

アナログ AI チップのオンライン学習に向けた 改良型デジタルバックプロパゲーション法の提案

A study for improved digital backpropagation towards online training for analog AI chip

金子竜也¹⁾, 山岸善治¹⁾, 百瀬啓²⁾, 浅井哲也²⁾¹⁾ 北海道大学大学院情報科学院²⁾ 北海道大学大学院情報科学研究院

E-mail:kaneko.tatsuya.cy@ist.hokudai.ac.jp

Abstract— An edge AI device, which has non-differentiable function for activation function (e.g. RAND chip) can not trained using standard backpropagation. On these devices we required Digital backpropagation (DBP); it is able to training neural network by predicting supervisor for hidden layer. Despite DBP can solve XOR or XNOR training task, it remains a challenging problem for more difficult tasks. The original DBP is incapable of classifying Iris dataset and MNIST dataset. This problem may be attributed to predicting supervisor for hidden layers. As the increasing number of output layer's node caused increasing difficulty to predict supervisor. We show improving MNIST classification result from about 46% to about 86% by introducing cluster node for output layer and partial connected layer between hidden layer and output layer.

Keywords— Machine Learning, Edge AI, Training Algorithm, Backpropagation, Online training

1 はじめに

現在隆盛期にある人工知能を支えてきたのは TPU [1] や数百ワット規模の GPU を有するクラウド型 AI に他ならない。これとは対照に、自動運転車やスマートフォンその他のスマート製品を含む現実世界での応用を目指したエッジ型 AI では Edge TPU や Jetson nano 等が推論エンジンとして利活用され始めている。しかし、従来のエッジ型 AI は潤沢な演算リソースを持つクラウド上で学習した AI モデルを縮小したものをダウンロードすることで成り立っている。エッジデバイス単体では推論機能のみを有しており、学習機能については人間やそれを取り巻く環境をオンラインで学習することが不可能である点など種々の問題を抱えている。エッジベースのアナログ AI チップの一つとして RAND チップ [2] が存在する。RAND チップのあるノードの出力はステップ関数により 0 or 1 となる。RAND チップ上に学習処理を実装する場合には、ステップ関数が微分不可能であることから SGD [3] に代表されるような通常の BP ではなく DBP (Digital BackPropagation) [4] が必要となる。以下に DBP について簡単な説明をまとめる。

1.1 DBP

図 1 に DBP による中間層の教師信号予測の概要図を示す。通常の逆伝播で得られる出力層の誤差を基に以下の式により決定因子 S_j を求める。

$$S_j = \sum_n (O_n - T_n) W_{nj} = \sum_n S_{jn} \quad (1)$$

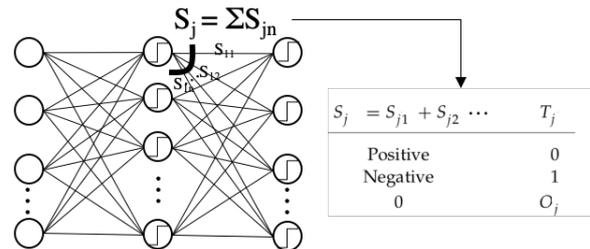


図 1: DBP による中間層の教師信号予測の概要図。各決定因子 S_{jn} の総和を取りその値により教師信号 T_j を予測する。

ここで O_n は出力層の出力値, T_n は教師信号であり, W_{nj} は中間-出力層間の重み, j は中間層のノード, n は出力層のノードに対応する値である。中間層の教師信号はこの決定因子の符号を基に決定される。DBP は出力層のノード数の増加に従い教師信号の予測難度も増加していくという問題点を抱えている。そこで、本論文ではニューラルネットワークの構造を変化させることで予測難度増加による識別精度の低下を解決できることを示す。

2 提案手法

DBP は大きなネットワーク構造が必要となる Iris データセットの識別や MNIST 分類問題に対しては十分な性能を発揮できないことが確認できた。本章では、この問題を解決するためにクラスタの導入と中間-出力層間部分結合の導入を提案する。

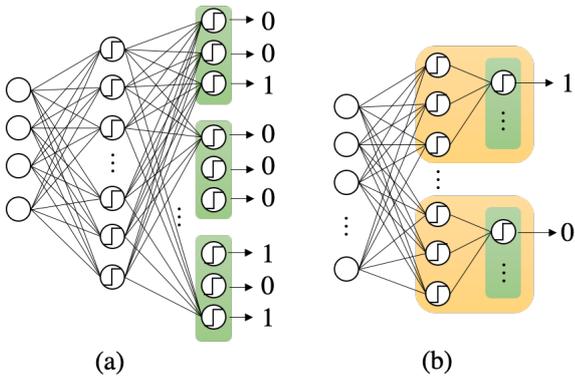


図 2: 教師信号予測精度向上のためのネットワークモデル図. (a) 出力層のノード数を 1 から N 個に増やすクラスタ化を行なったネットワーク. (b) 中間-出力層間の重みを全結合からクラス毎に部分結合させたネットワーク.

