

ニューラルネットワークの破壊実験によっ
て表現された神経心理学的モデルの考察

浅川 伸一

目次

第1章 序論	1
1.1 神経心理学における道具としてのニューラルネットワーク	1
1.2 脳の構成論的研究と人工脳損傷	2
1.3 ニューラルネットワークの特徴	3
1.4 新しい脳観	4
1.5 この分野の初期の研究が持つ欠点を克服するために	5
1.6 神経心理学のツールとしてのニューラルネットワークがみたすべき要件	5
1.7 アルゴリズムと計算論	6
1.8 本論文の構成	7
第I部 神経心理学の基礎概念及び推論方法への批判	8
第2章 局在性仮説批判	9
2.1 システムの等価性	10
2.2 システムの構造と解釈	11
第3章 二重乖離の原則批判	13
3.1 アルツハイマー病の呼称課題における視覚性および語彙性の誤りに関する二重乖離	13
3.1.1 モデル	13
3.1.2 視覚性障害仮説についてのシミュレーション	14
3.1.3 語彙の頻度効果に対するシミュレーション	15
3.2 生物, 非生物概念の二重乖離	15
3.2.1 神経心理学的症状	15
3.2.2 モデル	16
3.2.3 破壊実験	17
3.2.4 ニューラルネットワークによる障害の再解釈	18
3.3 機能的失文法と統辞的失文法の二重乖離	18
3.3.1 失文法 agrammatia の種類	18
3.3.2 モデル	19
3.3.3 シミュレーション	20

3.3.4	文法知識の創発と失文法	22
3.4	まとめ	23
第II部 ニューラルネットワークによる神経心理学的モデルの検討		24
第4章	2段階読字過程モデル	25
4.1	モデル	26
4.2	計算機実験	27
4.2.1	材料, 原画像	27
4.2.2	前処理	28
4.2.3	結果	30
4.3	破壊実験による脳損傷のシミュレーション	32
4.3.1	離断仮説	33
4.3.2	視覚性障害説	34
4.4	考察	34
第5章	単語の読みにおける労働の分割問題に対する解としてのエキスパート混合ネットワーク	36
5.1	失読症のニューラルネットワークモデル	37
5.2	英単語の特徴	37
5.3	3つの失読症	37
5.4	二重経路モデル	38
5.5	トライアングルモデル	38
5.6	論争	40
5.7	労働の分割問題	40
5.8	エキスパート混合ネットワークによる労働の分割問題の解	41
5.9	労働の分割問題の3つの実装モデル	43
5.9.1	素朴な労働の分割	43
5.9.2	競合的労働の分割	44
5.9.3	ガウシアン混合モデルによる労働の分割	44
5.10	数値実験	45
5.11	考察	46
第6章	大脳基底核と皮質領野の相互作用を仮定したモデルによるリハビリテーションのシミュレーション	48
6.1	はじめに	48
6.2	大脳基底核を含む二重経路モデル	49
6.2.1	大脳基底核の自己組織化	51
6.2.2	皮質系の経路と大脳基底核経路との統合	52

6.3	リハビリテーション	52
6.3.1	臨床症状とネットワークの破壊の関係	52
6.3.2	破壊と回復との関係	52
6.3.3	リハビリテーション	53
6.3.4	可能なリハビリテーション技法の組み合わせについて	54
6.3.5	概念系の損傷	55
6.3.6	媒介系の損傷	56
6.4	考察	57
第 III 部 総合考察		60
第 7 章	ニューラルネットワークを神経心理学へ適用する意味について	61
7.1	ニューラルネットワークモデルと症例研究	61
7.2	脳の構成論的研究の意義	61
7.3	人工脳損傷の意義	63
7.4	モデルの満たすべき要件	63
7.5	背景となる数学的な基礎の整備の必要性	66
第 8 章	結語	67
8.1	要約	67
8.2	おわりに —新しい脳観を求めて—	68
8.3	謝辞	69
付 録 A	エキスパート混合ネットワークの数学的解説	70
A.1	定式化	70
A.2	ridge 回帰	72
A.3	ME の確率的解釈	72
A.4	パラメータの推定	73
A.5	EM アルゴリズム	74

第1章 序論

1.1 神経心理学における道具としてのニューラルネットワーク

本論文は、神経心理学的症状を理解するためにニューラルネットワークを用いることを提案するものである。ニューラルネットワークを用いた神経心理学的研究とは、神経心理学に対してニューラルネットワークを応用することにより、計算論的観点から神経心理学的症状の理解を目指すものである。ニューラルネットワークを用いることによって神経心理学的症状のいくつかはまったく別の視点で捉え直すことができることを主張する。

自己増殖オートマトンの理論から脳とコンピュータの同型性に基けば、人間の行動とニューラルネットワークの挙動は同一視できる。すなわち、脳とコンピュータとは数学的には同じオートマトンである。von Neumann は晩年「機械は成長したり、増殖したりすることは可能か」という自己増殖問題を考えた(フォン・ノイマン, 1975)。彼は細胞モデル (cellular model) 上での機械の自己増殖問題を考えた。2次元の細胞モデルでは、無限の同じ構造をしたセル(細胞)が格子状に配列されている広い平面を仮定する。von Neumann は各細胞にチューリングマシン(有限オートマトン)を割り当て、無限に広い空間の中で機械が別の場所に自分と同じ機械を作ることができるか、すなわち自己増殖が可能かどうかを検討し、自己増殖が可能であることを示そうとした。von Neumann は最も簡単な細胞モデルの要素として McCulloch と Pitts の形式ニューロン (McCulloch & Pitts, 1943)(1.3節に簡単な説明を記した)を用いていることから、現在のコンピュータ (Neumann 型コンピュータすなわち有限オートマトン) と神経細胞の集合である脳とは同一視できる。2次元細胞モデルでの自己増殖問題の解は、脳の皮質上の任意の場所に任意の精度で計算可能なコンピュータが実現できることを意味していると考えることができる。

すなわち、細胞モデルで自己増殖が可能であるということは、神経心理学の分野で提案されているモデルはすべてニューラルネットワークモデルとして書き下すことができ、コンピュータ上に実現できることを意味する。いったんコンピュータ上に実現されたモデルを系統的に破壊することで神経心理学的症状の理解を試みるのが神経心理学の手段としてニューラルネットワークを用いる場合の標準的な手法となる。コンピュータ上では実際の患者には

要求できないような極端な要求や課題を課すことができ、その結果得られるデータも現象の理解に役立つ。こうして得られた知識や予測をもう一度症例の検討に戻って考えてみる、という研究手法がこの分野の研究を一步前に進める力になりうる。本論文ではこのような脳とコンピュータの同型性を利用した脳損傷のモデルを取り上げ、ニューラルネットワークモデルの神経心理学的応用可能性について議論する。

1.2 脳の構成論的研究と人工脳損傷

近年、脳はさまざまな方法で研究されている。fMRI に代表される機能的脳画像研究、ネコやサルにマイクロ電極を差し込んで細胞の動作を測定する電気生理学的手法、動物を用いた脳の破壊実験、脳波、薬理学的方法、神経心理学と呼ばれる障害を持った脳の観測、心理学実験などである。これらの方法に加えて脳の構成論的研究、すなわち、ニューラルネットワークと呼ばれる脳のモデルを作って、このモデルが実際の脳と同じ機能を果たしていると考えられるのならば、モデルの持っている機構が脳にも存在する可能性があるとする研究分野がある。

脳の構成論的研究とは、このようなモデルのひとつであるニューラルネットワークモデルを作ってコンピュータによるシミュレーションを行ない、シミュレーションを通して、脳の機能を類推してゆく研究である。脳の構成論的研究で重要なことは、解剖学的な事実や生理学的に分かっていることとはできるだけモデルに取り入れ、未だその役割や機能、計算論的意味付けが十分に分かっていないことについては大胆に仮定してモデル化を行なうことである。こうしてできあがったニューラルネットワークをシミュレーションによって研究するのが脳の構成論的研究である。

神経心理学的データは脳内の特定の部位と特定の認知機能との関連付けに対して、比較的単純な対応関係を明らかにしてくれるが、本論文の第一部で指摘するように、複雑な相互作用をする脳では、このような神経心理学的な考察だけでは十分とは言えない。

一方、脳の構成論的研究のツールであるニューラルネットワークは相互作用をする複雑なシステムとしての脳の振る舞いと、そのシステムが損傷したときの効果との関係を推論する手段を提供している。そのような推論が明示的で機能論的に検証できる、すなわちシミュレーションによる検証が可能になったことが重要なのである。

ニューラルネットワークを用いた脳損傷のシミュレーション研究では特定の認知機能を遂行するためのニューラルネットワークモデルをコンピュータプログラムとして実現し、構築されたニューラルネットワークの一部を破壊することによって対応する部位が損傷を受けたときに生ずる症状をプログラムの出力として表現することをめざしている。こうした研究はいわば人工脳

損傷とも言える研究である。人間の認知機構を理解する上でも、あるいは実際の脳損傷患者の症状を理解するためにも、コンピュータを用いた人工脳損傷研究は重要だと考えられる。倫理上の制約から実際の人間の脳を破壊して実験を行なうことは不可能だからである。ニューラルネットワークによる人工脳損傷研究は、実際の脳損傷患者を扱う神経心理学に対して強力な道具を提供していると言える。

1.3 ニューラルネットワークの特徴

ニューラルネットワーク—あるいはコネクショニストモデル、並列分散処理 (PDP) モデルとも呼ばれる—とは脳の神経細胞 (ニューロン) の動作を抽象化した表現であるユニットの集合を用いて人間の情報処理のモデル化をめざす研究分野である。脳内では莫大な数のニューロンが互いに密接に結合されており、ニューロンのネットワークを構成している。これがニューラルネットワークである。

ニューロンの動作を抽象化して考えれば比較的簡単な図式で理解できる。ニューロンとニューロンとの間はシナプス結合と呼ばれる結合によって結ばれている。実際のシナプスにはシナプス間隙と呼ばれるわずかな隙間があって、このシナプス間隙の間を神経伝達物質が放出 (入力を受け取る側から見れば吸収) されて情報伝達が行なわれる。シナプス結合には、この情報伝達の種類と効率に従ってシナプス結合強度が各シナプス毎に異なる。正の結合強度なら興奮性の、負の結合強度ならば抑制性の結合となる。結合が興奮性か抑制性かの違いは神経伝達物質の違いによる。情報を受け取る側のポストシナプスニューロンは、この伝達強度で重み付けられた他のニューロンからの情報を時間的、空間的に加算して、この値が一定の強度に達するとニューロンの電位が急激に変化する。このときニューロンが活性化したとか興奮したとかいう。ニューロンが活性化する値のことをしきい値という。しきい値が高くなれば活性化しにくく、反対に低ければ活性化しやすくなる。このニューロンの活性値がシナプスを介して他のニューロンに伝達されることによって脳内の活動、ひいては言語などの高次認知機能が発現する。

もっとも簡単な McCulloch と Pitts の形式ニューロン (McCulloch & Pitts, 1943) は、活性化した状態を 1 で表し、そうでなければ 0 の値をとる 2 値関数、従って真か偽かを表す命題関数と同一視できるモデルである。この場合のしきい値は 0 である。そのニューロンが活性化するか否かは、出力を送るニューロンの活性値にそのシナプス結合強度を掛けた値の合計値が 0 より大きいか否かで決まるというのが McCulloch と Pitts のニューロンモデルである。ニューラルネットワークとはこのような抽象化されたニューロンの表現 (ユニットと呼ばれることもある) を用いて人間の知的活動をモデル化する研究分野である。

人間の行なうさまざまな行為はすべてニューロンの活動とニューロン間の結合の強度として表現可能であると考えるのがニューラルネットワークである。ニューラルネットワークにおける特徴を挙げるとすれば、分散表現と統計的構造の漸進的学習、および相互作用の3点に要約できる (Farah, 1994; Plaut, 2001)。

分散表現: ニューラルネットワークにおいては、知識はそれぞれのユニット集団の活性化パターンとして表現される。例えば、ある単語の意味は別の単語の意味とは異なる活性化パターンとして表現されており、類似した概念は互いに類似した活性化パターンとして表現される。

統計的構造の漸進的学習: 長期的な知識はユニット間の関係、すなわちユニット間の結合強度としてネットワーク内に埋め込まれている。ユニット間の結合強度は学習によって徐々に変化する。すなわち、学習により外界から提供される情報(単語間の類似度や相互関係など)の統計的性質が徐々に獲得される。

相互作用: ユニットの結合は密接に連結されており、相互に影響しあう。すなわちユニット間の結合強度に応じて、互いに活性化パターンを強め合ったり、弱め合ったり、振動したりというような複雑な相互作用をする。

1.4 新しい脳観

コンピュータ上のモデルとして表現された神経心理学的現象について、脳内で対応する機能が存在するのかという疑問がある。脳の構成論的研究者の中には、情報論的必然性という概念を用いてこの分野の研究に理論的根拠を主張する人達もいる(中野, 1995)。情報論的必然性とは入出力関係が脳とモデルとの間で一致したとき、そのメカニズムも一致している可能性が高いというものである。入出力関係が複雑で巧妙であるほど、それを実現する情報処理の機構はそういくつもあらずが無いという主張である。この情報論的必然性が完全であるとは思えないが、むしろ脳の構成論的研究—シミュレーション—の意義は、今までの神経心理学的証拠から導き出された脳観に疑問を投げかけ、新しい脳観を求めていることではないかと思う。

山鳥(1985)などによれば、伝統的な神経心理学では二重乖離の原理(principle of double dissociation)、離断の原理(principle of disconnection)によって特定の認知機能の機能局在が論じられてきた(第一部の批判を参照)。この立場は特定の認知機能の責任領域が、任意の部位に局在していることを暗黙のうちに仮定している。しかし、ニューラルネットワーク研究の破壊実験では単一部位の損傷によって複数の認知機能障害を同時に説明できる可能性がある(第一部第3.1節を参照)。反対に、複数の部位を破壊または離断しなければ高次認知機能障害は発生しないかも知れない。特に高次認知機能におい

ては、シミュレーションによって確認してみるまで分からないという側面が強い。実際の患者に対して破壊実験を行なうことは不可能であるから、コンピュータ上での破壊実験は、病態の真の理解のためには欠かせない研究である。わが国においてはこのようなシミュレーション研究は極めて少ないと言わざるを得ない。

1.5 この分野の初期の研究が持つ欠点を克服するために

Hopfield (1985) や McClelland と Rumelhart ら (1986) の研究をきっかけに 1980 年代後半に起こった第 2 次ニューロブームと呼ばれる動きが始まった直後からニューラルネットワークを神経心理学へ応用する研究が始まった (Patterson, Seidenberg & McClelland, 1989; Farah & McClelland, 1991, など)。実際に Rumelhart らの本の中にも脳損傷とニューラルネットワークの関係の記述が見られる。1980 年代後半から 1990 年代の始めにかけての初期の研究では、この分野が開拓される黎明期であったという状況ゆえに、ニューラルネットワークモデルの応用可能性を強調するあまり、モデルの数学的検討が十分であるとは言えなかった。極論すればプログラムを書いて動きさえすれば論文になった時代である。現在ではこのような黎明期を過ぎ、より厳密な研究が必要とされている。すなわち、モデルの数学的な限界、計算論的な意味、適用範囲の限定などを考慮したモデルが求められている。本論文ではこのような点を考慮し、この分野が健全で実りあるものにするために、計算論的考察を重視している。

1.6 神経心理学のツールとしてのニューラルネットワークがみたすべき要件

上坂 (1993) などによれば、ニューラルネットワークは非常に記述能力が高い。そのためニューラルネットワークモデルを作れば、ほとんどの心理現象はシミュレートできてしまう。例えば 3 層パーセプトロンにおいて、中間層の数が十分に多ければ任意の精度で任意の関数を近似可能である。このことを 3 層パーセプトロンの関数近似機としての完全性という (Bishop, 1995)。

このようなパーセプトロンの強力な性質は、神経心理学ツールとしてのニューラルネットワークにとって重大な制約を課すことになる。すなわち、この数学的な拘束によって、外界の刺激入力に対して応答するような、心理学で用いられるパターン識別、認識、関数などは 3 層パーセプトロンで説明できることを意味する。これらの心理現象が説明可能であることは、単にシミュレーションを行なって動作するだけのモデルは、必ず作ることができること

を意味する。従って、単なるシミュレーションそのものを目的としたようなモデルは、生理学的、神経心理学的な妥当性を持たず、神経心理学の手段としてのニューラルネットワークの意味が無い。また1.5節でも述べたとおり、ただ動作するニューラルネットワークモデルを作れば論文になった時代は終わった。モデルの計算論的な意味付けができないようなニューラルネットワークモデルでは神経心理学に対して理論的な貢献は期待できないであろう。モデルが神経心理学に貢献するためには、以下のような要件を満たしている必要があるだろう。

1. 無矛盾性。生理学や機能的脳画像研究から得られたデータと矛盾してはならない。
2. 予測可能性。モデルは当該の症状を説明するだけでなく、症状が進行したときの予測や、新たな症状のおきる可能性を指摘できなければならない。
3. 説明可能性。ニューラルネットワークモデルによって複数の心理指標が同時に説明されなければならない。

出来上がったモデルが真に説得力を持ち、神経心理学の強力なツールとなるためには、計算論的意味を考察することが必要となる。

1.7 アルゴリズムと計算論

本論文では、ニューラルネットワークと計算論とを区別して考える。Marr(1982)の三つの説明水準に照らして考えれば、ニューラルネットワークはアルゴリズムの水準であって、計算論の水準でも、物理的実装の水準でもない。ニューラルネットワーク研究が成り立つのは、計算論の水準や物理的実装の水準がアルゴリズムの水準とは独立であることが前提だからである。1.1節で述べた脳とコンピュータの同型性の議論から、物理的実装については脳とニューラルネットワークとの違いを考慮する必要は無い。だが、アルゴリズムのレベルであるニューラルネットワーク研究は、計算論レベルの検討とは独立に成り立つため、計算論の水準での検討も考慮する必要がある。1.4節で述べた情報論的必然性という概念は、入出力の関係が巧妙であればある程、そこに至るアルゴリズムはそう多くはない、ということを主張しているだけであって、そこに計算論的必然性は含まれていない。従ってニューラルネットワークを神経心理学に応用する研究においては、単に特定のアルゴリズムを用いて動作するニューラルネットワークモデルを作るだけでなく、モデルの持つ計算論的意味が明確なモデル構築を目指す必要がある。この目安が上記の3つの要件と言うことになるだろう。ニューラルネットワークが単なるモデルではなく、神経心理学の理論として成り立つためには計算論的

な考察が必要である。こうした点から考えれば、神経心理学にとって計算論的な考え方は重要な要素である。本論文で展開する議論は神経心理学に対してニューラルネットワークを通してモデルの計算論的役割の理論的な根拠を与える試みである。

1.8 本論文の構成

本論文の第一部では、伝統的な神経心理学の原則をニューラルネットワークから見た場合の批判を取り上げる。二重乖離の原則に基づく局在性仮説は神経心理学の原則であり、基本的なモデル構築のための推論方法である。この原則および推論方法に対してニューラルネットワークから見れば別の視点が提供できることを示す。

第二部では、応用的なモデルの幾つかを取り上げ実際にどのような研究が可能かを議論する。具体的には、日本語の文字知覚のための2段階読字過程モデル、単語の読みにおける労働の分割問題の解としてのエキスパート混合ネットワーク、リハビリテーションへの応用と3種類のリハビリテーション技法へのニューラルネットワーク的な意味付けについてである。

本論文の第三部では総まとめとして第一部と第二部の議論を元に脳の構成論的研究の重要性を強調する。脳の構成論的研究と症例研究は車の両輪でなければならず、そのためにニューラルネットワークモデルが満たすべき条件を議論し、さらに数学的基礎の整備の必要性を主張する。

第I部

神経心理学の基礎概念及び推
論方法への批判

第2章 局在性仮説批判

第一部では伝統的な神経心理学の概念および推論方法と言ってもよい二重乖離 (double dissociation) の原則に基づく局在性仮説 (locality assumption) についてニューラルネットワークから見た批判を展開する。最初に認知モジュールが独立であると仮定する局在性仮説に対する批判を取り上げ、次に、二重乖離の原則を説明する別解釈が可能であることを3例示す。第一部の議論によって新しい神経心理学の見方であるニューラルネットワークモデル論の必要性を説明することにした。

神経心理学の分野では二重乖離の原則に基づく機能局在性仮説が重要な役割を果たしてきた。例えば第二部 5.3 節において説明するように、低頻度例外語の読みだけが選択的に障害される表層失読症と呼ばれる症状を呈する患者が存在する。これに対して、非単語が読めない音韻失読症と呼ばれる患者も存在する。表層失読症の患者は非単語を読むことができ、音韻失読症の患者は低頻度例外語を読むことができる。このようにある特定の認知機能 A だけに障害を示し、別の認知機能 B は正常に保たれている患者が存在し、逆に認知機能 B だけが選択的に障害され、認知機能 A は正常の範囲である患者も存在する場合、認知機能 A と認知機能 B とは二重に乖離しているという。二重乖離の現象が観察された場合、それぞれの認知機能に対応する脳内機構が存在し、それら脳内の認知機構は他方の影響を受けずに独立に機能すると主張するのが二重乖離の原則である (山鳥, 1985 など)。このような二重乖離を示す例は、上記の例外語と非単語の例ばかりでなく、日本語における仮名と漢字の乖離 (第二部 4 章の考察でも言及した)、意味記憶における生物と非生物の知識の乖離 (3.2 節)、具体語と抽象語の知識の乖離、自発話において名詞のみが損傷される失名辞失文法と助詞や冠詞が抜け落ちる機能的失文法との乖離 (3.3 節)、文章理解と文法性判断の乖離など、さまざまなものがある。こうした認知機能の独立性と脳の画像診断結果との対応を考察することが神経心理学における標準的な研究方法であった。

機能局在性仮説とは、心の機能が独立した下位モジュールによって構成されているとする考え方である (Fodor, 1983)。局在性仮説の背景には、脳内の認知機能は完全に独立したモジュールによって営まれており、各認知モジュールは専用のハードウェアによって実現されているという仮定がある。そしてこのモジュール間では、比較的簡単な情報を伝達するにすぎず、各モジュールは情報論的にカプセル化 (encapsulated) されていて、あるモジュールは他

のモジュールとは単純な情報を伝え合うだけであるという局在性の強い仮定に基づいて。従って、あるモジュールの損傷が別のモジュールの機能に影響を与えることはほとんどなく、あるモジュールの損傷は比較的単純なひとつの認知機能の低下として表出する、というのが局在性仮説である。

