

連載解説 「Deep Learning (深層学習)」 [第2回]

多層ニューラルネットワークによる深層表現の学習

Deep Representation Learning by Multi-Layer Neural Networks

麻生 英樹
Hideki Asoh

産業技術総合研究所
National Institute of Advanced Industrial Science and Technology, AIST.
h.asoh@aist.go.jp

Keywords: deep neural networks, stacked autoencoders, representation learning.

1. はじめに

This issue is important, because how information is represented can greatly affect how easy it is to do different things with it. David Marr “Vision”

「内部表現の発見・獲得・学習」, すなわち, さまざまな情報が混在し, 雑音で汚れている実世界の観測情報から, 本質的な情報やある課題(群)に必要な情報を抽出し, 処理しやすいように表現することは, 人工知能やパターン認識を始めとする知的情報処理における古くからの研究課題の一つである。この課題に対して, これまでにさまざまなアプローチでの研究が行われてきたが, 近年, 層の数が多い階層的なニューラルネットワーク(deep neural network: DNN)によってデータから抽象度の高い内部表現を獲得させる方法が, 深層学習(deep learning)として脚光を浴びている。そこで得られた内部表現を用いた手法が, 一般物体認識[Krizhevsky 12], 連続音声認識[Dahl 12], 自然言語処理, 化合物の活性予測, などさまざまな分野のコンペティションやベンチマークタスクで従来法を大きく引き離す性能をたたきだしていることがその直接的な理由である。

この連続解説の第1回では, 深層学習に用いられている代表的なモデルの一つであるリストリクティドボルツマンマシン(restricted Boltzmann machine: RBM)を中心に詳しい技術的解説が行われた。本稿では, もう少し歴史的・思想的な側面に重点を置きつつ, 階層的なニューラルネットワークを用いた深層学習について解説したい。

以下では, まず2~4章までで, 深層学習の基盤となる内部表現の学習と階層的なニューラルネットワークについてそれぞれ概説する。その後に5章で両者の関係と最近の深層学習への展開について解説し, 6章で課題と展望を述べる。内部表現の学習とニューラルネットワークにはどちらも歴史があり, 簡単に説明するのが難しい

のだが, 何とか頑張りたい。

2. 内部表現の重要性とデータからの学習

深層学習では, 層の数が多いニューラルネットワークによって, 観測データから本質的な情報を抽出した内部表現・潜在表現・特徴量を学習する。本章では, 深層学習の意義を理解するために, 内部表現の学習の重要性とこれまでの研究について述べる。

知的な情報処理とその学習における内部表現の重要性は, 人工知能, 認知科学, 機械学習, データ解析などの研究において古くから何度も指摘されてきた[Beugio 09, Beugio 12]。例えば, 音声認識や画像認識などのパターン認識においては, 通常, 入力である音声や画像などのパターン情報から特徴ベクトル(feature vector)を抽出する[Duda 00, 石井 98]。ここで認識に適した少数の良い特徴量が抽出できれば, それを並べた特徴ベクトルを線形識別器などの簡単な識別器でクラスラベルに対応付けることが可能になる。そして, 簡単な識別器のつくる空間が狭いことから, その空間の中での最適な識別器の探索, すなわち教師ありのデータを使った識別器の学習も容易になる。

多変量データ解析の分野では, 主成分分析や因子分析のように, データの変動の構造を説明する少数の変数を抽出することが, 研究の当初から重要な課題の一つであった[Anderson 84]。データの低次元化には, データを人間が見やすいように可視化して, その分布の様子を理解しやすくするという目的もあった。

こうした内部表現が有効であるのは, 観測データに, 観測対象の構造を反映した潜在的な構造があるためである。例えば, 高次元の観測データが低次元の部分空間内や低次元の多様体上に集中して分布するような場合には, データの分布に沿った良い座標系を取ることで, データの広がり=情報をあまり失わずに, 観測データを低次元の表現に変換することができる。データが高次元空間全体にわたって分布している場合でも, ある課題に必

要な情報は低次元空間内に収まっていることもある。また、データがいくつかのクラスタに分かれて分布している場合には、データをそれぞれのクラスタにまず分けてから、クラスタごとに処理をすることが有効であることが多い。

古典的 AI における機械学習の泰斗である Winston は機械学習における情報表現の重要性を指摘しているし [Winston 70], コンピュータビジョンの研究に大きな影響を与えた Marr もまた、視覚情報処理の本質を、二つの二次元画像から、原始スケッチ, $2\frac{1}{2}$ 次元スケッチ, 三次元モデル表現, といった, より高次元な内部表現を復元する表現変換であると定式化した [Marr 82]. Marr はまた、大脳皮質の情報処理一般についても、それが観測可能な情報から観測不可能だが生存に重要な本質的情報を推測することであると考えていた [Marr 70]. また、複数のタスクを同時に学習するマルチタスク学習や、あるタスクに対する学習結果をほかのタスクに転用する転移学習の研究においては、特定のタスクに適した情報表現ではなく、複数のタスクに適した汎用的な情報表現の獲得が重要な課題とされている [Baxter 95, Caruana 95, 神畠 10]. そのほかにも、強化学習における状態空間表現の獲得や、記号の創発 [JSAI 12], 低次元のセンシング結果からスパース性を利用して元の信号を復元する圧縮センシング [Donoho 06, 田中 10] などの研究も内部表現の学習と関連が深い。

内部表現の重要性はまた、パターン認識における醜いアヒルの子の定理 [Duda 00, Watanabe 69] や機械学習における No Free Lunch 定理 [Duda 00, Wolpert 96] とも関係している。醜いアヒルの子の定理では、個々の対象を表現するための特徴量を選択することなしに表現の近さに基づいた対象の分類は不可能であることが述べられている。また、No Free Lunch 定理では、すべてのタスクに対してはかより優れた性能を示す万能学習アルゴリズムは存在しないことが主張されているが、特定のタスクに対する優れた性能は、その学習アルゴリズムが用いている情報表現や特徴量に由来することが多い。このことは逆に、世の中にある多くの問題に適した汎用性の高い内部表現が得られれば、多くの現実問題に対して良い性能を示す学習器が作れる可能性があるということの意味している。

あるタスク、あるいは、タスク群に適した良い内部表現の構成（特徴工学 (feature engineering) などと呼ばれることもある）という問題に対して、古来から二つのアプローチが競合してきた。一つ目は、対象課題の性質の理解や領域知識に基づいて、人手によって内部表現をデザインするというアプローチである。例えば、画像認識のための特徴量としては、認識結果に影響を与えないような画像の平行移動や回転に対して値が不変となる特徴量の設計が行われてきた。最近の自然言語処理においては、ある構造（機械翻訳の出力文や推定されたタグ

系列）の良さを、対数線形モデル構造から抽出した特徴の重み付け和によって評価することがよく行われているが、そこでも、どのような特徴を用いるかは重要なヒューリスティクスになっている。

もう一つのアプローチは、データに基づいて、良い内部表現を学習させるものである。統計的機械学習の研究の多くは、タスクに依存した特徴量が得られた後の、タスクにあまり依存しない処理を対象としているが、良い内部表現 (= 特徴量) 自体をより生に近いデータから学習させるという研究も行われてきた。そこには、説明変数選択、カーネル設計、マルチカーネル学習、データ行列の低ランク分解、なども含まれる。最近では、自然言語処理などでも行われているように、人手で有望そうな候補をたくさんつくって、データを使った学習でその中から選択する、という、折衷的なアプローチも良く使われている。

内部表現を自動的に学習させる場合、何らかの客観的な評価尺度が必要になる。抽出した表現を用いて、予測や識別、分類などのタスクを行う場合には、そうした個々のタスクの性能で評価することが考えられるが、汎用的な特徴量の抽出のためには、以下のような評価の観点が使われている。

