1P-04

Forward 計算と Genetic Algorithm による Binary Weight 可変構造 DNN の最適化

渡部 直弘† 高木 俊平† 大塚 卓哉†† 北澤 仁志†

†東京農工大学 ††日本電信電話株式会社 NTT 先端集積デバイス研究所

1 はじめに

近年，Deep Neural Network(DNN) の研究が盛んに行われており，DNN のノード間の weight(重み) をバイナリ値とした BinaryConnect [1] が提案されている．weight を ±1, 0 の 3 値とすることで DNN における乗算を加減算に置き換えることができ，ハードウェア実装においては乗算器が不要となるため，資源の節約，識別器の小型化や省電力化が期待できる．しかし，Binary Weight DNN の学習では，Back Propagation 法 (BP) が必ずしも有効ではない．また，DNN では構造の決定も大きな課題となる．本稿では，Binary Weight DNN において Forward 計算と Genetic Algorithm を主とした非線形最適化手法により，BP を用いないで DNN の weight とネットワーク構造を同時に最適化する方法を示す．

2 Binary Weight DNN

BinaryConnect [1] の文献においては，weight を ±1 に制限して学習を行った場合でも，weight が実数値の場合に対し遜色ない結果となることが報告されている．しかし，予備実験の結果，表 1 に示すように我々の実験ではほとんど全てのデータで Binary Weight の BP の結果は大きく劣化した．そこで，Genetic Algorithm(GA)，Simulated Annealing(SA)，Particle Swarm Optimization(PSO) などの非線形最適化手法による改善を試みた．

3 DNN の Forward 計算と Genetic Algorithm による最適化手法

3.1 Forward 計算による最適化

BP を用いずに Forward 計算のみによる最適化の概略図を図 1 に示す．Forward 計算では，入力層から出力層へ向かい各ノード値とそのノードと結合している weight との積和演算を行う．ただし，weight はバイナリ値であるため乗算は加減算となる．Forward 計算後の学習データに対する識別率を目的関数とし，非線形最適化手法を用いて weight 及び，構造を更新する．この手法では BP を用いないため，活性化関数として微分不可能な関数を用いることが可能であり，ノードの計算には積和演算だけでなく，And，Or，Eor や max，min など任意の関数を用いることも可能である．また，パイプライン型のハードウェアアーキテクチャにより高速演算が可能である [2]．最適化手法として，GA，SA，PSO，Covariance Matrix Adaptation Evolution Strategy(CMA-ES)，Monte-Carlo Tree Search(MCTS) などが考えられるが，本研究ではバイナリ値でも適用できる GA，SA，PSO を組み合わせて適用した．

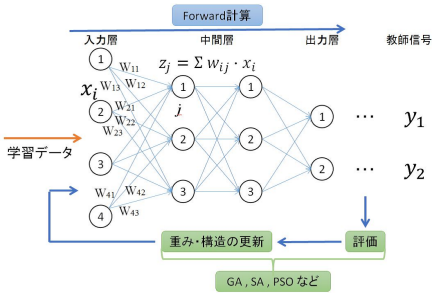

図 1 Forward 計算概略図

3.2 Genetic Algorithm

最適化手法の DNN への適用方法として，1 つの探索点を DNN のノード及び，weight の集合として定義する．GA においては 1 探索点を 1 個体とし，個体の各遺伝子が 1 つ 1 つのノードと weight を表現している．図 2 に DNN への適用方法及び，weight と構造の最適化概略図を示す．weight 値はバイナリの ±1 に加えて，weight を間引くことと同義である 0 を含めた 3 値である．また，あらかじめ定めた上限ノード数から，ノードの drop を行うことによりネットワーク構造を変化させる．なおノードの drop は入力層にも適用する．各遺伝子を GA で制御することにより，DNN の weight とネットワーク構造を同時に最適化していく．

