SE Report

連載 第2回

AIと安全工学

―AIの中核技術―

横浜国立大学　准教授

杉本　千佳 Chika Sugimoto

1. はじめに

人工知能（AI）技術は身近なサービスやシステムに導入され、例えばAI家電やAI型検索エンジンのように、実際に私たちの生活の中で使用されている。では、AIはどのような技術で、具体的にどのようなAI技術が用いられているのだろうか。第1回目の「AIの歴史と軌跡」の中で、第3次AIブームの到来には、機械学習の実用化と深層学習の登場という2つの大きな技術発展があったことを述べた。深層学習は第3次AIブームの中核となる技術であり近年特に注目を集めているが、実用的にはそれ以外の機械学習やベイズ統計学も中核技術として重要な役割を果たしている。そしてこれらの技術により、人間の頭脳が行うような複雑な学習をコンピューターで実現することが可能になり、大量のデータを学習して新たな知識を得ることができるようになった。本稿では、これらの技術について、数式を使うことなくそれぞれの学習方法の概要がつかめるように解説する。

まず、ベイズ統計はベイズの定理※1をもとにした主観確率※2を扱う統計学である。最初に「ある事象が発生する確率」を事前確率として設定し、新たな情報を得るとその確率を変更して事後確率とし、この更新を繰り返すことで本来起こるであろう事象の確率を導き出す。このように、最終判断は下さず可能性を数値で評価する。これにより、母集団の前提を必要とせず不完全情報環境下において、データが不十分でも未知の確率データを導出することが可能である。また、データ数が増えるにつれ精度が向上する。応用例として、迷惑メールフィルターやGoogleやマイクロソフトの検索エンジンがあり、ベイズ統計学が使われている。

一方、機械学習は客観確率※3を重視するネイマン・ピアソン統計学と呼ばれる正統派統計学をもとに誕生したが、ベイズ統計学と融合することで数々のAI技術を生んでいる。機械学習において、データを統計的に処理しパターン認識する際の数学的根拠を与えたのがベイズ統計学である。

深層学習を含む機械学習は、多くのデータを学習することにより知識を獲得し、識別や予測を可能にする技術である。機械学習手法にはサポートベクター―マシン（SVM）※4やニューラルネットワークなどがあるが、主に「教師あり学習」「教師なし学習」「強化学習」に分けられる（**図1**）。教師あり学習は正解を与えて学習させる方法で、回帰問題や分類問題に主に用いられる。教師なし学習は正解を与えずに学習させる方法で、データの分類やパターン抽出に用いられる。代表的なタスクに、クラスタリングや次元削減がある。強化学習は試行錯誤を通じて価値を最大化する行動を学習する方法であり、最適な行動の学習に用いられる。

例えば犬を判別する画像認識の場合、教師あり学習では、あらゆる犬種、色、大きさの様々な角度から撮られた画像データに「いぬ」

※1
ベイズの定理
ある事象に関連する可能性のある条件についての事前知識に基づいて、その事象の確率を記述するもの。
事象A、Bに対し、事前確率（事象が起きる前の事象Aが起きる確率）を$P(A)$、事後確率（事象Bが起きた後で事象Aが起きる確率）を$P(A|B)$とするとき次式で表される。

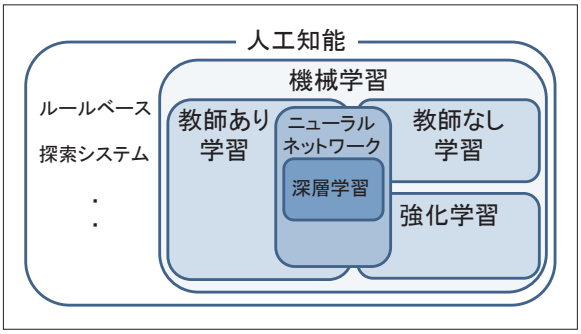
$$P(A|B)=\frac{P(B|A)\cdot P(A)}{P(B)}$$

※2
主観確率
事象生起の不確実性の程度を、主観的な信念や信頼度合いから評価した確率。

※3
客観確率
確率を頻度論で解釈し、ある事象が起きる比率が無限回試行を繰り返した際の極限値として定義されるもの。

※4
サポートベクター―マシン（SVM）
マージン最大化に基づき分類や回帰に用いられる教師あり学習によるパターン認識モデルの一つ。

図１　人工知能の概略図



のラベルを付けて AI に与え、犬の特徴を学ばせる。一方、教師なし学習では、犬や猫などの画像をラベルなしで AI に与えて学習させる。AI は各画像の特徴から類似点などを見つけて画像をグループに分類していく。このグループ化された画像に対し後付けでラベルを付与する。必ずしも「犬」「猫」のように期待した形で分類されるとは限らず、大きさの大小や色の違いで分類されることもある。このため、画像認識では教師あり学習が用いられることが多いが、高い認識精度の達成には大量のデータが必要であり、ラベル付与に多大な時間と労力を要するため、教師なし学習でグループ化の精度を向上させる工夫をしたり半教師あり学習が用いられたりする。

