

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 —

浅川伸一 <asakawa@twcu.ac.jp>

本章では、ニューラルネットワーク研究者と神経心理学研究者との共同研究を紹介している。これらの研究では、コンピュータ上に特定の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実現し、回路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している。さらに、計算機上でのモデルに対して、様々な操作を加えることで、その症状についての真の理解、病状予測、治療計画の立案、予後の予測など、有効な情報を引出すことができると考えられる。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事実上不可能であることを考えれば、ニューラルネットワーク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るだろう。ここでは、文献としては古典に属するが、この分野のパイオニアたちの初期の仕事を紹介する。いずれも、コンピュータ上に実現された回路網の破壊実験結果と患者の課題成績とを比較することで脳内機構を解明しようとした試みである。最初にパターソン、ザイデンバーグ、マクレランドら (Patterson, Seidenberg & McClelland, 1989) のトライアングルモデルを使った失語症研究を、次にファラーとマクレランド (Farah & McClelland, 1991) の意味記憶の構造に関する研究を紹介する。最後に、損傷からの回復過程もシミュレーションできることを示している。

1 ボックス-アンド-アロー・モデルを越えて

神経心理学の分野ではボックス-アンド-アロー・モデルを元にした考察と症例検討が行なわれることが多い。課題遂行に必要な下位機能をそれぞれのボックスで表現し、ボックス間を流れる情報を矢印で表現する。脳内の各領域で行なわれている処理がボックスに対応し、領域間の線維連絡が矢印に相当する。従って神経心理学症状は、ボックスで表現されている機能が損なわれる (損傷) か、または、ボックス間を結ぶ矢印が途切れる (離断) ことによって症状を記述しようとするのがボックス-アンド-アロー・モデルの神経心理学的解釈である。ここには、異なる症状が異なる脳内機構を反映しているという暗黙の了解が存在している。このようなボックス-アンド-アロー・モデルの考え方は一定の成果を挙げてきたし、現在も有効な手法である。

図 1 には典型的なボックス-アンド-アロー・モデルであるコルトハート (Coltheart, 1985) の二重経路 (dual route) 仮説を示した。欧米の失読患者の

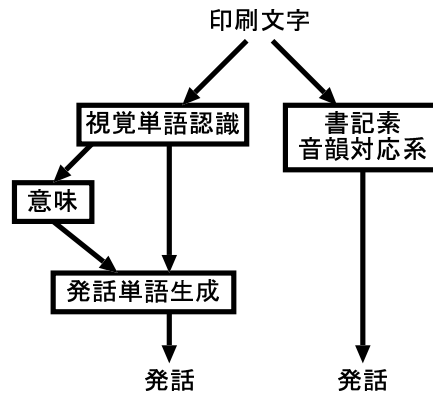


図 1: コルトハートの二重経路仮説

中には、単語の綴りと発音と関係が規則的な単語 (regular word) と不規則な単語 (irregular もしくは exceptional word) との間で乖離が見られる患者が存在する。このような乖離を説明するために考え出されたのが二重経路仮説である。図 1 の左の経路が傷害されると書記素音韻対応規則に従う読みしか出来なくなる表層失読 (surface dyslexia) と呼ばれる症状が現れる。一方、右の経路のうち特に意味系が傷害されると tall を long、her を woman などと読み誤る深層失読 (deep dyslexia) と呼ばれる症状が現れる。ただし、深層失読については無意味語の音読についても障害されることがこの症状を特定するときに重要な判断材料とされているので、深層失読については両経路の複合的な障害と考えられる。

ボックス-アンド-アロー・モデルは患者の症状を良く説明しているため有効であると考えられている。しかし、ボックス-アンド-アロー・モデルに対して以下のような 2 つの本質的な欠点を指摘できる。

1. モデル内を流れる情報についての言及が曖昧であるため、典型的な反応パターンがどのようにして生じ、反応パターンがなぜ一貫しているのかが説明できない。それぞれの経路における情報処理様式や神経回路の損傷によって生じる誤処理、誤動作のモデルを考える必要がある。
2. 伝統的な神経心理学では、2 重乖離の原理 (principle of double dissociation)、離断の原理 (principle of disconnection) など (山鳥 (1985)) によって特定の認知機能の機能局在が論じられる。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。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。

