G-009

オートエンコーダの生体妥当な誤差伝播学習則の実験的比較評価
An Experimental Evaluation for Extracting Features using Autoencoder by Biologically Plausible Learning Rules

大濱 吉紘†
Yoshihiro Ohama

1. はじめに

近年，深層学習とその応用に関する研究開発は，画像や言語の認識・生成に留まらず，多様な分野で精力的に進められている．この中で深層学習が果たす主な機能の 1 つは，大量かつ多次元のデータセットからの特徴量抽出である．これは多次元データを入力として，特徴量が表出する時間的・空間的に低次元化された情報表現が可能な階層を含む，多層構造の神経回路(DNN; Deep Neural Network)を，適切な評価関数の下で学習させることで実現される．最も基礎的な DNN とされるオートエンコーダ(AE; AutoEncoder)は，隠れ層に明示的な低次元の特徴抽出層(ボトルネック層)を設けた砂時計型の神経回路であり，入力情報と出力情報が一致するような恒等写像を学習させることで，ボトルネック層に低次元の特徴量を表出させるものである．

DNN の学習には，一般に勾配法による最小化アルゴリズムに相当する誤差逆伝播学習則(BPL; Back-Propagation Learning rule)が利用されるが，莫大な回数の繰り返し学習が必要となる．この原因の 1 つとして，神経回路の出力誤差を入力側に向かって伝播していく過程で，特定の階層での勾配が十分に小さな状態に陥って伝播誤差が消失し，学習が進まなくなる問題が知られている[1]．この問題を回避するために，深層学習の実応用にあたっては，神経素子の勾配が小さくなりにくいように活性化関数を工夫したり(例えば[2])，複数の階層間をバイパスするような神経回路構造としたり(例えば[3])することが多い．

一方で，生体の神経細胞は誤差信号を逆伝播できないと考えられており，新たに生体妥当な学習則を求める動きが広がりつつある．例えば，誤差の順方向伝播のみを用いる誤差順伝播学習則(FPL; Forward-Propagation Learning Rule)[4]及び目標誤差伝播学習則(TPL; Target Propagation Learning rule)[5]が提案されている．また別のアプローチとして，誤差伝播専用経路を用いるランダムシナプスフィードバック重み更新則(RSF; Random Synaptic Feedback weights)[6]及び直接フィードバック調整則(DFA; Direct Feedback Alignment)[7]が提案されている．このうち，FPL 及び TPL は AE での学習を想定しているが，RSF 及び DFA は AE での学習について報告がなされていない．

生体妥当な学習則の工学的応用として，AE の学習時に出力誤差を神経回路の上で逆伝播させる BPL だけでなく，順伝播させる FPL を並列計算で実行し，正味計算時間を削減する試みを我々は開始している[8]．並列実行が有効に働くためには，各伝播誤差による学習性能が同程度であることが望ましい．本稿では，手書き数字画像データセット MNIST に対する AE による特徴量抽出に，BPL，TPL，RSF，DFA を適用して学習型判別器による判別精度を比較し，並列実行が可能かどうかを実験的に確かめる．

† 株式会社豊田中央研究所　Toyota CRDL, Inc.

2. 実験方法

2.1 実験手順

本稿での実験手順を，図 1 に示す．手書き数字画像を訓練画像とテスト画像に分け，訓練画像を用いて比較対象とする学習則で AE を学習させて，特徴量抽出器を構成する．そして，この特徴量と手書き数字画像の正解ラベルから，判別器を学習させる．さらに，この特徴量抽出器を用いてテスト画像の特徴量を抽出して判別器に入力し，判別結果と正解ラベルを比較して，判別精度を算出する．このとき，異なる初期状態の AE を多数用意して，判別精度の統計的な差を評価することで，学習則間の比較を行う．

MNIST 手書き数字画像は正解ラベル付きの 28x28[pix]のバイナリ画像で構成され，本実験では訓練画像 55,000 枚，テスト画像 5,000 枚を用いる．また，AE の入出力層を 784 次元，ボトルネック層を 2 次元とする．

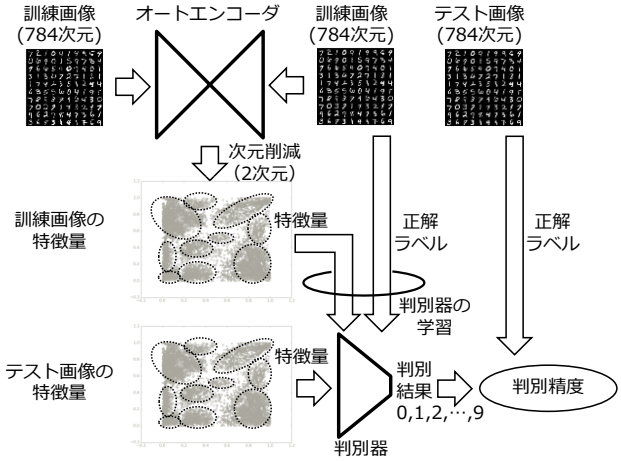

図 1　実験の流れ

2.2 比較対象とする学習則

入力をd次元ベクトルx，出力d次元のベクトルyとする非線形関数として AE を記述する．なお便宜上，AE は 5 層構造とする．

$$y = h(x)$$
$$h(x) = h_5\left(h_4\left(h_3\left(h_2\left(h_1(x)\right)\right)\right)\right) \quad (1)$$

