

Word2vec<sup>\*14</sup>は、分散表現を実装したアルゴリズムであり、学習した単語のベクトル表現(ワードベクトルファイル)を生成する。これを使うことで、ある単語に近い単語(意味が似ている単語)を探す、あるいは、各単語のベクトル同士の演算を行い、その結果に近いベクトルを見つけることも可能になる。例えば、「パリ」-「フランス」+「イタリア」というベクトル同士の演算を行った結果のベクトルは、「ローマ」のベクトルに近いものになるという。イタリアの首都がローマであるという事前の知識にはない場合でも、「イタリアの首都はどこ?」という質問に、「フランスの首都はパリ」という知識と、それぞれの単語の分散表現を使って「ローマ」と応答できれば、首都という概念を理解しているかのように見えるであろう。

### 画像認識を介した意味理解

ある画像に対して何が写っているのかをコンピューターに言語表現させたとき、人間と同じと判定できるのであれば、コンピューターは画像の「意味」を理解しているとみなせる。そこで、画像を介して異なる言語で表現させることで、翻訳が可能になる。例えば、日本語の「ネコ」という単語を使い画像を検索させ、表示された画像に対して英語で学習した画像認識を行わせることで"Cat"という英単語を得ることができる。さらに、様々な言語の単語に対して、同様の手順を繰り返すことで、画像を介して、複数の言語の単語を結びつけることが可能になり、画像の意味を理解したかのように見える[1]。

またVQA (Visual Question Answering) [2] 技術は、提示した画像の内容に関する質問に対して適切な回答を返す(例: バスの色は何色? → バスの色は赤色です)機能を深層ニューラルネットワークで実現し、コンピューターが画像の内容を理解しているかのように見えることを示した(図2-2-8)。

■ 図2-2-8 VQA (Visual Question Answering) のイメージ



### 身体性の獲得による概念の理解 (概念メタファー)

言語を通さなくても身体を通じた感覚として理解できる場合がある。「腑に落ちる」、「腹落ちする」などと表現する。このような身体や感覚を通じた理解の説明の一つとして「概念メタファー」がある。「概念メタファー」とは「ある概念領域を別の概念領域を用いて理解すること」<sup>\*15</sup>と

定義されている。概念領域でも最も基本的とされるものが肉体感覚に根差すものであり、これがあればメタファーを使って、様々な領域の概念が理解されるという考え方がある。AIにおいては、ロボットが「身体性」を獲得すれば、この肉体感覚を獲得したとみなすことができ、概念の理解を可能としたと考えることができる。例えば、人間が本来持つ「上下」などの肉体的感覚から、「成り上がり者」や「気分があがる」といった出世や気分に関する概念が理解できる。

### (3) 最新技術動向

#### マルチモーダル学習による対象の理解

あるものを複数の様式(モード)で扱うことを「マルチモーダル」と呼ぶ。機械学習において、いくつかの異なるモードのデータから学習して、統合的に処理するのがマルチモーダル学習である。画像からのキャプション生成や、キャプションからの画像生成の例のように、ディープラーニングは複数のモダリティを結びつける非常に強力な方法である。

視覚や言語、動作など、複数の様式で入出力が可能なインターフェースをマルチモーダル・インターフェースと呼ぶ。大成建設とエクサウィザーズがインターフェックス大阪に出展した液体計量ロボットアームは、視覚と触覚を組み合わせることで、液体の粘度を予測して注ぎ方を変化させることができる<sup>\*16</sup>。AIが複数のモードのデータを学習し、統合的に処理することで対象の特性をより正確に理解することができれば、音声と視覚により人の意図を適切に理解するなど、精度の向上が期待できる。

#### 文章における比喩表現の特定

言語表現に含まれる「メタファー(比喩)」をコンピューターが理解する試みもなされている。メタファーは日常的に使われる表現ではあるが、単語同士が無関係であることもあり、「彼は鬼だ」といった直接的な表現の場合には、文章構造や単語から意味の推測が難しい。しかし、EMNLP2018(Conference on Empirical Methods in Natural Language Processing)では、フルテキストを対象としたディープラーニングを使うことで、動詞のメタファーを特定できたとする論文が発表されている[3]。分散表現や文脈や比喩の理解などに加え、ロボットなどの深層強化学習の進展などもあり、コンピューターにおける「身体性」(Rodney Brooksの提唱によるEmbodiment)や「環境における身体性」(Rolf Pfeifer)<sup>\*17</sup>の獲得に向けた今後の発展が期待される。「身体性」については、「2.2.8 身体」を参照されたい。

#### 参考文献

- [1] Hewitt, John, et al. "Learning translations via images with a massively multilingual image dataset." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018.
- [2] H. Gao et al., "Are You Talking to a Machine? Dataset and Methods for Multilingual Image Question Answering", In Proc. of NIPS, 2015.
- [3] Gao, Ge, et al. "Neural metaphor detection in context." arXiv preprint arXiv:1808.09653 (2018).

\*14 Google Code Archive - Long-term storage for Google Code Project Hosting <<https://code.google.com/archive/p/word2vec/>>

\*15 George P. Lakoff, and Mark Johnson, Metaphors We Live By, University of Chicago Press, 2003. 邦訳「メタファーに満ちた日常世界—Metaphors We Live By」, 松柏社, 2013年

\*16 <https://aishinbun.com/clm/20190304/1982/>

\*17 Pfeifer Rolf, Christian Scheier. Understanding intelligence. MIT press, 2001. 邦訳「知の創成—身体性認知科学への招待」, 共立出版, 2001年