

ニューラルネットワークにおける誤差逆伝播法の学習性能向上のための重み初期値の設定方法

下 平 丕 作 士†

誤差逆伝播法は、多層前向きニューラルネットワークの学習に広く用いられている。しかし、誤差逆伝播法については、解が収束するのに要する時間が長く、解が収束しない場合があるという学習性能上の問題点がある。誤差逆伝播法の学習性能は重みの初期値に敏感であり、適切な重みの初期値を設定することにより、学習性能を向上させることができる。本論文では、ノードにおける情報伝達機構を表す式に基づいて、適切な大きさの重みの初期値を設定する新しい方法（OIVS法：Optimal Initial Value Setting Method）を提案している。数値実験によると、OIVS法は従来法にくらべて、解が安定して得られ、収束するのに要する反復回数が少なく、学習性能が著しく優れていることが分かった。たとえば、ランダム写像問題については、それぞれの方法で最も良い学習性能が得られた場合を比較すると、OIVS法は従来法に比べて、解が収束するのに要した反復回数の平均値は0.26倍、標準偏差は0.039倍となっている。OIVS法の特徴は、問題（入力次元数）に応じて適切な大きさの重みの初期値を設定できること、アルゴリズムがきわめて簡単であること、ノードごとに局所的な計算で設定できることである。

A Weight Value Initialization Method for Improving Learning Performance of the Back Propagation Algorithm in Neural Networks

HISASHI SHIMODAIRA†

The back propagation algorithm has been widely used for the learning in multilayer feedforward neural networks. However the performance of the algorithm has the problem that it takes long time to converge to the solution and sometimes it does not converge to the solution. The learning performance of the algorithm is sensitive to the initial values of the weights and can be improved by setting appropriate initial values to the weights. In this paper, we have proposed a new method (called OIVS method) for setting initial values to the weights using the equations which represent the information transformation mechanism at the node. The numerical experiments have shown that the learning performance of the OIVS method is remarkably better than that of the conventional method. The OIVS method has the features that (a) it can set appropriate initial values according to the input dimensions, (b) the algorithm is simple, and (c) the algorithm fits the local processing of neural networks.

1. ま え が き

多層前向きニューラルネットワーク（以下、ニューラルネットワークと呼ぶ）の学習において、誤差逆伝播法¹⁾が広く用いられている。しかし、誤差逆伝播法については、解が収束するまでの時間が長く、解が収

束しない場合があるという学習性能上の問題点がある。これらの問題点の解決方法として、アルゴリズムそのものを改良する。重みの初期値の設定方法を工夫するという二つの方法がある。本論文では、Rumelhartらの提案したもともとの誤差逆伝播法に関して、重みの初期値の設定方法について検討する。

誤差逆伝播法の学習性能は重みの初期値にきわめて敏感であることが知られており、性能向上のための初期値の設定方法の研究が行われてきた。従来の最も簡

† 日本メックス株式会社研究開発部
Research and Development Department, Nihon
MECCS Co. Ltd.

便な方法として、 δ を正の数として $(-\delta, \delta)$ の範囲の乱数を設定する方法が用いられてきたが、誤差曲面についての考察に基づいて、小さな範囲の値を設定することが推奨されている²⁾。これに対して Nguyen ら³⁾、賈ら⁴⁾、Chen ら⁵⁾、Kim ら⁶⁾、Kim ら⁷⁾などが、学習性能を向上させる重みの設定方法を提案している。

重みの初期値設定法については、学習時間ができるだけ短いこと、解が安定していることのほか、できるだけアルゴリズムが簡単であること、ニューラルネットワークの処理の局所性に適合していることなどが望まれる。この観点から既往の研究を評価すると、それぞれ一長一短があるが、ノードにおける情報伝達機構を表す関係式を用いた賈らの方法は注目値する。

本論文では、賈らが導いたノードにおける情報伝達機構を表す関係式を用いて、適切な大きさの重みの初期値を設定する、学習性能が優れた新しい方法を提案する。適切な大きさの重みの初期値を設定できるという意味で、提案方法を OIVS 法 (Optimal Initial Value Setting Method) と呼ぶこととする。

2章では、ノードにおける入出力の関係式と情報伝達機構を表す関係式について概説し、3章では、重みの初期値設定における問題点について述べる。4章では、ノードにおける情報伝達機構を表す関係式を用いて重みの初期値を設定する新しい方法を提案し、5章では、数値実験によりその有効性を検証する。6章では、既往の研究について概説し、提案方法の特徴について述べる。7章では、まとめと今後の課題について述べる。

