1L−10　パターン認識用ニューラルネットはいかにして構成できるか？

阿部重夫　鹿山昌宏　武長寛

(株)日立製作所　日立研究所

1．はじめに

教師データを用いてシナプスの重みを学習することにより容易にパターン認識用のニューラルネット[1]が構成できるため、パターン認識はニューラルネットの重要な応用分野となっている。しかしながら、アルゴリズムが不要であること、あるいはニューラルネットがブラックボックスとして扱えるというメリットが逆にディメリットとなることがある。例えば誤認識が生じたとき認識率を改善する方法はそのデータを追加して再学習するしか今のところ手が無い。このような問題を解決して実用的なニューラルネット技術とするために我々はニューラルネットの特性の解明を進めた。その結果ニューロンの飽和特性の意味、パターン認識用ネットが如何にして構成しうるかについて解明できた。

2．飽和特性はパターン認識に必須である

図1の3層ニューラルネットについて考える。図において

$$Zj(i)=f(Xj(i))= 1/(1+\exp(-Xj(i)/T)) \qquad (1)$$

$$Xj(i)= \sum_{k=1}^{n(i-1)+1} Wkj(i-1)Zk(i-1) \qquad (2)$$

$$Zn(i-1)+1=1$$

ここでZj(i):i層ニューロンjの出力

Xj(i):i層ニューロンjへの入力

Wkj(i-1):i-1層ニューロンkとi層ニューロンjとを結ぶシナプスの重み

n(i):i層のニューロン数

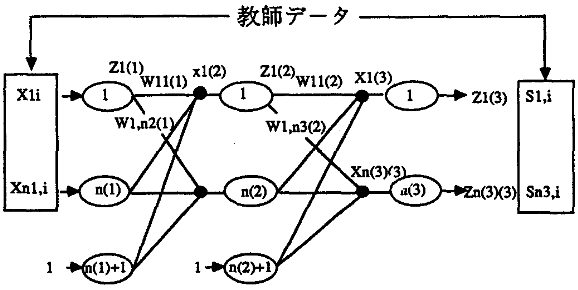

図1　3層ニューラルネット

Kolmogorovの定理によれば任意の連続関数は単調増加関数の四重の合成関数でいくらでも詳細に近似できる[2]。従って多層ネットの非線形関数は、飽和特性を持つシグモイド関数でなくてもよいはずである。飽和領域では微係数が0となるため学習の収束性を考えれば飽和特性のないほうがよいといえる。

これに対する答えは次の問題に答えることによって得られた。

中間層のニューロンの数が入出力のニューロンの数より少なくてもパターン分離は出来るか？

n(2) < n(3), n(2) < n(1)として中間層と出力層との関係を考える。n(3)個のパターン入力に対応する中間層出力及び出力層の入出力をZij(2),Xij(3),Zij(3), i,j=1,...,n(3)とすると

$$\{Zij(3)\}= \begin{bmatrix} 10 & 0 \\ 01 & 0 \\ \cdots\cdots & \\ 00 & 1 \end{bmatrix} \quad n(3)\times n(3)\text{次} \qquad (3)$$

ととれる。すなわち行列{Zij(3)}のランクはn(3)である。これに対して

$$\{Xij(3)\}=\{Wik(2)\}\times\{Zkj(2)\} \qquad (4)$$

のランクはn(2)<n(3)であるから行列{Wik(2)},{Zkj(2)}のランクはn(2)で従って{Xij(3)}のランクもn(2)となる。すなわちランクがn(1)の入力情報からn(2)に落ちた中間層情報に基いてランクn(3)の情報が再生できるかという問題となる。問題を解く鍵は飽和特性にある。ここでネットワーク出力1,0を1-ε, ε (ε>0)とし指定値の±ε以内に入ったとき収束したとしよう。このときZi(3)の値は、

$$1 \geqq Zi(3) \geqq 1-2\varepsilon \quad \text{あるいは} \quad 2\varepsilon \geqq Zi(3) \geqq 0 \qquad (5)$$

How can the Neural Networks for Pattern Recognition be Synthesized ?

Shigeo ABE, Masahiro KAYAMA,Hiroshi TAKENAGA
Hitachi, Ltd.

