

ニューラルネットワークの判別機能

吉永耕介・日置咲夫

1. まえがき

ニューラルネットワーク(教師付き学習法)は人間の脳で行われている情報処理の仕組みを模した情報処理ネットワークであり、その構造は細胞を模したユニットと呼ばれる素子とこれらの素子を結合するシナプスを模した結合係数とで構成される単純な構造のネットワークである。

ニューラルネットワーク(教師付き学習法)には、構造が単純ではあるが、排他的論理和のような非線形問題を、教師信号と学習信号を与えるだけで、処理手順を明示したプログラムなしで、解けるという利点がある。これらの利点を生かし、患者の種々の症状から病気を判断する医療診断システム、障害物を回避しながら目的地に到達するロボットシステムなどにニューラルネットワーク技術が応用されるようになってきている⁽¹⁾。

しかしながら、ニューラルネットワーク(教師付き学習法)を用いて情報を処理するとき、ネットワーク内の情報の流れが不明確であるという本質的な問題点がニューラルネットワークに存在する。そのために、情報を処理するのに必要な最小構造のニューラルネットワークの構築は試行錯誤によって行われてきた。これらの問題点を解決する手法として学習と忘却を繰り返すことにより学習がより強固となるという Hebb の学習則に基づく忘却過程⁽²⁾⁽⁴⁾が提案された。

提案された忘却過程では、予め一定の忘却量(以下、消去量という)を定めおかねばならないが、その一定の消去量を経験的に定めることは至難であるという難点がある。

本報告では、既存の忘却過程の難点を解決する適応型忘却過程を新しく提案し、その有効性を確認した。

提案した忘却過程の有効性を確認するため、エクセルのマクロ命令を用いてニューラルネットワーク(教師付き学習法)を構築して、教師・学習信号として排他的論理を求める問題を用いて実験を行った。その結果、提案した忘却過程はニューラルネットワーク(教師付き学習法)内の情報の流れを明確化するのに有効であることが明らかとなった。また、GRE得点を用い、卒業、落第の判別を提案した手法を適用したニューラルネットワークと良く知られている判別分析とで行い、ニューラルネットワーク(教師付き学習法)の判別機能を判別分析と比較評価した。その結果、提案した手法を適用したニューラルネットワーク(教師付き学習法)の方が判別分析より、僅かではあるが、優れていることが明らかとなったので報告する。

2. ニューラルネットワーク(教師付き学習法)の構築と問題点の提示

2.1 エクセルで構築したニューラルネットワーク(教師付き学習法)の概要

エクセルのマクロ命令で構築したニューラルネットワーク(教師付き学習法)⁽¹⁾(以下、ニューラルネットワークという)の操作画面を次図に示す。

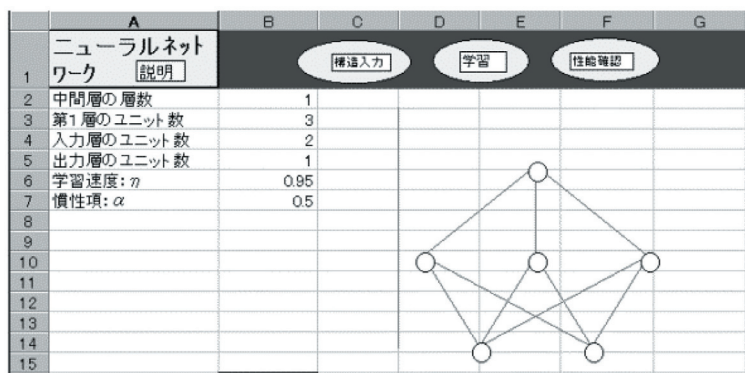


図1 ニューラルネットワーク(教師付き学習法)の操作画面

ニューラルネットワークの判別機能

図1の操作画面で「構造入力」、「学習」、「性能確認」は次のように動作する。

「構造入力」

構造入力をクリックし、ニューラルネットワークの構造を決定する。決定した構造が描画される。

「学習」

学習をクリックし、シート上に作成した教師・学習信号を指定すると、「構造入力」で決定したニューラルネットワークが学習を開始する。学習の過程が学習曲線で表される。

忘却過程を導入すると、結合係数が消去される状況が数値で表示される。

「性能確認」

性能確認をクリックし、シート上に作成した入力信号を指定すると、学習済みのニューラルネットワークが動作し、入力信号に対応した出力信号を出力する。

2.2 教師・学習信号

ニューラルネットワークの機能を確認すると共に、その問題点を提示するために表1に示す排他的論理和の変数 A、B を学習信号、A と B の排他的論理和 $A \text{ XOR } B$ を教師信号として用いる。

学習信号		教師信号
A	B	$A \text{ XOR } B$
1	1	0
1	0	1
0	1	1
0	0	0

表1 排他的論理和

2.3 ニューラルネットワークの構築

ニューラルネットワークの構造を入力層、中間層、出力層からなる 3 層構造とした。その構造を図 2 に示す。この構造をもつニューラルネットワークに表 1 のデータを適用し、その機能を確認する。

