

MPIを用いた学習データ並列化によるDBNの学習高速化

正木智史 佐藤公信
高知工業高等専門学校

1 はじめに

DBN(Deep Belief Networks)[1]は、深層学習のアルゴリズムを実装する人工知能の一種である。先行研究により、特に画像識別の分野において、DBNの有用性が証明されている[2]。DBNは、RBM(Restricted Boltzmann Machine)を多層にして教師なし学習を行った後、Logistic Regressionで教師あり学習を行い、効果的な学習を実現している。しかし、DBNは、学習データの入力細胞数、RBMの中間層の増加に伴って計算量が増大し、学習時間の増大は避けられないため、学習の高速化が課題である。

本研究では、DBNの学習をより高速化するため、MPI(Message Passing Interface)を用いて、RBMの学習だけでなく、LRの学習も並列化し、提案する手法の評価を行う。

2 Deep Belief Networks

DBN(Deep Belief Networks)は、多層のネットワークを特徴とするDeep Learningの一種であり、RBM(Restricted Boltzmann Machine)と、Logistic Regression(以下LR)により構成される。DBNの学習は、図1のような流れで行われる。

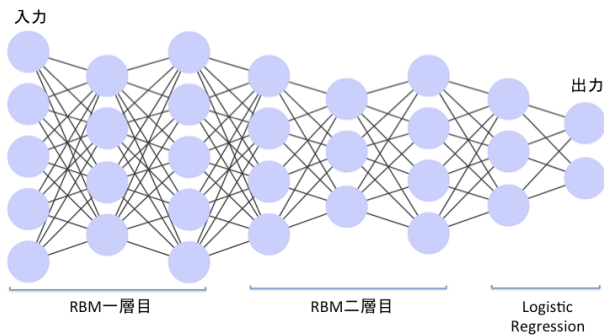


図1 DBNの学習

まず、多層のRBMにより教師なし学習を行い、その後、RBMの出力をLRに入力して教師あり学習を行う。

3 Message Passing Interface

MPI(Message Passing Interface)は、分散メモリ型の並列計算機における、メッセージ通信APIの規格である。分散メモリ型の並列計算機では、SPMD(Single Program Multiple Data)モデルによりプログラミングされる。それぞれのプロセスでは、同じプログラムが実行されるが、MPIを用いてプログラム上で並列処理を記述することにより、各プロセスで異なる処理を行う。

4 提案する並列化手法

提案する並列化手法について述べる。

MPIを用いて、学習データ群(図2: dataA, dataB, dataC)を、ランクと呼ばれる個々のMPIスレッドに振り分け、それぞれのランクにおいて学習パラメータの調整量を計算する。RBMにおける調整量の計算は、

$$\Delta w_{ij} = \epsilon (\langle v_i h_j \rangle_0 - \langle v_i h_j \rangle_k) \quad (1)$$

$$\Delta b_j = \epsilon (\langle h_j \rangle_0 - \langle h_j \rangle_k) \quad (2)$$

$$\Delta a_i = \epsilon (\langle v_i \rangle_0 - \langle v_i \rangle_k) \quad (3)$$

の各式により、学習データ毎に行う。ここで、 Δw_{ij} はウェイトの調整量、 Δb_j 、 Δa_i は、それぞれ隠れ素子と可視素子のバイアスの調整量、 ϵ は学習率、 v_i 、 h_j は、それぞれ可視素子と隠れ素子の状態である。また、LRにおける調整量の計算は、

$$\delta y_i = y_i - y'_i \quad (4)$$

$$\Delta w'_{ij} = \epsilon' \times \delta y_i \times x_j \quad (5)$$

$$\Delta b'_i = \epsilon' \times \delta y_i \quad (6)$$

の各式により、学習データ毎に行う。ここで、 $\Delta w'_{ij}$ はウェイトの調整量、 $\Delta b'_i$ はバイアスの調整量、 ϵ' は学習率、 x_j はLRへの入力データ、 y_i は正しい出力値、 y'_i は現在の出力値である。各ランクで計算された調整量は、学習データを全て学習し終えた後に、一つのランクに収集し、ウェイトとバイアスに加算する。その後、加算されたウェイトとバイアスを各ランクに配信し、再び学習を行う。

例として、dataA, dataB, dataCを学習する際の処理の流れを、図2のフローチャートに示す。

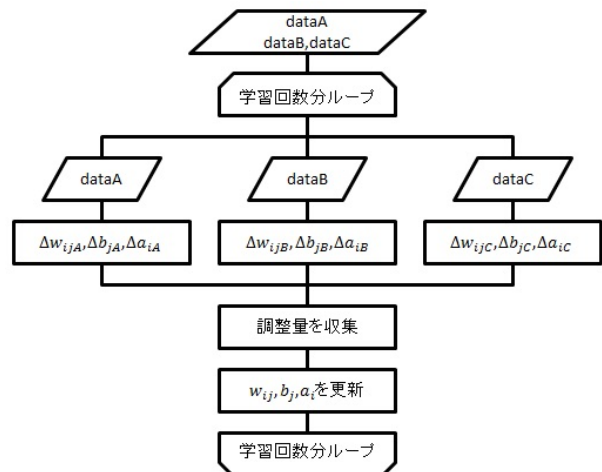


図2 MPIを用いた並列処理の流れ

5 性能評価

提案する並列化手法により並列化された DBN の実行結果と、CPU のシングルスレッドで実行した DBN の実行結果、それぞれの学習所要時間での比較により、提案する並列化手法の評価を行う。また、コア数の増加による影響を評価するため、クラスタシステムのコア数を変化させて、学習所要時間の計測を行う。