- 情報量：入力信号の情報量をできるだけ多く保存していること
- 独立性：相互の特徴量ができるだけ独立であること
- 説明性：どのような情報が抽出されているかが解釈・説明しやすいこと
- スパース性：少ない数の特徴量の組合せで個々の対象が表現できること
- 不変性：入力信号の特定の変換に対して変化しないこと
- ロバスト性：入力信号の微小変動に対して変化しないこと
- 滑らかさ：元の情報が変化したときに特徴量が滑らかに変化すること

これらの中には、互いに相反するようなものも含まれている。例えば、説明性や独立性を上げると、情報量は犠牲になる可能性がある。

次元圧縮による内部表現の学習法としては、多変量データ解析の分野で提案された主成分分析 (principal component analysis: PCA) がよく知られている。この手法は、圧縮後の分散が最大になる軸 (より一般には、元の信号空間の部分空間) にデータを線形写像することで次元圧縮を行うが、元のデータが多変数正規分布に従っていることを仮定すれば、分散最大と情報量最大は等価であり、元の情報の分布の情報量を最も良く保存するような写像になっている。類似の手法として、因子分析 (factor analysis: FA) では、主成分分析の軸を回転させて、より解釈しやすい特徴量が探索される。また、独立成分分析 (independent component analysis: ICA) では、

独立性の高い特徴量を抽出することで、線形な重み付き足合わせによって混合してしまった情報から、混合前の情報を復元できることが示された [Bishop 06].

個々の入力信号に対して疎な内部表現，すなわち，ゼロでない値を取る変数の数が少ない内部表現を得るスパースモデリングも，汎用的な内部表現学習への有力なアプローチである．例えば，Olshausen と Field は，自然画像から切り出されたパッチの集合を入力として，スパースモデリングを行い，動物の脳の視覚情報処理系にある特徴抽出器とよく似た特徴抽出器が学習されることを示した [Olshausen 96, Olshausen 97].

主成分分析などでは，信号空間の中の部分空間への線形の射影を求めるが，データが信号空間の中で曲がった低次元多様体上に分布している場合もよくある．このような場合に，すでに述べたように，多様体上の座標系を求めることができれば，あまり情報を失わずにデータを低次元のベクトルで表現することが可能になる．このような非線形の高次元多様体を学習する問題は，多様体学習 (manifold learning) などと呼ばれて盛んに研究されている [赤穂 08, Cayton 05].

3. 階層型ニューラルネットワーク

以下の二つの章では，深層学習の基盤となる階層型ニューラルネットワークの研究を概観する．ニューラルネットワーク (モデル) は，人間を筆頭とする動物の情報処理システムの中核である脳神経系を模倣した，並列分散学習的な情報処理モデルである [麻生 88, Bishop 06, Duda 00, 中野 05]. その研究は 1940 年代にまで遡ることができる．

我々の脳は，ニューロンと呼ばれる神経細胞がシナプスと呼ばれる結合部位を介して多数結合してできているネットワークである，ということが示されたのは 20 世紀の初頭だった．その発見を受けて，McCulloch と Pitts は，比較的単純な素子のネットワークによる情報処理のモデルについて考察している．彼らは，ニューロンを多入力 1 出力のしきい素子としてモデル化し，それを複数結合することで任意の論理関数が計算可能であることを示した [McCulloch 43]. その後，連続値のニューロンを用いてチューリングマシンの模倣ができることなども示されている [林原 90, Siegelmann 91].

単純な素子のネットワーク＝並列分散処理，と並んで脳神経系の情報処理の特徴とされるのは，学習能力，すなわち，経験を積むうちに，処理能力を適応的に変化させて，より良い環境情報の理解や行動が可能になっていくということである．Hebb は，条件反射のような行動変化において，神経細胞間のシナプス結合の変化が大きな役割を果たしていると考え，結合変化による学習のモデルを示した [Hebb 49]. さらに，こうした研究を踏まえて，Rosenblatt がパーセプトロン (perceptron)

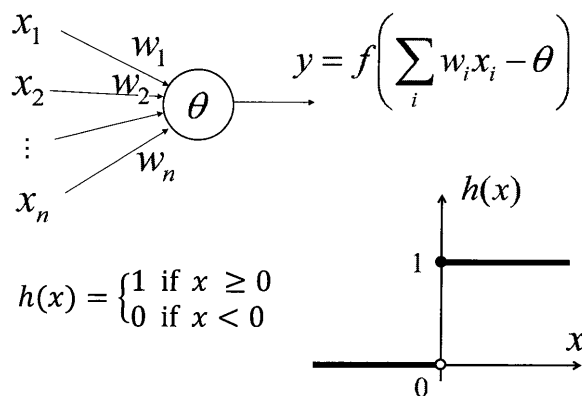


図1 しきい関数を出力関数とするニューロンのモデル

という脳の認知機能の学習的なモデルを提案した [Rosenblatt 58]. 1950 年代から 60 年代にかけて，脳神経系における情報処理に関する研究は非常に盛んになり，例えば，人工知能の大家として有名な Minsky や，経済学の大家である Hayek も，この時期に脳やニューラルネットワークによる情報処理の研究をしている [Hayek 52, Minsky 87].

ニューロンとそのネットワークを数理的にモデル化するやり方にはさまざまなものがある．例えばボルツマンマシンでは，ニューロンは確率的に動作する．この解説では，それぞれのニューロンを多入力 1 出力の非線形素子としてモデル化する研究を紹介する．ニューロンは多次元の入力 $\mathbf{x} = (x_1, \dots, x_n)$ を受け取り， y を出力する．その情報処理を式で書くと

$$y = f\left(\sum_{i=1}^n w_i x_i - \theta\right)$$

となる．ここで， x_i は \mathbf{x} の第 i 要素， w_i は第 i 番目の入力に対する結合の重みと呼ばれる数で，シナプスの信号伝達の仕方を表すパラメータである． θ はしきい値と呼ばれる．なお，常に 1 の値をとるダミー変数 x_0 を導入して， $\mathbf{x} = (x_0, \dots, x_n)$ ， $\mathbf{w} = (w_0, \dots, w_n)$ とすれば，上式は

$$y = f\left(\sum_{i=0}^n w_i x_i\right) = f(\mathbf{w}^T \mathbf{x}) \quad (1)$$

と書ける．ここで， $w_0 = -\theta$ である．すなわち，しきい値 θ をダミー変数からの結合の重みとみなすことで，式の形を簡単にできる．以下ではこの形の表現を用いる． \mathbf{w}^T はベクトルの転置を表し， $\mathbf{w}^T \mathbf{x}$ は重みベクトルと入力ベクトルの内積である．

関数 f は一般には非線形の関数であり，ニューロンの出力関数，活性化関数などと呼ばれる．パーセプトロンでは， f として，以下のようなしきい関数 (ヘビサイド関数, sign 関数などとも呼ばれる) が用いられた (図 1)．

$$h(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

複数のニューロンを結合したネットワークの構造＝アーキテクチャは，大きく，相互結合型 (mutually

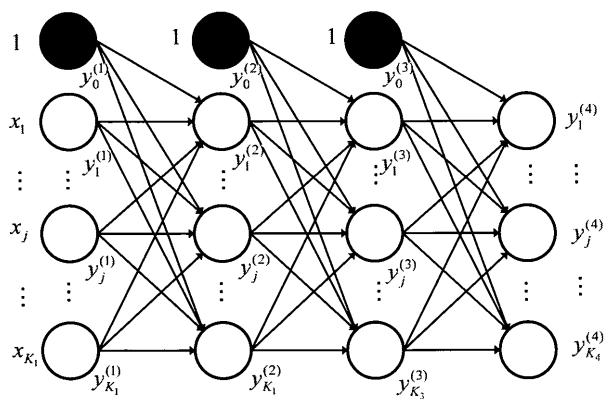


図2 4層の階層型ニューラルネットワーク

connect) と階層型 (layered) に分けることができる。相互結合型のニューラルネットワークは、ニューロン同士が双方向に結合している一般的なネットワークであり、信号はネットワーク内を繰り返し伝播していく。特に、ニューロン A からニューロン B への結合とその逆方向の結合の重みが同じ値を取る対称結合のネットワークが、連想記憶のモデルなどとしてよく研究された。ボルツマンマシンもまた、対称な結合をもつ相互結合型のニューラルネットワークの一種であるが、ニューロンが確率的に動作し、ネットワーク全体で確率分布を表現している点に特徴がある。