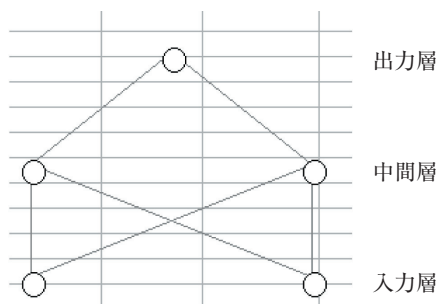


図 2 ニューラルネットワークの構造

図 2 の○が神経細胞の機能を模した多入力 1 出力の素子を表し、この素子はユニットと呼ばれている。○と○を結ぶ線は、神経細胞のシナプスに相当し、その線維が太ければ太いほど層から層へ伝達される情報量は大きくなる。シナプスの線維を三次元配列 W で表し、この W を結合係数という。したがって、結合係数の値が大きければ大きいほど次の層に伝わる情報量は多くなり、結合係数の値が小さければ小さいほど次の層に伝達される情報量は少なくなる。つまり、結合係数の値の大小が出力信号への寄与の度合いを表す。

図 2 に示したニューラルネットワークの構造は、次のようにして、決定した。

入力層のユニット数：

学習信号は変数 A 、 B の 2 つであるからユニット数を 2 とする。

出力層のユニット数：

排他的論理和の結果は 1 か 0 かのいずれかであることからユニット数を 1

とする。

中間層のユニット数：

中間層のユニット数が必要最小限の数であれば学習が収束し所期の機能をニューラルネットワークは発揮するが、1つでも少ないと教師信号と出力信号との差が小さくならないことがある。また、多すぎると過学習の可能性が出てくる。このため、最小のユニット数を求めるには、初期値として入力層のユニット数と同数を与え、逐次、増やしてゆく方法は、原始的ではあるが、確実な方法であると経験的に言われている⁽¹⁾⁽²⁾ことから、中間層のユニット数を入力層のユニット数と同数の2に選定した。

結合係数 W を用いて図2を表すと次図となる。

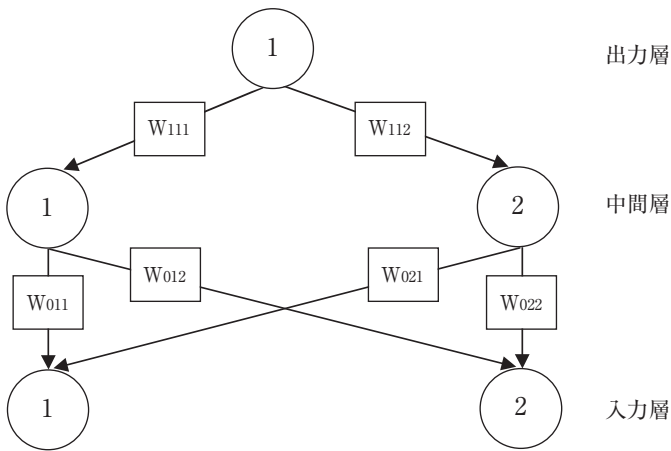


図3 結合係数を用いて表したニューラルネットワーク

図3で、結合係数 W_{011} の最初の添え字0は入力層を、次の添え字11は中間層の1番目のユニットと入力層の1番目のユニットとの結合を意味する。同様にして、 W_{021} は中間層の2番目のユニットと入力層の1番目のユニット

との結合を意味している。

また、 W_{111} の最初の添え字 1 は中間層を、次の添え字 11 は出力層の 1 番目のユニットと中間層の 1 番目のユニットとの結合を、 W_{112} は出力層の 1 番目のユニットと中間層の 2 番目のユニットを結合する係数を意味する。

ニューラルネットワークの各層の結合係数には、初期値として、人為的に発生した 0 から 1 の範囲の乱数が不規則に割り当てられている。

2.4 ニューラルネットワークの学習

初期状態のニューラルネットワークに教師信号と学習信号が入力されると、入力層のユニットは、入力された学習信号に処理を加えることなく、中間層へ伝達する。中間層のユニットに伝達される情報量は、入力層と中間層を結合する個々の結合係数の大きさに依存する。

中間層のユニットは、入力層のユニットから伝達された情報の総和が、ある閾値を超えたとき 1 に近い値を出力し、総和が閾値以下であるときは 0 に近い値を出力する。この中間層におけるユニットの働きは、細胞体の電位がある閾値を超えると伝導性のインパルスが発生する脳細胞の膜電位加重とインパルス⁽¹⁾との関係に相当し、ニューロンの動作関数と呼ばれている。動作関数には、図 4 に示すシグモイド関数が用いられる。

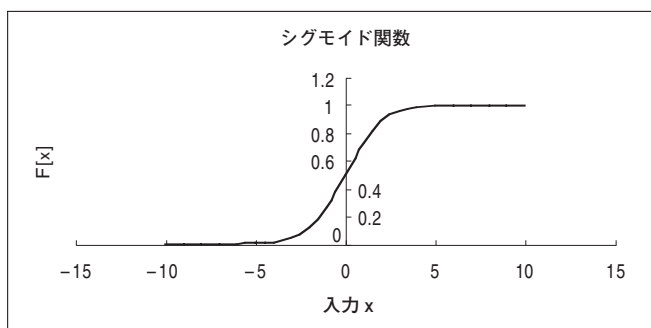


図 4 シグモイド関数

ニューラルネットワークの判別機能

このようにユニット自身は、入力された情報を中間層へそのまま伝えるか、または、入力された情報の総和と閾値をシグモイド関数で処理し、その結果得られた値を出力するだけの働きをする⁽²⁾。

出力層の出力信号と教師信号との差、誤差、を最小にする情報はユニット間に介在する結合係数の値の大小として蓄えられる。出力層からの出力信号と教師信号との差が最小になるようにユニット間の結合係数が、逐次、更新される過程をニューラルネットワークの学習という。

