

[研究ノート]

3層ニューラルネットワークにおける 誤差特性を利用した関数学習の加速化

田口 功* / 須貝 康雄**

Acceleration of Function Learning by the Use of an Error Characteristic in Three-Layer Neural Networks

Isao TAGUCHI and Yasuo SUGAI

Function learning for which generalization ability is used has been well documented. Errors occur when the general sigmoid function is used in three-layer neural networks. This paper demonstrates an improvement in the learning number by the use of an error rate.

The experiment described here showed that a territory of error is created in the neighborhood of point of inflection of the goal function. In addition, it is shown that the function of an output layer changed while the error territory did not.

*たぐち・いさお：敬愛大学国際学部助教授 情報処理論

Associate Professor of Computer Science, Faculty of International Studies, Keiai University; information processing.

**すかい・やすお：千葉大学大学院自然科学研究科教授 ニューラルネットワーク

Professor of Network Science, Graduate School of Technology, Chiba University; neural network.

1. はじめに

ニューラルネットワークを用いて関数を学習することは、よく知られている。どのような関数でも学習回数を増加させることによって誤差が徐々に減少すれば問題はない。また、学習を行うとき誤差の多く残る領域の学習率を大きく、誤差の少ない領域の学習率を小さくしてはどうか、という考えが生じる。参考文献 (5) では、規格化の仕方によって誤差特性を検討した結果、原関数の変曲点中心の誤差、1 回微分関数の変曲点中心の誤差、さらに誤差が減少するに従って 2 回微分変曲点中心の誤差が残って学習が行われるという実験結果を示した。

本研究では、文献 (5) に述べたように規格化を行った目標関数に対して、学習率一定のときと、誤差特性に応じて学習率を変化させた場合 (ベクトルの学習率と考えられる) の誤差特性を比較し、学習速度の比較を行う。さらに、本実験では、出力層にシグモイド関数を用いた場合、シグモイド関数の 2 回微分関数を用いた場合の比較実験を行う。シグモイド関数の 2 回微分関数の -2.01 倍を行うと、最大値は約 1.0 となり、最小値は約 -1.0 となることに注意をしておくと共に、本稿ではその定数を記号で mmm と記した。

誤差特性を利用した学習時の学習率は、絶対誤差の大きい場所を学習する場合は大きく、誤差が減少している点での学習率は小さくすることにより学習の加速化が行われたことを報告する。なお、実験では、シグモイド関数とシグモイド関数の 2 回微分関数を用意し、出力層に用いた。その結果、2 個の出力層関数に対して、出力関数には依存しない誤差の発生領域や加速化の良好な結果を得た。なお本実験は、一括更新のみで学習を行い、学習間隔は、すべて 0.0953 とした。

2. 3層ニューラルネットワーク

1. ネットワークの構成

最初に、図2.1に示すような3層ニューラルネットワークを考える。文献(5)に示したネットワーク構成と等しいので、詳細は省略する。

ここで、 $\mathbf{x}^{(p)}$ はp番目の入力パターンベクトル、 $t^{(p)}$ はそれに対応する目標値である。 $z_k^{(p)}$ は実際の出力とする。本実験では、出力層ニューロン数kは1である。pは目標関数の覚えこむデータ数によって異なる。 $V_{ji}^{(1)}$ は入力層と中間層の間の重み係数、 $W_{kj}^{(h)}$ は中間層と出力層の間の重み係数である。ここで、iは入力の1パターンあたりの入力数、jは中間層ニューロン数、kは出力層の1パターンあたりの出力数である。出力層にある素子は、入力uに対して、シグモイド関数を使用すると(2-1)式となる。学習は、

$$g(u) = 1/(1 + e^{-ku}) \quad (2-1)$$

$$E = \sum E^{(p)} = \sum \sum \{ (t_k^{(p)} - z_k^{(p)})^2 / 2 \} \quad (2-2)$$

の総誤差Eを最小化することである。ここで、 $t_k^{(p)}$ は目標値であり、 $t_k^{(p)} - z_k^{(p)}$ は、 $\mathbf{x}^{(p)}$ に対する3層ニューラルネットワークの出力と目標値の間の誤差となる。パターン $\mathbf{x}^{(p)}$ が入力として加わった場合、出力誤差 $E^{(p)}$ は、

