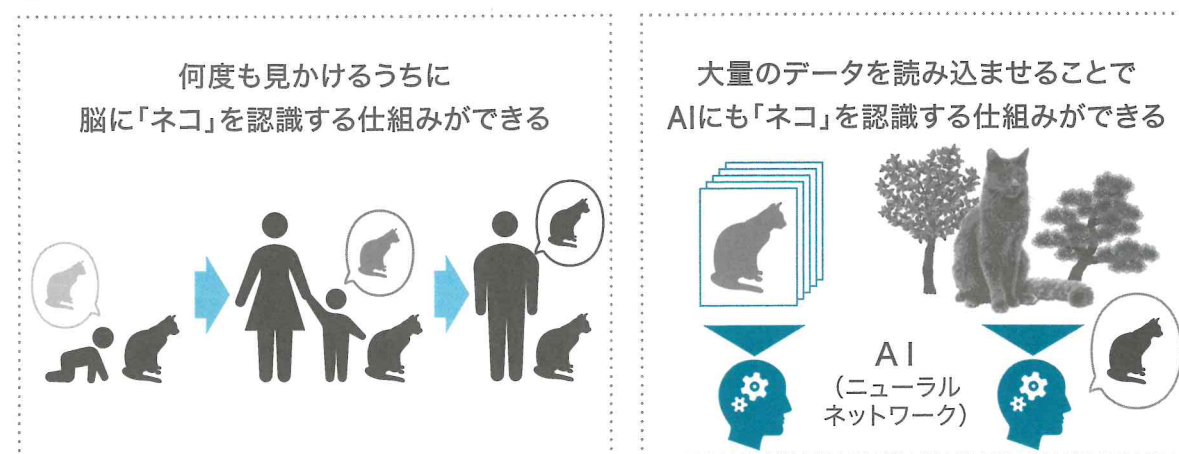


### 2.2.1 認識

誰かに「ネコ」の説明をするときには「耳が三角で……」と具体的な特徴を挙げるが、実際にネコを見たときには(特徴の確認なく)瞬時に「ネコ」と認識することができる。これは、何度もネコを見かけるうちに、対象の特徴からネコを認識する仕組みが脳細胞上にできあがるためである(図2-2-1)。従来のコンピューターによる認識は、個別の特徴を確認する方法が主流であったが、近年は大量の画像を読み込ませることで、特徴から対象を認識する仕組みを、神経細胞(ニューロン)を模したモデル上で実現する機械学習が台頭している。本節ではAIによる「認識」に焦点を当て、物体認識、音声認識、行動認識などの主要な認識技術と最新技術動向を紹介する。

■図2-2-1 人の認識、AIによる認識



#### (1) 概要

日本語の「認識」に相当する英語には、「Cognition」と「Recognition」がある。子どもが何度もネコを見かけるうちに「ネコが分かる(=ネコ概念を得る)」ようになることが「Cognition」であり、すでにネコ概念を持つ大人が一瞥して「ネコ」と判断するのが「Recognition」である。コンピューターでいう「認識」とは、この「Recognition」をいう。

「コンピューターの認識」とは、データの中に対象に関するパターンを見出す「パターン認識」処理といっても良い。データからある規則(パターン)を発見したり、規則(パターン)を使って、データがどのパターン=クラス<sup>※1</sup>に属するかを分類し、判定したりする「クラス発見とクラス分けの問題」であると考えれば分かりやすい<sup>※2</sup>。認識は、コンピューターの初期からの課題の一つである。人間の作業をコンピューターに代行させようとする、人間のように実世界を認識できる必要がある。このときの基本になるのが、人間の視覚や聴覚などに相当する機能だ。しかし、視覚や聴覚に対応する生データは、コンピューターからすれば背景が写り込むなど情報が過多であり、その中から対象に関連した特徴やパターンを抽出する手法が考えられてきた。

