論　文

選択的不感化法を適用した層状ニューラルネットの情報統合能力

森田　昌彦[†a)]　村田　和彦[†]　諸上　茂光[†]　末光　厚夫[†]

Information Integration Ability of Layered Neural Networks with the Selective Desensitization Method

Masahiko MORITA[†a)], Kazuhiko MURATA[†], Shigemitsu MOROKAMI[†], and Atsuo SUEMITSU[†]

あらまし　伝統的な層状ニューラルネットは，入出力関係が文脈に強く依存する場合や 2 つの独立な入力情報を統合する必要がある場合，学習および汎化の性能が著しく低下する．これは，「1 対多対応による荷重の平均化」という本質的な問題に起因すると考えられ，学習アルゴリズムをどのように改良しても解決することはできない．この限界を乗り越えるために，先に著者らが提案した「選択的不感化」という手法を層状ニューラルネットに適用し，その情報統合能力について検証した．その結果，このモデルは文脈依存的連合課題に関して，優れた学習能力をもつだけでなく，これまでにない高い汎化能力を示すことがわかった．これは，前述の平均化の問題が解消すると共に，2 種類の分散表現が局所表現を経ることなく統合されるためだと考えられ，ニューラルネットによる情報処理の可能性を大きく拡げるものである．

キーワード　神経回路網，多層パーセプトロン，分散表現，文脈修飾，汎化能力

1. まえがき

人間が普段行っているような知的情報処理は，文脈に大きく依存する場合が多い．つまり，与えられる情報が同じであっても，その場の様々な状況に応じて異なる結果を出力する必要がある．ここで「文脈」を種類の異なる情報と考えれば，より一般的に，二種類の情報を統合してその組合せに応じて適切な出力を選ぶような処理が必要と言うこともできよう．

ところが，ニューラルネット (以下，人工神経回路網の意味で用いる) の最も基本的なものと言える，入力層と出力層からなる 2 層モデル (単層パーセプトロン) は，このような文脈依存的処理や情報統合の能力が極めて乏しい．この欠点に対処するため，中間層を設けて誤差逆伝播法 [1] などにより学習する，いわゆる多層パーセプトロンが一般に用いられている．しかし，この方法は，ネットワークの規模が大きくなり学習すべき入出力関係が複雑になると，多数の中間層素子や膨大な量の計算が必要な上，しばしば無限に長い学習時間がかかる．

著者ら [2] は，このような難点が「1 対多対応による荷重の平均化」という本質的な問題に起因するものであって，古典的なニューラルネットの枠組みでは解決が困難であることを指摘すると共に，「選択的不感化」という文脈修飾の手法を用いた文脈依存的連想モデルにおいて，この問題が解決されていることを示した．但し，このモデルは非単調素子を用いた回帰型ニューラルネットであり，一般的なニューラルネットとはかなり異なるため，従来のモデルの問題点や，導入した手法の効果があまり明確ではなかった．そこで，本論文では，層状のニューラルネットモデルに選択的不感化による文脈修飾を適用し，それによる情報統合能力の向上について検証する．

2. 多層パーセプトロンの情報統合能力

2.1 単層パーセプトロンと平均化の問題

以下の課題を考える．まず，$p(>=2)$ 個のパターン $S^1, \ldots, S^p$ の他に，$q(>=2)$ 通りのパターン $C^1, \ldots, C^q$ があるとき，両者の組合せ (S^μ, C^ν) に応じて出力すべき目標パターン $T^{\mu,\nu}$ が与えられるという課題である．ここで，$T^{\mu,\nu}$ はすべて異なっていて

† 筑波大学大学院システム情報工学研究科，つくば市
Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Ten-nodai, Tsukuba-shi, 305-8573, Japan
a) E-mail: mor@bcl.esys.tsukuba.ac.jp

もよいし，重複があってもよいが，当面はすべて全く異なるパターンとする．また，以下では特に断りのない限り，情報はすべて 2 値 (±1) パターンによって分散表現されており，その次元 n は十分に大きいものとする．

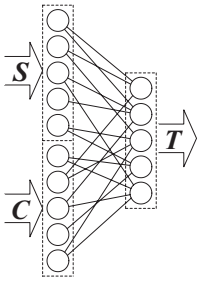
この課題は，S と C という 2 種類の情報を統合して目標パターン T を連想する課題と見なすこともできるし，通常の連想課題のように入力パターン S によって T が一意に決まるのではなく，C で表される文脈に依存して S と T との関係が変化する課題と見ることもできる．以下では，便宜上 S を入力パターン，C を文脈パターンと呼ぶが，ここで言う「文脈」とは「入出力関係に影響を与える情報」という程度の意味である．

さて，最初に 2 層のニューラルネットについて考えよう．この場合，図 1(a) のように S と C に対してそれぞれ n 個の素子を割り当て，$2n$ 個の素子からなる入力層から n 個の素子からなる出力層への結合荷重を学習する (以下，これを直和型モデルと呼ぶ) のが一般的である．ところが，この学習は，p および q が十分小さくない限りうまくいかない (**3.2** 参照)．その理由を簡単に説明すると，次のようになる．

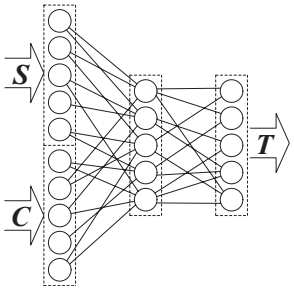
いま，各パターンについて均等に学習を行った結果，うまく学習ができたとしよう．このとき，仮に C を表す部分 (以下，文脈部と呼ぶ) と出力層との結合をすべて切断したとする．すると，出力されるパターンは，文脈パターン C^ν に関係なく S^μ のみによって決まるから，$T^{\mu,1},\ldots,T^{\mu,q}$ を平均化したパターン $\bar{T}^\mu$ かそれに近いパターンになっているはずである．

