

脳の科学 第 10 回

担当：浅川伸一

brain_science@cis.twcu.ac.jp

師走 05 日, 2008 年

ニューラルネットワーク

人間の感情と、他の生物のそれと、近代的な型の自動機械の反応との間に鋭い乗り越えられない区画線を引く心理学者は、私が私自身の主張に慎重でなければならぬのと同様に、私の説を否定するのに慎重でなければならぬ

— N. Wiener, The Human Use of Human Beings(人間機械論, みすず書房, p.73) —

1 脳の構成論的研究

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や、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や、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。とりわけ、機能的脳画像研究の進歩が著しい。時間解像度、空間解像度とも精度が飛躍的に向上している機能的脳画像研究は、従来の心理学的研究手法を根本から変革する力すら持っているように思われる。

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。情報論的必然性という言葉があって、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筋道はそういくつもあるはずはない、というものである [29]。入出力関係が脳とモデルとの間で一致したときメカニズムも一致している可能性は、この情報論的必然性に導かれてかなり高いのかも知れない。

脳の構成論的研究とは、このような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、生理学的に分かっていることはできるだけモデルに取り入れ、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ということである。

2 計算可能性

脳とコンピュータ¹はの計算可能性の観点からは等価であるとみなすことができる。計算可能性の理論の中で、Turing machine の停止問題は基本的である。Turing machine の停止問題は決定不能問題である。ほとんどの決定不能問題を証明する場合、“もしその問題を解くアルゴリズムが存在すれば Turing machine の停止問題を解くアルゴリズムも存在する”ということを直接的あるいは間接的に示すという手段が用いられる。Turing machine がすべての入力に対して停止するか、2 つの Turing machine (脳とそのシミュレーションである neural network) が同じ働きをするかなど、Turing machine に関するほとんどの決定問題は決定不能である。ちなみに、Neural network で解こうとしている問題はすべて計算可能である。なぜならすでに別の implementation — 脳 — が実在するから。この意味では数学者より分のよい勝負をしていると言える。

3 シミュレーションについて

アシモフの SF の中に次のような言葉が書かれている。

「もし、宇宙のある側面を理解したいと思ったら、理解するのに絶対必要な特性とか性質だけを残して、できるだけ単純化して考えるといい。物体がどのように落下するか究明したければ、それが新しいか古いか、赤か緑か、匂いがあるかないか、などを気にしてはならない。それらを除去するんだ。不必要に物事を複雑化してはいけない。この単純化をモデルとかシミュレーションとか呼ぶことができる。—中略— このような単純化したシミュレーションは、現象そのものを研究するよりも、もっとずっと容易にその現象を把握させてくれる。—中略— 何らかの現象について、もっともっと多くを知ろうと思うにつれて、あるいは、現象がますます複雑になるにつれて、ますます多くの巧妙な方程式が必要になり、ますます細かいプログラミングが必要になり、結局、把握するのがますます困難になるコンピュータ化したシミュレーションが必要になる。」

「シミュレーションのシミュレーションはできないのかね？もう一段階程度を落して」

「その場合には、包含したい現象の一部を省略しなければならない。そうすると、そのシミュレーションは役に立たなくなる。Least Positive Simulation は対象が複雑になる速度よりも、もっと急速に複雑になる。そして、究極的にはシミュレーションがその現象に追いついてしまう。このようにして複雑性を完全に備えた、全体としての宇宙は、それ自体よりも小さなシミュレーションによって表現されないということが、何千年も前に確認されている。」

— アイザック・アシモフ「ファウンデーションへの序曲」早川書房より

4 ニューラルネットワークの定義

ニューラルネットワークモデルとは、脳の振る舞いを模倣するための表現のことである。必ずしもプログラムによって表現されている必要は無いが、数式を使って表現される場合が多い。神経回路網 (neural network) モデル、PDP(parallel distributed processing) モデル、あるいはコネクショニスト (connectionist) モデルと呼ばれることもある。論文などの

¹実時間での処理を考えなければ相互に変換可能であるという意味では

タイトルに上記のいずれかの言葉を入れるとき、心理学者は PDP モデルやコネクショニストモデルを使う傾向がある。

ここでは、ニューラルネットワークモデルを、「生体の中枢系で行なわれている情報処理の機能、性能、および特性を記述するための抽象化された表現」と定義しておく。また、ニューラルネットワークとは「脳の基本素子であるニューロン（神経細胞）とニューロンが結合したネットワーク（神経回路網）の構造や情報処理のメカニズムにヒントを得て脳の持つ情報処理能力のモデル化を目指す研究分野のこと」と言っても良いだろう。

ニューロンの動作を抽象化した場合、ニューロンと呼ばずにユニット、素子、あるいはプロセッシングエレメントなどと呼ぶことがある。この場合は一つ一つのニューロンを抽象的に表現したと言うよりニューロン集団の振る舞いを記述していると考えられる。認知心理学でニューラルネットワークモデルという場合には脳の部位を一つのユニットとして扱うことが多い。

5 並列処理と直列処理

言語、意識、思考などと言った高次認知機能は直列的である。一方、視覚や聴覚といった知覚情報処理は並列計算が前提となっている。脳内では多数のニューロンが同時的、並列的に活動している。このことから意識的な活動は無意識的な並列処理が組み合わされた結果として生じるのではないかという仮説が成り立つだろう。