2. ノードにおける情報伝達を表す関係式

ニューラルネットワークにおける重みの役割は、入力情報を前向きに、誤差情報を逆向きに、重みに比例した大きさで結合するノードに伝達することである。各ノードは、受け取った入力情報を局所的に変換処理して出力し、結合する重みを介して他のノードに伝達する。したがって、重みの初期値としては、ノードにおける情報変換処理を円滑ならしめるような、大きさと符号を持つ値を設定する必要がある。そこで、本章では、各ノードにおける入出力の関係式と情報伝達機構を表す関係式について述べる。

ノードにおける入力の総和 net は、次式で表される。

$$net = \sum_{i=1}^n w_i x_i + w_0 \quad (1)$$

ここで、 x_i はノードへの入力、 w_i は入力の結合重み、 w_0 はしきい値、 n は入力の結合重みの数である。ノードの活性化関数としてシグモイド関数を用いることとすれば、出力 y は次式で表される。

$$y = f(net) = \frac{1}{1 + \exp(-net)} \quad (2)$$

ここで、 y は 0 から 1 の範囲の値をとる。層状に結合されたニューラルネットワークでは、上層のノードの出力が下層のノードの入力となる。また、入力層への入力は、0 から 1 の範囲の値に正規化される。したがって、隠れ層および出力層における入力空間は、 n 次元ユークリッド空間であり、その入力値の組は必ず入力空間内の単位超立方体 $(0, 1)^n$ の内部に存在することとなる。このような単位超立方体の対角線の長さは、次式で表される⁸⁾。

$$D = \sqrt{n} \quad (3)$$

式(2)において、 y の値として α が与えられたとすれば、

$$y = f(net) = \alpha \quad (4)$$

であり、式(1)と(2)から次式が得られる。

$$\sum_{i=1}^n w_i x_i + w_0 - f^{-1}(\alpha) = 0 \quad (5)$$

上式は重みによって決定される、入力空間における超平面 ($L_P(\alpha)$ と表す) であり、その上に存在する点に対応する出力は、すべて α となる。

式(2)のシグモイド関数は、 $net=0$ の近傍で線形に近く、 net が大きくなるにつれて 1 に、また小さくなるにつれて 0 に漸近する。 $net=0$ の近傍の範囲を活性領域、出力が 1 または 0 に漸近する範囲を飽和領域と呼ぶこととする。 ε を小さな正の数 (本論文では、0.1 とする) とすれば、これらの領域は次のように区分される。

$$\varepsilon \leq f(net) \leq 1 - \varepsilon \quad (6)$$

図 1 に、 $n=2$ の場合の入力空間を示す。この場合、式(5)で表される超平面は直線であり、図に示すように、 $y=1-\varepsilon$ 、0.5、 ε に対応して直線 $L_P(1-\varepsilon)$ 、 $L_P(0.5)$ 、 $L_P(\varepsilon)$ が得られる。

入力空間における活性領域に対応する幅 d (以後、活性領域幅と呼ぶ) は、超平面 $L_P(1-\varepsilon)$ と $L_P(\varepsilon)$ の間の距離であり、幾何計算により次式のように表される⁴⁾。

$$d = \frac{|f^{-1}(1-\varepsilon) - f^{-1}(\varepsilon)|}{\left(\sum_{i=1}^n w_i^2\right)^{1/2}} \quad (7)$$

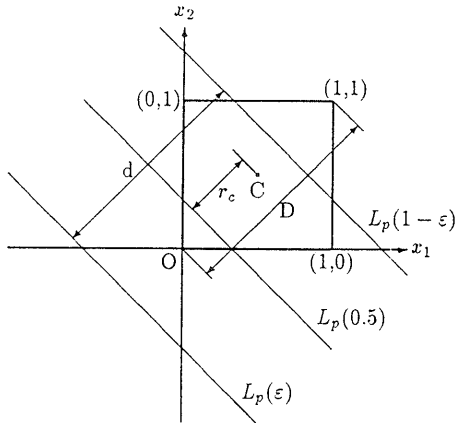


図1 2入力の場合の入力空間

Fig. 1 Input space in the case of 2 inputs.