2.5 学習の結果

(1) 学習曲線

出力層から出力された出力信号と教師信号との差の大きさは学習曲線として与えられる。

表1のデータを用いて、学習を10000繰り返したときに得られた学習曲線を図5に示す。

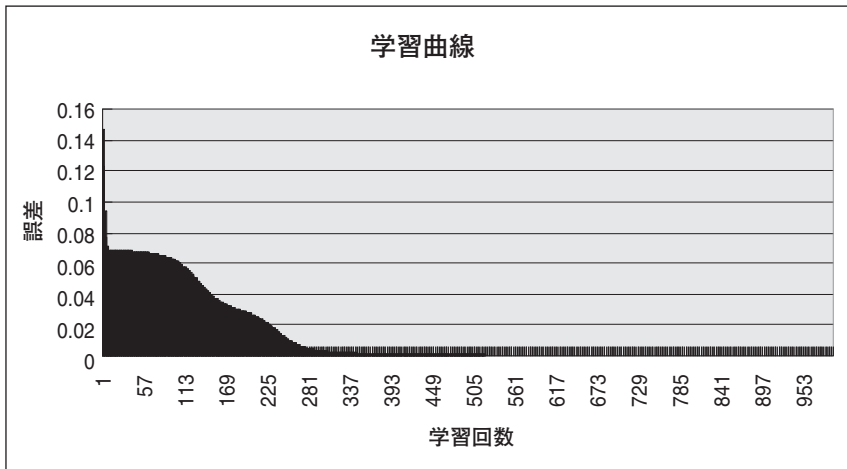


図5 学習曲線

図5から出力信号と教師信号との誤差は、学習開始当初は、大きな値であるが学習が進行するにつれて減少し学習回数280回程度で、0.02以下の値に抑えられることが分る。このようにある一定の値に誤差が収束することを学習の収束という。

なお、以下で示す図表の学習回数は、特に断らない限り、10000回である。

(2) 学習済み重み係数の値

教師信号と出力信号との誤差が最小になるように学習を繰り返して更新された結合係数(以下、学習済み結合係数という)を用いて図2のニューラルネットワークを表すと次図となる。

なお、学習済み結合係数の値は初期値として与えられた乱数値によって異なった値となる。すなわち、同一教師・学習信号を用いた実験であっても実験を繰り返すごとに発生する乱数の値が異なるので結合係数の初期値が異なる。このため、実験ごとに学習済み決定係数の値は異なる。しかしながら、学習済み決定係数の値が異なっても、学習済み結合係数に基づくニューラルネットワークの出力信号に差異はない。

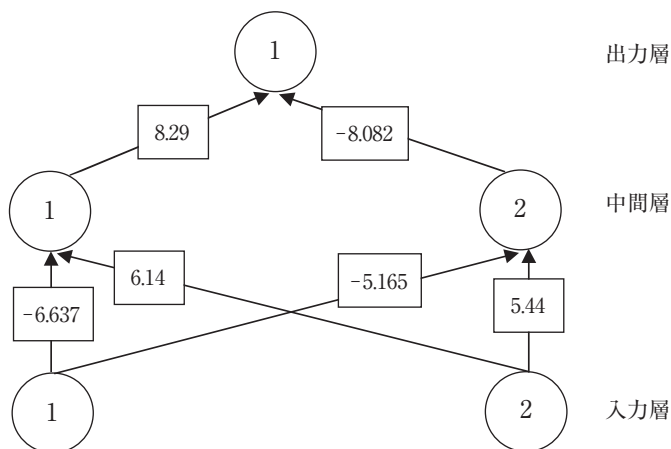


図6 学習済み重み係数の値の例

ニューラルネットワークの判別機能

図 6 から学習済み結合係数は、正或いは負の値となっており、その絶対値は、何れも、1 より大きいことが分る。これらの結合係数に排他的論理和を求める情報が蓄えられていることになる。

(3) 機能確認

図 6 に示した学習済みのニューラルネットワークの入力層に表 1 の変数 A、B を入力信号として与えたときに得られた出力信号を表 2 に示す。なお、表 2 では出力信号と教師信号との残差平方和、赤池の情報量⁽³⁾を計算するため教師信号を併せて示している。

教師信号	出力信号	残差
0	0.028392847	-0.02839
1	0.963239095	0.036761
1	0.970998949	0.029001
0	0.031569909	-0.03157
	残差平方和	0.003995
	赤池の情報量	-8.28428

表 2 出力層からの出力信号

表 2 から明らかなようにニューラルネットワークで得られる出力信号の精度は、だいたい正確であるという程度であり、ピッタリと 1 あるいは 0 とはならない。しかしながら、教師信号 1 に対応する出力信号は 0.963 と 0.970 であり 1 を表しているとみなすことができる。また、教師信号 0 に対応する出力信号は 0.028 と 0.031 であり 1 に比べて小さく 0 とみなすことできる。したがって、表 2 の出力信号は排他的論理和を表しているとみなされる。したがって、図 2 に示したニューラルネットワークは、排他的論理和を演算する能力を、学習によって、獲得したことになる。

また、残差平方和(=0.0039)、赤池の情報量(=-8.28)は共に小さな値であり、

学習済みニューラルネットワークの出力信号は、統計的に、妥当であると考えられる。

ニューラルネットワークの学習則は次のように考えられている⁽¹⁾。学習信号が与えられるとニューラルネットワークは、与えられた学習信号をパターンとしてとらえ、そのパターンに対応する出力信号の値が教師信号の値に近づくように、逐次、誤差逆伝播法に基づき結合係数を更新し、その値を各結合係数に分散して蓄える。

