

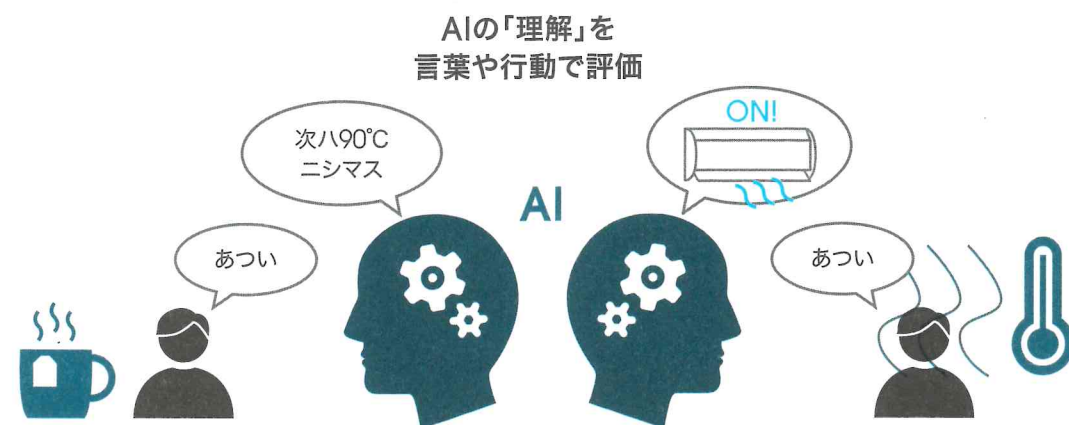
### 2.2.2 理解

人間であれば、赤い旗が海に立っていれば「遊泳禁止」であることや、相手が「あつい」と言ったときに気温が高ければ「暑い」、お茶を飲んでいれば「熱い」であることを「理解」するのは容易である。しかし、コンピューターが同様に「理解」するためには、画像や言語を「認識」するだけでなく、その背景や文脈(コンテキスト)を踏まえて分析する必要がある。本節では、AIによる「理解」に焦点を当て、言葉の意味理解、画像を介した意味理解などの主要技術と、マルチモーダル学習、比喩の理解などの最新技術を紹介する。

#### (1) 概要

「理解」とは「物事の道理や筋道が正しくわかること。意味・内容をのみこむこと。」「他人の気持ちや立場を察すること。」(デジタル大辞泉)である。しかし、人間同士であっても、自分が何かを理解できたか、相手が自分の気持ちを理解しているかを確認することは容易ではない。ましてや、コンピューターが何かを理解したか否かを外部から確認することは極めて困難である。そこでここでは、コンピューターが画像や言語を認識し、適切な応答をした場合に、コンピューターが「理解」したと表現する(図2-2-6)。

■図2-2-6 AIによる「理解」のイメージ



まず、言葉による「理解」であるが、言葉には様々な表現があり、ほかの言葉との関係も知る必要がある。例えば、「三毛猫」は「ネコ」に属し、「ネコ」は「ペット」に属するといった関係を知らなければ、「うちのネコは食事中」、「三毛猫がエサを食べている」、「ペットが何か食ってる」といった表現を同一の意味と判定することができない。こうした単語間の関係には、ハンドルと自動車といった部分と全体の関係、あるいは「同義語」などがある。こうした単語を分類し、体系づけた辞書を「シソーラス」と呼ぶ。AIがシソーラスを使うことで、人間の言葉や文章を分析し、異なる表現であっても同じ意味と判断して応答することができる。

また、「暑い」と「熱い」のように同音異義語の場合には、その場の状況やコンテキストを踏まえた応答が必要となる。一方、BERT<sup>※12</sup>に代表されるディープラーニングによる自然言語処理技術は、会話などのデータを膨大に学習させ、文章の穴埋めや後に続く文章の予測を行うことで、同音異義語の使い分けを含め、スムーズな文章生成、応答を可能としており、あたかもコンピューターが「理解」しているように見える。

※12 Bidirectional Encoder Representations from Transformers.「2.2.6 言語」で詳しく説明する。

画像に関しては、AIが画像上の何かを認識したとしても、その意味までは理解できていない。しかし例えば、自動運転車があらかじめ「ネコは可能な限り回避すべき障害物」、「段ボール箱は跳ね飛ばして良い障害物」という知識を与えられていれば、前方にネコを認識した場合に回避のフラグを立てることが可能である。実際にこの応答(ネコに対して回避行動をとった)を人間が目撃すれば、自動運転車が状況を理解して行動しているように見えるであろう。