しかし、この局在性仮説は単純すぎる (Farah, 1994)。1.3 節でも述べたとおり、モジュール間の相互作用はニューラルネットワークの特徴である。もし仮に、我々の認知機構の各モジュールが相互に関連し依存しあい、複雑な相互作用をするのであれば、一つの認知モジュールへの障害は他の認知モジュールにも影響を与える可能性がある (3.1 節も参照)。もちろん、脳損傷の効果が損傷を受けた認知モジュールだけに限局され、損傷の影響は情報論的にカプセル化された認知システムの中の要素だけに限定されると仮定する局在性仮説が正しい可能性はある。だがしかし、脳はそのような情報論的にカプセル化されたシステムであることの証拠が提出されたことは一度もない。むしろ最近の PET や fMRI に代表される機能的脳画像研究によって明らかになったことは、単純な課題を遂行しているときでさえ、多くの部位に活性化が見られることである。すなわち、特定の認知機能の検査結果とその認知システムの障害の部位を特定することの関連は、それ程明らかな関係にはないと思われる。

Arbib(1989) は次のように書いている。ニューラルネットワークの破壊実験、切除実験は重要な情報を与えてくれるが、その解釈には注意を要する。「ラジオから抵抗を取り外したらピーツという音が鳴ったからといって、その抵抗がピーの抑制中枢であるとはいえない。ほとんどの人は噂話が好きである、グループはある人がそのグループを離れると、その人の噂話をする。このときその人をそのグループの噂話抑制中枢とみなすようなものである。」

Arbib の記述から具体例を挙げてこのことを考察してみよう。

2.1 システムの等価性

Arbib(1989) からの引用が長くなるが、一部改変して記す。「下図 2.1 のニューロン n_1 とニューロン n_2 はいずれもしきい値が 0 であるとし、入力そのまま出力となるとしよう。ニューロン n_3 はニューロン n_1 や n_2 とは全く無関係にこの回路に信号を送っているとする。ニューロン n_4 はしきい値が 1.5 であるので、二つ以上の入力があったときに活性化する。したがって x_3 に入力がなければ、この回路全体の出力は x_1 と x_2 とが同時に入力されたときだけ活性化する。他方、もし x_3 からの入力があれば、この回路の出力は、 x_1 または x_2 からの入力があれば活性化する。すなわち n_3 は、 x_1 と x_2 とが AND 回路 (論理積) として動作するか OR 回路 (論理和) として動作するかを決定する制御ニューロンと見なすことができる。このシステムを実験者が外から見ると、 x_1 と x_2 はシステムへの入力ニューロンであることが分かる

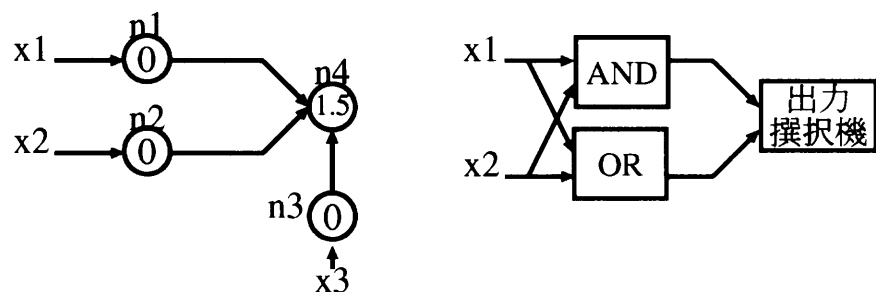


図 2.1: 同じ機能でも内部構造が異なる 2 つのシステム. 入出力は両者とも同じなので, 損傷実験をするか内部変数をモニターしないかぎり分離できない

が, どこか別のサブシステムから来ている AND と OR を計算する別々のサブシステムがあり, そのどちらかをシステムの出力として選択する出力選択機があると思つて, 右のようなブロック回路を描くことも可能である. すなわち外部から観察している限りこの二つのシステムは分離不可能である. 入出力は両者とも同じなので, 損傷実験をするか内部変数をモニターしないかぎり分離できない.]

神経心理学は, こうした破壊を取り扱う研究分野であり, 神経心理学的事実を考慮した構成論的研究であるニューラルネットワークは脳のメカニズムに迫る有力な研究手法であるということができよう. 1.4 節で紹介した情報論的必然性と神経心理学とは, この点において結びつく. 健常な脳だけを観察していることに相当する認知心理学実験だけでは図 2.1 の左右どちらのシステム図が正しいのか区別できない. 損傷実験に相当する神経心理学的観察が必要である. なんらかの方法で破壊実験(脳損傷), あるいは内部変数のモニターを行なわない限り図 2.1 の左右の区別はできない. 情報論的必然性の正当性が主張されるためには, 神経心理学的事実と脳の構成論的研究(ニューラルネットワークによるシミュレーション)と協調が必要である.

2.2 システムの構造と解釈

さらに Arbib(1989) からの引用であるが, 「一つのシステムの行動が完全にわかったとしても, そのことから構造が一義的に決まるわけではない. システム S を分析して, それが行動的に二つのシステム S_1 と S_2 の結合とみなすことができ, またそのように分解するとシステム S がよく説明できるにしても, それをもってシステム S の構造を S_1 と S_2 の二つのシステムに機械的に分解することはできない. もし心をいくつかの心的過程に分解できたとしても, その心的過程を脳の別々の部位に帰することはできないし, その逆も成り立たない. 観察された行動のもとにある内部相互作用の詳細を説明

する前に、内部構造と状態についての解剖学と生理学に留意しなければならない。

ある機能がないからと言って、その脳部位が無活動であるとは言えない。何かの異常によってニューロン n_3 が発火し続けるとニューロン n_4 はいつも OR 回路になり、AND 回路にはならない。したがってある機能が働かないということは、その神経回路が働かないということではなくて、ある仕方で働かないということなのである。すべてのニューロンが活動し、大筋においてすべての神経路が適切に結合していても、小さな結合やしきい値に異常があると異常行動となる。ラジオのピー音抑制の例を考えればわかりやすいだろう。脳構造の意味づけは慎重に行なう必要がある。」

上記のようにシステムの評価、とりわけ破壊の効果をそのシステムの機能や性能と結びつけるには慎重な調査、観察が必要である。だが、驚くべきことにこのような認識が生まれたのは比較的最近になってからのことである。神経心理学は、言語野の存在を示した Broca 以来 100 年以上の歴史を持ち、歴大な数の観測データがあるにもかかわらず、本節に示したようなシステムの解釈に関する考察が行なわれるようになったのはここ十数年でしかない。これから多くの研究がなされなければならない。

第3章 二重乖離の原則批判

3.1 アルツハイマー病の呼称課題における視覚性および語彙性の誤りに関する二重乖離

アルツハイマー病の初期症状の一つとして、物体呼称課題における障害が挙げられる (Tippett & Farah, 1998)。アルツハイマー病の患者は椅子のことをテーブルと答えたり、梨のことを果物といたりする。このことは意味記憶の構造を考える上で興味深い。同一カテゴリーに属する別のメンバーであると考えられる椅子とテーブルを間違えることと、具体的な事物を表す梨をより上位概念である果物と答えることが起こっている。このことからアルツハイマー病の意味記憶構造にはどのような障害を考えればよいのだろうか。

アルツハイマー病の患者の物体呼称課題における障害には、上記の意味記憶の障害説以外にも、二つの別の説明が存在する。一つめは、視覚刺激の質を低下させたときに、例えば写真を用いた物体呼称課題と線画を用いた物体呼称課題で、線画を用いた方が呼称成績が悪いのである。すなわち、視覚性の障害と解釈できる可能性がある。二つめは語彙性の障害の可能性である。物体呼称課題において、頻度効果が存在する。高頻度語の想起の方が、低頻度語の想起よりも成績がよい。

アルツハイマー病患者の物体呼称課題における成績低下は、意味性、視覚性、語彙性のいずれの障害によって引き起こされるのであろうか。個々の患者ごとに障害の場所が異なると言う説明の仕方ももちろん可能ではある。

しかし、Tippett と Farah(1998)によるニューラルネットワークを用いた研究によれば、意味記憶の障害のみによってアルツハイマー病の物体呼称課題における成績低下を説明できる可能性がある。ニューラルネットワークの特徴は知識の分散表現と相互作用を仮定する。そして、意味的な知識表象と視覚的および語彙的な知識表象は密接に関連しあっているのである。その結果、一つの構成要素、例えば意味記憶に障害があると、その障害の影響は、視覚性の知識にも語彙の知識にも影響を与えられからである。

3.1.1 モデル

彼女らのモデルを図 3.1 に示す。3つのユニット群、すなわち意味、名前、視覚のユニット群がある。実際には各層の間に中間層が存在していたのだが、

ここでは中間層の存在は本質的ではないので省略した。

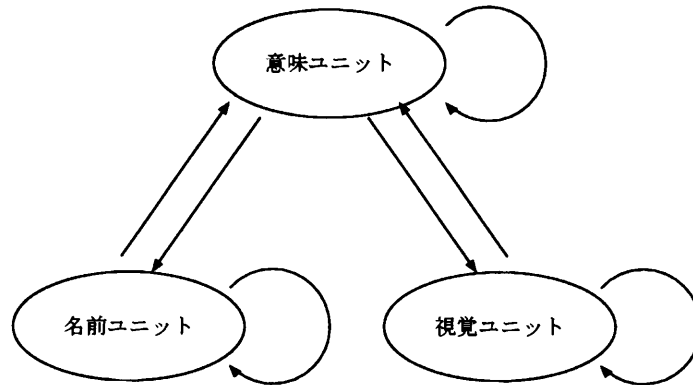


図 3.1: Tippet と Farah の用いたモデルの概略図

例えば視覚ユニット層に入力が与えられると、各ユニットの活性化(あるいは負の結合であれば抑制)は意味ユニット層に伝播する。意味ユニット層の活性化が、さらに視覚ユニット層と名前ユニット層の活性化に影響を与える。図中の矢印で示されているとおり各ユニット間の結合は層内、層間で双方向である。従って、あるユニットの活性は別のユニットの活性を引き起こし、全体として複雑な活性化パターンを示す。彼女らは視覚ユニット層(あるいは名前ユニット層)に入力を与えたときに、意味ユニット層と名前ユニット層(あるいは視覚ユニット層)に、特定の活性化パターンが示されるようにニューラルネットワークを訓練したのである。

3.1.2 視覚性障害仮説についてのシミュレーション

訓練後、意味層のユニットを除去することで脳損傷がシミュレートされた。脳損傷を受けていない完全なネットワークと脳損傷を受けたネットワークとに対して、視覚ユニットに入力を与えるときに、視覚刺激が完全である場合と、不完全な場合で名前ユニットに現われる活性化パターンがどのように変化するのが観察された。

結果は交互作用が観察された。すなわち脳損傷を受けていないネットワークに不完全な視覚刺激を与えても名前ユニットに表われる活性化パターンは学習したパターンに近いパターンが観察されたのに対して、意味ユニット層に損傷があるネットワークでは名前ユニットに表われるパターン、すなわち、視覚刺激を与えられたときに正しく物体呼称をすることが著しく障害されたのである。

3.1.3 語彙の頻度効果に対するシミュレーション

続いて、語彙の頻度効果を調べるために、高頻度語と低頻度語との違いがニューラルネットワークの訓練回数の違いとして表現された。視覚性障害仮説についてのシミュレーションと同じように、訓練後、意味層のユニットを除去することで脳損傷が表現された。

脳損傷を受けたネットワークの名前ユニットに、低頻度語を提示したときの成績は著しく低下するが、高頻度語を提示したときの成績はそれほど低下しなかったのである。すなわち、語彙性の障害と考えられてきたような症状を意味層の障害によって再現できたのである。

ニューラルネットワークを用いると、意味記憶の障害だけを仮定すれば、視覚性の誤りも語彙性の誤りも説明できる。すなわち従来から考えられてきた課題成績によるアルツハイマー病の障害の分類に全く新しい視点を与えることができるように思われる。

3.2 生物、非生物概念の二重乖離

認知心理学でしばしば話題になる記憶表象論争に関して、意味記憶は、個々の対象についてカテゴリーごとに構成されているのか、それとも意味記憶はモダリティーごとに構成されているのか、という論争がある。Farah と McClelland(1991) が行なったニューラルネットワークによる研究によれば、モダリティーに依存した意味記憶表象を考えれば、カテゴリーに基づく意味記憶表象は説明できる。

3.2.1 神経心理学的症状

実際の脳損傷患者の中には、動物や植物などの生物の知識について障害がある一方で、非生物の知識については健常のまま保たれている患者が存在する(Warrington & Shallice, 1984)。古典的な二重乖離の原則から、生物と非生物の知識の脳内での意味記憶には、生物と非生物とを独立に表象している意味記憶が存在すると仮定されてきた。しかし、Warrington と Shallice は、生物の知識と非生物の知識との間で選択的な障害が起こるのは、異なる感覚運動経路からの情報の重みの差異を反映しているためではないか、と述べている。すなわち、生物は主に感覚的な性質によって互いを区別することが多いが、非生物は主に機能によって分類される。ある動物、例えばヒョウは、他の肉食動物と比べて主に視覚的な特徴によって差別化される。これとは対照的に、机の知識については、他の家具との違いを記述するときには主に機能、すなわち何のために使うのか、によって差別化される。それゆえ、障害のある知識と健全に保たれている知識との違いは、生物-非生物の違いなのではな

く、対象を記述している特徴が感覚-機能の違いであるかも知れない。

Farah と McClelland(1991) のモデルは上記の感覚-機能仮説が意味記憶障害を説明できることを例示するために作成された。分散表現された2種類の意味記憶、すなわち視覚的意味記憶と機能的意味記憶とが損傷を受けたとき、生物-非生物の知識が選択的に障害されることを説明できる。

3.2.2 モデル

彼らのモデルを図 3.2 に示す。3つのユニット群、すなわち記憶を表現する意味記憶系と、入出力を表現する2つの周辺系、視覚ユニット群と言語ユニット群とがある。言語ユニット群と視覚ユニット群との間を除いて、全

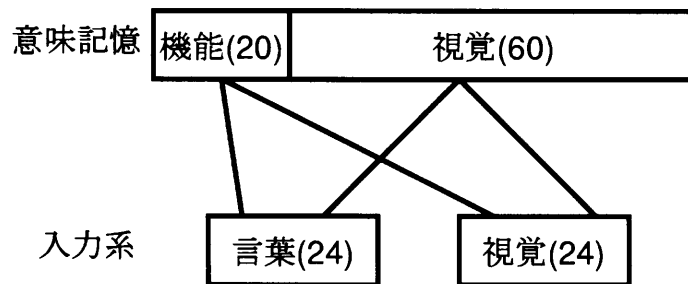


図 3.2: Farah と McClelland(1991) の意味記憶モデルの概念図。カッコ内の数字は数値実験で用いられたユニット数を表す。意味記憶内で機能的記憶と視覚的記憶のユニット数が異なるのは、彼らの論文中的実験 1 (心理実験) の結果を反映している。

てのユニットに群間および群内結合が存在する。

各ユニットの活性値は $(-1, +1)$ の間をとる実数値であり、しきい値は無い。各ユニットの活性値は同期的に更新された。すなわち、それぞれのユニットの活性値を計算するタイミングは全ユニットについて同時である。各周期ごとに、そのユニットに対して出力を送っている全てのユニットの活性値と結合係数からそのユニットの活性値が計算される。

それぞれ 10 個づつ、生物と非生物を表す刺激が乱数を用いて作成され、モデルに提示された。生物と非生物とを表す項目は、視覚情報と名前情報との比率が変えられた。生物項目では平均して 16.1 の視覚意味記憶ユニット、2.1 の機能的意味記憶ユニット。非生物では 9.4 の視覚意味記憶ユニット、6.7 の機能的意味記憶ユニットを用いて表現された。

誤差修正学習手続きによってネットワークが訓練された。絵画パターンが提示されたときに対応する意味記憶パターンと言語パターンが産出されるように、また、言語パターンが提示されたときには対応する意味記憶パターン

と視覚パターンが産出されるよに訓練された。

各訓練試行では、生物、もしくは非生物に対応する絵の名前が名前ユニット群あるいは視覚入力ユニット群に対して提示されそして、ネットワークは解が安定するまで 10 サイクルの活性値の更新が行なわれた。ユニット間の結合係数は学習についてはデルタルールが用いられた。

3.2.3 破壊実験

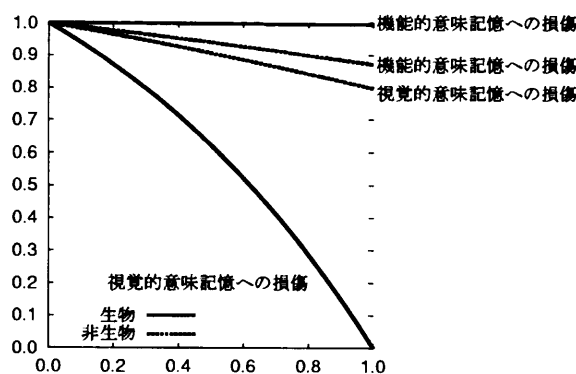


図 3.3: 生物-非生物別の意味記憶内の損傷の程度と課題成績との関係 (Farah と McClelland(1991) の表 3 と図 2 より改変). 縦軸は正解率, 横軸は損傷の程度を表している. 従って右に行けば行く程損傷が重篤になることを表し, 下に行けば行く程成績が悪かったことを示す. 視覚的意味記憶への損傷によって, 生物概念だけが選択的に傷害されていることがわかる. 彼女らは各条件で 0, 0.2, 0.4, 0.6, 0.8, 0.99 の 6 点について各 5 回づつしか数値実験を行っていないため実際の曲線は滑らかにならない. そこで, 指数関数に回帰させてプロットしなおした. 指数回帰を用いれば唯一のパラメータを変化させることで 4 つの条件に対応する曲線を描くことができる.

彼女らは, 実験 3 でモダリティに依存した意味記憶障害, すなわち, 耳で聞いたときには理解できるが, 目で見たときには特定の 카테고리についての知識に障害を生じる患者のシミュレーションを行なった. 実験 4 では, さらに, ネットワークに雑音を加える方法によって脳損傷を表現し, シミュレーションを行なった.

彼女らのシミュレーション結果は, モダリティ依存の意味記憶構造を用いれば, カテゴリ依存の障害を説明できることを示している. 図 3.3 は, 視覚的意味記憶への損傷によって生物概念だけが選択的に損傷を受けることが示されている. すなわち生物, 非生物という異なるカテゴリに属する事物は脳内で異なって表象されているのではなく, 意味記憶の構造は入力モダリ

ティーに依存して形成されているという結果が得られたのである。

3.2.4 ニューラルネットワークによる障害の再解釈

脳損傷を扱った認知障害の研究では二種類の実験を使って脳損傷患者の障害の場所を特定することが試みられてきた。

一つめは患者の示す誤りの種類を分析することである。例えば、物体の呼称課題において、視覚的に似ているものを答える誤りについては視覚性の誤りと分類し、患者の視覚機能に問題があると推論する。同様に意味性の誤りについては意味記憶に障害があると判断される。

二つめは課題の難易度を操作して障害の場所を特定することである。例えば呼称課題において、低頻度語の呼称成績が選択的に障害されていれば、語彙に障害があるとする。視覚刺激の質を低下させたときに、例えば写真を用いた物体呼称課題と線画を用いた物体呼称課題とで、線画を用いた方が呼称成績が悪ければ、視覚性の障害とみなす、などである。

ところが、本節で説明したようにニューラルネットワークモデルによる研究では、認知機能の局在を示す脳損傷患者のデータと、その認知機能を推論する伝統的な神経心理学的手法に疑問を投げかけている。ニューラルネットワーク研究によって、従来からの神経心理学的障害分類論に本質的な変更が迫られているように思える。

3.3 機能的失文法と統辞的失文法の二重乖離

3.3.1 失文法 agrammatia の種類

失文法患者には、発話の中でも品詞によって障害されやすさに不均衡があることが知られている。また、語順障害と語尾の欠落や誤用との二つに区別されるとする研究者もいる。ここでは、発話の流暢性と文法的な発話の産出は二重に乖離していることが要点である。すなわち

1. 表出が乏しい。内容語は産出できるが機能語が産出できないタイプ(電文体のような発話になる)形態的失文法
2. 内容語が乏しいが機能語は流暢で豊富である統辞的失文法

という二種類の患者が存在する。

実際、構文的に乏しく努力性の失文法的発話を示す患者では前方言語野(前頭葉)や深部の構造である島皮質、さらに側頭葉前部が損傷を受けることが知られており、流暢な発話における構文の障害では側頭葉、頭頂葉下部、および弓状回などに障害がある場合が多いとされている。

3.3.2 モデル

左半球のシルビウス裂によって二つの言語野、ブローカ野とウィルニッケ野は離されている。このシルビウス裂を開いてみると島皮質 (insula) という部分が現われる。島皮質はブローカ野とウィルニッケ野の中間に位置すると考えることができ、最近では言語の発話に関しても島皮質が関与している可能性も指摘されている (Dronkers, 1996)。ここでは島皮質の計算論的役割としてブローカ野とウィルニッケ野を結びつける役割の可能性を指摘しよう。島皮質で起こっていることが文章産出と言葉の聴理解に密接に関わっている可能性があり、大胆に仮説を構成すれば二重単純再帰型ニューラルネットワークである。二重単純再帰型ニューラルネットワークは言語産出と言語理解が密接に関わっていることを表す、おそらく最も単純なモデルである。我々が何かを話すときに起こっていることは、発話内容のプログラミングがブローカ野においてなされ、ブローカ野の指示に従って補足運動野や運動野を介して発話にいたる。いったん発話した内容は側頭平面にある第一次聴覚野を介してウィルニッケ野に入力される。つまり我々が話しているときには文章産出と文章理解の両者を同時に行なっていると推察される。その証拠に自分が話し