ニューラルネットワークモデルの存在意義はまさにこのような点にある。一旦学習が成立したネットワークを破壊することで症状を模倣させることができるのである。3 節に示す研究は図 1 の二重経路仮説に対して疑問を投げかける研究である。この研究で採用されたモデルでは、単一経路モデルでも異なる読みのパターンが再現できることが示される。パターソンら (Patterson, Seidenberg & McClelland, 1989) の言葉を借りれば二重経路仮説という‘セントラルドグマ’への挑戦である。

なお、コルトハートは自ら二重経路仮説に基づくニューラルネットワークモデルを提出していることを付記しておく (Coltheart, Curtis, Atkins & Haller, 1993; Coltheart & Rastle, 1994; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001)。

2 ニューラルネットワークによる脳損傷の表現方法

回路の一部を破壊することによって脳損傷の症状がコンピュータ上で再現できることを示す前に、脳損傷に対応するニューラルネットワークの破壊方法を整理しておく。それぞれの方法が、特定の病変に対応するため、破壊方法を適切に選択することが重要になってくる。

損傷をシミュレートするためには

1. 素子の除去
2. 結合の切断
3. 雑音の付加
 - (a) 入力そのものに雑音を加える場合
 - (b) 学習が成立した結合係数に雑音を加える場合
4. 活性値の制限
 - (a) 上限を下げる
 - (b) しきい値を高く設定する

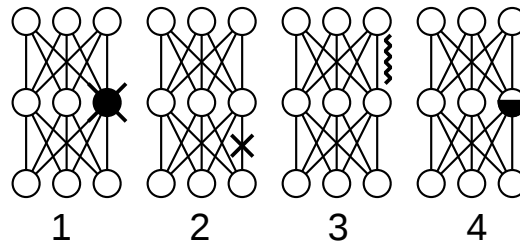


図 2: 損傷の種類

などが考えられる。

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、皮質にある神経組織そのものが損傷を受けた場合に相当する。例えば、3層のニューラルネットワークを考えた場合には、入力層の素子を取り除くことは入力が無視、あるいは半則空間無視や部分盲に相当する。一方出力層の素子を取り除くことは麻痺などの症状に対応すると考えられる。高次脳機能を扱った研究では中間層素子を取り除くことがよく行なわれている。中間層素子は刺激の内部表現である、あるいは、中間層素子は脳内に蓄えられた記憶表現である、と捉えれば取り除かれた中間層によって様々な機能障害が発生すると予想される。

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。神経心理学の用語では離断仮説に対応する。この場合、中継核のニューロンが損傷を受けた場合の症状、もしくは皮質下 (白質) の連絡線維に脳硬塞が発生した症状に対応する。

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。(a) の場合は回路への入力が悪化していることを表現し、(b) で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。

活性値に制限を設けることは損傷の結果、素子の性能が悪化した結果と解釈できる。上限を設定する方法には一定レベル以上の活性値を強制的にある値にしてしまう方法と関数の形を変化させる方法とがある。図 3 では、関数の形を最大 0.7 にしかならない場合 (a) と、しきい値を 5 に設定して活動しにくくした場合 (b) とのジグモイド関数 (S 字状の曲線) の概形が描かれている。オリジナルのジグモイド関数に比べてより大きな値でも活性化しにくいことが読み取れる。

ここに示した幾つかの方法と実際の脳損傷との対応を取ることは難しい。梗塞や血栓、あるいは外傷などでの損傷の程度によっては、上に挙げた複数の破壊方法を組み合わせて用いることも考えられる。

