7L-5

結合荷重のランダム変動によるニューラルネットワークの学習

久保　博隆　　三枝　亮　　橋本　周司

早稲田大学理工学部応用物理学科

１．はじめに

ニューラルネットワーク（以下ＮＮ）が入力に対して望む出力を得るためには、結合加重を調整する必要があり、その手法として誤差逆伝播法（ＢＰ法）[1][2]が多く用いられている。ＢＰ法は実用的には大変有用であるが、ＮＮの単純さに比べて学習アルゴリズムの計算コストは比較的高く、荷重の更新を同期的に行わなければならない。生命体の神経系ではヘブ学習[3]など、より単純なアルゴリズムが採用されていると思われる。我々は、最適化問題の解法の一つとして知られるランダムサーチのＮＮへの適用を検討している。ここではまず、ランダムに結合荷重を微小変動させ、その更新前後の誤差を比較して、変化の度合に応じて結合加重を更新する学習法について検討している。類似したアプローチとして、RasID[4]が挙げられるが、結合荷重を個々に更新している点や三値比較を行っている点で異なる。本学習法により、同期を取らずに荷重を更新することができ、ニューロン素子の出力が微分不可能なモデルにもこの学習法を適用することができる。

２.提案手法

図１に示されるように、ＮＮでは入力ベクトル $\boldsymbol{I}$、結合加重ベクトル $\boldsymbol{W}$ により、出力ベクトル $\boldsymbol{O}$ を得る。この $\boldsymbol{O}$ を、$\boldsymbol{I}$ に対する希望出力ベクトル $\boldsymbol{Y}$ に近づけることが学習である。近づける指針として、$\boldsymbol{O}$ と $\boldsymbol{Y}$ の距離の二乗誤差を誤差関数 e として用いることが多い。

$$e = \left|\boldsymbol{O}(\boldsymbol{I},\boldsymbol{W}) - \boldsymbol{Y}\right|^2 \quad (1)$$

N 個の（$\boldsymbol{I}_i$, $\boldsymbol{Y}_i$）$(1 \leqq i \leqq N)$ のセットに対し、各々の e_i の平均 E を最小化するために $\boldsymbol{W}$ を調整していくのが、ＮＮにおける学習である。

$$E = \frac{1}{N}\sum_{i=1}^{N} e_i = \frac{1}{N}\sum_{i=1}^{N}\left|\boldsymbol{O}_i(\boldsymbol{I}_i,\boldsymbol{W}) - \boldsymbol{Y}_i\right|^2 \quad (2)$$

検討した学習方法の概要を図２に示す。更新前の値を（t）、更新後の値を（$t+1$）で表すとし、$\boldsymbol{I}_i$ が入力されているとする。（希望出力：$\boldsymbol{Y}_i$）$\boldsymbol{W}(t)$ の任意の一成分 $W_j(t)$（$1 \leqq j \leqq M$, M は結合加重の本数）

Learning of Neural Networks
with Random Fluctuation of Weights
Hirotaka Kubo　kubo@shalab.phys.waseda.ac.jp
Ryo Saegusa　ryos@ieee.org
Shuji Hashimoto　shuji@waseda.jp
Department of Applied Physics,
School of Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku, Tokyo 169-8555 Japan

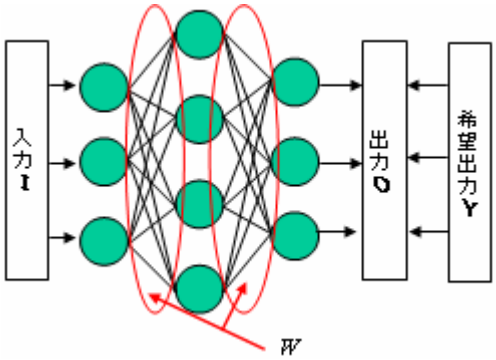

図１．学習対象のニューラルネットワーク

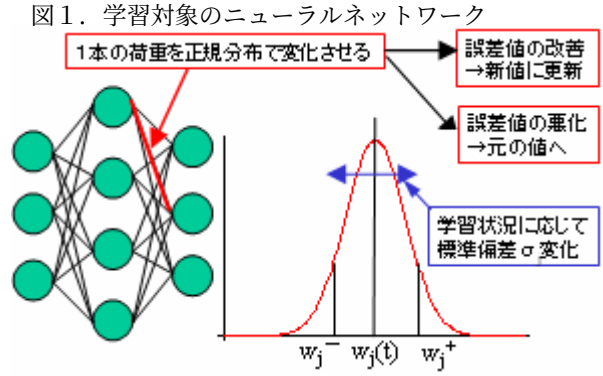

図２．学習方法概要

を、平均値０、標準偏差 $\sigma_j(t)$ の正規分布乱数で得た値 Δw を用い、W_j^+（$= W_j(t) + \Delta w$）に変化させる。$W_j(t)$ を W_j^+ に変化させたことにより与えられている（$\boldsymbol{I}_i$, $\boldsymbol{Y}_i$）のセットに対する誤差関数 $e_i(t)$ が e_i^+ に変化する。同様にして、$W_j(t)$ を W_j^-（$= W_j(t) - \Delta w$）に変化させた時の値 e_i^- を得られる。この $e_i(t)$、e_i^+、e_i^- の三値を比較し、以下に示す方法で更新値 $W_j(t+1)$ を得る。