ところが，$T^{\mu,\nu}$ は独立なパターンであるから，$\bar{T}^\mu$ は q が大きくなるにつれて μ によらないある一定のパターンに近づいていく．つまり，文脈パターン数 q が十分大きければ，入力パターンが何であっても出力パターンはほぼ同じパターンとなる．これは，入力部から出力層への結合荷重が無意味になるということであり，この問題を 1 対多対応による平均化と呼んでいる．

同様に，文脈部から出力層への結合も，入力パターン数 p が大きくなると無意味になる．従って，p および q が十分大きいとき，結合荷重をどんな学習則を用いてどんな値に設定しても，すべての $T^{\mu,\nu}$ を正しく出力することはできない．また，p や q がそれほど大きくなくても，$T^{\mu,\nu}$ の ν または μ に関する平均が μ または ν によらず一定の場合には，同じ議論が成り立つ．



(a) Single-layer perceptron



(b) Multilayer perceptron

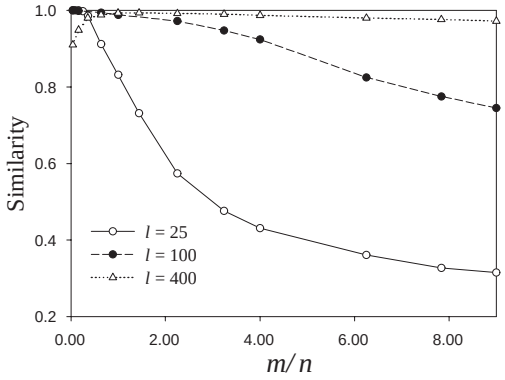
図 1　層状ニューラルネットの構造
Fig. 1　Architecture of layered neural networks.

2.2　3 層 BP モデルの学習容量

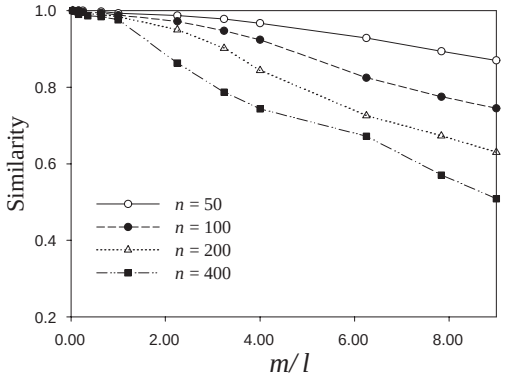
では，図 1(b) のように，中間層を加えた 3 層のニューラルネットはどうであろうか．この場合，入力層から中間層への結合荷重を学習するため，誤差逆伝播 (BP) 法 [1] を用いるのが一般的である．そこで，直和型モデルにこの方法を適用した場合 (以下，BP モデルと呼ぶ) について，数値実験を行った．

各部の素子数 (パターンの次元) は n であり，入力パターン S^μ および文脈パターン C^ν は同数 ($p=q$) ずつ，目標パターン $T^{\mu,\nu}$ は $m=pq$ 個，それぞれランダムに作成するものとする．学習は，出力の平均 2 乗誤差を見ながら十分な回数 (平均約 10000 回) 行ったが，誤差が十分に減少しない場合には，結合荷重の初期値を変えて最大 10 回までやり直し，その最良値を採用した．なお，学習時には，中間層および出力層の出力関数として -1 から 1 の連続値をとるシグモイド関数を用いたが，最終的な出力結果を得る際には各素子の出力を ±1 に 2 値化した．

図 2 に結果を示す．(a) のグラフの縦軸は，2 値化した出力パターン Y と目標パターン $T^{\mu,\nu}$ との類似度 (両ベクトル間の方向余弦で定義する) の平均値，横軸は m/n である．n は 100 に固定し，中間層の素子



(a) $n = 100$



(b) $l = 100$

図 2　BP モデルの学習性能
Fig. 2　Learning performance of the BP model.

数 l は 25, 100, 400 の 3 通りに変えた．一方 (b) は，$l = 100$ に固定して n を 50, 100, 200, 400 の 4 通りに変えたグラフであり，横軸は m/l である．

この図から，いくつかの興味深い事実が読み取れる．

（1） 学習能力は，主に中間層の素子数 l に依存する．また，l が一定ならば，n が大きいほど性能が低下する．

（2） 単純に考えて l が大きいほど学習容量は大きいはずであり，実際その傾向は見られる．但し，l が増えると学習がローカルミニマムに陥りやすくなり，誤差が 0 に収束することはない．意外なことに，学習パターン数 m が少ないとき，その傾向が顕著である．

（3） その結果として，平均類似度が 1 となる，つまりすべてのパターンを完全に学習できるのは，ごく限られた場合だけであった．つまり，平均類似度が 1 となる最大の m を学習容量と定義し，有限の学習時間で考えるならば，l および n が十分大きいときの学習容量はほとんど 0 と言える．

もちろん，類似度が 1 でなくても誤差が十分小さければ学習できたと見なすならば，容量は比較的大きく，l から $2l$ 程度はあると言える．しかし，この誤差は，十分に学習したパターンに関する誤差，すなわちサンプル誤差に相当するものであって，未学習のパターンに関する誤差 (汎化誤差) ではないことに留意すべきであろう．

2.3　3 層 BP モデルの汎化性能

では，多層パーセプトロンの汎化誤差はどの程度であろうか．ここでは，ごく簡単な 2 変数関数の近似，具体的には二つの角度 θ_1 と θ_2 からその差 $\theta_3 = \theta_1 - \theta_2$ を求めるという課題を例にとって論じる．普通の数値ではなく角度としたのは，値域に端がないようにするためであり，360° の整数倍の違いは無視 (例えば $-10°$ は 350° と同一視) して $0 \leqq \theta_k < 360$ $(k = 1, 2, 3)$ とする．

