

THE TRANSACTIONS OF THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS D-II

cic | **電子情報通信学会**
DII | **論文誌** VOL. J78-D-II NO.4 APRIL
情報・システムⅡ—情報処理 — 1995

情報・システムソサイエティ

社団法人 **電子情報通信学会**

THE INFORMATION AND SYSTEMS SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

選択的に学習を行うニューラルネット

正員 大江 修造[†]

Selective Training Neural Net
Shuzo OHEI, Member

[†] 東京理科大学工学部経営工学科, 東京都
Faculty of Engineering, The Science University of Tokyo, Tokyo, 162
Japan

あらまし バックプロパゲーション法は連続な関数を近似する汎化能力を有している。連続な関数が微分不可能な点を有していない場合は、バックプロパゲーション法の近似で支障は生じない。しかし、連続な関数が微分不可能な点を有している場合は、バックプロパゲーション法の近似では尖点を有する曲線に沿って分布しているデータの学習結果は振動を生じ、支障を来す。本論文はこの問題を解決するために、選択的学習方式を提案する。本論文の方法による結果は、特に、尖点の近傍における汎化能力の改善が著しい。

キーワード ニューラルネット, 汎化能力, バックプロパゲーション法, 選択的学習, 微分不可能な点

1. まえがき

ニューラルネットにおけるバックプロパゲーション法⁽¹⁾はニューラルネットにおけるモデルの中で最も実用的であるという高い評価^{(2),(3)}を得ていて、関連する研究結果も多い^{(4)~(6)}。ニューラルネットの応用は文字認識などの分類と関数補間の分野とに大別される。後者について、バックプロパゲーション法は任意の連続写像を、伝達関数として sigmoid-like な関数を用いることにより、任意の精度で近似的に実現できることが証明されている⁽⁴⁾。

関数補間にニューラルネットを利用する研究は相当程度行われてきてはいるものの、十分になされているとは言えない。特に、微分不可能な点、すなわち尖点を含んだデータの補間については未着手の状態にある。連続関数の補間はバックプロパゲーション法により、近似は可能であるが、尖点を表現することは不可能である。補間法は連続関数に沿って分布するデータの圧縮法として、データ表現に有効な方法である。さまざまな分野でデータ圧縮の要求がある。物理、化学の分野では混合物質系の状態を表示する状態図が広く利用されている⁽⁹⁾。状態図には尖点を含んだデータが多く、その実測データを代表データとしてニューラルネットの汎化能力により表現できれば、この分野におけるニュー

ラルネットの利用価値は大幅に増大する。

本研究では、バックプロパゲーション法において、教師付き学習を選択的に採用することにより、従来のバックプロパゲーション法では不可能であった尖点を有する曲線上のデータの補間を可能にした。本研究の方法では、出力層に複数のユニットを設け、各ユニットは尖点で挟まれた区間の出力値を担当させる。本研究で扱うデータは以下に示すデータである。すなわち、閉区間において定義される連続関数がこの区間内に有限個の尖点を有していて、この連続関数のグラフに沿って分布しているデータが与えられ、しかもそのデータの中には、その連続関数の尖点がすべて含まれている場合である。

このデータを尖点で挟まれたある一つの区間で学習させるには、その区間についてのみ教師付き学習を実行させ、他の区間については学習を行わせない。この学習法を採用することにより、担当する出力ユニットの出力に対して、担当外のデータによる影響を取り除くことができるので、尖点を有する曲線上のデータを補間する際に発生しやすい振動を抑制する効果もある。本方法はバックプロパゲーション法において、教師付き学習を選択的に採用する方法であるので、選択的バックプロパゲーション (Selective Back Propagation = SBP) 法と呼ぶ。

2. 選択的バックプロパゲーション (SBP) 法

本研究では、4層の階層構造型ニューラルネットを用いる。尖点を有する曲線上のデータの補間に際して、図1に示すニューラルネットを用いる。データ x , y において y が x に対して関数関係にあり、この関数が尖点を含む場合を考える。図1は、尖点 n 個を有する曲線で近似されるデータに対するものとする。 $y_1, y_2, \dots, y_i, \dots, y_{n+1}$ はそれぞれ2点の尖点に対する x