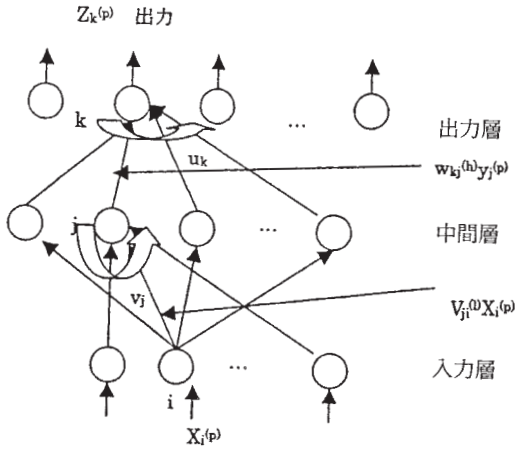
$$E^{(p)} = \sum (t_k^{(p)} - z_k^{(p)})^2 / 2 \quad (2-3)$$

となる。この $E^{(p)}$ を減少させるように $W_{kj}^{(h)}$ と $V_{ji}^{(1)}$ を調節する。 $\Delta W_{kj}^{(h)}$ と、 $\Delta V_{ji}^{(1)}$ は、

$$\begin{aligned} \Delta W_{kj}^{(h)} &= -\varepsilon \partial E^{(p)} / \partial W_{kj}^{(h)} \\ &= -\varepsilon \delta_k^{(h)} \cdot y_j^{(p)} \\ &= -\varepsilon (t_k^{(p)} - z_k^{(p)}) z_k^{(p)} (1 - z_k^{(p)}) y_j^{(p)} \end{aligned} \quad (2-4)$$

$$\begin{aligned} \Delta V_{ji}^{(1)} &= -\varepsilon \partial E^{(p)} / \partial V_{ji}^{(1)} \\ &= -\varepsilon \delta_j^{(1)} x_i^{(p)} \\ &= -\varepsilon (\sum \delta_k^{(h)} W_{kj}^{(h)}) y_j^{(p)} (1 - y_j^{(p)}) \cdot x_i^{(p)} \end{aligned} \quad (2-5)$$

図2.1 3層ニューラルネットワークの構成



となる。ただし、ここで、 ε は学習率である。

$$\delta_k^{(h)} = (t_k^{(p)} - z_k^{(p)}) z_k^{(p)} (1 - z_k^{(p)}) \quad (2-6)$$

$$\delta_j^{(l)} = (\sum \delta_k^{(h)} W_{kj}^{(h)}) y_j^{(p)} (1 - y_j^{(p)}) \quad (2-7)$$

とした（本稿では最急降下法を使用）。ここで、学習係数の誤差に応じた変化を取り入れる誤差に応じた学習係数を考える。学習数を p とし、誤差ベクトル EV を導入する。

$$E^{(1)} = (t_k^{(1)} - z_k^{(1)})^2, E^{(2)} = (t_k^{(2)} - z_k^{(2)})^2, \dots, E^{(p)} = (t_k^{(p)} - z_k^{(p)})^2 \quad (2-8)$$

である。平均を EA とすると、

$$EA = 1/p \cdot \sum E^{(p)} \quad (2-9)$$

となる。ここで

$$EV = [E^{(1)} E^{(2)} \dots E^{(p)}] / EA \quad (2-10)$$

とする。

図3.1.1 目標関数

$$f(X) = \exp(-X^2) \cos(2\pi X)$$

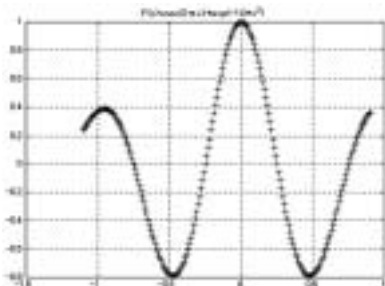
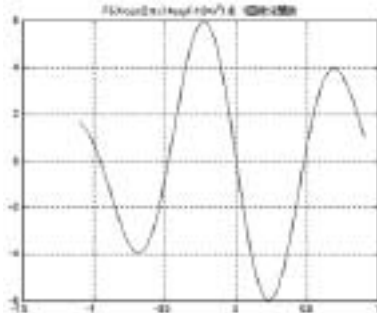