θ_k はそれぞれ n 次元の 2 値パターン $\Theta(\theta_k)$ で表現するが，汎化が起きうるよう，角度を連続的に変えると対応するパターンも連続的に変化しなくてはならない．ここでは最も単純に，Θ の n 個の成分を円環状に配置したとき，連続する半数を 1，残りを -1 とし，両者の境界を θ_k の値に応じて徐々に変えることにした．例えば $n = 360$ の場合，θ が 1° 変わるごとに $\Theta(\theta)$ の成分が 2 個ずつ変化することになる．

この課題を 2 層の直和型モデルおよび 3 層の BP モデルに学習させる．学習用サンプルは，(θ_1, θ_2) の組を m 個ランダムに選んで作成した．直和型モデルに関しては，直交学習を各サンプルにつき 30 回行ったが，学習回数を増やしても結果はほとんど同じであった．BP モデルの学習回数は，各サンプルにつき 10000 回としたが，こちらも学習回数の増加による結果の違いは見られなかった．

図 3 に，$n = 360$，$m = 100$ の場合の結果を示す．これは，θ_1 を 0° に固定し，θ_2 (横軸) を 0°〜360° まで変えた場合の出力パターン Y と正解パターン $\Theta(\theta_3)$ との類似度 (縦軸) を示したものであり，破線は直和型モデル，実線は BP モデル $(l = 200)$ を表す．

グラフから，直和型モデルでは，出力パターンがほとんどでたらめに変化していること，また BP モデルも，たまたま入力パターンが学習したものに近いときには正解に近いパターンを出力するが，それ以外では正解と大きく異なるパターンを出力していることがわかる．学習した 100 個のサンプルに関する平均類似度

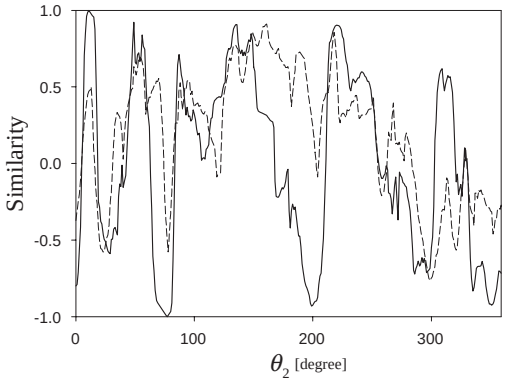


図 3　学習後のモデルの出力結果
Fig. 3　Output of the model after learning.

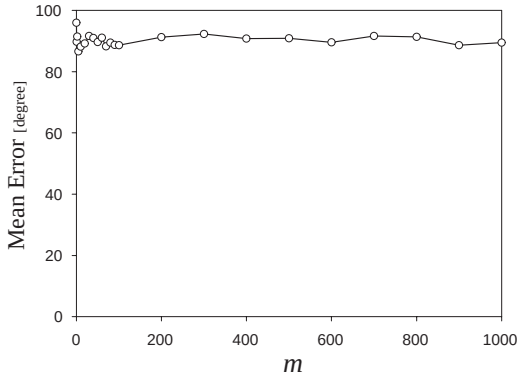


図 4　BP モデルの汎化誤差とサンプル数の関係
Fig. 4　Relationship between generalization error and sample size for the BP model.

は，直和型モデルが 0.56，BP モデルが 0.96 であったのに対し，すべての入力パターンに関する平均類似度は，両モデルとも 0.01 未満であった．

この大きな汎化誤差は，いわゆる過学習によるものでは全くない．なぜなら，中間層の素子数や学習回数を減らしても，サンプル誤差が増えるだけで結果はほとんど同じだったからである．では，サンプル数が少なすぎたからであろうか．

そこで，次に BP モデルについて学習サンプル数 m と平均誤差角 E の関係を調べた．ここで，$\Theta(\theta)$ と Y との類似度が最大となる θ を $\hat{\theta}$ としたとき，$|\hat{\theta} - \theta_3|$ を誤差角と定義する (Y と $\Theta(\theta_3)$ の類似度が 1 でなくても誤差角は 0° となる場合がある)．ランダムに選んだ 1000 通りのテスト入力に関する誤差角の平均値を求めて E としたが，これらのテスト入力に学習サンプルはほとんど含まれないから，E はほぼ汎化誤差と見てよい．なお，$n = 360, l = 200$ では計算時間がかかりすぎるため，$n = 90, l = 100$ で実験した．

結果を図 4 に示す．グラフの横軸は m，縦軸は E である．この図が示すように，サンプル数 m を増やしても，平均誤差は約 90° のまま全く減らない．つまり，少なくともこの課題に関して，多層パーセプトロンの汎化能力はほとんどないと言える．

2.4　多層パーセプトロンの限界

これまで，3 層 BP モデルを代表とする多層パーセプトロンは，中間層をもたない単層パーセプトロンには学習できない XOR 課題を学習できること，理論的には任意の連続関数を任意の精度で近似できることなどから，高い学習能力を有すると考えられてきたように思われる．経験的に多層パーセプトロンが苦手とする課題があることも知られているが，それは問題が難しすぎたから，あるいはモデルの構造，素子数，学習アルゴリズムなどとの「相性」が悪かったから，と見なされていたように感じられる．

しかし，上記の結果は，このような見方に強い疑問を投げかける．すなわち,「二つの変数の差を求める」という単純な課題であっても，学習サンプルを完全に学習するのは困難であるし，学習できたとしても汎化誤差が非常に大きい．もしすべての入力に対して誤差を抑えようとするならば，非常に多数の学習サンプルおよびそれとほぼ同数の中間層素子，そして気の遠くなるような学習回数と計算量が必要である．

はたして，このような性能の悪さは，たまたまこの課題に限ったことであろうか．あるいは，最も基本的な BP 学習則ではなく，様々に改良された学習アルゴリズムのどれかを用いればこのようなことはなかったのであろうか．