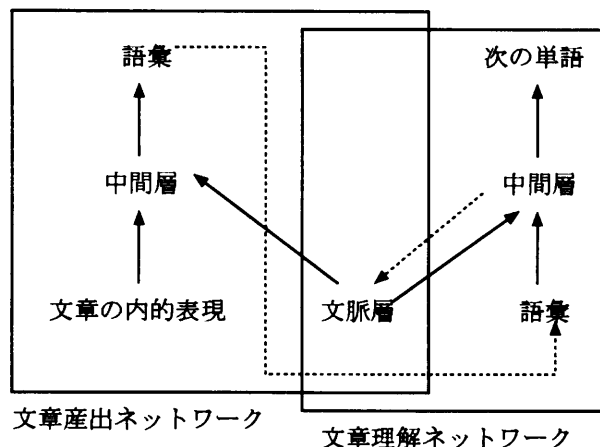


図 3.4: 2重経路単純再帰型ニューラルネットワークモデル

た言葉をマイクロフォンで録音し、一定の遅延をおいてヘッドフォンでその言葉を聴かせると言語産出が困難になる。このような心理実験課題とその効果のことを DAF (Delayed Auditory Feedback) という (Bradshaw, Nettleton & Geffen, 1971)。DAF 効果の存在が示していることは文章産出と文章理解とは密接にからみ合っており、切り離すことは難しいということである。

エルマンネットを使うと文章理解が可能である (Elman, 1990; Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996)。同じようにして入力刺激を一定の値に固定しておいて文脈層の変化によって文章産出をさせること

も可能である。このような方法をプロダクションSRNと言う。この二つのエルマンネットの文脈層を共有させるというモデルが二重単純再帰型ニューラルネットワークである。このモデルは大まかにブローカ野とウィルニッケ野という脳内の言語を司る領野とが文脈層(島皮質あるいは弓状回が候補になるだろうか)を介して結びついているということを表す、もっとも単純なモデルであると見なすことができよう(Dell, Schwartz, Martin, Saffran & Gagnon, 1997; Dell, Chang & Griffin, 2001)。二重単純再帰型ニューラルネットワークを使うことによって最も基本的な文章産出と文章理解の相互作用をモデル化することができる。図3.4に二重単純再帰型ニューラルネットワークを示した。

3.3.3 シミュレーション

文章産出ネットワークで生成された単語が次の時刻の聴理解を担当するエルマンネットである文章理解ネットワークへの入力となる。この二重単純再帰型ニューラルネットワークを用いて簡単な日本語の文章の産出と理解を訓練させた。訓練に用いた文型は全18文で以下の図3.5である。入力表現は、

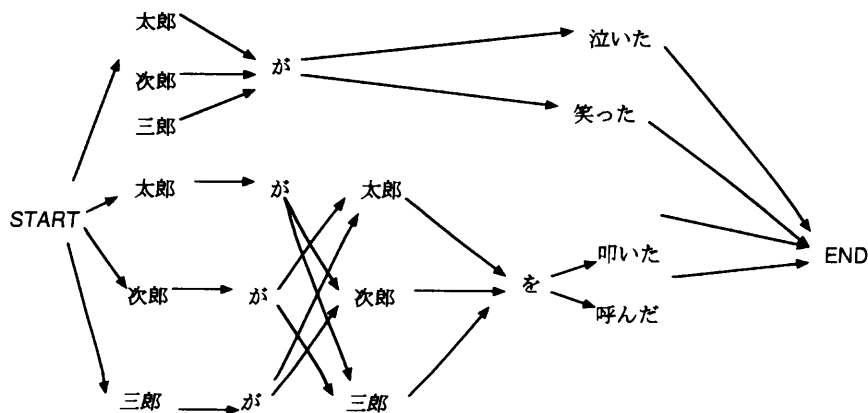


図3.5: 二重単純再帰型ニューラルネットワークの訓練に用いた文章。上が文型タイプ1, 下が文型タイプ2

主格太郎, 主格次郎, 主格三郎, 目的格太郎, 目的格次郎, 目的格三郎, 笑った, 泣いた, 呼んだ, 叩いた, の10ビットを0,1で表現した。例えば

「太郎が泣いた」→ 1,0,0 0,0,0 0,1,0,0

となる。

出力表現はEOS(文章の終わり), 太郎, 次郎, 三郎, 笑った, 泣いた, 呼んだ, 叩いた, を, が, の各ビットを0,1で表現した。例えば, 「太郎が泣いた」という文は, 時刻 $t=1$ で「太郎」を表すビットが1となり, かつ, 他

のビットはすべて0であるように0,1,0,0,0,0,0,0,0,0と表現された。次の時刻 $t=2$ では格助詞「が」を表すビットが1であり、かつ、他のビットは0となるように、0,0,0,0,0,0,0,0,0,1と表現された。以下、同様に時刻 $t=3$ では「泣いた」を表すビットが1であり、かつ他のビットはすべて0と表現され、最後の時刻 $t=4$ ではEOS(文章の終わり)を表現する最初のビットが1で他のビットがすべて0で表現された。すなわち「太郎が泣いた。」という文は、時刻 $t=1$ から $t=4$ までの時間刻みを用いて

0, 1,0,0, 0,0,0,0,0, 0,0 # ($t=1$)

0, 0,0,0, 0,0,0,0,0, 0,1 # ($t=2$)

0, 0,0,0, 0,1,0,0,0, 0,0 # ($t=3$)

1, 0,0,0, 0,0,0,0,0, 0,0 # ($t=4$)

などとなる。日本語の文章としては単純すぎるという反論は十分予想されるのだが、ここではいかに複雑な文章を産出、理解させるのかを目的にしているわけではなく、文章産出と文章理解の基本的な相互作用の在り方をシンプルに考えてみようという試みである。

二重単純再帰型ニューラルネットワークモデルでは、図 3.4 中の中央にある文脈層は、左側の文章産出ネットワークのための状態空間を遷移する。一方、図中の右側のネットワークは文章理解のためのエルマンネットであるから、次の単語を予測するために中央の文脈層が用いられる。すなわち図中の文脈層は文章産出の状態と単語予測のための状態とを同時に処理しなければならない。

学習の成立した二重単純再帰型ニューラルネットワークの文脈層では、どのようなことが起こっていたのであろうか。二重単純再帰型ニューラルネットワークの理論的解析についての研究は少なく、多くのことが分からずに残されている。従って文章産出と文章理解のエルマンネットで共有される文脈層で起こっていることは不明な点が多い。しかし、ニューラルネットワークが言語研究において有力な手段を提供していると考えられる長所の一つは、直接シミュレーションをして調べてみるができることである。

学習の成立した二重単純再帰型ニューラルネットワークに対して共有されている文脈層を破壊することによって人工脳損傷をおこさせてネットワークの振舞を観察した。人工脳損傷は、自由に、いつでも、どの場所でも破壊することができるので、神経心理学の有効な技法の一つになりうると考えられる。

シミュレーションの手続きは、中間層ユニット数を10にして、全18文のうちランダムに15文を学習データとして訓練した。残りの3文を用いて般化誤差を測定し、般化誤差が小さくなったパラメータを用いて文脈層のユニットを破壊した。

人工脳損傷によるシミュレーションから非常に興味深い結果が得られた。結果を二つのグラフ(図 3.6, 図 3.7)に示す。

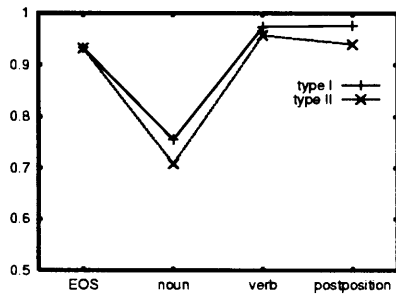


図 3.6: 失名辞タイプの損傷例. 縦軸は正解率を横軸は各品詞を表している. type I, type II の違いについては図 3.5 のキャプションを参照のこと

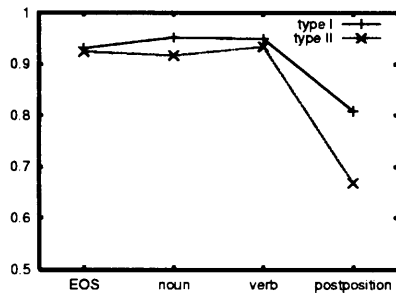


図 3.7: 形態的失文法タイプの損傷例. 縦軸は正解率を横軸は各品詞を表している. type I, type II の違いについては図 3.5 のキャプションを参照のこと

この二つのグラフは、人工脳損傷を文脈層に起こしたときに、各品詞の正解率を表している。すなわちグラフで高い位置にある品詞は人工脳損傷の影響を受けなかったことを表し、逆にグラフの下にきている品詞は成績が悪かったことを示している。両グラフの違いは、破壊した文脈層ユニットの違いである。すなわち別の文脈層ユニットを破壊すると別の品詞の成績が悪くなるのがこの二つのグラフから読み取れる。一方では格助詞が、他方では名詞が選択的に障害されている。

3.3.4 文法知識の創発と失文法

従来の神経心理学の枠組みでは、形態的失文法と統辞的失文法とは二重に乖離しているので、異なる脳内モジュールが障害を受けたと考えざるを得なかった。

ところが、本節の二重単純再帰型ニューラルネットワークの人工脳損傷のシミュレーションでは、同一モデルで二重に乖離した二つの失文法を、文脈

層ユニットに形成されたと考えられる文章産出と文章理解のために同時に利用される文法知識の障害として説明できるのである。しかも、この文法知識はアプライオリに与えられたものではなくニューラルネットワークの訓練の結果として文法知識が創発されたのである。結果を表す二つの図を見ると、従来の神経心理学的症例分類学をみなおす必要があると思われる。

3.4 まとめ

以上、二重乖離を示す神経心理学的データ3例をニューラルネットワークを用いて全く別の解釈ができることを示してきた。脳内で各認知モジュールは情報論的にカプセル化されていて各モジュール間は単純な情報が伝達されるだけであるという局在性仮説は、ニューラルネットワークモデルによって全く異なる見方が提供できる。ここで示したニューラルネットワークによるシミュレーション結果から考えれば、伝統的な神経心理学的理論のうちの幾つかは再検討されるべきではないだろうか。少なくとも二重乖離という現象を盲目的に当てはめることで、異なる認知モジュールを想定することはすべきではないだろう。

第II部

ニューラルネットワークによる
神経心理学的モデルの検討

第4章 2段階読字過程モデル

本章では視覚提示された文字を同定するためのニューラルネットワークモデルを提案し、計算機上での数値実験結果を報告する。ネットワークの一部を破壊することによって失読症に似た成績が現われることを示し、実際の症例への適用可能性を議論している。モデルの扱う範囲は視覚入力された文字の形態的特徴が抽象化された符号へと変換されるまでであり、破壊実験においては形態錯読をシミュレートすることを目的とした。読字過程を扱う心理学モデルでは音韻情報や意味情報へのアクセスとそれらの処理機構を扱っているものが多いが、本モデルはそれらのモデルの前段階に位置するものであって、音韻や意味の情報を扱っていない。

今日までに、多くの読みのモデルが認知心理学、神経心理学の分野で提出されて来ている (Geschwind, 1965; Morton, 1980; Coltheart, 1983; Farah, 1990; 御領, 1987, など)。中でも Dejerine が提唱し Geschwind (1965) によって再評価された視覚言語離断仮説は、今日でもなお純粋失読症状を説明するモデルとして妥当であろうと考えられる。しかし、近年異なる説明モデルも提出されている。たとえば、Farah (1990) は視覚障害が純粋失読を生起させる可能性を指摘している (視覚性障害説)。Farah の指摘の真偽 (河村と溝渕, 1998, の議論を参照) がどうであれ、視覚情報から音韻情報への変換過程の中の任意の機構に、損傷が発生することによって失読症が生じることは間違い無いだろう。

機能的脳画像研究 (たとえば櫻井, 1998) によれば極めて多くの部位が日本語文字の読字行動に関与している。このうち本章では紡錘状回後部から角回/縁上回の経路に着目してモデル化を行なった。事象関連電位を用いた研究では、顔・非顔の区別 (Jeffreys, 1989) や、文字・非文字の区別 (Nobre, Aillson & McCarthy, 1994) を表す波が紡錘状回後部で確認されている。すなわち紡錘状回後部は視覚入力をその特徴に従って大まかに分類する役割を担っていると考えられる。一方、Geschwind (1965) によれば角回は視覚文字-音韻変換の中継路ではなく、異種感覚記憶痕跡間の連合を営む領域である。この説が正しいとするならば、角回で表現されている情報は特定の感覚に依存しない抽象化された文字の表現でなければならない。本章ではこれらの事実と現在までにニューラルネットワーク研究で知られている2つの回路、自己組織化マッピングとパーセプトロンが対応すると考えてモデルを構築した。なおこのモデルは、工学の分野で考案された大規模ニューラルネットワーク CombNET-II

と基本的に同じ構造を持つ(堀田, 岩田, 松尾ら, 1992).

4.1 モデル

本章で提案するモデルは2段階(図4.1)で構成されている。入力パターンはこの2段階を逐次経過して処理される。

漢字の認識においては、繰り返し計算によるアルゴリズム、言い替えば、十分な時間が与えられれば完全な認識に至ることができるアルゴリズムを想定するのは不自然である。千字以上の文字が存在する漢字表記において、大分類を行う回路が脳内に存在しなければ、漢字文化圏の人間は、自身が考える速度で読むこと、または、読みながら考えるなどという行為は絶望的だと思われる。すなわち、脳内では、不完全であっても大まかな解を与える回路が存在すると考えることができる。原稿の誤字脱字を見付けることが難しいこと、あるいは、漢字を音読する場合の反応潜時が数百ミリ秒であることなどは、このようなワンショットアルゴリズムの機構を仮定することで説明できると考える。

第1段階では、入力された視覚パターンはその形態的特徴に従って高速に、大まかに分類される(以下、大分類と表記する)。大分類による入力文字画像のグループ分けは2次元平面に配置された大分類回路の素子の中のどの素子が活動するかによって定まる。大分類を担当する第1段階の回路は類似した入力刺激を2次元平面上の類似した位置に写像する機構である。類似した形態的特徴を持つ文字画像は同じ素子を活動させることになる、すなわち、同一グループに分類されることになる。入力刺激を2次元平面上に写像することは、大脳皮質が2次的に広がっていて、類似した特徴に反応するニューロンが皮質の近傍に分布するというトポグラフィックマッピングに対応すると考えられる。トポグラフィックマッピングは、視覚、聴覚、体性感覚など第一次感覚野の持つ特徴の一つであり、文字認識のためのモデルが持っていないからならぬ機構であると考えられる。本章では、第1段階での処理は自己組織化アルゴリズムによって訓練されると仮定した。実際に個々の文字を認識する前に入力文字の形態的特徴によって複数のサブグループに分類しておくことは、認識すべき文字の候補数を実質的な数に押える効果がある。第1段階で学習すべき文字の候補数を絞り込むことは、第2段階での認識機構において、学習に必要な時間を少なくし、かつ、学習の収束可能性を改善するという計算上の効率性が指摘できる。加えて、2段階の処理を仮定することは、全ての刺激入力を単一の大回路で処理する場合に比べて、ネットワークの経済性、すなわち、素子間の総結合数をより少なく押える効果も存在する。

第2段階での処理の目標は、入力画像の詳細な特徴に従って、視覚入力に対応する入力画像に依存しない文字コードに変換することである。本章では、第2段階の文字同定処理にパーセプトロン型の3層の回路を用いた。第2段

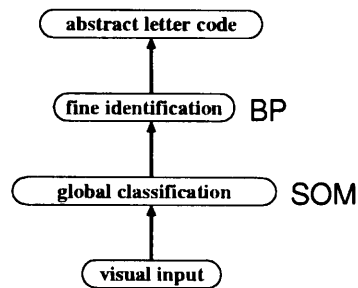


図 4.1: モデルの概念図

階の文字を同定する処理のために、第一段階の出力ユニット数と同数の 3 層フィードフォワード型ニューラルネットワークが用意された。第 2 段階のそれぞれのネットワークは第一段階の大分類処理によって、相対的に少数漢字を同定すれば良いことになる。このような 2 段階の処理を考えることで学習が容易になり、かつ、短時間で漢字の同定が可能になる。ただし、第 2 段階の各ネットワークは互いに同じ構成であり、同じ学習則に従って訓練されるものとした。第 1 段階の処理によって、第 2 段階の中のどの回路が活動すべきかが定まる。すなわち、大分類によって少数の候補に絞り込まれた入力画像が、第 2 段階を構成するネットワーク群の中の特定のネットワークのみを活性化し、そのネットワークへと信号を送ることになる。第 2 段階の各回路は、予め認識すべき文字の候補が少数に絞り込まれているので、認識までの時間、および認識の精度を向上させることが可能となる。第 2 段階での処理はバックプロパゲーション法 (Rumelhart, Hinton & Williams, 1986) によって訓練された。

いったん訓練が終了したネットワークに損傷を加えると、失読症、とりわけ形態失読に似た出力を生じることが期待できる。本章では 2 種類の損傷シミュレーションを行った。失読症状、特に、文字の形態的特徴を混同したり、類似した文字と誤読する症状を持つケースと、モデルとの挙動が一致するのであれば、本章で提案するモデルは日本語の読字モデルとして妥当性を持つと行うことができる。

4.2 計算機実験

4.2.1 材料、原画像

入力データには小学 1 年生から 6 年生までの学習漢字 1006 文字の TrueType フォント (Microsoft 社製) のフォント 2 種類、MS 明朝、MS ゴシックを用いた。各文字を縦横 32 × 32 ピクセルにラスターライズした画像を原画像として用いた。

4.2.2 前処理

ニューラルネットワークが文字の特徴を学習し、用いたフォントの微妙な特徴を学習してしまうことを避けるために、位置のずれの効果を減ずるために用いられるガウシアンフィルターを各文字に施した。ガウシアンフィルターを用いることは特徴検出器を構成することに相当する (Bishop, 1995)。

以下の図 4.2 のように完全に直交している 2 つの画像を考える。人間は容

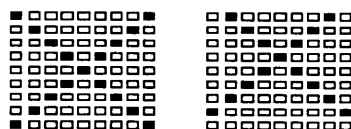


図 4.2: 人間には同じように見えるにもかかわらず、位置ずれのため完全に直交している 2 つの画像

易に左右の文字を同じ X であると認識できるのに対して、機械は全く別の文字だと認識する。このような事態を避けるためには、なるべく特定の画素に依存しない表現でなければならない。特定の画素に依存しない表現を用いる一つの方法としてガウシアンフィルターを用いられている。本章でもこの目的でガウシアンフィルターを導入した。ガウシアンフィルターを用いれば、ガウシアンフィルター用いなかったときにはほぼゼロであった上記二枚の画像の相関は、互いに類似するようになる (相関が高くなる)。

ガウシアンフィルターによってぼかされた文字画像は位置ずれに対して頑健であるが、高周波成分を除去することになり、元画像よりも識別性が低下してしまうという欠点がある。すなわち、ガウシアンフィルターは画像間の類似度を減少させることを意味する。一つの漢字を一本のベクトルで表現した場合、2 つのベクトル間の距離が小さくなる (図 4.3 中の f と \tilde{f} の関係)。このことは、漢字間の識別性を低下させ、学習を行う際に元画像を用いるより困難な課題をネットワークに課すことを意味する。

さらに、白黒反転した画像も同じ回路で扱えるように正準化と呼ばれる画像処理手法を導入した。正準化とは一種の射影であり、画像の平均がゼロになるように変換する操作である。(4.1) 式はこの操作を表している。

$$F(x) = \tilde{f}(x) - (f, 1), \quad (4.1)$$

ここで、 $(f, 1)$ は元画像とすべての要素が 1 であるベクトルとの内積である。(4.1) 式はガウシアンフィルターによってぼかされた画像 $\tilde{f}(x)$ から画像の平均値 $(f, 1)$ を引くことによって刺激画像 $F(x)$ が作成される手続きを表している。正準化の操作によって画像間の類似度 (あるいは画像間の距離、内積の値) は変わらない。元画像とガウシアンによってぼかされた画像、および正準化した画像との関係を図 4.3 に示した。図中の \tilde{f} はガウシアンフィルター

によってぼかされた画像を，また $F(\cdot)$ は正準化した画像を表す． $F(\cdot)$ 間の角度は元画像と同じであるが，一方ぼかした画像同士の角度は元画像よりも小さくなっている．以上のような前処理は，水晶体やレンズによする工学的なボケや網膜水平細胞間の結合（ガウシアンフィルタ）や，順応あるいは虹彩の調節による強度変化，アマクリン細胞の役割（正準化）などに対応することが想定でき，一次視覚野に至るまでの視覚情報処理過程に対応すると考える．図 4.4 は 2 種類のフォントについて上記の操作を行った例である．

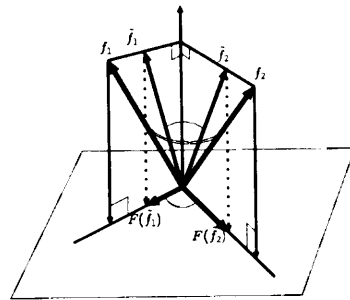


図 4.3: 原画像と前処理を施した画像との幾何学的関係

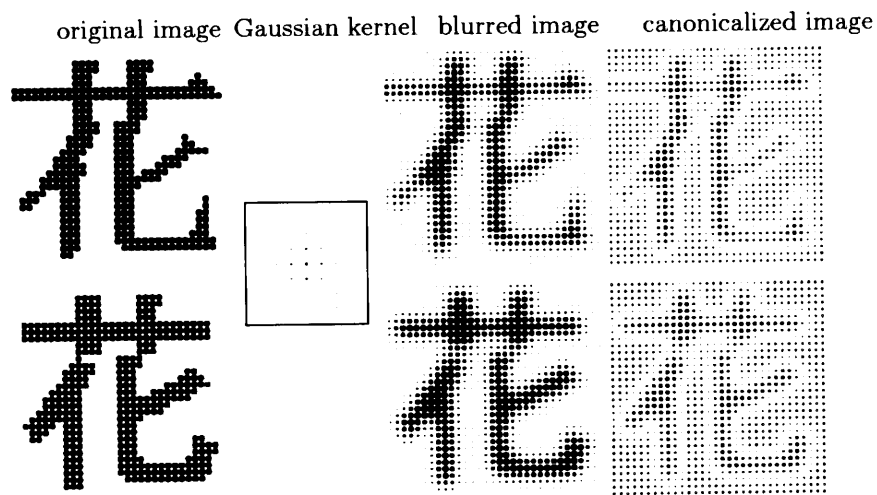


図 4.4: サンプル画像．画像内の各点の直径は絶対値の大きさを表している．白は正の値を，黒は負の値をそれぞれ表現している．

4.2.3 結果

第1段階

入力画像を大分類(グループ分け)するために Kohonen(1985)の自己組織化アルゴリズム(SOM)を用いた。SOMとは、競合学習の考え方を採用し、高次元の入力刺激を2次元平面へ射影する計算手続きである。本章では全漢字1006文字を49個(7×7)の正方格子状に配置された素子で構成される2次元平面へ写像させることにした。SOMの学習とは、各入力刺激に対して最大出力を与えるニューロン c と、その近傍関数 $h_{c,i}(t)$ とで定義されるニューロン群に対してHebb則(Hebb, 1949)を適用して訓練し、類似した入力刺激の特徴が類似した場所に投射されるような2次元のマップを形成することである((4.3)式を参照)。第一層の49個の各素子は入力次元である1024(32×32ピクセル)個の結合係数(コードブックベクトル)を持っており、学習の結果、これらの結合係数が刺激画像全体をグループ分けするような結合係数を獲得することになる。この学習が最適に行われたと仮定すれば、第2段階の各回路は20個強(=1006/49)の文字を認識すれば良いことになる。

