

移行学習のニューラルネットワーク・シミュレーションに
おける内部構造の分析

松尾 太加志 木藤 恒夫*

An Analysis of the Internal Structure of a Neural Network
Simulation of Shift Learning

Takashi MATSUO and Tsuneo KITO*

In an artificial neural network model, the internal structure of the network (weights and thresholds) represents the internal state of learning. When shift learning is simulated by the neural network, the initial state of the internal structure is a major factor determining the formation of a *mediational process*. In simulation 1, four initial state values were distributed at random. The results showed that the networks which performed best at reversal shift learning were those which had a symmetrical structure—"symmetry type"—and the worst were those which had a parallel structure—"parallel type"—or an asymmetrical structure—"asymmetry type"—. In simulation 2, the initial values were set systematically. According to the analysis of the performance and of the internal structure, the results were classified into three groups. The internal structure of the third group correlated with that of a *mediational process* system. Thus the connections between the input unit and the hidden unit were shown to influence the formation of a *mediational process*. While agreement with this type of psychological process was seen in the internal structure, the performance did not correspond to the results of the psychological experiment. However, the utility of an analysis of internal structure in research of neural networks was demonstrated.

ニューラルネットワーク・シミュレーションとは、人間の情報処理活動を脳内の神経活動に模した人工的な神経回路モデルによってシミュレーションをするものである。その簡単なモデルは、McCulloch & Pitts (1943) によって提案されたもので、今日利用されている神経回路モデルも、その原形は変わっていない。

* 久留米大学文学部 (Faculty of Literature, Kurume University)

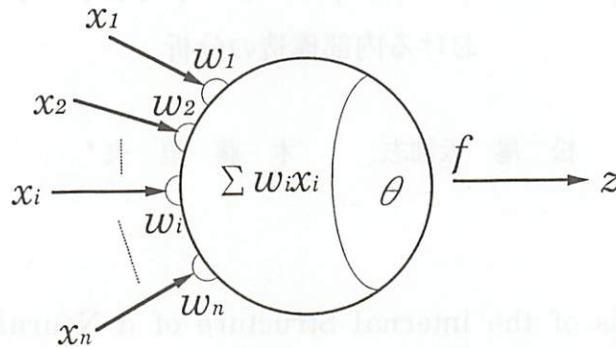


図1 人工的な神経回路モデル。神経細胞を多入力1出力の情報処理素子とみている。

実際の人間の神経回路網では、ひとつの神経細胞が他の複数の神経細胞とシナプス結合している。これら複数の神経細胞の興奮の発火頻度の総和が、ある閾値を越えると、結合を受けた神経細胞が発火する。ある神経細胞が、結合している神経細胞にどの程度影響を及ぼすかは一定ではなく、それぞれ異なった影響度、つまり、シナプス効率を持っている。神経回路モデルでは、神経細胞に相当するユニットとして、多入力1出力の情報処理素子を考える(図1)。このユニットに他のユニット n 個からの入力があり、その i 番目の入力信号を x_i とする。そのときのシナプス効率は、ユニット間の結合の程度として結合係数 w_i で表わす。入力値は信号 x_i とこの w_i を乗じたものとなり、入力の総和は $\sum w_i x_i$ となる。そして、ユニットの閾値を θ とすれば、このユニットの出力 z は、次の式で表現される。

$$z = f(\sum w_i x_i - \theta)$$

f は入出力関数で、階段関数やロジスティック関数などがとられる。階段関数をとった場合は、発火するかしないかを2値的に決める論理素子を実現でき、McCulloch & Pittsのモデルやパーセプトロン(Minsky & Papert, 1969)などで利用されている。バックプロパゲーション法(Rumelhart, Hinton & Williams, 1986)ではロジスティック関数が使われている。

人間の神経回路網では、シナプス結合に可塑性があり、シナプスの効率は変化する。それが人間の記憶や学習に関与していると言われている。神経回路モデルでも、シナプス効率に相当するユニット間の結合係数及び閾値を変化させることによって学習は行われる。言い換えると、神経回路モデルでの結合係数と閾値が人間の脳の状態をシミュレーションしている。したがって、結合係数と閾値の状態(以下、内部構造と呼ぶ)を分析することによって、そのモデルがどのような学習をしているかを知ることができる。

ニューラルネットワーク・シミュレーションを行う場合、学習開始時の内部構造は、ランダムで絶対値の小さな初期値に決められることが多い。しかし、これは人間の学習をシミュレーションする際に、必ずしも適切ではない。ランダムな初期値ということは、まだ何も学習していない状態を表わしている。生まれたばかりの赤ちゃんの学習をシミュレーションする場合であれば、問題はないかもしれないが、一般に

はそうではない場合が多い。とくに、過去に学習している経験や知識によって、学習結果が影響を及ぼすような場合は、学習開始時に初期値をランダムに決めるわけにはいかない。

Kendler & Kendler (1962) の移行学習の実験などはその典型である(図2)。2次元2種の組み合わせで4つの刺激が提示される弁別学習を行う。たとえば、大きさと明るさの次元で、それぞれ、「大」「小」及び「明」「暗」の2種で4つの組み合わせとなる。第1学習では一方の次元の1種類(たとえば、大きさと「大」)が正となる弁別学習を行い、第2学習では2つの移行学習に分かれる。ひとつは、逆転移行で、第1学習で学習した次元のもう1つの種類(「小」)が正となる学習である。この場合の移行は、第1学習と同じ次元内で正反応を逆転(逆転移行)するという次元内移行になる。もうひとつは、第1学習で学習した次元とは異なる次元の1種類(たとえば、明るさで「暗」)が正となる学習である。次元そのものが移行するため、次元間移行、または逆転ではないことから非逆転移行と言われる。

