

行列分解を基盤とした ディープニューラルネットワーク計算法

今倉 暁, 櫻井 鉄也
筑波大学

Matrix computation-based
deep neural network calculation

Akira Imakura, Tetsuya Sakurai
University of Tsukuba

E-mail: imakura@cs.tsukuba.ac.jp

1 はじめに

近年, 画像認識や信号処理など様々な分野においてディープニューラルネットワーク (DNN) が注目され活発に研究が進められている. DNN 計算は, 非線形活性化関数, バイアス項および正則化関数による非線形最適化問題として定式化され, その最適化には, 確率的勾配降下法に基づくバックプロパゲーション (BP) 法 [4] が標準的なアルゴリズムとして利用されている.

一方, 近年我々は全結合ニューラルネットワークに対して, 非線形非負値行列分解 (非線形 NMF) に基づく新しい最適化アルゴリズムを開発した [1, 2, 5]. NMF 型 DNN 計算法は BP 法に対して大きなミニバッチサイズを取ることができるため高い並列性を示すことが確認されている [2].

本稿では, NMF 型 DNN 計算法の概略を示し, 複数ネットワークによる並列化のアイデアを示し, 数値実験からその有効性を確認する.

2 ディープニューラルネットワーク (DNN) 計算

学習データセットの特徴量数を m_{in} , データサンプル数を n , 正解データの次元数を m_{out} とする. また, 教師データを $X \in \mathbb{R}^{m_{\text{in}} \times n}$, 正解データを $Y \in \mathbb{R}^{m_{\text{out}} \times n}$ とする. この時, 重み行列 $W_i \in \mathbb{R}^{m_i \times m_{i-1}}$ およびバイアス項 $\mathbf{b}_i \in \mathbb{R}^{m_i}$ を用いると, 全結合 DNN の目的関数は,

$$\begin{aligned} E(W_1, \dots, W_d, \mathbf{b}_1, \dots, \mathbf{b}_d, X, Y) \\ = D(Y, W_d f(W_{d-1} \cdots f(W_1 X + \mathbf{b}_1 \mathbf{1}^T) \cdots + \mathbf{b}_{d-1} \mathbf{1}^T) + \mathbf{b}_d \mathbf{1}^T) \\ + h(W_1, \dots, W_d, \mathbf{b}_1, \dots, \mathbf{b}_d), \end{aligned} \quad (1)$$

と書くことができる. ここで, $m_0 = m_{\text{in}}, m_d = m_{\text{out}}, \mathbf{1} = [1, 1, \dots, 1]^T \in \mathbb{R}^n$ であり, 関数 f は活性化関数と呼ばれる要素ごとの関数で, 一般的にシグモイド関数や ReLU: $f(X) = \max(0, X)$ などが用いられる. また, h は正則化項である.

Algorithm 1 The nonlinear semi-NMF based method

Input: Input dataset and ground truth X, Y , size of mini-batch s and regularization parameters λ_W, λ_Z

