

様式VI

## 博士学位論文調査報告書

論文題目 Algorithms for Accelerating Machine Learning with  
Wide and Deep Models  
(Wide・Deepモデルを用いた機械学習を高速化するための  
アルゴリズム)

申請者氏名 井田 安俊

最終学歴 平成26年3月  
早稲田大学大学院先進理工学研究科電気・情報生命専攻修士課程 修了  
令和3年3月  
京都大学大学院情報学研究科知能情報学専攻博士後期課程  
研究指導認定見込

学識確認 令和 年 月 日 (論文博士のみ)

論文調査委員 京都大学大学院情報学研究科  
(調査委員長) 教授 鹿島 久嗣

論文調査委員 京都大学大学院情報学研究科  
教授 田中 利幸

論文調査委員 京都大学大学院情報学研究科  
教授 山下 信雄

( 続紙 1 )

京都大学	博士 (情報学)	氏名	井田 安俊
論文題目	Algorithms for Accelerating Machine Learning with Wide and Deep Models (Wide・Deepモデルを用いた機械学習を高速化するためのアルゴリズム)		
(論文内容の要旨)			
<p>本論文は、近年、幅・深さともに巨大化する機械学習モデルを扱う際に必要となる莫大な計算コストを軽減するための、効率的な学習・推論を行うアルゴリズムについての研究結果をまとめたものであり、全5章から構成されている。</p> <p>第1章は序論であり、本研究の目的とその内容を概観している。社会における様々なサービスを実現するために、膨大なデータに基づく機械学習が活用されているが、対象の問題が複雑になるに従い、必要となるモデルも巨大化してきている。本研究では、機械学習モデルを、高次元の入力をもつデータを扱うための幅広 (wide) モデルと、複雑な入出力関係を表すための深い (deep) モデルに分類し、それぞれに対する計算コストの削減に取り組む。具体的には、幅広モデルに対しては特徴量にグループ構造がある場合の予測モデルの特徴選択と、行列分解の一種であるCUR分解に対して、効率的な学習手法を提案している。一方、深いモデルについては、確率的勾配降下 (SGD) 法の改良による深層ニューラルネットワーク (DNN) の効率化や、DNNの一種であるResidual Network (ResNet) の層を削減しモデルを圧縮する方法を提案している。</p> <p>第2章では、幅広モデルの典型例として、高次元の特徴量にグループ構造が与えられた状況において、グループ単位での選択とグループ内の特徴選択を同時に行うスパースグループラッソを取り上げ、その標準的な解法であるブロック座標降下 (BCD) 法を高速化したアルゴリズムを提案している。BCD法において特徴量グループがモデルに含まれるべきかの判定に用いる条件式を、その上界を用いて近似的に評価することによって、条件式の厳密評価を省略する方法と、さらに下界も用いることで、特徴の更新順を工夫することで、学習を高速化する方法を提案している。また、提案手法の理論的な保証も与えるとともに、数値実験によって実際のデータに対する有効性も検証している。</p> <p>第3章では、幅広モデルのもうひとつの例として、行列分解の一種であり、巨大なデータ行列から一部の行と列を選択して行列を近似するCUR分解を取り上げ、これに第2章と同様のアイデアを適用することによって、CUR分解に用いる座標降下法を高速に実行する方法を与えている。また、その有効性をデータを用いた実験によっても示している。</p> <p>第4章では、深いモデルの学習の効率化を扱っている。DNNの学習に用いられるSGD法の学習率を適応的に変化させることで、性能向上を狙っている。SGDにおいては、ミニバッチのサイズが小さいときには勾配のノイズが大きくなることで学習が不安定になるという問題があるが、これを解決するために、更新の分散を一定に保つように学習率を調整するという新たな発想に基づく方法を提案している。また、数値実験によって従来手法よりも高速に学習が行えることを示している。</p>			

第5章では、深いモデルの推論時の高速化を扱っている。特に、画像認識等のタスクで有効性が示されているDNNの構造であるResNetのモデル圧縮手法を提案している。提案手法では、予測精度をできるだけ犠牲にせずにDNNの層を削除するために、ResNetのユニットに優先度スコアと呼ぶパラメータを追加し、これも合わせて最適化し、得られた優先度スコアの低い層を削除するというアイデアを提案している。画像識別の標準的なデータセットを用いた実験では、精度を落とすことなく、層を大幅に削減できることが示された。