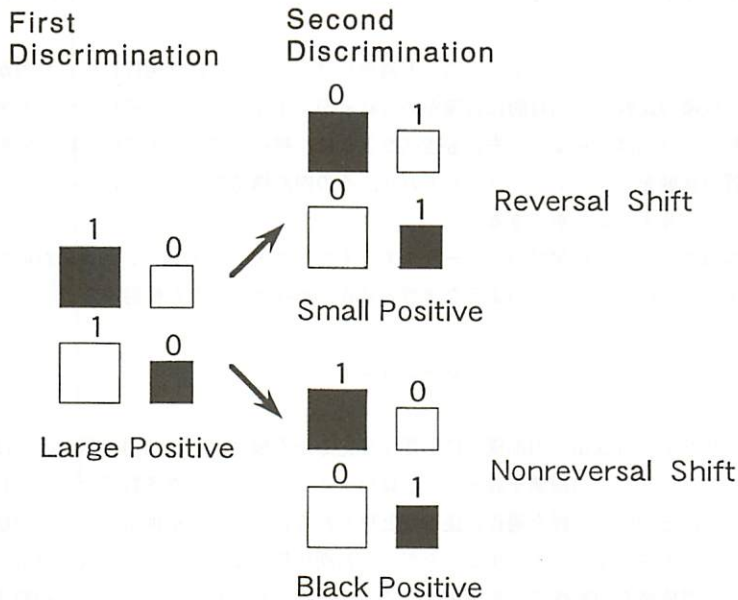


図2 移行学習の心理実験。Kendler & Kendler (1962) を改変。各図形の上の数字(0, 1)は、シミュレーションで、出力させる値に対応している。

このような学習を行う場合、2つの学習のタイプがあると言われている。ひとつは、単一ユニット過程説で、次元とは無関係に4つの刺激のうちどの2つが正であるかということで学習をする。そのため、第1学習から第2学習での移行の際、逆転移行では、2つの刺激とも変えることになるが、非逆転移行では1つはそのままでもうひとつだけを変えればよい。そのため、この単一ユニット過程による学習では非逆転移行での成績がよくなる。もうひとつは、媒介過程説で、大きさという次元を第1学習で学習をし、逆転移行では、それを媒介として学習を行う。したがって、逆転移行では、逆転させるだけでよいため速く学習ができる。しかし、非逆転移行では、その媒介となる次元を変える必要があるため、学習結果は悪くなる。

どちらの過程に従って学習を行うかは、発達によるところが大きい。年少児は単一ユニット過程に従っ

た学習を行うが、年長児や大学生では媒介過程に従った学習を行うことが多い (Buss, 1953; Kendler, Kendler & Wells, 1960)。

ここで重要なことは、この差が実験開始以前にどのような経験や知識を持っていたかにあるということである。つまり、年長児や大学生は、すでにある程度のことを経験してきており、大きさといった概念を学習できるレディネスができており、そのため、移行学習実験において、それを媒介として学習ができたものと思われる。一方、年少児の場合、まだ経験が浅いため、大きさといった概念の学習ができないために、単一ユニット過程の学習しかできなかったことになる。

一般にニューラルネットワークのシミュレーションではランダムな絶対値の小さな値を初期値とするが、移行学習のシミュレーションでは、単一ユニット過程の学習、つまり、非逆転移行のほうが速くなる学習しかシミュレーションすることはできない (松尾・木藤, 1991)。媒介過程による学習を実現するためには、第1学習に入る前のネットワークが媒介過程を学習できるような状態になってないといけない。つまり、内部構造の初期値をランダムに設定するのではなく、そのレディネスができている状態にしておくことである。

シミュレーションを行った学習が、単一ユニット過程に従ったのか、媒介過程に従ったのかは、逆転移行と非逆転移行の成績の比較でも相対的には確かめられるが、実際には内的な過程、つまりネットワークの内部構造を分析しなければならない。先にも述べたように、神経回路モデルでは、結合係数と閾値の内部構造が脳の学習の状態をシミュレーションしており、その内部構造を知ることによって、はじめて内的な学習過程について言及することができる。

したがって、本研究では、移行学習のニューラルネットワーク・シミュレーションにおいて、学習後の内部構造を分析することによって、どのような学習の過程が形成されたかを検討する。

シミュレーション1

どのような初期状態が学習成績や内部構造に影響を与えるかを検討する。初期状態の設定は、演繹的に定めるのではなく、ランダムに初期値を設定して計算機シミュレーションを多数回行う。そして学習回数と内部構造を分析し、その中から媒介過程に従ったと思われるようなものを抽出し、その初期値がどのような値であったかを帰納的に探ることとする。ただし、初期状態をまったくランダムに設定すると、どのような初期状態が学習成績や内部構造に影響を与えるのかわからないため、ここでは、初期状態に設定する値を限定 (4つの値) して行う。

【方法】

ネットワークは3層の階層型ネットワークで、各層のユニット数は、入力層で4 ($I_{00} \sim I_{03}$)、隠れ層2 (H_{00}, H_{01})、出力層1 (O_{00}) である。隠れユニット及び出力ユニットの入出力関数には、傾きパラメータ1.0のロジスティック関数を用いた。学習アルゴリズムは、慣性項のないバックプロパゲーション法を用いた。学習係数は、結合係数、閾値ともに0.5とした。なお、学習による結合係数と閾値修正は、1通りの学習パターンによるフォワード処理がなされた後に一括して行う。学習の終了条件は出力値と教師信号の差の絶対値の平均が0.05未満になることとし、学習の最大打ち切り回数は、2,000とした。