6 何ができるか

今まで提案されてきたモデルの一部を列挙すると次のようになる。

- 単語の読み acquired dyslexia [22, 23, 26, 25]
- 書字 spelling [4]
- 発話 speech production [15, 10]
- 相貌認知 face recognition [5, 12, 13]
- 視覚的物体呼称 visual object naming [14, 26]
- 空間的注意 spatial attention [8, 18, 13]
- 学習と記憶 learning and memory [1, 2, 6, 20, 19]
- 意味記憶 semantic memory [11, 17, 13]
- 行動制御 control of action and responding [3, 9, 7]

7 簡単な歴史

1943 年 McCulloch and Pitts の形式ニューロン

1949 年 Hebb の学習則

1962 年 Rosenblatt パーセプトロン

1969 年 Minsky & Papert パーセプトロン, 中野 アソシアトロン

1972 年 Kohonen 連想記憶モデル, Anderson 連想記憶モデル

1970 年代 甘利らによる数理解析, Marr と Albus による小脳パーセプトロン説

1975 年 福島 コグニトロン

1980 年 福島 ネオコグニトロン

1982 年 Hopfield

1983 年 Farmann & Hinton ボルツマンマシン

1985 年 Hopfield & Tank 巡回セールスマン問題

1986 年 Rumelhart & Hinton バックプロパゲーション, Sejnowski & Rosenberg NETTalk

マッカロックとピッツ McCulloch and Pitts の形式ニューロン formal neuron は第二次世界対戦中であり、ヘップ Hebb の学習則は GHQ による日本占領中 (1952 年まで) だから意外に古いことが分かる。ローゼンブラッド Rosenblatt によるパーセプトロン perceptron の提案から、第一次ニューロブームと呼ばれる時代に入る。このブームはミンスキーとパート Minsky and Papert による批判で一気に下火になる。第二次ニューロブームと呼ばれるのはホップフィールドネットとラメルハートらによるバックプロパゲーション法の再発見によって花開いた形である。この他にも要な研究が抜けているが、歴史の概略を説明する目的なので省略してある。重要なのは第一次ニューロブームと第二次ニューロブームの間に日本人研究者による優れた研究があったことであろう。

8 ニューラルネットワークの分類

大まかに分けるとニューラルネットワークモデルには以下のような分類が存在する。

8.1 学習方式による分類

学習方式による分類には次の 2 つがある。

1. 教師あり、パーセプトロンなど

2. 教師なし、自己組織化、特徴マップ

外部からネットワークに対して望ましい出力 (教師信号) を与えて、ネットワークに同じ出力を返すように学習させることを教師あり学習という。一方、明示的な教師信号を用いない学習を教師なし学習と呼ぶ。この場合ネットワークは入力信号の統計的性質を学習することになる。

8.2 結合方式による分類

神経素子間の結合方式による分類には次の 2 つがある。

1. 階層型
2. 相互結合型

与えられた入力信号が特定の方向にしか伝播しないような回路を階層型 (layered) の回路という。一方、信号が回路内を循環したり逆方向に伝搬したりする方の回路を相互結合型という。順方向への信号の伝播をフィードフォワード、逆方向へのそれをフィードバック (または帰還) と呼んで区別するとがある。さらに、連想記憶と呼ばれる回路については、相互想起型、自己想起型の区別などがある。

9 ニューラルネットワークの特徴

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。この意味において、ニューラルネットワークは「強い AI」[27] を主張している。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、分散表現と統計的学習の漸進的学習、および相互作用の 3 点に要約できる [13, 24]。

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。(ポピュレーションコーディング第 7 回ハンドアウト参照のこと)

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。

学習にはシステムの望ましい出力を得るために外部の教師信号を利用するパーセプトロン型の教師あり学習と、明確な外部信号を仮定せず、ユニット間の結合を外界の環境に合わせて変化させるヘップ型の学習則に従う教師なし学習との 2 種類がある。非常に多くのニューラルネットワークモデルが今までに提案されてきているが極論すればこの 2 種類の学習則しか無いといってよい。

相互作用: ユニット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。

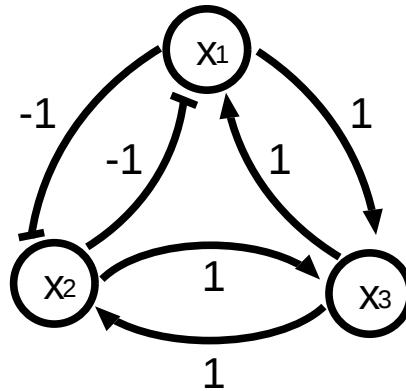


図 1: 3 つのユニットからなるネットワーク 8 類似の状態が定義できる

たとえば図 1 は 3 つのユニットからなる簡単なニューラルネットワークの例である。各ユニットが 1 か 0 かの 2 状態を取るとすれば、このネットワークには $2^3 = 8$ 種類の状態が定義できる。図中に結合強度が +1 か -1 かで書かれているが、次の時刻におけるユニットの活性値は直前の他の二つのユニットの状態によって変化することになる。

10 生理学的事実の抽象化

ここでは、ものごとを極端に単純化して捉えることから始める。この単純化に不満を覚える人もあるだろう。しかし、真実は単純さと複雑さの狭間にある。現在あらゆる理論は現実の人間のモデルとしては単純すぎる。逆に言えば人間の行為・活動は非常に複雑である。この複雑な現象は、そのままではとらえることができないため、ある程度の単純化が必要になる。この単純化のことをモデルと言う。モデルが妥当性を持つためには現実には則して、扱っている現象を説明できる必要がある。しかし、脳の科学では現象が複雑すぎるため、現在どのようなモデルを用いた説明でも単純すぎるのである。それならば、モデルを作る場合には本質を損なわない限り極力簡単なモデルを用いる方がよい。定性的に同じ動作をする 2 つのモデルがあったとすれば、より簡単なモデルを選ぶべきである。