ここで、分母は w_i を軸とする重み空間における重みベクトルの長さであることに注目されたい。また、単位超立方体の中心 C と活性領域の中央の超平面 $L_p(0.5)$ との距離 r_c は、次式で表される⁴⁾。

$$r_c = \frac{|w_0 + 0.5 \sum_{i=1}^n w_i|}{\left(\sum_{i=1}^n w_i^2\right)^{1/2}} \quad (8)$$

式(7)と(8)は、ノードにおける入力と出力の間の情報伝達機構を表す重要な式である。これらの式を用いて、ノードにおける情報変換を円滑ならしめる大きさと符号を持つ重みの初期値を設定することが可能となる。

3. 重みの初期値設定における問題点

本章では、2章で述べたノードの情報伝達機構を表す式に基づいて、筆者の経験も含めて既往の研究で指摘されている、重みの初期値設定における問題点について考察する。従来から用いられている最も簡単な重みの初期値設定法では、 δ を正の数として、 $(-\delta, \delta)$ の範囲の一律乱数を発生させて初期値としている。この方法では、次の場合に、解が収束しないか、または、解の収束が遅くなる。

- (1) 重みベクトルの長さが大きすぎる場合
- (2) 重みベクトルの長さが小さすぎる場合
- (3) 重みベクトルの長さは適切であるが、いくつかの重みが小さすぎる場合

(1) の場合は、式(7)から分かるように、活性領域幅 d が単位超立方体の対角線の長さ D より小さい場合である。活性領域を越えた入力 net があると、その

ノードの出力は1または0に近い値となり、シグモイド関数の飽和領域に存在することとなる。飽和領域におけるシグモイド関数の勾配は小さく、誤差曲面においては、この部分に対応して平らな台地状の領域が存在する²⁾。このような領域に入力が落ちた場合、曲面の勾配がきわめて小さいため、誤差逆伝播法における重みの修正項も小さくなり、解の収束が遅いか、または、収束しないこととなる^{4),9)}。

(2) の場合は、式(7)から分かるように、 d が D より大きい場合である。その程度が著しい場合には、 net が極めて小さな値となるため、式(2)から明らかなように、そのノードの出力はどのような入力に対しても0.5に近い値となる。このため、情報の変換機能が正常に作用せず、解が収束しないこととなる⁴⁾。

(3) の場合は、 d は D よりも大きく、活性領域が単位超立方体を覆っているが、いくつかの重みが非常に小さい場合である。前向きの入力情報の伝達は重みの大きさに比例し、また、誤差情報の逆伝播も重みの大きさに比例する。したがって、重みの初期値が小さ過ぎると、このような情報の伝達ができないため、その重みの成長が遅く、解の収束が遅いか、収束しないこととなる。このような事態は、ニューラルネットワークの規模が小さく、それぞれのノードと重みが重要な役割を持っている場合に生ずることが多い。従来法では、 $(-\delta, \delta)$ の範囲の乱数を発生させて、重みの初期値としているため、ゼロに近い小さな値も混入することとなる。

以上の問題点を解決するためには、全体として適切な大きさで（いいかえれば、ベクトルとして適切な長さを持ち）、個々の値も適切な大きさの重みを初期値として設定する必要がある。

4. 新しい重みの初期値設定方法の提案

本章では、3章で述べた問題点を解決する、新しい重みの初期値設定方法を提案する。(1)と(2)の問題点を解決するためには、入力空間における活性領域が適切な幅を持ち、かつ入力単位超立方体を覆うようにすればよい。(3)の問題点を解決するには、個々の重みが適切な大きさの値になるように制御すればよい。

具体的には、図2に示すように、活性領域幅 d を単位超立方体の対角線の長さ D より大きくとり、活性領域の中央の超平面 $L_p(0.5)$ が単位超立方体の中心 C を通るようにする⁴⁾。 d を D の k 倍にとるものとすれば、

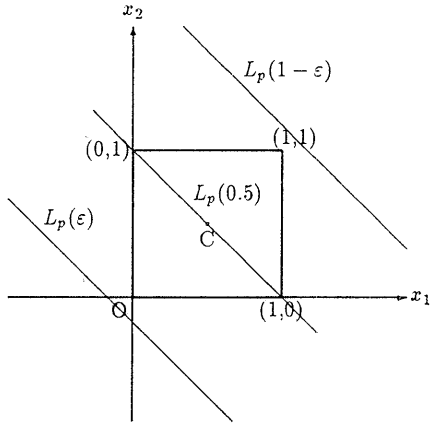


図2 単位超立方体と活性領域の望ましい位置関係
Fig. 2 Desirable positional relation between unit hypercube and activation region.

