

SAPT研究発表会

AIおよびディ-プラ-ニングの現状

一色 浩

(有)数理解析研究所

SAPT研究発表会、2019.08.19

目 次

1. ディープラーニングの衝撃
2. ニューラル・ネットワークの歴史
3. ニューラル・ネットワークの基礎演算
4. 相互結合ニューラル・ネットワーク
5. ニューラル・ネットワーク研究のフロンティア
6. 結言

参考文献、ビデオ教材

1. ディープラーニングの衝撃

最近になって連日ニュースが話題になっている。

画像認識、音声認識、機械翻訳などの世界で大きなブレークスルーが生まれた。

自動運転の技術が夢から現実になった。

ディープラーニングのしたこと：

人間には簡単にできるのに機械にはできなかつたことが可能に。
画像と音声認識、機械翻訳などの超複雑関係にブレークスルー。

いかなるブレークスルーか？：

人間のような柔軟性を機械に与える端緒が開かれた。

今後、工学の全てで柔軟性がキーコンセプトになろう。

現在、ディープラーニングは多層の階層型ニューラル・ネットワークを指すことが多い。

時系列を取り扱うネットワークや、階層型でない相互結合型のネットワークなど、今後急速に発展しよう。

AIとは何かについて述べる。NN以外にも、決定木、クラスタリング、ベイズ分類器、サポートベクターマシン(SVM)、画像、文字・音声認識などいろいろある。

日本の現状:AIを始めとする**ITで周回遅れ**

日本社会の問題を一言で言うと**閉鎖性**

ディープラーニングの特徴を挙げると

NNは人間の脳の働きを模したもの。

階層NNは入力と出力の非線形な関数関係を求める。

多層NNはどんな問題(関数関係)も解ける。

入出力変数が画像でも音声でも良い。

多層NNを最小値問題で解くと局所解に捕まる。

ヒントンが解決し人間よりも優れた画像認識を示した。

その際、人の視覚を模した畳み込みNNを導入した。

ヒントンによるディープラーニング → 大爆発

自己符号化器はノイズ削減に有効。

天気予測等の時系列には過去記憶するRNNがある。

数学が嫌いでもいい。パソコンの能力が低くてもいい⁵。

2. ニューラル・ネットワークの歴史

1873年、ゴルジはゴルジ染色法で神経組織を観察し、網状説を提唱。後にカハーンはこれを否定し、神経細胞(ニューロン)説を唱える。

1943年、マカロックとピツは脳の論理処理の計算モデル(形式ニューロンと呼ばれる素子からなる回路)を考え出した。

1957年、ローゼンブラットは「問題の解を与えるように訓練する」という学習の概念を導入した。

1960年、WidrowとHoffがWidrow-Hoff法(デルタール)を発表した。後に誤差逆伝播法と呼ばれる。

2006年、ヒントンらの研究チームが、制限ボルツマンマシンによるオートエンコーダの深層化に成功し、階層型ネットワークにも適用され、2015年に図形認識で人間を超える。

1979年、福島邦彦はネオコグニトロンを発表した。ディープラーニングにつながる先駆的研究である（畳み込みNN）。自ら学習することによってパターン認識能力を獲得（概念の形成）していくという自己組織化機能がある。

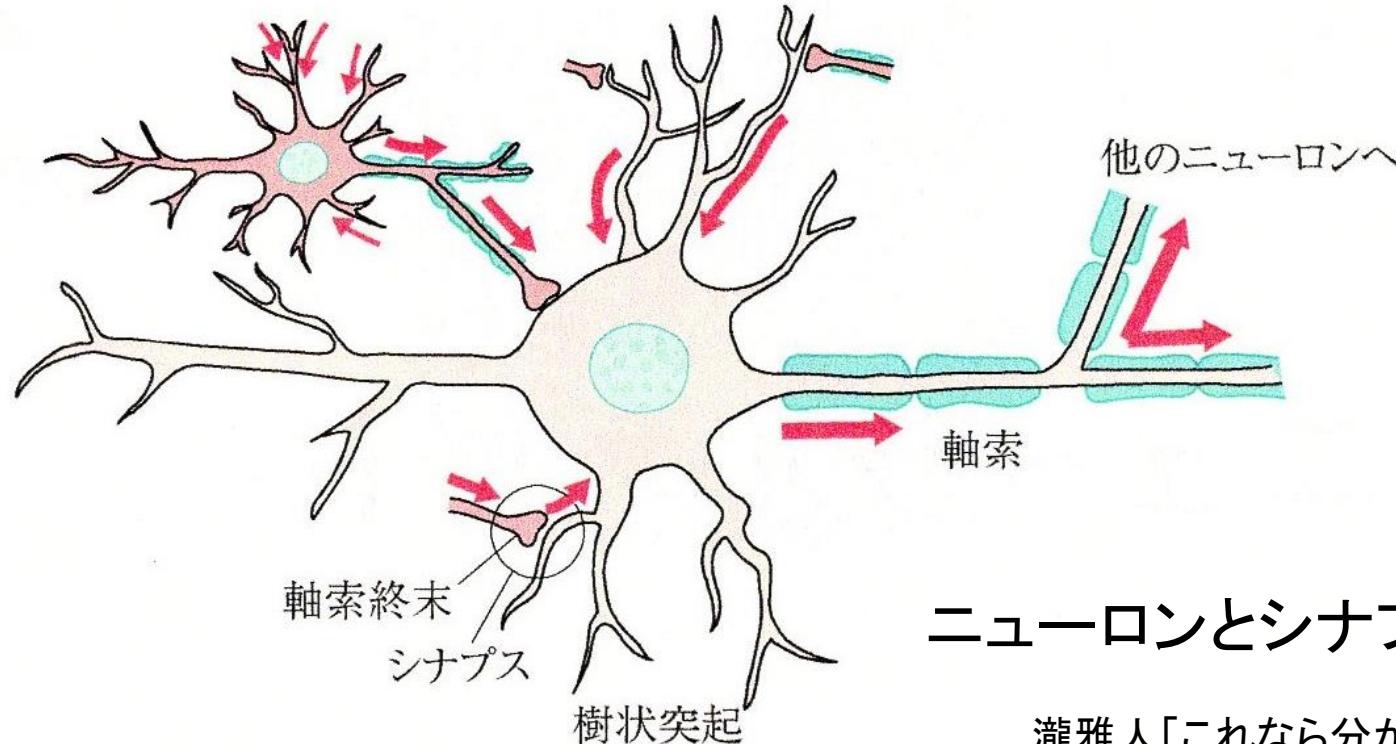
2.1 大脳の構造 ... ゴルジとカハーンの研究

ニューロンの形状を明らかにしたのはイタリア人のカミッロ・ゴルジで、1873年に銀とクロムを用いた細胞の染色法を編み出して神経細胞の観察に成功する。

スペインのサンティアゴ・ラモン・イ・カハールは、シナプスでは軸索終末と樹状突起は直接つながっていないで狭い隙間があることを主張した。ゴルジは継っていると考えた。光学顕微鏡では直接確認できないため見解が別れたが、結果的にはカハールが正しかった。

ニューロン同士が直接継っていない事には、信号処理上、深い意味がある。

人間の脳には、1000億個以上の下図のような神経細胞(ニューロン)がネットワークを形成している。



ニューロンとシナプス

瀧雅人「これなら分かる深層学習入門」図3.1

神経細胞の数は大脳で数百億個、小脳で1000億個、脳全体では千数百億個にもなる。

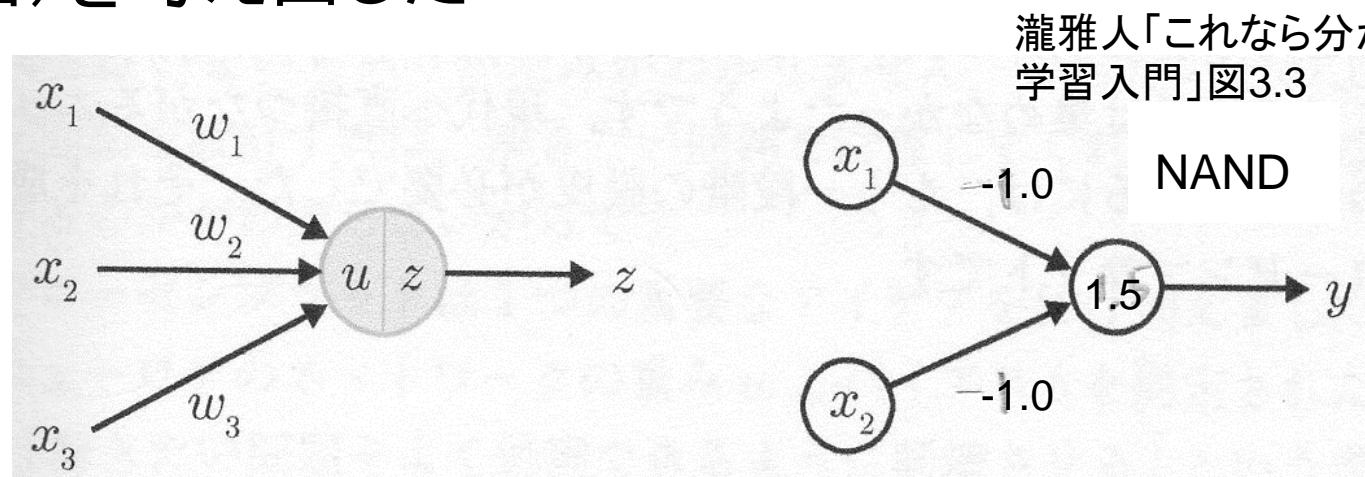
2.2 ニューラルネットワークの数理モデル

2.2.1 マカロックとピツのモデル

NNの起源は1943年のマカロックとピツの研究.