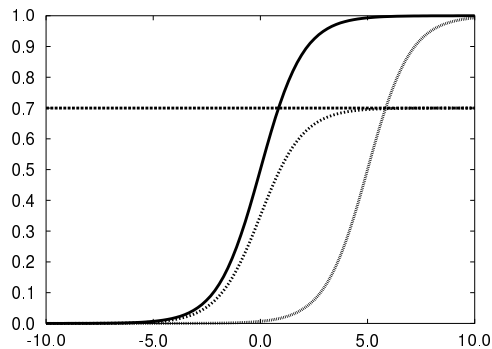


図 3: 活性値に制限を設ける方法。

3 トライアングルモデルによる失読症のシミュレーション — パターソンらの研究

パターソン、ザイデンベルク、マクレランド (Patterson, Seidenberg & McClelland, 1989) はニューラルネットワークモデルによって脳損傷による読みの障害—難読症を説明しようと試みている。彼らのモデルは 1 節で紹介した表層失読患者の典型的な読み誤りを例証できる。彼女らは、失読症患者の示す読み誤りの症状の中には、正しい発音と読み誤りとの間の音韻的特徴の類似性が関係していることを明らかにした。彼女らのモデルでは視覚提示された単語 PINT を */pAnt/* と読み間違えるような誤りが生じる¹。この種の誤りは従来、視覚的誤りとして分類されてきた。なぜなら英語では単母音文字 I は決して */A/* と発音れることはないので、患者は刺激文字 PINT を PAINT と見誤ったと解釈されてきたからである。ところが彼女らのモデルでは入力 PINT を与えても $\rightarrow$ */pAnt/* のような誤りを生じることがある。このことから、視覚的誤り分類されていた読みのエラーは、必ずしも視覚に起因する誤りとは言えないことを示している。PINT $\rightarrow$ */pAnt/* の読み誤りは、全く同じ読み誤りが実際の患者でも報告されており、興味深い結果と言える。

3.1 モデルの構成

モデルはルーメルハートとマクレランド (Rumelhart & McClelland, 1986) に基づいている (図 4)。各層の素子数は書記素層 400、中間層 200、音韻層 460 であり、各層から次層への結合が全結合である。すなわち書記素層 $\rightarrow$ 中間層間の結合数は $400 \times 200 = 80000$ 、中間層 $\rightarrow$ 音韻層間は $200 \times 460 = 92000$ の結合が存在する。書記素層に入力された単語が中間層を介して音素層へ出

¹一般には PINT を */pint/* と読み間違えるエラーの方が多い

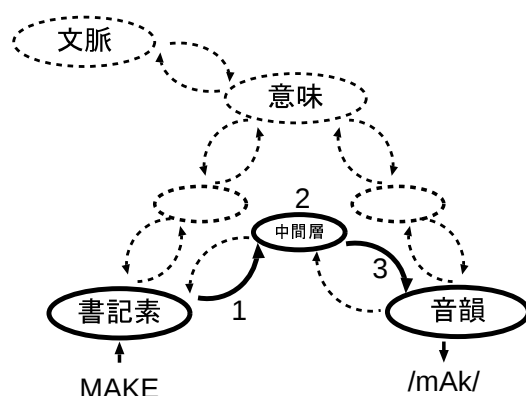


図 4: ザイデンバーグとマクレランドのトライアングルモデル。図中の実線の部分がパターンソンらのシミュレーションに関係する部分である。図中の数字は後述する損傷の場所を表している

力される。各素子の活性値は $[0, 1]$ 間の実数で、任意の単語は各素子の活性値の違いとして表現される。

3.2 入力表現

音素表現はルーメルハートとマクレランド (Rumelhart & McClelland, 1986) で開発された表現と同じで、連続する 3 音素が 1 素子として表現されている。この方法では、中央の文字の発音を表現するために、前後の文字も使用されおり、局所文脈情報表現 (あるいはトリプレット表現) と呼ばれることもある。トリプレット表現において MAKE という単語は [語の境界, M, A], [M, A, K], [A, K, E], [K, E, 語の境界] を表現する 4 つの素子の活性化として表現される。このような表現では素子数が (文字数 + 語の境界を表す記号) の 3 乗だけ必要になるが、彼らは表現の簡略化を行ない 16 素子にまで素子数を減らしている。たとえば [母音、摩擦音²、閉鎖音³] という表現は POST, SOFT などの単語で使われる表現である。このような 3 音素を 1 素子で表現することで文字位置情報を明示的に表現しないため入力層の素子数を文字数最長の単語に設定する必要が無い。