初期状態の結合係数及び閾値は、-0.1, -0.05, 0.05, 0.1の4つの値のうち、いずれかひとつがランダムに設定された。

入力信号および教師信号は、表1に示したように4つの学習パターンがあり、第1学習を行ったのち、

第2学習では、逆転移行条件（教師信号が第1学習とすべて逆転）と非逆転移行条件（2つの信号のみが変わる）の2通りの学習を行った。それぞれ、100回のシミュレーションを行った。

表1 学習に用いた入力信号及び教師信号

入力信号				教師信号			
I	00	01	02	第1学習		第2学習	
				03	逆転移行	非逆転移行	
1	0	1	0	1	0	1	
1	0	0	1	1	0	0	
0	1	1	0	0	1	1	
0	1	0	1	0	1	0	

【結果と考察】

図3に、各学習段階における学習回数の度数分布を示した。100例のうち、逆転移行において学習が2,000回で収束しなかった例が4例あり、これらは以下の処理の対象から除外した。表2にそれぞれの平均値と標準偏差を示した。ここでは、初期状態をある程度ランダムに定めたため、学習回数は、どの例においても非逆転移行のほうが逆転移行より少なく、学習成績の面だけからみると媒介過程の成立は確認されなかった。ただし、ばらつきをみると、第1学習や非逆転移行に比べ、逆転移行のばらつきが大きいことがわかる。これは、第1学習や非逆転移行では初期状態の違いが学習成績にそれほど大きな影響を与え

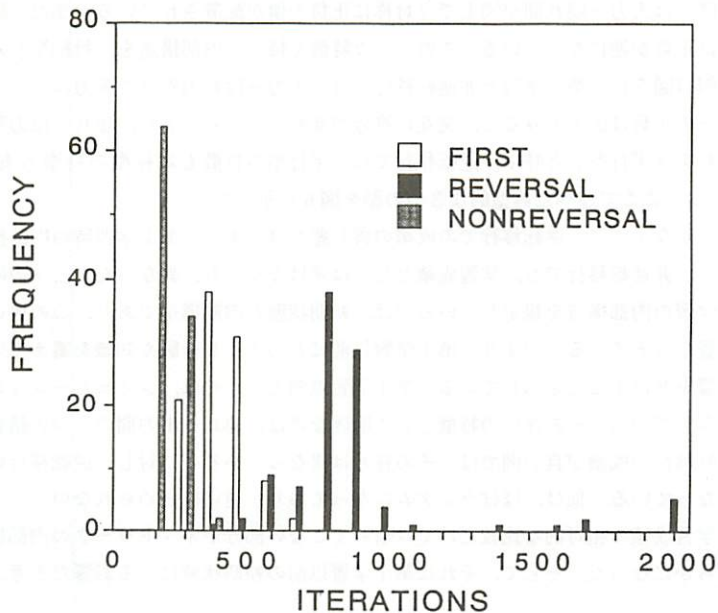


図3 各学習段階における学習回数の度数分布

表2 各学習段階における学習回数

	平均値	標準偏差
第1学習	475.89	88.74
逆転移行	891.60	196.87
非逆転移行	286.35	36.49

表3 各学習段階間の学習回数の相関係数

第1学習-逆転移行	0.167
第1学習-非逆転移行	0.753**
逆転移行-非逆転移行	0.126

** : $p < 0.01$

ないが、逆転移行では、その影響が大きく、初期状態の中には、逆転移行の学習に適したものとそうでないものが存在していることを示唆している。

各学習段階間で学習回数の相関をみると、表3のようになる。第1学習と非逆転移行での相関は有意に高いが、逆転移行では、非逆転移行との相関も第1学習との相関もほとんどみられない。非逆転移行の学習成績は、第1学習時の学習成績で予測はつくが、逆転移行での成績への影響は、第1学習時の学習成績からはわからない。逆転移行も非逆転移行も、第1学習の後に行う学習であるため、その学習回数の相関は高くてもいいはずだが、そうはなっていない。これは、第1学習時の学習回数には反映されていない要因が逆転移行に影響を与えているからだと考えられる。何回学習に要したかではなく、どのような学習を行ったかに起因しており、それを明らかにするには、ネットワークの内部構造を分析する必要がある。

したがって、逆転移行の学習回数が最も少なかったものと最も多かったものをそれぞれ3例ずつ抽出し、各学習段階における内部構造を分析した。内部構造をヒント表示に做って、図4及び図5に示した。両図を比較すると、その内部構造の特徴の違いが明らかに存在していることがわかる。

ここでは、結合係数の絶対的な値よりも正負の符号に着目する。逆転移行の学習成績が良い例(図4)は、第1学習以降では入力-隠れ間が図上で点対称に正負の値が配置されている。また、隠れ-出力間の2つの結合係数は正負が逆になっている。このような特徴を持った内部構造を「対称型」と呼ぶ。逆転移行の成績が悪い例(図5)の第1学習と非逆転移行では、入力-隠れ間が同じ入力ユニットでは同じ値を示しているが、その正負は図上でみると、交互に符号が変わっている。また、隠れ-出力間では同じ値を示している。これを「平行型」と呼ぶ。逆転移行では、平行型の特徴も対称型の特徴も有しておらず、「非対称型」と呼ぶ。ここで述べた典型的な3つの型を図6に示した。