神経生理学者で外科医のマカロックは、チューリングの影響を受けて、思考の計算モデルを作ろうと考えた。

神童のピツを共同研究者として、脳の論理処理の計算モデル(形式ニューロンと呼ばれる素子からなる回路)を考え出した。



形式ニューロン(左)と簡単なニューロン回路の例(右)

形式ニューロンには他の多数の形式ニューロンから0か1の信号 x_i が入る. 重みを w_i として総入力 u を

$$u = \sum_i w_i x_i \quad (3.1)$$

とする. このニューロンの出力 z を

$$v = \theta(u + b) = h\left(\sum_i w_i x_i + b\right) \quad (3.3)$$

とする. ここで h はヘビサイドの階段関数:

$$h(u + b) = \begin{cases} 1 & (u \geq -b) \\ 0 & (u < -b) \end{cases} \quad (3.2)$$

入力 u を出力 v に変換する関数を **活性化関数** という.

図の右はNAND. **NN**でいかなる論理演算も可能.

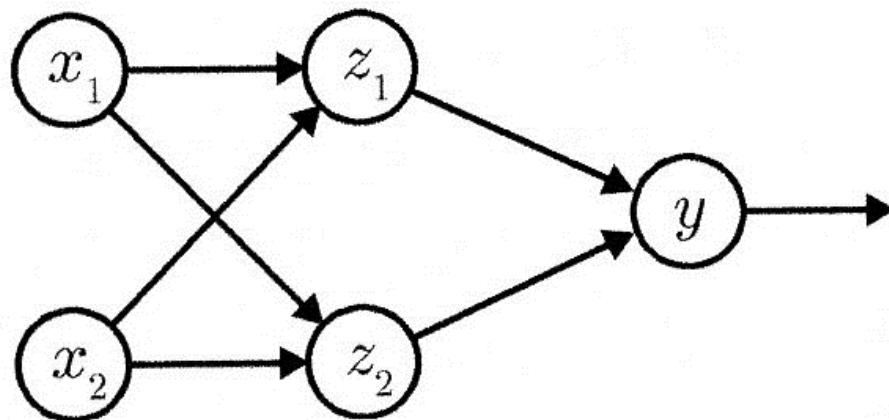
3.2.2 ローゼンブラットのパーセプトロン

前節ではパラメータ w や b は固定値であった。

ローゼンブラットはパラメータ w や b を「問題の解を与えるように訓練する」という**学習**の概念を導入した。

つまり、図のようなNNに**教師あり学習**をさせる。

ローゼンブラットはこのようなNNを**パーセプトロン**と呼んだ。

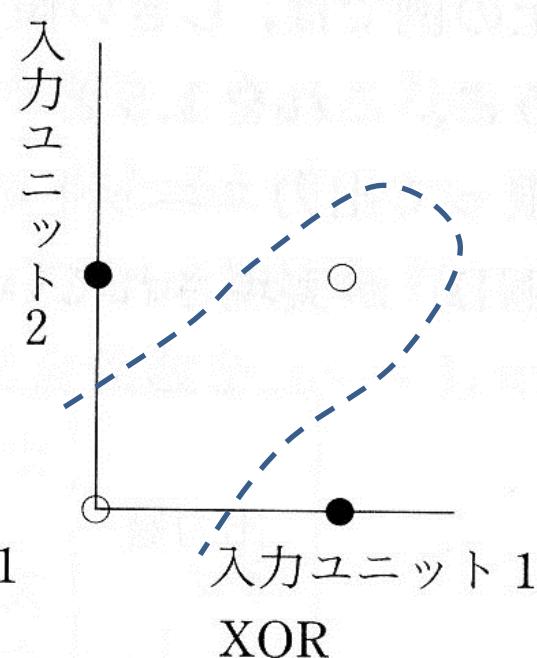
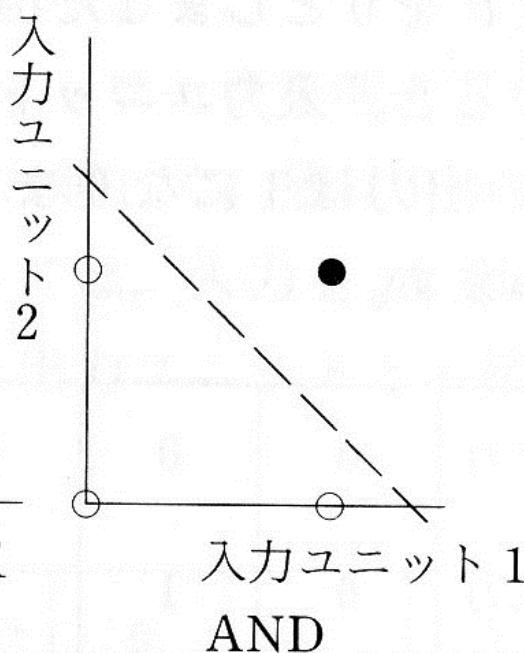
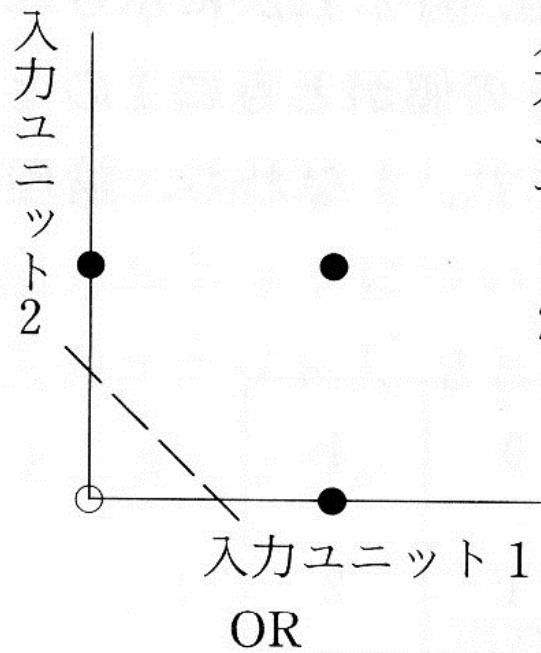


3層からなるパーセプトロンの例

人工知能と呼ばれる研究分野を立ち上げたミンスキー（ローゼンブラットの高校の同級生）は、入力と出力からなる2層パーセプトロンではXORのような線形分離不可能問題は解けないと示した。

これが一因となり60年代の終わりに第1次ニューラルネットブームが終焉した。

3層パーセプトロンならばXORを解ける。



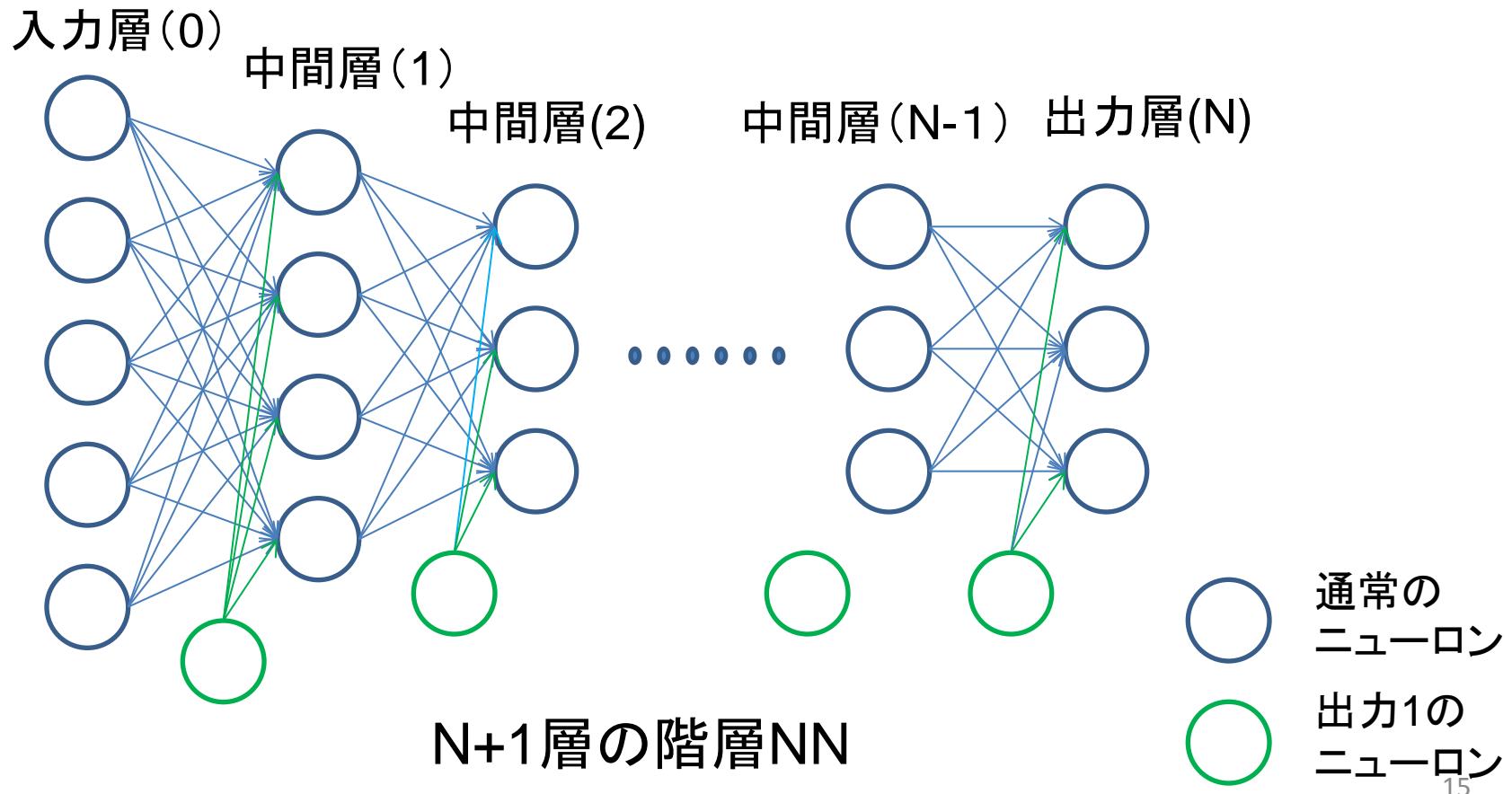
- 出力ユニットの値が 0
- 出力ユニットの値が 1

線型分離できる例(OR, AND)とできない例(XOR)