Output: Weight matrices and bias vectors $W_i, \mathbf{b}_i, i = 1, 2, \dots, d$

- 1: Set initial guess $W_1^{(0,0)}, W_2^{(0,0)}, \dots, W_d^{(0,0)}$ and $\mathbf{b}_1^{(0,0)}, \mathbf{b}_2^{(0,0)}, \dots, \mathbf{b}_d^{(0,0)}$
 - 2: Compute a low-rank approximation $X \approx U_1 \Sigma_1 V_1^T$
 - 3: **for** $k = 0, 1, \dots$ **do**:
 - 4: **for** $\ell = 0, 1, \dots, m/s - 1$ **do**:
 - 5: Set the index of mini-batches $\mathcal{J}_\ell^{(k)}$ and $X_\ell = U_1 \Sigma_1 V_1 (\mathcal{J}_\ell^{(k)}, :)^T, Y_\ell = Y(:, \mathcal{J}_\ell^{(k)})$
 - 6: Set $Z_i^{(k,\ell)} = f(W_i^{(k,\ell)} Z_{i-1}^{(k,\ell)})$ for $i = 1, 2, \dots, d - 1$, where $Z_0^{(k,\ell)} = X_\ell$
 - 7: Solve (approximately) semi-NMF (2)
with initial guesses $W_d^{(k,\ell)}, \mathbf{b}_d^{(k,\ell)}, Z_{d-1}^{(k,\ell)}$, and get $W_d^{(k,\ell+1)}, \mathbf{b}_d^{(k,\ell+1)}, \hat{Z}_{d-1}^{(k,\ell+1)}$
 - 8: **for** $i = d - 1, d - 2, \dots, 2$ **do**:
 - 9: Solve (approximately) nonlinear semi-NMF (3)
with initial guesses $W_i^{(k,\ell)}, \mathbf{b}_i^{(k,\ell)}, Z_{i-1}^{(k,\ell)}$, and get $W_i^{(k,\ell+1)}, \mathbf{b}_i^{(k,\ell+1)}, \hat{Z}_{i-1}^{(k,\ell+1)}$
 - 10: **end for**
 - 11: Solve (approximately) nonlinear LSQ (4)
with an initial guess $W_1^{(k,\ell)}, \mathbf{b}_1^{(k,\ell)}$, and get $W_1^{(k,\ell+1)}, \mathbf{b}_1^{(k,\ell+1)}$.
 - 12: **end for**
 - 13: Update $W_i^{(k+1,0)} = W_i^{(k,m/s)}$ and $\mathbf{b}_i^{(k+1,0)} = \mathbf{b}_i^{(k,m/s)}$ for $i = 1, 2, \dots, d$
 - 14: **end for**
-

DNN 計算では、目的関数 (1) を最小化するように、重み W_i およびバイアス項 \mathbf{b}_i を最適化する。

3 NMF 型 DNN 計算法

BP 法では、確率的勾配降下法に基づき、変数を $[W, \mathbf{b}] := [W_1, \dots, W_d, \mathbf{b}_1, \dots, \mathbf{b}_d]$ の微分値を元に、目的関数 (1) の最小化を行う。

$$[W^{(k+1)}, \mathbf{b}^{(k+1)}] = [W^{(k)}, \mathbf{b}^{(k)}] - \eta_k \frac{\partial E}{\partial [W, \mathbf{b}]}$$

一方、NMF 型 DNN 計算では、誤差関数 $D(\cdot, \cdot)$ を二乗誤差、活性化関数を ReLU とし、非線形行列分解を用いた交互最適化に基づき重み行列 W_i およびバイアスベクトル \mathbf{b}_i を層毎に最適化する。以下では NMF 型 DNN 計算法の概略を示し、並列ネットワークによる並列化法のアイデアおよび数値実験結果を示す。

3.1 アルゴリズムの概略

NMF 型 DNN 計算法は，出力層，中間層，入力層の 3 つの最適化ステップからなり，各最適化ステップは semi-NMF，非線形 semi-NMF および非線形最小二乗問題により求解される．第 k 反復目において，行列 $Z_i^{(k)} \in \mathbb{R}^{m_i \times n}$ を

$$\begin{aligned} Z_0^{(k)} &:= X, \\ Z_i^{(k)} &:= f(W_{i-1}^{(k)} \cdots f(W_1^{(k)} X + \mathbf{b}_1^{(k)} \mathbf{1}^T) \cdots + \mathbf{b}_{i-1}^{(k)} \mathbf{1}^T) + \mathbf{b}_i^{(k)} \mathbf{1}^T, \quad i = 1, 2, \dots, d-1 \end{aligned}$$

と定義する．以下では，NMF 型 DNN 計算法の各ステップの概要を示す．

出力層において， Z_{d-1} を用いると，

$$Y \approx W_d Z_{d-1} + \mathbf{b}_d \mathbf{1}^T = [W_d, \mathbf{b}_d] [Z_{d-1}; \mathbf{1}^T]$$