この分析で明らかのように、逆転移行での成績の善し悪しは、すでに第1学習時の内部構造によって決定している。また、非逆転移行でも、学習成績としては差はなくとも、異なったタイプの内部構造が存在している。第1学習の内部構造を規定しているのは、初期状態の内部構造であり、この違いが逆転移行の成績に大きな影響を与えている。つまり、第1学習以前にどのような経験や知識を蓄えていたかが、とくに逆転移行に影響を及ぼすことを示している。第1学習以前というのは、シミュレーション上では、初期状態の違いに依存している。その違いの特徴として明確なのは、隠れ-出力間の2つの結合係数の正負の符号である。逆転移行の成績が良い例では、その符号は異なっているのに対し、逆転移行の成績が悪い例では同じ符号になっている。他は、ほぼランダムになっており、違いは認められない。

このように、学習成績の相対的な比較だけでは現れてこない部分がネットワークの内部構造を分析することによって明らかになった。そして、それは第1学習以前の初期状態による影響だと考えられる。

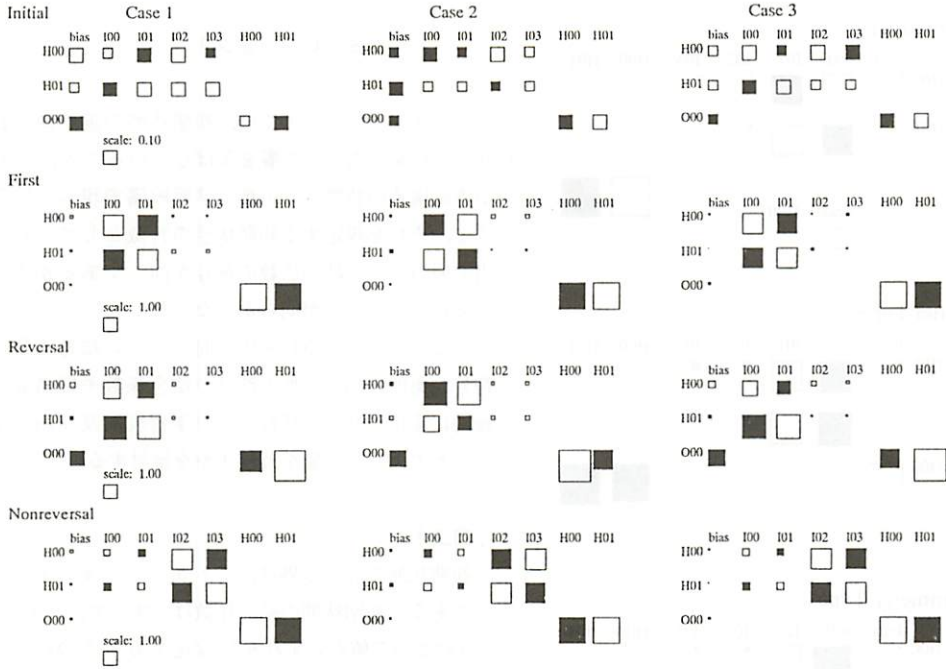


図4 逆転移行の学習成績が良い3ケースの内部構造。正方形の面積が結合係数および閾値 (bias と表記) の値を表している。正負の符号は、白がプラスで黒がマイナス。初期状態のみ、スケールが異なっている。

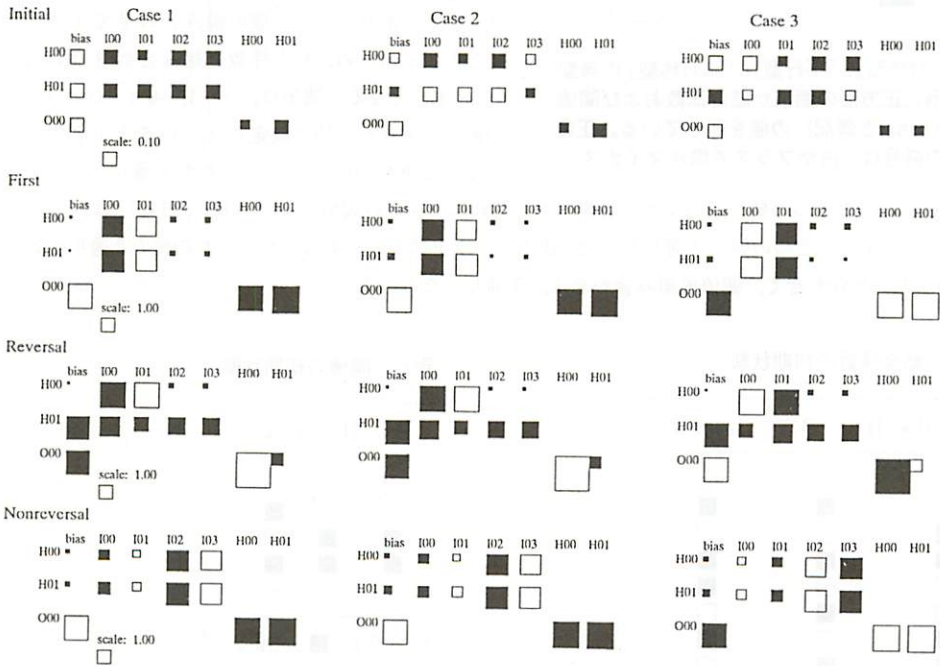
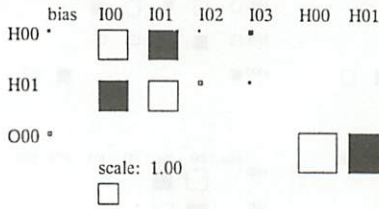
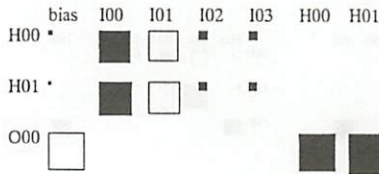