$$d = kD \tag{9}$$

ここで、 k は活性領域幅を制御し、ひいては、重みベクトルの長さを制御するパラメータである (以後、活性領域幅係数と呼ぶ)。その望ましい値は、実験的に定める。また、 $L_p(0.5)$ がC点を通るためには、式(8)から分かるように、式(9)を満足する $w_i (i=1, \dots, n)$ を求めた後、しきい値 w_0 を次式によって定めればよい⁴⁾。

$$w_0 = -0.5 \sum_{i=1}^n w_i \tag{10}$$

問題に対応して、 n と k が与えられたとき、式(9)を満足する重み $w_i (i=1, \dots, n)$ を次のようにして求める。ある範囲内で一様乱数を発生させて w_i^2 に対応させるものとすれば、その平均値 \bar{w}^2 は次式で表される。

$$\bar{w}^2 = \frac{1}{n} \sum_{i=1}^n w_i^2 \tag{11}$$

上式から、式(7)、(9)および(3)を用いて、次式が得られる。 \bar{w} は、正の値をとるものとする。

$$\bar{w} = \frac{b}{kn} \tag{12}$$

ここで、

$$b = |f^{-1}(1-\epsilon) - f^{-1}(\epsilon)| \tag{13}$$

w_i^2 の最小値と最大値を、それぞれ w_{min}^2 と w_{max}^2 とすれば、

$$w_{min}^2 \leq w_i^2 \leq w_{max}^2 \tag{14}$$

w_i^2 の最小値と最大値を、その平均値 \bar{w}^2 と係数 γ によって次のように表す。

$$w_{min}^2 = \bar{w}^2(1-\gamma) \tag{15}$$

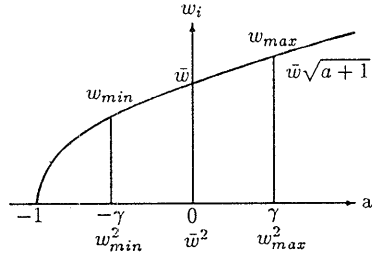


図3 式(14)、(17)および(19)の意味
Fig. 3 Meaning of Eqs. (14), (17) and (19).

$$w_{max}^2 = \bar{w}^2(1+\gamma) \tag{16}$$

ここで、 γ は重みの分布する範囲を制御する係数である (以後、重み分布幅係数と呼ぶ)。その望ましい値は、実験的に定める。式(14)、(15)および(16)から、次式が得られる。

$$-\gamma \leq a_i \leq \gamma \tag{17}$$

ここで、

$$a_i = (w_i^2 / \bar{w}^2) - 1 \tag{18}$$

重み w_i は正の値をとるものとすれば、上式から次式が得られる。

$$w_i = \bar{w} \sqrt{a_i + 1} \tag{19}$$

以上を要約すれば、 n, k, γ が与えられたとき、式(12)により $\bar{w} (>0)$ を求め、式(17)を満足する範囲で一様乱数 a_i を発生させ、これを用いて式(19)により w_i を求める。式(14)、(17)、(19)の意味を図3に示す。

このようにして求めた重みの2乗 w_i^2 の平均値は、式(11)によりあらかじめ定めた \bar{w}^2 とは、厳密には一致しない。しかし、 w_i^2 が式(15)と(16)を満足する範囲内で一様乱数によって定められるため、 n がある程度大きくなれば、その平均値はあらかじめ定めた \bar{w}^2 と近似的に一致し、式(11)は近似的に満足されることとなる。

以上の事項を、あるノードに結合する n 個の入力重み w_i を求めるアルゴリズムとして整理すれば、次のようになる。

- (1) 与えられた n と k によって、式(12)により \bar{w} を計算する。 \bar{w} は正の数とする。
- (2) 与えられた γ によって、式(17)の範囲内で一様乱数 a_i を発生させる。
- (3) a_i から、式(19)により、 w_i を求める。 w_i は正の数とする。(2)と(3)を n 回繰り返す。
- (4) 式(10)により、しきい値 w_0 を求める。

活性領域幅係数 (k) と重み分布幅係数 (γ) の望ましい値は、学習率 η 等と同様に、数値実験により定め

る。具体的には、これらの値をある幅で変えながら学習を行い、解の収束に要する反復回数が最も少なくなる値を探る作業を繰り返す、試行錯誤により定める。

上記の方法の特徴は、問題（入力次元数 n ）に応じて、ノードの情報変換機能が最適になるように重み全体の大きさを制御でき、また、個々の重みの分布も適切な範囲になるように制御できることである。この意味で、提案方法を OIVS 法 (Optimal Initial Value Setting Method) と呼ぶこととしている。