そうではない．上記の課題は，確かに従来の方法の弱点が極端に表れる例ではあるが，決して特殊な場合ではない．2 変数の関数近似課題など，独立な 2 種類の情報の広範な統合が必要な課題一般について，サンプル学習効率と汎化性能とはそもそも両立しないのである．

その根本的な原因は，2 層モデルの場合と同様に 1 対多対応による平均化の問題にある．つまり，出力パターンは中間層のパターンによって一意に定まるから，異なるパターンを出力するためには，中間層のパターンが異なっていなければならない．すると，入力層から中間層への結合に関して，2 層モデルの場合と同じ

議論が成り立つから，平均化の問題が生じないような表現，すなわち ν (μ) に関する平均が μ (ν) ごとに大きく異なるような表現を中間層に形成することが必要となる．

しかし，BP 学習は単に出力の誤差を局所的に減らそうとするだけで，特にそのような表現を選ぶ働きがあるわけではない．また，中間層の素子ができるだけ多い方がそのような表現を作りやすいが，一方で素子数の増加は，一般に学習がローカルミニマムに陥る確率を高める．そのため，この種の課題ではサンプルの学習が難しいのである．

更にこのことは，サンプルがうまく学習できたときには，各中間層素子が限られた (S^μ, C^ν) の組合せのみをコードしている，つまり中間層の表現が局所表現に近いことを意味する．その極端な場合として，中間層の表現を最初から局所表現にしてしまえば，少なくとも中間層素子数と同数のサンプルを完全に学習することが可能である．しかし，そうすると，入出力関係に特定の構造があってもそれが中間層に反映されないし，入力が多少変化すると中間層のパターンががらりと変わるから，学習したサンプルのごく近傍を除いて汎化がほとんど生じないことになる．

従って，学習アルゴリズムを改良してサンプル学習能力を高めようとすれば汎化能力が低下するし，汎化能力を保とうとすればサンプルがうまく学習できない．そして，課題が複雑になり，ネットワークの規模が大きくなるほど，このジレンマは顕著になる．

これまで BP モデルに様々な改良や拡張を加えたモデルが提案されてきたが，このような限界がはっきりと認識されていたかどうかは疑問である．学習アルゴリズムだけでなく，リカレント結合を含めてネットワークの構造をどう変えても，問題の本質的な解決にはならないからである．このことは，古典的なニューラルネットの枠組みに属さないモデル，具体的には素子の出力が $f(\sum w_i x_i)$ の形で表現できない場合には必ずしも当てはまらないが，今のところこの限界を越えられることが明確に示されたものは見当たらない．

3. 選択的不感化法を適用したモデル

以上の議論は，多層パーセプトロンの限界を示すと共に，局所表現を用いることなく 1 対多対応による平均化の問題を解消できれば，ニューラルネットの情報統合能力が飛躍的に向上する可能性があることを示している．そして，著者ら [2] は先に，非単調神経回路網において，比較的簡単な方法で平均化の問題が解消されることを明らかにした．それが選択的不感化による文脈修飾法である．

3.1 選択的不感化

ここでは，2 層モデルへの適用を前提として，この手法について説明する．まず，入力部の i 番目の素子の出力 x_i は，入力パターン S の i 番目の要素 s_i そのものではなく，

$$x_i = g_i(s_i - \bar{x}_i) + \bar{x}_i \tag{1}$$

で与えられるものとする．ここで，g_i はこの素子のゲインであり，入力の変化に対する感度を表す．また，$\bar{x}_i$ はこの素子の平均出力レベル，すなわちすべての入力に対する出力の平均値であるが，g_i と s_i の間に相関がなければ，s_i の平均と置き換えられる (両者が厳密に一致していなくてもほとんど問題ない)．例えば s_i が 1 と -1 を等確率でとり，g_i が s_i とは独立に決まるとき，$\bar{x}_i = 0$ としても良い．このとき $x_i = g_i s_i^\mu$ と表されるが，以下ではこのような場合について扱う．

素子のゲイン g_i は通常 1 であるが，これを 0 にすることを「不感化」と言い，n 個の素子のある一部分だけを不感化することを「選択的不感化」と呼ぶ．そして，不感化する素子の組合せを文脈ごとに変えるというのが選択的不感化に基づく文脈修飾法である．以下，これを積型文脈修飾と呼び，C^ν で積型修飾された S^μ を $S^\mu(C^\nu)$ で表す．なお,「積型」というのは「直和型」に対する略称であって，単に積の項を導入すればよいというものではない．各素子の感度が文脈に依存する点が本質的である．また，素子の出力が単純な積の形で表されるのは，平均出力が 0 の場合だけであることに注意されたい．

次に，不感化される素子の決め方であるが，一つの素子が複数の文脈で不感化されるべきである．そうでなければ，文脈の数だけ神経回路を用意し，それを切り替えて用いるのと同じになってしまうからである．また，文脈を表現する能力を最大にするために，ある素子が不感化を受ける文脈は素子ごとに独立に選ばれるべきである．そうすると，文脈は n 次元のゲインベクトル $G = (g_1, \ldots, g_n)$ によって分散表現されることになる．そして，そのような G の最も簡単な決め方は，文脈パターン C と同一視して，例えば $g_i = (1 + c_i)/2$ (c_i は C の成分) とすることである．

なお，以下ではこのように G と C を同一視するが，入力パターンが文脈パターンとしても使われ，しばし

ば $S \simeq C$ となる場合には注意が必要である．なぜなら，このとき s_i と c_i の間に相関が生じるため，$\bar{x}_i = 0$ が成り立たなくなってしまうからである．このような場合には，G の成分を適当にシャッフルするなど，g_i と s_i が無相関になるよう対策をとる必要がある．

ところで，積型文脈修飾を用いると，なぜ 1 対多対応による結合荷重の平均化が生じないのだろうか．これを議論することは本研究の直接の目的ではないが，簡単に説明するならば次のようになる．