評価は、DBN の学習時間について行い、評価に使用する学習データは、MNIST(手書き文字) データベースの画像 (60000 枚, 28×28 ピクセル, 入力細胞数:784, 中間層:800) を使用する。学習時間の測定には、MPI ライブラリの時間計測関数を使用し、RBM 一層目の学習 10 回, RBM 二層目の学習 10 回, LR の学習回数 10 回を 1 回とする、DBN の学習を 10 回行い、学習時間の平均をとる。

5.1 評価環境

評価には、高知高専で稼働中のクラスタシステム (以下 HPC) を使用する。

HPC の OS は、CentOS 6.5 であり、コンパイラは、Intel コンパイラ (バージョン:14.0.2.144), MPI は、IntelMPI を使用する。HPC は、HP ProLiant DL360p Gen8(CPU: Intel Xeon E5-2690v2(3.0GHz, 10 コア)×2 個, メモリ:64GB) を 1 ノードとする、全 16 ノードで構成される。ノード間は、Mellanox SX6036 Infiniband 36-port 56Gbps Switch で接続されている。提案する並列化手法により並列化した DBN の実行には、全 16 ノードを使用し、CPU のシングルスレッドの実行には、1 ノード 1 コアのみを使用する。

また、コア数の増加による影響評価では、コア数を 1, 20, 40, 80, 160, 320 と変化させて学習時間を測定する。

5.2 評価結果

MPI を用いて並列化した場合と、シングルスレッドの場合における、DBN の学習所要時間を表 1 に示す。

表 1 DBN における学習時間の比較

HPC(MPI)	シングルスレッド	高速化倍率
106.3[s]	6579[s]	61.89

表 1 より、MPI により並列化した場合の実行結果と、シングルスレッドでの実行結果とを比較すると、提案する並列化手法による高速化倍率は 61.89 倍であった。

次に、コア数を 1, 20, 40, 80, 160, 320 と変化させて学習時間を測定した結果を図 3 に示す。

図 3 より、コア数の増加に伴い、DBN の学習時間が短縮されている。

5.3 考察

表 1 の結果から、提案する並列化手法により、DBN の学習を 61.89 倍高速化できたが、コア数の増加に高速化倍率が正比例していない。これは、DBN の学習が一回終わるごとに行う、学習誤差の算出を並列化していないため、この部分の並列化により改善されると

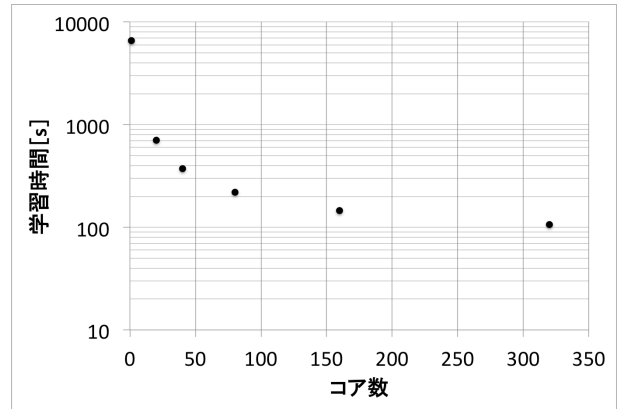


図 3 コア数による DBN の学習時間の変化
考えられる。学習誤差 ϵ は、次式により算出し、過学習を避けるために行っている。

$$\epsilon = \frac{1}{N} \sum_{k=1}^N \sum_{n=1}^{n_o} |T_n^k - O_n^k| \quad (7)$$

ここで、 N は総学習データ数、 n_o は出力細胞数、 T_n^k は n 番目の出力細胞の教師値、 O_n^k は n 番目の出力細胞の出力値である。また、ウェイトの調整量を各ランクから集める際などに行われる、MPI 通信回数の削減により、MPI 通信に伴う時間遅延も改善されると考えられる。

6 まとめ

本研究では、MPI を用いて学習データを個々の MPI スレッドに割り振り、RBM の学習と LR の学習を並列化し、DBN の学習時間を短縮する手法を提案した。

評価結果から、DBN の学習時間の短縮を実現でき、提案する並列化手法の有効性が確認された。また、コア数の増加による学習時間の短縮も確認された。

7 今後の展開

本研究では、16 ノード 320 コアの HPC を用いたが、今後は、よりコア数の多いクラスタシステムを使用し、提案する手法の評価を行う。

参考文献

- [1] Yoshua Bengio, Pascal Lamblin, Dan Popovici and Hugo Larochelle, "Greedy Layer-Wise Training of Deep Networks", *Advances in Neural Information Processing Systems* 19, pp.2-11, 2006
- [2] Hironobu Satoh, Kyoko Shibata, Tomohito Masaki, "Development of an Awakening Behavior Detection System with Kinect", *HCI International 2014 - Posters' Extended Abstracts Communications in Computer and Information Science Volume 435*, pp 496-500, June 2014
- [3] Noel Lopes, Bernardete Ribeiro and Joao Goncalves, "Restricted Boltzmann Machines and Deep Belief networks on Multi-Core Processors", in *The 2012 International Joint Conference on Neural Networks*, June 2012