

称号及び氏名 博士（工学） 土田 悠太

学位授与の日付 平成 25 年 3 月 31 日

論文名 「GPGPU を用いたニューラルネットワーク学習の  
高速化に関する研究」

論文審査委員 主査 吉岡 理文  
副査 市橋 秀友  
副査 中島 智晴  
副査 柳本 豪一

## 論文要旨

人工ニューラルネットワーク (Neural Network, NN) は様々な分野で応用されている。例えば長谷川らによる顔画像認識や、山川らによるガスセンサアレイを用いた匂いのパターン認識、また高橋らによる感情音声認識への応用などが挙げられる。しかし、NN の学習・訓練には、長時間を要するという問題点がある。一方で、近年、画像処理の発展によってグラフィックボードに搭載されている GPU (Graphics Processing Unit) の性能が格段に向上しているため、それを汎用的に用いることができる技術である GPGPU (General-Purpose computing on Graphics Processing Unit) が研究されている。GPU の特徴として、数百の演算回路が内蔵されている点があり、その性能を引き出すためには、これらが同時に利用されるように、計算処理を並列化する必要がある、それによって高速化が可能である。そこで、本論文では、NN 学習アルゴリズムを並列化が可能のように変更し、GPGPU を用いることで学習に要する時間を短縮することを目的とする。

NN はニューロンと呼ばれるユニットを複数個結合した構造を持ち、多変数の入力信号から非線形関数に従う多変数の信号を出力する機械学習手法である。ニューロンは、複数の信号を入力し、その入力信号の数に対応する重み係数との積和を入力とするシグモイド関数の値を出力とするものである。今回対象とする NN は、入力層、中間層、出力層の 3 層構造で、中間層と出力層部分にニューロンが存在する 3 層パーセプトロンである。入力信号と対応する教師信号の組をパターンとすると、複数の異なるパターンを学習する場合はその度に NN の出力信号とそれぞれの教師信号の誤差が小さくなるように重み係数の修正が行われる。これを繰り返す行うことで、学習が進められる。学習に要する時

間とは、各層に存在するニューロンの個数と、パターン数、およびその繰り返し回数の積に依存する。

一方、近年、画像処理専用の演算装置である GPU を汎用的に利用する、GPGPU がある。3DCG や動画処理の発展、またコンピュータゲームの進歩に伴って、GPU は汎用の CPU (Central Processing Unit) より処理速度が向上しており、今後より発展が期待される技術である。GPGPU は、GPU の特徴である複数個実装された単純な演算回路および高速なメモリといった資源を有効に活用する技術であり、複数の処理を同時に処理するルーチンの最小単位であるスレッドを準備することで、高速化が容易に行える点が利点として挙げられる。

NN を並列化する手法に関する従来手法として、Witbrock らによる SIMD 並列コンピュータによる実装、大道らによるホップフィールドネットワークの実装、山森らや Singer, 吉田らによる複数計算機を並列に構成したニューラルネットワーク学習法が提案されている。GPGPU による並列計算の実装は、複数計算機を用いた場合と比較してメモリ構成の違い等があるため、単純ではない。NN の GPU への実装手法は、Oh らによって提案され、その後 Davis によってこの提案が体系的にまとめられている。しかしながら、これらの研究では、学習済みの NN に対しての GPGPU の利用を想定しており、学習に関する高速化は提案されていない。そこで、本論文では、NN の学習に必要な時間を短縮するために、学習過程を並列化し、GPGPU に実装することで、高速化する手法を提案する。本論文で対象となる、学習中の計算に必要な時間を 3 つに大別すると以下になる。1 つ目は、異なるパラメータを持つ NN を学習するのに必要な時間である。2 つ目は、NN 内部のニューロンの計算に伴う時間である。同一の層に含まれるニューロンの計算は、入力と重み係数の値が異なるものが、同一の計算式で行われ、その計算を繰り返し実施する必要がある。3 つ目は、1 つの NN に対し、複数のパターンを提示するのに必要な時間がある。これら 3 つの時間のうち、1 つ目の時間を短縮するには、ニューラルネットワーク全体を複数個並列化させたものが考えられる。これは同期等特別な手法が必要で無いため本論文において、従来手法の代表としての比較手法とする。本論文では、2 つ目の時間を短縮するために NN 内部の並列化手法、および 3 つ目の複数パターンを同時に学習する手法について述べる。以下、本論文の構成の概要を述べる。