書記素層での情報表現も音韻表現と同様に、1 番目の文字に 10 文字、2、3 番目の文字に 10 文字づつからなるテーブルを乱数を用いて作成し、最終的にこの表現を 400 個の書記素層素子に変換して書記素層における情報表現としている。上記の連続する 3 文字から構成される局所文脈類似法により単語の類似性が表現されている。例えば MAKE は MADE や MATE に類似

²f, v, ɱ など

³p, b, t, d, k, g など

度が大きく、MILE や SMALL とは隔てて表現されることになる。

3.3 学習

結合係数、しきい値ともに $[-0.5, +0.5]$ で乱数によって初期化された。学習はバックプロパゲーション法によって行なわれ、学習に用いた単語は一音節語 2897 語で計 250 エポックで学習が成立するまで繰り返された。各語の頻度情報は対数変換して用いられた。すなわち最高頻度語の THE は 0.93、最低頻度語では 0.05 であった。これは 250 エポック中それぞれ 230 回と 12 回学習することを意味する。

実行手順は、最初に入力パターンが書記素層に提示され、その活動が中間層から音韻層へと伝播される。

パターンソンらは音韻誤差得点 (ターゲットと音韻層に現れた出力との差の 2 乗和) が単語呼称課題の潜時と正確さとに関係すると仮定して、心理実験データとの比較を行なっている。誤差得点が低いことは速くて正確な反応が生じることを意味する。すなわちノイズが少ないということが正確な反応を意味し、誤差が少ないということがより速く反応基準に達すると仮定している⁴。

モデルの学習中の振舞いで注目すべき点は、頻度効果と規則性効果との交互作用についてである。心理実験では、低頻度語に比べて高頻度語の方が視覚提示された単語を発話するまでの潜時が短いことが知られている (頻度効果)。規則性効果とは不規則語の方が反応潜時が長いことを指す。頻度効果と規則性効果との交互作用とは、成人の心理実験結果では、高頻度語の場合には規則性効果が生じないが、低頻度語では規則性効果が認められることを指す。モデルの訓練初期には、高・低頻度語とも大きな規則性効果が認められたが、250 回の学習終了後、高頻度語での規則性効果が消失した。このことは、心理実験の結果と定性的に一致すると見なすこともできる。

3.4 破壊実験

学習成立後の図 4 に数字で示した 3 つの部位を破壊した。損傷のシミュレーションでは該当部位の素子または結合を 0.1, 0.2, 0.4, 0.6 の割合で 0 にすることでなされた。図 5 に中間層の破壊 (図 5 左) と中間層から音韻層へ結合の離断 (図 5 右) の 2 種類の損傷の影響を示した。図 5 から以下のようなことが読み取れる。

- 中間層の破壊と中間層から音韻層への結合を離断した場合とでは同じ傾向にあること

⁴反応時間のニューラルネットワークモデルについてはマクレランドの GRAIN モデルがある (McClelland, 1993)