前節の方法で前処理(ガウシアンフィルターと正準化の操作)された1024次元の入力文字画像を表すベクトル x が全コードブックベクトル m_i ($i=1\dots 49$)と比較され、最小のユークリッド距離 $|x - m_i|$ を与えるコードブックベクトルを決める。このコードブックベクトルのことを最適ノード(添字 c で表す)という。すなわち、49本のコードブックベクトルのうち、最も入力画像に類似しているコードブックベクトル m_c が選択される。数式で表現すれば

$$c = \underset{c}{\operatorname{argmin}} |x - m_i|, \quad (4.2)$$

となり、入力画像 x とコードブックベクトル m_i との距離の絶対値を最小にするコードブックベクトル m_c が定められる。SOMの学習は次の(4.3)式によって更新された。

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)], \quad (4.3)$$

ここで、 t は学習回数を表す離散変数であり、 $h_{ci}(t)$ は近傍関数である。近傍関数とは、刺激画像と最も近いコードブックベクトル m_c の周辺にあるコードブックベクトルのことであり、コードブックの2次元的な配置の中で、類似した特徴をもつ画像どうしは近くの領域に投射されるというトポロジカルマップを実現するために用いられる。(4.4)は用いた近傍関数の定義である。

$$h_{ci}(t) = \begin{cases} \alpha(t) & \text{for all } i \in (|m_c - m_i| \leq 3) \\ 0 & \text{otherwise,} \end{cases} \quad (4.4)$$

ここで、 $\alpha(t)$ は学習係数である。(4.3)式と(4.4)式によって入力画像に最も近い最適コードブックベクトルとその周辺(3以内)にあるコードブックベク

トルが刺激画像との距離を減少させるように学習が繰り返される。2000回の学習期間中に最適コードブックベクトルとその近傍のニューロンが同一の刺激に対して類似した出力を与えるように学習が行われるようになった。学習は収束するために必要な(4.4)式中の学習係数 $\alpha(t)$ は以下のように

$$\alpha(t) = 0.9 \left\{ 1 - \frac{t}{2000} \right\}, \quad (4.5)$$

学習回数 t の関数として単調減少関数として定義した。(4.5)式の右辺の内は学習回数 t が 2000 に近づくけば近づくほど 0 に近付き、従って学習の効果が徐々に減衰することを表している。

第一段階の結果を図 4.5 に示した。例えば図 4.5 内の左上隅のノードに‘しんによう’検出器と考えられるノードが形成されたことが分かる。このノードの結合係数が最適コードブックベクトルである漢字は、出力値順に並べると‘通適適達選連運遺遠造道過追送近建述進速退短辺返迷逆遊店法’であった。実際‘店’と‘法’とを除けばすべて‘しんによう’が部首である漢字である。これらの漢字は互いに類似しているが発音が全く異なることは日本語を母語とする人間には自明であろう。一方、図 4.5 内の右から 2 列目の中程には‘いと

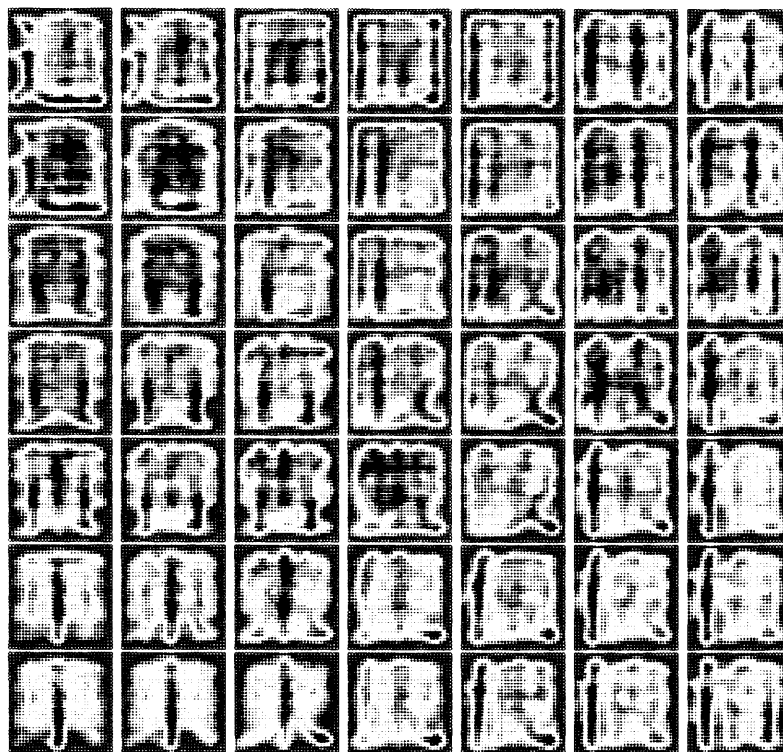


図 4.5: SOM による大分類結果

へん’を示すノードが形成されている。以上のように刺激画像の持つ統計的な

性質を反映する形で、2次元のコードブックベクトルマップが形成され、刺激画像をその視覚的特徴によって大まかに分類する2次元マップが形成されたと言える。本論文では上記のような漢字の形態による大分類を行なっている領野が脳内のいずれかに存在すると想定している。

第2段階

大分類の結果から漢字を49群に分類し、パーセプトロン型の回路網に読みを学習させた。分類の難しい文字(最適コードブックベクトルと2番目に大きな値を生じるベクトルとの差が小さい文字)については、複数のグループに同一の文字が複数のグループに属するようにした。回路の学習にはバックプロパゲーション法を用いた。第2段階の各素子はその出力がジグモイド関数によって決まるようにした。49個のサブネットのそれぞれは入力層の素子数1024個、中間層の素子数128個、出力層の素子数16個とした。教師信号はJIS X 0208コードを表す16ビットをそのまま用いた。

第2段階の領域として本モデルで想定している角回が異種感覚記憶痕跡間の連合を営む領域であるのなら(Geschwind, 1965)、角回で表現されている情報は特定の感覚に依存しない抽象化された表現である。ここでは視覚刺激と無関連な抽象的なコードであるJIS X 0208を用いることにした。

JIS X 0208では一文字を2バイトで符号化している。2バイトで表現可能な文字の総数は $65536 (= 2^8 \times 2^8)$ 文字であるがJIS X 0208では最上位ビットが常に0であるため可能な文字集合は16384文字 $(= 2^7 \times 2^7)$ である。すなわち第2段階の各回路は、出力層における最上位ビット2つは、冗長であることを学習しなければならなかったことを意味する。

モデルの最終的な認識率は0.892であった。正しく認識できなかった文字は、全2012文字中、MS明朝フォントが135文字、MSゴシックが82文字であった。

4.3 破壊実験による脳損傷のシミュレーション

損傷をシミュレートするためには(1)結合の切断、(2)素子の除去、(3)雑音の付加、の3種類が考えられる(浅川, 2001b)。結合の切断は、素子間の結合係数を0にすることに等しく、ニューロン(もしくは皮質)間の線維連絡が切断された場合に対応する。素子の除去は回路から一部の素子を取り除くことを意味し、皮質上のニューロンが損傷を受けた場合に相当する。雑音の付加にはさらに細かく2種類の方法が考えられる。(3a)入力そのものに雑音を加える場合と、(3b)学習が成立した結合係数に雑音を加える場合とである。前者は回路への入力が劣化していることを表現し、後者は何らかの外生要因によってシナプス結合が変性している状況に対応する。本章では離断仮説に

対応する (1) と、視覚入力正常であるにもかかわらず入力情報が変容してしまう (3b) とを取り上げた。

4.3.1 離断仮説

Geschwind(1965)の視覚言語離断仮説は、上記の結合の切断で実現できると判断できる。第2段階の回路の任意の結合を切断すれば視覚言語離断仮説と直接対応すると考えられる。この仮説を検証するため第2段階の回路内の任意の結合係数をランダムに0にして認識率を調べた。図4.6に結合係数を0で置き換える切断率を0.1から0.9まで変化させたときの認識率を示した。図4.6は、結合係数をゼロにする割合が大きければ大きい程急速に認識率が低下する。言い替えれば、障害の程度が重ければ重い程、認識率が低下することを示している。

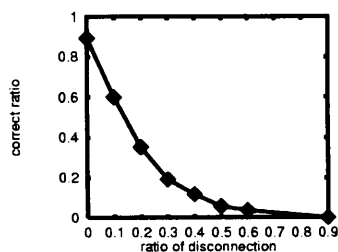


図 4.6: 離断仮説によって生ずる失読の数値実験結果。中間層の破壊率の関数として認識率が描かれている。各点は100回の実験結果の平均である

図4.6によれば、切断率とモデルの示す成績との関係は単調減少の傾向にあるが、変化は一定ではない。切断率が0.3付近まで、急激に認識率が低下する。このことは損傷の程度が軽い場合でも認識不能な文字が多い(読字可能な文字が減少する)ことが分かる。もし実際の損傷あるいは病変が本モデルで記述しているとおりならば、以下の2点が演繹できる。

1. 本モデルの第2段階に相当する領野に進行性の病変が存在すれば、患者の読字能力は急激に悪化する。
2. 損傷による「読み」の成績低下が図4.6のようになるとすれば、損傷部位の面積と課題成績とは直線的ではなく、損傷の程度が少なくても急激に成績が低下すると予想される。すなわち、失読症の患者間で成績を比較する場合、小さな損傷領域の差異が大きな成績差となって現われる可能性がある。

4.3.2 視覚性障害説

Farah (1990) は純粹失読の患者には、ある種の視知覚的な障害を持つものがあることを報告している。この仮説を調べるために雑音の付加によるシミュレーションを試みた。ただし、視覚入力に雑音を加わると仮定する前述の方法 (3a) では他の視覚機能にも影響が及んでしまうので、純粹失読の定義からシミュレーションとしては適当ではない。ここでは本ニューラルネットワークモデルの結合係数に雑音を加える (3b) によって数値実験を行った。

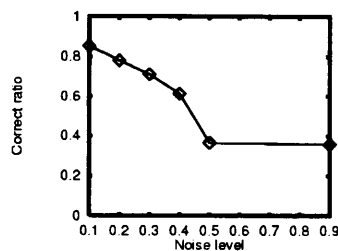


図 4.7: 視覚性障害説によって生ずる失読の計算機実験結果。付加された雑音の大きさの関数として認識率が描かれている。各点は 100 回の実験結果の平均である

図 4.7 に結果を示す。離断仮説によるシミュレーションと対照的に、視覚性障害説では、雑音の程度に対して頑健な傾向がある。雑音のレベルが高くなければ成績が悪くなるが図 4.6 と比較すれば成績低下の割合は少ない。このことから、障害の影響が 2 つの型の失読症で異なることが予測できる。

4.4 考察

現在までに知られている事実から計算機上で動作するモデルを作成すると文字数候補の多さによる学習時間、認識時間の増加を抑える意味でも 2 段階回路網を採用した方が有利であると考えている。現在の機能的脳画像技術では本モデルでの予測を検証できるような精度が得られていない。そのためモデルと直接対応する回路を脳内に発見することはしばらくできないと考えられるが、敢えて予想される対応部位を検討すると次のようになろう。櫻井 (1998) などによれば読字過程の PET 画像では角回の deactivation が見られると言う。文字や顔などは、高速処理が要求される視覚刺激である。このような時間的制約の中で正確に入力刺激を分類する能力をもつニューラルネットワークでは、できるだけ単純な処理過程を考えるべきである。仮に大分類回路が紡錘状回後部だけでなく紡錘状回後部と角回との協応作業であると仮定するならば、最適コードブックベクトルを選択する際に互いに他の素子を抑制することになり、最適な素子だけが活性化されることになる。その場合、全体の

活性値は閉眼時よりも減少すると考えられる。すなわち機能的脳画像研究で見出された角回での deactivation は大分類回路が存在する可能性を示唆すると解釈可能であろう。deactivation の扱いについては機能的脳画像研究者間でコンセンサスが得られていない問題で、単なるアーチファクトあるいは計算の便宜上存在するものだと考えている研究者もいる。ここに帰納的(実験的)研究の限界があり、本論文のような演繹的な研究(序論でも述べたとおり脳の構成論的研究)方法を必要としていると考える。角回での deactivation 問題は、角回の計算論的意味を考察することにより、検証すべき仮説となり得るからである。

最後にかな漢字乖離問題に触れておく。現在までに提案されたモデルの中にはかなと漢字との乖離を主張しているものが数多い(例えば岩田,1996)。しかし、ここではかなと漢字との処理経路を分けることはしなかった。理由は経済性である。もし、一つの経路で、かなと漢字との両者が同時に扱えるのなら、敢えて二つの経路にわけると必要はないと考えた。これは脳は同一の構成原理で成り立っているとの主張と合致する。同様の考え方は最近、老人研グループの研究でも示されている(伏見,伊集院,辰巳,2000;伊集院,伏見,辰巳,2000)。大分類回路による分類は確率的に振舞うことになるので、回路内で局所的にかなと漢字とをそれぞれ選択的に処理する部位が発生することは十分に考えられる。しかし、その場合でもモデルの構成原理はかなと漢字とで同一である。同一原理によって構築された脳が外部入力 of 性質を反映して機能分化を起こす可能性までは否定できない。

第5章 単語の読みにおける労働の 分割問題に対する解として のエキスパート混合ネット ワーク

ニューラルネットワークを神経心理学へ応用する研究分野では、いったん計算機上に実現されたニューラルネットワークモデルを系統的に破壊することで神経心理学的症状の理解を試みる研究手法が試みられている。本章ではこのような研究のうち特に読みのモデルである失読症の研究におけるニューラルネットワーク的な実装を議論する。

本章は以下のような構成になっている。まず、一般的な失読症の読みのモデルの持っていなければならない要点を整理し、つぎに主要なニューラルネットワークによる単語の読みのモデルである二重経路モデル (Dual Route Cascade model) (Coltheart, Curtis, Atkins & Haller, 1993; Coltheart & Rastle, 1994) とトライアングルモデル (triangle model) (Plaut, McClelland, Seidenberg & Patterson, 1996b) とを概観する。さらに両モデルにおける論点に触れ、二つのモデル間での論争を紹介し、トライアングルモデルにおけるもっとも大きな特徴である労働の分割 (division of labor) 問題についてのニューラルネットワーク的な実装の問題を取り上げる。そして、エキスパート混合ネットワーク mixtures of experts (Jacobs, Jordan, Nowlan & Hinton, 1991; Jordan & Jacobs, 1994) と呼ばれるニューラルネットワークモデルにおいて二つのモデルは統一的に記述することができることを指摘する。エキスパート混合ネットワークによる説明を用いれば、二重経路モデルとトライアングルモデルとの間に本質的な違いはない。両者の間に存在する違いとは、分散パラメータによる入力空間の領域分割の大きさという量的な問題に帰結され、本質的には二重経路モデルとトライアングルモデルは同じエキスパート混合ネットワークという、より一般的な範疇のニューラルネットワークモデルとして同一視できることを示す。

5.1 失読症のニューラルネットワークモデル

語彙辞書へアクセスは排他的論理和と同じ構造を持つものがある。例えば mother, father, female, male という単語は意味構造に、性別 (女性であれば 1 で男性であれば 0) という属性を持ち、親という属性 (親であれば 1 そうでなければ 0) も持つものと仮定すると、最初の音が /m/ であるか否かについては排他的論理和で表現される。従って意味から単語の音声へと変換する機構には少なくとも排他的論理和を解く能力が備わっていなければならない。

排他的論理和は線形分離不能なので 2 層のパーセプトロンでは解けないことが知られている (浅川, 2001a)。従って、単語の音読課題のシミュレーションには 3 層以上のパーセプトロン型ネットワークが用いられてきた。

5.2 英単語の特徴

英語ではおなじ a でも make と cat では発音が異なる。a がどのように発音されるかはその単語の前後の文字によって定まる。英単語には、文字と音との間の対応関係が規則的である規則語 (regular words) と規則に従わない不規則語 (irregular words または exception words) とが存在する。多くの単語と読みを共有している単語を一貫語 (例えば hint, mint, print, lint) と言い、例外語を非一貫語と呼ぶ (例えば pint) ことがある。

ニューラルネットワークにおける基本的なアイデアは単語は 3 つの要素 (あるいは脳内の部位と言っても良い) に分散表現されていると仮定することである。3 つの要素は書記素、音韻、意味と呼ばれる。書記素とは視覚提示された単語の脳内表現を指し、音韻とは音声単語の脳内表現、意味とは視覚や聴覚などの感覚入力に依存しない抽象的な表現である。

5.3 3 つの失読症

大きく分けて読みの障害には 3 種類ある。音韻失読症 (phonological dyslexia)、表層失読症 (surface dyslexia)、深層失読症 (deep dyslexia)、である。

音韻性失読症の患者は実在する単語は読むことができるし、非単語を書き取ることもできるし、復唱することもできるが、非単語を読むことができない。例えば must は読むことができても実際には存在しないが発音可能な単語 must を発音することができない。

表層失読の患者は、一貫語や非単語を読むことができるが、低頻度の例外語を読むことができない (例えば yacht)。表層失読の患者は視覚性の誤り (dog を dot と言ったりする) もするが意味の誤りはない。後述する二重経路モデルによる音韻失読の説明では、書記素から音韻へ変換 (直接経路という) の障

害と考えることができる。

深層失読の患者は音韻失読の患者と同じく非単語を読むことができない。このことに加えて、深層失読の患者は意味性の錯読を示す。例えば dog を cat と言ったりする。また、視覚性の誤りを示すことがある。ときには上記二つの混ぜ合わさった読みの誤り、たとえば sympathy を orchestra と言ったりする（おそらく sympathy と綴りの似ている symphony を介して）。この種の患者は、抽象語と具体語の成績にも差があることがある。具体語である「椅子」よりも抽象語である「真実」の読みの成績が悪い。二重経路モデルによる深層失読の説明では、書記素から音韻への直接経路の障害の他に意味経路の障害が加わったとされる。一方、やはり後述するトライアングルモデルに沿って考えれば、深層失読は、直接経路の障害によって生じた、意味系への間接経路との労働の分割問題の結果とみなすことができる。労働の分割問題の意味については節を改めて概説する。

5.4 二重経路モデル

Coltheart ら (Coltheart, Curtis, Atkins & Haller, 1993; Coltheart & Rastle, 1994) によって開発された記号処理モデルである二重経路モデルでは、印刷文字を音韻へ変換するための明示的な規則に基づく直接経路（一貫語が処理される）と、規則にあてはまらない単語（非一貫語）を読むための単語ベースのルックアップテーブルを持つ間接経路（語彙経路）とから構成されている。印刷された文字を読むときには、直接文字を音韻へと変換する直接経路と、例外語の語彙テーブルまたは意味を介して発話に至る間接的な経路との2つの経路を仮定するのが二重経路モデルである。失読症、あるいは難読症患者のパフォーマンスを示す特徴である一貫語と例外語の乖離から二重経路モデルが考案された。二重経路モデルの枠組みによれば、一貫語の読みの特徴から書記素音韻対応規則 (Grapheme Phonology correspondig rule) が導き出され、非単語の読みを説明するために（間接経路である）語彙経路が仮定されている。

5.5 トライアングルモデル

これに対して、トライアングルモデルでは、そのようなルールあるいはスイッチの存在を仮定しない (Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg & Patterson, 1996b)。代わりに、トライアングルモデルでは同時的、相互作用的処理が仮定される。トライアングルモデルにおける直接経路では、一貫語と高頻度の例外語が学習される。一方、低頻度の例外語は意味系に依存すると仮定される。すなわちトライアングルモデルにおける直接経路では一貫語と高頻度例外語が学習され、従って直接経路は単語の頻度効果

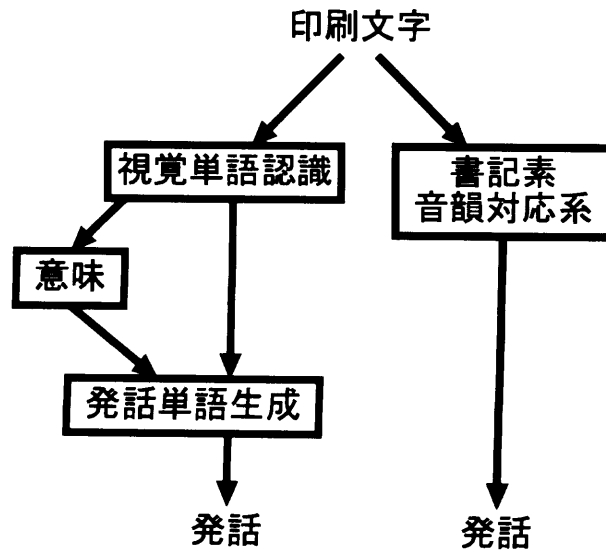


図 5.1: 二重経路モデル

に、すなわち単語の統計情報 (生起確率) に敏感である。一貫語および高頻度例外語と低頻度例外語との処理の違いには労働の分割と呼ばれる作用が関与する。2つの経路間の離散的なスイッチを持つ代わりに、トライアングルモデルでは同時的、相互作用的处理である労働の分割が行なわれる。

トライアングルモデル (Plaut, McClelland, Seidenberg & Patterson, 1996b) では英語の単音節語 2998 語を読むことが可能であった。入力表現は、オンセット、母音、コーダという3つの部分から構成される母音中心表現と呼ばれる表現を用いている。この入力表現の工夫により非単語の読みについても人間と同程度の能力があることが示された。最近ではトアイアングルモデルで用いられた母音中心表現を自動的に学習させる試みもなされている (O'Reilly & Munakata, 2000)。