一方、階層型ニューラルネットワークは、ニューロンが複数の層に分かれており、ある層のニューロンは、その層よりも前の層のニューロンから結合を受ける。それぞれの結合には結合の重みが割り当てられている。図2に4層の階層的ニューラルネットワークの例を示した。入力層の信号は複数の中間層（隠れ層とも呼ばれる）を経て出力層へと一方向に伝播されながら、式(1)に従って変換されていく（入力層は、通常、入力信号をそのまま出力する、すなわち、 $y_i^{(1)} = x_i$ である。そのため、図2のネットワークを3層と数えることもある）。

図では、ある層のニューロンは一つ前の層のニューロンからしか結合を受けていないが、それよりも前の層から結合を受けることもできる。また、一つ前の層のすべてのニューロンから結合を受けているが、一部のニューロンだけから結合を受けるように構成することもある。その場合、そのニューロンに結合している前層のニューロンの集合は、当該ニューロンの受容野 (receptive field) と呼ばれる。

こうしたニューラルネットワークは多層 (multi-layer) ニューラルネットワーク、あるいは多層パーセプトロンとも呼ばれる。ネットワークの中の結合の重みをすべてまとめて W と書くと、階層型ニューラルネットの出力は、 $y = g(x, W)$ のように入力ベクトルと結合の重みによって決まる。

4. 階層型ニューラルネットワークの学習

本章では、階層型ニューラルネットワークの学習について述べる。階層型ニューラルネットワークの学習は、大きく、教師あり学習と教師なし学習に分けることができる。教師あり学習とは、入力 x とそれに対する正解出力 t のペアの集合 $\{x_j, t_j\}, j=1, \dots, M$ を学習用データとして、ニューラルネットワークに x_j を入力したときの出力が t_j に近づくように結合の重み W を修正する。一方、教師なし学習では、入力信号の集合 $\{x_j\}$ だけから学習を行う。以下では、まず、教師あり学習の例として誤り訂正学習 (パーセプトロン学習) と誤差逆伝播学習 (error-backpropagation learning) を紹介し、その後で、教師なし学習の例として競合学習 (competitive learning) を紹介する。誤差逆伝播学習や競合学習は、深層学習でも用いられている。

4.1 誤り訂正学習

最も単純な階層型ニューラルネットワークは、入力層と出力ニューロン一つだけのものである。入力層は入力値をそのまま出力するだけなので、情報処理と学習は出力ニューロンでのみ行われる。この情報処理モデルは単純パーセプトロン (simple perceptron) などと呼ばれる。出力関数にしきい関数を用いた単純パーセプトロンの情報処理は、入力信号の空間を超平面で二つに分割して、一方に対して1を、もう一方に対して0を出力するものである。このような処理は入力空間の線形分離と呼ばれる。

単純パーセプトロンの学習は、入力 x とそれに対する正解出力 t のペアの集合 $\{(x_i, t_i)\} (i=1, \dots, M)$ を学習用データとして、各入力に対して正解を出力するような結合重み $w = (w_0, \dots, w_n)$ を求める問題である。この問題を解くための学習アルゴリズムとして、誤り訂正学習則 (パーセプトロン学習則) がよく知られている。

誤り訂正学習則では、以下のように、学習用データの入力ごとに、パーセプトロンの出力が誤り、すなわち正解と一致しない場合には重みを少しずつ修正することを繰り返す。

- (1) 学習データの中の入力 x_i に対して、現在の重み w を用いて出力 y を計算する。
- (2) y が正解 t_i と一致していたら何もしない。
- (3) y が正解 t_i と一致していない場合には、重み w を $w \leftarrow w + \eta(t_i - y)x_i$

のように更新する。 η は一回の更新量を決める小さな正の数で、学習係数と呼ばれる。

このように、入力が与えられるごとに少しずつ学習を進めるような学習はオンライン学習と呼ばれる。これに対して、すべての学習用データに対する修正量を計算した後で、一気に修正を行うような学習をバッチ学習と呼ぶ。

誤り訂正学習則は、すべての学習用データに対して正

解を出力するような重み w が存在する場合（このような場合を「データが線形分離可能である」という）には、学習中の重みと正解を与える重みとの距離が重みの更新ごとに単調減少することから、有限回の修正の後に収束することを証明できる（パーセプトロン学習の収束定理）[中野 05]。

単純パーセプトロンの情報処理能力はあまり大きくないが、その前に前処理層（連合層などと呼ばれた）を加えることで、情報処理能力を拡張することができる。すなわち、入力信号を固定結合のニューロンの層で一度非線形変換した後に、その出力を単純パーセプトロンで識別学習させることが考えられる。これは、一般化線形モデルの一種とも考えることができる。

前処理層のニューロンを十分にたくさん使えば、入力層からの結合重みをランダムに決めた場合でも、前処理層の出力ベクトルが線形分離可能になる確率を高めることができる。しかし、前処理層の結合重みが固定であること、すなわち、特徴があらかじめ決められていることによる限界は存在する。Minsky らが 1960 年代後半に、前処理層を導入した場合でも、前処理層の結合の重みに自然な制約を入れた場合には、入力を画像と見た場合の連結性の判定など、パーセプトロンでは計算できない問題があることを理論的に示したこともあり [Minsky 87]、パーセプトロンを始めとするニューラルネットワークの研究はしだいに下火となっていった。

4.2 誤差逆伝播学習

パーセプトロンの学習則の限界は、出力層のニューロンへの結合の重みしか修正できないということであった。これに対して、出力層以前の中間層（隠れ層とも呼ばれる）のニューロンへの結合重みも修正できる学習法としてよく知られているのが 1980 年代に Rumelhart, Hinton, Williams によって提案された誤差逆伝播学習である [Rumelhart 86]。

誤差逆伝播学習法では、ニューロンの出力関数を不連続なしきい関数から微分可能なシグモイド関数 $f(x) = 1/(1+e^{-x})$ に変更して、出力層での誤差評価 $R(W)$ への任意の結合重み w の寄与を、偏微分係数 $\partial R(W)/\partial w$ として計算できるようにした。これにより、最急降下法などの通常の最適化手法によって、出力と正解との間の誤差が小さくなるようにすべての重みを修正することが可能になった。

層の数を N 、第 n 層の第 k ニューロンから第 $n+1$ 層の第 j ニューロンへの結合の重みを $w_{k,j}^{(n+1)}$ とするとき、最急降下法（厳密には確率的降下法）による結合重みの修正規則は以下ようになる。

$$w_{k,j}^{(n+1)} \leftarrow w_{k,j}^{(n+1)} - \frac{\eta \partial R(W)}{\partial w_{k,j}^{(n+1)}}$$