図3.1.2 目標関数の1回微分関数

$$f'(X) = 2\pi \sin(2\pi X) \exp(-X^2) - 2X \cos(2\pi X) \exp(-X^2)$$



3. 計算機実験

1. 学習関数の特性

本稿でニューラルネットワークを利用して、関数学習をする検査関数として次の関数を使用することとした。目標とする関数は、

$$f(X) = \exp(-X^2) \cos(2\pi X)$$

を使用した。この関数は、文献(2)で示した関数の1次元版である。関数 $f(X)$ を図3.1.1に示す。ガウス関数と比較すると、学習が簡単には行われぬのが特徴である。また、その微分関数

$$f'(X) = 2\pi \sin(2\pi X) \exp(-X^2) - 2X \cos(2\pi X) \exp(-X^2)$$

の関係を図3.1.2に示した。さらに $f(X)$ の2回微分関数は、

$$f''(X) = \exp(-X^2) \{8.0\pi X \sin(2\pi X) - 2\cos(2\pi X) + 4X^2 \cos(2\pi X) - 4\pi^2 \cos(2\pi X)\}$$

となる。

4. 一括学習による誤差特性と出力関数

1. 学習率を一定とした場合の学習

(1) 出力層シグモイド関数の場合

最初に、出力層に普通のシグモイド関数を用い、規格化を行った結果を示す。0.0953の学習間隔、目標関数の縮小率 $syuku=0.54$ 、平行移動量 $heiko=0.45$ 、学習率 $myu=1.0$ 一定、2,500回の学習を行った。以下にその出力特性と誤差特性を示す。+印の線が実際の出力特性であり、実線が目標関数を示す(図4.1.1)。

図4.1.1 出力特性
出力層シグモイド関数

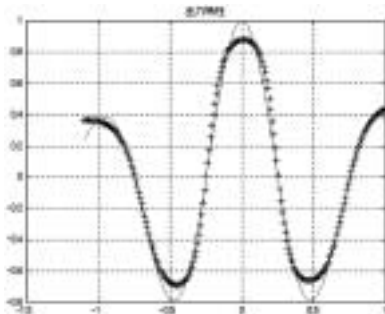


図4.1.2 誤差特性
出力層シグモイド関数

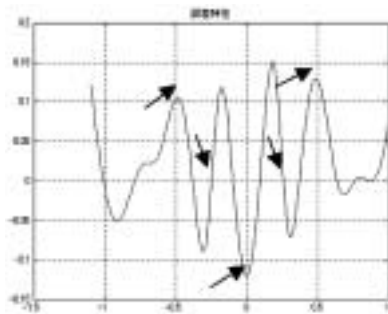


図4.1.3 出力特性 出力層
シグモイド関数2回微分関数

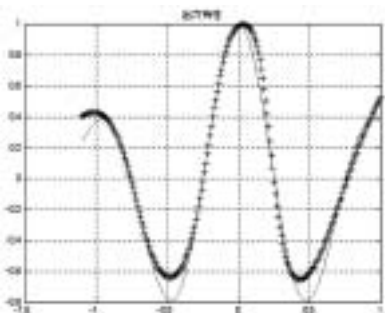
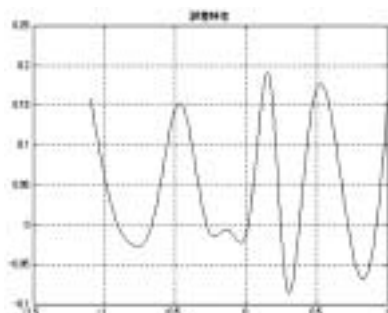