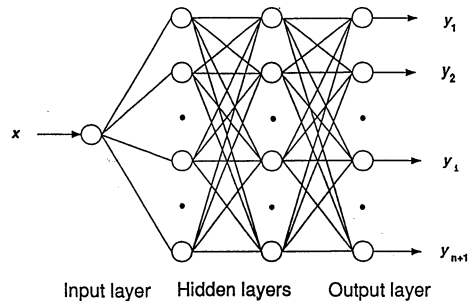


図1 教師付き学習の選択的採用のための4層ニューラルネット
Fig. 1 Four layers network for selective training with supervisor.

で挟まれた区間における y の出力値を示す。特定の区間の学習を行わせる場合は、その特定の区間についてのみ教師付き学習を実行し、他の区間については学習を行わせない。すなわち、出力ユニットにおける誤差信号をゼロとする。

SBP 法においては、出力ユニットに対して、選択的に教師データによる重み値およびしきい値の修正を行うので、一般化デルタルール

$$\Delta p_{pj} = \eta \delta_{pj} o_{pj} \quad (1)$$

において、

$$\delta_{pj} = S_{pj} o_{pj} (1 - o_{pj}) (t_{pj} - o_{pj}) \quad (2)$$

として、選択係数 S_{pj} を採用する。出力ユニット j に対して重み値の修正を行う場合は $S_{pj} = 1$ とし、重み値の修正を行わない場合は $S_{pj} = 0$ とする。

ここに、 Δp_{pj} は入・出力ペア p における j 番目から i 番目のユニット間の重みの変化量である。 δ_{pj} は入・出力ペア p における j 番目のユニットの誤差信号である。 o_{pi} および o_{pj} は入・出力ペア p における i 番目および j 番目の出力パターンである。 η は定数である。 t_{pj} はパターン p における j 番目の教師信号である。

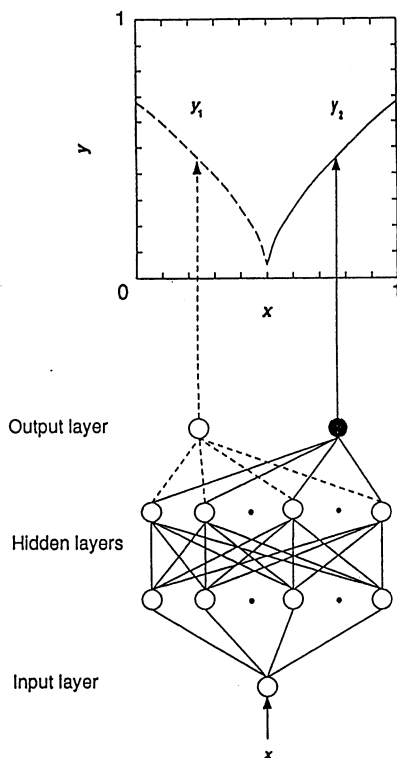


図2 教師付き学習の選択的採用
Fig. 2 Selective training with supervisor.

しきい値は重み値と同様に修正を行う。隠れ層と隠れ層間および隠れ層と入力層間における重み値の修正はバックプロパゲーション法と同じ方法により行う。

SBP 法とバックプロパゲーション法との相違点は出力層と出力層に隣接する隠れ層間の重み値としきい値を選択的に修正する点にある。SBP 法では隠れ層と隠れ層間および隠れ層と入力層間における重み値としきい値の修正方法はバックプロパゲーション法と同じである。この結果、SBP 法はバックプロパゲーション法を線形的に合成したものとはならない。それ故、尖点を有する曲線上のデータの補間を単一のニューラルネットにより実現できることになる。

2.1 SBP 法のアルゴリズム

選択的学習法を、図2により説明する。出力層の両ユニットはグラフ上の矢印で示した部分の補間を「担当」する。図において入力層への x のデータが●印で示された出力ユニットの「担当」する区間内にあるときは●印の出力ユニットについては教師付き学習を実行し、○印の出力ユニットについては学習を執行させない。