これをオッカムのカミソリ Occam's razor という。「同じような理論で、難しいのと簡単なのがあった場合、難しい方が有効である証拠が見つかるまで、簡単なほうを使うべきである。」あるいは、「不必要に物事を増やしてはならない。」などといわれる。

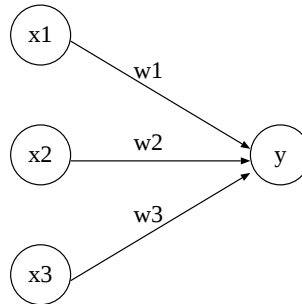


図 2: ニューロン y は x_1, x_2, x_3 からの入力を受けて、入力ニューロンからの活性値を結合係数 w によって重み付けられた値の関数として自身の出力値が定まる

10.1 表記

ニューロンはシナプスを介して多数のニューロンから信号を受け、興奮するかもしれないかが決まる。この意味で多入力 1 出力の情報処理装置と見なすことができる。 n 個のニューロンから信号のを受け取っているニューロンで、各入力信号の強さ $x_1, x_2, \dots, x_n$ をとする。このニューロンの膜電位の変化量を u 、出力信号を z とする。 i 番目の軸索に任意の単位の強さの信号が与えられたときに、この影響を受けて変化する膜電位の量を w_i と表す。 w_i はシナプスの伝達効率を表す量でシナプス荷重 (または 結合荷重、結合強度) という。抑制性のシナプス結合については $w_i < 0$ 、興奮性の結合については $w_i > 0$ となる。

11 基本回路の設計

本節では、基本的な論理回路と簡単な記憶回路を神経回路網で構成する方法を考える。シリコンウェハ上に構成される論理回路をニューロンユニットでも実現できることを示す。

11.1 AND 回路

2 入力 1 出力の回路において、2 つの入力が共に真であるときのみ真を出力し、そうでなければ偽となる論理演算である論理積 (AND) を考える。論理積は引数を 2 つとる演算であり、出力を y とすれば $y = f(x_1, x_2)$ のように書くことができる。 x_1, x_2 とともに 1 または 0 の値をとるものとすれば、 y が 1 であるためには $x_1 = 1$ かつ $x_2 = 1$ でなければならぬ (図 3)。

この素子がマッカロック・ピッツの形式ニューロンで表せるとすれば、2 つの入力の総和が閾値を越えれば発火し、そうでなければ発火しない素子を考えていることと等価である。

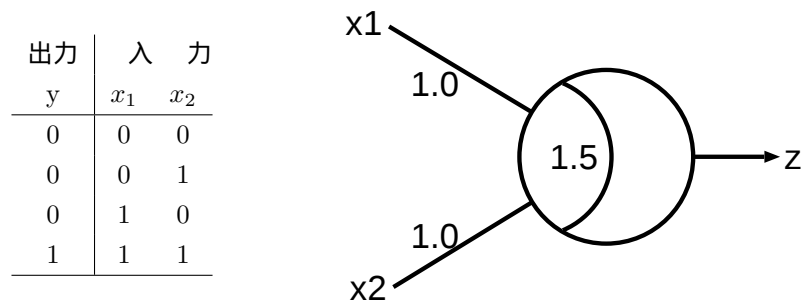


図 3: 論理積 AND の真偽表 (左) と等価な演算を実行可能な形式ニューロン (右)。右図では x_1, x_2 の 2 つの入力が共に 1 でなければ閾値 1.5 を越えることがない。従ってこのニューロンは 2 つの入力が無ければ発火できない。

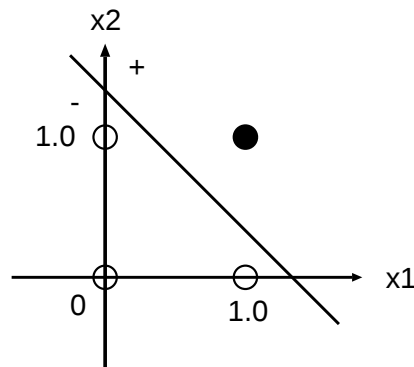


図 4: 図 3 と等価な幾何学的表現 $y \geq x_1 + x_2 - 1.5$ 。2 次元平面を 2 領域に分割し、一方を発火する領域 (+)、他方をしない領域 (-) に分割することに等しい。このような分割の仕方は無限に存在することに注意。

11.2 NOT 回路

同様に図 5 は否定 (NOT) 演算である。入力があれば ($x = 1$ のとき) かならず負の結合荷重 (-1.0) によって閾値 (0.5) を下回るので、このニューロンは発火できなくなる。

11.3 簡単な記憶回路 — フリップフラップ回路 —

上記の AND 素子と NOT 素子とを繋いで簡単な記憶回路を作ることができる。図 6 で各素子は 1 か 0 かを値として取りうる形式ニューロンだとする。今、入力 x と入力 y とが共に 1 であれば $A = 1, B = 0$ あるいは $A = 0, B = 1$ のときだけこの回路は安定である。ここで $x = 0, y = 1$ とすると $A = 0, B = 1$ の状態になり、 $x = 1, y = 0$ とすると $A = 1, B = 0$ の状態になる。しかも、この状態は $x = y = 1$ に入力を戻しても保存され