排他的論理和の場合、学習によって4つの入力パターン(1,1)、(1,0)、(0,1)、(0,0)に関する情報と、そのパターンに対応する出力信号0,1,1,0に関する情報を6個の結合係数に分散して蓄える。そして、学習済みニューラルネットワークに入力信号が与えられると、ニューラルネットワークは蓄えているパターンの中から入力信号に対応したパターンを判別し、そのパターンに対応する出力信号を出力することになる。

3. 問題の提示

図6に示す学習済みのニューラルネットワークの入力層に、例えば、入力信号として1と1を与えたとき、出力層からの出力信号が0.032となる過程を入力層から出力層へ結合係数を順にたどって確認することは不可能である。すなわち、個々の結合係数が出力に寄与する度合いを明らかにすることが出来ないという問題点が、本質的に、ニューラルネットワークに存在する。このために、図2に示したニューラルネットワークの構造が排他的論理和を学習するのに必要な最小構造であるか否かを判断することができない。

ニューラルネットワーク内での情報の流れが不明確であるという問題が生ずる理由は出力信号と教師信号との誤差を最小にする情報が、個々の結合係数に、分散して蓄えられることにあると考えられている⁽²⁾。

この問題点を解決する試みの一つとして、入力層に余分なユニットを設け

ニューラルネットワークの判別機能

て学習と無関係な余分な学習信号をニューラルネットワークに与えて学習させることを試みる。

試みの結果、もし、学習と無関係な余分なユニットに係わる結合係数の値が、他の結合係数の値より小さな値となり、かつ、学習済み結合係数に基づく出力信号が排他的論理を表すのであれば、余分な学習信号に関する情報を担う結合係数と排他的論理の学習に関する情報を担う結合係数とを区分できるはずである。その結果、ニューラルネットワークの構造が排他的論理和を学習するのに必要な最小構造であるか否かを判断することができるはずである。

このように考えて、表1のデータに学習とは無関係な一定の値0.1をもつ変数Cを加え学習信号とする。そのデータを表3に示す。

学習信号			教師信号
A	C	B	A XOR B
1	0.1	1	0
1	0.1	0	1
0	0.1	1	1
0	0.1	0	0

表3 学習と無関係な変数Cを加えた教師・学習信号

表3のデータを用いてニューラルネットワークに学習させ、その結果得られた学習済み結合係数を図7に、学習曲線を図8に、出力信号を表4に示す。

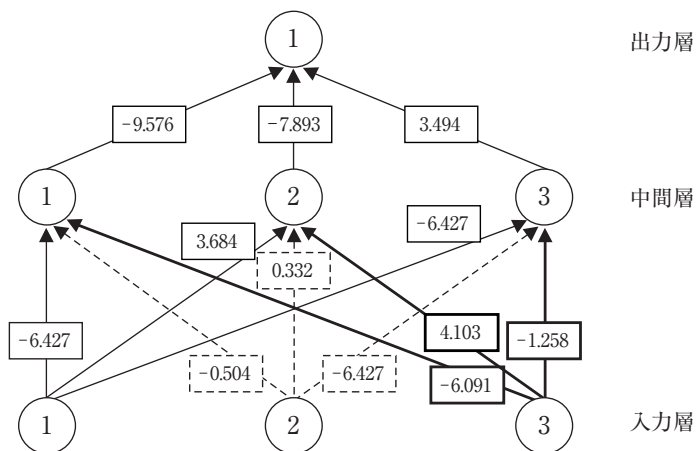


図7 学習と無関係な余分な学習信号を加えたときの学習済みネットワーク

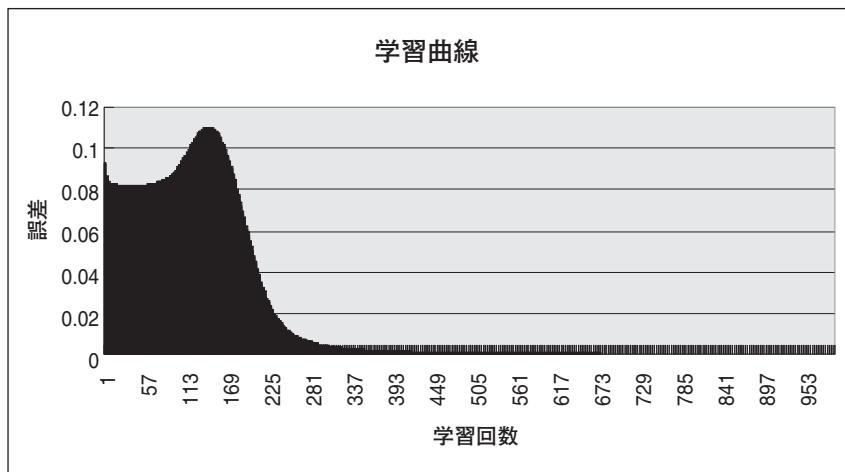


図8 学習と無関係な余分な学習信号を加えたときの学習曲線

教師信号	出力信号	残差
0	0.030536841	-0.03054
1	0.969798404	0.030202
1	0.971316745	0.028683
0	0.027592805	-0.02759
残差平方和		0.003429
赤池の情報量		-6.89593