3. ニューラル・ネットワークの基礎演算

3.1. 誤差逆伝播学習 (NNの理論ではこれが一番難しい)

誤差逆伝播学習： 誤差（教師データと出力データの差）
をゼロにする学習をする



$n+1$ 層の j ユニットへの入力:

$$u_j^{(n+1)} = \sum_i w_{ji}^{(n)} v_i^{(n)} + b^{(n)}, \quad n = 0, 1, \dots, N-1$$

$n+1$ 層の j ユニットの出力:

$$v_j^{(n+1)} = h(u_j^{(n+1)}) = h\left(\sum_i w_{ji}^{(n)} v_i^{(n)} + b^{(n)}\right)$$

誤差関数: **これを最小にする**

$$E = \frac{1}{2} \sum_j (t_j - v_j^{(N)})^2$$

$$\frac{\partial E}{\partial w_{ji}^{(N-1)}} = -(t_j - v_j^{(N)}) \frac{\partial v_j^{(N)}}{\partial w_{ji}^{(N-1)}} \quad \frac{\partial v_j^{(N)}}{\partial w_{ji}^{(N-1)}} = h'(u_j^{(N)}) v_i^{(N-1)}$$

したがって

$$\frac{\partial E}{\partial w_{ji}^{(N-1)}} = -(t_j^{(N)} - v_h^{(N)}) h'(u_j^{(N)}) v_i^{(N-1)}$$

$$\frac{\partial E}{\partial b^{(N-1)}} = -(t_j^{(N)} - v_h^{(N)}) h'(u_j^{(N)}) \cdot 1$$

誤差の定義
と計算法：

$$\delta_j^{(N)} = -\frac{\partial E}{\partial u_j^{(N)}} = -\frac{\partial E}{\partial v_j^{(N)}} \frac{\partial v_j^{(N)}}{\partial u_j^{(N)}}$$
$$= (t_j^{(N)} - v_j^{(N)}) h'(u_j^{(N)})$$

$$\delta_j^{(n+1)} = -\frac{\partial E}{\partial u_j^{(n+1)}} = -\frac{\partial E}{\partial v_j^{(n+1)}} \frac{\partial v_j^{(n+1)}}{\partial u_j^{(n+1)}}$$
$$= -h'(u_j^{(n+1)}) \sum_k \frac{\partial E}{\partial u_k^{(n+2)}} \frac{\partial u_k^{(n+2)}}{\partial v_j^{(n+1)}}$$

$$= h'(u_j^{(n+1)}) \sum_k \delta_k^{(n+2)} w_{kj}^{(n+2)}$$

誤差 $\delta_j^{(n)}$ は上の階層から決まるこれが**誤差逆伝搬法**の名前の由来.

勾配の計算法:

$$\frac{\partial E}{\partial w_{ji}^{(n)}} = \frac{\partial E}{\partial u_j^{(n+1)}} \frac{\partial u_j^{(n+1)}}{\partial w_{ji}^{(n)}} = -\delta_j^{(n+1)} v_i^{(n)}$$

$$\frac{\partial E}{\partial b^{(n)}} = \frac{\partial E}{\partial u_j^{(n+1)}} \frac{\partial u_j^{(n+1)}}{\partial b^{(n)}} = -\delta_j^{(n+1)} \cdot 1$$

最急降下法によるウェイト w_i とオフセット b の修正量:

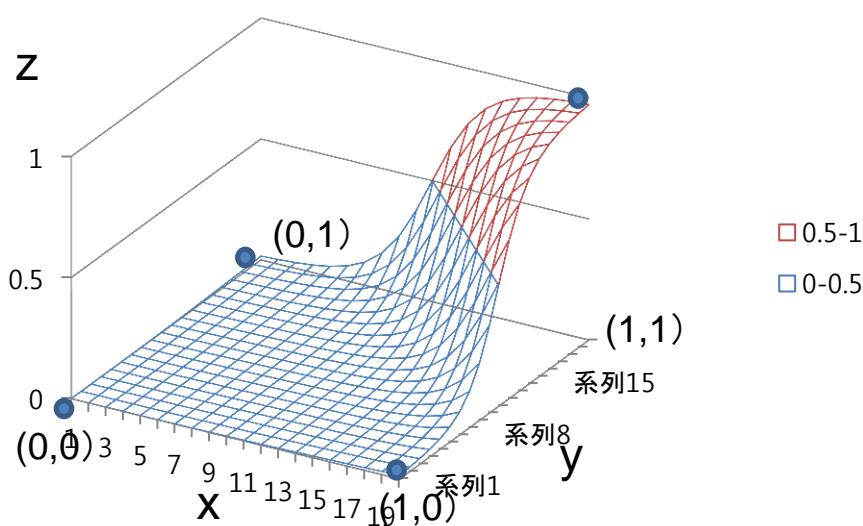
$$\Delta w_{ji}^{(n)} = -\frac{\partial E}{\partial w_{ji}^{(n)}} \alpha \quad \Delta b^{(n)} = -\frac{\partial E}{\partial b^{(n)}} \beta$$

3層NNによるAND、OR、XORの学習

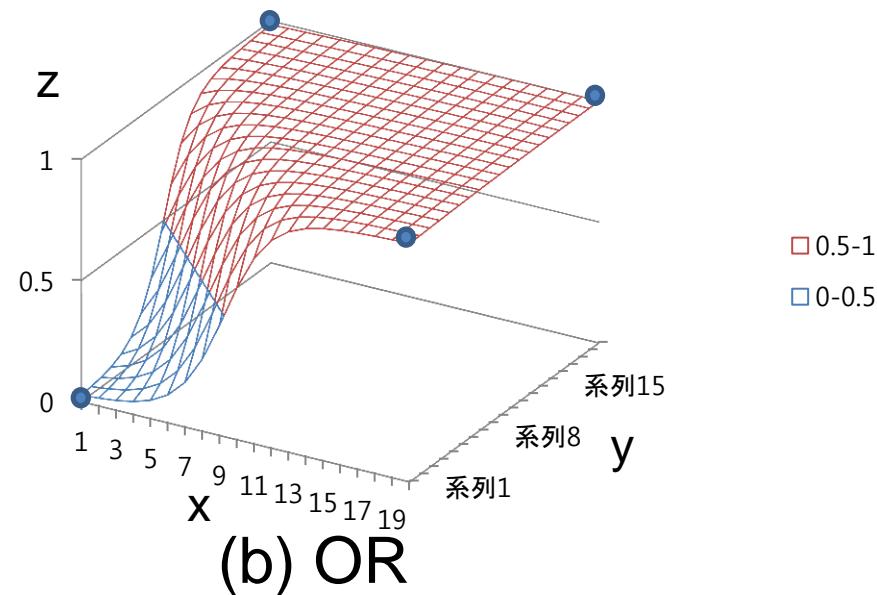
出力結果

入力値		出力値		
ユニット0	ユニット1	ユニット0	ユニット1	ユニット2
		AND	OR	XOR
0	0	0	0.02	0.04
1	0	0.02	0.98	0.97
0	1	0.02	0.98	0.97
1	1	0.98	1	0.04