5. 数値実験

本章では、XOR 問題とランダム写像問題¹⁰⁾について、従来法と OIVS 法により重みの初期値を設定し、数値実験を行った結果を述べる。誤差逆伝播法における重みの修正は、すべての入力パターンと目標パターンについて重みの修正量を加算して、一括して修正する方法を用いた。誤差逆伝播法で使用するパラメータ（学習率 η ）と、初期値設定において使用する係数（従来法では重みの分布幅 δ 、OIVS 法では重み分布幅係数 γ 、活性領域幅係数 k ）を変化させて、様々な組合せでシミュレーションを行い、最も良い学習性能が得られる範囲を調べた。誤差逆伝播法における慣性項の係数 α は、0.9 とした。学習性能の評価尺度としては、50 組の重み初期値についてシミュレーションを行い、解が収束するのに要した反復回数（重みの修正回数）の平均値 (m) と標準偏差 (σ)、および解が収束しなかった場合の数を用いた。解の収束判定尺度としては、(1パターン当たり、出力層の1ノード当たりの出力誤差の2乗和の平均値)=0.01 を用いた。1000 回反復してもこれをクリアしなかった場合は、収束しなかったものとした。なお、使用した言語は Microsoft C であり、乱数は RAND を用いて生成した。乱数のシードは、各シミュレーションごとにシステム時刻に関連させて、ランダムな値を設定した。

(1) XOR 問題 ネットワーク構造を、入力層 2ノード、隠れ層 2ノード、出力層 1ノードとし、入力パターンと出力パターンとして、XOR 問題の入出力関係 4パターンを与えた。予備の数値実験によると、学習率 η は 0.7 前後の値で最も良い結果が得られることが分かったので、すべての場合について $\eta=0.7$ とした。図 4 に、OIVS 法で $k=4.0$ とし、 γ を変化させた場合の学習性能の変化を示す。図 5 に、OIVS 法で $\gamma=0.9$ とし、 k を変化させた場合の学習性能の変化を示す。図 6 に、従来法で、 δ を変化させ

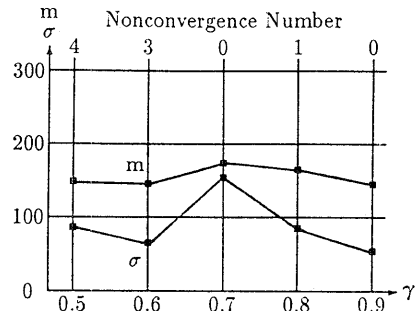


図 4 OIVS 法 ($k=4.0$) による結果: XOR 問題における重み分布幅係数 (γ) と反復回数の関係

Fig. 4 Results by OIVS method ($k=4.0$): Relation between weight distrib. width coeff. (γ) and iteration number in XOR problem.

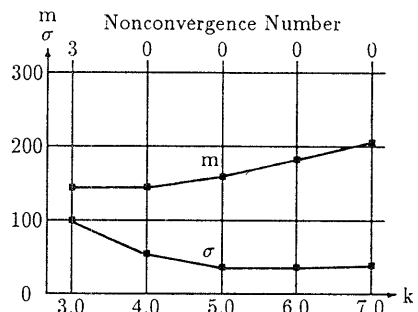


図 5 OIVS 法 ($\gamma=0.9$) による結果: XOR 問題における活性領域幅係数 (k) と反復回数の関係

Fig. 5 Results by OIVS methods ($\gamma=0.9$): Relation between activ. region width coeff. (k) and iteration number in XOR problem.

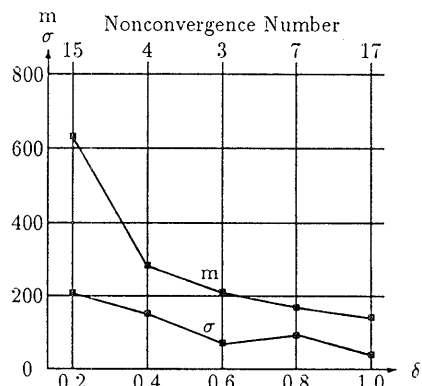


図 6 従来法による結果: XOR 問題における重みの分布幅 (δ) と反復回数の関係

Fig. 6 Results by conventional method: Relation between weight distrib. width coeff. (δ) and iteration number in XOR problem.