トライアングルモデルでは、次のような心理実験結果を説明することができた。

1. 高頻度語の音読潜時は低頻度語の音読反応時間よりも速いこと
2. 一貫語は非一貫語よりも音読反応時間が速いこと
3. 頻度効果との交互作用があること。規則語-例外語の効果は低頻度語の方が効果が大きく、高頻度語では規則語-例外語の効果が小さいこと

などである。特に低頻度の例外語を規則化して発音する誤りが見られた。すなわち、音韻失読は直接経路への損傷の結果生じると見なすことができ、モデルの出力結果は患者の検査結果とも一致していることが示された。

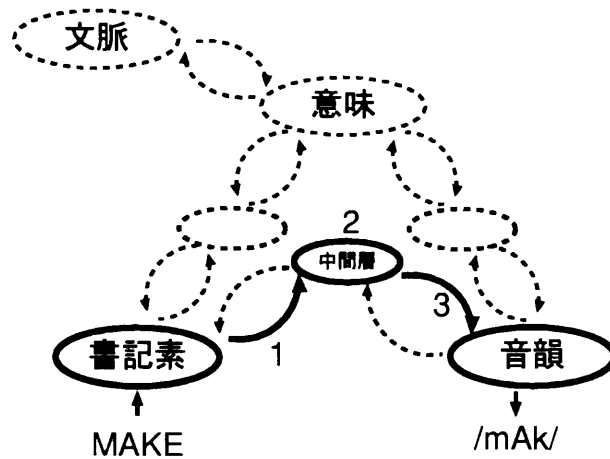


図 5.2: トライアングルモデル. 書記素から音韻への直接経路は3層の、意味経由の間接経路は5層のネットワークと見なせることに注意

トライアングルモデルの直接経路は3層の、意味層を介した間接経路は5層のニューラルネットワークと見なせることに注意されたい。

5.6 論争

二重経路モデルかトライアングルモデルかという論争の論点は、単一経路か二重経路かという議論ではなく、直接経路と間接経路の処理の違いに関してである。

二重経路モデルの特徴は明示的で記号处理的なルールすなわち単語毎のルックアップテーブルを用いることである。すなわち二重経路モデルでは2つの経路間の離散的なスイッチが仮定されている。これに対して、トライアングルモデルではそのようなルールあるいはスイッチの存在を仮定しない (Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg & Patterson, 1996b)。代わりに、トライアングルモデルではニューラルネットワークの特徴である同時的、相互作用的处理が仮定される。2つの経路間の離散的なスイッチを持つ代わりに、ニューラルネットワークアプローチでは同時的、相互作用的な労働の分割が行なわれる。

5.7 労働の分割問題

トライアングルモデルはニューラルネットワークの特徴である相互作用をする。いったん書記素層へ入力された文字は、意味層を介する間接経路と音

韻層に直接出力を送る直接経路との両方の影響を受ける。ある単語がどちらの経路をたどって読まれるかは、単語ごとに、また、個人ごとに異なるとされる。直接経路では規則語と高頻度例外語とが処理され、低頻度例外語は間接経路である意味層のサポートを必要とする。このように単語ごとに2つの経路の影響が異なって表現されることを労働の分割問題という。

表層失読は、この労働の分割問題によって説明される。労働の分割による説明では直接経路に損傷がある場合、間接経路を経由した読みは労働の分割の程度によって不完全な読みが生じるからである。

ただしトライアングルモデルでは労働の分割問題を完全に実装したわけではない。Plautら(Plaut, McClelland, Seidenberg & Patterson, 1996b)は図5.2の実線で描かれた部分を実装し、図中の数字が描かれている部分を破壊することで音韻失読の症状を再現したにすぎない。完全な三角形が実装されたわけではなく、図5.2中の実線部分で説明できないことを労働の分割と読んでいるにすぎない。

5.8 エキスパート混合ネットワークによる労働の分割問題の解

ここで、あらためてニューラルネットワークモデルに単語の読みを学習させることの意味を考えてみよう。ニューラルネットワークに単語の読みを学習させるということは、書記素から音韻への変換規則を学習させることである。低頻度例外語にエラーが大きいのは、他の大部分の単語に共通する書記素-音韻対応規則を学習し、その結果を適用しているからであって、ニューラルネットワークの見地からすれば正しい一般化と解釈することができる。すなわち、未学習のデータに対して、学習によって獲得した書記素-音韻対応規則を適用しているという意味である。むしろ、高頻度例外語は学習のしすぎ、すなわち過学習である。規則語と例外語との頻度効果の交互作用は、このような説明が可能である。そして、直接経路と間接経路との労働の分割問題は2つのネットワーク間の競合あるいは協調作用とみなすことができるだろう。

もし、書記素-音韻対応規則を学習し、非一貫語と一貫語を自動的に分類して学習できるアルゴリズムが存在すれば、表層失読と音韻失読という神経心理学的症状を説明するための労働の分割問題を解決できるモデルになるだろう。この語彙の自動分類機構を実現するのがエキスパート混合ネットワーク Mixtures of experts (以下 ME と略記) とよばれるニューラルネットワークモデル (Jacobs, Jordan, Nowlan & Hinton, 1991; Jordan & Jacobs, 1994) である。

ME とは入力データ空間をいくつかの小領域に分割し、その分割された各領域に対してひとつのニューラルネットワークを割り当てることによって、複雑な問題の解を求めるため手法である。複雑な問題を分割して小領域に区切

ることによって、1つの大きなニューラルネットワークを学習させるよりも効率の良い学習をさせようというモデルである。幾つかに分割可能な課題からなる訓練データに対して、一つの階層型ネットワークをバックプロパゲーションによって訓練したとしたら、各課題間の干渉作用によって学習が遅くなり、貧弱な一般化しか得られない。予め訓練データセットが幾つかの下位課題に分割されていることが分かっているのならば、どのエキスパートネットワークがどのエキスパートに属するかをきめるゲーティングネットワークと、各々の課題を処理する複数のエキスパートネットワークを用いて、大きな改善が見込める。MEにおける学習とは、入力空間を分割の仕方を発見させ、分割された各小領域に属するデータに対する最適な答えを見つけ出すことである。

MEはエキスパートネットワークとゲーティングネットワークから成り立っている教師あり学習の一手法である。ゲーティングネットワークは問題空間を分割するために用いられ、各エキスパートネットワークは分割された領域内での局所的な解を出力する。MEは階層的な問題空間の分割とエキスパートの割り当てが可能である。2階層の場合のMEを図5.3に示した。

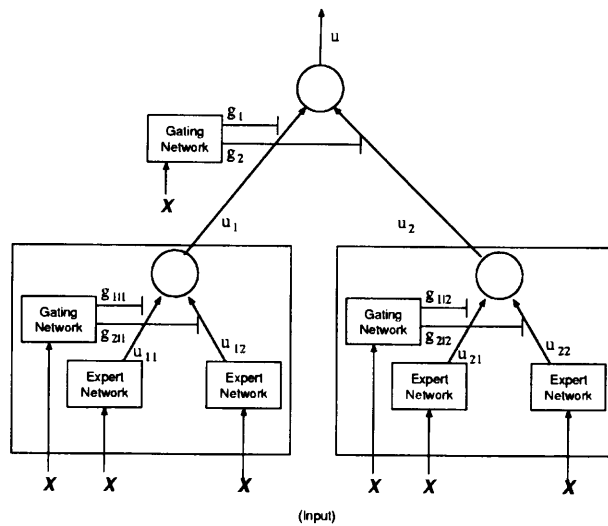


図 5.3: 2段階のエキスパート混合ネットワーク。各エキスパートは単純なフィードフォワード型のネットワークであり、すべてのエキスパートは同じ入力を受け取り同じ数の出力ユニットを持っている。ゲーティングネットワークもフィードフォワード型のネットワークであり、エキスパートネットワークと同じ入力を受け取る。図中の g はゲーティングネットワークの出力(確率)を表し、すべてのゲーティングネットワークの和は1となる。 u はエキスパートの出力である。エキスパートネットワークの出力はゲーティングネットワークによって重み付けられた合成変量となる。

トライアングルモデルは直接経路と意味層を介した間接経路という 2 つのエキスパートネットワークを持つ ME とみなすことができる。ME において出力が混合されることと労働の分割問題は直接対応すると考えられる。すなわち、トライアングルモデルにおける労働の分割問題は ME におけるゲーティングネットワークによる領域の分割そのものである。実際のトライアングルモデル (Plaut, McClelland, Seidenberg & Patterson, 1996b) では、意味経路は実装されていない。論文中では putative semantic pathway を用いたとしか書いていない。すなわち図 5.2 中の実線部分だけがシミュレートされただけである。

ME は入力データ空間をいくつかの小領域に分割し、その分割された各領域に対してひとつのニューラルネットワークを割り当てることによって複雑な問題の解を求める手法である。この領域の分割方法については次の 3 つの実装が考えられる。以下では各々の特徴を簡単に述べる。

5.9 労働の分割問題の 3 つの実装モデル

5.9.1 素朴な労働の分割

もっとも直観的な労働の分割問題の実装には次の誤差関数を使うことであろう。

$$E^c = \left| d^c - \sum_i p_i^c o_i^c \right|^2 \quad (5.1)$$

ここで o_i^c は訓練データ c におけるエキスパートネットワーク i の出力、 p_i^c は出力ベクトルに対するエキスパートネットワーク i の貢献率、 d^c は訓練データ c における教師信号である。この誤差関数は教師信号とエキスパートネットワークの出力の合成値とを比較することになる。誤差を最小にするために、各エキスパートネットワークはその出力を、他のエキスパートネットワークの出力の合成値と教師信号の差によって定義される残差を減じるように振る舞う。あるひとつのエキスパートネットワークの結合係数が変化した場合、残差誤差が変化し、したがって他のすべてのエキスパートネットワークの誤差関数の微分値が変化する。このように密接なエキスパートネットワーク間の結び付きによって、それぞれのエキスパートネットワークが協調するようになる。しかし、個々の訓練データに対して、数多くのエキスパートネットワークが関与する解が生じやすいことになるだろう。すなわち、この場合には労働の分割というより労働の協調が生じることとなるだろう。

5.9.2 競合的労働の分割

別個のエキスパートネットワークの出力を線形結合する代わりに、各々の訓練データについてただ一つのエキスパートネットワークが使われるような確率判断をする(ゲートによって信号の伝達を制御する)ゲーティングネットワークを考えることができる。この場合の誤差関数は教師信号と実際の出力ベクトルとの差の自乗和の期待値

$$E^c = \left\langle |d^c - o_i^c|^2 \right\rangle = \sum_i p_i^c |d^c - o_i^c|^2 \quad (5.2)$$

とすることができる。この誤差関数の元では各エキスパートネットワークが出力ベクトル全体を生成するように要請される。結果として与えられた訓練データに対するローカルなエキスパートネットワークの目標は他のローカルなエキスパートネットワーク内部の結合係数の変化には直接影響されないであろう。しかし、それでもなおいくらかの間接的な影響がある。なぜなら、もし他のエキスパートネットワークの結合係数が変化すればゲーティングネットワークがそのエキスパートネットワークに割り付ける責任(信用割り付け問題)が変化するからである。もし、ゲーティングネットワークとエキスパートネットワークとがこの誤差関数の元で勾配降下法によって訓練されるのなら、システムは各訓練データに対して一つのエキスパートネットワークを割り当てるようになるだろう。あるエキスパートネットワークの出力が他のすべてのエキスパートネットワークの誤差の加重平均よりも少ない誤差を与えるときにはかならず、各エキスパートネットワークの誤差を重み付けする方法を決定するゲーティングネットワークの出力によって、そのエキスパートネットワークの責任は増大する。そして、そのエキスパートネットワークの出力が加重平均よりも悪いときには必ずその責任は減少する。

5.9.3 ガウシアン混合モデルによる労働の分割

ゲーティングネットワークの出力を、与えられたデータ空間で定義された確率エキスパートネットワークの存在確率であるとみなすのは理に適っている。

$$E^c = -\log \sum_i p_i^c \exp\left(-\frac{1}{2\sigma} |d^c - o_i^c|^2\right) \quad (5.3)$$

$$\frac{\partial E^c}{\partial o_i^c} = - \left[\frac{p_i^c e^{-\frac{1}{2\sigma} |d^c - o_i^c|^2}}{\sum_j p_j^c e^{-\frac{1}{2\sigma} |d^c - o_j^c|^2}} \right] (d^c - o_i^c) \quad (5.4)$$

すなわち各ゲーティングネットワークの出力が入力空間全体を分割するための関数になっている。

$\sigma \rightarrow 0$ の極限では Dirac のデルタ関数となり、特定のパターンにしか応答しないニューラルネットワークを作ることができる。逆に σ を大きくすれば

多くのパターンに応答するニューラルネットワークになる。前者を低頻度非一貫語、後者を一貫語、高頻度非一貫語ととらえることができるだろう。

さらに、意味層を介した間接経路のネットワークでは、非一貫語に特化した局在化した領域分割がなされていると考えれば、労働の分割問題が説明できる。各々の低頻度例外語ごとに限局された極限では、その単語のみに応答するルックアップテーブルと同一視できる。すなわち ME モデルの観点からトライアングルモデルを解釈し直すとすれば、なぜ低頻度非一貫語は意味経路を通るのかを見通しよく説明することができる。さらに、領域の分割を行なう適切なゲーティングネットワークをもつ ME により、二重経路モデルにおけるルックアップテーブルも、トライアングルモデルにおける労働の分割問題も統一的に記述できる。この意味において、二重経路モデルとトライアングルモデルの間に本質的な違いはない。両者の間に存在する違いとは、分散パラメータ σ^2 による入力空間の領域分割の大きさという量的な問題に帰結され、本質的には二重経路モデルとトライアングルモデルは同じ ME という、より一般的な範疇のニューラルネットワークモデルとして同一視できるのである。ME の数学的説明については付録 A 章を参照されたい。

以上 3 つの労働の分割問題に関する解を見てきたが、最も一般的で記述力があるのはガウシアン混合モデルによる労働の分割である。加えて、トライアングルモデルにおける労働の分割問題にもっとも近い解を与える可能性があるのもエキスパート混合ネットワークである。従って、本章ではエキスパート混合ネットワークによる ME の実装を用いて労働の分割を記述できるかどうかを議論する。

5.10 数値実験

Plaut ら (Plaut, McClelland, Seidenberg & Patterson, 1996b) の用いた 2998 単語を ME と通常のバックプロパゲーション法による学習させたものに、Plaut らの用いた非単語リスト (Glushko list) を提示した場合の正解率を表 5.1 に示した。ME の場合は 2 階層のアーキテクチャとし、それぞれ 2 つづつのエキスパートを用いた。バックプロパゲーション法による学習では Plaut らの用いた中間層ユニット数と同じ 100 個の場合と中間層 30 個の場合とを示した。また、平均二乗誤差 (MSE) を 0.05 と 0.03 の 2 種類学習させた場合の結果とを比較のために示した。

一貫語の方が非一貫語よりも成績がよいのは全てのモデルで共通していた。これは人間のデータでも示されていることであり、全てのモデルはある程度人間のデータを再現できているとみなすことができよう。労働の分割問題の解としてはどれを選んでも大差が無いように見える。したがってモデルの優劣をきめるのは、課題成績だけではなくその背後にあるモデルの持つ本質を議論しなければならないだろう。ある程度万能なアルゴリズムを用いれば任

表 5.1: Glushko(1979) の非単語リストによる般化テスト結果 (%)

	一貫語	非一貫語
人間	93.8	78.3
PMSP96	97.7	72.1
ME(2 × 2)	93.0	69.7
bp3(中間層 100, MSE=0.03)	90.7	53.5
bp3(中間層 100, MSE=0.05)	95.3	58.1
bp3(中間層 30, MSE=0.05)	88.4	58.1

PMSP96 は Plaut ら (1996) の結果. bp3 は通常の 3 層のバックプロパゲーション法による解を示している.

意の精度で望む関数を実現できることはノーフリーランチ定理が主張しているところでもある (Duda, Hart & Stork, 2001). この意味において, ME は労働の分割問題を自然に取り入れた実装になっており, 同じ成績を示すのであれば最も自然なモデルだと言うことができる.

また, バックプロパゲーション法によるモデルでは, いずれの場合でも非一貫語の読みの成績が他のモデルよりも低かった. 特に中間層のユニット数が 100 で, 平均二乗誤差 MSE が 0.03 と学習データセットを完全に学習させたモデルにおいては, 他のどのモデルよりも成績が悪かった. このことは一つの大きなアーキテクチャによるモデルでは, 遅い学習と貧弱な一般化がおこるとされる知見 (Jordan & Jacobs, 1994) を再確認したものと解釈することもできよう.

5.11 考察

意味層を介した間接経路のネットワークでは, 例外語に特化した局在化した領域分割がなされていると考えれば, 労働の分割問題が説明できる. 分散パラメータを $\sigma^2 \rightarrow 0$ とすれば領域はディラックのデルタ関数に等しくなり, 各々の低頻度例外語ごとに限局された極限では, その単語のみに応答するルックアップテーブルと同一視できる. すなわち ME モデルの観点からトライアングルモデルを解釈し直すとするれば, なぜ低頻度例外語は意味経路を通るのかを見通しよく説明することができる. ME により, 二重経路モデルにおけるルックアップテーブルも, トライアングルモデルにおける労働の分割問題も統一的に記述できる. この意味において, 二重経路モデルとトライアングルモデルの間に本質的な違いはない. 両者の間に存在する違いとは, 分散パラメータ σ^2 による入力空間の領域分割の大きさという量的な問題に帰結され, 本質的には二重経路モデルとトライアングルモデルは同じ ME という,

より一般的な範疇のニューラルネットワークモデルとして同一視できるのである。

第6章 大脳基底核と皮質領野の相互作用を仮定したモデルによるリハビリテーションのシミュレーション

6.1 はじめに

本章では失語症の回復過程に関するニューラルネットワークモデルを提案し、言語治療の技法に理論的な根拠を与えることを目的とする。すなわち、いったん学習が成立したニューラルネットワークに対して、その一部を破壊することで脳損傷を表現し、ニューラルネットワークの再学習とリハビリテーションを同一視することで失語症の回復過程の記述を試みた。

佐藤(2001)によれば、失語症の言語機能改善のアプローチには次の3つの方法がある。

1. 再活性化 (reactivation): 通常の情報処理過程の強化を図るもの
2. 再編成 (reorganization): 活性化できる情報が限られている場合、異なる情報処理過程を強化して同じ言語出力に至るようにするもの
3. 再学習 (relearning): 活用できる情報が失われてしまった場合、新たな情報の獲得を行なうことで機能の改善を図るもの

本章では、これら3つのリハビリテーション技法を用いたとき、失語症患者の内部でどのような変化が起こっているのかをニューラルネットワークを用いて表現する。

概念(あるいは意味記憶)から言語産出に至る責任領野は、ある程度明らかになっている。たとえば、Damasio と Damasio(1992)によれば、脳における言語の処理には3つの機構が存在する。単語が表す概念を蓄える概念系、単語・文章生成系、および、この2つをつなぐ媒介系の3種類である。しかし、この3種類の系に加えて、他の感覚入力からの経路を仮定しなければ、再編成や再学習をモデル化することはできないと考える。

他の感覚入力から言語産出に至る経路の候補としては大脳基底核があげられる。大脳基底核の障害では Parkinson 病や Huntington 病など運動制御系

に参与することが知られているが、大脳基底核は小脳と同じように高次認知機能に重要な役割を果たすことが示されている (Middleton & Strick, 1994). 大脳基底核の損傷によって学習や記憶などの高次認知機能に影響が現われることが明らかになってきている (Gabrieli, 1995; Jackson & Houghton, 1995). 視覚と運動の協調も大脳基底核が関与するという証拠もあり、ニューラルネットワークモデルも提案されている (Jackson & Houghton, 1995). このように大脳基底核は大脳皮質と協調しながら複雑な認知処理過程を媒介していると考えられる. 大脳基底核内部では運動制御系の回路とは並列的に高次認知処理系が存在し、大脳基底核内での横の線維連絡がなく、運動制御系とは独立の回路が存在すると考えられる (銅谷, 2000). 概念系、単語・文章生成系、媒介系の皮質上の領野は、運動野や運動前野と直接に結合し、また、左大脳基底核や左視床前部の核を介して間接的にも結合している. この二重の経路の存在は、失語症患者のリハビリテーションにおいて、皮質と皮質下の両経路を通じて治療が行われることに対応すると思われる. すなわち新たに単語を学習するときには、その意味や他の感覚モダリティとの相互結合を介して新たな回路が形成されると考えられる. 大脳基底核の損傷による仮名特異的な読字障害の症例が報告されている (橋本, 柴崎 & 利島, 2001) ことから、大脳基底核が読字過程に重要な役割を果たしていると考えられる.

本章では、上記のような神経心理学的事実から帰納される言語処理過程、すなわち意味記憶から言語産出系に至る 2 つの経路を持つニューラルネットワークモデルを定義し、言語治療による回復過程を記述する.

ニューラルネットワークの観点から上記の 3 つのリハビリテーションを考えれば、1. 再活性化の場合は、概念系から単語・文章生成系への経路の再訓練とみなすことができる. 一方、2. の再編成は、音韻手がかりや絵画・文字マッチングによるリハビリテーションと対応させて考えることができ、大脳基底核経由で学習が進むと考えられる. 3. の再学習は、リハビリテーションによって大脳基底核内部での自己組織化による応答の変化として考えることができよう.

本章では、第 2 節に大脳基底核を含む二重経路を仮定したニューラルネットワークモデルを提案し、第 3 節ではこのモデルによってニューラルネットワークの破壊実験後の失語症の回復過程とリハビリテーション技法との関連を考察している.