出力層での誤差評価 $R(W)$ が二乗誤差 $(1/2) \sum_j (t_j - y_j^{(N)})^2$ であるとき、具体的な修正規則は

$$w_{k,j}^{(n+1)} \leftarrow w_{k,j}^{(n+1)} - \eta \delta_j^{(n+1)} y_k^{(n)}$$

と書ける。ここで $y_k^{(n)}$ は第 n 層の第 k ニューロンの出力である。 $\delta_j^{(n+1)}$ は、合成関数の微分法に従い、出力層での誤差 $(t_j - y_j^{(N)})$ から計算される $\delta_j^{(N)}$ から出発して、以下のように再帰的に計算される：

$$\begin{aligned} \delta_j^{(N)} &= -(t_j - y_j^{(N)}) y_j^{(N)} (1 - y_j^{(N)}) \\ \delta_k^{(n)} &= \left\{ \sum_{j=1}^{K_{n+1}} \delta_j^{(n+1)} w_{k,j}^{(n+1)} \right\} y_k^{(n)} (1 - y_k^{(n)}) \end{aligned}$$

この計算過程が、出力層での誤差を前の層に伝播させていく形になっているため、誤差逆伝播学習と呼ばれている。

同様の考え方は 1960 年代にも提案されていたが（例えば [Amari 67]）、この方法では、誤差を極小にする重みの値しか求まらない（局所収束）という問題点があるため、あまり注目されてはいなかった。しかし、Rumelhart らがこの手法をいくつかの具体的問題に適用して、学習がうまく進み、興味深い内部表現が得られることを示したことをきっかけとして注目が集まり、さまざまな問題への適用が試みられると同時に、最急降下法以外の最適化手法の適用などさまざまな学習アルゴリズムも提案されるようになった。

4.3 競合学習

教師あり学習は、入力と出力の間の関数関係の学習であるが、教師なし学習は入力信号の分布の特徴を捉える学習である。競合学習は、最も初期に研究された教師なし学習法であり、入力信号のクラスタリングを行う。入力層と出力層からなる 2 層のネットワークに入力 x を与えると、出力層のニューロンはそれぞれ出力を出す。その中で最大の出力を出しているニューロンが、その入力に所属するクラスを表していると考えられる。

アイディアを簡単に示すために、入力信号の長さ $|x|$ と出力ニューロンの結合の重みの長さ $|w|$ がすべて 1 である場合を考える。このとき、結合の重み w と入力 x の内積は、 w と x の角度を ϕ として $\cos(\phi)$ になる。すなわち、入力ベクトルの方向が結合の重みベクトルの方向に近いほど値が大きくなる。

競合学習では、ある入力ベクトル x に対して最大の出力を出すニューロンを選び、そのニューロンへの結合を以下のように更新する：

$$w \leftarrow (w + \eta x) / |w + \eta x|$$

これは、 w を x のほうに少し近づける操作になっている。更新後も w の長さは 1 になる。最大出力を出すニューロンだけが学習することから「競合学習」と呼ばれた。入力信号がいくつかのクラスタに分かれて分布しているとき、入力が与えられるたびに重みを更新していくと、最終的に出力ニューロンの結合の重みはいずれかのクラスタの中心に近づいていき、入力 x に対して、それが所属しているクラスタに対応する出力ニューロンが最大出力を出すようになる。このことは、入力信号の中の頻出

パターンを検出しているともいえる。

ここで紹介したものは、最も単純な競合学習のアルゴリズムだが、より一般の場合に対するアルゴリズムや、クラスタリングの精度や学習速度を上げるためのさまざまな工夫が行われている [Duda 00, 中野 05]。

5. 階層型ニューラルネットワークによる内部表現の学習

本章では、階層型ニューラルネットワークによる内部表現の学習について述べる。まず、階層型ニューラルネットワークと内部表現学習の関係について述べた後で、深層学習に用いられる層の数が多いニューラルネットワークを学習させる方法として、畳込みネットワーク (convolutional neural network: CNN) と積層自己符号化器 (stacked auto-encoder) について述べる。

5.1 誤差逆伝播学習による内部表現の学習

誤差逆伝播学習の最大の特徴は、多層のニューラルネットワークの学習を可能にした点である。これはすなわち、中間層の出力としてタスクに適した内部表現を得ることができる、ということの意味している。Rumelhartらが1986年に発表した論文の主張点もそこにあった [Rumelhart 86]。その研究では、5層のニューラルネットワークを用いて、二つの同型の家系図における人間関係を学習させている。その結果、中間層のニューロンにおいて、二つの家系図に共通な人間関係情報と家系図ごとに異なる人名情報とが分離され、効率良く表現されていることが示された。

さらに、Sejnowskiらは、当時としては大規模な3層のニューラルネットワークを用いて英語の発音を学習させるデモ-NETtalk—を行った。入力層に7文字分のテキストを入力し、中央の文字の発音を出力させるタスクで、中間層には80個のニューロンを用いた。学習には当時のミニコンピュータを用いて1週間程度かかったといわれている。学習過程での出力を音声合成器に入力して、人間の子供が発音を獲得する学習発達過程を思わせるような印象的なデモに仕上げこともあり、おおいに注目を集めた*1。また、得られた内部表現をデータ解析手法を使って分析した結果、音韻種別による階層的なクラスタ構造が存在していることが明らかになった [Sejnowski 87]。

このように、誤差逆伝播学習によって、生に近い入力データから課題に適した内部表現が学習的に獲得できる可能性が示されたことから、階層的なニューラルネットワークを、文字認識、音韻認識、ソナー信号処理など、さまざまなタスクに応用することが試みられることとな

った。手書き数字認識や漢字認識などのタスクでは、生の画像を入力とした場合でも、人手で作成された特徴量をを入力とした場合とほぼ同等の性能が得られることが示されている。

5.2 多層のネットワークの学習

階層型のニューラルネットワークを応用する場合の問題の一つが、層の数をいくつにするべきかである。原理的には十分に多くのニューロンをもつ中間層が1層あれば、任意の入出力関係を実現できることが証明されている [Cybenko 89, Funahashi 89]。しかしながら、タスクによっては、中間層1層では非常に多くのニューロンが必要となり、層の数を増やすほうが性能が上がることも期待される。例えば、Håstedらは、線形しきい素子を使った深さ k の回路で計算できる論理関数の中には、深さ $k-1$ の回路で実現しようとする非常に多くの素子が必要になるものがあることを示している [Håsted 86]。また、視覚系神経回路網などの研究から、生物の脳でも複数の階層をもつ特徴抽出が行われていることが示唆されている。

そこで、誤差逆伝播学習を層が多いネットワークに適用することが試みられたが、その結果明らかになったことは、一部のタスクに特化したケースを除くと、局所収束の問題と、入力層に近い層で最急降下法の勾配が小さくなって学習が進まなくなるという問題があるために、層の数が多いニューラルネットワークの学習は難しいということであった。そこに、Vapnikによるサポートベクタマシンの提案が追い打ちをかけた [Cortes 95]。サポートベクタマシンは、局所解への収束の問題がなく、カーネル法を使うことで複雑な非線形の識別課題にも対処が可能であり、NISTの手書き数字認識を始めとするさまざまなベンチマークタスクでそれまでの研究を上回る性能を示した。また、ニューラルネットワークを確率分布モデルの一種として解釈し、より一般的な統計的機械学習の枠組みの中で研究することも盛んになった [Bishop 06]。こうして、誤差逆伝播学習によって盛り上がった階層型ニューラルネットワークの応用研究もまたしだいに下火となった。

しかし、2006年頃から、再び、多層のニューラルネットワークを用いた情報処理の研究が盛んになっている。その最初のきっかけをつくったのは、Hintonらによるリストリクテッドボルツマンマシンを積み重ねた多層の自己符号化器 (auto-encoder) の構築であった [Hinton 06]。そこでは、全部で九つの層をもつ自己符号化ネットワークが、顔画像、手書き文字、文書の単語ベクトルなどの生データに適用され、データに潜在する構造が特徴として抽出されることが示されている。ここで、多層のネットワークを層ごとに教師なしで事前学習させて、その結果を初期値として、教師あり学習を行うというアイデア、層ごとの事前学習 (layer-wise