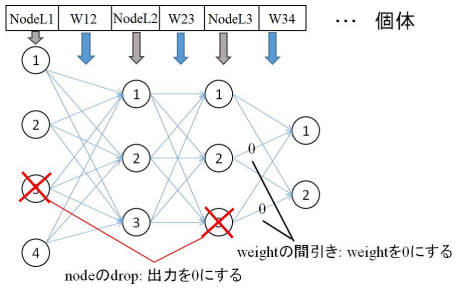

図 2 DNN 構造最適化

3.3 Simulated Annealing

SA の考え方に基づき，識別率の変動が小さくなったら，GA の操作である突然変異の比率を上げるようにし，GA と SA を組み合わせた．

3.4 Particle Swarm Optimization

構造の変更はできないが，近傍を探索して解を改善する機能は PSO の方が優れていると考えられるため PSO も用いた．

4 実験結果

4.1 使用データ

本実験では 3 つの人工データと 5 つの実データを使用した．各データについて以下にまとめる．(*nf*:特徴量の次元)

(a):minLast5．人工データ．実数の時系列データにおいて，現在値が過去 5 入力の最小値である時刻を検出．*nf*=10．(現入力と過去 9 入力を 10 特徴量とする．)

Optimization of Binary-Weight Variable-structure DNN by Forward Calculation and Genetic Algorithm

† Naohiro Watanabe, Shunpei Takaki, Hitoshi Kitazawa (Tokyo University of Agriculture and Technology)

†† Takuya Otsuka (NTT Device Technology Labs, NTT Corporation)

(b):Q4_3120．人工データ．0〜3 の 4 値の時系列データにおいて直前 4 入力が 3，1，2，0 となる時刻を検出．nf=10.
(c):sinPsinR．人工データ．Gaussian ノイズを付与した 2 つのサイン波を足し合わせた波形から谷の部分を判別．nf=10.
(d):Parking．駐車場の監視カメラ映像より背景差分で得られた連結領域について移動物体か日照変動かを判別．nf=10.
(e):South．(d) とは異なる場所の同様のデータ．nf=10.
(f):cp822x．加速度計を用いたインフラの劣化検出データ．3 つの構造物より得られた x 方向加速度の 256 点 FFT パワースペクトルの log．nf=128.
(g):cp822y．同上 y 方向．nf=128.
(h):cp822z．同上 z 方向．nf=128.

4.2 提案手法による学習結果

4.2.1 学習過程

データ (a) に対して，SA を組み合わせた手法と組み合わせない手法の結果を 図 3 に示す．DNN の構造は 3 層で，中間層のノード数は 20 とした．また，GA のパラメータとして，親個体数 100，交叉・突然変異を行い生成する子個体数 200，一点交叉法，突然変異率 3〜10%，ルーレット選択とした．図 3 より，SA を組み合わせ突然変異率を変化させた場合では局所解からの脱出効果を確認できた．

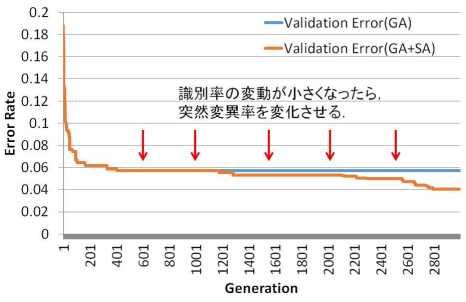

図 3　提案手法による学習過程

4.2.2 構造最適化の効果

データ (b) は識別に必要な情報が現時刻付近の 4 特徴量のみである．バイナリ DNN で，構造の最適化を行わない BP で学習させた場合，本データに対しては全く識別ができなかった．一方，構造の最適化を行う提案手法の学習結果では，入力層ノードのうち必要な 4 特徴量以外のノードが drop され，100% の識別ができていた．提案手法にはこのように学習データに合わせて DNN の構造を最適化する効果があることを確認した．