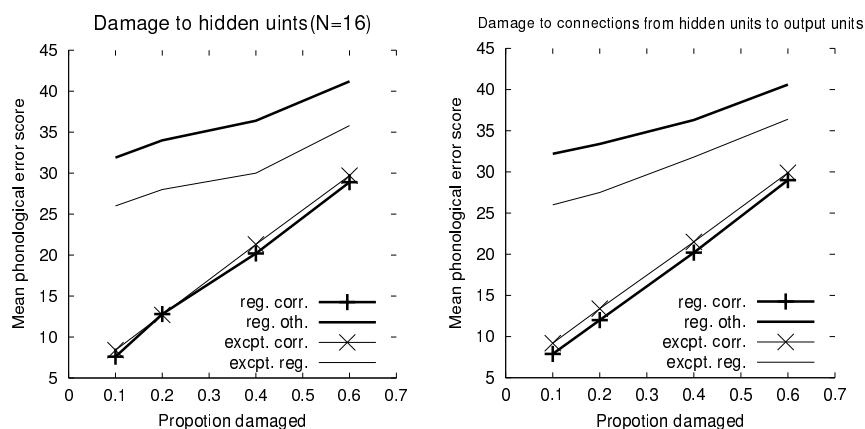


図 5: 破壊実験結果、右が中間層を破壊した場合、左は中間層から音韻層への結合を離断した場合。reg.corr.:規則語を正しく発音した場合、reg.oth.:規則語を正しく発音できなかった場合、except.corr.: 不規則語を正しく発音した場合、except.reg.: 不規則語を規則語の発音規則で発音した場合。例えば COVE を /kOv/ と発音すると reg.corr. に、/kUv/ と発音すれば reg.oth. に相当する。一方、MOVE を /mUv/ と発音すると except.corr. に、/mOv/ と発音すれば except.reg. に分類された。Patterson ら (1989) Table 7.3 を元で作成

- 音韻誤差得点は、破壊の程度関数として単調に増加すること
- 正しい発音の方が損傷の影響が大きいこと。すなわち損傷の程度が大きければ、正しい発音と誤りとの差が小さくなる。
- 規則語の正解と読み誤りとの差よりも、不規則語の正解と読み誤りとの差の方が小さいこと。

実験 2 では、訓練中に提示されなかった新奇語や音韻対応規則には従ってはいないが実際の英語には存在しない非単語を読ませることが行なわれた。モデルが新奇語を読むことができることは、モデルが一般的な発音規則を習得できたことを意味するが、非単語を読むことができることについては注意が必要である。図 1 の右の書記素音韻対応系が障害された場合、非単語の読みは一般に図 1 の左の経路を通過しなければならないと考えられてきたからである。そして彼らのモデルには図 1 の左の経路に対応するニューラルネットワークが実現されていない。このことは従来のた読みの誤りの分類に対して、加えて、背景となるボックス-アンド-アロー・モデルによる症例の検討に対して再考をせまるものであると言えよう。

彼女らはさらに、図 5 で示した結果が単語の頻度効果 (実験 3) による説明より、音韻的な類似性から説明できる (実験 4) 可能性を示唆している。

これらの結果と表層失読の患者の以下のような読みの特徴との対応が示唆される (Patterson et al., 1989, p.172, type II の症状)。

- 規則語の音読より不規則語の音読の方が成績が悪いが、規則語の音読も障害されている。
- 音読の潜時は異常に遅い。一単語を音読するのに苦労している。
- 非単語の音読が障害されている。
- 規則化 (不規則語を規則語のような音読規則に当てはめて発音する) だけでは誤りの大部分を説明できない。
- 意味性の誤りは認められない

モデルの成績と患者の症状との間の上記のような類似点から、ニューラルネットワークモデルが神経心理学に対して有効な道具になり得ると考える。

4 記憶障害あるいは意味記憶の構造について —ファラーとマクレランドの研究

意味記憶は、個々の対象についてカテゴリー毎に構成されているのかそれとも、それとも意味記憶はモダリティー毎に構成されているのか、についてファラーとマクレランド (Farah & McClelland, 1991) が行なった研究を紹介する。彼らのモデルによれば、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ことを意味し、認知心理学でしばしば話題になる記憶表象論争に対する一つの解答を与えている。

4.1 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)。ウォリントンとシャリスは、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。

ファラーとマクレランド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-非生物の知識が障害されることを説明できる。

4.2 モデル

彼女らのモデルを図6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、全ての

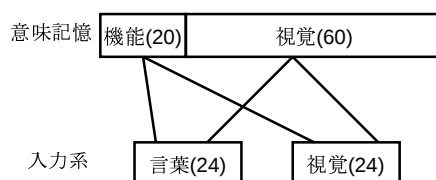


図6: ファラーとマクレランド (1991) の意味記憶モデルの概念図。カッコ内の数字は数値実験で用いられたユニット数を表す。意味記憶内で機能的記憶と視覚的記憶のユニット数が異なるのは、彼らの論文中的実験1 (心理実験)の結果を反映している。

