

## 論文の内容の要旨

論文題目 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)

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を実行する際にかかる計算コストが膨大になりつつあり、学習の高速化は喫緊の課題として重要である。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、実験的に確認する。

### (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

はじめに、機械学習において最も典型的な設定である凸損失関数を用いた有限次元における正則化付き経験損失最小化問題を取り扱う。滑らかな目的関数を持つ凸最適化問題において、Nesterov 加速法は通常の勾配法より優れた収束率、即ち、少ない反復数での収束性を持つが、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。本研究では近年提案された確率勾配の分散縮小法により Nesterov 加速法を確率化する方法を提案し、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サイズで Nesterov 加速法と同じ収束率が実現される事を示す。現在では分散縮小法の研究は更に進み、より優れた手法が提案されているが、本研究はミニバッチの並列計算で真に確率的勾配降下法の計算量を改善する加速法の存在を示した初の試みである。この手法に続き、リスタート法と呼ばれる Nesterov 加速法の強凸関数と一般の凸関数に対する適応性をリスタートタイミングに集約させるアイデアを取り入れた手法を提案する。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、あるいは一般の凸関数に対しても理論が拡張

される。またデータの特徴の疎性の逆数に比べ条件数が大きい場合にアルゴリズムの総計算量がデータの次元に非依存となる性質を示す。

続いて、最適化対象の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する。DC 関数の最小化は制限ボルツマンマシン等多くの機械学習問題を包含する重要な関数のクラスである。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である事を示し、いくらかの追加条件下では収束性が改善される事を示す。主な適用対象としてレコメンデーションや教師無しアンサンブルの分野で重要とされる制限ボルツマンマシン及びその拡張である深層ボルツマンマシンの学習を考察し、ボルツマンマシンの学習においては本手法が Expectation-Maximization 法の拡張である事を示す。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する。

## (2) 関数勾配を用いた新しい機械学習手法の構築

以上では有限次元モデルを学習するための最適化手法を扱ったが、データが複雑な場合、機械学習モデルには強い非凸性が求められ、その度合いやパラメトライズの仕方に最適化の難易度も大きく依存する。具体的には勾配が小さな領域での学習の停滞や表現力に欠ける局所解への収束等が起きうる。そこで、本論文ではいくつかの非凸モデルの学習は確率測度を最適化するための変数変換写像の学習と見做せるという点に着目し、関数勾配を用いた確率測度の最適化、言い換えれば変数変換写像の最適化手法を提案する。本手法は一般に広大な無限次元空間で振る舞うが、単純な変数変換写像から反復的に複雑度を向上させていき、後述の問題設定において変数変換写像についての滑らかさに応じた収束性を備え、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る。本手法の各反復は Residual Network という近年提案された深層ネットワークと類似の構造を持つ層を積み重ねる過程として実現される。Residual Network は画像の分類問題等で優れた性能を発揮する事が多くの実験から確認され、理論的な解析が求められているが、提案手法は Residual Network に対するいくつかの観測結果と整合しており、より踏み込んだ理論解析のための手掛かりになり得る。更に、より深い Residual 型ネットワークの学習の一つの実現にもなっている。

本手法の応用例については三種の問題設定（無限アンサンブル学習、敵対的生成ネットワークの学習、勾配ブースティング法の学習）を取り上げ、その各々に対して具体的なアルゴリズムの実現化方法と収束解析及び得られる変数変換写像に関する理論解

析を与える。更に、数値実験的に有効性も確認する。一つめの問題設定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であり、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直接学習する。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る。二つめの問題設定は 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の改良である。WGAN は入力データの分布と生成ニューラルネットワークによるノイズ分布の押し出し分布との (近似的) Wasserstein 距離を最小化するようにネットワークを学習する手法で、現在ニューラルネットワークを用いた生成モデルの学習において標準的に用いられている。WGAN の目的関数は非凸性が強く最適化の困難が生じると考えられるため、WGAN の学習が停滞した後に得られた生成用ネットワークを初期値として提案手法を適用し、より複雑な生成ネットワークを学習することを試みる。生成ネットワークが変数変換写像の空間において最適化余地を残していた場合に、そのギャップを効率的に埋めることが期待されるが、実際に画像データセットを用いた数値実験において、生成ネットワークの更新数について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) を改善する事を確認する。三つめの問題設定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である。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、提案法はこの性質をより直接的に捉えた勾配ブースティング法である。通常の勾配ブースティング法と比べ、関数勾配の近似に反復的に深化するニューラルネットワークを適用しているところが異なり、最適化の効率の違いを生んでいる。本論文では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の著名な実装である LightGBM を上回る事を確認し、その有効性を示す。