(ｉ) 三値の最小値が $e_i(t)$ の場合
値を変化させたことにより、誤差値が悪化してしまっているので、値を更新前のものに戻す。

$$W_j(t+1) = W_j(t) \quad (3)$$

また、ローカルミニマムに陥っているとも考えられるので、探索域を広げるため標準偏差の最大値 σ_{max} を超えないように、(0,1]の定数 α を用いて $\sigma_j(t)$ を増加させる。

$$\sigma_j(t+1) = \begin{cases} \sigma_{max} & (if\ \ \sigma_j(t)/\alpha > \sigma_{max}) \\ \dfrac{\sigma_j(t)}{\alpha} & (otherwise) \end{cases} \quad (4)$$

(ii) 三値の最小値が e_i^+、または e_i^- の場合
値を変化させたことにより誤差値が改善されているので、

最小値の e_i^+（e_i^-）に対応する W_j^+（W_j^-）を新たな W_j とする。

$$W_j(t+1)=\begin{cases} W_j^+ & (if\ e_i^- > e_i^+) \\ W_j^- & (otherwise) \end{cases} \qquad (5)$$

また、最適解範囲を飛び越えて探索を行わないように、探索域を狭めるため、標準偏差の最小値 $\sigma_{\min}$ を下回らないように、式（4）と同一の定数 α を用い、$\sigma_j(t)$ を減少させる。

$$\sigma_j(t+1)=\begin{cases} \sigma_{\min} & (if\ \alpha\cdot\sigma_j(t)<\sigma_{\min}) \\ \alpha\cdot\sigma_j(t) & (otherwise) \end{cases} \qquad (6)$$

この更新則を希望入出力セット（$\boldsymbol{I}_i$, $\boldsymbol{Y}_i$）ごとに、全ての結合加重 j（1≦j≦M）に対して行う。結合加重の学習の順番が、学習に影響を与えないように学習順序はランダムで与える。また、同様の理由により、希望入出力セット（$\boldsymbol{I}_i$, $\boldsymbol{Y}_i$）の学習の順番もランダムで与える。

３．評価実験

本学習法を実際の問題に適用して有効性を確認した。3.1 では中間層に微分不可能なニューロンを含む場合の NN の学習、3.2 では Fisher のアヤメ分類問題[5]を学習対象とした。

3.1　中間層に微分不可能な素子を含む NN の学習

中間層の出力関数がステップ関数、出力層の出力関数がシグモイド関数の NN を準備し、２入力１出力の論理関数 XOR を学習させた。実験条件を表１、結果を表２に示す。ここで学習成功率とは、用意した初期結合加重 100 個のうち、学習終了サイクルまでに誤差関数 E が目標値以下になった割合を意味するものとする。表２から読み取れるように、BP 法では学習不可能な微分不可能なニューロンを含む NN を、本学習法では学習することができる。

表１．微分不能素子を含む NN の学習（XOR）の実験条件

入力層	2
中間層	40
出力層	1
出力関数（中間層）	$f(x)=\begin{cases}1 & (if\ x>0)\\ 0 & (otherwise)\end{cases}$
出力関数（出力層）	$1/\{1+\exp(-x)\}$
標準偏差初期値 σ	0.001
標準偏差最小値 σmin	0.0008
標準偏差最大値 σmax	0.0015
標準偏差更新係数 α	0.9
誤差関数目標値	0.01
学習終了サイクル	5000

表２．微分不能素子を含む NN の学習（XOR）の実験結果

	学習成功率（%）	平均学習終了サイクル
本学習法	85	321

3.2　Fisher のアヤメ分類問題

この問題は花から計測される４個のデータから、３種のアヤメのいずれに該当するかを分類する、よく知られた問題である。150 個あるアヤメのデータのうち、100 個を訓練用データ、50 個を判別用データとし、この 100 個・50 個の組を 10 種類、また、初期荷重 10 種類を準備し、10×10 の 100 個の初期条件の平均及び最大判別率を求めた。ここで判別率とは、学習終了後のＮＮが 50 個の判別用データを正しく分類出来た率である。尚、該当種類を示す方法として、該当種類の希望出力を 1.0、非該当種類の希望出力を 0.0 として学習を行ったが、実際には３出力中最大のものを NN の判別結果とした。実験条件を表３、平均及び最大判別率を表４に示す。比較として、表３にはＢＰ法の判別率も示す。本学習法によりＮＮが改善されていることが確認できる。

表３．アヤメ分類問題における NN の実験条件

入力層	4
中間層	4
出力層	3
出力関数	$1/\{1+\exp(-x)\}$
標準偏差初期値 σ	0.001
標準偏差最小値 σmin	0.0008
標準偏差最大値 σmax	0.0012
標準偏差更新係数 α	0.9
学習終了サイクル	100

表４．アヤメ分類問題における NN の判別率

	平均判別率（%）	最大判別率（%）
ＮＮ未学習	34	72
本学習法	48	96
ＢＰ法	47	100

４．まとめ

ランダムサーチを利用したニューラルネットワークの学習方法を提案し、ＢＰ法では学習不能である微分不能素子を含むＮＮに適用が可能であることを示した。また、Fisher のアヤメ分類問題にてその性能を確認した。今後の課題としては、パラメータの決定法の検討と、他問題への適用が挙げられる。

参考文献
[1] D. Rumelhart and J. McClelland 1986,
"Learning internal representation by error propagation,"
Parallel Distributed Processing 1, 318-362, MIT Press.
[2] 長尾智晴，最適化アルゴリズム,昭晃堂（2000）
[3] Hebb.D.O.
“The Organization of Behavior”
Willy, New York, 1949
[4]平澤宏太郎，東郷和幸，胡敬炉，大林正直，邵寧，村田純一
”ニューラルネットワークの適応的ランダム探索最適化手法
-RasID-“, 計測自動制御学会論文集, 34, 8, 1088-1096 (1998)
[5] Fisher, R. A.
“The use of multiple measurements in taxonomic problems.”
Annals of Eugenics 7: 179-188, 1936.