6.2 大脳基底核を含む二重経路モデル

意味記憶から言語産出系に至る大脳基底核を含む二重経路は浅川 (2002) のモデルを使って図 6.1 のように考えることができる. 概念系から単語・文章生成系へは直接教師信号が利用できるため、誤差逆伝播法などの教師つきの学習が可能であるとする. 一方、大脳基底核については明確な教師信号が存

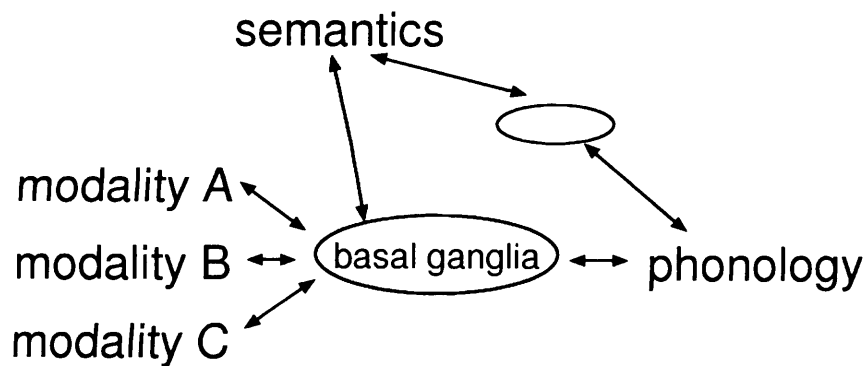


図 6.1: 大脳基底核を含む 2 重経路モデル

在しない。そこで大脳基底核では自己組織化によってモデル学習が起こると考える。実際 Mahurin(1998)は大脳基底核のモデルとして自己組織化アルゴリズムを提案しているので、自己組織化アルゴリズムを仮定するのは不自然なことではないと思われる。銅谷(2000)によれば、皮質から大脳基底核への入力部位にあたる線状体(striatum)から出力部位に相当する黒質網様部(substantia nigra pars reticulata)および淡蒼球(globus pallidus)への結合は抑制性であり、黒質網様部、淡蒼球から視床(thalamus)への結合も抑制性である。この二重の抑制の結果、大脳基底核の神経回路は視床から皮質への回路を興奮させる働きをもっていると考えられる。この生理学的事実を考慮し、大脳基底核から単語・文章生成系への結合はすべて正になるようにした。このために大脳基底核ユニットの興奮状態を +1 で表現し、静止状態を -1 とすることにした。さらに大脳基底核内のユニットの抑制的な相互作用を導入することでスパースコーディングを実現し、入力データの構造を学習させることとした。

本章では大脳基底核の学習に条件つき主成分法(Conditional Principal Component Analysis: CPCA)型の Hebb 則(O'Reilly & Munakata, 2000)を使うことにした。CPCA 則を用いる理由は、異なる条件ごとに異なる内部モデルが形成されることを意図したからである。大脳基底核の損傷によって仮名特異的な読字障害の症例が報告されている(橋本, 柴崎 & 利島, 2001)ことは、CPCA 則によって仮名だけ選択的に応答するような特化した学習がなされたと解釈することができるからである。

CPCA 則による学習に加えて、ユニット間の相互作用として一度に活性化されるユニット数を k 個に限定する kWTA (k -winners-take-all) 回路(Majani, Erlanson & Abu-Mostafa, 1989)を実装し、脳内でのスパースコーディングを実現した。kWTA の実装方法は、層内の全てのユニットに対して抑制性の信号を送る 1 つのインターニューロンと抑制性の自己結合を介して k 個のユ

ユニットが活性化されるようになるまで繰り返し計算を行ない、 k 個の勝者ユニットについてだけ CPCA 則による学習を行った。このようなインターニューロンの候補としては黒質緻密部 (substantia nigra pars compacta) が考えられる。大脳基底核では黒質緻密部から線状体へのフィードバックが存在するからである (木村, 2000)。

十分な数の大脳基底核ユニットを用意し、わざとどの特徴にも反応しない、いわゆる dead unit を作ることができる。この dead unit が後述する再編成によるリハビリテーションのときに重要な役割を果たすと考えられる。

図 6.1 に示した拡張モデルを使って、概念系が活性化したときに対応する単語・文章生成系の 1 つのユニットが活性化するように訓練された。各層におけるユニット数は、概念系 25、単語・文章生成系 12 とした。結果の視覚化

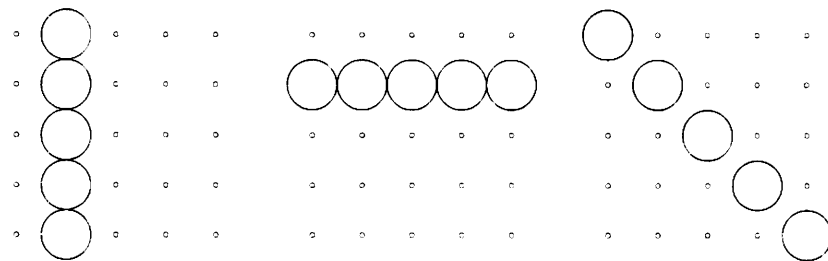


図 6.2: 用いた入力刺激の 3 例。それぞれのパターンが 1 つの概念に対応すると仮定した

を容易にするために、概念系 (入力層) の 25 個のユニットを 5×5 の正方形とみなし、縦横の各線分 10 と 2 つの対角線の計 12 パターンをそれぞれひとつ概念を表すものとして用いることにした (図 6.2)。このような入出力表現は、現実とはかけ離れていると思われるが、脳内表現がどのようになっているかはまったく不明なため、ここでは人工的なデータを使ってモデルのおおまかな振る舞いを把握することを目的とした。

単語・文章生成系 (出力層) には 12 個のユニットを用意し各概念に対応するただ 1 つのユニットだけが活性化するように訓練された。概念系から媒介系を通して単語・文章生成系への経路は前述のとおり 3 層のパーセプトロンなので誤差逆伝播則を使って学習を行った。学習にはモーメント項や崩壊法などの項は用いなかった。

6.2.1 大脳基底核の自己組織化

前述の CPCA 則と kWTA 回路を用いた大脳基底核ユニットの自己組織化による情報表現と、解くべき課題との間には自明な解が存在する場合がある。大脳基底核のユニット数を 12 としたとき、 k の値 (kWTA 回路の勝者数) を

1とすると、各概念に対して1つの大脳基底核ユニットが活性化されるようにすればよい。k=2のときには2つ複合概念を1つのユニットで表現することにし、2つの大脳基底核ユニットの論理積で1つの概念を表現するようにすればよい。このように考えればk=1,2,...,11までの値で入力情報を表現することが可能である。もっとも効率良く入力データの構造を獲得していれば、解くべき問題は入力データを12に分類することであるため、4ビット($2^4 = 16$)の分類問題であるから、最低4個の大脳基底核ユニットを用いればよい。ここでは脳内のスパースコーディングを考慮して12個の大脳基底核ユニットを用意し、同時に活性化するユニット数をk=2とした場合の結果を用いることにした。

6.2.2 皮質系の経路と大脳基底核経路との統合

概念系から媒介系を通して単語・文章生成系に至る皮質の経路が言語処理には主要な経路であることを考慮して、皮質系の経路と大脳基底核の経路の学習率に差をつけた。学習率は皮質上の媒介系0.1、大脳基底核0.01とした。

6.3 リハビリテーション

6.3.1 臨床症状とネットワークの破壊の関係

前節で説明した学習方式によってモデルに学習を行なわせ、すべての刺激に対して正しく応答するようになるまで学習を繰り返した。この学習終了時のネットワークの結合強度を健常者のモデルとみなす。そして図6.1に示したニューラルネットワークの一部を破壊することによって失語症をモデル上で表現した。概念系への損傷の場合には破壊されたユニットからの出力を0とすることによって損傷を表現した。概念系への損傷は言語の意味そのものの表象に障害が発生したことを意味することとした。媒介系への損傷についても同様にユニットの出力を0にすることで損傷を表現した。

6.3.2 破壊と回復との関係

上述のとおりニューラルネットワーク上での脳損傷のシミュレーションを実施し、破壊後残されたユニットのみを用いてニューラルネットワークを再訓練した。すなわち再訓練においては破壊されていないで、残されたユニットのみを用いて最初の訓練時と同じ学習方式で学習を行わせた。損傷の程度が重篤であればいくら再訓練を行なってもすべてにおいて正解できるわけではない。以下に示す結果は、10000回を上限として再訓練を行わせた結果である。このニューラルネットワークの再訓練をリハビリテーションと同一視

することにする。

6.3.3 リハビリテーション

6.1節で示したとおり、失語症の治療には3つの方法がある。本章で提案したモデルに対応させて考えれば、1. 再活性化の場合は、概念系から単語・文章生成系への経路の再学習とみなすことができる。一方、2. の再編成は、音韻手がかりや絵画・文字マッチングによるリハビリテーションと対応させて考えることができ、大脳基底核経由で学習が進むと考えられる。3. の再学習は dead unit を利用した再学習で、リハビリテーションによって大脳基底核内部での kWTA 回路による自己組織化による応答の変化として考えることができよう。

図 6.3 は学習が成立したネットワークに対して、概念系に損傷を加えた後、媒介系のみを再学習させたときの結合強度の変化を示している。図 6.3 では左上 3 行 3 列のユニットの出力を 0 にして (すなわち破壊された概念系のユニット数は 9) 再学習させたときの媒介系のユニットへの概念系のユニット群からの結合係数を表している。図 6.3 から左上 3 行 3 列の結合係数は変化

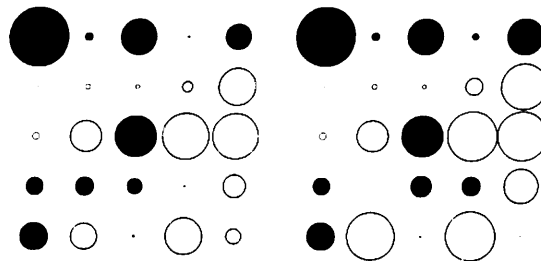


図 6.3: 概念系から媒介系への結合強度損傷前(左)と再学習後の結合強度(右)の変化

せず、残りのユニットへの結合係数が変化していることが分かる。このような結合係数の変化は 1. の再活性化に相当するものとみなすことができるだろう。

音韻手がかりによって失語症患者の呼称課題成績が向上する現象(佐藤, 2001)は、大脳基底核のユニットを活性化させることに相当すると考えることができる。課題学習である皮質経路に比べて大脳基底核はモデル学習であり、入力情報の表現に主眼がおかれている。このことは概念系に損傷が起きた場合に、以前とは異なる情報から最大限に情報を活用するように再編成がおこなわれると考えられる。図 6.4 の左が損傷前の結合係数であり、図 6.4 の右は左上 3 行 3 列の概念系のユニットが損傷を受けた場合の変化を表している。概念系(入力)が変化したことに対応して大脳基底核ユニットが残された

部分の結合係数の大きさを強調するような再構成がなされたと見なすことができる。

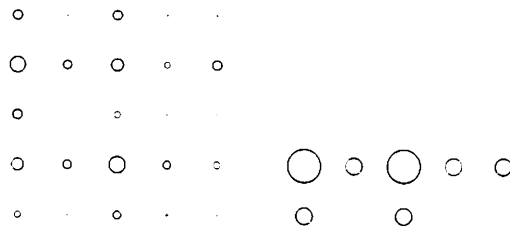


図 6.4: 再編成の結果と考えられる大脳基底核ユニットの例. 左上 3 行 3 列に損傷を加えた場合の再学習 (右)

絵画と提示語とのマッチングによる再学習の治療方法を大脳基底核ユニットを介した再学習によって説明できる可能性がある。図 6.5 に示したように、入力情報の条件つき確率を再学習することで 1 つの概念 (入力) だけでなく複数の概念 (入力) に応答するように大脳基底核ユニットが再構成されたとみなすことができる。図 6.5 は、図 6.4 と同様に左上 3 行 3 列の意味系ユニット

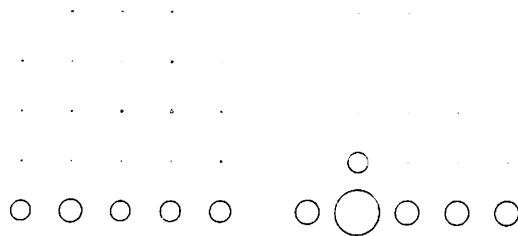


図 6.5: 再学習の結果と考えられる大脳基底核ユニットの例. 左上 3 行 3 列に損傷を加えた場合の学習前 (左) と再学習後 (右)

が破壊された場合の大脳基底核ユニットの再学習結果を示している。破壊前の図 6.5 左では 1 番下の行に対応する概念に応答するように訓練されていた大脳基底核ユニットが損傷後の再訓練で最下行の線分と右から 2 列目の垂直線分 (損傷のため左上 3 行 3 列の入力はないので) の 2 つの概念を表象するようになったと解釈可能である。

6.3.4 可能なリハビリテーション技法の組み合わせについて

再活性化、再編成、および再学習という 3 つの技法は損傷の程度にしたがって順序付けされていると考えることができる。しかしこれら技法の組合せを

考えることも、可能であろう。損傷後に残されたユニットを再訓練することと、新しいユニットを用いて再訓練することを組み合わせることが可能であると考えられる。このように考えれば、3つの技法の組み合わせ数だけ再訓練の方法が可能であると考えられる。“実際の患者さんの言語機能の改善過程は、複数の回復メカニズムが働くとするのが妥当なことが多いのです。たとえば呼称能力の改善に、単に呼称練習を行っても効果は低いものにとどまります。音読など他の情報処理過程を活性化することにより、音韻情報が強化された結果として呼称能力が改善する事例があります。この場合、活性できる情報を使って再編成を行う「再活性化+再編成」と表現することが可能だと思います(佐藤,2002,私信)。”このことから3つのリハビリテーション技法の組み合わせを考えることができよう。

本章では再活性化を皮質上の媒介系を経由する経路の再訓練としてとらえ、再編成を大脳基底核における自己組織化の変化、再学習を大脳基底核を経由する皮質下の経路の再訓練としてとらえることにした。このようにして、ニューラルネットワークの再訓練とリハビリテーションを同一視することでどの程度改善が可能かをシミュレートすることにした。以下に示す結果は再編成と再学習とを合わせた再訓練の結果を示してある。再編成と再学習とはともに皮質下の大脳基底核を含む学習であるため、再編成単独や再学習単独では提案したニューラルネットワークモデルにおける再訓練の効果がなかったためである。

6.3.5 概念系の損傷

概念系で用いた25個入力信号のうち、1, 2, 4, 9, 16個のユニットを0にすることで概念系の損傷を定義した。損傷が1個の場合は可能な25箇所すべて、損傷が2個の場合は連続した2個を、4, 9, 16個の場合にはそれぞれ正方形の領域を破壊した。シミュレーションにおいては、学習回数10000回を上限に学習を打ち切り誤差をプロットした。図6.6は概念系のユニットに損傷を加えたときの再訓練の結果を示している。図6.6は横軸に破壊したユニットの数をとり縦軸に損傷の程度の割合を示した。横軸の数は破壊したユニット数なので右へ行けば行くほど損傷が広範囲に広がっていることを表している。この図の縦軸の値は、損傷の重篤度をニューラルネットワークの標準誤差の平均値として表現している。上に行けば行くほど再訓練の効果がなかったことを示している。リハビリテーションを行わなかった場合が+の記号で表されているが、意味系のユニットを16個破壊して再学習を行わせた場合、もっとも成績が悪い。すなわちリハビリテーションが行われなければほとんど正解できないということを示している。一方、再活性化によるリハビリテーションすなわち、ニューラルネットワークモデルの再学習を行った場合、損傷の程度はきわめて低いものになる。もっとも重篤な16ユニットが破壊された場合でも、成績の低下率はほとんどなく、ほぼ100%回復して

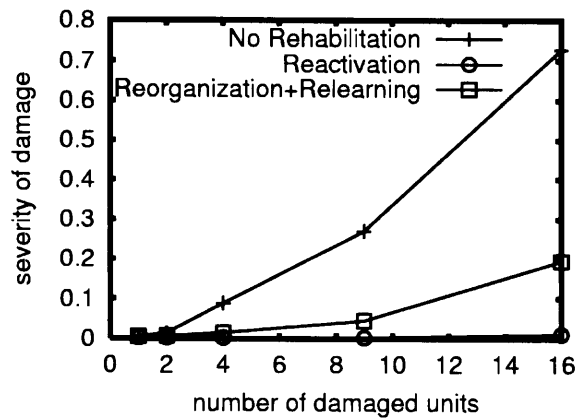


図 6.6: 意味系の損傷からのリハビリテーションの結果

いと解釈できる。再編成と再学習とを組み合わせたリハビリテーションの効果は、まったくリハビリテーションの行なわれなかった場合と再活性化によるリハビリテーションの結果中間に位置する。まったくリハビリテーションが行なわれなかった場合に比べれば、再編成と再学習を組み合わせたリハビリテーションの効果があったとみなすことができるが、もっとも重篤な損傷である 16 個のユニットを破壊した場合にはやはり成績の低下 (図では誤差の上昇) が見られた。図 6.6 から、再訓練を行わなかった場合に比べてどのリハビリテーション技法でも改善がみられるのが分かる。すなわち概念系の障害が重篤な場合でさえ、ある皮質上の経路 (再活性化) の再訓練であれ、皮質下の経路である大脳基底核経路 (再編成+再学習) の再訓練であれ、ある程度回復することが分かる。このことから概念系の損傷の場合、治療によって大幅に言語機能が回復する可能性があるかと予想できる。

6.3.6 媒介系の損傷

図 6.7 は媒介系のユニットを損傷させたときの再訓練の結果である。媒介系のユニット数を 5 個、大脳基底核のユニット数を 12 個とした。図 6.7 では任意の媒介系のユニットを破壊したときの再訓練による改善の様子を示している。

図 6.7 では破壊した媒介系のユニット数を横軸にとり、縦軸は図 6.6 と同様に、損傷の重篤度をニューラルネットワークの標準誤差の平均値として表現している。したがって上に行けば行くほどリハビリテーションの効果がなかったことを示している。いずれの方法も右に行けば行くほど縦軸の数値が大きくなっているため、損傷が重篤であればあるほど成績は悪くなることを示している。この図でもリハビリテーションを行わなかった場合にもっとも

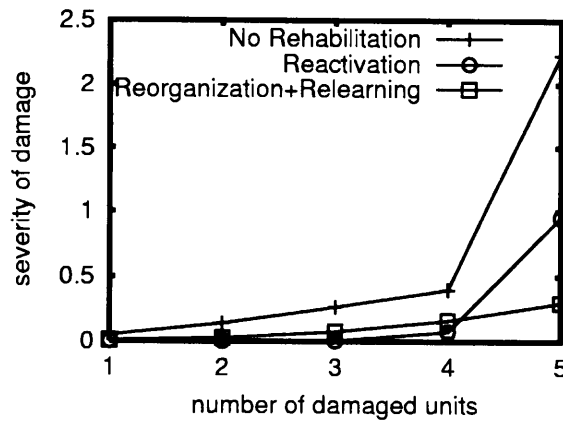


図 6.7: 媒介系の損傷からリハビリテーションの結果

誤差が大きくなる (したがって正解率は落ちる) ことを示している。媒介系のユニットを破壊したとき、ユニットが 4 つ破壊されたときと 5 つ破壊されたときとで、リハビリテーションの効果が逆転していることがわかる。すなわち媒介系のユニットを 4 つ破壊した場合、その差は小さいものの再活性化のリハビリテーションの効果がほうが、再編成と再学習とを組み合わせたリハビリテーションの効果よりもすぐれている。しかし、媒介系のユニットを 5 つ破壊した場合のリハビリテーションの効果では、再編成と再学習とを組み合わせたリハビリテーションのほうが標準誤差の平均値が小さくなり、再活性化によるリハビリテーションでは急にリハビリテーションの効果が少なくなっている (すなわちグラフ上では上にきている)。媒介系のユニットがまったく利用できない損傷数 5 の場合には、理論的に再活性化による効果がなくなり、再編成や再学習の効果がほうが改善することがわかる。このことから、媒介系の障害が重篤な場合には、再活性化によるリハビリテーションより、再編成や再学習のリハビリテーションを行った方が有益であることが示唆される。

6.4 考察

本モデルを用いれば、リハビリテーション訓練中の失語症患者の内部でどのような変化が起こっているのかを説明でき、リハビリテーション技法である、再活性化、再編成、および再学習において、媒介系、および大脳基底核のユニットの再構成、再学習を視覚化して表現することが可能であることを示した。本モデルによって言語治療の技法に理論的な根拠を与えることができるのではないかと考えている。

とくに、媒介系の損傷の場合、損傷が重篤であれば既存の皮質経路である

媒介系そのものを再訓練するよりも、他の感覚モダリティーを利用する再編成や再学習を行った方がリハビリテーションの効果が高いという点は現実の症状とも対応すると考えられる。

本章で提案した二重経路モデルの妥当性を検証するには、図 6.6, 6.7 から予測されることが実際の臨床例でみいだされれば検証されたことになるのではないだろうか。すなわち損傷が重篤な患者のリハビリテーション効果では、再編成と再学習とを組み合わせたリハビリテーションのほうが改善がみられ、損傷の軽度な患者では再活性化の効果のほうが改善がみられるとモデルでは予測できる。このような症例群がみいだされるのであれば、本章で提案したモデルを支持する結果といえるだろう。

また、levodopa の投与によって Parkinson 症の患者の認知的、および運動遂行の改善がみられることが知られている。このことから、大脳基底核は dopamine の支配を受けていると考えられる。本論文で提案したモデルに従えば、再編成と再学習による言語治療は levodopa の投与によって促進されることが予想される。本モデルでは大脳基底核における学習とは異なるモダリティー間の連合の再構成としてとらえられ、大脳基底核を介した経路を積極的に用いるリハビリテーションに対して levodopa の投与は促進的に働くと考えられるからである。

神経心理学研究にニューラルネットワークを用いる研究の利点として、モデル化されるまで意識化されなかったさまざまな変数や動作規則を明示的に指定しなければならないという点があげられる。シミュレーションを行う際には、必要なすべての変数とモデルの従う動作方程式を検討しなければならない。本章で用いた例では、概念系のユニット数、媒介系のユニット数、大脳基底核のユニット数、単語・文章生成系のユニット数、概念系、媒介系、大脳基底核のそれぞれにおける情報表現、モデルの従う学習則、学習終了とみなす判定基準、媒介系と大脳基底核との学習率の差、破壊実験における損傷の表現方法、再学習時の学習率などである。現在までのところ、これらの変数を決めるための神経心理学的証拠や生理学的証拠は存在しない。すなわち、これらの変数や動作規則が正しいかどうかを判定できない。それでもなお、本章で示したような失語症とそのリハビリテーションにおける計算論的アプローチ、あるいはシミュレーション研究が必要だと考える。適切な理論的モデルは、失語症の記述や言語セラピー計画の立案などがより適切かつ精緻になされるための枠組みとなる。

神経心理学研究にニューラルネットワークを用いるもう一つの意義として検証可能性があげられる。ボックス・アンド・アローモデルでは曖昧なままになっていた情報の表現や動作方程式を明示的に表現したことによって、本モデルが正しいかどうかを検証可能だからである。例えば本章で用いた CPCA 則と kWTA 回路は間違っているかもしれない (Mahurin, 1998)。将来別のもっと有力なアルゴリズムが開発され、新しい学習則に取って代わられるか