図4.1.4 誤差特性 出力層
シグモイド関数の2回微分関数使用



ここで、図4.1.2において矢印の位置のx軸は、目標関数の変曲点および1回微分関数の変曲点を示す。

(2) 出力層2回微分シグモイド関数のとき

出力層には、シグモイド関数の2回微分関数を用いた。この場合も誤差特性を利用した重み係数更新を行った(図4.1.3)。

2回微分関数の最大値を1とするために $mmm = -2.01$ とした。また、 $syuku = 0.20$ 、 $heiko = -0.02$ 、基礎学習係数 $= 0.16$ とした。8,000回の学習を行った(図4.1.4)。

(3) 出力層関数と誤差特性

実験では(1)および(2)による結果、出力層関数によらない誤差特性を示した。実際には、 $syuku$ 、 $heiko$ を色々変化させて特性を取った結果、図4.1.2、図4.1.4のような特性がほとんどであった。すなわち、誤差特性は、目標関数の変曲点中心誤差、1回微分関数の変曲点中心誤差が最後まで残る結果となった。出力層関数をシグモイド関数にした場合とシグモイド関数の2回微分関数を使用した場合、残る誤差の多い領域は変化しないという結果となった。

5. 誤差特性を利用した学習速度の改善

1. 出力層シグモイド関数のとき

(1) 学習率 $myu = 0.33$ 一定のとき

出力層には、シグモイド関数を用い、4,000回、6,000回、8,000回の学習を行った。その誤差特性をそれぞれ図5.1.1、図5.1.2、図5.1.3に示す。

学習回数が4,000回のときは、目標関数の変曲点で誤差が多いことがわかる。

学習が6,000回になると、目標関数の変曲点中心の誤差が減少し、1回微分関数の変曲点中心の誤差が新たに現れてくるという結果となった。さらに、8,000回に学習回数を増加させると、全体的に誤差は減少すると共

図5.1.1 学習率一定 4,000回

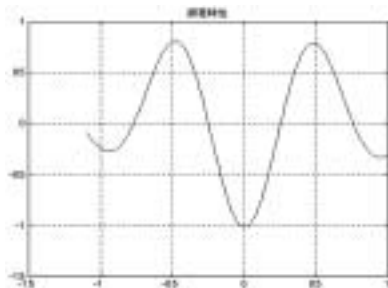


図5.1.2 学習率一定 6,000回

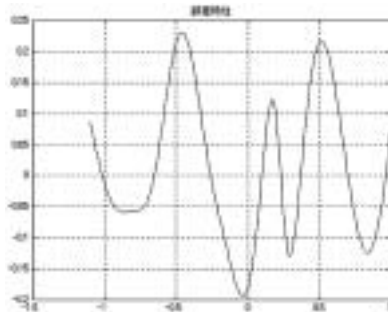


図5.1.3 学習率一定 8,000回

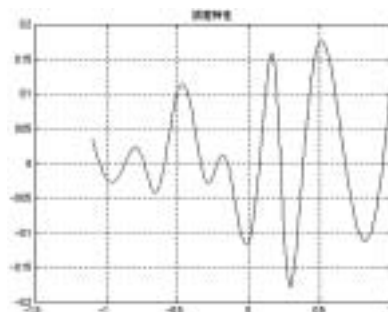


図5.1.4 2乗誤差特性を利用 500回

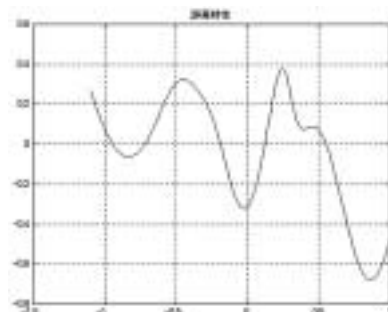


図5.1.5 2乗誤差特性を利用 1,000回

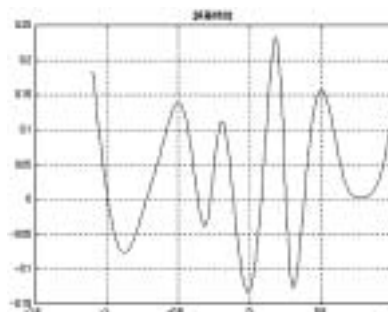
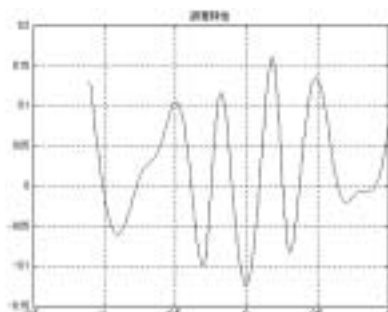