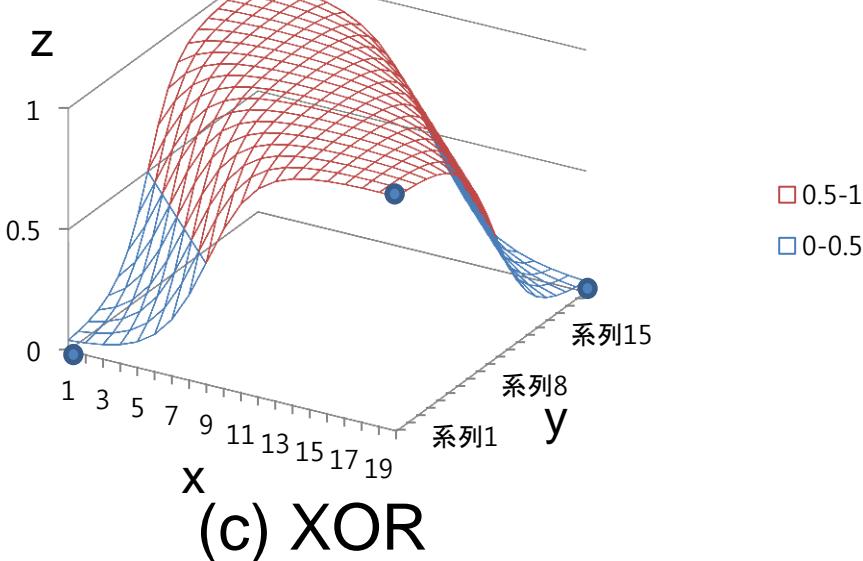
3層NNによるAND、OR、XORの学習



(a) AND



(b) OR



(c) XOR

AND, OR, XOR で
できる曲面

3.2. 新しいニューラルネットワーク

最近NNが大きな話題

ブレークスルーのキーワードは**深層学習**

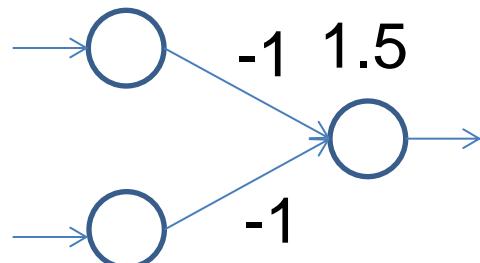
人工知能の中核技術になる勢い

これまでのNN

汎化能力(未学習のデータに正解する能力)が
非常に低い

NNはいかなる問題も解けることが数学的に保証

NNはNAND回路を作れる。NAND回路を用いればいかなる論理回路も構成できる。

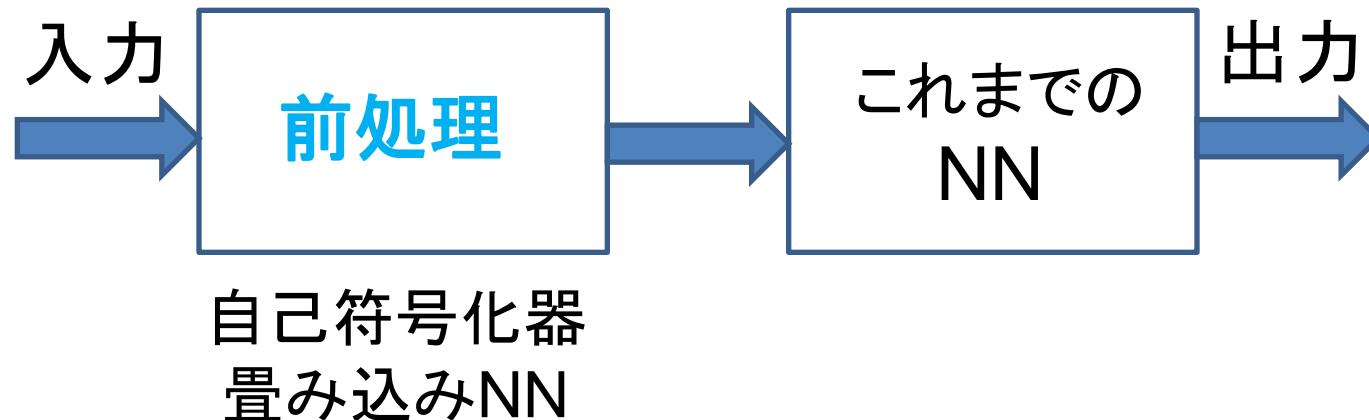


NAND回路

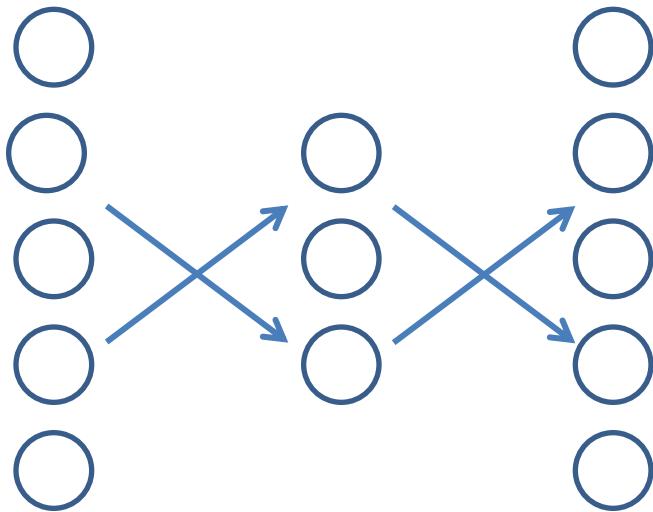
深層とは...層数が増えたということ

層数が増えても誤差関数を最小にする(グローバル・ミニマムに到達させる)方法が開発された。

新しいアーキテクチャの導入：



3.3. 自己符号化器(Autoencoder)



ニューロン数:
入力層 = 出力層 > 中間層

学習:
教師データ = 入力データ
中間層に圧縮されたデータ

自己符号化器の利用

ノイズ除去:

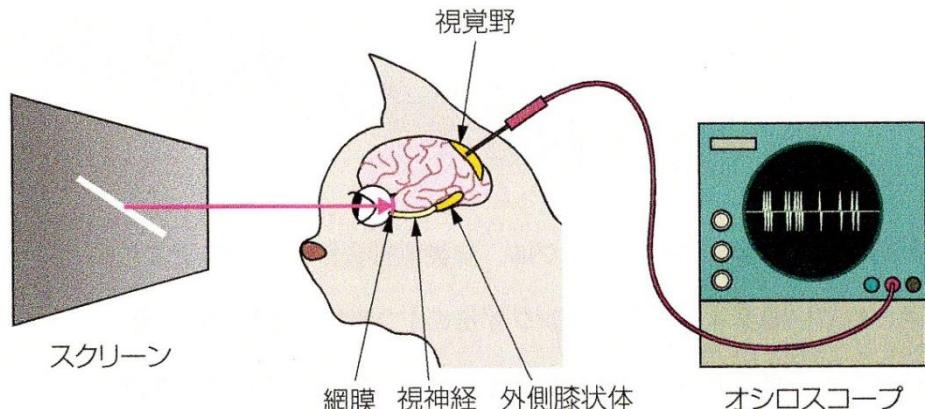
- (1) 正常データで学習させた自己符号化器にノイズありデータを入力するとノイズが落ちる。
- (2) 入力データにノイズありデータ, 教師データにノイズなしデータを与えて学習させると, ノイズ除去フィルタができる。

故障, 異常診断:

正常データで学習させた自己符号化器にノイズのレベルを超えた異常のあるデータを見せると, 異常データを出力するので, 正常な入力データと出力データの差を取ると, 異常が検知される.

3.4. 署み込みNN(Convolution NN)