深層学習は、ニューラルネットワークの中間層（隠れ層）を深く多層にし、表現・学習能力を高めた機械学習の一手法である。ニューラルネットワークや深層学習は教師あり学習を発展させたものであるが、教師なし学習や強化学習にも応用されている。教師なし深層学習の学習メカニズムは教師あり深層学習のそれとは全く異なるため、本稿では教師あり深層学習について述べる。また、強化学習と深層学習を組み合わせた深層強化学習が近年飛躍的な技術発展を遂げ注目されているため、強化学習の項で触れたい。

2. 機械学習の方法と応用

（1）教師あり学習

教師あり学習では、入力データと正解データが紐づいた学習データをコンピューターに与えて、その関係性を表すモデルを学習する。得られたモデルをもとに、分類問題や回帰問題の予測タスクを行うことができる。

ここで、安全工学において大切な保守を考えてみよう。機械やシステムを安全安心に運用するためには、その故障率を算出して故障する前にメンテナンスすることが必要となる。そのため、出来るだけ正確に故障するタイミングを予測したい。そこで、設置環境や利用状況、稼働時間などの多数のデータを説明変数、稼働状況データを目的変数として、稼働状況が正常から異常に変化するタイミングを見つけようと考える。その手段がデータの学習であり、故障時を含む長期間の大量の稼働データを収集し、機械学習を用いて説明変数と目的変数の関係を学習させることで、その関係を表すモデル関数を導き出す。この学習では、多くの機械学習のアルゴリズムで「勾配降下法」[※5]が用いられている。学習はトレーニングにより最適なパラメータ値を自動で獲得することであり、その評価に損失関数を用いる。損失関数は実際の正値と予測値との差異を表す関数で、モデル関数に対して損失関数で表わされた損失の大きさが最小になるようにパラメータを少しずつ変化させ、トレーニングによりこの値を最小にするパラメータの組合せを求めるのが学習の目的である。勾配降下法では、現在の場所から勾配する方向に任意の距離だけ進むことを繰り返し、損失関数の値を減らすことで効率よく対象パラメータの最適値を求めることができる。このようにして説明変数と目的変数の関係を表すモデル関数を求めることで、故障タイミングの予測が可能になる。もちろん、正確なモデ

※5
勾配降下法
データから計算した誤差関数の勾配を降下していくようパラメータを更新する最適化アルゴリズム。

ルの作成と故障タイミングの算出には、十分な量の良質なデータが必要であることは言うまでもない。故障時のデータがある場合には、こうした教師あり学習が可能になる。

(2) 教師なし学習

教師なし学習では、正解ラベルがついていない学習データをコンピューターに与えて、データそのものが持つパターンや構造を見つけ出す。主なタスクに「クラスタリング」と「次元削減」があり、データ間の類似度に基づいてデータをいくつかのクラスタに分けてデータ構造を明らかにしたり、多次元からなる情報をその意味を保持したままより少ない次元の情報に落とし込むことでデータを圧縮し、データに内在する重要な情報を明確にすることができる。

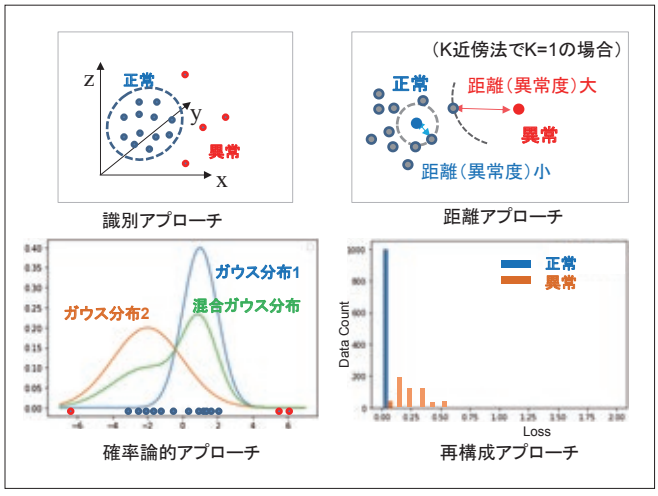
安全工学では、機械やシステムの異常検知は重要なタスクである。機械などの故障やデータ中に表れる外れ値を早期に検知・推測するために、しばしば異常検知手法が活用される。異常検知は、データセット内の正常データ群が持つ特徴とは異なった特徴をもつ観測結果や推定される異常パターンなどを識別することを言う。また、外れ値検知では機械学習を使って正常時には現れない外れた値を見つけ出す。正常時のデータを学習し、このモデルに当てはまらないデータを「異常」と判定する。この検出すべき「異常」は一定とは限らず、多種多様な形で出現する可能性がある。また、「正常」に対して一般的に出現確率が限られる。よって、「異常」を網羅的にモデリングすることは難しい。