も知れない。しかし、この点にこそ、すなわち、古くて間違っただモデルが新しくより現実的なモデルに置き換えられる可能性にこそ、真の意味でモデル論的な価値がある。このような検証可能性を持たないモデルでは現象の理解についての進歩が望めないからである。

本章では3つのリハビリテーション技法である再活性化、再編成、再学習に焦点をあてモデル化を試みた。本モデルでは説明できない未解決な問題として、特定の治療方法との関連や個々の症例への適用などがあげられる。これらの問題に対する一般的で応用可能性のあるモデルが望まれる。そしてこのような議論なしには失語症とそのリハビリテーションに対する真の理解は生まれえないといえることができるだろう。

第III部

総合考察

第7章 ニューラルネットワークを 神経心理学へ適用する意味 について

7.1 ニューラルネットワークモデルと症例研究

ニューラルネットワークモデルによる神経心理学の検討の特徴は、現在分かっている生理学および神経心理学的事実をできる限り採り入れてモデルを構築し、未だ明らかでない点、疑わしい点には大胆な仮説を設けてモデルを作ることである。モデルの検証にはコンピュータによるシミュレーションを用いる。コンピュータ上では実際の患者には要求できないような極端な要求や課題を課すことができ、その結果得られるデータも現象の理解に役立つ。こうして得られた知識や予測をもう一度症例の検討に戻って考えてみる、という研究手法がこの分野の研究を一步前に進める力になりうるだろう。モデルだけが一人歩きした測定無き理論 (theory without measurement) も現象 (症例) の記述のみに終始した理論無き測定 (measurement without theory) も、どちらも研究としては不完全である。このうちの理論的考察を行う際に、推論の道具としてニューラルネットワークを用いることが本論文で主張した神経心理学に対するニューラルネットワークの適用である。ニューラルネットワークモデルによる脳の構成論的研究と、実際の症例研究とは車の両輪をなしていなければならないと考える。

7.2 脳の構成論的研究の意義

序論でも述べたとおりニューラルネットワークの特徴は分散表現と統計的構造の漸進的学習、および相互作用の3点に要約できる (Farah, 1994; Plaut, 2001)。また、同じく序論で述べたとおり脳とコンピュータの同型性がある。このような特徴を持ったニューラルネットワークを破壊する人工脳損傷によって、さまざまな認知機能の障害をコンピュータ上にモデル化できる。本論文では、人間の認知機能とニューラルネットワークモデルプログラムとを同一視し、かつ、ニューラルネットワークを部分的に破壊することと脳損傷を同一視することによって、神経心理学に対して新しい視点を提供できることを示してきた。その適用範囲は、神経心理学の伝統的な推論方法である二重

乖離の原則に基づく局在性仮説への批判から、具体的なリハビリテーション技法に対する理論的意味付けまで、多くの概念を説明することが可能である。

第一部で示してきたように、ニューラルネットワークモデルによる推論を行なうことで従来の神経心理学的症例分類学をみなおす必要があるのではないかと思える。意味記憶の障害だけを仮定すれば、視覚性の誤りも語彙性の誤りも説明できる。すなわち従来から考えられてきた課題成績によるアルツハイマー病の障害の分類に全く新しい視点を与えることができるように思われる。同じく第一部で示したように、生物-非生物の二重乖離に対して、一認知機能の局在を示す脳損傷患者のデータ—その認知機能を推論する伝統的な神経心理学的手法に疑問を投げかけている。同様にして、二重単純再帰型ニューラルネットワークの人工脳損傷のシミュレーションでは同一モデルで二重に乖離した二つの失文法を、文脈層ユニットに形成されたと考えられる文章産出と文章理解のために同時に利用される文法知識の障害として説明できる。しかも、この文法知識はアプリアリに与えられたものではなくニューラルネットワークの訓練の結果として文法知識が創発されたのである。ニューラルネットワーク研究によって、従来からの神経心理学的障害分類論に本質的な変更が迫られているように思える。神経心理学の道具としてニューラルネットワークモデルを用いた脳の構成論的研究の意義はまさにこの点にあると思われる。

ニューラルネットワークモデルの記述力の高さによって、5章で見てきたように、論争の一部に結着をつけるような、より高い視点から現象を観察できるようになることもニューラルネットワークモデルを用いることの大きな利点である。エキスパート混合ネットワークを導入することで、失読症のモデルである二重経路モデルとトライアングルモデルにおける直接経路と間接経路の処理の違いに関する論争は、より一般的な読みのエキスパート混合ネットワークの一形態にすぎないということが導き出せる。二重経路モデルにおけるルックアップテーブルの存在も、トライアングルモデルにおける労働の分割問題も、エキスパート混合ネットワークによる領域分割の仕方として記述可能である。すなわちエキスパート混合ネットワークにより、この論争に結着をつけることができると考えられる。

この例は、科学における論争の結着方法を思い起こさせる。すなわち、論争の争点となっている問題は、新しいモデル—新しい宇宙観あるいは世界観と呼んでもよい—によって統合された再解釈がなされるということが科学史においてしばしば起こっている。例えば、ガリレオがアリストテレスの物理学を覆し、コペルニクスがトレミーの理論を覆し、アインシュタインの理論がニュートンのそれを特殊な場合として含んでいたように、同じようなことが二重経路モデルとトライアングルモデルとの間の論争においても起こっていたと考えられる。このことは、伝統的な科学の知識体系に新しい宇宙観、世界観を導入し、既成事実を考え直すことがいかに大切かを教えてくれている。

と思えるのである。

7.3 人工脳損傷の意義

ニューラルネットワークを用いた人工脳損傷研究では特定の認知機能を遂行するためのニューラルネットワークモデルをコンピュータプログラムとして実現し、構築されたニューラルネットワークの一部を破壊することによって対応する部位が損傷を受けたときに生ずる症状をプログラムの出力として表現することをめざしている。倫理上の制約から実際の人間の脳を破壊して実験を行なうことは不可能であるから、人工脳損傷研究は、神経心理学に対して強力な道具を提供していると言えよう。例えば4章に示した2段階読字過程モデルでは視覚性障害仮説と離断仮説によって生じる障害の程度が量的に異なることを予測している。また、例えば6章に示したように障害の程度を変化させて、どのリハビリテーション技法が効率良く治療を促進するかを予測することができる。このように、人工脳損傷とリハビリテーションによる回復との関係を論じることができるようになったのもニューラルネットワークによるシミュレーションの大きな効用であろう。

伝統的な神経心理学では、二重乖離の原理による機能局在仮説に従って、特定の認知機能の機能局在が論じられてきた。第一部3章3.3節の二重単純再帰型ニューラルネットワークによる2種類の失文法の説明で示したとおりニューラルネットワーク研究の破壊実験では単一部位の損傷によって複数の認知機能障害を同時に説明できる可能性がある。反対に、複数の部位を破壊または離断しなければ高次認知機能障害は発生しない場合も存在するかも知れない。特に高次認知機能においては、シミュレーションによって確認してみるまで分からないという側面が強い。人工脳損傷研究は、神経心理学の要となる可能性を秘めている。

7.4 モデルの満たすべき要件

序論でも論じたとおり関数近似機としての完全性というパーセプトロンの強力な性質は、神経心理学にとって拘束を課すことになる。すなわち、この数学的な拘束によって、多くの心理現象がニューラルネットワークモデルとして実現可能である。換言すれば、シミュレーションを行なうだけのニューラルネットワークモデルは必ず作ることができることを意味する。従ってシミュレーションそのものを目的としたモデルは意味を持たない。モデルが、説明可能性や妥当性を有し説得力を持った神経心理学のツールとなるためには、モデルの計算論的意味付けが可能でなければならない。

すなわち、序論1.6節でも述べたとおりニューラルネットワークはあくまでツールであって理論ではない。どのように考えれば神経心理学的症状を矛

盾無く説明できるのか、あるいは、認知心理学的実験データをどのように解釈すべきなのか、といった問題に対して手段を提供しているにすぎない。

ニューラルネットワークは Marr(1982) の言う計算論の水準ではない。特に個々の現象をシミュレートする場合には、計算論的な考察がなされないことも多い。1.4 節で紹介した情報論的必然性も、入出力の関係が巧妙であればある程、そこに至るアルゴリズムはそう多くはない、ということを手帳しているだけであって、そこには計算論的考察が欠けている。ニューラルネットワークというアルゴリズムの側面だけが強調される傾向があるが、計算論レベルの説明も忘れてはならない。そうでなければニューラルネットワークモデルはポストホックな現象の記述のみに終始し、発展が望めないだろう。計算論の水準を視野にいれたモデルによって真の理論となると考える。

個々のモデルにおいては、1.6 節で述べた要件が満たされている必要がある。そうでなければ、モデルが神経心理学に貢献することはできないだろう。1.6 節の要件をもう一度記すと

1. 無矛盾性. 生理学や機能的脳画像研究から得られたデータと矛盾してはならない。
2. 予測可能性. モデルは当該の症状を説明するだけでなく、症状が進行したときの予測や、新たな症状のおきる可能性を指摘できなければならない。
3. 説明可能性. ニューラルネットワークモデルによって複数の心理指標が同時に説明されなければならない。

である。この要件が、これからのニューラルネットワークを用いた神経心理学モデルに求められている条件であると考えられる。

1 の無矛盾性はニューラルネットワークモデルにとって必須である。例えば印刷された文字を読むことを考える場合、角回、縁上回、紡錘状回後部の存在を無視してモデルを構成するのは無意味である(4章)。ニューラルネットワークを通して角回、縁上回、紡錘状回後部の計算論的役割を考察することがニューラルネットワークを用いた神経心理学の要点である。同様に、大脳基底核のモデル化を行なう場合、線状体から黒質網様部および淡蒼球への結合、黒質網様部、淡蒼球から視床への結合などの生理学的事実を考慮しなければモデルとしても意味が無い(6章)。

ニューラルネットワークモデルは生理学や機能的脳画像研究の計算論的意味を賦与するための道具となり得ると考える。PET や fMRI に代表される機能的脳画像研究では、ある実験課題を解決するために必要な部位の活性化の有無が確認されるだけである。換言すれば、その部位がどのような機能をはたしているのかの意味付けを確かめる方法ではない。脳内の一部位に対して、その部位の計算論的な意味付けをするためには、その部位が他の部位とどのような神経結合をしているのかを知らなければならない。シナプス結合

には興奮性の結合と抑制性の結合の2種類しかない。この事実は、ある部位と他の部位とのシナプス結合の強さが微妙に変わることによって、多様な高次認知機能が実現されていることを示していると思われる。このことから導き出される予想は、脳内の一部の部位の活性化だけを測定しているだけでは脳の計算論的意味付けができないのだ、ということである。脳の各部位の計算論的役割を知るためには部位間の神経結合の程度(結合係数)を知る必要がある。いまのところ非侵襲的に神経結合の強度を測定する技術は存在しない。まさにこの点にこそニューラルネットワークモデルによる脳の構成論的研究の重要性がある。シミュレーションによって初めてその部位の計算論的役割が明らかにできる可能性があるからである。

2の予測可能性については、本論文では第三部4章に示したとおり離断仮説による読みの障害と視覚性障害仮説による読みの障害とでは損傷の効果に対して傾向が異なることが予測される。また同様に、6章に示したとおり3種類のリハビリテーションの効果と損傷の程度についての予測も可能である。実際の症例と本論文で提出したモデルからの予測との比較をする研究は今後の課題である。1.1節でも述べたとおり、コンピュータ上では実際の患者には要求できないような極端な要求や課題を課すことができ、その結果得られるデータも現象の理解に役立つ。こうして得られた知識や予測をもう一度症例の検討に戻って考えてみる、という研究手法がこの分野の研究を一步前に進める力になりうると思われる。症例の検討は続けて行かなければならない。症例がモデルに変更を迫ることはあっても、モデルが症例に変更を要請することは考えられないからである。

3の説明可能性は重要な概念である。1.6節で述べたとおりニューラルネットワークは万能関数近似機としての性格を持っている。単に心理学的データの入出力関係を近似するためのニューラルネットワークモデルは必ず作ることができる。ニューラルネットワークモデルを作ることが重要なのではなく、神経心理学的症状をどれだけ矛盾無く説明できるかが重要なのである。ただし、パターン認識の分野では、すべての事例について、完全に正解を与える一般的モデルは存在しないことが知られている(Duda, Hart & Stork, 2001)。このことはニューラルネットワークモデルにもう一つの制約を与えている。すなわち、すべての現象を説明可能な万能モデル(完全なシミュレーション)を作る努力をするよりも、個々の認知課題をより良く説明するモデルの作成を目指すべきである。単に神経心理学的症状をシミュレートするだけではなく、モデルの持つ数学的性質、および能力の限界が明確でなければならない。今後この分野のモデルの目指すべき目標は、モデルによって説明される神経心理学的症状とは何か、そのモデルによって説明可能な症例はどのような症状であるかを明確にしたものでなければならないだろう。

7.5 背景となる数学的な基礎の整備の必要性

ニューラルネットワークを用いた神経心理学モデルの検討には計算論的研究が不可欠である。5章で示したように個々の症例を説明できるだけでなく、複数のモデルをその特殊な場合として含むような一般的なモデルの構築を目指す努力をすべきである。そのためには単にシミュレーションだけではなく計算論のレベルの考察が重要になってくるだろう。

ところがこの種の研究は極端に数が少ないのが現状である。確かにニューラルネットワーク研究はここ十数年で理論的な整備が進んだ。雑誌 Neural Networks, Neural Computation, Biological Cybernetics, IEEE transaction of Neural Networks など工学関係の論文誌では、ニューラルネットワークの数学的な議論がなされ、その整備がなされてきた。ところが神経心理学への応用についてはほとんど研究が存在しない。工学的には、実際に動作するモデルが重要なのであり、そのモデルが部分的に破壊されたときの挙動については、工学者の守備範囲外なのである。

今後、神経心理学的症状を説明するための数学的な理論の整備が必要である。本論文で展開した議論だけでは不十分であり、モデルの数学的検討は急務である。そしてモデルの数学的的検討は、すべての神経心理学者にとって役立つ共有財産となるだろう。

第8章 結語

8.1 要約

ニューラルネットワークの破壊実験によって表現された神経心理学的モデルの考察を行った。ニューラルネットワークを破壊することでさまざまな認知機能の障害をコンピュータ上に再現できる。人間の認知機構を理解する上でも、あるいは実際の脳損傷患者の症状を理解するためにも、コンピュータを用いたいわば人工脳損傷研究は重要だと考えられる。ニューラルネットワークによる人工脳損傷研究は、実際の脳損傷患者を扱う神経心理学に対して強力な道具を提供していると言えよう。

ニューラルネットワークは非常に記述能力が高いためにニューラルネットワークモデルを作れば大抵の心理現象はシミュレートできてしまう。この数学的な拘束により、ほとんどの心理現象はニューラルネットワークで説明できることを意味する。ほとんどの心理現象が説明可能であることは、単にシミュレーションを行なって動作するだけのモデルは必ず作ることができることを意味する。従ってそのようなモデルは生理学的、神経心理学的な妥当性を持たず、意味が無い。モデルが神経心理学に貢献するためには、無矛盾性、予測可能性、一般化可能性、の要件を満たしていなければならないことを議論した。

第一部では、神経心理学の伝統的な考え方である二重乖離の原則に基づく局在性仮説に対してニューラルネットワークから見た批判を行なった。局在性仮説と二重乖離の原則は神経心理学の原則であり、基本的なモデル構築のための推論方法だと言ってよい。この二つの原則が、ニューラルネットワークから見れば別の見方が提供できることを示した。局在性仮説は脳内で各認知モジュールがカプセル化されていて、局所的な損傷が別のモジュールに影響を与えることはほとんどないことが大前提となっているが、この考え方は単純すぎる (Farah, 1994)。二重乖離の原則は、ニューラルネットワークの相互作用的な性質によって、別の説明が可能であることを、アルツハイマー病の呼称課題における視覚性の誤りと語彙性の誤りに関する二重乖離、生物と非生物概念の二重乖離、および、機能的失文法と統辞的失文法との二重乖離、の例を用いて説明した。

第二部では具体的なニューラルネットワークモデルを議論した。日本語の文字知覚のための2段階読字過程モデル、単語の読みにおける労働の分割問

題の解としてのエキスパート混合ネットワーク、リハビリテーションへの応用と3種類のリハビリテーション技法へのニューラルネットワーク的な意味付けについてである。2段階読字過程モデルでは、離断仮説と視覚性障害仮説という2つの仮説に基づくシミュレーションを行ない、これら2つの仮説から演繹される現象が異なる課題成績を予測できることを示した。単語の読みにおける労働の分割問題の解としてのエキスパート混合ネットワークでは二重経路モデルとトライアングルモデルという論争の論点について、そして、労働の分割問題について、エキスパート混合ネットワークを用いることで統一的な見方が提供できることを示した。リハビリテーションへの応用と大脳基底核と皮質領野の相互作用を仮定したモデルによるリハビリテーションのシミュレーションでは、回復過程とニューラルネットワークの再訓練を同一視した視点からリハビリテーション技法へのニューラルネットワーク的な意味付けについての考察を行なった。

第三部ではそれまでの議論をふまえニューラルネットワークの神経心理学への適用を試みた。モデルだけが一人歩きした測定無き理論も症例の記述のみに終始した理論無き測定も、どちらも研究としては不十分であり、このうち理論的な側面を重視した考え方が本論文で主張したニューラルネットワークを用いた神経心理学の適用である。ニューラルネットワークを用いた神経心理学的モデルの考察は、実際の症例研究と車の両輪をなしていなければならないと考える。

本論文では、人間の認知機能とニューラルネットワークモデルプログラムとを同一視し、かつ、ニューラルネットワークを部分的に破壊することと脳損傷を同一視することによって、神経心理学に対して新しい視点を提供できることを示してきた。その適用範囲は、神経心理学の指導原理である二重乖離の原則に基づく局在性仮説への批判から、具体的なリハビリテーション技法に対する理論的意味付けまで、多くの概念を説明することが可能である。

8.2 おわりに —新しい脳観を求めて—

序論1.4節では新しい脳観が求められていることを指摘し、第一部では二重乖離の原則に基づく局在性仮説を批判した。本論文では、この批判から出発してニューラルネットワークモデルによる神経心理学的事実の説明が可能なることを示してきた。ニューラルネットワークの視点に立てば、伝統的な神経心理学的症状に対してまったく別の見方、いわば新しい脳観を見ることができると思われる。

第二次ニューロブームといわれた1980年代からすでに十五年以上が経過し、ニューラルネットワークという言葉も一般的な言葉になってきた。しかし、本論文で記述したようなニューラルネットワークの神経心理学への応用は、とりわけ我国においてはほとんどなされて来なかった。なぜこれほどまで

に重要な考え方 —ニューラルネットワークによる神経心理学的事実の説明— が、我国においてほとんど紹介されずに残っていたのかは信じがたい程である。一つの理由は、詳細な症例研究が最優先されるべきであるという神経心理学者のコンピュータおよびシミュレーションに対する偏見があるのではないかと思う。もう一つの理由はニューラルネットワーク研究者が神経心理学に対する応用に目を向けて来なかったということであろう。本論文ではこのような事実に鑑みて、神経心理学とニューラルネットワーク研究の橋渡しをしようとしたものである。お互いの長所と短所を認めて相互に補完しあう関係が確立すれば、必ずや実り多い研究成果が産まれるものと信じている。

神経心理学は Broca 以来 100 年の伝統があり症例研究も積み重ねられてきた。一方、ニューラルネットワークを用いた神経心理学データの検討を行う研究は 1980 年代後半に起こった第 2 次ニューロブーム以降に確立された考え方であり、理論的な整備はこれからの課題である。これからこの分野で多くの研究がなされなければならない。そしてこの理論的研究こそが神経心理学、さらに人間の認知機構を探究する認知心理学にとって必要な研究となると信じている。

8.3 謝辞

本論文は —より正確には私は— 数多くの人達に支えられてでき上がったものである。今まで私をささえてくれた人たちに感謝したい。特に、筆者に神経心理学とニューラルネットワークとの関係が実り多い分野であることに気づかせてくれた早稲田大学文学部福沢一吉先生には感謝の気持ちを表明しておきたい。

付録A エキスパート混合ネットワークの数学的解説

ここでは Jordan & Jacobs(1994) に従ってエキスパート混合ネットワーク (ME) の概説を行なう。

A.1 定式化

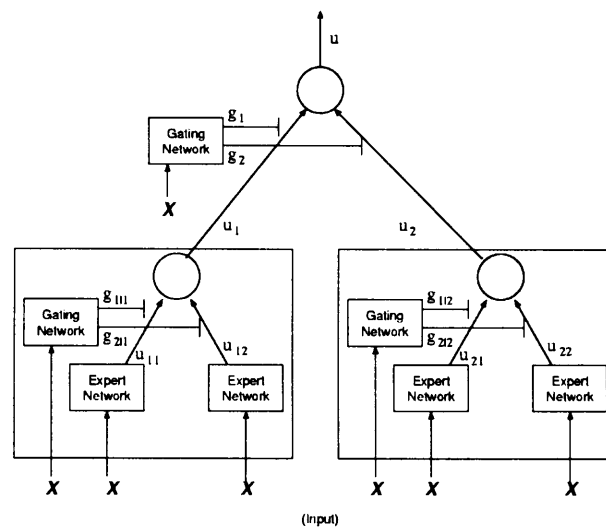


図 A.1: 2 階層のエキスパート混合ネットワーク

入力ベクトルが m 次元の実数値をとるベクトル \mathcal{R}^m で、出力が n 次元の実数値をとるベクトル \mathcal{R}^n で定義された要素であるとする。観測されたデータを $\mathcal{X} = \{x^{(t)}, y^{(t)}\}$ とする。

ME の木構造において非末端部にはゲーティングネットワークがあり、1つの数値 (確率と考えて良い) を出力する。一方、末端部分にはエキスパートネットワークがある。各エキスパートネットワークは入力ベクトル x に対して出力ベクトル μ_{ij} を出力する。各々の出力ベクトルは木を登って行きゲーティングネットワークによって混合される。

エキスパートネットワーク (i, j) は入力ベクトル x に対する出力としてベクトル μ_{ij} を出力する。

ここで u_{ij} は結合係数行列で、入力 x は切片項を表す常に 1 を出力する固定の要素を含む。関数 f のことを入力と出力を結びつけるという意味でリンク関数という。 f は問題によって線形でもシグモイド関数でも、あるいはガウシアン関数でも良い。