*1 このデモの音声は YouTube で聴くことができる。

pretraining) が導入された (前回の解説記事 [安田 13] を参照)。

さらに、2007 年には Bengio らが、階層的なニューラルネットワークを用いた自己符号化器を積み重ねる積層自己符号化器を提案した [Bengio 07]。また、大量の学習用データの存在を背景として、画像認識などのタスクに特化した結合構造をもつ多層のネットワークの学習も復活し、ベンチマーク課題で良い成績を収めている [Krizhevsky 12]。以下ではまず、タスクに特化した構造をもつ畳込みニューラルネットワークについて説明し、その後で、自己符号化器と積層自己符号化器について説明する。

5.3 畳込みニューラルネットワーク

層の数が多いニューラルネットワークをうまく学習させるためのアイデアとして、タスクに応じた結合構造をあらかじめつくり込むことで、結合重みの自由度を減らし、学習を容易にするというものがある。こうしたネットワークの例としては、福島 [福島 79] や LeCun [LeCun 89] によって、手書き文字などの画像認識のために提案されたものがよく知られており、現在では畳込みニューラルネットワークと呼ばれている。

CNN の源流の一つである、福島が提案したネオコグニトロン (neocognitron) の構造と原理を図 3 に示す [福島 79, Fukushima 13]。ネオコグニトロンは、S 層 (単純細胞層) と C 層 (複雑細胞層) を組み合わせた 2 層の神経回路を基本モジュールとして、このモジュールを複数個 (図 3 では四つ)、階層的に接続した構造をしている。S 層、C 層のニューロンは局所的な受容野をもつ、すなわち、前の層のニューロンのうち、局所的な一部だけから結合を受けている。図 3 上の中の間円錐は、その頂点にあるニューロンが円錐の底面内にあるニューロンから結合を受けていることを表している。さらに、S 層のニューロンは複数の細胞面 (図 3 上の各層の中の矩形) に分かれており、同じ細胞面内のニューロンは、異なる位置の受容野をもつが、結合重みの値は共有している。こうした構造は、生物の脳の視覚神経系の構造からヒントを得て考案された。

S 層のニューロンへの結合重みを、入力信号を提示しながら教師なしの競合学習によって更新すると、S 層のニューロンは、一つ前の層の出力の中に存在する特徴的なパターンを検出する特徴抽出器となる。一つの細胞面内では結合重みの値が同じになるように学習が進むため、前の層にある特徴パターンの位置によって、細胞面内の対応する場所の特徴抽出細胞が出力を出す。また、競合学習の効果により、異なる細胞面には異なる特徴抽出器が構成される。この情報処理が、入力信号に対する複数の局所的フィルタ関数の畳込み演算になっていることが、CNN の名前の由来である。学習の結果、入力に近い層では局所的なエッジなどの特徴が抽出され、後段

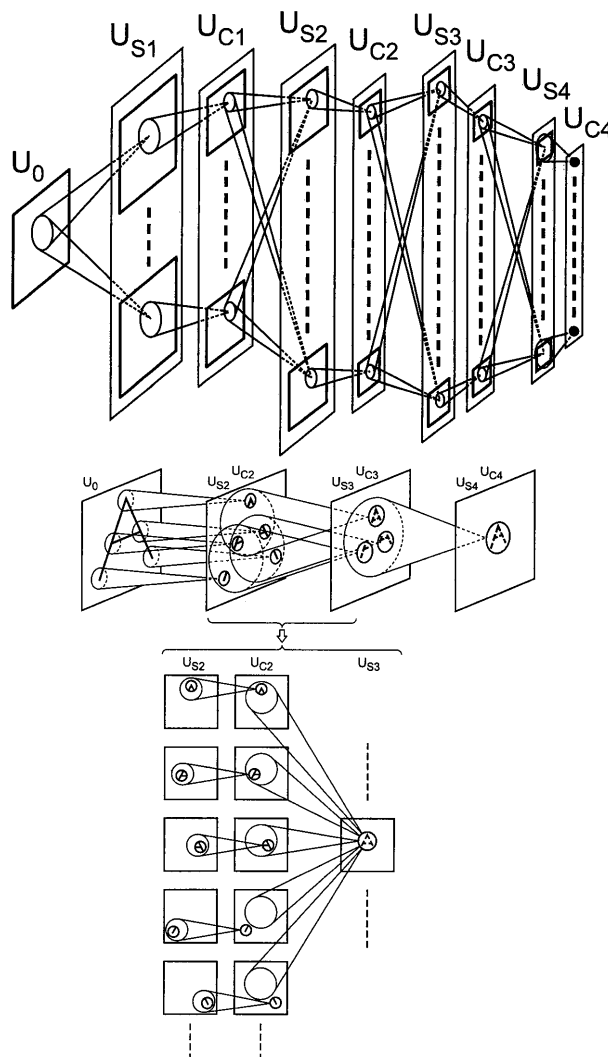


図 3 ネオコグニトロンの構造と原理 (図版提供: 福島邦彦氏)

にいくほど、前段の特徴が複数組み合わせられた、より大域的な特徴が抽出されるようになる (図 3 下)。

一方、S 層から C 層への結合は固定結合であり、C 層のニューロンは、その細胞の受容野内にある S 層のニューロンの中のどれか一つでも大きな出力を出した場合に出力を出すように設定された。この処理は S 層のニューロンの出力を集めて (プールして) C 層の出力とすることから、プーリング (pooling) と呼ばれる。この結果として、C 層のニューロンの出力は S 層への入力パターンの多少の位置変動があっても変化しない。また、通常、C 層のニューロン数は S 層のニューロン数よりも少なく設定されるため、S 層の出力をダウンサンプリングしたことにもなっている。

すでに述べたように、画像認識においては、入力画像内で文字の位置がずれたり、多少変形していたりしても認識結果が変わらないことが望ましいことが多い。CNN は、上記のような構造によって、そうした位置ずれや多少の変形に不変な内部表現を学習することができる。さらに、重みの共有や一部の結合の固定によって、

実質的に学習しなければならない結合の重みの数（自由度）が大幅に減っていることから、学習が容易になっている。畳込み層のニューロンの結合の重み更新には、ネオコグニロンでは教師なしの競合学習、LeCun によって提案された LeNet では、教師ありの誤差逆伝播学習を改良した方法が用いられている。

近年、Krizhevsky らは、大規模な CNN で、階層的にタグ付けられた大規模画像データベース ImageNet のデータを使った大規模一般物体認識コンペティション ImageNet Large Scale Visual Recognition Challenge (ILSVRC) に参加し、2012 年には物体認識と物体検出の 2 部門で、他手法に大幅な差をつけて優勝した。このことは、深層学習の高い可能性を印象付けることとなった。なお、畳込み層の学習には、事前学習なしの誤差逆伝播学習を用いており、NVIDIA の GPU を使って実装している [Krizhevsky 12]。また、畳込み構造は、画像だけではなく、音声などの時系列信号や自然言語テキストに対しても有効であることが知られている。

5.4 積層自己符号化器

誤差逆伝播学習は、入力と出力の関係を学習するために用いられるが、Cottrell らは、図 4 のような砂時計型（ワイングラス型ともいわれる）のネットワークを使い、入力信号自身を教師信号として誤差逆伝播学習させることによって情報圧縮表現を得ることを提案した [Cottrell 88]。図 4 の \hat{x}_i は、ネットワークの出力であり、この値が対応する入力 x_i と一致するように学習を行わせる。