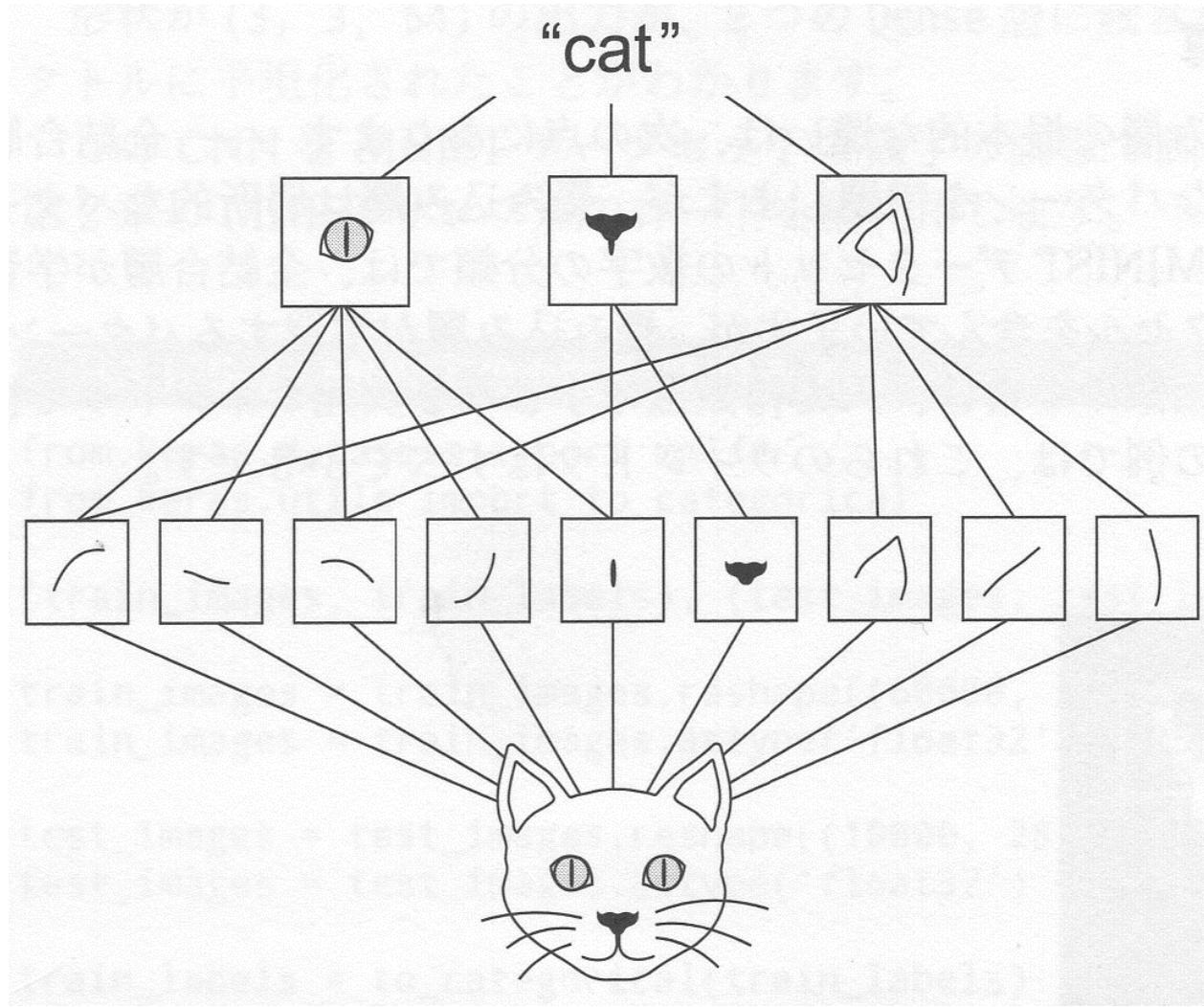
ヒューベルとウィーゼルの動物視覚野の階層仮説



ヒューベルらの実験
斜線に反応する細胞
山下、ディープラーニング、
図3.1

単純型細胞と複雑型細胞が階層構造をつくる。
単純型細胞は特徴を捉える(コンボリューション)。
複雑型細胞は特徴の大きさや位置のズレを吸収する
(プーリング)。

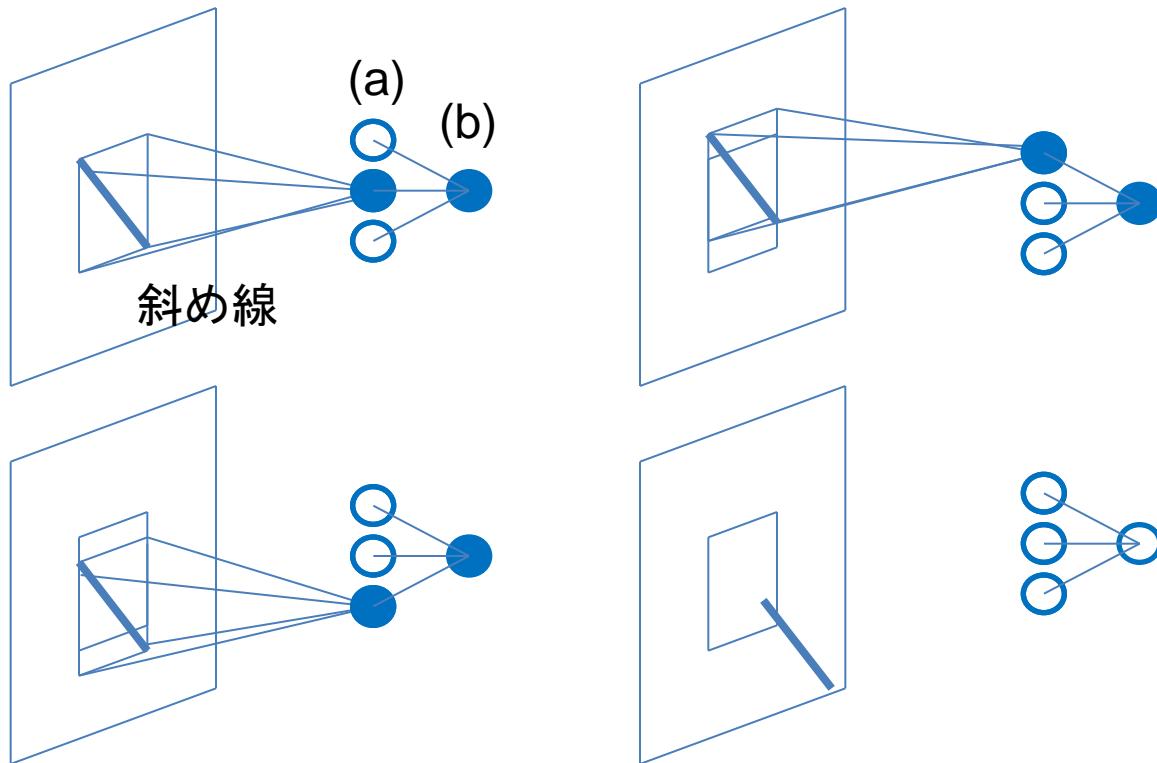
情報は階層的に処理する。
まず画像の細かい特徴を捉える。
上の階層に送られて情報が統合される。



微細特徴→部品→全体 に統合される様子

F. Chollet著、(株)クイーブ訳、巣籠悠輔監訳、「PythonとKerasによるディープラーニング」、p136、図5-8

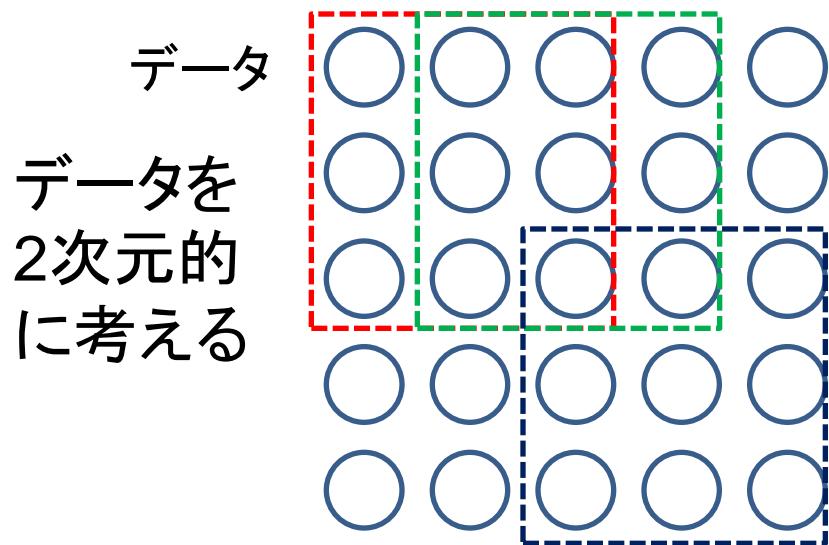
単純型細胞(a)では、パターンが小さな枠で囲まれた領域にぴったり現れた時にのみニューロンが発火する。右下以外の図では発火するが右下図では発火しない。複雑型細胞(b)があればパターンがずれた時でも発火する。



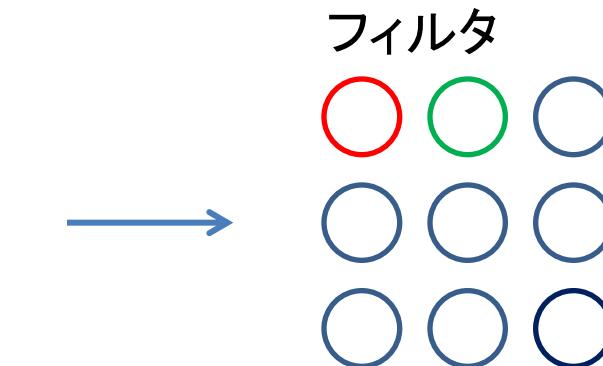
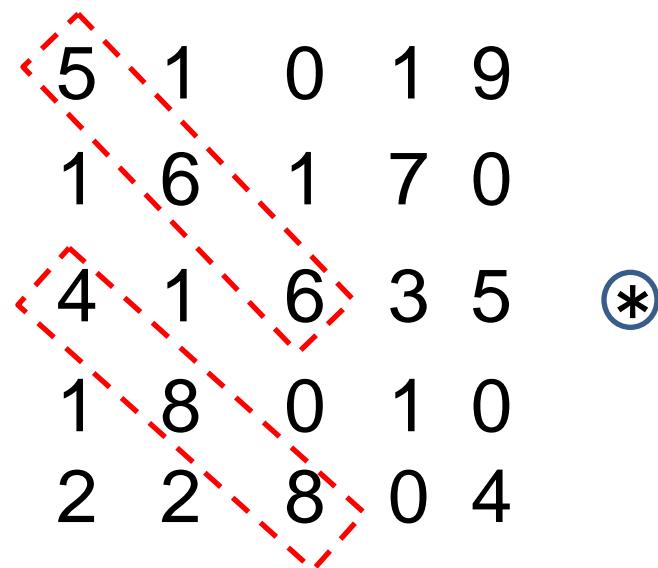
「特徴の抽出をする単純型細胞」と「ズレの吸収をする複雑型細胞」

コンボリューション演算

瀧雅人「これなら分かる深層
学習入門」図8.4



斜め線の検出



$$\begin{aligned}
 26 = & 5 \times 2 + 1 \times 0 + 0 \times (-2) \\
 & + 1 \times 0 + 6 \times 2 + 1 \times 0 \\
 & + 4 \times (-2) + 1 \times 0 + 6 \times 2
 \end{aligned}$$

$$\begin{array}{ccc}
 2 & 0 & -2 & 26 & 6 & -6 \\
 0 & 2 & 0 & 0 & -4 & 8 \\
 -2 & 0 & 2 & -2 & -14 & -4
 \end{array}
 = \begin{array}{ccc}
 26 & 6 & -6 \\
 0 & -4 & 8 \\
 -2 & -14 & -4
 \end{array}$$

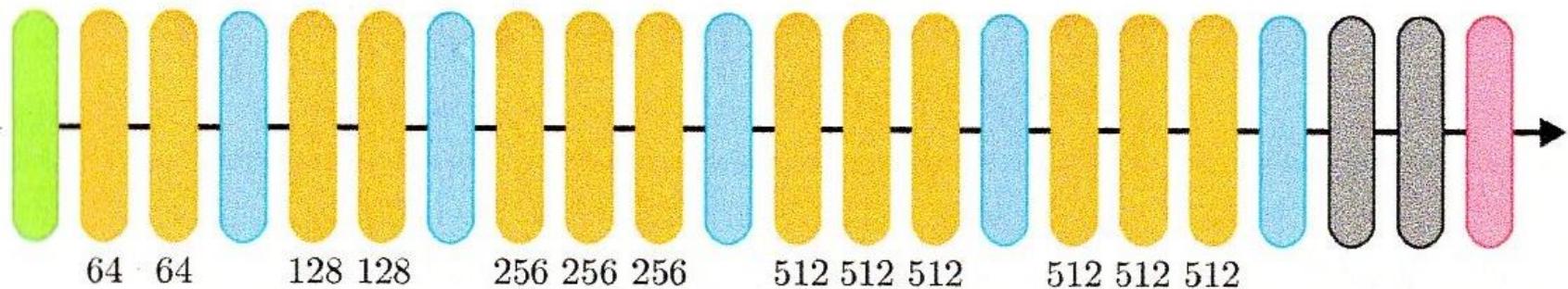
コンボリューションの重要性：
特長が学習により捉えられるよう
になった。

コンボリューションの特長：
重み行列はスパースであるので
過学習に陥りにくい。

プーリング層の役割は畳み込み層の出力を束ねるだけだから重み学習はしない。学習中もプーリングの式は変更されない。

ネットワーク構造

典型的なCNNの例を図に示す(オックスフォードのVGG)。



VGGの一例。オレンジ色は畳み込み層、青はマックスプーリング層。灰色は全結合層、そして赤は出力のソフトマックス層。数字は各畳み込み層のフィルタの幅(上の層ほどフィルタ幅が大きい...特徴の統合)。