したがって AE は，各層の変換$h_l(s_l)$を直列接続したものである．

$$s_{l+1} = h_l(s_l)$$
$$y = h_5(s_5)$$
$$s_1 = x \qquad (l = 1,2,3,4) \quad (2)$$

各層間の変換には様々な手法が提案されているが，ここでは最も単純なものとして，前向き結合神経回路モデル(FNN; Feed-forward Neural Network)を考える．

$$s_l = W_l h_{l-1} \quad (3)$$

ここで，W_lは各層の全ての神経素子同士の結合の強さを表すも

ので結合パラメータ，結合重み，結合荷重などと呼ばれる．またその構造から，FNN は全結合型神経回路モデルとも称される．さらに各層では，以下の非線形変換が行われる．

$$h_l = \sigma(s_l) \quad (4)$$

σは神経素子での単調増加の非線形変換であり，活性化関数と呼ばれる．

次に，AE の学習を実現するための問題設定について述べる．入力xがd次元のベクトルをN個並べた行列であるような，いわゆるデータセットである場合を考える．

$$x = [x_1, x_2, \cdots, x_N]^T \quad (5)$$

このとき，出力y及び各層h_lの入力s_lもまた，同様に行列となる．

$$y = [y_1, y_2, \cdots, y_N]^T$$
$$s_l = [s_{l,1}, s_{l,2}, \cdots, s_{l,N}]^T \quad (6)$$

AE の学習とは，できるだけ入出力が一致するような結合パラメータW_lを見つけることである．入出力の一致度を測る指標$J(x, y)$を，例えば次の二乗誤差の総和のコスト関数で定義する．

$$J(x, y) = \frac{1}{2}\sum_{n=1}^{N} \|x_n - y_n\|^2 \quad (7)$$

コスト関数Jに含まれるyはW_lの関数であるため，学習はJを最小化する結合パラメータ$\widehat{W}_l$を見つける問題に帰着される．

$$\widehat{W}_l = \underset{W_l}{\operatorname{argmin}} J \quad (8)$$

BPL は，この問題を勾配降下法で解くことに相当する．

TPL では AE が前学習済みであれば，出力信号yをフィードバックして AE に入力して得られる各層の差分$\Delta h_{l,n}^{TPL}$により，逆伝播誤差を近似できることを利用し，結合パラメータを更新する．

$$\widehat{w}_{l,i+1} = \widehat{w}_{l,i} + \varepsilon \sum_{n=1}^{N} h_{l,j,n} \left.\frac{\partial \sigma}{\partial s_{l+1}}\right|_{s_{l+1}=s_{l+1,n}} \Delta h_{l+1,n}^{TPL}$$
$$\Delta h_{l,n}^{TPL} = h_l(y_n) - h_l(x_n) \quad (9)$$

ここで，εは学習係数と呼ばれる微小な正定値であり，iは繰り返し学習の更新ステップ数を表している．

RSF は，ランダム重みの誤差伝播専用経路B_lを，神経回路モデルに追加し，次式で結合パラメータを更新する．

$$\widehat{w}_{l,i+1} = \widehat{w}_{l,i} + \varepsilon \sum_{n=1}^{N} h_{l,j,n} \left.\frac{\partial \sigma}{\partial s_{l+1}}\right|_{s_{l+1}=s_{l+1,n}} \Delta h_{l+1,n}^{RFW}$$
$$\Delta h_{l,n}^{RFW} = B_l \Delta h_{l+1,n}^{RFW}$$
$$\Delta h_{5,n}^{RFW} = x_n - y_n \quad (10)$$

この更新則により，BPL による$\widehat{w}_{l,i}$の修正方向からかい離せずに学習は収束するということが，実験的に示されている．

RSF では，安定した学習を行うためにランダム重みB_lの値は小さなものとする必要がある．しかし，オートエンコーダのような多段に重ねた神経回路モデルでは，B_lの各要素が 1 を下回る小さな値である場合に，出力層から離れるにつれて誤差が小さくなり，学習が進まなくなる．この問題を回避するために，DFA が提案されている．

$$\widehat{w}_{l,i+1} = \widehat{w}_{l,i} + \varepsilon \sum_{n=1}^{N} h_{l,j,n} \left.\frac{\partial \sigma}{\partial s_{l+1}}\right|_{s_{l+1}=s_{l+1,n}} \Delta h_{l+1,n}^{DFA}$$
$$\Delta h_{l,n}^{DFA} = B_l \Delta h_{5,n}^{DFA}$$
$$\Delta h_{5,n}^{RFW} = x_n - y_n \quad (11)$$