中間層のニューロン数が入力層のニューロン数よりも小さくなっているため、中間層の出力には、入力層の信号を情報をできるだけ保存しつつ低次元化した表現が得られる。各層のニューロンの出力関数をシグモイド関数にせず、入力の重み付き和をそのまま出力するようにすれば、主成分分析と同様の処理となる。さらに、入江らは、2 層目と 4 層目のニューロンの数を多く、3 層目のニューロンの数を少なくした 5 層のネットワークを用いることで、より非線形性の強い符号化が実現できることを示している [DeMers 93, 入江 90]。例えば、高次元の空間の中の低次元の多様体上にデータが集中している場合には、3 層目の出力として、その多様体に沿った座標系でのデータの表現が得られる。こうした研究は多様体学習の先駆けともいえる。

このようなネットワークは、自己符号化器 (auto-encoder) と呼ばれている^{*2}。自己符号化器は、タスクによらない汎用の内部表現を得る手段として興味深いものであったが、情報圧縮の能力自体はほかの符号化法と比べてそれほど高くはなかったため、実用的にはあまり注目されなかった。しかし、それから 20 年近く経った 2006 年頃に、Bengio らは、自己符号化器を、層の多いネ

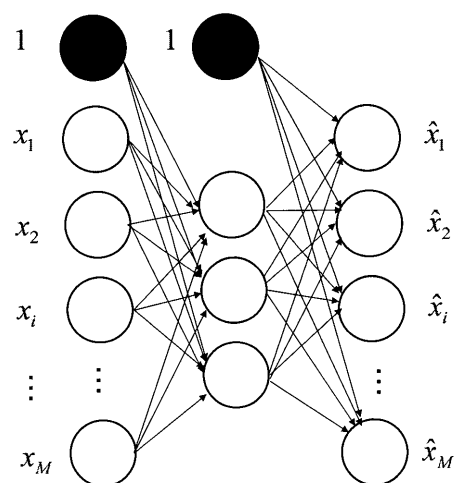


図 4 砂時計型ネットワーク

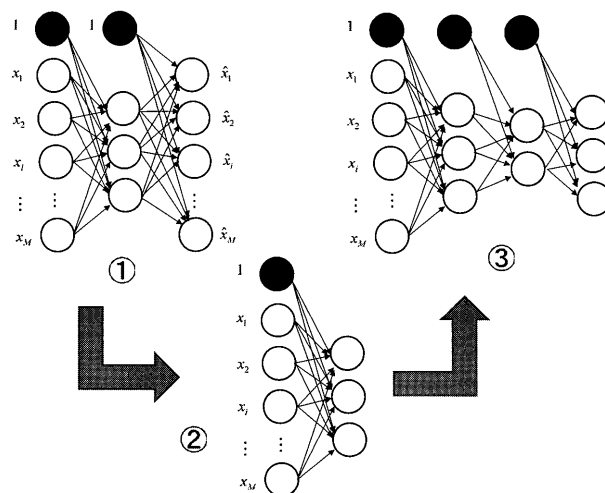


図 5 積層自己符号化器の構成法

ットワークの事前学習手法として用いることを提案した [Bengio 07]。

自己符号化器を使って深いネットワークを構築することは（いわれてみれば）簡単である。まず、自己符号化器を学習させる（図 5 の①）。学習が終了したら、自己符号化器の復号部分（出力層）を取り除き、符号化部分（入力層から中間層まで）を残すことで、入力信号を情報圧縮表現に変換するネットワークが得られる（図 5 の②）。その圧縮表現を入力信号として、さらに自己符号化学習を適用することで、より圧縮された内部表現を得る（図 5 の③）。このようにして、自己符号化学習を再帰的に繰り返すことで、得られるネットワークの符号化部分を積み重ねていくことで、多層の階層的ネットワークを得ることができる。こうして構築される多層ネットワークは積層自己符号化器 (stacked auto-encoder) と呼ばれている。この方法で多層のネットワークを構築した後に、最終層の出力を使った識別ネットワークを追加し、全体として教師あり学習をさせることも行われている。

積層自己符号化器は、RBM を積み重ねた DNN と同

^{*2} 自己符号器とも訳される。

様に、いろいろな課題に適用されている。中でも最近有名になったものは、スタンフォード大学・GoogleのNgらのグループが行った、大量の画像からの特徴抽出器の学習実験である[Le 12]。そこでは、畳込みネットワークのような構造（ただし、重みの共有は行っていない）をつくり込んだ3層のサブネットワークを三つ重ねた9層のネットワークを、1000万本のYouTubeビデオから各1枚ずつ切り出された1000万枚の画像を入力として学習させた。各サブネットワークは自己符号化器の手法で教師なし学習させている。その結果、人の顔やネコの顔、瓶などさまざまな種類の物体に特異的に応答するニューロンが形成されたことが報告されている。また、それらの特徴量を入力としたロジスティック回帰モデルを追加して、ImageNetのデータでネットワーク全体の教師あり学習を行い、積層自己符号化器を使った事前の教師なし学習の有効性を検証している。

§1 スパース自己符号化器

自己符号化器は情報量を保存するような中間表現を学習するが、その他の評価尺度を導入することも可能である。そのためには、復号誤差に、望ましい性質に対応する正則化項（ベイズ的な枠組みにおいては特徴変数の事前分布）を加えた評価関数を最適化するような学習を行えばよい。

一例として、スパース（疎）な中間表現を得ることがあげられる。特徴集合がスパースであるとは、個々の入力情報を表すために必要な特徴（基底・辞書項目などとも呼ばれる）の数が少ないということである。別の言い方をすれば、ある入力情報から計算される特徴量のほとんどがゼロになり、少数の特徴量だけが値を取るということである。入力刺激に対する神経活動の計測などから、動物の脳においては、感覚信号の内部表現としてスパースな特徴量セットが用いられていると考えられている。すでに述べたように、Olshausenらは、自然画像の集合に対するスパースな基底を構成した結果として、動物の視覚系に見られる特徴抽出器とよく似た基底が得られることを示している。

スパースな内部表現を得るための正則化項としては、さまざまなものが提案されているが、例えば、隠れ層の重みのL1ノルムの和を用いるL1正則化があげられる。スパース化の制約を導入することで、入力層よりも多い数の中間層ニューロンを用いた場合でも、恒等写像が学習されてしまうことを防ぐことができる。ちなみに、誤差逆伝播学習の評価関数にL1正則化項を適用してスパースな中間層表現を得るというアイデアは、石川真澄の研究が源流の一つである[石川 90]。

§2 雑音除去自己符号化器

層の多いニューラルネットワークを実現するための、層ごとの学習手法としては、RBMを用いる方法と自己符号化器を用いる方法が代表的である。それらの得失としては、自己符号化器のほうが学習や積み重ねる際の手

間が小さいが、性能はRBMのほうが良い、と考えられている。

Vincentらは、性能の差の由来について検討を行い、性能改善のために、確率分布 $q(\mathbf{x}'|\mathbf{x})$ によって真の信号 \mathbf{x} を雑音で汚した観測信号 \mathbf{x}' に変換し、 \mathbf{x}' を入力して \mathbf{x} を復元するような自己符号化器を提案した[Vincent 10]。このネットワークは、雑音除去自己符号化器(denoising auto-encoder)と呼ばれている。雑音分布 q としては、入力信号が連続値の場合には加法的な正規分布雑音、入力が0/1の2値信号に近い場合には、ランダムに選んだ変数の値を0か1に確率的に変換するSalt-and-Pepperノイズなどが用いられる。

雑音除去自己符号化器は、RBMを用いた学習との関連が高く、両者の関係も研究されている。また、層ごとの事前学習に単純な自己符号化器を用いた場合よりも良い性能を示すことが実験的に示されている。このほかにも、入力の微小変動に対して安定な内部表現を得るための縮小自己符号化器(contractive auto-encoder)などの拡張が提案され、認識課題などでの有効性が実験的に示されている。