まず，入力層のパターンは S が同じでも C によって異なるから，直和型修飾の場合のように同一の S が複数の T と連合されることはない．これは，S 全体だけでなく，$m/2$ 個以上の成分からなるどの部分パターンについても言える．また，C は入力層の感度を変えるだけであって，文脈部と出力層との間に直接結合があるわけではない．そして $S^\mu(C^\nu)$ $(\mu = 1, \ldots, p)$ と C^ν との相関は 0 であり，間接的にも C が T と 1 対多に連合しているとは言えないから，平均化は生じないのである．

なお，不感化された素子が平均値以外の値 (例えば，素子が 0〜1 の値を取るときに 0) を出力したならば，$S^\mu(C^\nu)$ と C^ν の間に相関が残るため，平均化の問題は解消しない．ここで言う不感化とは入力の変化に対する感度を 0 にすることであって，単に素子を不活化する (出力を 0 にして計算から除外する) のとは意味が異なるゆえんである．

3.2 モデルの構造と学習容量

2 層のニューラルネットに選択的不感化法を適用したモデル (以下，積型モデルと略称) の構造を図 5(a) に示す．出力パターン $Y = (y_1, \ldots, y_n)$ は，

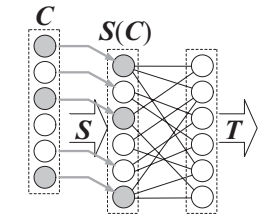
$$y_j = \mathrm{sgn}\left(\sum_i w_{ji} x_i\right) \tag{2}$$

で与えられる．w_{ji} は入力層の i 番目の素子から j 番目の出力素子への結合荷重，関数 $\mathrm{sgn}(u)$ は $u > 0$ のとき 1，それ以外で -1 を取る符号関数である．

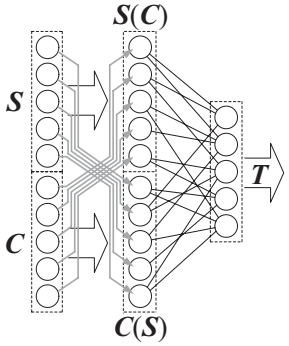
本モデルでは，入力層のパターン間の直交性が低いため，学習則として直交学習を用いる．すなわち，パターンを入力するたび

$$\Delta w_{ji} = \varepsilon(t_j^{\mu,\nu} - \sum_k w_{jk} x_k) x_i \tag{3}$$

に従って結合荷重を更新する．ここで，$t_j^{\mu,\nu}$ は目標パターン $T^{\mu,\nu}$ の成分，ε は正定数 (以下の実験では 0.3) である．



(a) product-type modification model



(b) mutual modification model

図 5　選択的不感化法を適用したモデルの構造
Fig. 5　Architecture of the models with the selective desensitization method.

このモデルに関して，**2.2** と同じ数値実験 (但し $n = 1000$) を行った．m/n に対して Y と T との平均類似度をプロットした結果を図 6 に示す．学習回数は各パターンにつき 100 回 (実線) または 1000 回 (破線) である．比較のため，直和型モデルについての結果も点線で示した．

グラフから，直和型モデルの学習容量がほとんど 0 であるのに対し，積型モデルはそれより圧倒的に大きな容量をもつことがわかる．また，学習回数が 100 回のとき $1.4n$，1000 回のときは $1.7n$ と容量が増加しており，学習回数を更に増やしていけば，2 層ニューラルネットの最大学習容量である $2n$ に限りなく近づくと思われる．このことは，積型モデルが 1 対多対応による平均化の問題から完全に逃れていることを示している．

3.3 相互修飾モデル

直交学習では，入力層のパターン間に相関があってもかまわないが，類似したパターンが多いほど，そしてその類似度が高いほど必要な学習回数が増える．積型修飾の場合，S が同じで C だけが異なるとき，類似度の期待値が 0.25 となるが，この程度の値でも m/n

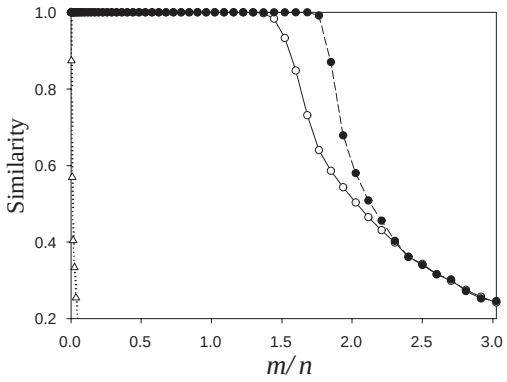


図 6　積型修飾モデルの学習容量
Fig. 6　Learning capacity of the product-type modification model.

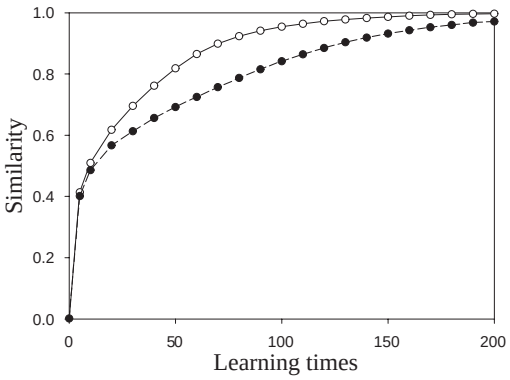


図 7　相互修飾モデルの学習速度
Fig. 7　Learning rate of the mutual modification model.

