

話題のLLMとディープ・ラーニングの関係

[ご購入はこちら](#)

新井 正敏

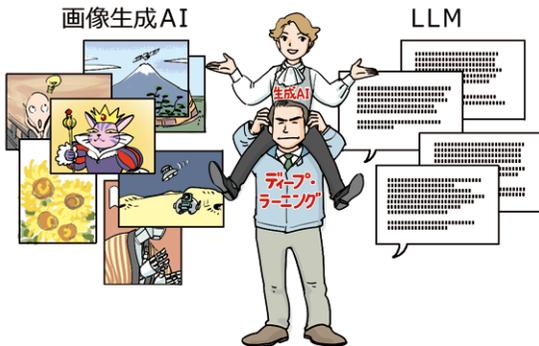


図1 今はやりの生成AI技術はディープ・ラーニングの技術に支えられている

学習モデル構築の工程で、ディープ・ラーニングのバックプロパゲーションが、LLMモデルのパラメータ更新に用いられる

最近では、さまざまな開発現場でAI技術が取り入れられるようになりました。特に今注目を浴びているのは、ChatGPTに代表されるLLM(大規模言語モデル)など生成AI技術です。この生成AI技術には、特に学習モデル構築の工程において、2010年代後半に話題になったディープ・ラーニングの技術が利用されています(図1)。

AIを使った開発で特につまづきやすいのは、学習モデルの構築だと思います。十分な知識がないまま学習モデルを構築しようとすると、うまく構築できるかどうかは運任せになってしまいがちです。

本特集では、ブラックボックスになりがちなディープ・ラーニングの仕組みについて、数学の観点からメカニズムを解説します。また、実際にモデルをフルスクラッチで実装することにより、手を動かしながら理解します。(編集部)

年々大規模化が進むLLM

LLMは、Large Language Modelの略称です。大量のテキスト・データで訓練された学習モデルであり、自然言語処理タスクを遂行するために設計されています。まずは、ChatGPTを例に、背後にあるLLMの規

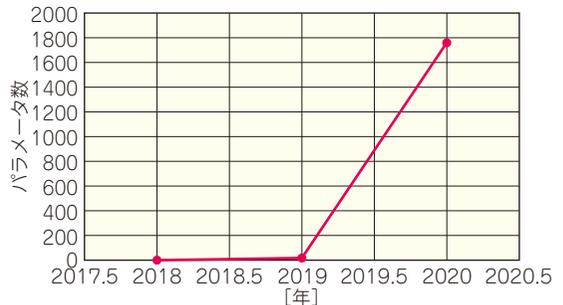


図2 LLMモデルのパラメータ数の推移

模を見てみましょう。

ChatGPTは、OpenAIのGPT(Generative Pre-trained Transformer)シリーズの一部として開発されており、さまざまなバージョンが存在します。それぞれのバージョンには異なる規模のモデルが用意されています。

次に、主要なモデルのリリース年とモデル規模を示します。

- GPT-1(2018年): 最初のモデルであり、約1億1700万個のパラメータ
- GPT-2(2019年): 約15億個のパラメータを持ち、発表当時はその生成能力が注目された
- GPT-3(2020年): 約1750億パラメータの大規模モデルであり、その性能により一躍注目を集めた
- GPT-4(2023年): 具体的なパラメータ数は公表されていないが、GPT-3よりもさらに大規模で高度なモデルとなっていると言われている

図2に示すように、ここ数年で指数関数的にモデルの大きさが増していて、この勢いはますます大きくなっていくものと思われます。

● LLM普及の背景…開発環境の進化

この大規模な言語モデルが実現可能になった背景として特に重要なのはハードウェア:GPU(Graphics Processing Unit)の進化とソフトウェア環境: CUDA(Compute Unified Device Architecture)の貢献が大