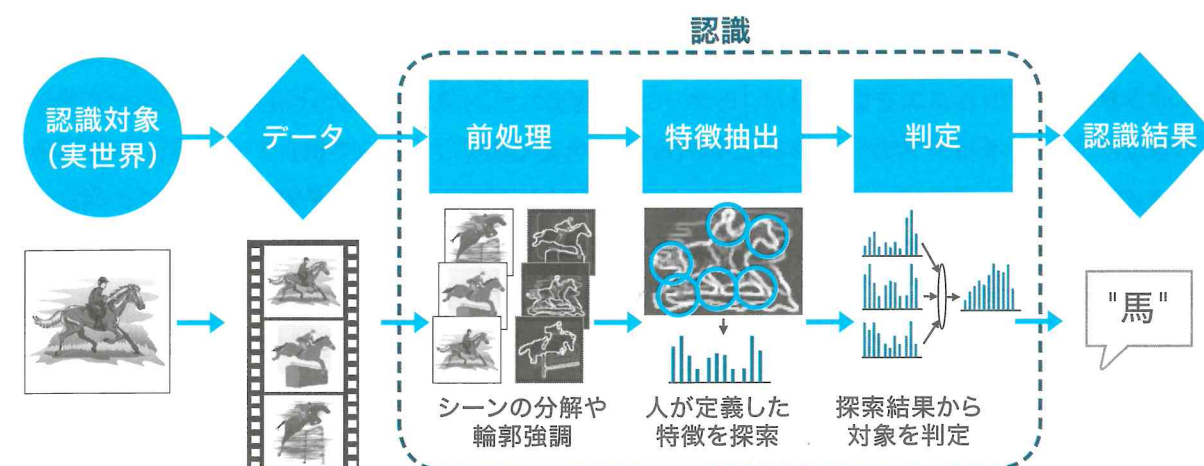
認識技術は認識対象に基づいて、「物体認識」、「文字認識」、「顔認識」、「行動認識」、「(自然)言語認識」などに分類される。また、認識するデータ構造に着目すれば「画像(静止画)認識」、「映像(動画)認識」、「音声認識」などに分類されることもある。

※1 クラスとは、ネコやイヌや花などの、カテゴリを指す。

※2 M. Bishop ("Pattern Recognition and Machine Learning", Springer, 2006)によると、「パターン認識の分野は、データから自動的に規則(regularities)をコンピューターのアルゴリズムを利用して発見すること、発見された規則を用いて、データを複数のカテゴリに分類するなどの、アクションを起こすこと、に関連している」と定義される。

認識技術の一例として「画像認識」の概要を説明する。画像認識は、大きく、アルゴリズムによる手法(図2-2-2)とディープラーニング(図2-2-3)の2つがある。アルゴリズムによる手法では、対象を表す「特徴」を人間が与えてアルゴリズムによりデータから「特徴量」を計算する「特徴抽出」部分と、算出された特徴量から統計的手法などを用いて対象を特定する「判定」部分から構成される。特徴抽出を容易にするためにデータからノイズや揺らぎを除去したり、不要な部分を削除するフィルタリングをしたりといった前処理部分を含むこともある。

■図2-2-2 アルゴリズムによる認識



ディープラーニングによる手法は、学習を繰り返すことによって、データに潜在する特徴の抽出を行うことができ(詳細は「2.3 ディープラーニング」参照)、そのことで画像認識システムが高い認識率を達成し、AIブームの火付け役ともなった。2012年の画像認識ソフトウェアのコンテストであるILSVRC (ImageNet Large Scale Visual Recognition Competition ; 大規模な画像データのコレクションであるImageNetを利用した認識技術のコンテスト)で、Geoffrey Hintonらのトロント大学(University of Toronto、カナダ)のチームがAlexNetと呼ばれるディープラーニングを応用したシステムにより、それまでのアルゴリズムを利用した特徴抽出方式の画像認識システムに対して10%以上低い約16%という誤り率を達成した。

■図2-2-3 ディープラーニングによる学習と認識