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個ずつ、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うに訓練された。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絵画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。

4.3 破壊実験

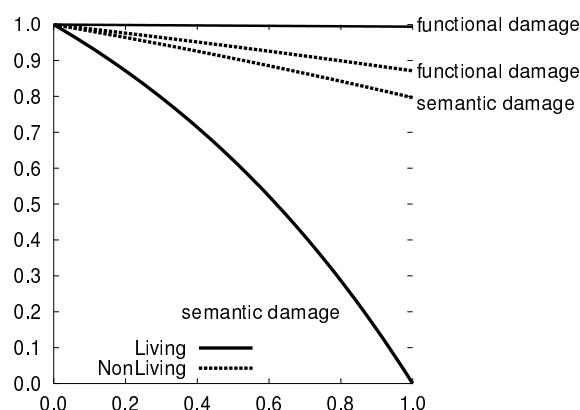


図 7: 生物-非生物別の意味記憶内の損傷の程度と課題成績との関係 (ファラーとマクレランド (1991) の表 3 と図 2 より改変)。彼らは各条件で 0, 0.2, 0.4, 0.6, 0.8, 0.99 の 6 点について各 5 回づつしか数値実験を行なっていないため実際の曲線は滑らかにならない。そこで、指数関数に回帰させてプロットしなおした。指数回帰を用いれば唯一のパラメータを変化させることで 4 つの条件に対応する曲線を描くことができる。

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、目を見たときには特定のカテゴリー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い。実験 4 では、2 節で説明した雑音を加える方法によるシミュレーションを行なっている。

結果から、モダリティ依存の意味記憶構造だけを用いて、カテゴリー依存の障害を説明できることが確認された。

5 回復過程のシミュレーション

ここでは回復過程をニューラルネットワークにおける再学習問題と捉えて議論する。なお本節では、高次認知機能に焦点をあてるため、入力系に異常はない、すなわちニューラルネットワーク回路に入力される情報は、損傷前後で変化がないものと仮定する。

一般に梗塞などによって一時的に認知機能に障害を呈しても、リハビリテーションによって回復する場合と回復不可能な場合とがある。回復不能な場合には、次のようなことが予想されよう。

1. 損傷によって損なわれた認知機能を補完するだけの学習能力がない。
2. 加齢の影響で学習速度が遅い。
3. 回復不能な程度に該当部位が損傷を受けている
4. ある種の病変によって神経細胞が壊死するために受ける能力の低下と再学習による効果とが拮抗する状態になってそれ以上学習が進まない。

ここでは (1) の場合に絞って考察することにする。(2) の場合が想定されるのは、長年に渡る追跡調査が事実上不可能なので入院時だけでは変化が観察されないという可能性を考慮したものである。(3) の場合は片麻痺などの運動機能の障害の場合に観察されることであるが、高次認知機能の場合には、完全に機能しない場合だけではなく、部分的には答えることができたり、正解できる場合と出来ない場合があることがある。(4) のような進行性の病変については今回取り上げない。実際の臨床医の判断では特定の症例では、発症後 1 ヶ月程度で予後が予想できるそうである。このことは (2) の可能性も否定できないが、加えて、損傷によって脳内の表象が変化し、問題を再学習するときに困難な場合があることを予想させる。

5.1 モデル

(1) の認知機能を補完するだけの学習能力がない場合、あらかじめ構造化された形の脳内表現の一部が利用不能になった状態であると考えられる。発症後の再学習過程では利用できるユニットが少なくなることで、発症前と異なる内部表現を獲得する必要がある。本来ならシナプス結合を維持するために使われていた栄養因子を残存するユニットが利用することによって再学習がおこると考える。

5.2 理論的限界について

簡単のため出力ユニットが 1 個しかない場合のパーセプトロンを考える。すなわち入力パターン x を 2 分割する問題を考える。 $x \in G^+$ であれば正、 $x \in G^-$ であれば負を出力するように学習が成立しているものと仮定する。学習すべき問題が n 次元空間で線形分離可能であっても、損傷によって m 個のニューロンが欠落することで、 $n - m$ 次元空間で線形分離可能か否かに帰着する。このときパーセプトロンの循環定理によって患者は誤った学習をする。すなわち発症後の回復過程では、理論的な上限が存在し、患者が到達できる正解率の上限は $1 - \frac{1}{2} P(G^+ \cap G^-)$ で与えられる。