そこで、基本的に教師なし学習を用いて異常検知を行うことになる。異常検知のためのアルゴリズムは、概ね次の4つに大別される。正常データと異常データを区分する境界を直接推定して識別するアプローチ、異常データは正常データから距離が離れているものと考えデータ間の距離に基づき異常検知を行うアプローチ、正常データの確率分布を推定して生起確率が低い事象を異常とみなす確率論的アプローチ、正常データを再構成するように訓練されたモデルによりうまく再構成できない場合に異常とみなす再構成アプローチである（**図2**）。それぞれの代表的な手法に、1-class SVM、K 近傍法、混合ガウスモデル（GMM）、オートエンコーダーがある。各手法はいずれも良く知られた手法でありここでは割愛する。

(3) 強化学習

強化学習では、与えられた環境における価値を最大化するように「エージェント」を学習させる。エージェントとは、ある条件下において行動する主体を指す。エージェントは、自身の行動に対する環境からの「状態」と「報酬」のフィードバックをもとに、将来得られる「価値」を最大化する「方策」を導き出す。状態はエージェントが存在する仮想空間、環境の情報であり、エージェントが行動する度に更新される。報酬はある行動の結果としてエージェントに与えられるフィードバックを指し、価値は一連の行動全体を経て生まれる効用である。この価値を Q 値、あるいは状態行動価値と呼び、これをアルゴリズムにより学習して最適な方策を導出する。

図 2　4 つの異常検知アプローチ



このように、強化学習は外部からの入力に対応して自律的に制御することを可能にする学習手法である。強化学習には、環境を把握して取るべき行動を決定する「特徴量抽出」と、時系列での各行動の影響を考慮して行動パターンを決定する「時系列データ生成」という2つの過程がある。深層学習はこの2つの処理過程に対応した強みを持つアルゴリズムを持つ。3章で示すように、深層学習は入力されたデータから自律的に特徴抽出することが可能で、時系列データの扱いも可能である。このため、強化学習に深層学習手法を組み合わせた深層強化学習は高い性能を見せている。

近年では、強化学習は自動運転や自律制御ロボットなどの基盤技術として注目を集めている。様々な環境で動作させる中で、強化学習により自律的に環境に応じた最適な車の運転やロボットの操作が可能になる。また、囲碁や将棋などのゲームは、将来の価値を最大化すること、つまり勝利することが目的であり、今の手が最終的な勝ち負けにつながるように打つ。このため、これらのゲームには強化学習が有効であり、プロ棋士を破ったコンピューター囲碁プログラムの AlphaGo[1] には深層強化学習が組み込まれている。場合の数が10^{360}に及ぶといわれる囲碁の巨大な環境において深層学習技術を応用して学習に成功させ、深層学習と強化学習の融合の可能性を示した。

3. 深層学習

ニューラルネットワークは入力層、出力層、隠れ層から構成され、層と層の間の各ニューロン同士のつながりの強さが重みで表わされる。深層学習はニューラルネットワークから発展し、複数の説明変数と目的変数が入出力される入力層と出力層との間に中間変数が媒介する多数の隠れ層が存在する構造を持つ。多数の説明変数を持つデータに対し、多層構造により何階層もの処理を経ることで特徴抽出と抽象化を行い、入力データの特徴を抽出して識別を可能にする。しかし、多階層の処理を行うために学習計算量が階層数により指数関数的に増加し、層が深くなると計算コストが膨大になるという問題が生じた。この問題を解決したのが、従来とは異なる学習メカニズムの「誤差逆伝播法」※6により学習する方法である。誤差逆伝播法では勾配降下法とは逆に、入力パターンをネットワークに順伝播して出力値を得、その出力値と教師データの値との差異を計算し、逆算して重み更新量を出力層側から計算して全体を見直すことで、ネットワーク全体を学習する。この際、誤差逆伝播法では最小化したい損失関数の勾配（微分）を効率的に計算することができる。これにより、膨大な変数の処理が可能になり、実用可能な手法になった。