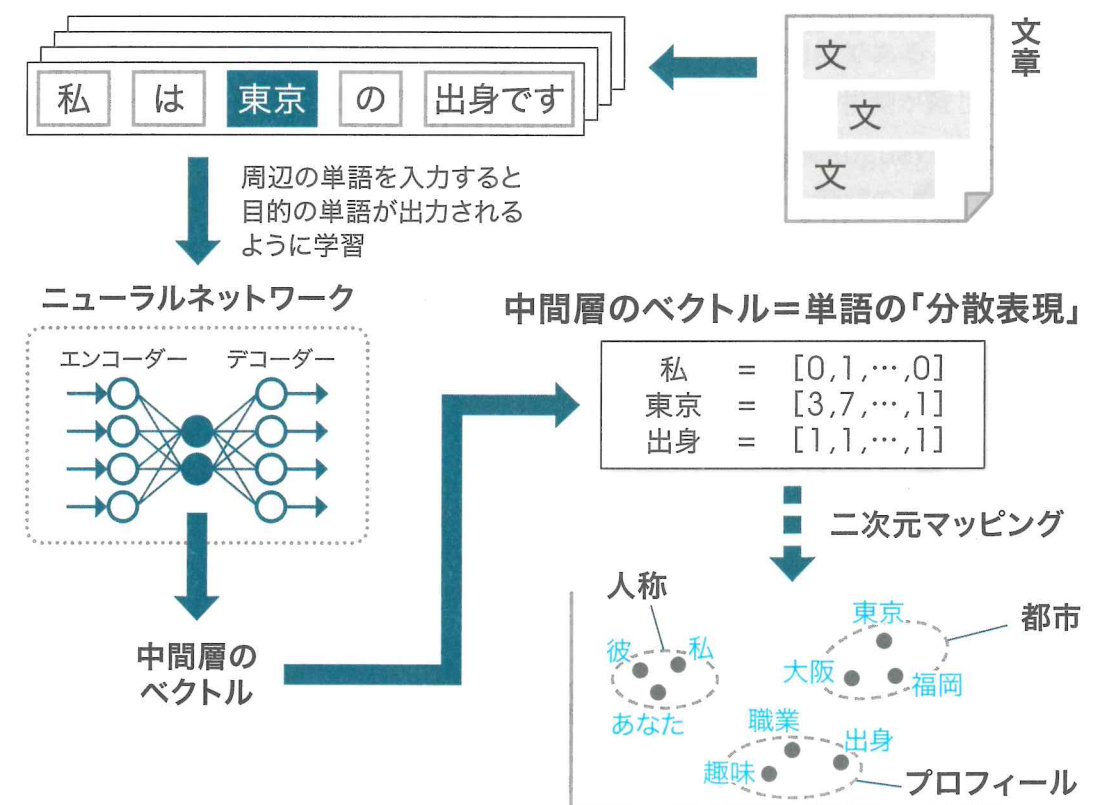
#### (2) 主な技術

ここでは、AIによる「理解」に焦点を当て、分散表現による言語の意味理解、画像を介した理解、身体性の獲得による概念の理解、などを紹介する。

##### 分散表現による言語の意味理解

文章における単語の意味は、その単語の周辺の単語や文脈に依存し、コンピューターがこれらを考慮した応答ができれば意味を理解しているとみなせる。このような意味の理解に資する技術に分散表現がある。分散表現とは、簡単にいうと、与えられた文章において、単語が出現する周辺の単語の情報を使って、その文章における単語の特徴を多次元のベクトルで表したものである<sup>※13</sup>。分散表現の求め方の代表的なものに、大量の例文データを使い、ニューラルネットワークに分散表現を学習させる手法がある(図2-2-7)。分散表現が意味を表すという根拠となるのは、意味の似ている単語は、似ている文脈(文章中の単語の並び)に出現するという「分布仮説」である。意味の似ている単語の分散表現は、ベクトルとして距離が近い分散表現に対応する。

■図2-2-7 ニューラルネットワークによる分散表現の学習の例



※13 ターゲットとなる単語の周囲の単語の情報(コンテキスト)から、目的とする単語の情報を得られるような変換行列を、コンテキストを入力とし、単語を出力とするニューラルネットを用いて、ターゲットとする単語の事後確率が最大になるように学習することで求める。得られた変換行列から単語の分散表現が得られる。