1P－1

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
日本メックス株式会社　研究開発部

1．まえがき

多層前向きニューラルネットワークの学習に用いられている誤差逆伝播法は，学習時間と解の安定性という学習性能についての問題点がある．学習性能は，適切な重みの初期値を設定することにより向上させることができる．本報告では，買ら[1]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を設定する新しい方法を提案する．

2．ノードにおける情報伝達を表す関係式

ノードにおける入力x_iと出力yの関係は，図１の式(1)と(2)で表される．nは入力の次元数，w_iは重み，w_0はしきい値である．x_iの値は[0, 1]であるため，入力値の組は入力空間における単位超立方体の内部にある．単位超立方体の対角線の長さは，式(3)で与えられる．出力αが式(4)のように与えられたとき，これを満足する入力値の組は入力空間における超平面であり，式(5)のように表される．正の微小な数εによってシグモイド関数を，式(6)のように活性領域と飽和領域に区分する．入力空間における対応する超平面を$L_p(1-\varepsilon)$，$L_p(0.5)$，$L_p(\varepsilon)$で表す（図２参照）．$L_p(1-\varepsilon)$と$L_p(\varepsilon)$の間の距離dは式(7)で，単位超立方体の中心Cと$L_p(0.5)$の間の距離は式(8)で表される[1]．

3．新しい重み初期値の設定方法（ＯＩＶＳ法）

提案する方法では，重みベクトルの長さを制御し，式(9)のように，活性領域の幅dが単位超立方体の対角線の長さのk（活性領域幅係数と呼ぶ）倍になるような重みを生成する．さらに，そのような重みから決まる$L_p(0.5)$が単位超立方体の中心を通るように，しきい値を式(10)によって求める．

上記のような重みを生成する方法は，次の通りである．重みの２乗の平均を$\bar{w}^2$とすれば，式(11)が成立する．式(11)と式(7)，(9)，(3)から，式(12)，(13)が

$$net = \sum_{i=1}^{n} w_i x_i + w_0 \tag{1}$$

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

$$D = \sqrt{n} \tag{3}$$

$$y = f(net) = \alpha \tag{4}$$

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

$$\varepsilon \le f(net) \le 1-\varepsilon \tag{6}$$

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

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

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

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

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

$$\bar{w}^2 = \frac{b^2}{k^2 n^2} \tag{12}$$

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

$$w_{i,min}^2 \le w_i^2 \le w_{i,max}^2 \tag{14}$$

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

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

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

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

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

図１　本文中の式

得られる．重みの２乗の最小値と最大値を，$\bar{w}^2$とγ（重み分布幅係数と呼ぶ）によって式(15)と(16)のように表せば，式(14)が成立する．これらの式から，式(17)と(18)が得られる．式(18)から式(19)が得られる．

アルゴリズムとしてまとめると，次の通りである．

(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．$\bar{w}$は正とする．

(2) 与えられたγによって，式(17)の範囲内で一様乱

A Weight Value Initialization Method for Improving Learning Performance of the Back Propagation Algorithm in Neural Networks
Hisashi Shimodaira
Nihon MECCS Com. Ltd.
1-22-10 Nishishinbashi, Minato-Ku, Tokyo, 105

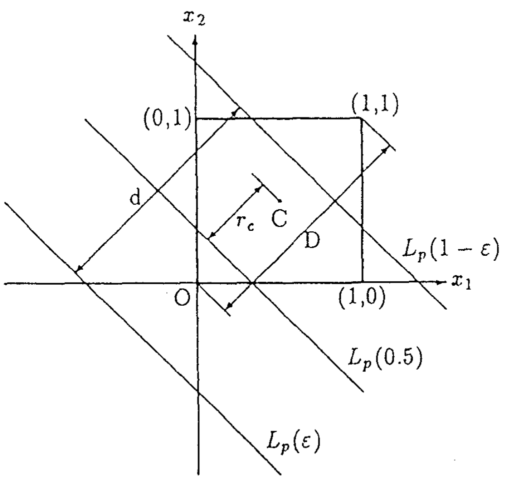

図2　2入力の場合の入力空間

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.

4. 数値実験

ランダム写像問題について, ＯＩＶＳ法と従来法の比較を行なった. (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, 50組の重み初期値についてシュミレーションを行なった. 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. 重みの修正には, 一括修正法を用いた. 誤差逆伝播法における学習率ηは0.7, 慣性項の係数αは0.9 とした. 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ヤーしなかった場合は, 収束しなかったものとした. 使用する係数を変化させて, 様々な組合せでシュミレーションを行い, 最も良い学習性能が得られる範囲を調べた. kとγは, 入力層と隠れ層について共通の値を用いた.

図3と図4にＯＩＶＳ法の結果 (m：反復回数の平均値, σ：反復回数の標準偏差) を, 図5に従来法の結果を示す. 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. 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. 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.

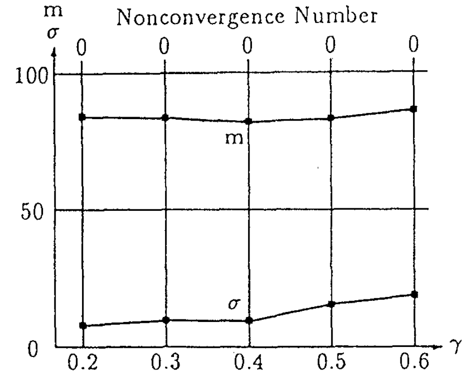

図3　OIVS 法 (k=2.5) による結果：ランダム写像問題における重み分布幅係数 (γ) と反復回数の関係

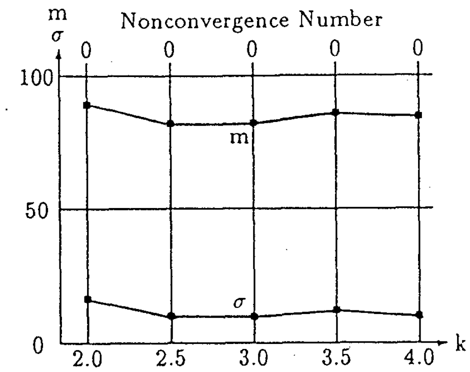

図4　OIVS 法 (γ=0.4) による結果：ランダム写像問題における活性領域幅係数 (k) と反復回数の関係

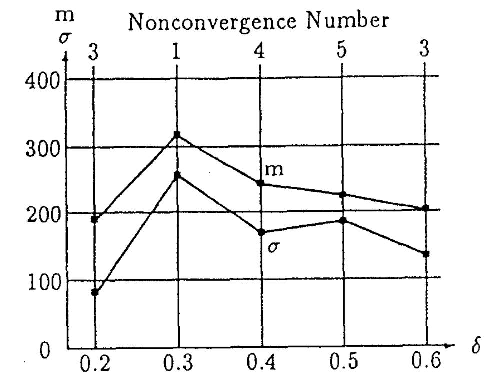

図5　従来法による結果：ランダム写像問題における重み分布幅 (δ) と反復回数の関係

5. むすび

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.

参考文献　1) 賈棋 他: 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, Vol. J73-D-II, No. 8, 1990