が大きければかなりの学習回数を要する．しかもこれは異なる C の間に相関がない場合であり，C が類似しているような場合には $S(C)$ の類似度が更に高くなって，学習がより難しくなる．

この問題を緩和するために考案したのが，図 5(b) に示す相互修飾モデルである．これは，S を C で積型修飾すると同時に，C を S によって積型修飾し，両者を連結したパターン $(S(C), C(S))$ を入力層に与えるというものである．こうすると，S または C の一方が同じときの入力層のパターン間の類似度が平均 0.125 まで低下するため，前述の積型モデルよりも学習しやすいと考えられる．また，S と C の対称性がよいため，一方を文脈として扱うのではなく，2 種類の情報を対等に統合するモデルとして見たとき，より適切であろう．

図 7 は，相互修飾モデル (実線) と積型モデル (点線) の学習速度を，$m = 1.6n$ の場合について比較したものである．縦軸は平均類似度，横軸は学習回数を表す．この図から，相互修飾モデルの方が積型モデルよりも学習速度が速いことがわかる．但し，前者は後者に比べて入力層の素子数および結合荷重の数が 2 倍である一方，冗長性が高いため学習容量はそれほど増えないから，入力層の素子数 $(2n)$ に対する相対容量は低い．このように，両者には一長一短があるが，性質に大きな違いはない．

3.4 汎化性能

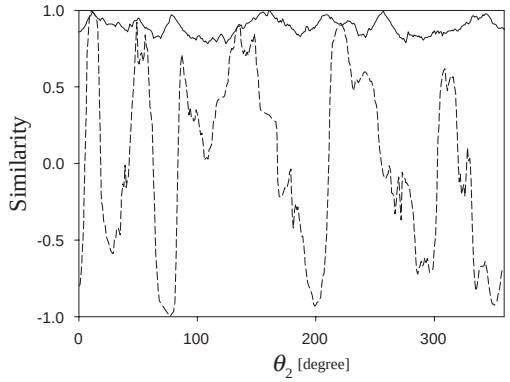
2.3 と同様の実験を相互修飾モデルについて行い，その汎化能力を調べた．相互修飾モデルを用いたのは，この課題では S と C が同種の情報だからであるが，積型モデルでも結果に大差があるわけではない．なお，S および C として用いるパターンが共通であり，しかも隣り合う成分間の相関が強いので，**3.1** で述べた理由から c_i と g_i の対応を完全にランダムに組み替える必要がある．

実験方法は **2.3** と同じであるが，学習回数は各サンプルにつき 30 回とした．m の値によってはもっと多数回の学習をしないとサンプル誤差が 0 にはならないが，汎化誤差に関しては，これ以上学習回数を増やしても同じかわずかに減少する程度であった．

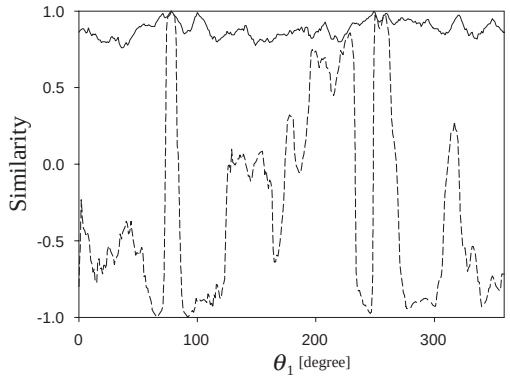
図 8 に，$n = 360$，$m = 100$ の場合の結果を示す．(a) は θ_1 を固定して θ_2 を変えた場合，(b) は θ_2 を固定して θ_1 を変えた場合である．比較のため，**2.3** における BP モデルの結果も破線で示した．学習サンプル数 $m = 100$ というのは (S, C) が取りうる 360×360 通りのパターンの約 0.08%に過ぎないが，相互修飾モデルは常に正解に近いパターンを出力しており，強い汎化が生じていることがわかる．

表 1 に各モデルの平均類似度をまとめた．この表から，相互修飾モデルと従来のモデルとの間には，特に汎化能力において大きな差があることがわかる．

次いで学習サンプル数 m と平均誤差角 E の関係を図 9 に示す．(a) は m を横軸，E を縦軸としてプロットしたグラフ，(b) は (a) の横軸を平均サンプル間隔 $\Delta\theta \equiv 360^\circ/\sqrt{m}$ に変えてプロットし直したものである．実線は $n = 360$ の場合，破線は $n = 90$ とした場合である (点線は BP モデル)．この図から，BP モデルとは異なり，相互修飾モデルの誤差 E は，m の増加と共に減少することがわかる．しかも，m が約 50



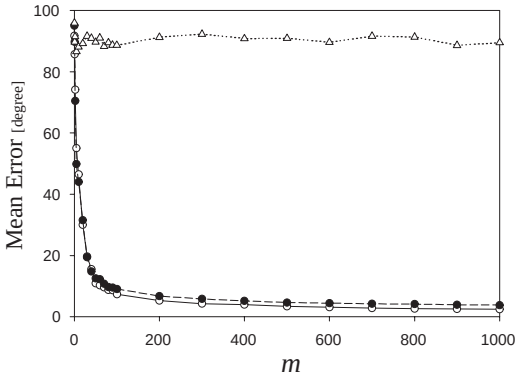
(a)



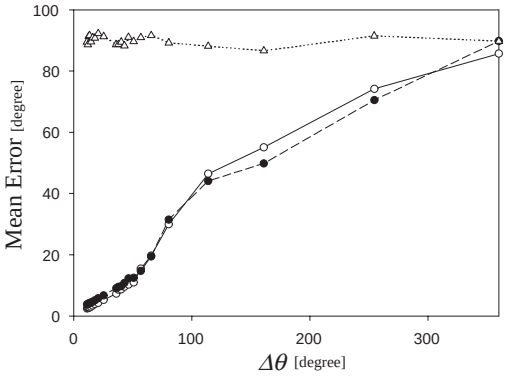
(b)

図 8　学習後のモデルの汎化性能

Fig. 8　Generalization performance of the model after learning.



(a) vs. number of samples



(b) vs. average sample interval

図 9　汎化誤差とサンプルサイズの関係

Fig. 9　Relationship between generalization error and sample size.

表 1　各モデルの平均類似度

Table 1　Mean similarity for each model.