4.3 提案手法と他識別器の比較

提案手法と他識別器による 8 つのデータに対する識別率を比較した結果を 表 1 に示す．識別器は以下の通りである．
(1)SVM: cost=10，gamma=0.1
(2)BP(real):実数値 DNN を BP で学習．
(3)BP(binary):バイナリ DNN を BP で学習．
(4)BP+GA(binary):バイナリ DNN を BP で学習 GA で改善．
(5)BP+PSO(binary):バイナリ DNN を BP で学習 PSO で改善．
(6)GA(binary):バイナリ DNN(乱数初期値) を GA で改善．
(●)BP+GA(binary):(4) の識別器をテストデータで最適化．
なお，最適パラメータはデータにより異なるため，数通り試して最良値を示した．表 1 より，バイナリ DNN で比較した際，ほぼ全てのデータに対して提案手法である (4)，(5)，(6) は BP による学習 (3) よりも優れていたが，実数値を使用した BP(2) や SVM(1) と比較すると識別精度が劣化した．原因としてバイナリ化と最適化手法が考えられる．そこで，binary weight での解の有無を調査するためテスト用データを用いて最適化を行った．結果として実数値の学習に近い値が得られたことから，binary weight でも解は存在すると考えられる．今後，最適化手法の改善を進めることで，現在 DNN の計算に多用されている GPU に置き換わる，小型省電力な識別器の実現が期待できる．

5 まとめ

本稿では，Forward 計算と Genetic Algorithm により Binary Weight DNN の weight と構造の同時最適化手法を示した．また，Binary Weight DNN での BP による学習と提案手法による学習の比較では，提案手法による学習で識別精度が向上することを示した．

参考文献

[1] M.Courbariaux, et al., ”BinaryConnect:Training Deep Neural Networks with binary weights during propagations”, Advances in Neural Information Processing System，2015
[2] 高木ら,”Genetic Algorithm に基づく Binary Weight 可変構造 DNN の高速学習ハードウェア”, 情処全大 79 回，2017

表 1　各手法の識別率比較

データ	(a)minLast5		(b)Q4_3120		(c)sinPsinR		(d)Parking	
	Validation	Test	Validation	Test	Validation	Test	Validation	Test
(1)SVM		0.0780		0.0140		0.0060		0.1272
(2)BP(real)	0.0065	0.0180	0.0000	0.0000	0.0321	0.0380	0.0153	0.1152
(3)BP(binary)	0.1237	0.1270	0.0745	0.0700	0.0494	0.0460	0.0694	0.0762
(4)BP+GA(binary)	0.0479	0.0470	0.0000	0.0000	0.0403	0.0400	0.0277	0.0550
(5)BP+PSO(binary)	0.1005	0.0990	0.0595	0.0550	0.0479	0.0450	0.0310	0.0792
(6)GA(binary)	0.0403	0.0390	0.0000	0.0000	0.0463	0.0390	0.0143	0.1279
(●)BP+GA(binary)		0.0350		0.0000		0.0370		0.0172
データ	(e)South		(f)cp822x		(g)cp822y		(h)cp822z	
	Validation	Test	Validation	Test	Validation	Test	Validation	Test
(1)SVM		0.0266		0.0000		0.0800		0.0450
(2)BP(real)	0.0249	0.0412	0.0000	0.0000	0.0000	0.0600	0.0000	0.0340
(3)BP(binary)	0.0613	0.0779	0.0767	0.0800	0.1026	0.1090	0.0938	0.1400
(4)BP+GA(binary)	0.0487	0.0385	0.0484	0.0420	0.1005	0.1530	0.0938	0.1400
(5)BP+PSO(binary)	0.0579	0.0807	0.0563	0.0670	0.1026	0.1090	0.0832	0.1330
(6)GA(binary)	0.0546	0.0550	0.0525	0.0530	0.0955	0.1500	0.1056	0.1320
(●)BP+GA(binary)		0.0091		0.0000		0.0010		0.0140