モデルの改善、Residual Network型の勾配ブースティング法の提案、といった各種応用に展開している。これら提案手法の有用性は、理論評価だけでなく、数値実験による従来手法との比較を通して十分に検証されている。機械学習における計算と汎化という問題を、有現次元・無限次元の最適化という観点から包括的に研究し、理論と実用の両面から種々の問題を解決し、数理情報学の発展に大きく寄与している。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。