	サンプルパターン	全パターン
直和型モデル	0.56	0.00
BP モデル	0.96	0.00
相互修飾モデル	1.00	0.88

以上のとき，E の値は平均サンプル間隔 $\Delta\theta$ にほぼ比例しており，$n = 360$ の場合で $E \simeq 0.2\Delta\theta$ という関係であった．

なお，グラフからも読み取れるように，n を 360 から 90 に減らしても，誤差はそれほど増えない (増加分は，量子化誤差の差と思われる)．一方 n を大きくしても，実験した範囲では差はほとんど見られず，必要な学習回数にも差はなかった．従って，量子化誤差が問題となるほど m を増やさない限り，これ以上 n を増やすことには意味がない．逆に，多少の誤差を許容するならば，素子数もサンプル数もかなり少なくてよいと言える．

4. 考　　察

4.1 選択的不感化法の効果

なぜ選択的不感化法によって汎化能力が向上するのだろうか．実は，これは適切な問いかけではない．なぜなら，ニューラルネットに汎化能力をもたらしているのは分散表現であって，選択的不感化ではないからである．

つまり，入出力に分散表現を用いることによって期待される汎化能力が，BP モデルでは情報統合の過程で失われていた，という見方が妥当であろう．これに対して選択的不感化法を用いたモデルでは，分散表現された 2 種類の入力情報が，局所表現やそれに近い表現を経由することなく分散表現のまま統合されるため，

本来の汎化能力が失われないのである．

この議論からわかるように，積型モデルはすべての課題に対して高い汎化性能を示すわけではない．例えば，不連続点を含む関数の近似や，特定の入力のとき出力が例外的な値をとる場合には，うまく学習できなかったり，汎化誤差が増大する．このような課題に対しては，積型モデルに中間層を導入する，あるいはBP モデルの入力層や中間層で選択的不感化を行うことが有効かもしれない．

このように，選択的不感化法の一つの大きな効果は，従来の層状ニューラルネットが発揮できなかった分散表現に基づく情報処理の潜在的能力を引き出す点にあるが，同様のことが回帰型のニューラルネットについても言える．例えば Elman ネット [3] のように 3 層ニューラルネットに回帰結合を加えたモデルは，理論上は任意の有限オートマトンを学習し模擬する能力をもつ．しかし実際には，1 対多対応による平均化の問題から逃れられないため，オートマトンが複雑かつ大規模になると，必要な素子数が増大して有限時間で学習に成功することはまずないし，仮に学習できたとしても汎化が生じない．

この問題に対しても，選択的不感化法は有効である．すなわち，非単調神経回路網に適用すれば分散表現のみに基づいて任意の有限オートマトンが模擬できることが既に示されている [2] し，Elman ネットに適用しても大きな効果があると考えられる．

そのほか，選択的不感化法には，パターン間の距離的関係を文脈に依存して変化させるという効果がある (直和型修飾では，距離的関係は変えられない点に注意)．このことは分散表現に基づく情報処理にとって非常に重要であるが，詳しくは別の機会に論じたい．

4.2 脳との関連性

ニューラルネットの工学的な価値は，脳がどのような原理に基づいているかと直接的には関係ない．しかし，もし選択的不感化による文脈修飾が脳でも行われているとすれば，この手法の有効性やそれを用いたモデルの高い将来性を間接的に示すことになるであろう．そして，著者らは実際この仮説は正しいと考えている．但し，本論文は脳の解明を直接の目的とするものではないので，ここではその根拠となる事柄を簡単に列挙するにとどめたい．なお，下記の項目 (2)〜(4) については，別の論文 [4] でより詳しく論じている．

（1） 積型および相互修飾モデルで用いている直交学習は，BP 学習に比べるとはるかに単純であり，生物学的妥当性も高い．直交学習を式 (4) に従ってそのまま実行するのは無理だとしても，例えば素子の入出力特性に非単調性 (単調な素子の組合せでも等価的に実現可能である) がある場合，Hebb 型の学習則によって近似的に直交学習が実現できる [5], [6]．

（2） 選択的不感化による文脈修飾を行うために必要な信号線の数は，素子数 n のオーダーであって，n^2 ではない．脳では離れた領域にあるニューロン同士の結合がごく限られていることを考えると，このことは重要な意味をもつ．

（3） 選択的不感化の神経メカニズムには，シナプス前抑制や神経修飾物質など，いくつもの候補が挙げられる．また，脳で選択的不感化が行われているという仮説に反する生理学的知見は，今のところ全く見当たらない．

（4） 前頭葉には刺激が同じでも文脈に依存して活動が変化するニューロンがよく見られるが，これらの多くは選択的不感化によって説明することができる．また，文脈依存的記憶課題実行時の下側頭葉ニューロンに関するデータ [7] は，選択的不感化法に基づくモデルの挙動と非常によく合致する．

（5） **2.3** や **3.4** で取り上げた課題は，実は脳における自己中心座標から他者中心座標への変換問題を念頭に置いたものでもある．例えば眼球がサッケード運動を行ったとき，視覚的注意の外部空間における位置は，眼球運動の前後で変わらない，すなわち網膜座標上では眼球の移動角だけ逆側にシフトすることが知られている [8]．眼球運動の方向と大きさは上丘の 2 次元神経場における局在興奮によって表現されると考えられているが，仮に視覚的注意の網膜座標における位置も同様に表現されており，両者の引き算で眼球運動後の注意の位置が求められているとするならば，この問題は先の課題とほぼ同じである (この仮定の妥当性は疑問であるが，いずれにせよ座標がニューロン集団の活動パターンで表現されているならば，計算には本質的に同じ困難さがある)．これ以外にも，本論文で扱ったような課題は脳の情報処理のいたるところで現れる．脳はこれらを解いているわけであり，その手段として選択的不感化法を用いている可能性は十分にあると言えよう．

（6） 脳において，異なる種類の情報の統合に関係すると見られる領域の一つに，海馬がある．例えば，ラット海馬のいわゆる場所ニューロンは，場所だけでなく課題や状況によって反応が変化するし，海馬を破壊し