表 4 学習と無関係な余分な学習信号を加えたときの出力信号

図 8、表 4 から次のことが明らかとなった。

学習と無関係な余分な学習信号をニューラルネットワークに与えて学習させた場合であっても、337 回程度の学習で出力信号と教師信号との誤差は 0.02 以下に抑えられる (図 8)。

学習済み結合係数に基づく出力信号は排他的論理和であるとみなすことができる。そして、その値は、残差平方和(=0.0034)、赤池の情報量⁽³⁾(=-6.895)は共に小さく、統計的に、妥当であると考えられる(表 4)。

図 7 おいて、入力層の 2 番目のユニットが学習と無関係な余分な学習信号が与えられたユニットであり、そのユニットに係わる学習済み結合係数を点線で囲って表している。その値は、他の学習済み結合係数と比較して無視できるほど小さな値とはなっていない。すなわち、学習と無関係な余分な学習信号与えて学習させ、ニューラルネットワーク内の情報の流れを明確化しようとする試みは失敗した。その理由は、学習と無関係な余分な学習信号に関する情報と出力信号と教師信号との誤差を最小にする情報は、共に、各結合係数に分散して蓄積されるからであると考えられることができる。

4. 適応型忘却過程の提案

情報が分散されて各結合係数に蓄積されるのを抑え、ニューラルネットワーク内の情報の流れを明確化する手法として忘却過程の導入が提案されている⁽²⁾⁽⁴⁾。文献(2)(4)で提案されている忘却過程は、一定の間隔で、経験的に予め定めた一定の値を結合係数から消去する手法である。その考え方は、学習と消去とが繰り返されると、Hebbの学習則⁽²⁾により、幾つかの結合係数は消去より学習の効果が強く反映して結合がより強固になる。その一方で、学習より消去の効果が強く働き結合はより弱くなりほとんどゼロとなる結合係数が存在するようになる。その結果、情報の流れが明確化されるという考えに基づいている。文献(2)、(4)で提案されている忘却過程の問題点は、消去量を一定の値に経験的に予め決定しておく必要があることである。

消去量が大きすぎると情報を担うべき結合係数が消去され学習が収束しないおそれがあり、また、学習が収束する場合であっても、収束するまでに時間がかかる。逆に、小さすぎると結合係数間に情報が分散されて蓄積され情報の流れが明確化されない。このため、消去量を、経験的に、一定の値に定めることは至難であるという難点がある。そこで、学習の進行につれ変化する結合係数の値に適応して消去量が定められる適応型忘却過程を次のように提案する。

① 忘却回数の設定。

例えば、学習回数が100の倍数(忘却回数という)になったら消去を行うというように消去を実行する学習回数を予め定めておく。

② ニューロンネットワークに学習を開始させる。学習回数をカウントし忘却回数になったら学習を中断し、次の③、④、⑤を実行する。

③ 中間層、出力層ごとにその結合係数の最大値 W_{\max} を求める。

④ 中間層、出力層ごとに消去量 ε を次式により定める。

$$\varepsilon = \begin{cases} |W_{\max}|/100 & (|W_{\max}| \geq 1) \\ |W_{\max}|/10 & (|W_{\max}| < 1) \end{cases}$$

- ⑤ 中間層、出力層ごとに結合係数 W_{ij} の一部を次式により消去する。

$$W_{ij} = \begin{cases} W_{ij} - \varepsilon & (W_{ij} > 0) \\ W_{ij} + \varepsilon & (W_{ij} < 0) \end{cases}$$

- ⑥ ②に戻る。

提案した忘却過程は、学習の進行につれて変化する結合係数の値に適応して消去量が定められ、過大な値となることはないという特徴がある。

忘却回数を100とし提案した忘却過程を適用して図7に示したニューラルネットワークに表3の教師・学習信号を与え学習させて得られた学習済み結合係数を図9に、学習曲線を図10に示す。

なお、学習回数を10000回としているので、学習が終了するまでに、提案した忘却過程を100回繰り返して実行したことになる。

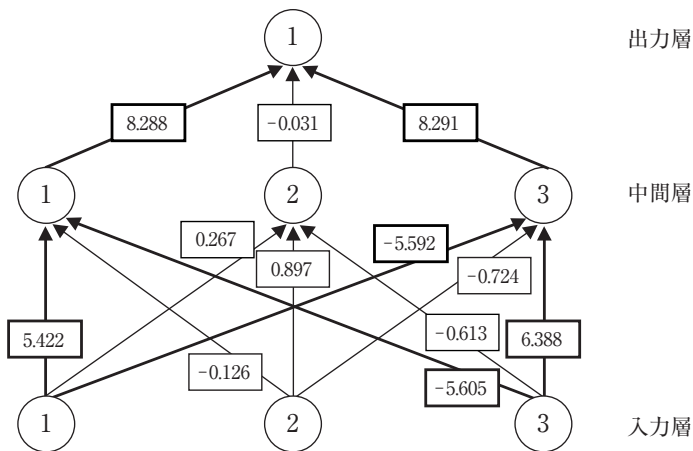


図9 学習信号に学習と無関係な余分な学習信号を加えたときの学習済み結合係数

図 9 では、1 以上の値の結合係数を太線の枠と線で、1 以下の結合係数を細線の枠と線で示している。

図 7 と図 9 を比較して次のように結論される。図 9 では、図 7 とは異なり、学習と無関係な余分な学習信号が与えられた入力層の 2 番目のユニットに係わる結合係数は 1 以下の小さな値であり、他の結合係数に比べ、出力信号に寄与する度合いは小さいと判断できる。このために、入力層から出力層への情報の流れを、出力信号への寄与が小さいとみなされる結合係数を迂回して、辿ることができる。すなわち、提案した忘却過程により情報の流れが明確化されたと結論できる。