SBP 法のアルゴリズムを図3に示す。出力ユニットの数 $(n+1)$ をデータの尖点の数 (n) に対して決定する。 x の区間に対応して教師付き学習か否かを決定する。この決定は入力データに加えた判別用のデータにより行う。必要な出力ユニットにのみ教師付き学習の教師データを設定する。

尖点を有する曲線上のデータの学習能力を検証するために関数式(3)から得られたデータを用いた。学習データは x を0から1まで0.1おきに变化させて式(3)から求めた y の値を用いた。但し、 x が0.4から0.6の範囲では0.02おきに变化させた。

$$y = \{(x-1/2)^2\}^{1/3} \quad (0 \leq x \leq 1) \quad (3)$$

学習結果による補間の結果を図4に示す。図4において、 y の値は0.05から0.95の範囲に正規化して、シグモイド関数における0, 1近傍の影響を除外した。実線部分は選択的学習の結果であり、点線部分は無学習の結果である。各点線部分は接続している各実線部分を延長した形状をしている。

2.2 SBP 法による学習

尖点を有する曲線上のデータの学習の手順を示す。

- ① 隠れ層(2層)のユニット数を決定する。ユニット数は、データにおける尖点の数 (n) により最適な個数を実験的に決定する。本研究では二つの隠れ層のユニット数は同数とし、 $n=1, 2$ の場合は10とし、 $n=3$ の場

```

if number of cuspidal points is n
  then number of output units is n+1
for (first output unit; last output unit)
  if output unit( $y_i$ ) is corresponding closed interval( $x_i$ )
    then train with supervisor's data
    else do not train
for (first output unit; last output unit)
  if training with supervisor's data
    then change the threshold of the unit
    else do not change the threshold of the unit
for (first unit; last unit at hidden layer 2)
  if training with supervisor's data
    then change the weights between the connections
    else do not change the weights between them
for (first unit; last unit at hidden layer 2)
  change the threshold of the unit
for (first unit; last unit at hidden layer 1)
  change the weights between the connections
for (first unit; last unit at hidden layer 1)
  change the threshold of the unit
for (first unit; last unit at input layer)
  change the weights between the connections

```

図3 選択的バックプロパゲーション法のアルゴリズム
Fig. 3 Algorithm of selective training neural net.

合は15とした。

② 学習回数は学習誤差が相対誤差で1%以下となることを目標に50万回とした。50万回の学習後、出力値に誤差が1%以上のものが存在する場合、再度学習を行わせ、すべての教師データの誤差が1%未満になるまで学習を行わせる。

2.3 SBP法による補間

2.2で示した学習で得られたニューラルネット構造を用い、以下の手順で補間を行う。

① 補間する x の入力値を、ニューラルネットの入力層のユニットに入力する。

② 出力層の各出力ユニットからの出力値の中から、入力値 x の区間に対応する「担当」の出力ユニットを教師(使用者)が選び、ニューラルネットからの出力値とする。

3. 実験結果

本方法の適用結果を尖点の個数 n が1, 2, 3の場合

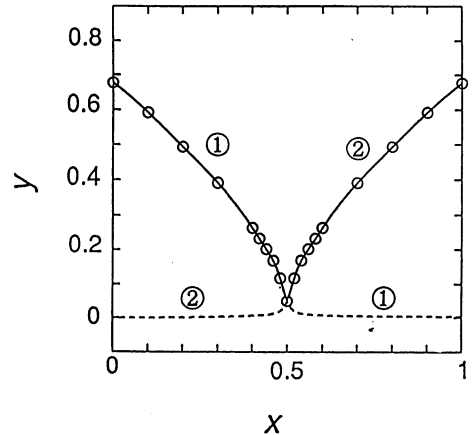


図4 教師付き学習の選択的採用による結果
Fig. 4 Result by selective training with supervisor.

