多層高次神経回路網の能力に関する一考察

非会員 馬 立 新 (鹿 児 島 大 学)
 非会員 小 田 幹 雄 (久 留 米 工 業 高 等 専 門 学 校)
 正 員 宮 島 廣 美 (鹿 児 島 大 学)

A consideration on ability of multi-layered higher order neural networks

Ma Lixin, Non-member (Kagoshima University), Mikio Oda, Non-member (Kurume National College of Technology), Hiromi Miyajima, Member (Kagoshima University)

In this paper, the numerical simulations are performed in order to compare ability of HONNs with one of the traditional model in the case where the same number of parameters is used, and effectiveness of HONNs is shown.

キーワード：多層高次神経回路網, 関数近似, 2重螺旋問題, 時系列予測

1. ま え が き

高次の神経回路網は, 従来モデルの一般化したモデルとして定義され, 新しい展開が期待されている^[1]. これまでにも, いくつかの応用において, 優れた能力をもつことが知られている. しかしながらこれらの場合では, いずれもパラメータ数が異なる場合の比較であり, 同程度の場合の能力の比較については, ほとんど行われていない. そこで本稿では, パラメータ数 (重みの個数) を同じ程度にした場合に, 高次多層神経回路網と従来モデルとの能力について数値シミュレーションにより比較検討を行った.

2. 高次神経回路網と誤差逆伝搬法

多数の神経素子を組合せた, 3層2次の高次神経回路網を考える. $y_i^{(m)}$ を第 m 層 ($m=1,2$) $i(i=1,\dots,N_m)$ 番目の素子の出力とすると,

$$y_i^{(m)} = F \left[\sum_{j=1}^{N_{m-1}} w_{ij}^{(m)} y_j^{(m-1)} + \sum_{j=1}^{N_{m-1}-1} \sum_{k=j+1}^{N_{m-1}} w_{ijk}^{(m)} y_j^{(m-1)} y_k^{(m-1)} - \theta_i^{(m)} \right] \quad (1)$$

$$F[x] = \frac{1}{1 + \exp[-x]} \quad (2)$$

となる. 式 (1) の第1項は従来の一次結合であり, 第2項は任意の2入力の積をとり, それに重みを与えたものである. また, $w_{ijk}^{(m)}$ は i 番目素子への入力 $y_j^{(m-1)}$ と $y_k^{(m-1)}$ の積に対する重みであり, $\theta_i^{(m)}$ はしきい値である. 第0層は入力層であり, $y_i^{(0)}$ は入力信号を表す. 学習パターンとして, 入力パターン $y_i^{(0)} (i=1,\dots,N_0)$ と出力層の素子が出すべき望ましい出力 $d_i (i=1,\dots,N_2)$ が与えられるとき, 式 (1) により得られた神経回路網の実際の出力 $y_i^{(2)}$ と期待出力の2乗誤

差の総和 E を表すと次のようになる.

$$E = \frac{1}{2} \sum_{i=1}^{N_2} (y_i^{(2)} - d_i)^2 \quad (3)$$

神経回路網が望ましい出力を出すようにするためには, E を評価関数としてそれを小さくするように重みを逐次修正していく方法が取られる. このような最急降下法の考え方は, 誤差逆伝搬 (BP) 法^[1]と呼ばれており, 高次の場合にも同様に適用できる. この場合, 以下の修正式にしたがって重みを更新する.

$$w_{ij}^{(m)}[t+1] = w_{ij}^{(m)}[t] - \eta \frac{\partial E}{\partial w_{ij}^{(m)}} \quad (4)$$

$$w_{ilk}^{(m)}[t+1] = w_{ilk}^{(m)}[t] - \eta \frac{\partial E}{\partial w_{ilk}^{(m)}} \quad (5)$$

($j=1,\dots,N_{m-1} : l=1,\dots,N_{m-1}-1 : k=l+1,\dots,N_{m-1}$) ここで t は学習回数を表し, η は収束速度を決めるパラメータである. 入力の積の項に対する重みがすべて零の場合が従来モデルである.

3. 数値シミュレーションと評価

ここでは, パラメータ数を同じ程度にした場合の従来モデルと2次モデルの能力を数値シミュレーションにより比較する. 問題としては, 関数近似, 二重螺旋, カオス時系列を扱う.

〈3.1〉 関数近似 目標の関数は, 次の3つとする.

$$z = \frac{(xy)^2 + 1}{2} \quad (6)$$

$$z = \frac{4\sin(\pi x) + 2\cos(\pi y)}{12.0} + 0.5 \quad (7)$$

$$z = 1.9[1.35 + e^x \sin(13(x - 0.6)^2) e^{-y} \sin(7y)] \quad (8)$$

各関数において、入力変数は x, y , 出力変数は z である。関数 (6),(7),(8) に対して重みの総数は 45,45,84 個、学習用データの数は、100,100,500 個とする。ここに入力データは区間 $[-1,1]$ からランダムに選び、出力は区間 $[0,1]$ に正規化する。テストケースでは、区間から一様にそれぞれ、529,529,1024 個選ばれた入力データに対して、回路網の出力と理想出力の比較を行った。表 1 において、A の行は 2 次モデルに対する結果、B の行は従来モデルに対する結果である。学習回数は、関数 (6),(7) が 10000 回であり、関数 (8) が 30000 回であり、結果として、100 回の平均、最小、最大の 2 乗誤差をそれぞれ示している。関数 (6),(7) については、2 次モデルが優れ、関数 (8) については、従来モデルがやや優れている。

