

### 2.2.3 学習

顧客が10万人いれば年齢、性別、職業なども多様であり、その行動もまちまちと考えられるが、一見共通点のない顧客が同じ購買行動を示すことがある。顧客データを収集・分析することで何らかの法則を見い出せば、同じ購買行動を示す顧客をターゲットとして広告を打つなど、ビジネスに活かせる可能性がある。データ分析を行うには、従来は数学的なモデルで近似する統計学的手法が主流であったが、近年はコンピューターを駆使し、様々な手法でデータを分類し、法則を見出す「機械学習」も台頭している。本節では、AIにおける「学習」に焦点を当て、機械学習の基本的考え方と、新旧の機械学習のアプローチの手法を説明し、学習自動化などの最新技術動向を紹介する。

#### (1) 概要

コンピューターにおける「学習 (Learning)」は、特に「機械学習 (Machine Learning)」と呼ばれる。機械学習とは、データの背後に潜む規則性や特異性を発見することにより人間と同程度あるいはそれ以上の学習能力をコンピューターで実現しようとする技術<sup>\*18</sup>である。機械学習の処理は、大きく「学習」ステップと「推論」ステップに分かれる。

「学習」とは、事前に与えられたデータからモデルをつくることをいい、「学習」に使われるデータを特に「学習データ」と呼ぶ。学習が終わったモデル（「学習済みモデル」とも呼ばれる）には、学習データの統計的な傾向や規則性が反映されている。モデルの出力は、判定結果や、分類対象のクラスや、予測結果を示す数値になる。また出力の確かさ（確率）を出力する場合もある。

これに対して「推論」とは、実用環境で得られるデータなどを使い、学習済みモデルを使って判定、分類、予測などを行うことをいう。

#### (2) 主な技術

ここでは、機械学習について、学習手法による分類と、代表的な機械学習アルゴリズムを説明する。

##### 学習手法による分類

学習には様々な手法があるが、大きくは、教師あり学習 (Supervised Learning)、教師なし学習 (Unsupervised Learning)、強化学習 (Reinforcement Learning) の3つに分類できる。

表2-2-2、図2-2-9にその概要を説明する。

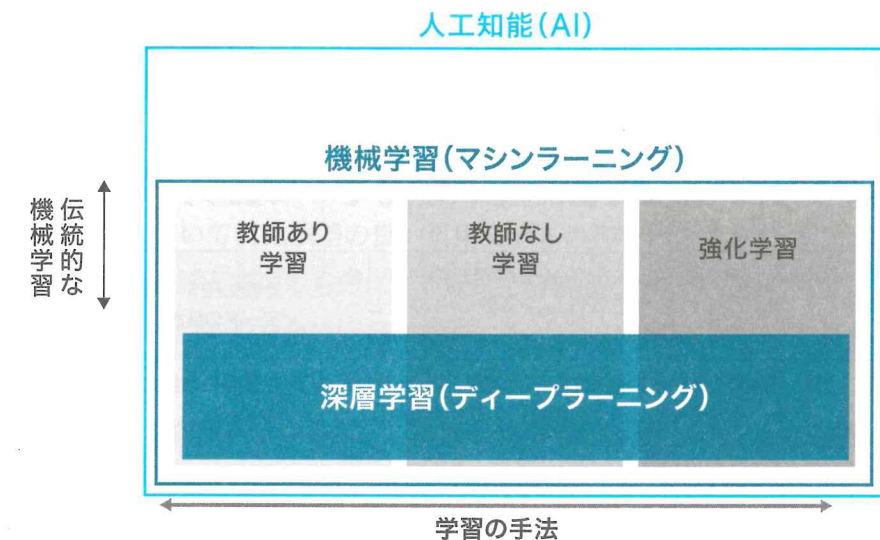
■表2-2-2 学習の手法の3分類

教師あり学習	正解値などのラベルが付与された学習用のデータを利用して学習済みモデルを作成し、推論時に、ラベルのない新たなデータに対して分類や予測を行う。自然言語処理、スパムメールフィルタリング、手書き文字認識などの応用がある。
教師なし学習	ラベルが付与されていないデータから、隠されている構造や規則性を記述するための特徴を学習する。クラスタリングやデータ次元圧縮技術などがこれに含まれる。
強化学習	強化学習では、ゴールや目的を仮定せず、学習を行う「エージェント」が状態を観察し、環境からの報酬 <sup>*19</sup> を最大化するように試行錯誤しながら行動を選択することで学習を行う。あらかじめ報酬を最大化する条件などを定義する必要がなく、システム自身に試行を繰り返させることで学習が進む。ロボット操作などに使われることが多い。

\*18 JST (研究開発の俯瞰報告書) システム・情報科学技術分野 (2019年) : 2.1.1 機械学習 <<https://www.jst.go.jp/crds/report/report02/CRDS-FY2018-FR-02.html>>

\*19 報酬とは、行動の結果が良かったか悪かったかに関する情報を指す。

■図2-2-9 学習の手法の関係



出典:PWC、Strategy&分析「機械学習のアプローチ」に加筆

なお、機械学習には一般に、大量の学習データが使用される（数百万から数千万の場合もある）。例えば画像認識で教師あり学習を行う場合、正解ラベルのついた対象画像を大量に用意する必要がある。データを増やす工夫として、前処理の段階で、一つの正解ラベル付きの画像データに対し、反転画像や拡大・縮小した画像を生成して学習データに追加する方法がある。

また、すでにある領域で学習済みのモデルをほかの領域に流用する「転移学習」という方法もある。例えば医療画像診断において、ImageNetと呼ばれる大規模画像データセットによってすでに画像分類問題用に学習済みのモデルを初期モデルとして使用し、早期がん検出のために再学習させることで、少ない画像データでも高性能な診断ができることが試みられている。

##### 代表的な機械学習アルゴリズム

機械学習のアルゴリズムは、大きく「ディープラーニング」とそれ以前の「伝統的な機械学習」に分類できる。代表的なものを以下に示す。

■表2-2-3 代表的な機械学習のアルゴリズム

名称	概要	特徴
ディープラーニング	神経細胞（ニューロン）を模倣した処理を深く多層に重ねたネットワークをモデルとし、学習させるアプローチ。非線形性を導入するために層間に活性化関数を挟む（詳細は2.3参照）。	・特徴量を自ら発見できる。問題によっては人間を超える性能を達成。 ・大量なデータ及び前処理が必要。
ロジスティック回帰(Logistic Regression)	AかBのどちらかしか起こらない場合などに、事象がどちらに分類できるかの確率を計算する。	・特徴量の重みからその特徴の結果への貢献度が分かる。 ・外れ値などに過剰反応しやすい。
サポートベクトルマシン (Support Vector Machine: SVM)	データを何らかの特徴に基づき2分類する際、双方のデータ群を等しい距離（マージン）で区切る平面を機械学習により発見する手法。	・未学習データに対しても高い識別性能がある。 ・仮説の設定や特徴の選択が必要。
ベイズ推定 (Bayesian Inference)	確率の初期状態（事前確率）に対して、得られたデータにより確率（事後確率）を計算（ベイズ更新）して推定を行う。	・データを増やすことで精度を向上できる。 ・学習データで仮定した確率分布以外には未対応。