プーリング層

複雑型細胞に対するプーリング層 (pooling layer) の説明をする。

プーリング層の役割は単純型細胞の出力を合算して入力位置の平行移動に対してロバストにすること。

最大プーリングでは、領域中の最大の画素値を代表値とする。最大プーリング以外に、平均プーリングなどがある。

3.5. 学習済みモデルと転移学習

多層のCNNを用いた自然画像の学習には多くのコストと計算時間をする。

もしも、学習済のモデルが他の学習に役立たないならば、コストが高すぎて、魅力が半減する。

しかし、ニューラルネットには高い汎化性能があり、深いネットワーク構造では、中間層にさまざまな特徴量を獲得していると期待される。

そこで、学習済のモデルの出力側の層を取り除いて、中間層からの出力を入力画像の表現として用いることが考えられる。

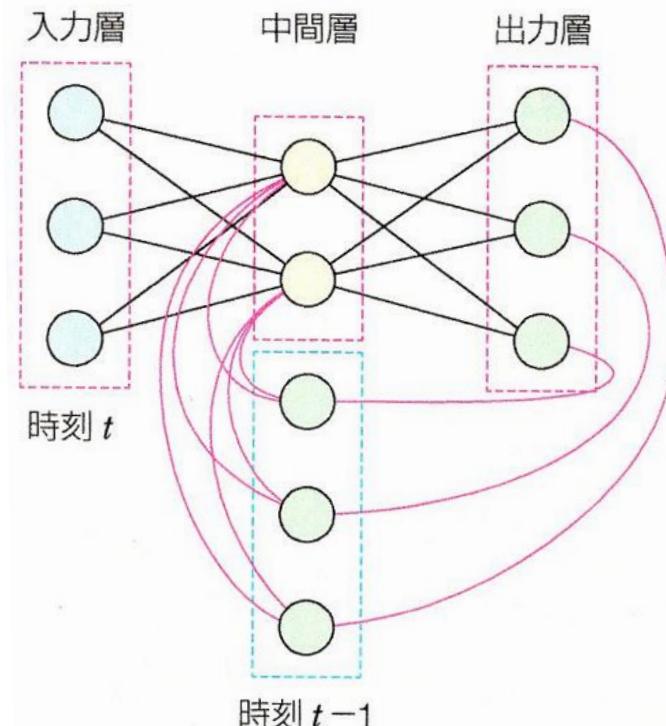
このようにしても、学習済みパラメータが初期値となって、学習がスムースに進行する。これを転移学習という。

3.6. 時系列を処理するNN

3.6.1. リカレントNN(RNN)

いろんな種類が提案
されている。

過去のデータを反映
するために、前時刻
の出力を中間層に結
合する。更に中間層
のデータを後ろへシフ
トする。中間層の
ニューロン数が増す
ので、計算困難になり
易い。

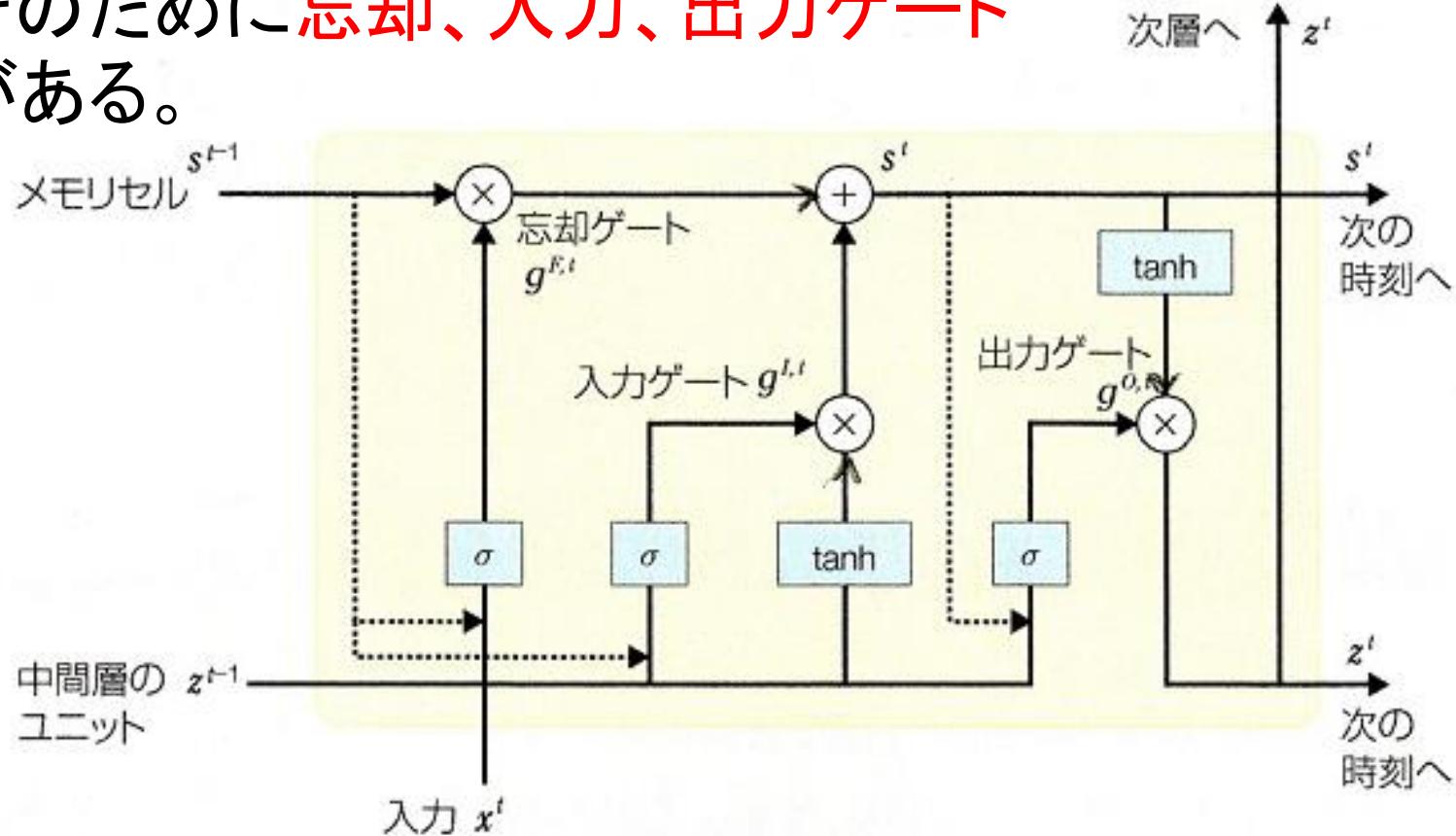


Jordan Network
山下、ディープラーニング、
図6.2

3.6.2 長短期記憶NN(LSTM)

関連の少ない記憶を減らし効率的に長期の記憶を保持する。

そのために忘却、入力、出力ゲートがある。



長短期記憶(LSTM)NN

山下、ディープラーニング、
図6.7

4. 相互結合ネットワーク

4.1. ヘブ学習

神経回路の学習には教師なしと教師ありがある。

教師なし学習: • ある刺激に対して、どの神経細胞が反応すべきかは明示されない。

• 幼児は自然に言葉を覚える。

• ある刺激に対して他より強い反応を示す細胞があり、繰り返し刺激を受けると、その細胞だけが反応するようになる。

教師あり学習: • 特定の出力(教師データ)をするように指導される学習をする。

• 文字や計算は教えられて覚える。

• NNではアソシアトロンとかBPがある。

アソシアトロン: 覚えたパターンの一部を示されても、全体を再現できる(中野)。

BP : 誤差逆伝播学習階層型NN(ラメルハートがBPと命名)

学習規則：ヘブの学習規則、デルタルール

ヘブの学習規則

ある神経細胞が興奮すると、その細胞に信号を送っている細胞のシナプス結合が強化される。パターン p を見せられた時の学習の数理モデルは

$$\Delta_p w_{ji} = v_j(p) \cdot v_i(p)$$

この学習では、学習は相関 $v_j(p) \cdot v_i(p)$ のみに関連する。
デルタルール

教師信号と出力信号の差を最小にする。パターン p を見せられた時の学習の数理モデルは

$$E(p, \mathbf{w}, \boldsymbol{\theta}) = \frac{1}{2} (t_j(p) - v_j(p))^2 \rightarrow \min$$