図5 逆転移行の学習成績が悪い3ケースの内部構造。正方形の面積が結合係数および閾値 (bias と表記) の値を表している。正負の符号は、白がプラスで黒がマイナス。初期状態のみ、スケールが異なっている。

Symmetry Type



Parallel Type



Asymmetry Type

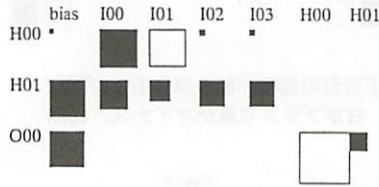


図6 「対称型」、「平行型」、「非対称型」の典型例。正方形の面積が結合係数および閾値 (bias と表記) の値を表している。正負の符号は、白がプラスで黒がマイナス。

は、+0.1, -0.1の2つの値で次のように設定した。隠れユニットの閾値の2つ (H₀₀, H₀₁) は同じ値になるように設定し、2つの値の2通りとした。出力ユニットは1つであるため、2つの値の2通りとなる。先の0.0の場合と合わせて、閾値の組み合わせは、5通りとなる (表5)。

表4 結合係数の初期状態

I* - H*	H ₀₀ - O ₀₀	H ₀₁ - O ₀₀
□	□	□
□	■	■
■	□	□
■	■	■
□	□	□
□	■	□
■	□	■
■	■	□

□ : +0.1 ■ : -0.1

シミュレーション2

シミュレーション1では、初期状態の違いが逆転移行の成績に大きな影響を及ぼしていること、それが内部構造の特徴として第1学習段階で現れていること、それを規定する初期状態の特徴として隠れ-出力間の2つの結合係数の符号が同一であるか否かがあげられることが明らかになった。

ここでは、その隠れ-出力間の2つの結合係数の符号の条件に加え、他の部分の結合係数や閾値も系統的に変化させて、移行学習の学習成績及び内部構造にどのような影響を及ぼすかを検討する。

[方法]

初期状態の設定以外は、シミュレーション1と同じである。初期状態の結合係数は、すべて、+0.1と-0.1の2つの値のいずれかに設定する。その組み合わせは、表4に示したようになる。入力-隠れ間 (I* - H*) の8ヶ所は2つの値のいずれか一方に統一した。隠れ-出力間の結合は2ヶ所 (H₀₀ - O₀₀, H₀₁ - O₀₀) あり、2つの値の組み合わせで4通り設定した。そのため、結合係数の組み合わせは8通り (2 × 4) となる。閾値は、+0.1, -0.1, 0.0の3つの値のいずれかの値に設定した。0.0をとる場合、すべての閾値を0.0に統一し、これを1通りとした。後

表5 閾値の初期状態

H ₀₀	H ₀₁	O ₀₀
□	□	□
□	□	■
■	■	□
■	■	■
•	•	•

□ : +0.1 ■ : -0.1
• : 0.0

結合係数の組み合わせ数（8）と閾値での組み合わせ数（5）の組み合わせで、すべての組み合わせは40通りとなる。ここでは、初期状態はランダムに設定されていないため、シミュレーションの回数はすべて1回ずつとした。

〔結果と考察〕

表6に、各学習に要した学習回数および各学習終了後のネットワークの内部構造の分類パターンを示した。逆転時の学習回数は、どの条件でも非逆転移行のほうが逆転移行より少なかった。内部構造は、シミュレーション1に倣ってヒントン表示によって対称型、平行型、非対称型の3つに分類した。なお、学習終了条件に満たなかったもの（表中、「—」と表記している）は、その分類には含めていない。この表では、その学習成績と内部構造をもとにさらに3つの群に分類をしている。

第1群での初期状態の特徴は、隠れ—出力間の結合係数の符号が同一であるということである。学習成績は、どの学習段階をとっても他の2群よりも劣り、逆転移行では、2,000回までで学習が収束しなかった。内部構造は、第1学習、非逆転移行とも平行型を示し、シミュレーション1で示した逆転移行の成績が悪い例（図5）と同様の結果を示している。

第2群と第3群では、隠れ—出力間の2つの結合係数の初期値は正負の符号が異なっているが、閾値においては、第3群ですべて0であるのに対し第2群は0ではない。両群とも第1学習と非逆転移行においては、学習成績も内部構造もほとんど差はなかった。学習回数は第1群よりもかなり少なく、内部構造も対称型を示した。しかし、逆転移行においては、両群に大きな違いがみられる。学習回数は、第3群は第2群の半分以下になっており、内部構造では、第2群は非対称であるのに対し、第3群では対称型を示した。

これら3つの群の内部構造は各群内ではほとんど同じ型をしており、各群よりそれぞれ1例ずつを図7に示した。移行時に内部構造がどのように変化するかをみると、非逆転移行においては、どの群でも同様の特徴を示している。第1学習時には、入力—隠れ間の入力ユニット I_{00} と I_{01} の結合係数の絶対値が大きくなっているのに対し、非逆転移行では、 I_{02} と I_{03} のユニットに対する結合の係数が逆に大きくなっている。そして、隠れ—出力間の結合係数にはほとんど変化はない。この学習が媒介過程に従ったものであると仮定すると、この移行は次元間移行であると言え、媒介とするものを変化させていることになる。