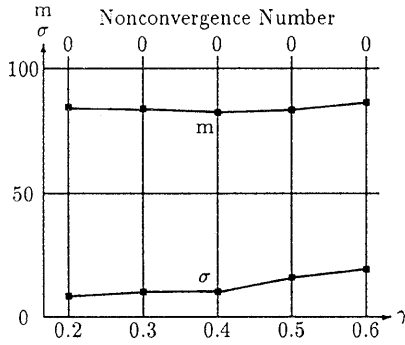


図 7 OIVS 法 ($k=2.5$) による結果: ランダム写像問題における重み分布幅係数 (γ) と反復回数との関係
 Fig. 7 Results by OIVS method ($k=2.5$): Relation between weight distrib. width coeff. (γ) and iteration number in random mapping problem.

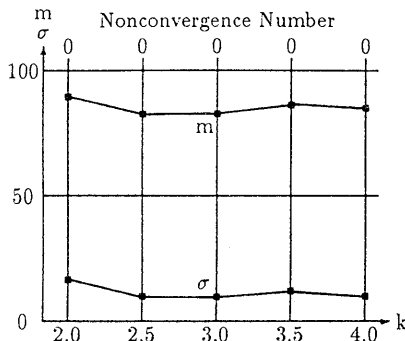


図 8 OIVS 法 ($\gamma=0.4$) による結果: ランダム写像問題における活性領域幅係数 (k) と反復回数との関係
 Fig. 8 Results by OIVS method ($\gamma=0.4$): Relation between activ. region width coeff. (k) and iteration number in random mapping problem.

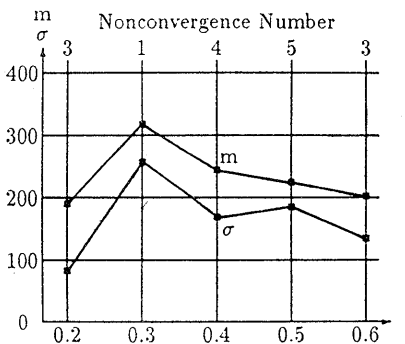


図 9 従来法による結果: ランダム写像問題における重み分布幅 (δ) と反復回数との関係
 Fig. 9 Results by conventional method: Relation between weight distrib. width (δ) and iteration number in random mapping problem.

た場合の学習性能の変化を示す。

OIVS 法では、非収束の場合が少ない。 γ が大きい方が学習性能が良く、 k については 4.0 ないし 7.0 で良い結果が得られている。最も良い結果は、 $\gamma=0.9$, $k=4.0$ の場合である。従来法では、 δ が大きくなるにつれて m と σ は小さくなるが、非収束の場合が増加している。最も良い結果は、非収束の場合が最も少ない $\delta=0.6$ である。それぞれの方法で最も良い学習性能が得られた場合を比較すると、OIVS 法は従来法に比べて、 m は 0.69 倍、 σ は 0.76 倍となっている。

(2) ランダム写像問題¹⁰⁾ (0,1) の範囲で発生させた一様乱数を用いて、入力パターンと出力パターンとして 6 組のデータを作成して、学習させた。ネットワークの構造を、入力層 10 ノード、隠れ層 5 ノード、出力層 10 ノードとした。予備の数値実験によると、学習率 η は 0.7 前後の値で最も良い結果が得られることが分かったので、すべての場合について $\eta=0.7$ とした。入力層と隠れ層で n の値が異なるが、 k は共通の値を用いた。 γ の値も共通の値を用いた。入力層と隠れ層についての平均的な n は、6.7 である。図 7 に、OIVS 法で $k=2.5$ とし、 γ を変化させた場合の学習性能の変化を示す。図 8 に、OIVS 法で $\gamma=0.4$ とし、 k を変化させた場合の学習性能の変化を示す。図 9 に、従来法で、 δ を変化させた場合の学習性能の変化を示す。

OIVS 法では、図に示した範囲で収束しない場合はゼロであり、 m と σ には大きな変化はなく、安定した性能を示している。また、係数の値がここで示した範囲を外れても、学習性能はそれほど大きく劣化していない。最も良い結果は、 $\gamma=0.4$, $k=2.5$ の場合である。従来法では、収束しない場合がかなりあり、 m と σ の変化が激しい。最も良い結果は、非収束の場合が最も少ない $\delta=0.3$ であると考えられる。それぞれの方法で最も良い学習性能が得られた場合を比較すると、OIVS 法は従来法に比べて、 m は 0.26 倍、 σ は 0.039 倍となっている。

