

2023. 4 ブログ：『ディープラーニング 学習する機械』を読んで（２）、の詳細
（→ <http://www.1968start.com/M/blog/index3.html#2304c> ）

『ディープラーニング 学習する機械』を読んで（２）

中所武司

■このエッセイのきっかけ

情報処理学会誌（2022. 11）に、2018 年のチューリング賞を受賞したヤン・ルカンの著書「ディープラーニング 学習する機械」（384 ページ、講談社 2021）が紹介されていた。

私の卒論・修論（1968～1971）の研究との関連で読んでみることにした。

■懐古趣味的コメント

前回は、主に自伝的な内容の 1 章と 2 章を読み、その要約とコメントを以下に記述した。

・2023. 3 『ディープラーニング 学習する機械』を読んで（１）

＊ブログ： <http://www.1968start.com/M/blog/index3.html#2303>

＊詳細： [http://www.1968start.com/M/essay/2303DeepBook12. pdf](http://www.1968start.com/M/essay/2303DeepBook12.pdf)

今回は、3 章の「単純な学習機械」を読み、その要約とコメント（→★）を記述する。

なお、私の著作物の引用先の URL ページを示すときは、下記の略称を用いる。

【卒論】「条件反射の生体工学的解析」（1969）

[http://www.1968start.com/M/bio/olduniv/soturon. htm](http://www.1968start.com/M/bio/olduniv/soturon.htm)

【修論】「思考過程の数学的表現と模擬実験」（1971）

[http://www.1968start.com/M/bio/olduniv/shuuron. htm](http://www.1968start.com/M/bio/olduniv/shuuron.htm)

第 3 章 単純な学習機械

3-1 発想の源アメフラシ

- ・軟体動物のアメフラシのシナプス接続による適応の性質は、学習機械のモデルになる。
エラに触れるとエラを引っ込めるが、慣れると反応しなくなるという現象は、
接触検出とエラ反応制御の二つのニューロン間のシナプスの伝達効率の低下によるもの。
- ・この学習メカニズムは、神経系を有するほぼすべての生物に存在する。

＊アメフラシ：	1.8 万個（ニューロン数）
＊ショウジョウバエ：	25 万個（同上）
＊マウス：	7100 万個（同上）＜シナプス数：10 億個＞
＊ウサギとタコ：	5 億個（同上）
＊犬とブタ：	22 億個（同上）
＊オランウータンとゴリラ：	320 億個（同上）
＊人間：	860 億個（同上）＜シナプス数：約 150 兆個＞

3-2 学習と誤差の最小化：その一例

- ・熟練ドライバの運転中のハンドル操作を観測し、道路の中央からのずれの距離 x とハンドルの角度 y の組み合わせデータを何千組も集めて学習セットとして使用し、 x から y を求める関数 $f(x)$ が見つければ自動運転に使える。これが教師あり学習。

3-3 x の関数 y を予測する $f(x)$ を求める

- ・完璧なドライバならば、 $F(x) = w * x$ (w は傾き) となる。
Python プログラムでは、`def f(x, w): return w * x` と記述する。
学習セットでは、すべての点を通る直線は見つけれないので、すべての点の「できるだけ近く」を通る線を選ぶ。
具体的には、すべての点の誤差の二乗や絶対値の平均値を最小にするように w を選ぶ。
そのために誤差の平均値をコスト関数 $L(w)$ で表現し、確率的勾配降下法を用いる。
これは確率近似法とも呼ばれる。

3-4 数学好きのための補足

- ・教師あり学習で訓練される機械は、以下の基本原理に従う：
 1. 学習セットを集める。
 2. モデルの提示：パラメータ w をもつ第一の関数 $f(x, w)$ を提示。
 3. コスト関数 $C(x, y, w)$ を提示。その誤差と誤差の平均値 $L(w)$ を計算。
 4. コスト関数 $L(w)$ を最小化する関数 $f(x, w)$ のパラメータの値を求める。
通常は、確率的勾配降下法を使用。

3-5 ガリレオとピサの斜塔

- ・高所から石を落とすと、落下距離は落下時間の 2 乗に比例する。
ガリレオが実験によってこの法則を見つけた方法は、機械学習と同じだ。

→★経験から学ぶ教訓や訓練で身につける技術など、人間の学習も同じでは。

3-6 画像その他を認識する

- ・規則を発見する原理は画像にも適用される。入力画像は数値の集まりに過ぎない。
1000*1000 ピクセルの画像は各ピクセルの濃淡値を示す 100 万個で表現される。
- ・カメラ搭載自動車の自走訓練に適用される。
- ・教師あり学習は、同じ原理で動作する：
 1. 入力 x ：画像、音声信号、文字列など
 2. 所望出力 y ：入力 x に対する理想的出力
 3. 出力 y_p ：機械が生成した応答