逆転移行の場合は、3つの群でそれぞれ異なっている。第1群では収束しなかったが、2,000回の学習終了時の結合係数はすべて同じ符号になっている。第2群では入力ユニット I_{00} 、 I_{01} と隠れユニットの一方のユニットとの間の結合をほぼ保ちながら、他の部分はかなり形が崩れている。第3群は、入力—隠れ間の結合係数はほとんど変化しておらず、隠れ—出力間の結合係数が対称的に正負が逆転する学習を行っている。これは、文字通り反応を逆転することになり、媒介型の学習に対応している。

すなわち、第3群においては、入力—隠れ間の結合が媒介過程を示していると考えられる。そのため、逆転移行ではこの部分の変化がなく、同じ次元を媒介とすることができ、反応する部分（隠れ—出力間）で逆にすればよい。非逆転移行では、第1学習で使った次元の媒介過程を使うことはできず、次元を変えなければならない。そのため、入力—隠れ間の部分で上述のような変化が起こることになる。さらに、第1学習と非逆転移行で結合係数の絶対値が大きくなる部分が完全に分離されており、それぞれの次元に対応していることがわかる。つまり、第1学習で学習された次元は、入力ユニット I_{00} と I_{01} に対応して媒介過程が成立しており、非逆転移行でのもうひとつの次元は、入力ユニット I_{02} と I_{03} に対応して媒介過程が形成されている。

表6 すべての条件での学習回数及び内部構造。表中、「—」とあるのは、学習終了条件を満たさなかったもの。

分類群	初期状態						学習結果					
	結合係数			閾値			学習回数			内部構造の分類パターン		
	I ₁ -H ₁	H ₀₀ -O ₀₀	H ₀₁ -O ₀₁	H ₀₀	H ₀₁	O ₀₀	第1学習	逆転移行	非逆転移行	第1学習	逆転移行	非逆転移行
1	□	□	□	□	□	□	547	--	318	平行型	--	平行型
	□	■	■	□	□	■	547	--	318	平行型	--	平行型
	□	□	□	■	■	□	539	--	318	平行型	--	平行型
	□	■	■	■	■	■	539	--	318	平行型	--	平行型
	■	□	□	□	□	□	534	--	318	平行型	--	平行型
	■	■	■	□	□	■	534	--	318	平行型	--	平行型
	■	□	□	■	■	■	532	--	318	平行型	--	平行型
	■	■	■	■	■	■	532	--	318	平行型	--	平行型
	□	□	□	.	.	.	527	--	318	平行型	--	平行型
	□	■	■	.	.	.	527	--	318	平行型	--	平行型
	■	□	□	.	.	.	521	--	318	平行型	--	平行型
	■	■	■	.	.	.	521	--	318	平行型	--	平行型
	□	■	■	□	□	□	519	--	318	平行型	--	平行型
	□	□	□	□	□	■	519	--	318	平行型	--	平行型
	□	■	■	■	■	□	515	--	318	平行型	--	平行型
	□	□	□	■	■	■	515	--	318	平行型	--	平行型
■	■	■	□	□	□	514	--	318	平行型	--	平行型	
■	□	□	□	□	■	514	--	318	平行型	--	平行型	
■	■	■	■	■	□	513	--	318	平行型	--	平行型	
■	□	□	■	■	■	513	--	318	平行型	--	平行型	
2	■	□	■	■	■	□	393	992	274	対称型	非対称型	対称型
	■	■	□	■	■	□	393	992	274	対称型	非対称型	対称型
	■	□	■	■	■	■	393	992	274	対称型	非対称型	対称型
	■	■	□	■	■	■	393	992	274	対称型	非対称型	対称型
	□	□	■	□	□	□	393	1048	272	対称型	非対称型	対称型
	□	■	□	□	□	□	393	1048	272	対称型	非対称型	対称型
	□	□	■	□	□	■	393	1048	272	対称型	非対称型	対称型
	□	■	□	□	□	■	393	1048	272	対称型	非対称型	対称型
	■	□	■	□	□	□	392	1012	273	対称型	非対称型	対称型
	■	□	□	□	□	□	392	1012	273	対称型	非対称型	対称型
	■	□	■	□	□	■	392	1012	273	対称型	非対称型	対称型
	■	■	□	□	□	■	392	1012	273	対称型	非対称型	対称型
	□	□	■	■	■	□	392	1030	273	対称型	非対称型	対称型
	□	■	□	■	■	□	392	1030	273	対称型	非対称型	対称型
□	□	■	■	■	■	392	1030	273	対称型	非対称型	対称型	
□	■	□	■	■	■	392	1030	273	対称型	非対称型	対称型	
3	□	□	■	.	.	.	389	457	275	対称型	対称型	対称型
	□	■	□	.	.	.	389	457	275	対称型	対称型	対称型
	■	□	■	.	.	.	389	457	275	対称型	対称型	対称型
	■	■	□	.	.	.	389	457	275	対称型	対称型	対称型