図5.1.6 2乗誤差特性を利用 1,500回



に、その色が強くなるという結果となった。

(2) 学習率を誤差に応じて変化させたとき

ここでは $\text{myu}=0.33$ 一定とし、学習率を $\text{myu} * \text{EV}$ ((2-7) 式) に従って学習係数を調整) とし、学習係数を変化させた場合の誤差特性を、図5.1.4、図5.1.5、図5.1.6に示す。学習回数は、それぞれ500回、1,000回、1,500回とした。学習率一定の場合の8,000回特性と誤差特性を利用した場合は、1,500回の誤差特性に近い特性となっている。学習率一定で8,000回の学習を行った結果は、最大絶対値誤差は約0.17~0.18であり、1,500回の学習を行い、誤差特性を利用した場合は、最大絶対値誤差は約0.16~0.17となっており、6,500回の学習回数の改善が行われるという結果となっている。学習速度は非常に改善されている。

2. 出力層シグモイド関数の2回微分関数使用のとき

(1) $\text{myu}=0.16$ 一定のとき

ここでは、出力層の関数をシグモイド関数からその2回微分関数(-2.01倍)を用い学習を行った。学習回数は、4,000回、6,000回である。結果を図5.2.1、図5.2.2に示す。

4,000回から6,000回と学習が進むにつれて誤差が減少する結果となり、さらに、目標関数の変曲点に誤差が多いことがわかる。またこれは、学習の初期状態では出力層に、シグモイド関数を用いた場合には、目標関数の

図5.2.1 4,000回 学習率0.16一定

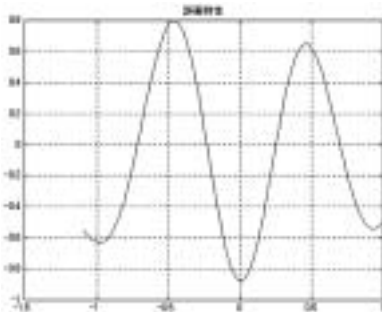
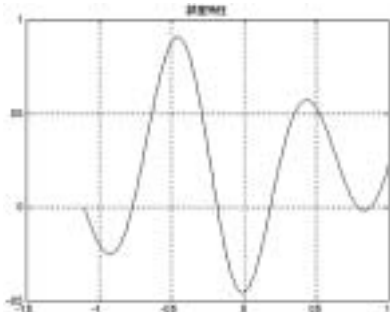


図5.2.2 6,000回 学習率0.16一定



変曲点に誤差が多いという結果と同一となっている。

(2) $\mu=0.16$ 一定とし、誤差特性を学習率に考慮したとき

ここでは、(1) に対して μ は一定として、5.1.と同様、2乗誤差特性を考慮した学習率を計算し、入力値に応じて変化をもたせた。学習回数を2,000回とした場合の誤差特性を図5.2.3に示す。最大絶対値誤差は0.9程度であり、目標関数の変曲点で誤差が多い。さらに、図5.2.4には、学習率を0.16一定とした場合の入出力特性を示し、図5.2.5には $\mu=0.16$ とし、誤差特性を考慮した学習率を使用した場合の入出力特性を示した。図5.2.4と図5.2.5を比較すると、学習の程度としては図5.2.4が進んでいる。誤差特性から考えると図5.2.1と図5.2.2の中間的な誤差特性が、図5.2.3の誤差