(3) 考察 XOR 問題では、従来法に対する OIVS 法の改善効果は、それほど大きくはない。これは、 n が小さいため、式 (14) により発生させた $w_i^?$ が、式 (11) を満足する度合いが小さい場合がかなりあるためである。ランダム写像問題では、OIVS 法の学習性能は従来法に比べて著しく優れており、OIVS 法の有用性が裏付けられている。

OIVS 法における γ と k の望ましい範囲をみると、

γ については, XOR 問題では [0.7, 0.9], ランダム写像問題では [0.2, 0.6] である. k については, XOR 問題では [4.0, 7.0], ランダム写像問題では [2.0, 4.0] である. この結果から, 入力の次元数 n により, 重み初期値のベクトルの長さおよび分布幅の最適値が異なることが分かる.

6. 既往の研究の概要と OIVS 法の特徴

Nguyen ら³⁾は, 隠れ層の各ノードが近似しようとする関数のある部分を分担するという仮定のもとに, 隠れ層のノード数と関連付けて重みの大きさを決める方法を提案している. 賈ら⁴⁾は, δ を正の数として, $(-\delta, \delta)$ の範囲で発生させた一様乱数により仮の w_i を求め, 式(7)を用いて式(9)の k の値が与えられた値になるように, 適当な正の数を乗じて w_i を決める方法を提案している. しきい値は, 式(10)により求めている. Chen ら⁵⁾は, 隠れ層と出力層の間の重みからなるマトリックスのランクにより場合分けをし, 代数的に重みの最適値を求める方法を提案しているが, そのアルゴリズムはきわめて複雑である. Kim ら⁶⁾は, 重みの大きさの下限が学習率によって決まることを示唆している. そして, $(-\delta, \delta)$ の範囲で一様乱数により仮に求めた重みを正の方向にずらして重みを定め, 良好な結果を得ているが, ずらす量をどのように決めるのかについては明確に示されていない. Kim ら⁷⁾は, 入力層と隠れ層の間の重みを, 入力パターンに対応した値になるように線形計画法により定める方法を提案しているが, そのアルゴリズムはかなり複雑である. これらの研究の中で, Nguyen ら³⁾と Kim ら⁶⁾は, 重みをすべて正の値としていることが注目される. ただし, Nguyen らは, しきい値は従来法と同様にして定めている.

本論文で提案した OIVS 法は, 賈ら⁴⁾の研究で示された式(7)と(8)を用いて, 重みベクトルの長さが単位超立方体の対角線の長さとの関係で適切な値になるようにし, 活性領域の中央の超平面が単位超立方体の中心を通るようにしきい値を決めるというアイデアに基づいている. しかし, 賈らの方法は, 基本的には重みベクトルの長さのみを制御の対象とするのに対し, OIVS 法は, 重みベクトルの長さのみならず, 個々の重みの大きさと分布範囲をも制御するという考えに基づいており, 重み決定のアルゴリズムは全く異なっている. これにより, 5章の数値実験の結果からも明らかのように, 入力次元数 n に応じた, より適切な

重み初期値の設定が可能になる. また, 賈らの方法では, 重みを $(-\delta, \delta)$ の範囲の一様乱数により定めているため, ゼロに近い小さな値が混入することが避けられない. このような重みを用いて式(10)によりしきい値を求めているため, しきい値がゼロに近い小さな値になることが避けられない. OIVS 法では, 重みはすべて正の値としているため, このような事態は起こらず, 3章で述べた解の収束を妨げる場合(3)は起こらない.

OIVS 法と既往の研究で提案された方法の学習性能の直接的な比較は行っていないが, 従来法を介した間接的な比較によれば, これらと同程度もしくはそれ以上の学習性能が得られている. 賈らの方法については, 同一の問題 (XOR 問題およびランダム写像問題) について, 文献 4) に示されている収束条件で, OIVS 法により数値実験を行い, その結果を文献 4) に示された結果と比較した. 収束条件は, XOR 問題についてはパターン平均 2 乗誤差=0.005 であり, ランダム写像問題についてはパターン平均 2 乗誤差=0.01 である. XOR 問題については, 50 回のシミュレーションのうち 1000 回反復しても収束条件を満足しなかった回数は, 賈らの方法 ($\eta=1.2, \alpha=0.9$) では 5 回 (文献 4) の図 6 の (b)) であるのに対し, OIVS 法 ($\eta=0.7, \alpha=0.9$) では, 図 4 と図 5 に示したように適切な k と γ の値を用いれば, ゼロであった. また, ランダム写像問題 (入力層のノード数: 5, 隠れ層のノード数: 7, 出力層のノード数: 5, 学習パターン数: 9) については, 50 回のシミュレーションのうち収束しなかった回数は両者ともにゼロであり, 収束に要した反復回数数の平均値は, 賈らの方法 ($\eta=1.2, \alpha=0.9$) では 96.4 回 (文献 4) の図 7 の (b)) であるのに対し, OIVS 法 ($\eta=1.2, \alpha=0.9, k=1.0, \gamma=0.7$) では 91.7 回であった. この結果から, OIVS 法は賈らの方法に比べて, 解の収束に要する反復回数はそれほど減少しないが, 解が収束しない場合を避けられるという利点があることが分かる.