→★親が入力と所望出力を提示する幼児の学習もこの原理では。

3-7 ローゼンブラットとパーセプトロン

- ・本節では、単純な線形分離器だったパーセプトロンについて述べる。
- ・当時、心理学や生物学の分野での脳の機能やニューロンの相互接続の仕組みの研究では、ニューロンを複数の突起を持つ星の形で表現した。
- ・各突起の先は、一本を除いて樹状突起（入力部）が形成され、接触領域（シナプス）を介して上流ニューロンとつながり、残りの一本は下流ニューロンへの唯一の出力部である。
- ・ニューロンは、上流からの電気信号を処理し、必要に応じて下流に電気信号を送る。これは一連のパルスで構成され、スパイクと呼ばれる。
- ・その頻度は、ニューロン活動の強度として数値で表された。

→★このニューロンの構造は、卒論（1969 年）の図 1-4 に掲載している（【卒論】参照）

2022.1 のエッセイ「脳の神経細胞における力学的情報伝達の発見」にも掲載：

<http://www.1968start.com/M/essay/2201synapseMech.pdf>

- ・1943 年、ウォーレン・マカロックとウォルター・ピッツは、サイバネティクスと神経科学の研究の中で、このニューロンを単純化した数理モデルを提案した。ニューロンは活性の場合は 1，非活性の場合は -1 を出力するとして、上流ニューロンの出力の加重和がしきい値を越えれば 1，そうでなければ -1 を出力する。このモデルでは、脳をバイナリニューロンの論理計算による論理的推論機械とみなした。

→★このモデルは、卒論の図 2-1「McCulloch-Pitts の神経回路網のモデル」に掲載。

2023.1 のエッセイ「人工知能分野の勃興～第一次 AI ブーム』について」にも掲載：

<http://www.1968start.com/M/essay/2301firstAIboom.pdf> (p.3)

- ・1957 年、米国の心理学者ローゼンブラットはこのモデルからパーセプトロンを考案した。この中で、加重和の重みを修正して誤差を調整し、機械を適応させていく手順を提案した。
- ・このアイデアの参考とした、脳内の学習によりシナプスの伝達効率が変わるという発想は、19 世紀末のスペインの神経解剖学者の研究によるものだった。
- ・さらに、カナダの心理学者ドナルド・ヘップが 1949 年の著書で提案した、2 つのニューロンをつなぐシナプスは両方のニューロンが同時に活性化すると強化されるという仮説を参考にしている。
- ・この仮説はヘップ学習と呼ばれ、1960 年代にその妥当性が認められ、1970 年代には、アメフラシの研究の中で、その生化学的なメカニズムが解明された。

→★私の卒論の参考文献の 15 番目にヘップの下記の著書を挙げている（【卒論】参照）

『D. O. Hebb: A textbook of psychology, W.B.Saunders company, 1958.

（白井ほか訳、行動学入門、紀伊国屋、1964）』

- 1957 年、ローゼンブラットの機械は完成した。巨大な金属製キャビネットのよう。何千本もの配線がスパゲッティのよう。入力画像をキャプチャする光電管のグリッドと重みを表す電動つまみは何百個もあり。つまみは小型電気モータ付きの可変抵抗器。
- パーセプトロンの電子回路は、可変抵抗器で重みづけされ、光電管のグリッドの出力電圧の加重和を計算し、しきい値を越えれば出力ランプが点灯する。
- パーセプトロンの革新的な点は、画像とその所望出力が提示されるたびに重みが自動調整されるという学習手順と、それをパターン認識に応用したことだ。

→★パーセプトロンについては、修士論文で次のように引用している。【修論】参照）

＜引用箇所：p. 15 の「1.4 節 思考に関連した工学的諸研究について」＞

『こうした結合係数の変化と、閾値関数をその骨組みとした、パターン認識、学習機械の研究も多く行われ、その代表的なものとして、パーセプトロン[34]がある』

（34 番目の参考文献） M. Minsky & S. Papert: "Perceptrons" The MIT Press, 1969.