図 6 と図 9 とを比較して、次のように結論される。図 9 では、他の結合係数と比較して無視しえるほど小さな値を持つ余分な結合係数が存在するが、図 6 のニューラルネットワークでは、他の結合係数と比較して無視しえるほど小さな値を持つ結合係数は存在しない。このことから、図 6 は排他的論理和を学習するのに必要な最小構造のニューラルネットワークであると考えられる。

なお、最小構造のニューラルネットワークに忘却過程を適用して学習させても、特定の結合係数が、他の結合係数に比べ、無視しえるほど小さな値になることはない。

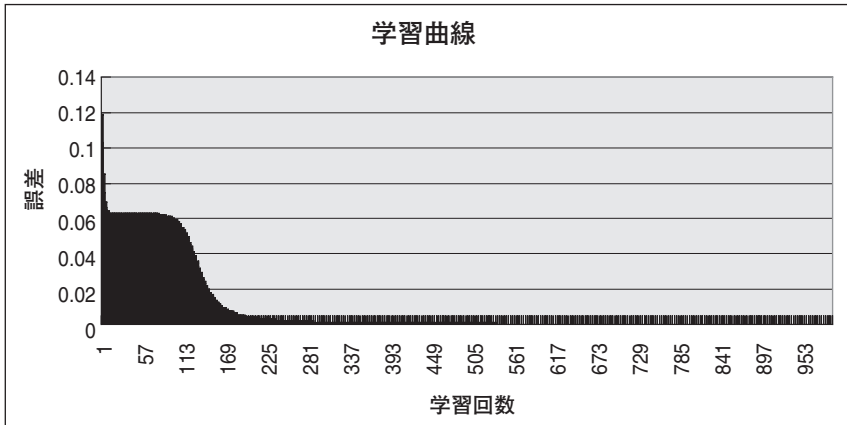


図 10 忘却過程を適用したときの学習曲線

図 8 と図 10 を比較すると、両図とも、学習回数 225 回位で、出力信号と教師信号との誤差は 0.02 以下に抑えられており、忘却過程の導入によってニューラルネットワークの学習は阻害されることはない、すなわち、提案した忘却過程での消去量は学習の収束を阻害する程過大な値となることはないと結論できる。

図 9 に示した学習済みニューラルネットワークに表 3 の学習信号を入力信号として与えて得られた出力信号の値を表 5 に示す。

教師信号	出力信号	残差
0	0.030845163	-0.03085
1	0.970793938	0.029206
1	0.970758028	0.029242
0	0.034946593	-0.03495
残差平方和		0.003881
赤池の情報量		-6.40055

表 5 忘却過程を適用したときの出力信号

表 4 と表 5 を比較すると、表 5 の残差平方和、赤池の情報量は、共に、表 4 の値より僅かではあるが大きくなっている。その理由として、結合係数が

担うべき出力信号と教師信号との誤差を最小にする情報が忘却過程の導入により、僅かではあるが、消去される。その結果、表 5 の残差平方和と赤池の情報量が、表 4 の値より、大きくなったと考えられる。

5. 判別分析との比較

口頭および筆記 GRE 得点を用いて、ニューラルネットワークの判別機能をよく知られている判別分析の判別機能と比較して評価する。

5.1 判別分析による判別

(1) 使用するデータ

文献 (5) に示されている口頭および筆記 GRE 得点 (表 6) データとして使用する。

連番	卒業(第1群)		落第(第2群)	
	口頭	筆記	口頭	筆記
1	750	590	740	680
2	360	600	670	600
3	720	750	560	550
4	540	710	540	520
5	570	700	590	540
6	520	670	590	700
7	590	790	470	600
8	670	700	560	540
9	620	730	540	630
10	690	840	500	600
11	610	680		
12	550	730		
13	590	750		
平均点	598.46	710.76	576	596

表 6 口頭および筆記 GRE 得点

ニューラルネットワークの判別機能

なお、表6のデータをシェフェの方法⁽⁶⁾で検定すると次の結果が得られる。

帰無仮説：「卒業（口頭）と落第（口頭）の母平均は等しい」を有意水準5%でシェフェの方法で検定すると

シェフェの統計量：0.669077 < 棄却値：2.912242

となり、帰無仮説は受容される。

帰無仮説：「卒業（筆記）と落第（筆記）の母平均は等しい」を有意水準5%でシェフェの方法で検定すると

シェフェの統計量：3.41871 > 棄却値：2.912242

となり、帰無仮説は棄却される。

この検定の結果、卒業（口頭）と落第（口頭）の母平均に有意差は有意水準5%で認められないことから卒業、落第の判定は筆記の得点によってなされたと考えられる。

(2) 判別分析の結果

表6に判別分析⁽⁵⁾を適用し得られた結果を表7に示す。

表7の卒業（第1群）で落第（第2群）と誤判別されているのは筆記の得点が平均点710.76より低い590と600で生じており、落第（第2群）で卒業（第1群）に誤判別されているのは筆記の得点が平均点596より高い680と700で生じている。この判別結果とシェフェの方法による母平均の検定結果とから落第（第2群）の連番6は口頭で卒業（第1群）の平均点598.46に近い590を得点し、筆記では卒業（第1群）の平均点710.76に近い700を得点していることから卒業（第1群）に判定すべきであった。