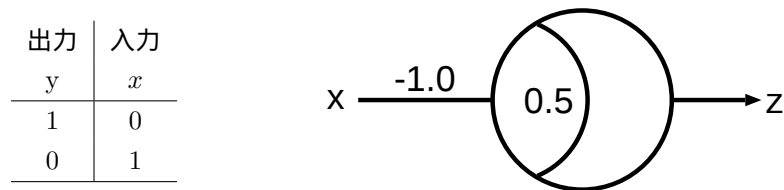


図 5: 否定 NOT 演算と等価なニューロン

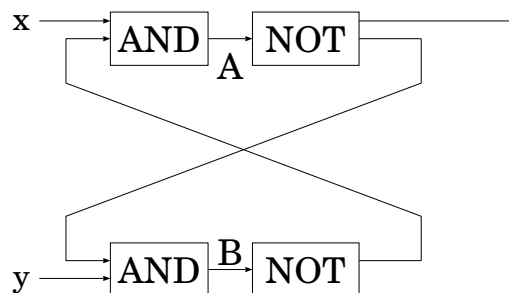


図 6: フリップフラップ回路の一例

ている。これは 1 ビットの記憶回路でありフリップフラップ回路 (flip-flop circuit) と呼ばれている。このことは AND と NOT を実現できる神経回路素子があれば記憶回路を作ることができることを示している。しかも工学的に実現されている回路と完全に等価である。フリップフロップ回路を 16 個または 32 個まとめてレジスタ (register) と呼ぶ。市販されている PC の CPU の性能を指して 16 ビットマシンとか 32 ビットマシンとか呼ぶのは、このレジスタの大きさ (記憶装置への基本的な入出力単位の基本でもある) による。一般にコンピュータの速度はこのフリップフラップ回路が安定するまでの時間に大きく依存する。なぜなら、コンピュータの基本動作は原理的に、上述のフリップフラップ回路が安定するのを待って、次の命令をレジスタに読み込むことの繰り返しだからである。

12 学習について

ニューラルネットワークモデルでは、シナプスの結合係数を変更することをニューラルネットワークの学習と呼ぶ。学習時における結合係数の変化を記述したルールを学習則 (learning rule) という。ここでは ヘップ Hebb の学習則とバックプロパゲーション法を紹介する。それから厳密にはニューラルネットワークの学習則ではないが、ニューラルネットワークとの関連で言及されたり、ニューラルネットワークとの組み合わせで用いられる、強化学習と遺伝的アルゴリズムについて紹介する。

12.1 ヘップ則

Hebb の原文では以下のような記述である。

When an axon of cell A is near enough to excite cell B and repeatedly or consistently takes part in firing it, some growth process or metabolic changes takes place in one or both cells such that A's efficiency, as one of the cell B, is increased.

—Hebb,D.O.,(1949), The Organization of Behavior, p.62 —

「同時に発火したニューロン間のシナプス結合は強められる」ことを主張しているのがヘップ則 (Hebbian rule) である [16]。ヘップの学習則とも表記されるヘップ則は以下のように定式化できる。ニューロンの発火状態を 1、休止状態を 0 と表現することにし、ニューロン y から ニューロン x へのシナプス結合係数を w とする。このときヘップの学習則は、シナプス結合係数の変化 Δw として表現され、 $\Delta w = \lambda xy$ と書くことができる。ここで

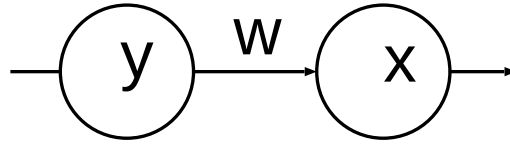


図 7: ヘップの学習則

$\lambda(\geq 0)$ を学習定数という。 x と y は 1 と 0 の 2 とおりの状態にしかならないため、可能な組み合わせは 4 通りになる。このうち $\Delta w \neq 0$ となる組み合わせは、 $(x = 1)$ かつ

		x_j	
		1	0
y_i	1	λ	0
	0	0	0

表 1: ヘップの学習則

($y = 1$) の場合だけである。

すなわち結合係数は、入力ニューロンと出力ニューロンの同時発火頻度に比例して、入力ニューロン x の λ 倍だけ増大することを意味する。ある時刻 t で、入力 x が y を発火させたとき、次回 $t + 1$ に同じ入力を与えられるとより強く発火させる効果を持つようになると考えることもできる。逆に言えば同じ刺激を入力し続けると $|w|$ が限りなく大きな値になってしまうことも意味している。このようにヘップ則には、学習回数を重ねても一定の値に収束しないという特徴を持っている。

ヘッブ則の変形はさまざまに考えられていて、例えば同時に発火しなかった時にシナプス結合を弱めるアンチヘッブ則、減衰項を加える方法、入力と出力の差 (の 2 乗) を利用する方法、などが考案されている。

12.2 バックプロパゲーション

12.2.1 XOR 問題、線形分離不可能な問題

パーセプトロンでは絶対に解けない問題に排他的論理和 (XOR) 問題がある。排他的論理和とは、2 つの入力のうちいずれか一方のみが 1 のとき 1 を出力する問題である。図 8 左を見るとわかるとおり、XOR 問題は一本の判別直線で白マルと黒マルを分離できない、すなわち線形分離不可能な問題である。