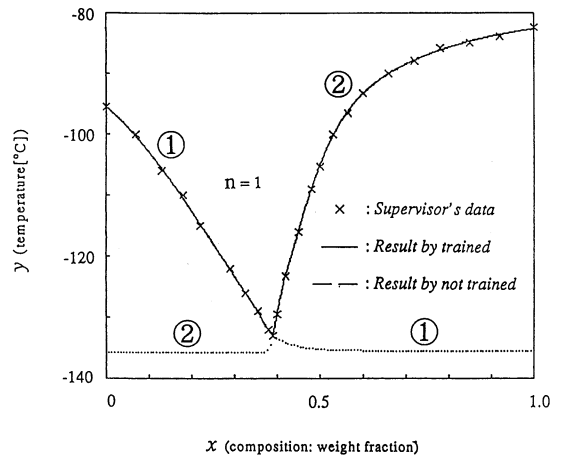


図5 選択的バックプロパゲーション法による補間結果
($n=2$ の場合: データ; Methanol-H₂Sの固液平衡)
Fig. 5 Result by selective training neural net.
($n=2$: Solid-liquid equilibrium data for methanol + H₂S system)

につき以下に示す。出力データの正規化は

正規化された値

$$= 0.05 + (0.95 - 0.05)$$

$$\times (\text{原データ} - \text{最小値}) / (\text{最大値} - \text{最小値})$$

によった。最大値、最小値は原データから正規化のために採用した値である。

(1) $n=1$ の例

尖点は $x=0.391$, $y=-133.0$ である。出力ユニット数は2であるが、尖点における誤差は区間(1) [0, 0.391]では0%, 区間(2) [0.391, 1]では0.1%となった。平均誤差は区間(1)では、0.19%, 区間(2)では、0.18%となった。誤差の最大値は区間(1)では、0.5%, 区間

(2)では、0.4%であり、誤差の最小値は区間(1)、区間(2)ともに、0.0%となった。図5に補間結果を示す。図5から、無学習の結果(点線部分)を確認することができる。無学習による曲線部分が約-136でほぼ一定となっているのは、出力値の正規化による結果である。

(2) $n=2$ の例

尖点は $x=0.700, y=-105.0$ と $x=0.820, y=-125.0$ である。出力ユニット数は3である。最初の尖点における誤差は区間(1) $[0, 0.700]$ では1.0%、区間(2) $[0.700, 0.820]$ では0.1%となった。次の尖点における誤差は区間(2) $[0.700, 0.820]$ では0.2%、区間(3) $[0.820, 1]$ では0.0%となった。平均誤差は区間(1)では0.64%、区間(2)では0.35%、区間(3)では0.17%となった。誤差の最大値は、区間(1)では1.9%、区間(2)では0.7%、区間(3)では0.4%であり、最小値は区間(1)、区間(3)ともに0.0%となったが、区間(2)は0.1%となった。図6に $n=2$ の補間結果を示す。

(3) $n=3$ の例

尖点は $x=0.126, y=8.8$ 、 $x=0.535, y=10.7$ および $x=0.756, y=13.4$ である。出力ユニット数は4である。最初の尖点における誤差は区間(1) $[0, 0.126]$ では0.1%、区間(2) $[0.126, 0.535]$ では0.5%となった。次の尖点における誤差は区間(2) $[0.126, 0.535]$ では0.2%、区間(3) $[0.535, 0.756]$ では1.5%となった。最後の尖点における誤差は、区間(3) $[0.535, 0.756]$ では0.1%、区間(4) $[0.756, 1]$ でもやはり0.1%となった。

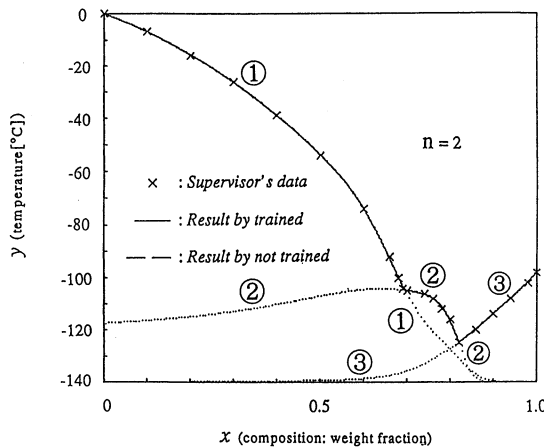


図6 選択的バックプロパゲーション法による補間結果 ($n=2$ の場合: データ: Methanol-H₂O の固液平衡)
Fig. 6 Result by selective training neural net. ($n=3$: Solid-liquid equilibrium data for methanol + H₂O system)