となることが期待される．ここで，活性化関数が ReLU であるため， $Z_{d-1} = f(W_d Z_{d-2} + \mathbf{b}_d \mathbf{1}^T) = \max(W_d Z_{d-2} + \mathbf{b}_d \mathbf{1}^T, O) \geq 0$ である．従って，非負制約付きの行列分解により， W_d, \mathbf{b}_d および Z_{d-1} を計算することを考える． k 反復目の近似解 $W_d^{(k)}, \mathbf{b}_d^{(k)}, Z_{d-1}^{(k)}$ を初期近似解として，正則化項付き semi-NMF

$$\begin{aligned} [W_d^{(k+1)}, \mathbf{b}_d^{(k+1)}, \widehat{Z}_{d-1}^{(k+1)}] \\ = \arg \min_{W_d, \mathbf{b}_d, (Z_{d-1} \geq 0)} \|Y - [W_d, \mathbf{b}_d] [Z_{d-1}; \mathbf{1}^T]\|_F^2 + \lambda_W \|[W_d, \mathbf{b}_d]\|^2 + \lambda_Z \|Z_{d-1}\|^2, \end{aligned} \quad (2)$$

を計算することにより， $W_d^{(k+1)}, \mathbf{b}_d^{(k+1)}$ および $\widehat{Z}_{d-1}^{(k+1)}$ を更新する．

中間層 ($i = d-1, d-2, \dots, 2$) において， $Z_i = f(W_i Z_{i-1} + \mathbf{b}_i \mathbf{1}^T)$ より，

$$\widehat{Z}_i^{(k+1)} \approx f(W_i Z_{i-1} + \mathbf{b}_i \mathbf{1}^T)$$

が期待される．従って，(2) と同様に，正則化項付きの非線形 semi-NMF

$$\begin{aligned} [W_i^{(k+1)}, \mathbf{b}_i^{(k+1)}, \widehat{Z}_{i-1}^{(k+1)}] \\ = \arg \min_{W_i, \mathbf{b}_i, (Z_{i-1} \geq 0)} \|\widehat{Z}_i^{(k+1)} - f([W_i, \mathbf{b}_i] [Z_{i-1}; \mathbf{1}^T])\|_F^2 + \lambda_W \|[W_i, \mathbf{b}_i]\|^2 + \lambda_Z \|Z_{i-1}\|^2, \end{aligned} \quad (3)$$

を計算することにより， $W_i^{(k+1)}, \mathbf{b}_i^{(k+1)}$ および $\widehat{Z}_{i-1}^{(k+1)}$ を更新する．

入力層では， $Z_0 = X$ であるため，最適化問題は非線形の最小二乗問題

$$[W_1^{(k+1)}, \mathbf{b}_1^{(k+1)}] = \arg \min_{W_1, \mathbf{b}_1} \|\widehat{Z}_1^{(k+1)} - f(W_1 X + \mathbf{b}_1 \mathbf{1}^T)\|_F \quad (4)$$

となる．

最後に，得られた重み行列 $W_i^{(k+1)}$ およびバイアスペクトル $\mathbf{b}_i^{(k+1)}$ を用いて， $Z_i^{(k+1)}$ を更新する．実用上の観点から，アルゴリズムは学習データの部分サンプルを逐次的に用いるミニバッチ学習が行われる．また，計算コストの削減および数値安定性の改善のため，教師データの低ランク近似を用いる．NMF 型 DNN 計算法のアルゴリズムは Algorithm 1 のように示される．

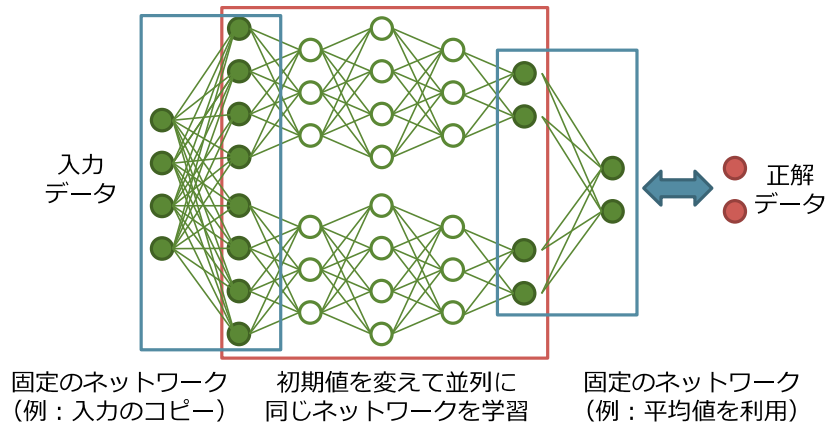


図 1: 複数ネットワークによる並列化.