5.3 制約つき誤差逆伝播学習法

従来の誤差逆伝播法での学習方程式は、最急降下法

$$\frac{dw_{ij}}{dt} = -\frac{\partial E}{\partial w_{ij}} \quad (1)$$

のように表されてきた。この学習方程式は、乳児が言語を獲得するときの様子に例えることができ、損傷からの回復をモデル化する場合には不向きである。そこで損傷後には、脳内での特定の栄養因子が働かず新たにシナプスを形成するのにコストがかかることを考慮して、損傷後の再学習時には $\sum |w_{ij}| = \text{一定}$ のような拘束条件を設けた学習方程式を提案する。具体的には再学習時は、 w の更新時に

$$w_{ij} \rightarrow \frac{w_{ij}}{\sum w_{ij}} \quad (2)$$

のような規格化を行なう。上式は、Malsberg のモデルを階層的なネットワークに適用したと見ることができる。

5.4 数値例

Rumelhart ら (Rumelhart, Hinton & Williams, 1986) のパリティ問題を例にとって数値実験を行なった結果を図 8 に示す。パリティ問題とは ON-OFF の 2 値をとる入力に対して ON の数が偶数ならば 0、奇数ならば 1 を出力する問題である。例えば ON を 1、OFF を 0 と表現すれば $[0000] \rightarrow 0$ 、 $[1011] \rightarrow 1$ のような出力を与える規則を学習することを意味する。入力層の

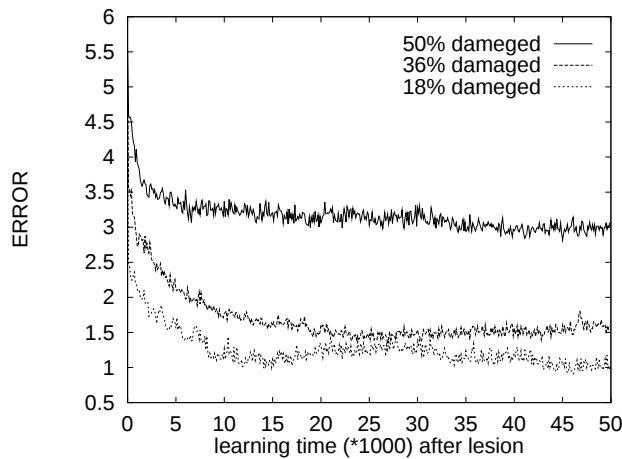


図 8: 制約つき誤差逆伝播法による学習曲線

素子数 4、中間層 8、出力層 1 のニューラルネットワークに予め誤差逆伝播法を用いてパリティ問題を解くように訓練した。図 8 は一旦学習が成立した後、

任意の中間層を削除し、上記の誤差更新式に従って学習させたときの誤差の減少の様子を学習回数の関数としてプロットしたものである。損傷の程度が大きいほど誤差は減少せず、しかも一定量以上誤差が減少しないことを示している。この傾向は損傷を受けた中間層の数が多ければ多い程顕著である。

高次認知機能の損傷から回復過程のモデルとして、内部状態空間の縮退による限界を考察する他に、シナプス結合強度の和が損傷前とそれと比べて変化しないという制約の元にネットワークを訓練する方法を紹介した。この種の研究は極端に数が少なく、実際のリハビリテーション過程との対応は今後の課題である。