$$\Delta_p w_{ji} = -\eta \delta(p)_j$$

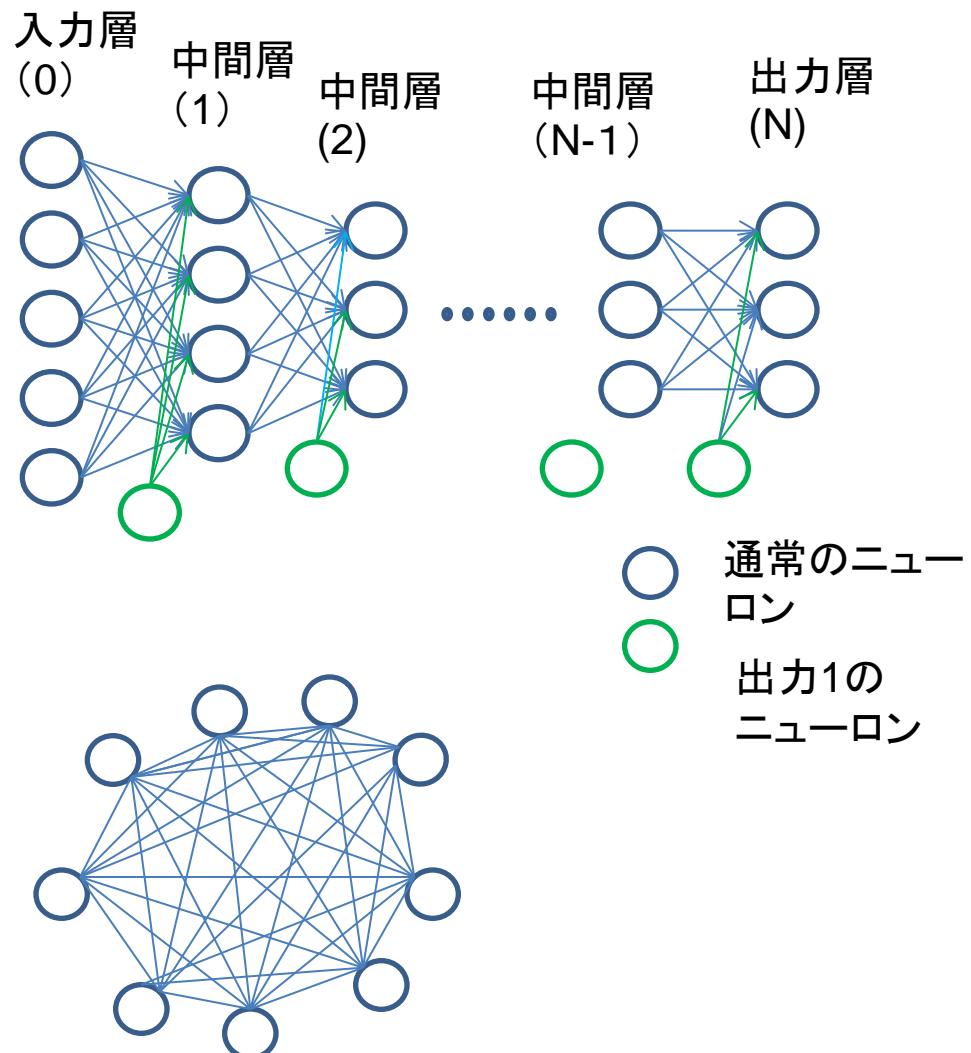
$$\delta(p)_j = t_j - v_j$$

4.2. 相互結合ニューラル・ネットワークとは

階層型NN

相互結合型NN

図のように全部が相互につながっているものを全結合型と呼ぶが、部分結合でも良い。



4.3. ホップフィールド・マシン

相互結合型のネットワークは非線形の回路網なので、ある初期状態から出発させると、いくつかの平衡状態のどれかに収束する。

どの平衡状態に収束するかは、初期状態による。

物理学者のホップフィールドは、ある種の最適化問題(巡回セールスマン問題など)が相互結合型ネットワークで見事に解けることを示し、ラメルハートに命名されたBPとともに前回のニューロブームの契機となつた。

相互結合型ネットワークに、物理学における物質の磁気的性質のスピニンモデルを適用したものである。

巡回セールスマン問題：

「一人のセールスマンが、 N ヶ所の都市をそれぞれ一度だけ訪問して、再び出発点に戻る経路の中で最小にするものを求めよ」…都市数 N の増加とともに組み合わせ数が指数関数的に増加する…組み合わせ最適化問題

強磁性体の統計力学：

「強磁性体の各状態のエネルギー関数を定義し、それにより各状態の確率を求め、確率最大(エネルギー最小)の状態が平衡状態と考える」

$$E(\mathbf{S}) := -h \sum_i S_i - J \sum \sum_{i \neq j} S_i S_j$$

$$P(\mathbf{S}) := \frac{1}{Z} \exp(-\beta E(\mathbf{S})), \quad Z = \sum_{S_1=\pm 1} \sum_{S_2=\pm 1} \cdots \sum_{S_N=\pm 1} \exp(-\beta E(\mathbf{S}))$$

$$\beta = 1/(k_B T)$$

巡回セールスマン問題(Traveling Salesman Problem)

	都市1	都市2	都市3	都市4
順番1	0	0	1	0
順番2	2	0	0	0
順番3	0	3	0	0
順番4	0	0	0	4
戻る	0	0	1	0

コスト関数(QUBO形式)

$$L = \frac{1}{2} \sum_{a,i,j} d_{i,j} n_{a,i} (n_{a+1,j} + n_{a-1,j})$$

$d_{i,j}$: 距離、 $n_{a,i}$: 各マス目のQUBO変数、 a : 順番、 i : 場所

制約条件
同じ列に1はひとつ: $\sum_a n_{a,i} = 1$

同じ行に1はひとつ: $\sum_i n_{a,i} = 1$

最終的なコスト関数:

$$L_1 = \frac{1}{2} \sum_{a,i,j} d_{i,j} n_{a,i} (n_{a+1,j} + n_{a-1,j}) + k_1 \sum_i (\sum_a n_{a,i} - 1)^2 + k_2 \sum_a (\sum_i n_{a,i} - 1)^2$$

(1) ユニット i から j への結合係数 w_{ij} と、ユニット j から i への結合係数 w_{ji} は等しい。

$$w_{ij} = w_{ji} \quad (1)$$

(2) ネットワークの状態遷移は非同期的に行われ、時刻 t にランダムに選ばれたユニット j が

$$v_j(t+1) = \phi(u_j(t)) \quad (2a)^*$$

$$u_j(t) = \sum \sum_{j \neq i} w_{ji} v_i(t) + \theta_j \quad (2b)$$

で遷移し、ユニット i は

$$v_i(t+1) = v_j(t) \quad (3)**$$

で遷移する。 $\phi(u)$ はユニットの入出力関数で

$$\phi(u) = \begin{cases} 1 & \text{when } u > 0 \\ 0 & \text{when } u < 0 \end{cases} \quad (4)$$

スピニンモデルとのアナロジから、ネットワークのエネルギー $E(t)$ を

$$E(t) = -\frac{1}{2} \sum \sum_{j \neq i} w_{ji} v_j(t) v_i(t) - \sum_j \theta_j v_j(t) \quad (5)$$

と定義する。時刻 t において、上の(2)で述べた状態遷移が起こるものと仮定すると、変化するのは $v_j(t)$ のみであり

$$E(t+1) \leq E(t) \quad (7)$$

が導けるので、 $v_j(t)$ は $E(t)$ の極小値を与える状態に収束する。

4.4. ボルツマン・マシン

山下りの過程に、一定の確率で山登りを取り入れる。

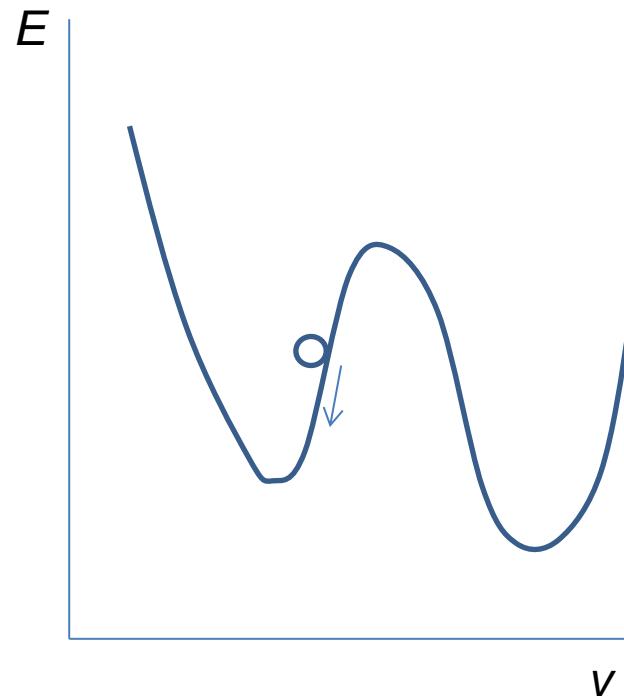


図1a ホップフィールド・マシン
(常に斜面を下る)

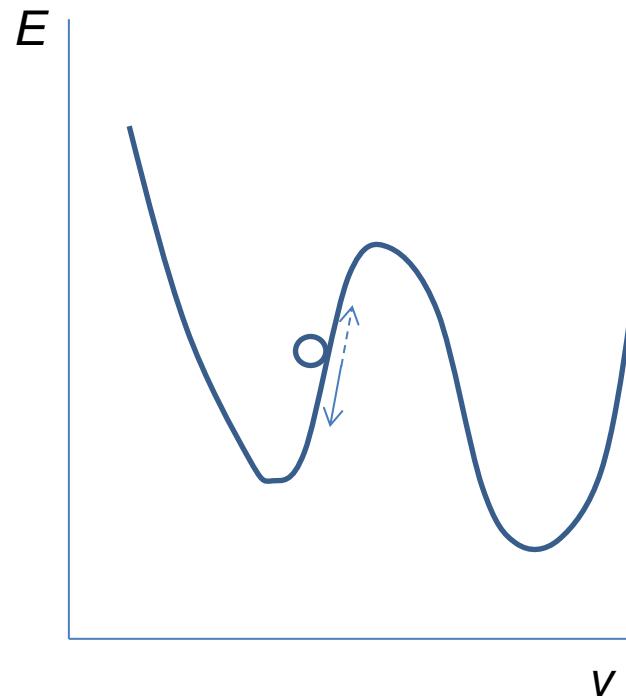


図1b ボルツマン・マシン
(斜面を上がったり下がったりする)

一定温度でボルツマン・マシンを動かし続けると、その温度での平衡状態になる。ネットワークの各状態の出現確率は、その状態のエネルギーの関数になる。平衡状態で得られる状態 V_a の確率は、その状態のエネルギーを E_a とすると、ボルツマン分布で与えられる。

$$P^-(E_a) = \frac{1}{Z} \exp(-E_a(t)/T) \quad (9a)$$

$$Z = \sum_a \exp(-E_a(t)/T) \quad (9b)$$

低温では、最小エネルギーの状態の出現確率がほぼ1となり、他の状態の出現確率はほぼ0となる。すなわち、ボルツマン・マシンは最小値検出器となる。

しかし、温度が低いと、斜面を登るような状態遷移が起こり難くなる。

そこで、高温から初めて徐々に温度を下げて行って、なるべく早く平衡状態に達するようにする。これを焼き鈍しと呼ぶ。

上ではボルツマン・マシンの動作を述べた。ボルツマン・マシンの学習とは、「平衡状態において、ある状態が可視ユニットに出現する確率が、与えられた確率となるように、ネットワークの結合係数およびオフセットを決定すること」である。