2.1 クラスタの導入

活性化関数にステップ関数を用いた場合にはシグモイド関数等を用いた場合と異なり 0 or 1 の出力となることから複数の出力ノードが 1 を出力するという識別ミスが発生しやすくなる. このことが識別精度の低下を招いていると考えられることから, 図 2 中 (a) に示すように出力ノードの 1 つが 1 つのクラスに対応するのではなく複数個で 1 つのクラスに対応するというクラスタの導入を行った. クラスタは奇数個のノードで構成され最終的な出力は各クラス内で多数決を取ることで得られる. クラスタ化により冗長性の確保が可能となる.

2.2 中間-出力層間部分結合

DBP の本質的な問題点として出力層のノード数が増えるに従い中間層の教師信号の予測精度は不確かなものになるという点が挙げられる. これは教師信号の予測を逆伝播された値の総和を符号化することで求めていることに起因する. この教師信号の予測能力低下による識別精度の低下は, 出力層のノード数が 3 である Iris データセットの場合にクラスタの導入で回避可能であった. しかし, 10 個の出力ノードを持つ MNIST データセットの場合には回避不可能であることが確認され, この解決法として図 2 中 (b) に示すように中間層と出力層間を全結合にするのではなくクラス毎に独立した結合をする部分結合方式を導入した.

3 評価

評価は Iris データセット, MNIST データセットで行なったが紙面の都合上 MNIST 評価について記述する. MNIST データサイズは 14×14 に縮小したものを用了. ネットワークの各ノード数は入力層 196, 中間層 64×10 , 出力層 $10 \times$ クラスサイズである. 活性化関数は中間層, 出力層ともにステップ関数である. ミニバッチサイズはオンライン学習を前提としていることから 1 である. DBP の学習率は 0.01 とし, 比較対象である浮動小

表 1: MNIST 認識精度. 試行回数は各 5 回である.

	mean	median	max	min
BP (float)	93.56	93.10	94.90	92.80
BP (fixed)	92.90	91.58	92.90	90.10
DBP (original)	46.96	46.80	50.00	44.80
DBP (cluster:1)	75.00	75.50	77.30	71.70
DBP (cluster:3)	84.32	84.90	86.20	82.50
DBP (cluster:5)	85.60	85.40	87.60	83.20
DBP (cluster:31)	78.74	79.40	80.10	76.10

数点方式 BP と 16 bit 固定小数点方式の学習率は 0.25 とした. 比較対象のネットワークのノード数は入力層 196, 中間層 64, 出力層 10 であり, 活性化関数はシグモイド関数である. 表 1 に認識精度を示す. この認識精度は学習データ数が 6000, テストデータ数は 1000 である時に BP は 30epoch, DBP は 6epoch 学習をしたものである. クラスタと部分結合の導入によって識別精度が約 47% から約 86% にまで向上できたことが確認できた. クラスタサイズの増加は冗長性の確保につながり, 汎化性能向上に結びつくことから識別精度の向上に寄与していると考えられる. 一方であまりに大きなクラスサイズでは教師信号の予測難度の増加を引き起こすために識別精度の低下を招いていると予測される.

4 まとめ

本論文では, DBP の出力ノード数の増加に伴う教師信号の予測精度低下という問題点を解決するためにクラスタの導入と部分結合の導入について提案を行なった. 本提案により従来の DBP 実装では約 47% だった MNIST の識別精度が約 86% に向上した.

謝辞

本研究で多大なご支援を頂きましたヌヴォトン テクノロジージャパン株式会社に深謝致します. また, この成果は, 国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものです.

参考文献

- [1] Jouppi, N., Young, C., & Yoon, D., et. al., (2017). In-Datcenter Performance Analysis of a Tensor Processing Unit. *Proc. the 44th Annual International Symposium on Computer Architecture*, pp.1 – 12.
- [2] Mochida, R., Kouno, K., & Gohou, Y., et. al., (2018). A 4M Synapses integrated Analog ReRAM based 66.5 TOPS/W Neural-Network Processor with Cell Current Controlled Writing and Flexible Network Architecture. *2018 IEEE Symposium on VLSI Technology*, pp. 175–176.
- [3] Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. *Proc. COMPSTAT'10*, pp. 177–186.
- [4] Oohori, T., Naganuma, H., & Watanabe, K. (2007) A new backpropagation learning algorithm for layered neural networks with nondifferentiable units. *Neural Comput.* pp. 1422–1435.