また、卒業（第1群）の連番1は口頭で平均点598.46より高い750点を得点

しているが、筆記では平均点 710.76 より低い 590 を得点している。口頭の母平均は卒業(第 1 群)と落第(第 2 群)とで有意差はないと考えられることから筆記の得点が低い卒業(第 1 群)の連番 1 は落第(第 2 群)に判定すべきであったと考えられる。

連番	卒業(第1群)		落第(第2群)	
	判別関数の推定量	判別	判別関数の推定量	判別
1	-2.536	第2群	0.110	第1群
2	-0.562	第2群	-1.901	第2群
3	2.220	第1群	-2.872	第2群
4	1.841	第1群	-3.654	第2群
5	1.422	第1群	-3.291	第2群
6	0.771	第1群	1.336	第1群
7	3.939	第1群	-1.038	第2群
8	0.991	第1群	-3.162	第2群
9	2.074	第1群	-0.472	第2群
10	4.953	第1群	-1.167	第2群
11	0.671	第1群		
12	2.376	第1群		
13	2.782	第1群		

表 7 表 6 の判別分析の結果

5.2 ニューラルネットワークによる判別

(1) 使用するデータ

ニューラルネットワークで扱うことのできる数値は 0 から 1 の範囲の値であることから卒業を 1、落第を 0 で表し、そして口頭、筆記試験の得点をそれぞれの最大値で規格化して表した表 8 のデータを教師・学習信号として用いた。

ニューラルネットワークの判別機能

教師信号	学習信号	
	口頭	筆記
1	1	0.702381
1	0.48	0.714286
1	0.96	0.892857
1	0.72	0.845238
1	0.76	0.833333
1	0.693333	0.797619
1	0.786667	0.940476
1	0.893333	0.833333
1	0.826667	0.869048
1	0.92	1
1	0.813333	0.809524
1	0.733333	0.869048
1	0.786667	0.892857
0	0.986667	0.809524
0	0.893333	0.714286
0	0.746667	0.654762
0	0.72	0.619048
0	0.786667	0.642857
0	0.786667	0.833333
0	0.626667	0.714286
0	0.746667	0.642857
0	0.72	0.75
0	0.666667	0.714286

表 8 教師・学習信号

(2) 学習の結果

入力層、中間層、出力層の3層構造とし、それぞれの層のユニット数を2、5、1としてニューラルネットワークを構成し、表8の教師・学習信号を与えて、学習回数50000、忘却回数100として学習させた。その結果得られた学習済み結合係数を表9に示す。

なお、学習回数を50000回としても、学習に要する時間は20秒程である。

		W_{111}	W_{112}	W_{113}	W_{114}	W_{115}			
		-12.120	3.203	-7.457	-7.838	-1.933			
W_{011}	W_{012}	W_{021}	W_{022}	W_{031}	W_{032}	W_{041}	W_{042}	W_{051}	W_{052}
7.927	-27.966	-6.605	2.886	5.345	-17.087	5.369	-17.883	-0.048	-3.174

表 9 学習済み結合係数

表 9 に基づき、結合係数の絶対値が 5 以上であれば太線で、それ以下であれば細線でニューラルネットワークを表すと、図 10 となる。

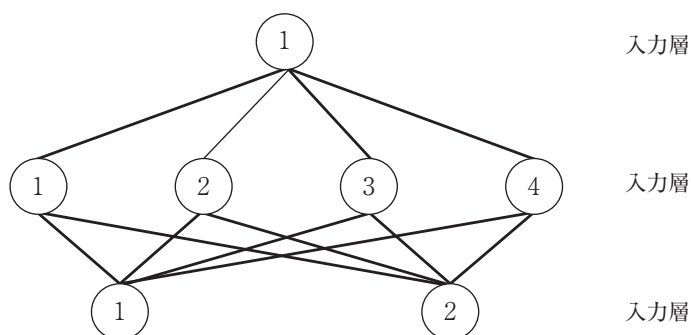


図 11 GRE 得点を学習したニューラルネットワーク

図 11 から、中間層の 2 番目のユニットと出力層とを結ぶ結合係数の値は 3.203 であり、その値は他の結合係数の値に比べて小さく、出力への寄与の度合いは小さい。したがって、GRE 得点を学習し、卒業、落第を判別するのに必要なニューラルネットワークの最小構造は中間層のユニット数を 3 として構成できることが分る。

図 11 のニューラルネットワークに表 8 の学習信号を入力信号として与えたときの出力信号を表 10 に示した。

出力信号	残差
4.99687E-08	1
0.876737377	0.123263
0.989909658	0.01009
0.996995571	0.003004
0.990262949	0.009737
0.972317938	0.027682
0.999300263	0.0007
0.786792748	0.213207
0.995562111	0.004438
0.999342333	0.000658
0.783701959	0.216298
0.998385124	0.001615
0.998606507	0.001393
0.003820865	-0.00382
1.01604E-06	-1E-06
2.99106E-07	-3E-07
5.48741E-08	-5.5E-08
6.52276E-08	-6.5E-08
0.983747018	-0.98375
0.019742439	-0.01974
1.3333E-07	-1.3E-07
0.058061828	-0.05806
0.003528123	-0.00353
残差平方和	2.079979
赤池の情報量	17.99904

表 10 出力信号

5.3 比較・評価

表 10 において、0.7 以上の値の出力信号を第 1 群(卒業)に、それ以下の値の出力信号を第 2 群(落第)に置き換え、表 7 と表 10 をまとめて表すと表 11 となる。