5. ニューラルネットワーク研究のフロンティア

5.1. 事象抽出と性能分析(燃料電池劣化)

K. Fukui, S. Akasaki, K. Sato, J. Mizusaki, K. Moriyama, S. Kurihara, M. Numao, Visualization of Damage Progress in Solid Oxide Fuel Cells, *Journal of Environment and Engineering*, Vol.6, No.3, 2011

固体酸化物燃料電池(SOFC)は高効率低環境汚損の発電システムであるが、実用化の最大の問題は性能劣化の推定。

SOFCは使用中にAcoustic Emission (AE)を起こす。これをモニターし評価することで劣化の進展を把握できると考えられる。

イベント切り出しで取り出したAEの周波数スペクトルを求め、NNの一種であるKernel Self-Organizing Map (Kernel SOM)で分類し、AEイベントと損傷進行の4ステージとの関連付けに成功した。

Kernel SOMはユークリッド距離ではなく、Kullback-Leibler (KL) divergenceを用いる。その結果、スペクトルの違いを分類に反映できる。

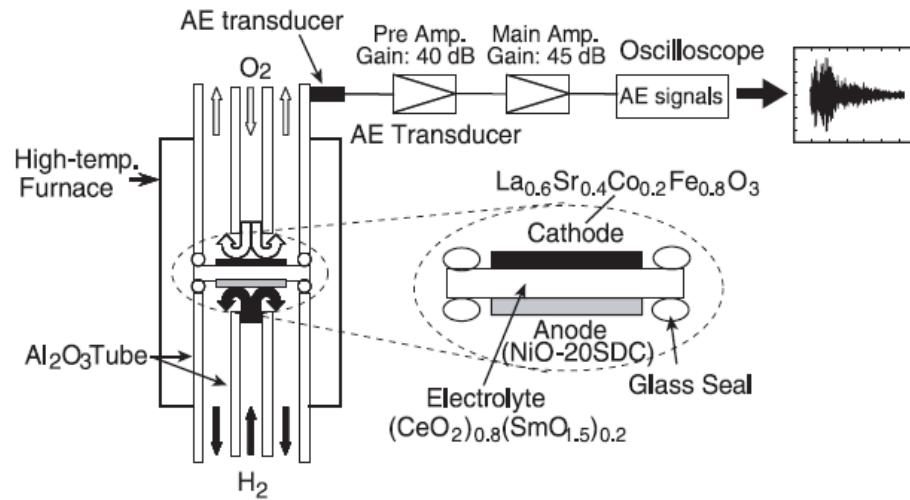


Fig. 3 SOFC damage test and AE measurement apparatus.

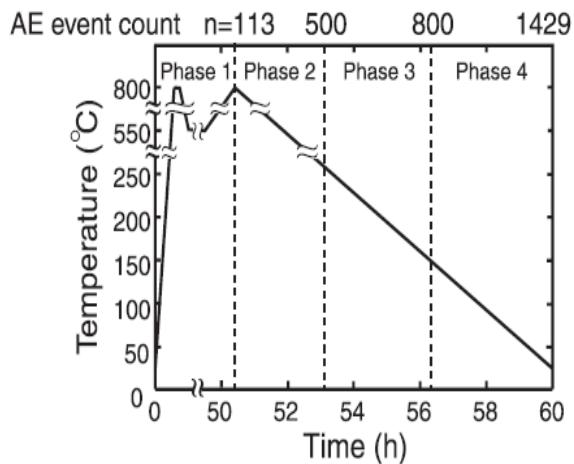


Fig. 4 Correspondence between AE occurrence process phase and temperature history.

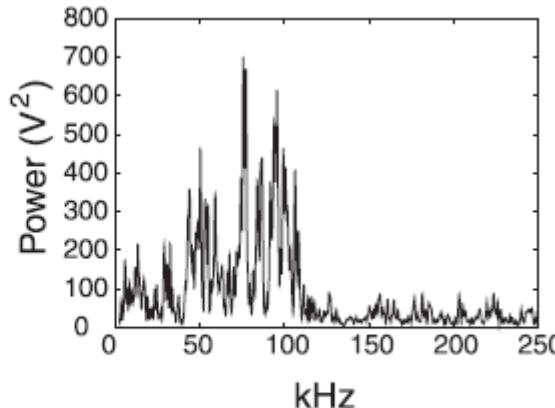
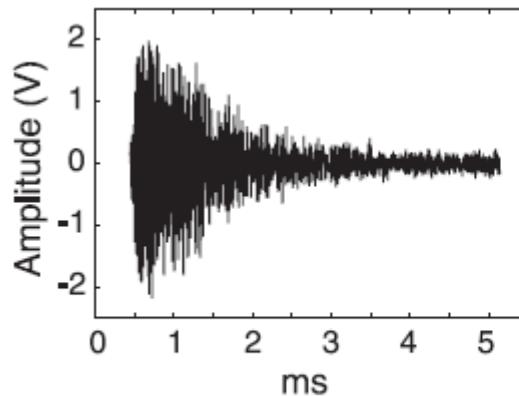


Fig. 8 An example of an extracted AE wave and frequency spectrum.

5.2. 変分自己符号化器

Variational Autoencoder徹底解説

<https://qiita.com/kenmatsu4/items/b029d697e9995d93aa24>

類似画像(変分画像)の作成をする。

自己符号化器の入力画像に $I_S = \{I_{S_1}, I_{S_2}, \dots, I_{S_M}\}$ を、教師画像に $T_S = \{I_{S_1} + \Delta I_{S_1}, I_{S_2} + \Delta I_{S_2}, \dots, I_{S_M} + \Delta I_{S_M}\}$ を与えて学習する。

学習済の自己符号化器の入力画像に $I_T = \{I_{T_1}, I_{T_2}, \dots, I_{T_N}\}$ を、出力画像として $T_T = \{I_{T_1} + \Delta I_{T_1}, I_{T_2} + \Delta I_{T_2}, \dots, I_{T_N} + \Delta I_{T_N}\}$ という変分画像を出力する。

例えば、30歳の日本人男性、 $M+N$ 人のうちの M 人の30歳時の顔画像と50歳児の顔画像の学習から、残りの N 人の男性の50歳時の顔画像を推定する。

この考え方を以下のように発展させることが可能である。

自己符号化器の中間層の出力 $Z_S = \{Z_{S_1}, Z_{S_2}, \dots, Z_{S_M}\}$ から
入力ニューロンに対する中間層ニューロン出力の事前確率
を推定できる。また、中間層ニューロン出力に対する自己符
号化器出力ニューロンの事前確率を推定できる。

例えば正規分布を仮定すれば、これらの事前確率をそれ
ぞれの平均と分散というパラメータで与えられる。

試験画像を与える代わりに、これらのパラメータに適当に変
分を付与すると、新しい画像が得られる。

6. 結言

ニューラル・ネットワーク(NN)の現状を概観した。

NNは、人間が容易にできるのに機械ではなかなかできないような超複雑な関係の解明にブレークスルーを実現した。

そのようなシステムは数式表現も可能であるが、ネットワーク表現が向いている。

このようなネットワークは一種の数学的補間とみなせるが、確率マシンと見る方が本質的であろう。

ネットワークの係数は単なる数と見るよりは、変数間の確率的関係に変換できる統計量と見ることができる。

これにより機械は人間のような柔軟性を獲得した。

ディープラーニングは一種のブラックボックスであるが、統計学に基づく「ベイズ推論による機械学習」では、過学習がない上に、データ量を増やすと精度向上が保証される。

須山敦志、ベイズ推論による機械学習入門、講談社機械学習スタートアップシリーズ(2017)

ニューラルネットワークと人間の脳との関係については、ニューラルネットワークが生物学的な意味で人間の脳の近似であると断言できない。

例えば、人間の脳は計算に弱い…掛け算、割り算のレベルは暗算できない。人間の脳は非常に簡単で、創発現象が人間の脳の働きを説明できる？

池谷裕二、単純な脳、複雑な「私」、講談社ブルーバックス(2013)

参考文献

ディープラーニング以外の機械学習を知りたい人には

- [1] 福井謙一、Pythonと実例で学ぶ機械学習 識別・予測・異常検知、オ-ム社(2018)
- [2] 中川祐志、情報工学 機械学習、東京大学工学教程編纂委員会、丸善(2015)

入門者には

- [3] 計算機アルゴリズム研究会（一色、遠藤）、C言語による実用ニューロ・コンピューティング、ラッセル社(1992)
- [4] 山下隆義、イラストで学ぶディープラーニング、改訂第2版、講談社(2018)

理論に興味のある人には

- [5] 滝雅人、これならわかる深層学習入門、講談社(2017)

フレームワークを動かしたい人には

[6] 大田満久ほか、TensorFlow開発入門、翔泳社、(2018)

フレームワークの使い方を知りたい人には

[7] 巢鴨悠輔[監訳]、(株)クイープ[訳]、Francois Chollet [著]、
PythonとKerasによるディープラーニング、マイナビ(2018)

ビデオ教材

基本的な考え方を知りたい人には

[1] Geoffrey Hinton, Neural Networks for Machine Learning,
2016

https://www.youtube.com/watch?v=cbeTc-Urqak&list=PLoRI3Ht4JOcdU872GhiYWf6jwrk_SNhz9

ディープラーニングのインパクトを知りたい人には

[2] Yoshua Bengio, The Rise of Artificial Intelligence through
Deep Learning, TEDxMontreal

<https://www.youtube.com/watch?v=uawLjkSI7Mo>