3-8 25 ピクセルのグリッド

- パーセプトロンの原理
 - * 入力：5*5（25 個、 $x(i)$ ）のピクセルの画像、黒ピクセルは+1，白ピクセルは-1
 - * 各ピクセル $x(i)$ は、重みパラメータ $w(i)$ を経由して出力部（加重和）につながる。
 - * 重みの初期値は 0 なので、加重和は 0，出力は-1 となる。

3-9 CとDを区別する

- 文字の C と D を区別する訓練の例で、重みを調整する学習手順を述べる。
- 各文字のサンプルを提示し、正解の場合は何もせず、間違えた場合は重みを調整する。加重和が C はプラス、D はマイナスになるように重みを変える。
 - C が入力で加重和が 0 未満の場合、入力が +1 の重みを増加し、-1 の重みを減らす。
 - D が入力で加重和が 0 以上の場合、入力が +1 の重みを減らし、-1 の重みを増やす。
- 重みの増減は少しずつ行い、重みの設定が安定状態に収束すれば、学習は終わる。
- ローゼンブラットのパーセプトロンは、重さ数トンの機械だった。
- この手順は、システムの調整可能な重みをパラメータとするコスト関数の最小化である。

→★この段階では、2 つのニューロンをつなぐシナプスは両方のニューロンが同時に活性化すると強化されるというヘップ学習（3-7 節）は使われていない。

3-10 教師あり学習と汎化

- 学習機械では、学習に使用しなかったサンプルにも正解となる汎化が重要。
- C と D について位置、大きさ、フォントが異なる数百枚～数千枚のサンプルで訓練すれば不思議とサンプル以外でも識別できるようになる。

→★卒論で研究対象とした犬の条件反射における汎化と分化を思い出した。

ある条件刺激（1000 ヘルツの音）に対して条件反射が形成されると、それに類似した刺激（500 ヘルツの音）に対しても反応をおこすようになる現象を汎化という。

その後、1000 ヘルツの音で餌を与え、500 ヘルツの音で餌を与えないことを繰り返すと、後者の音では唾液が分泌されないようになる現象を分化という。

3-11 パーセプトロンの限界

- 位置、大きさ、フォントのばらつきが大きいサンプルでは学習は難しい。
この限界は、すべての線形分離器に共通である。
- 排他的論理和の例で、 $(0, 0)$ 、 $(1, 1)$ と $(0, 1)$ 、 $(1, 0)$ を分離する直線は引けない。
- 1960 年代、自然画像の被写体認識などのパターン認識には使えないことが判明した。
1969 年のシーモア・パパートとマーヴィン・ミンスキーの本「パーセプトロン」により AI 研究は初めての冬の時代を迎えた。

→★修論におけるこの著書の記述については、3-7 節参照。

3-12 解決策：特徴の抽出

- 解決策は、入力画像とニューロン層の間に、入力画像に含まれる特定パターンの有無を検出する**特徴抽出器**を配置することである。
- 20*20 ピクセルの画像で、文字の角部や線の終端部有無を検出して、フォントや書体の違いに関係なく識別する方法である。
- C は線の両端に終端部を持つ開いた曲線で、D は 2 つの角部を持つ閉じた図形なので、第 1 層で終端部検出ニューロンと 2 つの角部検出ニューロンにより、角部と終端部の数を数えるだけで、位置、大きさ、字体に関係なく、C と D を区別できる。
- この方法は、特徴抽出の中間ニューロンが大量に必要になり、最終ニューロンへの入力数が膨大になるため、2013 年以降は使われなくなった。
- そのため、パーセプトロンのニューロン層を多層にして特徴抽出する方式が考案された。
- もっとスマートな方法として、サポートベクターマシン SVM がある。
これは、1996～2010 年の主要な分類方法だった。
- SVM はカーネルマシンの一種で、2 層ニューラルネットワークとみなせる。
第 1 層は、学習セットと同数のユニットを持ち、入力はいずれも学習サンプルと比較するのにカーネル関数が使用される。第 2 層は、パーセプトロンと同様に加重和を計算する。
- SVM はサンプル数が少なくても信頼性が高く、1990 年代はカーネル法が圧倒した。
- 2 つのニューロン層ですべての計算が可能なが数学的に証明されている。
- 2010 年代初頭まで、パーセプトロン流のアーキテクチャは機械学習の主流であり続けた。
入力は手作業で設計した特徴抽出器で処理後、統計的学習法を使った分類システムに通す。

→★3-7 節と 3-11 節で言及した 1969 年発行の著書「Perceptrons」では、
p.231 の 13.2 節「Other Multilayer machines」で、多層マシンに言及している。

3-13 まとめ

- ・パーセプトロンは教師あり機械学習の端緒となり、汎化の特性がある。
- ・しかし、特徴抽出器の設計が大変だった。
- ・著者は、特徴抽出器を手作りするのではなく、それを訓練する方法を考えていた。
それが、多層ニューラルネットワークとディープラーニングの扉を開くカギだった。

以上