ゲーティングネットワークは線形であり ξ を変数とする以下の式のように表される。

$$\xi_i = v_i^T x \quad (\text{A.1})$$

ここで v_i は結合係数ベクトルである。最上位のゲーティングネットワークの i 番目の出力は ξ_i の softmax 関数として定義される。

$$g_i = \frac{e^{\xi_i}}{\sum_k e^{\xi_k}} \quad (\text{A.2})$$

g_i は正であり、すべての g_i について合計すると必ず 1 になる。すなわちソフトマックス関数は入力空間を「柔らかく」分割するものである。同様にして下位のゲーティングネットワークも線形であり、出力 ξ_{ij} は以下のように定義される。

$$\xi_{ij} = v_{ij}^T x \quad (\text{A.3})$$

このとき

$$g_{j|i} = \frac{e^{\xi_{ij}}}{\sum_k e^{\xi_{ik}}} \quad (\text{A.4})$$

は 2 段目の i 番目のゲーティングネットワークにおける j 番目のゲーティングネットワークの出力である。ここでも $g_{j|i}$ は正であり各 x に対して足し合わせると 1 になる。

非末端部 (ノード) における出力は、そのノードの下にあるエキスパートネットワークの出力に各ゲーティングネットワークの重みをかけた出力になる。すなわち第 2 層のノード部における出力は

$$\mu_i = \sum_j g_{j|i} \mu_{ij} \quad (\text{A.5})$$

となり、木の最上位層ゲーティングネットワークの出力は

$$\mu = \sum_i g_i \mu_i = \sum_i g_i \sum_j g_{j|i} \mu_{ij} \quad (\text{A.6})$$

となる。 g も μ も入力 x に依存している、ゆえに、全体の出力は入力ベクトル x の非線型関数となる。

A.2 ridge 回帰

ソフトに分割された領域とは、領域間に重なりがあると言う意味である。この重複の性質を理解するために2つのエキスパートからなる1つの階層を考えてみることにする。この場合、ゲーティングネットワークは g_1 と g_2 との2つの出力を持つ。ゲーティング出力 g_1 は

簡単のためゲーティングネットワークが2つの場合を考える。

$$g_1 = \frac{e^{\xi_1}}{e^{\xi_1} + e^{\xi_2}} \quad (\text{A.7})$$

$$= \frac{1}{1 + e^{-(v_1 - v_2)x}} = 1 - g_2 \quad (\text{A.8})$$

と表される。上式は $v_1 - v_2$ というベクトルの方向によって決まるシグモイド関数である。ゲーティング出力 g_2 は $1 - g_1$ に等しい。ある x が与えられると全体の出力 μ は $g_1\mu_1 + g_2\mu_2$ となる。すなわちエキスパートの重みづけ平均。 $g_1 = g_2 = 1/2$ というリッジに沿って2つのエキスパートは等しく貢献する。

リッジから離れるに従って1つのエキスパートの影響が大きくなり、他方の影響は小さくなる。

$$g_1 = g_2 = \frac{1}{2} = \frac{1}{1 + e^{-(v_1 - v_2)x}} \quad (\text{A.9})$$

$$1 = \frac{2}{1 + e^{-(v_1 - v_2)x}} \quad (\text{A.10})$$

$$e^{-(v_1 - v_2)x} = 1 \quad (\text{A.11})$$

$$-(v_1 - v_2)x = 0 \quad (\text{A.12})$$

という計算により $v_1 - v_2$ に直交するベクトルの方向では2つのエキスパートの影響は等しくなる。リッジを横切る滑らかさの度合は $v_1 - v_2$ というベクトルの大きさによって決まる。もし $v_1 - v_2$ が大きければリッジ関数は2つの領域を明確に分離することになり、エキスパートの重み付された出力は線形に近くなる。もし、 $v_1 - v_2$ が小さいのなら各エキスパートはリッジの両側において意味のある程度まで貢献し、滑らかなマップになる。 $v_1 - v_2$ の値が大きい程分離の程度が大きくなり、この重みが0になる極限では両エキスパートは等しく貢献することになる。

A.3 ME の確率的解釈

ME の階層を確率的に解釈することができる。データが環境によって生成されるメカニズムは x を y へ写像する回帰が木の末端にくる意志決定の入れ子構造と仮定。意思決定は多次元確率変数としてモデル化。すなわち、す

すべての x について、 g_i の値を、最初の意味決定に関する多次元確率と解釈し、 $g_{j|i}$ を g_i が与えられたときの条件付き多次元確率とする。

ME の出力ベクトル y を、入力データベクトル x とデータの密度関数を表すパラメータ θ とを用いて条件付確率として定式化することができる。

$$P(y|x, \theta) = \sum_i g_i(v_i, x) \sum_j g_{j|i}(v_{ij}, x) p_{ij}(y|x, \theta_{ij}) \quad (\text{A.13})$$

$$P(y|x, \theta^0) = \sum_i g_i(v_i^0, x) \sum_j g_{j|i}(v_{ij}^0, x) P(y|x, \theta_{ij}^0) \quad (\text{A.14})$$

ここで θ_{ij}^0 は分布の密度関数を定めるパラメータベクトルである。

例えば P_{ij} データが多次元正規分布に従い、共分散行列が $\Sigma = \sigma^2 I$ (I は単位行列) であるならば、データ x と θ とが与えられたときの密度関数は次のように書くことができる。

$$P(y|x, \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \sum_i g_i \sum_j g_{j|i} e^{-(1/2\sigma^2)(y-\mu_{ij})^T(y-\mu_{ij})} \quad (\text{A.15})$$

分散パラメータ σ^2 は μ_{ij} を中心とする円の大きさ (半径) を定めるものと解釈することが出来る。 $\sigma^2 \rightarrow 0$ の極限ではディラックのデルタ関数となる。

y が離散的な 2 値関数、すなわち多次元ベルヌーイ試行ならば

$$p(y|x, \theta) = \sum_i g_i \sum_j g_{j|i} u_{ij}^y (1 - u_{ij})^{1-y} \quad (\text{A.16})$$

となる。

A.4 パラメータの推定

ME の学習を考えるために ME の木構造のノードにおける事後確率 h を定義する。 g_i と $g_{j|i}$ とは事前確率と呼ぶことにする。 $g_i, g_{j|i}$ は入力データベクトル x だけで決まり、対応する出力ベクトル y の知識なしに定まるからである。事後確率は入力ベクトル x と出力ベクトル y の両方が分かったとき定義される。ベイズの定理を用いて i 番目のノードの事後確率は次のように定義される。

$$h_i = \frac{g_i \sum_j g_{j|i} P_{ij}(y|x)}{\sum_i g_i \sum_j g_{j|i} P_{ij}(y|x)} \quad (\text{A.17})$$

$$h_{j|i} = \frac{g_{j|i} P_{ij}(y|x)}{\sum_j g_{j|i} P_{ij}(y|x)} \quad (\text{A.18})$$

g_i と $g_{j|i}$ は (A.2), (A.3) 式で定義されている。

また、 h_i と $h_{j|i}$ との積、すなわち結合事後確率 h_{ij} を定義しておく。この量はエキスパートネットワーク (i, j) がデータを生成したと見なすことができる確率。

$$h_{ij} = \frac{g_i g_{j|i} P_{ij}(y|x)}{\sum_i g_i \sum_j g_{j|i} P_{ij}(y|x)} \quad (\text{A.19})$$

N 個のデータ集合 $\mathcal{X} = \{(\mathbf{x}^{(t)}, \mathbf{y}^{(t)})\}_1^N$ から次の対数尤度関数が得られる.

$$l(\theta; \mathcal{X}) = \sum_t \ln \sum_i g_i^{(t)} \sum_j g_{j|i}^{(t)} P_{ij}(\mathbf{y}^{(t)} | \mathbf{x}^{(t)}) \quad (\text{A.20})$$

A.5 EM アルゴリズム

EM アルゴリズムによって ME のパラメータを求めることを考える. まず尤度関数が簡単になるような隠れ変数を定義する. 変数 z_i はすべての i の中のただ 1 つの z_i だけが 1 となり, 他の全ては 0 であるとする. 同様に変数 $z_{j|i}$ は 1 つの $z_{j|i}$ だけが 1 であり, 他はすべて 0 であるとする. これらの隠れ変数は確率モデルにおける決定に対応したラベルであると解釈できる. また z_i と $z_{j|i}$ との積である z_{ij} を定義しておく. z_{ij} は特定のエキスパートネットワークを示すラベルであると解釈できる. もし, ラベル $z_i, z_{j|i}, z_{ij}$ が分かれば尤度最大化問題は各エキスパートネットワークの回帰問題に分離され, ゲーティングネットワークについては分割された集合の分類問題に切り離され, これらの問題は互いに独立に解くことができる. もちろん隠れ変数は未知であるが, 隠れ変数と観測可能なデータとを結びつける確率モデルを特定することができる. この確率モデルは z_{ij} の項を用いて

$$P(\mathbf{y}^{(t)}, z_{ij}^{(t)} | \mathbf{x}^{(t)}, \theta) = g_i^{(t)} g_{j|i}^{(t)} P_{ij}(\mathbf{y}^{(t)} | \mathbf{x}^{(t)}) \quad (\text{A.21})$$

$$= \prod_i \prod_j \left\{ g_i^{(t)} g_{j|i}^{(t)} P_{ij}(\mathbf{y}^{(t)} | \mathbf{x}^{(t)}) \right\}^{z_{ij}^{(t)}} \quad (\text{A.22})$$

と書くことができる. この確率の対数をとると次の完全データの尤度が求められる.

$$l_c(\theta; \mathcal{Y}) = \sum_t \sum_i \sum_j z_{ij}^{(t)} \ln \left\{ g_i^{(t)} g_{j|i}^{(t)} P_{ij}(\mathbf{y}^{(t)}) \right\} \quad (\text{A.23})$$

$$= \sum_t \sum_i \sum_j z_{ij} \left\{ \ln g_i^{(t)} + \ln g_{j|i}^{(t)} + \ln P_{ij}(\mathbf{y}^{(t)}) \right\} \quad (\text{A.24})$$

上式で表現された完全データ尤度と (A.20) で定義された不完全データ尤度の関係に注意されたい. 指示変数 z_{ij} を用いることで対数を加算記号の内側に持って来ることが可能になる. このことによって最大化問題を簡単に行うことができる.

完全データ尤度の期待値をとることで EM アルゴリズムの E ステップを

$$Q(\theta, \theta^{(p)}) = \sum_t \sum_i \sum_j h_{ij}^{(t)} \left\{ \ln g_i^{(t)} + \ln g_{j|i}^{(t)} + \ln P_{ij}(\mathbf{y}^{(t)}) \right\} \quad (\text{A.25})$$

と定義する。ここで以下の事実を用いた。

$$E \left[z_{ij}^{(t)} | \mathcal{X} \right] = P \left(z_{ij}^{(t)} = 1 | \mathbf{y}^{(t)}, \mathbf{x}^{(t)}, \boldsymbol{\theta}^{(p)} \right) \quad (\text{A.26})$$

$$= \frac{P \left(\mathbf{y}^{(t)} | z_{ij} = 1, \mathbf{x}^{(t)}, \boldsymbol{\theta}^{(p)} \right) P \left(z_{ij}^{(t)} = 1 | \mathbf{x}^{(t)}, \boldsymbol{\theta}^{(p)} \right)}{P \left(\mathbf{y}^{(t)} | \mathbf{x}^{(t)}, \boldsymbol{\theta}^{(p)} \right)} \quad (\text{A.27})$$

$$= \frac{P \left(\mathbf{y}^{(t)} | \mathbf{x}^{(t)}, \boldsymbol{\theta}_{ij}^{(p)} \right) g_i^{(t)} g_{j|i}^{(t)}}{\sum_i g_i^{(t)} \sum_j g_{j|i}^{(t)} P \left(\mathbf{y}^{(t)} | \mathbf{x}^{(t)}, \boldsymbol{\theta}^{(p)} \right)} \quad (\text{A.28})$$

$$= h_{ij}^{(t)} \quad (\text{A.29})$$

すなわち z_{ij} の期待値は事後確率 h_{ij} になる。同様にして $E \left[z_i^{(t)} | \mathcal{X} \right] = h_i^{(t)}$, $E \left[z_{j|i}^{(t)} | \mathcal{X} \right] = h_{j|i}^{(t)}$ が導出できる。

EM アルゴリズムの M ステップでは $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(p)})$ をエキスパートネットワークパラメータとゲーティングネットワークパラメータに関して最大化する必要がある。隠れ変数の導入によりエキスパートネットワークパラメータは $h_{ij}^{(t)} \ln P(\mathbf{y}^{(t)})$ をとおしてだけ Q に影響し、ゲーティングネットワークパラメータは $h_{ij}^{(t)} \ln g_i^{(t)}$ と $h_{ij}^{(t)} \ln g_{j|i}^{(t)}$ とをおしてだけ Q に影響する。従って M ステップは以下のような個別の最大化問題に帰結する。

$$\boldsymbol{\theta}_{ij}^{(p+1)} = \arg \max_{\boldsymbol{\theta}_{ij}} \sum_t h_{ij}^{(t)} \ln P_{ij}(\mathbf{y}^{(t)}) \quad (\text{A.30})$$

$$\mathbf{v}_i^{(p+1)} = \arg \max_{\mathbf{v}_i} \sum_t \sum_k h_k^{(t)} \ln g_k^{(t)} \quad (\text{A.31})$$

$$\mathbf{v}_{ij}^{(p+1)} = \arg \max_{\mathbf{v}_{ij}} \sum_t \sum_k h_k^{(t)} \sum_l h_{l|k}^{(t)} \ln g_{l|k}^{(t)} \quad (\text{A.32})$$

個々の最大化問題はそれ自身尤度最大化問題に帰結する。

引用文献

- Arbib, M. A. (1989). *The metaphorical Brain 2: Neuran Networks and Beyond*. John Wiley and sons. (邦訳 ニューラルネットワークと脳理論 第2版, 金子隆芳訳, 1992, サイエンス社).
- 浅川伸一, (2001a). ニューラルネットワークの数理的基礎. 守一雄, 都築誉史, 楠見孝 (編著), *コネクショニストモデルと心理学* 第12章, (pp. 166-203). 北大路書房.
- 浅川伸一, (2001b). 脳損傷とニューラルネットワークモデル — 神経心理学への適用例 —. 守一雄, 都築誉史, 楠見孝 (編著), *コネクショニストモデルと心理学* 第5章, (pp. 51-66). 北大路書房.
- 浅川伸一, (2002). 認知神経心理学モデルとしてのニューラルネットワーク — 失語症をめぐる —. *電子情報通信学会ニューロコンピューティング研究会技術報告書 NC2001-84(2002-1)*, 101(615), 17-24.
- Bishop, C. (1995). *Neural Networks for Pattern Recongnition*. Oxford University Press.
- Bradshaw, J. L., Nettleton, N. C. & Geffen, G. (1971). Ear differences and delayed auditory feedbac: Effects on a speech and a music task. *Journal of Experimental Psychology*, 91, 85-92.
- Coltheart, M. (1983). *The right hemisphere and disorders of reading*. Academic Press, London. In A.W.Young(Ed.), *Functions of the right cerebral hemisphere*.
- Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distriputed-processing approaches. *Psychological Review*, 100(4), 589-608.
- Coltheart, M. & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1197-1211.
- Cun, L., Denker, J. S. & Solla, S. A. (1990). Optimal brain damage. In Touretzky, D. S. (Ed.), *Advances in Neural Inforamation Processing Systems (Denver)* (pp. 589-605). San Mateo: Morgan Kaufmann.

- Damasio, A. R. & Damasio, H. (1992). Brain and language. *Scientific American*, 267, 88-95.
- Dell, G., Schwartz, M., Martin, N., Saffran, E. & Gagnon, D. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104.
- Dell, G. S., Chang, F. & Griffin, Z. M. (2001). Connectionist models of language production: Lexical access and grammatical encoding. In M. H. Christiansen & N. Charter (Eds.), *Connectionist Psycholinguistics* chapter 7, (pp. 212-243). Westport, CT: Ablex Publishing.
- 銅谷賢治, (2000). 大脳基底核のモデル. 甘利俊一, 外山敬介 (編), 脳科学大辞典 3.11, (pp. 607-610). 朝倉書店.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384, 159-161.
- Duda, R. O., Hart, P. E. & Stork, D. G. (2001). *Pattern Classification, 2nd Edition*. John Wiley and sons. (邦訳 パターン識別, 尾上守夫監訳, 2001, 新技術コミュニケーションズ).
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking Innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press. (邦訳「認知発達と生得性」, 乾, 今井, 山下訳, 共立出版).
- Fahlman, S. E. & Lebiere, C. (1990). The cascade-correlation learning architecture. In D. Touretzky (Ed.), *Advances in Neural Information Processing Systems*, Volume 2 (pp. 524-532). Morgan-Kaufman.
- Farah, M. J. (1990). *Visual Agnosia*. MIT press, Cambridge, MA.
- Farah, M. J. (1994). Neuropsychological inference with an interactive brain: A critique of the locality assumption. *Behavioral and Brain Sciences*, 17, 43-104.
- Farah, M. J. & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120(4), 339-357.

- 堀田健一, 岩田彰, 松尾啓志, 鈴木宣夫. (1992). 大規模ニューラルネット CombNET-II. *電子情報通信学会論文誌 J75-D-II(3)* pp. 545-553.
- 伊集院睦雄, 伏見貴夫, 辰巳格, (2000). 漢字・仮名で書かれた単語・非単語の音読に関するトライアングル・モデル (2). *失語症研究*, 20(2), 127-135.
- 岩田誠, (1996). *脳とことば — 言語の神経機構*. 共立出版.
- Jackson, S. & Houghton, G. (1995). Sensorimotor selection and the basal ganglia: a neural network model. In J. C. Houk, J. L. Davis & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* chapter 17, (pp. 337-367). MIT press.
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J. & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, 3, 79-87.
- Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, 78, 193-202.
- Jordan, M. I. & Jacobs, R. A. (1994). Hierarchical mixtures of experts and the em algorithm. *Neural Computation*, 6, 181-214.
- 河村満, 溝渕淳, (1998). 読みの脳内機構. 荻阪直行 (編), *読み — 脳と心の情報処理* (pp. 185-216). 朝倉書店.
- 木村實, (2000). 大脳基底核. 甘利俊一, 外山敬介 (編), *脳科学大辞典* 5.5.1, (pp. 333-337). 朝倉書店.
- Kohonen, T. (1985). *Self-Organizing Maps*. Springer-Verlag.
- MacKay, D. J. C. (1991). Bayesian model comparison and backprop nets. In J. Moody, S. Hanson & R. Lippmann (Eds.), *Advances in Neural Information Processing Systems*, Volume 4 (pp. 839-846). Morgan-Kaufman.
- Mahurin, R. K. (1998). Neural network modeling of basal ganglia function in parkinson's disease and related disorders. In R. W. Parks, D. S. Levine & D. L. Long (Eds.), *Fundamentals of Neural Network Modeling — Neuropsychology and Cognitive Neuroscience* chapter 14, (pp. 331-355). MIT Press.
- Majani, E., Erlanson, R. & Abu-Mostafa, Y. (1989). On the k-winners-take-all network. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems*, Volume 1 (pp. 634-642). San Mateo, CA: Morgan Kaufmann.
- Marr, D. (1982). *Vision*. W. H. Freeman: MIT Press.

- McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133. in Neurocomputing, Anderson, J.A. and Rosefeld, E., Neurocomputing, MIT press, chapt 2, 1988.
- Middleton, F. A. & Strick, P. L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science*, 266, 458-461.
- Morton, J. (1980). *The logogen model and orthographic structure*. Academic Press. In U. Firth(Ed.), Cognitive processes in spelling.
- Murata, N., Yoshizawa, S. & Amari, S. (1994). Network information criterion — determining the number of hidden units for an artificial neural network model. *IEEE Transactions on Neural Networks*, 5(6), 865-872.
- 中野馨 (1995). 脳をつくる — ロボット作りから生命を考える. 共立出版.
- Nobre, A. C., Aillson, T. & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature*, 372, 260-263.
- O'Reilly, R. C. & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding in mind by simulating the brain*. MIT Press.
- Patterson, K. (1990). Alexia and neural nets. *神経心理学*, 6(2), 90-99.
- Patterson, K., Seidenberg, M. S. & McClelland, J. L. (1989). Connections and disconnections: acquired dyslexia in a computational model of reading processes. In R. G. M. Morris (Ed.), *Parallel Distributed Processing: Implications for Psychology and Neurobiology* chapter 7, (pp. 131-181). Oxford University Press.
- Plaut, D. C. (2001). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. In M. H. Christiansen & N. Charter (Eds.), *Connectionist Psycholinguistics* chapter 8, (pp. 244-278). Westport, CT: Ablex Publishing.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S. & Patterson, K. (1996a). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56-115.

- Plaut, D. C., McClelland, J. L., Seidenberg, M. S. & Patterson, K. (1996b). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56-115.
- Plaut, D. C. & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10(5)*, 377-500.
- Rumelhart, D. E., McClelland, J. L. & The PDP Research Group (Eds.). (1986). *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*. Cambridge, MA: MIT Press.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Volume 1 chapter 8, (pp. 318-362). Cambridge, MA: MIT Press.
- 櫻井靖久, (1998). Pet からみた読字の神経機構. 荻阪直行 (編), *読み —脳と心の情報処理* (pp. 217-238). 朝倉書店.
- 佐藤ひとみ, (2001). *臨床失語症学*. 医学書院.
- Seidenberg, M. S. & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96(4)*, 523-568.
- Seidenberg, M. S., Petersen, A., Plaut, D. C. & MacDonald, M. C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22(1)*, 48-62.
- Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L. & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *20(6)*, 1177-1196.
- Tippett, L. J. & Farah, M. J. (1998). Parallel distributed processing models in alzheimer's disease. In R. W. Parks, D. S. Levine & D. L. Long (Eds.), *Fundamentals of Neural Network Modeling: Neuropsychology and Cognitive Neuroscience* chapter 17. MIT press.
- 上坂吉則, (1993). *ニューロコンピューティングの数学的基礎*. 近代数学社.
- フォン・ノイマン (1975). *自己増殖オートマトンの理論*. 岩波書店. (原典: *Theory of Self-Reproducing Automata*, John von Neumann, edited and completed by Arthur W. Burks, 1966, the University Illinois Press).

Warrington, E. K. & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107, 829-854.

山鳥重, (1985). 神経心理学入門. 医学書院.