6 まとめ

ニューラルネットワークモデルを神経心理学の分野に応用する場合について私見を述べておく。

6.1 機能的脳画像研究との関連

以前の神経心理学の進歩は遅かったと聞いている。患者の死後の剖検結果でしか特定の脳機能の責任領域を特定できなかったからだ。これが脳の可視化技術の進歩によって状況は激変した。機能的核磁気共鳴法 (fMRI)、陽電子造映法 (PET)、近赤外線分光法 (NIRS) などを用いて特定の認知課題を遂行している間の脳の賦活状態を見ることができるようになってきた。これらの研究で明らかになったことは、単純な課題遂行中でさえ、従来考えられてきた担当領域以外に極めて多くの部位が活性化することと、従来の担当領域であると考えられて来た部位が必ずしも活性化せずむしろ活動を低下させる可能性も考えられることである。活動低下は一見矛盾した結果のように捉えられがちである。機能的脳画像の研究者たちは、それぞれの手法が持つ独自の技術的問題点や、賦活画像どうしの引き算を行なう際に用いられる統制課題との関連などを考慮しなければならないと考えている。しかし、ニューラルネットワークの視点で考えれば、不思議なことではない。その領域への入力が抑制性の結合であると考えればよいだけである。このようにニューラルネットワークの視点にたてば、活性低下の問題は矛盾無く説明できる。

6.2 神経心理学者との共同研究の奨め

現在分かっている事実を可能な限り採り入れるのは当然のこととして、まだ明らかでない点、疑わしい点には大胆な仮説を設けてモデルを作ることが重要である。検証は計算機が行ってくれる。計算機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ることができ、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、という良好な関係がこの分野の研究をさらに一歩前に進める力になるだろう。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片手落ちである。ニューラルネットワーク研究と神経心理学研究とは共同研究が望まれている分野であり、筆者も神経心理学者との共同研究を呼びかけたい。

6.3 紹介したモデルのその後

はじめに述べたとおり、ここで紹介した研究はこの分野の古典に属する。今日的視点から見れば、ここで紹介する研究にはそれぞれ問題点が指摘できるが、ここでは敢えて言及しなかった。主な理由は、この章の目的が研究の紹介であって批判ではないからである。ここではオリジナルの研究を紹介し意義を感じとっていただければと思う。

それぞれの著者による修正モデルも提案されている (Plaut, McClelland, Seidenberg & Patterson, 1996; Harm & Seidenberg, 1999; Seidenberg, Petersen, Plaut & MacDonald, 1996; Seidenberg & McClelland, 1989; Seidenberg, Plaut, Petersen, McClelland & McRae, 1994)。特に最近の研究では、もう一つの読みのモデルであるアトラクタネット (Hinton & Shallice, 1991; Plaut & Shallice, 1993) の概念を採り入れている。アトラクタネットでは、失読症患者が音読に異常に時間がかかる問題—逐次読み (letter-by-letter reading) の現象を説明できると考えられ、興味深いニューラルネットワークモデルである。対応する日本語の読みのモデルとして伊集院ら (2000) を挙げておく。

文献

山鳥重 (1985). 神経心理学入門. 医学書院.

Coltheart, M. (1985). Cognitive neuropsychology and the study of reading. In M. I. Posner & O. Marin (Eds.), *Attention and Performance XI* chapter 1, (pp. 3–37). Hillsdale, NJ: Erlbaum.

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-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.

- Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (2001). Drc: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256.
- 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.
- Harm, M. W. & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist model. *Psychological Review*, 106(3), 491–528.
- Hinton, G. E. & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98(1), 74–95.
- Ijuin, M., Fushimi, T., Patterson, K., Sakuma, N., Tanaka, M., Tatsumi, I., Kondo, T. & Amano, S. (2000). A connectionist approach to naming disorders of Japanese in dyslexic patients. In *Proceedings of the 6th International Conference of Spoken Language Processing* (pp. 32–37). Beijing.
- McClelland, J. L. (1993). Toward a theory of information processing in graded, random, interactive network. In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience* (pp. 655–688). MIT Press, MA.
- 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., McClelland, J. L., Seidenberg, M. S. & Patterson, K. (1996). 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., 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.

Rumelhart, D. E. & McClelland, J. L. (1986). On learning the past tenses of english verbs. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Volume 2 chapter 18, (pp. 216–271). Cambridge, MA: MIT Press. (この章は邦訳書では省略されている).

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.

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