

1. 画像生成AIの現状と核医学での可能性



松原 佳亮 秋田県立大学システム科学技術学部経営システム工学科

画像生成AIの変遷と現状

ChatGPTなどに代表される「生成AI」は、その名前のおりテキストや画像を「生成」する人工知能 (artificial intelligence: AI) の総称である。なかでも画像を生成するAIのことを「画像生成AI」といい、Stable Diffusion, Midjourney, DALL-Eといったテキストから画像を生成するAIサービスが登場したことにより、われわれにとって身近になった。しかし、本来「画像生成AI」という言葉は、テキストから画像を生成するAIに限定するものではなく、もっと広義に画像を出力・生成するAI全般のことを指す言葉である。本稿においても、後者の定義に従って「画像生成AI」について解説していく。

昨今のAIブームの火付け役となった技術として、深い層を持つニューラルネットワークモデルを用いたディープラーニングが挙げられる。特に、動物の視覚野の仕組みを模した畳み込みニューラル

ネットワーク (convolutional neural network: CNN) を使ったモデルにより、画像認識の精度が大幅に改善された^{1), 2)} ことがブームのきっかけの一つとされている。その後、入力した画像から異なるタイプの画像を生成する仕組みとして敵対的生成ネットワーク (generative adversarial network: GAN)³⁾ が提案されたことにより、画像を変換・生成するAIやそのアプリケーションの研究開発が爆発的に進んだ。例えば、物体の輪郭だけ描かれた線画からその物体の完全なイラストを復元したり⁴⁾、写真と同様の風景について有名な画家が描いたかのような画像をその写真から変換生成したりすることも可能となった⁵⁾。医用画像分野においても、MR画像からCT画像の変換生成など、従来は困難であったアプリケーションもGANにより可能になった⁶⁾。一方で、GANは生成器と判別器の2つのモデルを競合させながら学習するため、学習が不安定になりやすいなどの問題があった。

ところが、昨今の「画像生成AI」で

は、CNNやGANといった技術とは違った技術が使われており、それらは大別すると以下の2つに分けられる。1つはTransformerおよびそれによる事前学習であり、もう1つは拡散モデルである。以降はこれら2つの技術について解説する。

まずTransformer⁷⁾は、もともと文章などの自然言語処理のためのモデルとして提案されたものであり、Attentionという仕組みを搭載したモデルである。Attentionでは、ベクトルとして表現されたデータ (テキストの場合は単語など) 同士の内積を計算することで、データ同士の類似度・関連度、言い換えればAIが目標とするタスクを達成するために、データのどこに注目すべきかを直接計算している。例えば、図1のように「好きな動物は」という文章を理解する上では、「好き」と「動物」との間の修飾-被修飾の関係を認識する必要があるが、Attentionの仕組みを使うことで、こうした関係を直接計算により認識することができる。このように、文章におけるコンテキストを効率良く理解できることが

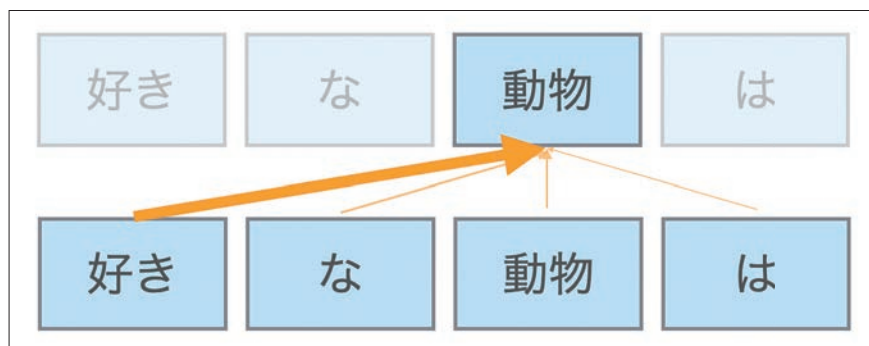


図1 Attentionによる修飾-被修飾関係の認識の一例