□ : +0.1 ■ : -0.1 . : 0.0

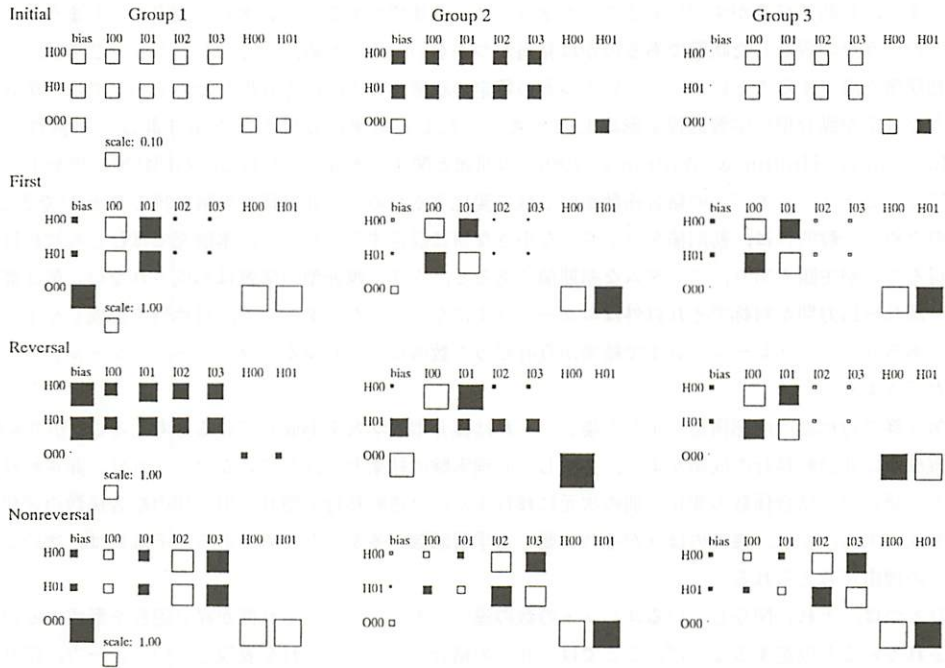


図7 第3群での内部構造。正方形の面積が結合係数および閾値（biasと表記）の値を表している。正負の符号は、白がプラスで黒がマイナス。初期状態のみ、スケールが異なっている。

第1群や第2群では、見かけ上第1学習と非逆転移行において媒介過程と考えざるをえない。しかし、第3群では、その内部構造において媒介過程が成立したと考えることができる。

全体的考察

シミュレーション1では、初期状態が第1学習の成績には影響を与えなくとも、その内部構造に影響を与え、それが、逆転移行での学習成績や内部構造を規定していることが明らかになった。シミュレーション2では、それを確かめるために初期状態を系統的に変化させて行った結果、媒介過程の形成に入力-隠れ間が関与していることが示唆された。

シミュレーション1の第1学習や非逆転移行での100回のシミュレーションの学習成績のばらつきは偶然誤差の範囲内であったと考えられるが、逆転移行ではそうではなかったと考えられる。各ユニットの初期状態をランダムな小さな値に設定しているが、これは必ずしも何も学習していない状態を表わしているとは言い切れない。実験者の意図したことに関しては何も学習していない状態ではあるが、その初期状態が実際の学習結果に大きな影響を与えている。本研究の場合、第1学習や非逆転移行では、それが学習成績に影響を与えなかったが、逆転移行では影響を与える程度であった。とくに、本研究で用いたバックプロパゲーション法の場合、ローカルミニマムに落ち込むことがあり、シミュレーション1では4例において、また、シミュレーション2においては第1群において、そのような結果が認められた（もっとも、学習最大回数を増やせば収束はする）。そのため、工学的な応用の場面では効果的な初期値の設定の仕方が

様々に検討されている(仁木, 1989; 賈・戸田・臼井, 1990; 五反田, 1993)。

つまり, 初期値自身がすでにあることを学習している状態である。ランダムに設定してしまうと, それがどのような学習をした状態であるのかは見当がつかない。そのため, シミュレーション2では, 系統的に初期値を設定することによって, いくつかの特定の状態を人工的に作り出した。その中では, 第3群のような設定が媒介型の学習過程を形成したと考えられる。結果が良好だった第3群は, 対称性の破壊(Rumelhart, Hinton & Williams, 1986)の問題と関連があると思われる。初期状態の値をすべて同じにしまうと, すべての結合係数が同じ様に変化するために, 非対称性の解は得ることができない。そのため, 一般的には, 初期値をランダムな小さな値に設定する。しかし, 本研究ではむしろ対称性の解を得ることが主眼であり, ランダムな初期値であると, うまく媒介型の学習は形成されない。第3群の場合, 隠れ-出力間が対称でそれ以外はニュートラルになっていたパターンで, 対称性を形成しやすかったのであろう。シミュレーション1で結果が良好だった数例は, いわゆる「ラッキーイニシャル」ということができよう。

第3群で対称型の内部構造を示した場合, それは媒介型の学習を形成していると考えることができたが, 学習成績は非逆転移行の成績がよく, 必ずしも心理実験の結果と一致していない。これは, 非逆転移行で入力-隠れ間の結合係数の変化(別の次元に移行する)と逆転移行で隠れ-出力間の結合係数の変化(反応を逆にする)では, 後者のほうがそれに要した学習回数が多かったためである。それには, 次のような2つの理由が考えられる。

ひとつは, それに関与しているユニットの数の違いである。入力-隠れ間が媒介過程を形成するのに利用されていると仮定するならば, ここでは, 8つの結合によって, それを表現している。一方, 隠れ-出力間では, 2つの結合によって表現している。したがって, 前者のほうが並列分散処理としては効率がよく, 変化に要する時間は少なくてすむ。

もうひとつは, 媒介過程以外の学習の問題である。ここでは, 媒介過程が成立するかどうか重点を置いて初期値を設定した。しかし, 媒介過程による学習がたとえスムーズにできたとしても, 反応を逆にする点に関しては何も操作はしていない。逆転移行では反応を逆にするという操作が学習できてないといけない。したがって, ここでのネットワークは, 反応を逆にするという学習をしていないのかもしれない。そのためには, 反応を逆にするという学習をさせるか, バックプロパゲーション法自身のアルゴリズムを改良する(松尾, 1992)しなければならない。