となるから、これに対するXi(3)は

$$\infty > Xi(3) \geqq -T \log(1/(1-2\varepsilon)-1)$$

あるいは $-T \log(1/2\varepsilon-1) \geqq Xi(3) > -\infty$ (6)

となる。これより{Xij(3)}のランクがn(2)でも{Zij(3)}のランクはn(3)となりうるのである。すなわち、中間層の数は入出力ニュ-ロンの数より少なくても分離はできるのである。このように飽和特性がパターン分離に重要な役割を果たしているのである。以上のように収束判定を行うとすると1入力n出力の1層のニューラルネットの能力に関して次のことを証明できる。

一つの変数Z(1)= [0,1]のn個の連結成分は1層のニューラルネットで分離できる。

一つの変数Z(1)=[0,1]の異なる連結成分は1層のニューラルネットでは同一パターンに分類出来ない。

2番目の性質はMinskyのPerceptronの能力に関する証明と同じ意味を持つ。

3.パターン認識は3層ネットで行えるか?

文献2)によれば任意の連続関数は3層ニューラルネットで構成できる。これに対し文献3)では関数近似を行うときは3層より4層の方が近似の精度、及びニューロン数の点から有利であると述べている。しかしながらどのようにすればネットが構成されるかについては分かっていない。本節では次のことを示す。

n次元デ-タをm個のパタ-ンに分類することを考える。n次元空間にk個の超平面が存在しこれらの超平面によりm個の各パタ-ンが単連結領域に分離できるとき、入力n,中間層k,出力mの3層ニューラルネットでパターン分離が可能である。

証明の概略を以下に示す。k個の超平面を表現する一次方程式の係数をn個の入力とk個の中間層の間の重みWij(1)として設定する。これにより分離平面上の点に対応する中間層出力は1/2となり各パターンに対応する点では1/2とはならない。従って重みを定数倍することにより各パターンに対する中間層出力を1,あるいは0とすることができる。

中間層と出力の重みはパターンごとに合成できi番目の出力ニュ-ロンではパタ-ンiが入力のとき出力が1でそれ以外のときは0になればよい。ここでXi(3)≧αのとき1に対応しXi(3)≦-αのとき0に対応するとすると、パターンiの領域を決定するのに貢献している中間層ニュ-ロンと出力ニュ-ロンと間の重みを中間層出力が1のとき2α，0のとき-2αとすればよい。(証明終わり)

あるパタ-ンが複数の単連結領域からなるときは、各単連結領域に対して上記手順を行い4層目でそれらを合成すればよい。なおこの証明では証明を簡単にするために、中間層出力を1,0としたが実際にはその必要はない。

図2の2次元パタ-ンについて上記手順を行うと図3のように合成できる。このときパターンIIは単連結でないのでパタ-ンIIのみ4層となる。また平面P1はパタ-ンIを分離するのに貢献していないからP1に対応する中間層ニュ-ロンはパタ-ンIに対応する出力ニュ-ロンに接続されていない。

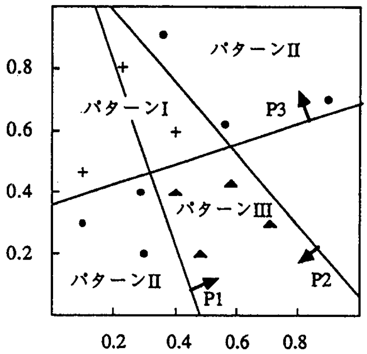

図２　平面上の点の分離

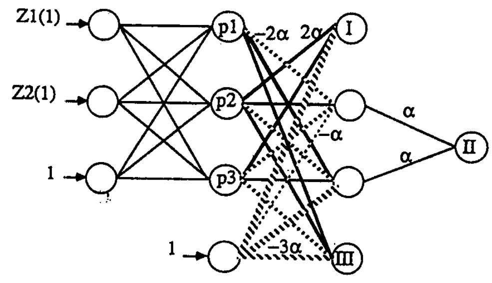

図３　図２のパターンを分離するネット

4.おわりに

パタ-ン認識に対する多層ニュ-ラルネットの飽和特性は必須であること示した。またパタ-ンが超平面により単連結領域に分離できるときは3層ニューラルネットで構成できることを明らかにした。これに基いた学習の高速化[4]に関しては別報に譲りたい。

参考文献

1. D.E.Rumelhart et al,"Parallel Distributed Processing",Vol. 1,2,MIT Press, Cambridge,Mass
2. K.Funahashi,"On the Approximate Realization of Continuous Mapping by Neural Networks," Neural Networks, Vol. 2, No. 3, pp183-192,1989
3. D.L.Chester,"Why Two Hidden Layers are Better Than One," Proc. IJCNN-90-WASH-DC,pp I-265 - I-268,January 1990
4. 阿部重夫、"多層ニューラルネットの並列順伝播学習法、" 人工知能基礎論研究会(SIG-FAI),1990年6月