6. おわりに一課題と展望

深層学習が現在注目されている最大の理由は、パターン認識タスクや予測タスクにおける性能の良さであるが、深層学習によって得られる内部表現がなぜ良い性能につながるのか、どのようなタスクにどのような構造のネットワークが適しているのか、などについてはまだ十分に理解が進んでいるとはいえない。得られた内部表現の解釈についても、内部表現の多変量解析を行う、深い層のニューロンの値を固定して浅い層のニューロンの値をサンプリングする、ある特徴ニューロンが最も強く反応する入力信号を合成する、などの方法で解釈されているが、解釈が困難なものも多いことが想像される。

また、現在獲得されている内部表現のほとんどが、特徴ベクトルの形のものであることも、個人的には不満を感じる点である。特徴ベクトルはパターン認識課題や予測課題には適しているが、言語の意味のように、組合せ的な構造をもつ潜在情報の表現としては不十分である。そうした組合せ的構造をもつ情報をニューラルネットワークでどのようにして扱うのかは長年の課題であるが、深層学習によってその課題が解決されているようには今のところ見えない。例えば、画像を入力として、クラスラベルではなく、画像の内容を説明する文章を出力するようなDNNがつくれれば、その中間層には意味に対応するような内部表現が獲得されているはずであるが、そのようなネットワークはまだ存在していないように思われる。

こうした意味では、誤差逆伝播学習が提案された時代に多層ニューラルネットワークの課題とされていた、学

習の局所収束, 中間表現の解釈困難性, 組合せ構造への対処, などが深層学習の技術によってクリアに解決されているわけではない。今回のブレークスルーを生んだとされている「層ごとの教師なし学習」も, 実問題でうまくいくことが経験的に知られているが, 基本的には貪欲学習であり, 結果として得られる内部表現が最適なものになっているという保証はない。

現在の深層学習の研究の盛り上がりは誤差逆伝播学習のブームを偲ばせるものであるが, これが1980年代のように一過的なブームとして終わるのか, それとも知的情報処理に関するより本質的な発見やブレークスルーにつながっていくのかを現時点で予想することは難しい。しかしながら, 1980年代と異なるのは, 多くの分野でデータがあふれており, 計算機のパワーも比較にならないほど増大しているということである^{*3}。深層学習が適用できそうな課題は非常にたくさんあり, それぞれで高い性能を示す可能性は高い。そうした中で, 学習アルゴリズムやその理論的解析がさらに洗練されていくことが期待できる。また, ニューラルネットワーク以外のアプローチによる内部表現学習手法の巻き返しも予想される。

2013年3月には, 将棋の現役トッププロ棋士がコンピュータに敗北するという象徴的な出来事があった[電王戦 13]。これは, 一見すると深層学習とは関係がないことのように思われるが, 将棋のプログラムの重要な要素は現在の局面の優劣を評価する盤面評価関数であり, そのために鍵となるのは, 評価関数を構成する盤面特徴量の獲得と選択である。そこにおいては, すでに, プロ棋士の棋譜を学習用データとして, 機械的に構成された非常に多くの特徴量から機械学習(ある種の徒弟学習)によって重み付けて組み合わせられた特徴量が, 将棋が強いプログラマが領域知識を用いて人手で構成した特徴量を凌駕している[保木 07]。

こうしたことは, 誤差逆伝播学習の頃と現在とのもう一つの違いを示唆している。それは, 知的情報処理のタスクが高度化しているということである。具体的には, 手書き数字認識から一般物体認識へ, 音韻認識から連続音声認識へ, 形態素解析や統語解析から談話理解や統計的機械翻訳へ, などの例があげられる。その結果として, タスクに適した特徴量を人手で構成することはますます困難になっている。したがって, 今後さまざまな分野において, 大量のデータに基づく機械学習によって獲得・選択された特徴量が, その分野の専門家によって考えられた特徴量を凌駕していくことは間違いない。特徴量をデータから構成する深層学習が, そうしたトレンドを強く推し進めることも考えられる。また, ほぼ同時期に, Robot Scientist [King 04] などに代表されるような, 人工知能による科学的仮説生成と検証の研究が盛んになっ

てきていることも示唆的である。これはまた, 技術的特異点(Technological Singularity)にもつながっていくのかもしれない[JSAI 13]。

大量のデータと大規模並列処理による高度な内部表現獲得能力を得た人工知能は, 果たしてどこまで人間の知能に迫り, それを超えていくことができるのか, 研究のさらなる進展を期待したい。

謝 辞

図3に使用した図版を提供していただいた福島邦彦氏に感謝いたします。本研究の一部はJSPS科研費23240043の助成を受けた。

◇ 参 考 文 献 ◇

- [赤穂 08] 赤穂昭太郎: カーネル多変量解析—非線形データ解析の新しい展開, 岩波書店 (2008)
- [Amari 67] Amari, S.: Theory of adaptive pattern classifiers, *IEEE Trans. Electronic Computers*, Vol. EC-16, No. 3, pp. 299-307 (1967)
- [Anderson 84] Anderson, T. W.: *An Introduction to Multivariate Statistical Analysis*, Wiley (1984)
- [麻生 88] 麻生英樹: ニューラルネットワーク情報処理—コネクショニズム入門, あるいは柔らかな記号に向けて, 産業図書 (1988)
- [Baxter 95] Baxter, J.: Learning internal representations, *Proc. 8th Int. Conf. on Computational Learning Theory*, pp. 311-320 (1995)
- [Bishop 06] Bishop, C.: *Pattern Recognition and Machine Learning*, Springer-Verlag (2006) (元田 浩 ほか監訳: パターン認識と機械学習 (上・下), シュプリンガー・ジャパン (2008))
- [Bengio 07] Bengio, Y., Lamblin, P., Popovici, D. and Larochelle, H.: Greedy layer-wise training of deep networks, *Advances in Neural Information Processing Systems 19*, pp. 153-160, MIT Press (2007)
- [Bengio 09] Bengio, Y.: Learning deep architectures for AI, *Foundations and Trends in Machine Learning*, Vol. 2, No. 1, pp. 1-127 (2009)
- [Bengio 12] Bengio, Y., Courville, A. and Vincent, P.: Representation learning: A review and new perspectives, arXiv:1206.558v2 (2012)
- [Caruana 95] Caruana, R.: Learning many related tasks at the same time with backpropagation, *Advances in Neural Information Processing Systems 7*, pp. 656-664 (1995)
- [Cayton 05] Cayton, L.: Algorithms for manifold learning, *Technical Report CS2008-0923, UCSD* (2005)
- [Cortes 95] Cortes, C. and Vapnik, V. N.: Support-vector networks, *Machine Learning*, Vol. 20, pp. 273-297 (1995)
- [Cottrell 88] Cottrell, G. W. and Munro, P.: Principal component analysis of image via backpropagation, *Proc. SPIE 1001 Visual Communications and Image Proc.* '88, pp. 1070-1076 (1988)
- [Cybenko 89] Cybenko, G.: Approximation by superpositions of a sigmoidal function, *Mathematics of Control Signals and Systems*, Vol. 2, pp. 303-314 (1989)
- [Dahl 12] Dahl, G. E., Yu, D., Deng, L. and Acero, A.: Context-dependent pre-trained deep neural networks for large vocabulary speech recognition, *IEEE Trans. on Audio, Speech, and Language Processing* (2012)
- [DeMers 93] DeMers, D. and Cottrell, G. W.: Nonlinear dimensionality reduction, *Advances in Neural Information Processing Systems 5*, pp. 580-587, Morgan Kaufmann (1993)
- [電王戦 13] 第2回電王戦, <http://ex.nicovideo.jp/denousen2013/> (2013)
- [Donoho 06] Donoho, D.: Compressed sensing, *IEEE Trans. on Information Theory*, Vol. 52, No. 4, pp. 1289-1306 (2006)
- [Duda 00] Duda, R. O., Hart, P. E. and Stork, D. G.: *Pattern*