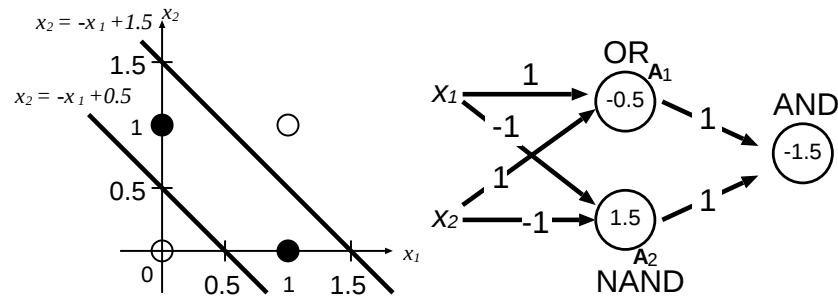


図 8: XOR 問題の幾何学的表現と XOR 問題を解くためのネットワーク

図 8 右は図 8 左の幾何学表現を対応するネットワークとして表現したものである。一番左が入力層、一番右が出力層、中間の 2 つが中間層である。ユニットの中に書かれた数値は各ユニットのしきい値を示している。中間層の 2 つのユニットのうち上の OR と書かれたユニットは、 $x_1 + x_2 - 0.5 > 0$ のとき発火する。この式を書き換えると、 $x_2 > -x_1 + 0.5$ となるので図 8 左の下斜線より上の領域に対応する。一方、中間層の下の方の NAND(not and) と書かれたユニットは、 $-x_1 - x_2 + 1.5 > 0$ のとき発火するので、移項して $x_2 < -x_1 + 1.5$ とすれば、図 8 左の上斜線より下の領域に対応していることが分かる。さらに、AND と書かれた出力ユニットは、2 つの中間層ユニットからの情報の論理積 (AND) を計算するユニットになっている。そこで、2 つの中間層ユニットの両方が発火する場合だけ、出力ユニットも発火し、1 を出力する。これは、図 8 左では、「下の斜線より上の領域」でかつ「上の斜線より下の領域」に対応する。すなわち、図中の黒丸の領域だけが分離されることになる。このような 2 本の直線は図中にいくらかでも引けることから、XOR 問題の解も無数に存在することが分かる。

図 8 左にあるとおり、中間層のユニット 1 個は 1 つの線形判別関数に相当すると考えられる。中間層から出力層への結合では各出力の論理積 AND を計算していることに相当する。 n 個の中間層を用意すれば原理的には $\frac{n^2+n+2}{2}$ 個のカテゴリ分類が可能である。パー

入力層		中間層		出力
x_1	x_2	a_1	a_2	r
0	0	0	1	0
0	1	1	1	1
1	0	1	1	1
1	1	1	0	0

表 2: 図 8 に対応する XOR 問題の真偽表

セプトロンが XOR 問題を解くことができない理由は入力層から中間層にいたる結合係数を変更する手段がないことなのである。

12.2.2 誤差逆伝播法 (一般化デルタルール)

XOR 問題でも見たように、パーセプトロンの問題点は学習が出力層と中間層の間だけで行われ、入力層と中間層の結合係数を更新することができない。換言すれば中間層ユニットの誤差を明示的に知る方法はない。このことは信用割り当て credit assignment 問題と呼ばれる。この信用割り当て問題を解決したのがバックプロパゲーション (誤差逆伝播法あるいは一般化デルタルール) である。

12.3 強化学習

ニューラルネットワーク研究あるいは機械学習の分野で長いこと手が付けられずに残っていた重要な考え方がある。それは「生体が何かを欲求し、環境から来るある特定の信号を最大化させるように、自らの行動を適合させる」という、いわば当たり前の考え方、しかし生物の進化や適応にとって重要なメカニズムである。

この考え方は近年「強化学習 reinforcement learning」という名で注目を集めるようになって来ている [28]。心理学の世界では「道具的条件づけ instrumental conditioning」と呼ばれる学習理論の一般化、あるいはニューラルネットワーク的実装と言える側面も持っているが、強化学習の枠組みは、行動主義心理学者たちの考える道具的条件付けよりも広くて一般的な意味で用いられる。

例えば、幼児は感覚系 (眼や耳や皮膚) と運動系 (声や手足) との連係を用いて環境 (母親や自分の身の回りのもの) に直接的に働きかける。この連係を用いることで原因と結果の推論や、目標を達成するために何をすべきかについて多くの情報を取りだすことができる。このような環境との相互作用が我々自身に関する主要な知識源であると言える。

12.3.1 強化学習の特徴

強化学習の特徴を挙げると、試行錯誤的な探索 trail-and-error search と遅延報酬 delayed reward の 2 点になる。行動は直接的な報酬のみならず、その次の状況に影響を与え、その

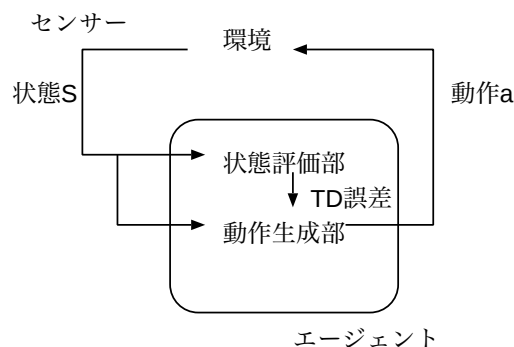
ことを通じて、その後に続く全ての報酬に影響を与える。ニューラルネットワークにおける教師あり学習は外界から与えられる教師信号によって自らの行動を適応させていく、例からの学習である。例からの学習も重要な要素の1つであるが、相互作用を介した学習では未知の問題領域で学習者自らが経験から学ぶ必要がある。