3.2 複数ネットワークによる並列化

DNN 計算の並列学習として、図 2 に示すような複数のネットワークを組み合わせた構造を考える。このような複数ネットワークの学習は、各ネットワークを並列に学習可能であり、また中間層のユニット数が増えることで高い認識性能を発揮することが期待される。また、並列ネットワーク数を p とすると、各中間層のユニット数が p 倍された大きなネットワークの部分構造になっているため、大きなネットワークを学習する際の前処理としても利用することができる。

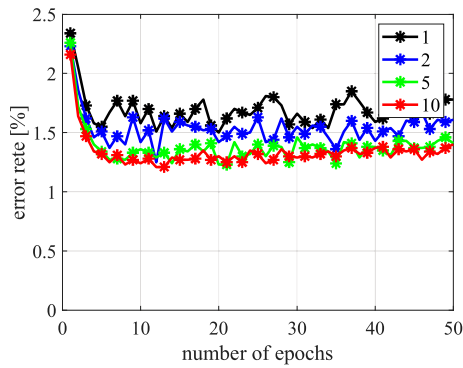
3.3 数値実験

本節では、手書き数字認識 (MNIST) [3] を用い、NMF 型 DNN 計算法 (Algorithm 1) を BP 法と比較する。画像サイズは 28×28 であり、学習データ数は 60000、テストデータ数は 10000 である。DNN の層構造として、 $[768 - 1000 - 500 - 10]$ のネットワークを最大 10 個用いた並列ネットワークを考える。

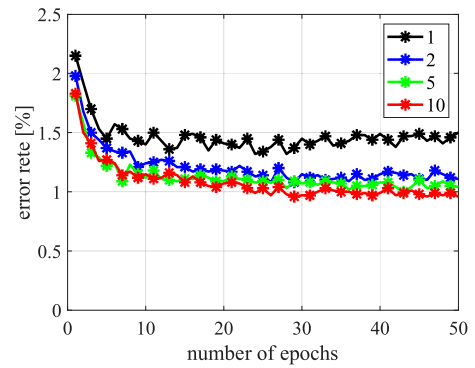
並列ネットワーク数 $p = 1, 2, 5, 10$ による BP 法および NMF 型 DNN 計算法の収束履歴を図 2 に示す。実験結果から、複数ネットワークを用いることで認識性能の改善が可能であることがわかる。また、BP 法 (図 2(a)) と NMF 型 DNN 計算法 (図 2(b)) を比較すると NMF 型 DNN 計算法がより認識性能の改善効果が高いことがわかる。

4 まとめ

本稿では、近年提案された NMF 型 DNN 計算法の概略を示し、複数ネットワークによる並列化のアイデアを示した。また、MNIST を用いた数値実験から複数ネットワークを用いることで NMF 型 DNN 計算法の認識性能が改善できることが示された。



(a) BP 法



(b) NMF 型 DNN 計算法

図 2: 複数ネットワークによる並列化.

謝辞

本研究は JST/ACT-I (No. JPMJPR16U6), JSPS 科研費 (Nos. 17K12690, 18H03250, 19KK0255) の助成を受けたものです。

参考文献

- [1] R. Arai, A. Imakura, T. Sakurai, An improvement of the nonlinear semi-NMF based method by considering bias vectors and regularization for deep neural networks, *International Journal of Machine Learning and Computing (IJMLC)*, **8** (2018), 191–197.
- [2] A. Imakura, Y. Inoue, T. Sakurai, Y. Futamura, Parallel implementation of the nonlinear semi-NMF based alternating optimization method for deep neural networks, *Neural Processing Letters*, **47** (2018), 815–827.
- [3] Y. LeCun, The MNIST database of handwritten digits, <http://yann.lecun.com/exdb/mnist>.
- [4] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning Representations by Back-Propagating Errors *Nature*, **323** (1986), 533–536.
- [5] T. Sakurai, A. Imakura, Y. Inoue, Y. Futamura, Alternating optimization method based on nonnegative matrix factorizations for deep neural networks, In: A. Hirose, S. Ozawa, K. Doya, K. Ikeda, M. Lee, D. Liu. eds, *Neural Information Processing, ICONIP 2016, Lecture Notes in Computer Science*, **9950** (2016), 354–362.