この方法では，$\Delta h_{l,n}^{DFA}$を計算するために，出力層の誤差である$h_{5,n}^{DFA}$を直接用いている．つまり出力層の誤差出力層から各層への直接的なランダム重みの誤差伝播専用経路としてB_lが用意されており，RSF で見られるような伝播誤差が小さくなって学習が進まなくなる問題を回避している．

3. 実験結果

特徴抽出器である AE を 9 層の FNN で構成し，各層の神経素子数を入力層から順に，784，500，250，2，250，500，784 とした．また活性化関数は，入出力層及びボトルネック層は線形関数とし，その他の層はシグモイド関数とした．そして，制限付きボルツマンマシンを用いた Hinton の前学習法[9]により異なる初期状態を持つ 100 個の AE を用意した．また，判別器は多クラス判別 SVM を用いた．

2.1 節で述べた実験手順により，BPL，TPL，RSF，DFA について判別精度を比較した結果を，図 2 に示す．この結果から，TPL は BPL で学習した AE と同程度の判別精度を達成する特徴量を抽出できた．一方，RSF 及び DFA については，BPL 及び TPL に対して有意に劣っていた．このため，BPL と並列に実行する枠組み[8]の学習則に TPL は利用可能と考えられるが，RSF と DFA は学習を妨げる可能性があり，何らかの工夫が必要であることが示唆された．

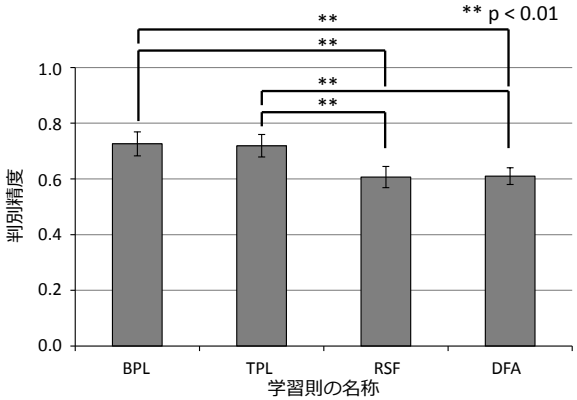

図 2　学習則ごとの判別精度の比較結果

4. おわりに

本稿では，AE の学習時に BPL と並列計算可能な誤差伝播学習則を実行して学習の正味計算時間を削減する手法を想定し，生体妥当な学習則の利用可能性を実験的に評価した．その結果，TPL について利用可能である見込みを得た．

参考文献

[1] Hochreiter. S., Bengio, Y., Frasconi, P., Schmidhuber, J., “Gradient flow in reccurent nets: the difficulty of learning long-term dependencies”, In: Kremer, C., Kolen, J. F. (eds.) Dy-namical Recurrent Neural Networks. IEEE Press. (2001).
[2] Nair, V., Hinton, G. E., “Rectified Linear Units Improve Restricted Boltzmann Machines”, Proc. ICML 2010, pp.807-814 (2010).
[3] Srivastava, R. K., Greff, K., Schmidhuber, J., “Highway Networks”, arXiv:1505.00387 (2015).
[4] Ohama, Y., Fukumura, N., Uno, Y., “A Simplified Forward-Propagation Learning Rule Applied to Adaptive Closed-Loop Control”, Proc. ICANN 2005, pp. 437-443 (2005).
[5] Bengio, Y., “How auto-encoders could provide credit assignment in deep networks via target propagation”, arXiv:1407.7906. (2014).
[6] Lillicrap, T. P., Cownden, D., Tweed, D. B. Akerman, C. J., “Random synaptic feedback weights support error backpropagation for deep learning”, Nature Communications 7, 13276 (2016).
[7] Nøkland, A., “Direct Feedback Alignment Provides Learning in Deep Neural Networks”, Proc. NIPS 2016, pp. 1037-1045 (2016).
[8] Ohama, Y., Yoshimura, T., “A Parallel Forward-Backward Propagation Learning Scheme for Auto-Encoders”, Proc. ICONIP 2017, pp. 125-136 (2017).
[9] G. E. Hinton, and R. R. Salakhutdinov, “Reducing the Dimensionality of Data with Neural Networks”, Science, Vol.313, Issue 5786, pp.504-507.