学習成績の面においては, 心理実験と一致はしなかったが, その内部構造においては, 心理的過程と一致することが示された。内部構造を分析することは, ニューラルネットワークにおいて何を学習したかを明らかにするには有効な手段であると考えられる。Rumelhart, Smolensky, McClelland & Hinton (1986)は, いくつかの部屋のスキーマを家具などの記述子間でどのように結合しているかを重み行列で示している。また, 吉野・石川(1992)は, 忘却付き構造学習によって, 学習した結果の内部構造を示している。それによると, 5×5 のマトリックス図形において大小を識別するモジュールとして形成されたものは, 隠れユニット2つと入力ユニット2つに対して, 我々が示したように対称型をしている。吉野らの実験は, 隠れユニットの数が忘却によって減少していくのであるが, 最終的にこのような形のモジュールに落ち着いたことを示した。また, 内部構造ではなく, 内部表現を積極的にとらえた研究例も数多い(Sejnowski & Rosenberg, 1987; 入江・川人, 1990; 吉田・浅野・松田, 1991)。

本来, ニューラルネットワークの特徴のひとつは, 入力と出力にある値を入力すれば, そのネットワーク内部で何がなされているかをあらかじめ考えておく必要がないということにある。しかし, 行動主義的に内的な過程を無視してしまうのは, 工学的な応用においては, それで満足できる場合もあるかもしれ

ないが、心理学的に人間の学習をシミュレーションすることによって、人間の学習過程を明らかにすることに目的をおくのであれば、内的過程を無視するわけにはいかない。したがって、ニューラルネットワークの研究において、その内的過程を積極的に分析していくことが必要になってくるであろう。

参考文献

- Buss, A.H. 1953 Rigidity as a function of reversal and nonreversal shifts in the learning of successive discrimination. *Journal of Experimental Psychology*, 45, 75-81.
- 五反田博 1993 単一出力3層ネットワークの初期値設定に関する考察 信学技報, NC92-128.
- 入江文平・川人光男 1990 多層パーセプトロンによる内部表現の獲得 信学論 (D-II), J73-D-II, 1173-1178.
- 賈棋・戸田尚宏・白井支朗 1990 ニューラルネットワークにおける逆伝搬学習アルゴリズムの初期値設定に関する一考察 信学論 (D-II), J73-D-II, 1179-1185.
- Kendler, H.H. and Kendler, T.S. 1962 Vertical and horizontal processes in problem solving. *Psychological Review*, 69, 1-16.
- Kendler, T.S., Kendler, H.H., and Wells, D. 1960 Reversal and nonreversal shifts in nursery school children. *Journal of Comparative and Physiological Psychology*, 53, 83-88.
- 松尾太加志 1992 バックプロパゲーション法における入出力関数の符号反転: 移行学習のシミュレーション, 佐賀女子短期大学研究紀要, 26, 69-75.
- 松尾太加志・木藤恒夫 1991 移行学習のニューラルネットワーク・シミュレーション, 佐賀女子短期大学研究紀要, 25, 1-10.
- McCulloch W.S. and Pitts, W. 1943 A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- Minsky, M. and Papert, S. 1969 *Perceptrons*. Cambridge, MA:MIT Press. (齊藤正男訳 [パーセプトロン] 東京大学出版会, 1971)
- 仁木和久 1989 2値パターン誤差逆伝搬学習の性質と学習の高速化手法 信学技報, NC89-4.
- Rumelhart, D.E., Hinton, G.E. and Williams, R.J. 1986 Learning internal representations by error propagation. In J.L. McClelland, D.E. Rumelhart and The PDP Research Group (Eds.), *Parallel Distributed Processing, Volume 1*. Cambridge: MIT Press. (甘利俊一監訳 [PDPモデル] 産業図書, 1989)
- Rumelhart, D.E., Smolensky, P., McClelland, J.L., and Hinton, G.E. 1986 Schemata and sequential thought processing. In J.L. McClelland, D.E. Rumelhart and The PDP Research Group (Eds.), *Parallel Distributed Processing, Volume 2*. Cambridge: MIT Press. (甘利俊一監訳 [PDPモデル] 産業図書, 1989)
- Sejnowski, T.J. and Rosenberg, C.R. 1987 Parallel networks that learn to pronounce English text. *Complex Systems*, 1, 145-168.
- 吉田富美男・浅野一志・松田基一 1991 内部表現を利用した誤差逆伝搬法 信学技報, NC91-121.
- 吉野健一・石川真澄 1992 忘却付き構造学習を用いたモジュール構造ニューラルネットワークの自動タスク分割の試み 神経回路学会第3回全国大会講演論文集, 84-85.

北九州大学文学部紀要

(人間関係学科)

第 1 巻

畑中健一教授退職記念号

目 次

畑中 健一教授略歴・業績	1
畑中 健一・野井 英明：北部九州における 最終氷期最盛期の花粉群集	9
松尾太加志・木藤 恒夫：移行学習のニューラルネットワーク・ シミュレーションにおける内部構造の分析	15
亀山 徳彦：大分市丹生台地に認められる後期更新世の高海水準期について ——とくに阿蘇4火砕流堆積時を中心にして——	29
倉崎 取：北九州大学の弓道部選手の体格について	37
保坂恵美子：旧産炭地における内発的地域づくりの現状と課題について ——せちばるおこしを中心に——	41
山内 隆久：障害者に対する態度の変容におよぼす対人的接触の効果 ——接触相手・情報に対する態度と視覚障害者に 対する態度との相互影響過程——	59
青木多寿子・山内 隆久：青年における身近な他者への 役割期待の違いと性差(II)	71
安増 朗：時間軸における一音節弁別のための主要素に関する一考察	81

北九州大学文学部

1994年3月発行