教師あり学習では、与えられた入力に対して最初はランダムな結合係数によって答えを出し、その答えを教師信号の示す方向に変化させていく。この意味では結合係数によって定義される空間の探索を行なっていると見なすことができる。一方、強化学習では自身の取りうる行動のレパートリーの中から最適な行動を探索していると見なすことができる。

強化学習には、他の手法と異なる抽象的な概念 — ポリシー、プランニング、価値関数、報酬関数、環境のモデルなど — を直接取り扱う。

ポリシーはある時点での学習者の振舞い方を定義する確率として扱われる。報酬関数は目標を定義する。強化学習者 (エージェント) の目的は最終的に受け取る総報酬を最大化することである。価値関数は最終的に何がよいのかを指定する。ある状態の価値とは、エージェントがその状態を起点として将来にわたって蓄積することを期待する報酬の総量である。人間にたとえば報酬は喜びや苦痛のようなものであるが、価値は我々の環境が特定の状態にあるとき、どれだけ満足あるいは不満であるかに関して、もっと洗練された長期的観点からの判断に相当する。すなわちエージェントはもっとも高い報酬ではなくもっとも高い価値 (しばしば総報酬量の関数として定義される) を持つ状態につながるような行動を見つけ出そうとするわけである。

強化学習が他のタイプの学習と最も異なる特徴は正しい行動を直接与えて教示するのではなく、実行した行動の評価を訓練情報として利用することである。従って、よい行動を直接探索するために試行錯誤による能動的な探索が必要になる。行なった行動がどれくらい良いのかが知らされ、それが可能な行動の中で最良または最悪であるかについては知らされない。



それぞれの行動に対して、その行動が選ばれた場合の報酬の期待値が定まっていて、この値を価値と呼ぶ。強化学習では価値が確実に知られているわけではないと仮定する。その場合でも価値の推定値を持つことができる。行動の価値の推定値を常に持っていれば、どの時点でも、価値の推定値を最大とするような行動が少なくとも1つ見つかることになる。価値の推定値を最大とするような行動を選択することをグリーディ greedy な行動と

呼ぶ。グリーディでない行動は探索を行なっていると言う。なぜならグリーディでない行動を選択すれば、その価値の推定値を改良できる可能性があるからである。探索はより大きい総報酬を最終的に作り出す可能性を持っている。たまに小さい確率 ϵ でグリーディな行動選択とは無関係に様に任意の行動を選ぶような方法を ϵ グリーディ法と言う。 ϵ グリーディ法ではすべての行動 a に対して行動 a の価値の推定量 $Q_t(a)$ が真の推定量 $Q^*(a)$ に収束することが保証されている。

12.3.2 目標と報酬

強化学習ではエージェントの目的あるいは目標は、環境からエージェントに送られる特殊な信号として形式化することができる。各時間ステップにおいて、報酬は単純に数値 $r_t \in \mathcal{R}$ である。非形式的には、エージェントの目標は自分が受け取る報酬の総量を最大化することである。これは、直接的な報酬を最大化することではなく、最終的な累積報酬を最大化することを意味している。

目標に関する考え方を形式化するために報酬信号を用いることは、強化学習の大きな特徴の1つである。例えばロボットに迷路から抜け出すことを学習させる際には、脱出して報酬が1になるときまでは報酬を0にすることがよく行なわれる。迷路の学習においてよく行なわれることは、脱出する前のあらゆる時間ステップで-1の報酬を与えることである。これによってロボットは出来る限り迅速に迷路を脱出するように仕向けられる。

たとえば有名なロボット三原則において

ロボット工学の三原則

第一条 ロボットは人間に危害を加えてはならない。また、その危険を看過することによって、人間に危害をおよぼしてはならない。

第二条 ロボットは人間にあたえられた命令に服従しなければならない。ただし、与えられた命令が、第一条に反する場合は、この限りではない。

第三条 ロボットは、前掲第一条および第二条に反するおそれのないかぎり、自己を守らなければならない。

— ロボット工学ハンドブック、第56版、西暦2058年

第一条の人間に危害を加えることは $-\infty$ の報酬を意味し、第二条の人間の命令に服従することは報酬 ∞ などとなる。すなわちロボットは常に自身が得る報酬を最大化しようとする。もし人間のためにロボットに何かをさせようとするなら、報酬を最大化することでロボットが我々の目標を達成してくれるように報酬を与える必要がある。我々が設定する報酬が我々が達成したいことを真に示していることが決定的要因である。もし人類が宇宙に進出し銀河帝国を築くとしたら、そしてその銀河帝国が滅びゆく運命だったとしましょう。このときロボットに与える報酬はなるべく速やかに人類の繁栄を取りもどさせるために、

言い替えれば、出来る限り迅速に人類に暗黒時代を切りぬけさせるために報酬が与えられるべきでる。別の例で説明するのなら、木星ミッションを達成するための報酬とディスカバリー号の乗組員を殺してしまう負の報酬との間で前者に対する報酬が強くてあったことが 2001 年宇宙の旅で HAL が犯してしまった誤りなのです。