平均誤差は区間(1)では0.07%、区間(2)では0.73%となり、区間(3)では0.69%となり、区間(4)では0.05%となった。誤差の最大値は区間(1)では0.1%、区間(2)では1.5%、区間(3)では1.7%、区間(4)では0.1%であり、最小値は区間(1)、区間(4)がともに0.0%となったが、区間(2)は0.2%、区間(3)は0.1%となった。図7に $n=3$ の補間結果を示した。

図6、図7において、無学習による曲線部分は、「担

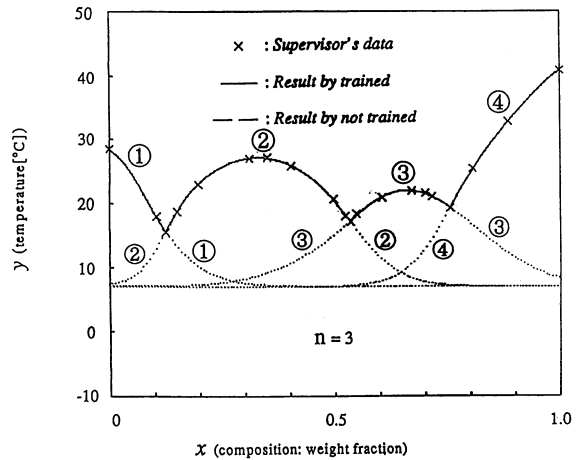


図7 選択的バックプロパゲーション法による学習結果 ($n=3$ の場合: データ: phenol+2-methyl-2-propanol の固液平衡)
Fig. 7 Result by selective training neural net. ($n=3$: Solid-liquid equilibrium data for phenol+2-methyl-2-propanol system)

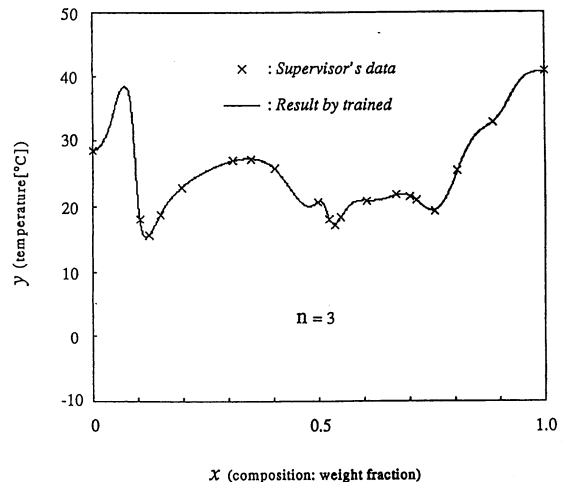


図8 従来のバックプロパゲーション法(4層)による学習結果
Fig. 8 Result by conventional back propagation method. (four layers)

当」区間の出力値を延長した形になっていることが認められる。この部分に教師付き学習をさせると、その影響によって、他の「担当」区間の出力値が期待どおりにならないことが理解できる。

図7において、尖点は本学習法により明確に表現できていることがわかる。ところが、同データを従来法により学習させると図8に示すように、尖点を明確に表現できない。確かに、尖点を含んだデータを近似できることは図8からわかるが、あくまでも近似にすぎない。図8において、データの学習はされているものの、併せて振動も見られる。特に x の区間 $[0, 0.126]$ では大幅な振動が生じている。小規模な振動は各区間 $[0.126, 0.535]$, $[0.535, 0.756]$, $[0.756, 1]$ において認められる。

4. 考 察

補間結果から、選択的学習による効果として、次の3点を確認した。

① 教師付き学習を円滑に行わせる。教師付き学習の選択的採用を行わずに、従来のバックプロパゲーション法により尖点を含むデータを学習させると、データ相互間の学習により干渉を生じ、学習が円滑になされない。これに対して選択的に教師付き学習を採用することにより、この干渉を除外することができる結果を得た。