特性と考えられる。こう考えると、学習回数で言えば、誤差特性を取り入れた2,000回の学習が、一定学習率の5,000回学習にほぼ相当する形になったと思われる。ここでも誤差特性を考慮した学習速度の改善が行われた。したがって、出力層にシグモイド関数の2回微分関数を用いた場合でも、誤差特

図5.2.3 2,000回 誤差特性考慮

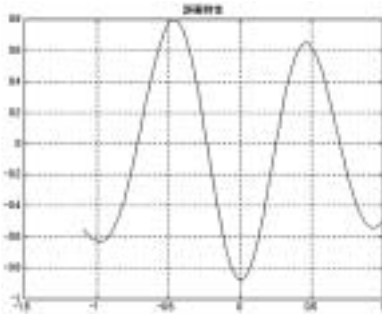


図5.2.4 6,000回 学習率一定
入出力特性

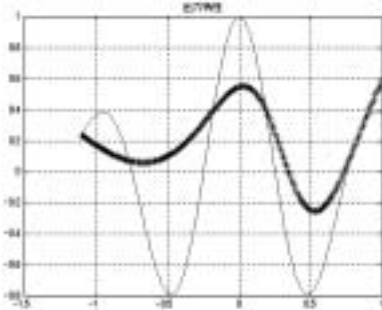
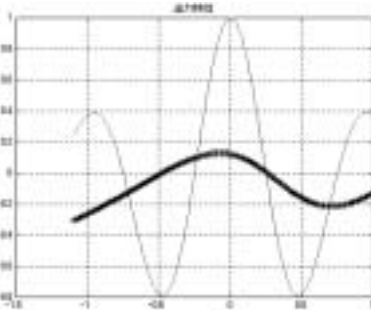


図5.2.5 2,000回 誤差特性考慮
入出力特性



性を考慮した学習率を使用した場合、誤差特性が改善された結果となった。

6. おわりに

関数学習において、条件が悪い場合には、誤差が非常に大きくなる。その場合には、規格化が大切な処理となる(文献〔3〕)。本研究では、規格化をした場合の誤差の発生する領域について最初に考察した。その結果、目標関数の変曲点中心の誤差、目標関数の1回微分変曲点中心の誤差が学習中に残っていることを確認した。誤差が少なくなるに従い、目標関数の2回微分変曲点中心誤差も確認された。これは、出力層にシグモイド関数とシグモイド関数の2回微分関数を用いて実験を行った。その結果、出力層の関数には影響せず、目標関数の変曲点中心の誤差、目標関数の1回微分変曲点中心の誤差が最後まで残った形で学習が行われるということを確認した。

次に、誤差特性を利用し、学習係数を一定とした場合と myu (基礎学習係数)は一定として(2-7)式を用い、学習に対して誤差が多いところは学習係数が大きく、誤差の少ないところは小さくなるように2乗誤差ベクトルを利用し、学習を行った(2乗誤差がちょうど平均値の場合1とする)。この場合も出力層にシグモイド関数とシグモイド関数の2回微分関数を用いて実験を行った。その結果、両関数とも(2-8)式を用いて学習を行った場合、学習速度が大幅に改善された。本研究から、学習時、刻々変わる誤差ベクトルを学習係数とし、利用することによって、学習速度は、改善されるという結果が得られた。

(参考文献)

- (1) 王婷・須貝康雄「非線形多変数関数近似のためのウェーブレットニューラルネットワーク」『電気学会論文誌C』第120巻第2号、2000年、185-193ページ。
- (2) 松井伸之・石見憲一「しきい値ゆらぎをもつニューロンモデルを用いた、階層型ニューラルネットワーク」『電気学会論文誌C』第114巻第11号、1994年、185-193ページ。

- (3) 田口功・須貝康雄「3層ニューラルネットワークの関数学習における誤差領域」『敬愛大学国際研究』第10号、2002年、63-93ページ。
- (4) 計測自動制御学会編：ニューロ・ファジィ・AIハンドブック、88-90ページ。
- (5) 田口功・須貝康雄「3層ニューラルネットワークの関数学習の規格化による誤差変化」『敬愛大学国際研究』第12号、2003年、83-104ページ。