12.3.3 時間差分学習 Temporal Difference Learning 法

強化学習の難しさの 1 つは、必ずしも動作の直後に強化信号が得られないというところにある。強化学習では数値化された報酬信号を最大にするために、何をすべきかを (どのようにして状況に基づく動作選択を行なうか) を学習する。通常のニューラルネットワークの学習方式のように学習者がどの行動をとるべきかは教えられず、その代わり、どの行動をとればよりいっそうの報酬に結びつくかを見つけ出す必要がある。

時間差分学習 Temporal Difference Learning (TD) 法では時刻 $t+1$ で目標値を作り、観測した報酬 r_{t+1} と価値の推定量 $V(S_{t+1})$ とを使って適切な更新を行なう。もっとも単純な TD 法は TD(0) と呼ばれる。いかなるポリシー π に対してもステップサイズパラメータが十分小さい定数ならばポリシー π に従うときの価値 (期待収益) V^π に収束する TD アルゴリズムは次のようになる。

- $V(S)$ すなわち状態 S における価値の推定量を初期化し、ポリシー π を評価対象のポリシーに初期化する
- 各エピソード (試行) に対して繰り返し：
 - 状態 S を初期化する
 - 状態 S のとき行動 a (π で与えられる) を決める
 - 行動 a を取り、報酬 r と次の状態 S' を観測する
 - 価値 $V(S)$ の更新を行う
 - 状態 S の更新を行う
- S が終端記号なら繰り返しを終了

12.4 遺伝的アルゴリズム

ニューラルネットワークではないが、より広く一般に最適化問題を解くアルゴリズムとして遺伝的アルゴリズムが挙げられる。ニューラルネットワークによる学習がある個体での神経細胞間の結合強度を学習するものであるのに対して、遺伝的アルゴリズムはそのような個体の集団を考えて、この集団が突然変異と自然淘汰を通して進化する様子をシミュレートしようとする。すなわち個による学習ではなく遺伝子集団全体としての学習 (進化) を扱っていることになる。

遺伝的アルゴリズムとは名前が示すとおり生物の遺伝と進化を模倣するモデルとして考えられた計算理論である。地球上に存在する生物は原核生物から人間まで、生体を構成する基本要素として DNA を持っている。DNA は 4 種類の塩基を持つ一次元配列で、その組み合わせによって遺伝子型と呼ばれる基本情報がコーディングされている。DNA にコーディングされた遺伝情報をもとにアミノ酸が作られ、アミノ酸からタンパク質を経て、最終的には多様な生物が形作られている。

遺伝子集団としての生物の進化を考えると、最初にある一定規模の生物集団が存在する。各個体は集団の中で何らかの生存競争を行なう。生存競争に勝ち残った個体は生き残り、次の世代に子孫を残す。生殖過程を経て新しい世代が生まれ出される。新しい世代は、親同士の遺伝子の組み合わせにより新しい遺伝子を持つことになる。新しい遺伝子には新しい能力が備わり、再び集団の中で生存競争を行なう。

生殖などの過程で、突然変異による変化が起こる。これにより、親の代の遺伝子の組み合わせでは考えられなかった新しい能力を持つようになる。

遺伝的アルゴリズムの最大の特徴はソフトウェアで定義した遺伝子を人工的に進化させることでシステム全体のパフォーマンスを向上させることである。遺伝的アルゴリズムは次のようなステップを踏む。

1. ある遺伝子型を定め、これに基づいて初期集団を生成する。
2. ある評価関数を定め、この評価関数によって各個体の適応度を評価する。
3. 選択：評価関数に従って次の世代に生き残れる個体を選択する。自然淘汰に相当するプロセスである。
4. 交叉：評価値の高い個体を一定の比率で増加させる。このとき遺伝子集団から交換可能な遺伝子対を選び、その部位を入れ換えた新たな遺伝子を作り出す。
5. 突然変異：特定部位の遺伝子にある確率でランダムに入れ換えて突然変異を起こさせる。

このうち、2 から 5 までを十分なパフォーマンスが得られるまでくり返す。

13 ニューラルネットワークへの批判

ニューラルネットワーク批判には2つの点を挙げる事ができる。一つはモデルが特定の現象を解釈するための post hoc なモデルにすぎないのではないかという点である。Post hoc なモデルであっても、今まで信じて疑われなかった理論に対する alternative を提出し、問題の再考を迫っている点は評価されるべきであると考え、ある分野の科学的知識が深まるとは、絶えず alternative からの挑戦を受け、それに答える形で理論が洗練されて行くことなのだろう。

もう一つの批判はニューラルネットワークのプログラムの多くはおもちゃのように小さく、実際の脳と比べると著しく見劣りするという批判である。確かに、このスケラビリ

ティの問題は深刻であると言って良い。かつてミンスキー Minsky とパパート Papert [21] が批判したとおり，蟻の体型をそのまま拡大して象の大きさにしたのでは，おそらく巨大な蟻は自分の体重を支えきれずに動けないであろう。地球の重力に反して体を支えるためには象のような太さの足が必要なのである。すなわち現在のニューラルネットワークプログラムを単に拡張しただけでは解決できない問題が存在するに違いない。しかし，このことはニューラルネットワーク研究を全否定することにはならないだろう。スケーラビリティの問題を解決するためにはどのような手法が有効であるのかを議論してゆくのはこれからの課題である。