*3 同じことは, 1960年代との対比として, 1980年代にもいわれていた。

- Classification*, 2nd Edition, John Wiley & Sons (2001) (尾上 守夫 監訳: パターン識別, 新技術コミュニケーションズ (2001))
- [福島 79] 福島邦彦: 位置ずれに影響されないパターン認識機構の神経回路モデル—ネオコグニトロン—, 信学論 (A), Vol. J62-A, No. 10, pp. 658-665 (1979)
- [Fukushima 13] Fukushima, K.: Artificial vision by multi-layered neural networks: Neocognitron and its advances, *Neural Networks*, Vol. 37, pp. 103-119 (2013)
- [Funahashi 89] Funahashi, K.: On the approximate realization of continuous mappings by neural networks, *Neural Networks*, Vol. 2, No. 3, pp. 183-191 (1989)
- [Håstad 86] Håstad, J.: Almost optimal lower bounds for small depth circuits, *Proc. 18th Annual ACM Symposium on Theory of Computing*, pp. 6-20 (1986)
- [林原 90] 林原香織, 山下雅史, 阿江 忠: シグモイド関数の連続性/離散性とニューラルネットワークのマシン能力について, 信学論, Vol. 73-D-II, No. 8, pp. 1220-1226 (1990)
- [Hayek 52] Hayek, F.: *The Sensory Order: An Inquiry into the Foundation of Theoretical Psychology*, University of Chicago Press (1952)
- [Hebb 49] Hebb, D. O.: *The Organization of Behaviour: A Neurophysiological Theory*, Wiley (1949) (鹿取廣人ほか 訳: 行動の機構—脳メカニズムから心理学へ, 岩波書店 (2011))
- [Hinton 06] Hinton, G. E. and Salakhutdinov, R.: Reducing the dimensionality of data with neural networks, *Science*, Vol. 313, pp. 504-507 (2006)
- [保木 07] 保木邦仁: 将棋における局面評価の機械学習—探索結果の最適制御—, 第 10 回情報論的学習理論ワークショップ (IBIS2007) 予稿集, pp. 145-149 (2007)
- [入江 90] 入江文平, 川人光男: 多層パーセプトロンによる内部表現の獲得, 信学論, Vol. 73-D-II, No. 8, pp. 1173-1178 (1990)
- [石井 98] 石井健一郎, 上田修功, 前田英作, 村瀬 洋: わかりやすいパターン認識, オーム社 (1998)
- [石川 90] 石川真澄: 忘却を用いたコネクショニストモデルの構造学習アルゴリズム, 人工知能学会誌, Vol. 5, No. 5, pp. 595-603 (1990)
- [JSAI 12] 特集「記号創発ロボティクス」, 人工知能学会誌, Vol. 27, No. 6, pp. 544-625 (2012)
- [JSAI 13] 特別企画「シンギュラリティの時代: 人を越えゆく知性ととも」, 人工知能学会誌, Vol. 28, No. 3, pp. 424-471 (2013)
- [神鷹 10] 神鷹敏弘: 転移学習, 人工知能学会誌, Vol. 25, No. 4, pp. 572-580 (2010)
- [King 04] King, R. D., Whelan, K. E., Jones, F. M., Reiser, P. G. K., Bryant, C. H., Muggleton, S. H., Kell, D. B. and Oliver, S. G.: Functional genomic hypothesis generation and experimentation by a robot scientist, *Nature*, Vol. 427, pp. 247-251 (2004)
- [Krizhevsky 12] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, *NIPS 2012* (2012)
- [Le 12] Le, Q. V., Ranzato, M. A., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J. and Ng, A. Y.: Building high-level features using large scale unsupervised learning, *Proc. 29th Int. Conf. on Machine Learning* (2012)
- [LeCun 89] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard R. E. and Jackel, L. D.: Backpropagation applied to hand-written zip code recognition, *Neural Computation*, Vol. 1, No. 4, pp. 541-551 (1989)
- [Marr 70] Marr, D.: A theory for cerebral neocortex, *Proc. Royal Society of London, Series B*, Vol. 176, No. 1043, pp. 161-234 (1970)
- [Marr 82] Marr, D.: *Vision, A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman and Company (1982) (乾 敏郎ほか 訳: ビジョン—視覚の計算理論と脳内表現—, 産業図書 (1987))
- [McCulloch 43] McCulloch, W. and Pitts, W.: A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, Vol. 7, pp. 115-133 (1943)
- [Minsky 87] Minsky, M.: *Perceptrons: An Introduction to Computational Geometry*, Expanded Edition, MIT Press (1987) (中野 馨, 阪口 豊 訳: パーセプトロン, パーソナルメディア (1993))
- [中野 05] 中野良平: ニューラル情報処理の基礎数理, 数理工学社 (2005)
- [Olshausen 96] Olshausen, B. A. and Field, D. J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature*, Vol. 381, pp. 607-609 (1996)
- [Olshausen 97] Olshausen, B. A. and Field, D. J.: Sparse coding with an overcomplete basis set: A strategy employed by V1?, *Vision Research*, Vol. 37, pp. 3311-3325 (1997)
- [Rosenblatt 58] Rosenblatt, F.: The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, Vol. 65, No. 6, pp. 386-408 (1958)
- [Rumelhart 86] Rumelhart, D. E., Hinton, G. E. and Williams, R. J.: Learning representations by back-propagating errors, *Nature*, Vol. 323, No. 9, pp. 533-536 (1986)
- [Sejnowski 87] Sejnowski, T. J. and Rosenberg, C. R.: Parallel networks that learn to pronounce English text, *Complex Systems*, Vol. 1, pp. 145-168 (1987)
- [Siegelmann 91] Siegelmann, H. T. and Sontag, E. D.: Turing computability with neural nets, *Applied Mathematics Letters*, Vol. 4, pp. 77-80 (1991)
- [田中 10] 田中利幸: 圧縮センシングの数理, *IEICE Fundamentals Review*, Vol. 4, No. 1, pp. 39-47 (2010)
- [Vincent 10] Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.-A.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, *J. Machine Learning Research*, Vol. 11, pp. 3371-3408 (2010)
- [Watanabe 69] Watanabe, S.: *Knowing and Guessing: A Quantitative Study of Inference and Information*, John Wiley & Sons (1969) (村上陽一郎ほか 訳: 知識と推測: 科学的認識論, 東京図書 (1987))
- [Winston 70] Winston, P. H.: *Learning Structural Descriptions from Examples*, MIT AI Technical Reports MIT/LCS/TR-76 (1970)
- [Wolpert 96] Wolpert, D. H.: The lack of a priori distinctions between learning algorithms, *Neural Computation*, Vol. 8, No. 7, pp. 1341-1390 (1996)
- [安田 13] 安田宗樹: ディープボルツマンマシン入門—ボルツマンマシン学習の基礎—, 人工知能学会誌, Vol. 28, No. 3, pp. 474-485 (2013)

2013 年 5 月 14 日 受理

著者紹介



麻生 英樹 (正会員)

1981 年東京大学工学部計数工学科卒業。1983 年同大学院工学系研究科情報工学専攻修士課程修了。同年, 通商産業省工業技術院電子技術総合研究所(現独立行政法人産業技術総合研究所)入所。1993 ~ 94 年ドイツ国立情報処理研究センター客員研究員。現在, 独立行政法人産業技術総合研究所知能システム研究部門上級主任研究員。学習能力をもつ知的情報処理システムの研究に従事。電子情報通信学会, 日本神経回路学会などの各会員。