OIVS 法の特徴は, 問題 (入力次元数) に応じて適切な大きさの重みの初期値を設定できること, アルゴリズムがきわめて簡単であること, ノードごとに局所的な計算で設定できることである.

7. むすび

本論文では, ノードにおける情報伝達機構を表す式に基づいて, 適切な大きさの重みの初期値を設定する

新しい方法 (OIVS 法) を提案した。数値実験によると, OIVS 法は従来法に比べて, 著しく学習性能が優れていることが分かった。OIVS 法の特徴は, 問題 (入力次元数) に応じて適切な大きさの重みの初期値を設定できること, アルゴリズムがきわめて簡単であること, ノードごとに局所的な計算で設定できることである。

今後の研究課題としては, 入力次元数に応じた活性領域幅係数と重み分布幅係数の望ましい値の範囲を調べて, OIVS 法を実用に供する際の参考資料を作成することである。

参 考 文 献

- 1) Rumelhart, D. E. et al.: *Parallel Distributed Processing*, Vol. 1, Chapt. 8, MIT Press (1986).
- 2) Hush, D. R. et al.: *Error Surfaces for Multilayer Perceptrons*, *Proc. IJCNN (Seattle)*, Vol. I, pp. 759-764 (1991).
- 3) Nguyen, D. et al.: *Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights*, *Proc. IJCNN*, Vol. III, pp. 21-26 (1990).
- 4) 賈棋ほか: ニューラルネットワークにおける逆伝搬アルゴリズムの初期値設定に関する考察, *電子情報通信学会論文誌 (D-II)*, Vol. J73-D-II, No. 8, pp. 1179-1185 (1990).
- 5) Chen, C. et al.: *Improving the Training Speed of Three-layer Feedforward Neural Nets by Optimal Estimation of the Initial Weights*, *Proc. IJCNN (Singapore)*, pp. 2063-2068 (1991).
- 6) Kim, Y. K. et al.: *Weight Value Initialization for Improving Training Speed in the Back-*

propagation Networks, *Proc. IJCNN (Singapore)*, pp. 2396-2401 (1991).

- 7) Kim, L. S.: *Initializing Weights to a Hidden Layer of a Multilayer Neural Network by Linear Programming*, *Proc. IJCNN*, pp. 1701-1704 (1993).
- 8) Hect-Nielsen, R.: *Neurocomputing*, p. 42, Addison-Wesley (1989).
- 9) Lee, Y. et al.: *The Effect of Initial Weights on Premature Saturation in Back-Propagation Learning*, *Proc. IJCNN (Seattle)*, Vol. I, pp. 765-770 (1991).
- 10) Kung, S. Y.: *An Algebraic Projection Analysis for Optimal Hidden Units Size and Learning Rates in Back-Propagation Learning*, *Proc. IJCNN*, Vol. I, pp. 363-370 (1988).

(平成 6 年 1 月 20 日受付)

(平成 6 年 6 月 20 日採録)



下平丕作士 (正会員)

昭和 44 年東京都立大学工学部建築工学科卒業。昭和 46 年同大学修士課程修了。同年日本電信電話公社 (現株式会社) 入社。武蔵野電気通信研究所および建築部建築技術開発室において, 構造工学, 数値解析, CAD, データベース, AI 等の研究とシステム開発に従事。平成 4 年から日本メックス (株) において, 建物・設備管理へのコンピュータの応用に関する研究・開発に従事。データ工学, データベース, 知識工学, エキスパートシステム, ニューラルネットワーク, GA, 図形処理, CG 等に興味を持っている。工学博士。電子情報通信学会, 人工知能学会, 日本建築学会各会員。