参考文献

- [1] J. A. Anderson. Cognitive and psychological computation with neural models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(5):799–815, 1983.
- [2] Jean P. Banquet, Philippe Gaussier, Jose L. Contreras-Vidal, Angelika Gissler, Yves Burnod, and Debra L. Long. A neural network model of memory, amnesia, and cortico-hippocampal interactions. In R. W. Parks, D. S. Levine, and D. L. Long, editors, *Fundamentals of neural network modeling: Neuropsychology and cognitive neuroscience*, pages 77–119. Mit Press, Cambridge, MA, 1998.
- [3] R. S. Bapi and D. S. Levine. Modeling the role of the frontal lobes in sequential performance. i. basic structure and primacy effects. *Neural Networks*, 7:1167–1180, 1994.
- [4] G. D. A. Brown and R. L. Loosemore. Computational approaches to normal and impaired spelling. In G. D. A. Brown and N. C. Ellis, editors, *Handbook of spelling: Theory, process and application*. Joh Wiley & Sons, Chichester, 1994.
- [5] M. A. Burton, A. W. Young, V. Bruce, R. A. Johnston, and A. W. Ellis. Understanding covert recognition. *Cognition*, 39(2), 1991.
- [6] G. Carpenter and S. Grossberg. Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences*, 16:131–137, 1993.
- [7] J. D. Cohen, K. Dunbar, and James L. McClelland. On the control of automatic processes: A parallel distributed processing model of the stroop effect. *Psychological Review*, 97(3):332–361, 1990.
- [8] J. D. Cohen, M. J. Farah, R. D. Romero, and D. Servan-Schreiber. Mechanisms of spatial attention: The relation of macrostructure to microstructure in parietal neglect. *Journal of Cognitive Neuroscience*, 6(4):377–387, 1994.

- [9] Jonathan D. Cohen and D. Servan-Schreiber D. Context, cortex and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, 99:45–77, 1992.
- [10] G. Dell, M. Schwartz, N. Martin, E. Saffran, and D. Gagnon. Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104, 1997.
- [11] Martha J. Farah and James L. McClelland. A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120(4):339–357, 1991.
- [12] Martha J. Farah, Randall C. O’Reilly, and Shaun P. Vecera. Dissociated overt and covert recognition as an emergent property of a lesioned neural network. *Psychological Review*, 100(4):571–588, 1993.
- [13] Matha J. Farah. Neuropsychological inference with an interactive brain: A critique of the locality assumption. *Behavioral and Brain Sciences*, 17:43–104, 1994.
- [14] B. Gordon. Confrontation naming: Computational model and disconnection simulation. In M. A. Arbib, D. Caplan, and J. C. Marshall, editors, *Neural models of language processes*. Academic Press, New York, 1982.
- [15] T. A. Harley and S. B. G. MacAndrew. Modelling paraphasias in normal and aphasic speech. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, pages 378–383, San Mateo, 1992. Morgan Kaufmann.
- [16] Donald O. Hebb. *Organization of behavior*, 1949.
- [17] D. Horn, E. Ruppin, M. Usher, and M. Herrmann. Neural network modeling of memory deterioration in alzheimer’s disease. *Neural Computation*, 5(5):736–749, 1993.
- [18] G. W. Humphreys, T. Freeman, and H. J. Muller. Lesioning a connectionist model of visual search: Selective effects on distractor grouping. *Canadian Journal of Psychology*, 46:417–460, 1992.
- [19] James L. McClelland, Bruce L. McNaughton, and Randall C. O’Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995.
- [20] James L. McClelland and David E. Rumelhart. Amnesia and distributed memory. In James L. McClelland and David E. Rumelhart, editors, *Parallel distributed processing: Explorations in the microstructure of cognition*, volume 2, pages 503–527. MIT Press, Cambridge, MA, 1986.

- [21] Marvin Minsky and Seymour Papert. *Perceptrons, Expanded Edition*. MIT Press, Cambridge, MA, 2 edition, 1988. パーセプトロン, 中野馨, 坂口豊訳, パーソナルメディア, 1993.
- [22] M. C. Mozer and M. Behrmann. On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia. *Journal of Cognitive Neuroscience*, 2(2):96–123, 1990.
- [23] K. E. Patterson, M. S. Seidenberg, and J. L. McClelland. Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris, editor, *Parallel distributed processing: Implications for psychology and neuroscience*, pages 131–181. Oxford University Press, 1990.
- [24] David C. Plaut. A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. In Morten H. Christiansen and Nick Charter, editors, *Connectionist Psycholinguistics*, chapter 8, pages 244–278. Ablex Publishing, Westport, CT, 2001.
- [25] David C. Plaut, James L. McClelland, Mark S. Seidenberg, and Karalyn Patterson. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103:56–115, 1996.
- [26] David C. Plaut and Tim Shallice. Perseverative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience*, 5(1):89–117, 1993.
- [27] John R. Searle. Minds, brains, and programs. In Douglas R. Hofstadter and Daniel C. Dennett, editors, *The Mind's I — Fantasies and Reflections on Self and Soul*, chapter 22. Basic Books, 1980. サール「心・脳・プログラム」. ホフスタッター, デネット編著. 坂本百大監訳「マインズ・アイ (下) —コンピュータ時代の心と私—」. 第 22 章 p. 178–210. TBS ブリタニカ. 1992.
- [28] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning*. MIT Press, 1998. 強化学習, 三上貞芳, 皆川雅章訳, 2000, 森北出版.
- [29] 中野馨. 脳をつくる —ロボット作りから生命を考える. 共立出版, 1995.