深層学習は急速な発展を遂げ、その手法やネットワーク構造は進化しているが、最初に登場した基本形は畳み込みニューラルネットワーク（CNN）[2] である。CNN は深層学習の代表的手法であり画像認識にしばしば適用されるが、その名の通り、画像処理でよく利用される「畳み込み（convolution）」という手法が使われている。隠れ層を構成する畳み込み層は画像の局所的な特徴を抽出して際立たせ、「プーリング層」は局所的な特徴をまとめてフィルタリングし抽象化する。畳み込み層とプーリング層を組み合わせた隠れ層の処理を何層も重ねることで、画像の特徴を抽象化してパターン学習を進め、全結合層を経て最終的な識別結果を出力する。誤差逆伝播法により、CNN の隠れ層の随所に埋め込まれた重み付けパラメータ全体を一度に調整して学習することが可能になった。このように、深層学習では高次元データを識別するアルゴリズム自体が、アルゴリズムのどこをどう調整するかを特定して学習を行う自律的な学習

※6
誤差逆伝播法
勾配法によるニューラルネットワークの学習において、重み付けの勾配を連鎖律で求める方法。

手法が大きな特徴である。また、人間が判断基準を教えなくてもコンピューター自身が自ら特徴を抽出するため、人が試行錯誤して有用な特徴量を設計し入力データとして与える必要がなくなったことも大きな進歩である。

深層学習は画像認識や音声認識の領域で従来手法と比較し大幅な精度向上を達成するという成果をあげ、CNN の登場以後盛んに研究開発が進められている。学習データとしては、時間の経過とともに値が変化していく可変長の時系列データを扱うことも増えている。そこで、時系列データの学習に適したモデルとして、ある時点の入力がそれ以降の出力に影響を及ぼすことが可能となるよう、ある隠れ層の出力を再び隠れ層に入力として利用する再帰的構造を持つ再帰型ニューラルネットワーク（RNN）が導入された。しかし、RNN に長い時系列データを学習させる場合、遠い過去の隠れ層出力を反映させるのが難しいという問題がある。このため、中間層の状態を保持するパラメータ要素として時刻変化とともに保持すべき情報と忘れるべき情報をもち、忘却ゲートと呼ばれる関数が時間経過による各要素の変化を制御することで、長期の時系列データにも対応できるように改良された LSTM（Long Short Term Memory）[3] が利用されるようになった。さらに近年、時系列データを順次処理する必要がないという特徴を持つ Transformer※7 [4] が、並列化に容易に対応でき、大きなデータセットで効率的なトレーニングが可能なモデルとして登場する。発表当初は主に自然言語処理分野で使用されていたが、画像認識分野でも Transformer のアーキテクチャが用いられ CNN を凌駕する成果をあげたことから、その応用分野は広がっている。

4. まとめ

第3次 AI ブームの中で機械学習および深層学習の技術は大きく進展している。いまや機械学習の応用技術は、オープンソースの Python※8 などの AI 開発言語でパッケージ化されている。パッケージには機械学習のアルゴリズムがはじめから組み込まれているため、機械学習のメカニズムを深く理解していなくとも AI 開発が可能であり、実用化が進められている。発展の背景には、コンピューター性能の飛躍的な向上と、IoT 並びにクラウドの普及によるビッグデータ取得・蓄積・利用の容易化がある。AI が進展する土壌が整ってきたことにより、AI の様々な技術開発と理論的な解析が進められている。

一方で、こうした研究開発では深層学習技術が万能ではないことを念頭に置く必要がある。新たなモデル開発が、既存技術の延長でコンピューターパワーに頼る力まかせの成長では限界がくると考えられる。今後技術的なブレークスルーが起こらない限り、これまでの AI ブームと同様に冬の時代が来ることは避けられないだろう。また、AI 技術は倫理的な問題をはらんでいるため、慎重な適用が求められる。有用なアプリケーションにおいて AI 技術を適切に活用していくとともに、多方面からの AI 技術の検証が必要である。

参考文献

1） Silver, D., Huang, A., Maddison, C. et al., Mastering the game of Go with deep neural networks and tree search, Nature 529, 484–489, 2016.
2） LeCun Y., Haffner P., Bottou L., Bengio Y., Object Recognition with Gradient-Based Learning. In：Shape, Contour and Grouping in Computer Vision. Lecture Notes in Computer Science, 1681, 319-345, Springer, Berlin, Heidelberg., 1999.
3） Hochreiter, S., Schmidhuber, J., Long Short-Term Memory, Neural Computation, 9（8）, 1735–1780, 1997.
4） Ashish Vaswani, et. al., Attention Is All You Need, arXiv：1706.03762v5, 2017.

すぎもと●ら

生体計測工学、知覚情報処理、アフェクティブ・コンピューティングなどの研究分野において、生体・行動・環境情報の認識とその応用に関する研究に従事。東京大学大学院新領域創成科学研究科助教を経て、2010年から横浜国立大学大学院工学研究院准教授。

※7
Transformer
高性能で汎用性が高い深層学習手法の一つ。

※8
Python（パイソン）
優れた科学技術計算ツールとして機械学習分野で使われる、インタープリタ型の汎用プログラミング言語。