② 教師データの中に、教師データ全体の分布を考慮した場合、誤差の比較的大きいと思われるデータが存在している場合に、補間に際して振動の発生を抑制する。教師付き学習のみのニューラルネットを用い、データの補間を行った場合、データの中に誤差の比較的大きい値が存在すると学習が進むにつれ、補間結果に振動が発生する。それに対し、教師付き学習を選択的に採用したニューラルネットを用いて、データに対して補間を行った場合、誤差の大きな部分においては、最小2乗法を用いた近似結果に似た補間結果となる。このような補間結果が得られる要因は次のように考えられる。「担当」出力ユニットにおいて、「担当外」の出力ユニットにおける無学習のために、「担当外」の区間において「担当」区間の曲線を延長した形となったが、この延長部分が逆に「担当」区間の振動を抑制する働きを有していると考えられる。

③ 過学習による補間結果の振動を抑制する。教師データの学習を繰り返すことによって、教師データと出力値の誤差は小さくなるが、教師データと教師データ間の間で振動が生じやすい。従来のバックプロパゲ

ーション法の場合、全教師データが学習に反映されるため、誤差は減少するが、隣接する教師データ間の補間値の誤差が増大する結果となる。しかし、教師付き学習の選択的採用によって、教師付き学習の過学習を無学習によって抑制させる働きがある。

5. む す び

本研究は尖点を有する曲線上のデータを単一のニューラルネットにより一括して補間することが可能であることを示した。それは従来のバックプロパゲーション法に対して教師付き学習を選択的に採用することにより可能となった。特に、選択的学習の反面としての無学習は、尖点を有する曲線上のデータの補間に際して、「担当」区間への干渉を防ぐ機能を有していると考えられる。このために、補間の際の振動を抑制し、教師付き学習を円滑に行わせる働きがあると考えられる。

尖点を有する1価関数のグラフ上のデータに対して、SBP法を適用したが、尖点を有する多価関数のグラフ上のデータに対しても、同様にSBP法の適用の可能性が予想されるので、今後の研究課題として検討していきたい。

文 献

- (1) Rumelhart D. E., McClelland J. L. and the PDP Research Group: "Parallel Distributed Processing", Rumelhart D. E., Hinton G. E. and Williams R. J., "Learning Internal Representations by Error Propagation", Volume 1, Chapter 8, The MIT Press, Cambridge, Massachusetts (1986).
- (2) 上坂吉則: "学習認識機械としてのニューラルネットへの期待と課題", 電学論(D), **111**, pp. 2-9 (1991-01).
- (3) 上坂吉則: "ニューラルネットと学習可能性", 信学誌, **74**, pp. 943-948 (1991-09).
- (4) 船橋賢一: "ニューラル・ネットワークによる連続写像の近似的実現について", 信学技報, **MBE88-9** (1988).
- (5) 戸田尚広, 萩原克幸, 白井支朗: "多層ニューラルネットワークによるデータフィッティング-情報統計量による構造決定", 信学技報, **NC89-100** (1988).
- (6) 栗田多喜夫: "ニューラルネットにおけるモデル選択の試み", 信学技報, **PRU89-16** (1989).
- (7) 高橋祥兼: "曲線連続写像を有限学習データから汎化・近似する3階層ニューラルネット", 信学論(D-II), **J76-D-II**, 7, pp. 1458-1462 (1993-07).
- (8) 米倉達広, 宮崎慎也, 鳥脇純一郎: "恒等写像モデルおよびPPNに基づく4層ニューラルネットの情報集約機能の解析", 信学論(D-II), **J74-D-II**, 10, pp. 1398-1410 (1991-10).
- (9) Landolt-Bornstein: "Zahlenwerte Und Funktionen aus Physik·Chemie·Astronomie·Geophysik·Technik", 2-b -I, pp. 3-404-3-484, Springer-Verlag (1962).

(平成6年11月15日受付)