表 1 関数近似の結果
Table 1. Result for function approximations

		平均誤差	最小誤差	最大誤差
(6)	A	0.00001575	0.0000272	0.00018930
	B	0.00024965	0.00010307	0.00057698
(7)	A	0.00020131	0.00008310	0.00049346
	B	0.00034212	0.00016617	0.00059953
(8)	A	0.00600602	0.00321706	0.01225440
	B	0.00540734	0.00387050	0.00700634

〈 3 2 〉 二重螺旋問題^[2] はじめに 2 次元平面上に二重螺旋を形作る二つのデータ群 (例えば白または黒) をそれぞれ 97 個づつ準備する。これを、入力が 2 変数 (平面上の点)、出力が 1 変数 (どちらのデータ群か) からなる 3 層神経回路網を使って学習する。学習後の回路網に、2 次元平面上の任意の点を与えて結果が白か黒か判断し、パターン分離の状態を図示する。この問題は、従来モデルではよい結果を示さないことが知られている。結果については、文献 3 を参照されたい。このような複雑なパターン分離問題に対しては、従来モデルより 2 次モデルの方が優れている。

〈 3 3 〉 Mackey-Glass のカオス時系列予測問題^[2] この問題は、現在利用可能な時系列の値を使って将来の値を予測する。この時系列が遅延 $\tau = 17$ によって定義された Mackey-Glass の微分方程式は次のように示される。

$$\frac{dx_t}{dt} = \frac{ax_{t-17}}{1 + x_{t-17}^{10}} - bx_t \dots\dots\dots (9)$$

この場合、時系列 x_t はカオス時系列となる。学習段階では $t=103$ から 1103 までのデータを使い、 $t-18, t-12, t-6$ と t 時刻の 4 つを入力とし $t+18$ 時刻の値を予測する回路網を構築する。テスト段階では、 $t=3603$ から $t=4103$ までのデータに対して、回路網出力と理想出力の比較を行った。表 2 は結果である。表 2 の A, B 等の記法は表 1 と同じ使い方である。学習回数は 10000 回で、重みの個数を 111 と 205 としシミュレーションを行った。表 2 により、同じ重み数を使った場合、2 次モデルの方が高い学習能力を示していることが分かる。

〈 3 4 〉 少数の学習データセット による関数近似 問題としては〈 3.1 〉の関数近似と同じであるが、学習用データを少数にした場合の汎化能力を比較する。次の関数を用い

表 2 カオス時系列予測問題に対する結果
Table 2. Result on prediction for time series

		平均誤差	最小誤差	最大誤差
重み 111	A	0.00033714	0.00015560	0.0007830
	B	0.00048877	0.00026108	0.0009895
重み 205	A	0.00011335	0.00007213	0.0002187
	B	0.00044150	0.00021179	0.0007648

て、5 個の入出力データを与える。ここでは、これをスプライン関数で補間した結果^[4] を目標関数として比較する。

$$\begin{aligned} &1 - 2x && x \leq -3 \\ &28 + 25x + 9x^2 + x^3 && -3 \leq x \leq -1 \\ &26 + 19x + 3x^2 - x^3 && -1 \leq x \leq 0 \\ &26 + 19x + 3x^2 - 2x^3 && 0 \leq x \leq 3 \dots (10) \\ &-163 + 208x - 60x^2 + 5x^3 && 3 \leq x \leq 4 \\ &157 - 32x && 4 \leq x \end{aligned}$$

5 個の入出力データを $(-3, 7), (-1, 11), (0, 26), (3, 56), (4, 29)$ とする。入出力は 1 変数、中間層の素子数を、1 次では、4 個、2 次では 2 個とする。5000 回の学習を行い、テストデータにより比較した結果を図 1 に示す。従来モデルの中間層の素子数を変えても結果の改善は見られなかった。この場合、2 次ではかなりうまく近似が行われている。

図 1 スプライン関数の近似
Fig. 1. Approximation for a spline function

4. ま と め

関数近似、パターン分離 (二重螺旋)、時系列予測の異なるタイプの 3 つの問題に対して、同じ程度のパラメータ数をもつ従来モデル、2 次モデルによりシステム同定を行った。結果として、多くの場合で、2 次モデルが優れていることが分かった。

(平成 10 年 10 月 23 日受付)

文 献

(1) D.E.Rumelhart and J.L.McClelland." Parallel Distributed Processing," MIT Press, Cambridge, 1986
(2) S.Fahlman, M.white" The Carnegie Mellon University Collection of Neural Net Benchmarks," from ftp.cs.cmu.edu.
(3) " 高次ランダム神経回路網のパターン間の距離の変換特性" 宮島, 馬, 重井, 電気学会論文誌 C, vol. 118-C, 1516~1521(1998)
(4) 桜井義." スプライン関数入門", 東京電機大学出版局