連番	卒業(第1群)		落第(第2群)	
	判別分析	ニューラルネットワーク	判別分析	ニューラルネットワーク
1	第2群	第2群	第1群	第2群
2	第2群	第1群	第2群	第2群
3	第1群	第1群	第2群	第2群
4	第1群	第1群	第2群	第2群
5	第1群	第1群	第2群	第2群
6	第1群	第1群	第1群	第1群
7	第1群	第1群	第2群	第2群
8	第1群	第1群	第2群	第2群
9	第1群	第1群	第2群	第2群
10	第1群	第1群	第2群	第2群
11	第1群	第1群		
12	第1群	第1群		
13	第1群	第1群		
誤判別数	2	1	2	1

表 11 ニューラルネットワークと判別分析との比較

表 11 から次のことが分る。誤判別数を比較すると、ニューラルネットワークでの誤判数が 1 個少ない。したがって、ニューラルネットワークの方が判別分析より、僅かではあるが、優れていると考えられる。

なお、落第(第 2 群)の連番 6 は、判別分析とニューラルネットワークのいずれでも卒業(第 1 群)に判別されている。また、卒業(第 1 群)の連番 1 は、落第(第 2 群)と判別されている。これらの結果は、シェフェの方法による検定結果と符合し合理である。

また、判別分析にはマハラノビスの距離を用いているので、データの非線

ニューラルネットワークの判別機能

形性が加味されて判別が行われている。

5.4 GRE 得点を学習するのに必要な最小構造ニューラルネットワークの学習結果

GRE 得点を学習するのに必要なニューラルネットワークの最小構造は中間層のユニット数を 3 として構成できることが分ったので、ニューラルネットワークの構造を 3 層構造とし入力層、中間層、出力層のユニット数をそれぞれ 2,3,1 とし、表 8 の教師・学習信号を用いてニューラルネットワークに学習させた。その結果得られた学習済み結合係数を表 12 に示した。

	W_{111}	W_{112}	W_{113}		
	-10.254	4.592	-12.439		
W_{011}	W_{012}	W_{021}	W_{022}	W_{031}	W_{032}
6.497	-23.683	-5.626	9.129	7.852	-28.327

表 12 学習済み結合係数

表 12 では、表 9 とは異なり、1 以下の値を持つ結合係数はみあたらない。

表 12 に示した学習済み結合係数をもつニューラルネットワークに表 8 に示した学習信号を入力信号として与えると、次表に示す出力信号が得られた。

出力信号	残差
2.96186E-07	1
0.924732104	0.075268
0.996622827	0.003377
0.999466149	0.000534
0.997575096	0.002425
0.991451774	0.008548
0.999918645	8.14E-05
0.881721372	0.118279
0.998986129	0.001014
0.999909979	9E-05
0.880808486	0.119192
0.999755217	0.000245
0.999781338	0.000219
0.004232639	-0.00423
2.023E-06	-2E-06
9.25434E-07	-9.3E-07
3.64316E-07	-3.6E-07
3.76882E-07	-3.8E-07
0.995321507	-0.99532
0.016968422	-0.01697
5.73189E-07	-5.7E-07
0.061605549	-0.06161
0.002820034	-0.00282
残差平方和	2.028727
赤池の情報量	17.4252

表13 出力信号

表13の教師信号に対応する出力信号は、表8の出力信号と比べ、より1に近い値となっている。また、教師信号0に対応する値はより0に近い値となっている。また、表13の残差平方和の値を表10の値と比較すると、表10では2.0799、表13では2.0287となっており、僅かではあるが、表13の値が小さくなっている。

ニューラルネットワークの判別機能

以上の結果から入力層、中間層、出力層のユニット数をそれぞれ 2,3,1 とした 3 層構造のニューラルネットワークは、表 6 に示した GRE 得点を学習し、卒業、落第を判別するのに必要な最小構造であると考えられる。

(まとめ)

ニューラルネットワークにはネットワーク内の情報の流れが不明確であるという本質的な問題点がある。この問題を解決する手法として、新しく適応型忘却過程を提案した。そして、提案した手法の有効性を確認するため、ニューラルネットワークに提案した忘却過程を適用し排他的論理和問題を学習させた。その結果、ネットワーク内の情報の流れが明確となり、提案した手法が有効性であることが明らかとなった。

また、GRE 得点を用い、卒業、落第の判別を提案した手法を適用したニューラルネットワークとマハラノビスの距離を用いた判別分析とで行った。その結果、ニューラルネットワークでの誤判別数が、1 個だけ、少なく、提案した手法を適用したニューラルネットワークの方がマハラノビスの距離を用いた判別分析より、僅かではあるが、優れていることが明らかとなった。

なお、適応型忘却過程の着想は 1998 年 3 月に行った。

(文献)

- (1) 八名、鈴木：“ニューロ情報処理技術”、海文堂出版 (1992 年)
- (2) 市川 紘：“階層型ニューラルネットワーク”、共立出版 (1993 年)
- (3) 田中、垂水、脇本：“パソコン統計解析ハンドブックⅡ”、共立出版 (1992 年)
- (4) 石川 真澄：“忘却を用いたコネクションモデルの構造学習アルゴリズム”、人工知能学会、Vol.5, No.5, pp595-602 (1990 年)
- (5) 中村 慶一訳：“応用多変量解析”、森北出版 (1976 年)
- (6) 豊田 秀樹：“違いを見ぬく統計学”、講談社 (1996 年)