第7章は結論であり、本論文で得られた成果を要約している。即ち本論文は、実世界の問題を扱うために幅・深さともに巨大化する機械学習モデルにおける計算量の問題を解決するために、効率的な学習・推論を行うアルゴリズムを提案するものであり、これらの有効性を理論的・実験的に示したものである。本論文では最後に、将来の展望として、本論文では扱わなかった問題設定への拡張や、さらなる理論的な解析、ならびに、幅広かつ深いモデルへの展開を挙げ、本論文を結んでいる。

(続紙 2)

(論文審査の結果の要旨)

本論文は、近年、幅・深さともに巨大化する機械学習モデルを扱う際に必要となる莫大な計算コストを軽減するための、効率的な学習・推論を行うアルゴリズムについての研究に取り組んだものであり、得られた主な成果は次の通りである。

1. 幅広モデルの典型例として、高次元の特徴量にグループ構造が与えられた状況において、グループ単位での選択とグループ内の特徴選択を同時に行うスパースグループラッソがあるが、その標準的な解法であるブロック座標降下 (BCD) 法では、特徴量グループがモデルに含まれるべきかの判定に用いる条件式の評価を行うためには、全特徴量を処理しなければならないという問題があった。また、行列分解の一種であり、巨大なデータ行列から一部の行と列を選択して行列を近似するCUR分解においても、不要な行・列の判定を行う条件式の評価に、行列全体を処理しなければならないという問題があった。本論文では、条件式に含まれる値の上界を用いて条件式の評価を効率的に行うことによって、スパースグループラッソにおいては各グループ内の特徴量数に比例した時間で、あるいはCUR分解においてはひとつの行あるいは列のサイズに比例した時間で、条件式の評価を行う高速化を行った。また、上界に加えて下界も用いることで、特徴の更新順に優先度を与え、さらに学習を高速化した。また、提案手法が一定の条件のもとで従来のBCD法と同様の解を与えることの理論的な保証も与えるとともに、数値実験によって実際のデータに対する有効性も検証した。

2. 深層ニューラルネットワーク (DNN) の効率的な学習には、ミニバッチに基づく確率的勾配降下 (SGD) 法が用いられるが、その学習率の設定が性能に大きく影響を与える。たとえば、ミニバッチのサイズが小さいときには勾配のノイズが大きくなることで学習が不安定になるという問題があった。本論文では、この問題を解決するために、更新の分散を一定に保つように学習率を調整するパラメータ更新法を提案した。また、数値実験によって従来手法よりも高速に学習が行えることを示している。

3. DNNでは層の数を増やすことによって、複雑な入出力関係を実現しているが、このような巨大なモデルでは、学習時のみならず、推論時にもモデルサイズ、とくにネットワークの深さが推論時の計算時間に大きく影響する。本論文では、画像認識等のタスクで有効性が示されているResidual Network (ResNet) モデルにおいて、予測精度をできるだけ犠牲にせずに層の数を削減して推論効率を高めるために、ResNetのユニットに優先度スコアと呼ぶパラメータを追加することで、これも合わせて最適化し、得られた優先度スコアの低い層を削除するという新たなモデル圧縮手法を提案した。画像識別の標準的なデータセットを用いた実験では、精度を落とすことなく、層を大幅に削減できることを示している。

以上、本論文は実世界の問題を扱うために幅・深さともに巨大化する機械学習モデルにおける計算量の問題を解決するために、効率的な学習・推論を行うアルゴリズムを提案するものであり、これらの有効性を理論的・実験的に示した結果をまとめたものであり、学術上・実応用上寄与するところが少なくない。よって、本論文は博士 (情報学) の学位論文として価値あるものと認める。また、令和3年2月24日に実施した論文内容とそれに関連した口頭試問の結果、合格と認めた。なお、本論文のインターネットでの全文公表について支障がないことを確認した。

要旨公開可能日： 年 月 日以降