た動物は，複数の情報の組合せを学習する課題や文脈ごとに異なる連合を学習する課題がうまくできない [9]．そして，嗅内野 (EC)→ 歯状回 (DG)→CA3→CA1 という海馬体の層状構造 (但し EC から CA3 へは直接投射もある) は，相互修飾モデル (図 5) との整合性が高い．しかも，DG から CA3 へ投射する苔状線維はかなり特殊な性質をもっており，CA3 ニューロンの感度を調節するのに好都合である．例えば，ニューロンの興奮性を調節する作用があると考えられる Zn^{2+} を大量に含有し，シナプス活動に伴い放出する [10]．また，1 個の DG ニューロンは巨大なシナプスを介してごく少数の CA3 ニューロンとだけ結合している．これらの事実は，CA3 において選択的不感化による文脈修飾が行われている可能性を示唆する．

5. むすび

従来の多層パーセプトロンは，2 種類の独立な情報の統合が必要な課題に関して，サンプルの学習容量と汎化性能の両方を高くすることはできない，という限界があることを示した．また，層状ニューラルネットに選択的不感化による文脈修飾を適用したモデルは，この限界を越えられることを示した．特に汎化性能に関して，二つの分散表現を分散表現のまま統合しているため 2 次元的な汎化が生じ，課題によっては学習に必要なサンプル数を大幅に減らせることがわかった．

本研究の結果は，選択的不感化法の有効性を示すと共に，ニューラルネットによる情報処理の可能性を大きく拡げるものである．著者らは，これによって分散表現の利点を生かし，記号処理に基づく古典的人工知能の限界を打破することも視野に入れている．また，**4.2** で述べたことなどから，選択的不感化による情報統合が脳の神経回路でも行われていると考えているが，もしそうであるならば，そこには脳のように高度かつ柔軟な知能システムを実現する一つの重要な鍵があるに違いない．

今後，本モデルの学習および汎化能力について更に解析を進めると共に，上述したような可能性について，ニューラルネットの工学的研究と脳の計算論的研究の両面から追求していきたいと考えている．また，本モデルの特長を生かした応用や，3 つ以上の情報を統合するための具体的方法の検討も今後の課題である．

謝辞　本研究の一部は，文部科学省の科学技術振興調整費による「文脈主導型，認識・判断・行動機能実現のための動的記憶システムの研究」の一環として行われた．また，同科学研究費補助金基盤研究 (B)(No.15300068) および特定領域研究 (No.16016207) の補助を受けた．

文　献

[1] D.E. Rumlehart, G.E. Hinton, and R.J. Williams, “Learning representations by back-propagating errors,” Nature, vol.323, pp.533–536, 1986.

[2] 森田昌彦，松沢浩平，諸上茂光，“非単調神経素子の選択的不感化を用いた文脈依存的連想モデル，” 信学論 (D-II), vol.J85-D-II，no.10，pp.1602–1612，Oct. 2002.

[3] J. L. Elman, “Finding structure in time,” Cognitive Science, vol.14, pp.179–211, 1990.

[4] 末光厚夫，諸上茂光，森田昌彦，“下側頭葉における文脈依存的連想の計算論，” 信学論 (D-II), vol.J87-D-II，no.8，pp.1665–1677，Aug. 2004.

[5] 森田昌彦，吉澤修治，中野　馨，“非単調ダイナミクスを用いた構造をもつパターンの連想記憶，” 信学論 (D-II), vol.J75-D-II，no.11，pp.1884–1891，Nov. 1992.

[6] 大本智幸，小泉耕二，岡田真人，“非単調ニューロンを用いて相関学習された連想記憶モデルの S/N 解析，” 日本神経回路学会誌，vol.8，no.3，pp.86–93，2001.

[7] Y. Naya, K. Sakai, and Y. Miyashita, “Activity of primate inferotemporal neurons related to a sought target in pair-association task,” Proc. Natl. Acad. Sci. USA, vol.93, pp.2664–2669, 1996.

[8] O. Hikosaka, S. Miyauchi, and S. Shimojo, “Orienting of spatial attention—its reflexive, compensatory, and voluntary mechanisms,” Cognitive Brain Research, vol.5, pp.1–9, Dec. 1996.

[9] A.D. Redish, Beyond the Cognitive Map, MIT Press, 1999.

[10] S. Ueno, M. Tsukamoto, T. Hirano, K. Kikuchi, M.K. Yamada, N. Nishiyama, T. Nagano, N. Matsuki, and Y. Ikegaya, “Mossy fiber Zn^{2+} spillover modulates heterosynaptic N-methyl-D-aspartate receptor activity in hippocampal CA3 circuits,” J. Cell Biol., vol.158, pp.215–220, 2002.

（平成 16 年 1 月 5 日受付，5 月 30 日再受付）

森田　昌彦　（正員）

昭 61 東大・工・計数卒．平 3 同大大学院博士課程了．日本学術振興会特別研究員，東京大学工学部助手を経て，平 4 筑波大学電子・情報工学系講師．現在，同大大学院システム情報工学研究科助教授．生体の情報処理機構および神経回路網の研究に従事．平 5 日本神経回路学会研究賞，平 6 同学会論文賞，平 11 日本心理学会研究奨励賞受賞．

村田　和彦　（学生員）

平 14 筑波大・工学システム学類卒．平 16 同大大学院システム情報工学研究科博士前期課程了．現在，PFU（株）勤務．在学中，神経回路モデルの研究に従事．

諸上　茂光　（学生員）

平 12 筑波大・工学システム学類卒．現在，同大大学院博士課程システム情報工学研究科在学中．神経回路モデルの研究に従事．

末光　厚夫　（正員）

平 10 筑波大・工学システム学類卒．平 15 同大大学院博士課程了．現在，同大学院システム情報工学研究科研究員．神経回路モデルによる脳の記憶機構の研究に従事．