第 2 章では、まず比較手法として、1 つ目の時間を短縮する手法について述べる。階層型 NN を学習する際に、重み係数の初期値によって二乗誤差が最小値に収束しない現象が発生する。通常は試行錯誤によって最適な重み係数の集合を探索する必要がある。そこで、重み係数の初期値が異なる NN を複数用意し、それらを GPGPU によって並列に処理する方法が考えられ、これを比較手法とする。次に NN における各層の内部に存在するニューロンを並列化することで、処理時間を短縮する手法を提案する。各層の計算はそれぞれの前段の層の値がすべて揃うまで、進めることが出来ない。しかし各層の内部に位置するニューロンの出力値は、それぞれの値を用いることなく計算する事が可能なため、計算上独立である。そこで、並列化する計算の最小単位を、入力値と重み係数から 1 つの値を出力するまでのニューロンの計算とする。各層におけるニューロンの個数の最大個数をスレッドの個数として指定する。また、前の層から出力される値をすべて次の層で使用するために、各層毎に同期を取る必要がある。これらの手法の有用性を検証するために、並列化の有無を変えて数値実験を行い、処理速度を比較する。数値実験では各層のニューロンの個数およびパターン数を変えて処理速度を調査する。結果として、提案手法は、CPU と比較して約 20 倍の高速化が達成され、比較手法と比べても、6.4 倍の高速化に成功したことを示す。

第3章では、複数のパターンを同時に学習する手法について提案する。第2章において述べた手法では、各ニューロンの処理は並列で行われているが、それらのニューロンに接続されている重み係数を更新する処理は逐次的に処理する手法をとっていた。そこで、パターン毎に更新される重み係数に注目して並列化するように、逆伝播段階の計算を改良する。具体的には、同一層に含まれる個々のニューロン内に含まれる重み係数の個数と、各層のニューロン個数との積の数だけ同時に計算する手法を提案する。順伝播の段階ではパターン毎に同一の重み係数を用いて、出力層の値を計算し、逆伝播ではパターン毎に重み係数を求めてそれらの和を用いて更新を行うアルゴリズムに変更する。そのため、従来の個々のパターンに対し逐次的に重み係数を更新するアルゴリズムおよび、パターン毎の出力と教師信号の誤差の和を用いるバッチ型とは異なるため、乱数によってテストデータを作成し、最適値に収束するかどうかを確認する。結果として、従来法と同様に収束することが確認され、また第2章の手法と比較して、約6倍の高速化を達成したことを示す。

第4章では、提案手法の有効性を検証するために、入力信号の個数が多い例として、ツイッターの情報をを用いた異常気象情報の捕捉を取り上げる。ツイッターとは、情報サービスの一種で、140文字以内の「ツイート」と呼ばれる短文を投稿し、共有することができるサービスである。ツイートには様々な情報が含まれ、中には災害情報も含まれており、そのようなツイートから、いち早く災害情報を取得するために、前章までの提案手法が有用であると考えられる。本論文では、2012年1月11日に鳥取で発生した大雪情報の検出を試みる。単語は、形態素解析エンジン MeCab を用いて、「名詞」と「形容詞」を抽出し、全ツイートで使用頻度が大きなものを採用する。ツイート毎に含まれるこれら単語の有無をベクトルにしたものを特徴ベクトルとして NN に入力し、収集したツイートが、単語「鳥取」を含むツイートを「陽性」、他を「陰性」とする。入力する特徴ベクトルには「鳥取」を含まない状態で学習し、その効果を検証する。ここで用いられる特徴ベクトルの次元数は 890 におよび、通常の学習手法では非常に時間がかかる。この実データを用いて本手法でどの程度時間が短縮されるかを比較する。結果として、このツイートをを用いた推定の正解率が 65%であることを示す。また学習時間が、通常の CPU を用いた手法で 2 時間必要であるのに対し、提案手法では約 4 分で処理が完了し、CPU と同様に学習が可能でかつ高速であることから、本手法の有用性が確認出来たことを示す。

最後に第5章では、本研究の総括と今後の展望について述べる。

## 審査結果の要旨

本論文は、人工ニューラルネットワーク (NN) 学習の GPU による高速化手法に関する研究成果をまとめたものであり、以下の成果を得ている。

(1) NN における各層の内部に存在するニューロンを並列化することで、処理時間を短縮する手法を提案した。結果として、提案手法は、CPU と比較して約 20 倍の高速化が達成され、比較手法と比べても 6.4 倍の高速化に成功した。

(2) 複数のパターンを同時に学習する手法について提案した。この手法は、従来法のアルゴリズムとは異なるため、学習の収束性を確認した。結果として、従来法と同様に収束することが確認され、

また(1)の手法と比較して、約6倍の高速化を達成した。

(3) 提案手法の有効性を検証するために、入力信号の個数が多い例として、ツイッターの情報をを用いた異常気象情報の捕捉を取り上げた。結果として学習時間が、通常のCPUを用いた手法で2時間必要であるのに対し、提案手法では約4分で処理が完了し、CPUと同様に学習が可能でかつ高速であることから、本手法の有用性が確認出来た。

以上の諸成果は、NN学習の高速化を達成するものであり、高度な学習能力を持つNNの実用化に大きく貢献するものである。また、申請者が自立して研究活動を行うのに必要な能力と学識を有することを証したものである。