

2021 年 8 月

SoC1249

Scaling to AI Breakthroughs

By Sean R. Barulich (Send us [feedback](#))

AI 革新にあわせた規模拡張

AI 開発者は AI を進化させ、その偏向やデータの問題、方法論の不備など、実用化を阻む様々な問題に対処するべく奮闘している。業界関係者が実用的な AI 技術を求めるなか、研究者やテクノロジー企業は AI システムの柔軟性、精度、推論力を高める新たな戦略を模索中だ。先端を行く開発者たちは、大規模モデルの作成と高性能コンピューティングシステムの進歩が同時に進めば、AI の次なる飛躍につながると想像している。AI 企業や研究者は既存の AI モデルと学習データセット、専用ハードウェアの規模の拡大が先進技術の性能と効率を向上させ、自然言語処理や安全性が必須のコンピュータービジョンなど、幅広い分野で新たなチャンスを生み出すと考えているのだ。

2021 年 6 月上旬に開催された北京智源人工知能研究院(BAAI)の年次総会において、所属研究者らは自然言語処理、テキスト生成、画像生成、画像認識など、あらゆるタスクを実行するマルチモーダル・ディープラーニングモデル「悟道 2.0」(Wu Dao 2.0)を発表した。このモデルは 1 兆 7500 億のパラメータを学習しており、米国 OpenAI 社の GPT-3 (Generative Pre-trained Transformer 3) 言語モデルに比べ、桁違いの大きさになっている。BAAI を率いる Zhang Hongjiang 博士は、大規模な AI モデルとさらに強力な計算システムが、汎用人工知能やその他 AI の進歩の重要な要素技術になると述べている。米国 Google をはじめとする他の AI 開発者も同様の考えを示し、大規模 AI モデルとデータセット、高性能コンピューティングシステムのさらなる進歩がディープラーニングを大きく前進させ、現在の欠点に対応してくれると主張している。し

研究者やテクノロジー企業は AI システムの柔軟性、精度、推論力を高める新たな戦略を模索中だ。

かし、業界として大規模なモデルやデータセットに移行するには、時間効率の良い学習技術や専用 AI チップのイノベーションが必要になってくるだろう。実際、複数の研究グループが、グラフィックス・プロセッシングユニットや特定用途の集積回路など、業界標準 AI アクセラレータの性能を最適化する、新たなアルゴリズムに取り組んでいる。たとえば米国 Massachusetts Institute of Technology (MIT) の Computer Science and Artificial Intelligence Laboratory のチームは、高速かつ効率的な AI の実現にむけ、これまで使われてきた行列の乗算演算を改良する最適化アルゴリズムを開発した。乗算と加算の手順を省き、既存の行列乗算演算の 100 倍、現行の近似法の 10 倍の速度でアルゴリズムが実行できるようになった。まだ低性能の中央処理装置を使った小規模なテストしか行っていないが、付帯的なランタイムなく配置でき、低ビット幅の整数計算に必要な依存コンポーネントをすべて含んでいるという。重要なのは、このアルゴリズムを使った新しいハードウェアを実装すると、現行のアクセラレータに比べて大幅に効率が上がる可能性だ。このイノベーションがソフトウェアやハードウェアの新たなソリューションにつながり、自然言語処理や物体認識など、現在の AI タスクが抱えるパフォーマンス障害の解決に一役買うことになるかもしれない。

最新の方法論の進展に加え、新たな AI モデルや計算アーキテクチャにつながる研究が登場し、AI 分野に破壊的な進歩をもたらすことになりそうだ。たとえば Google と米国 Harvard University の研究グループは、人間の脳の大脳皮質をごく一部ながら検索で

きる 3D マップを開発した。神経回路を正確に表現したこのマップは、人間の脳全体量の 100 万分の 1 を示しているに過ぎないが、細胞 5 万個、シナプス 1 億 3000 万個、その他の様々な細胞構成要素を含み、これまでに作られた最も完成した脳地図の 1 つになっている。さらに大きな神経回路マップの開発には何年も要すると思われるが、今回の小規模マップは、人間の脳構造を擬したニューロモーフィックな AI ハードウェアや新しいアルゴリズムの技術革新を促す可能性をもっている。脳に近いニューロモーフィック・アーキテクチャの採用は、従来のディープラーニング技術よりも優れた推論力と柔軟性を備えた AI モデルを実現する、1つの方法になるかもしれない。とはいえ、ニューロモーフィック技術が破壊的な進歩をもたらす可能性はあっても、その採用には、現在使われているチップやソフトウェア、高性能コンピューティングシステムなどの技術に大幅な変更が必要になる。AI モデルの推論力や柔軟性、精度、スピードが大きく向上すれば、日常的なユーザーの AI システムとの接し方も変わってくる。そうしたシステムはもっと会話的、予測的、直感的に使えるようになるだろう。そのため、研究者たちは、未来の AI システムが労働や日常生活にどう影響するようになるかの概念化に努めている。Google の研究チームは先日、AI 搭載の検索エンジンがいかに信頼性の高い情報を提供し、自然言語の理解を通じて質問をさらに良く理解するようになるかを発表した。具体的には、既存モデルでは回答能力に限界があるため、従来の情報理論と、事前に学習させた大規模な言語モデルの統合を提案している。他の研究グループも同様の戦略を提示しており、この新アプ

ローチがうまくいけば、エンドユーザーに専門レベルの正確な分析結果を提供する、使いやすいクラウド型プラットフォームが実現するかもしれない。こうしたシステムは、正確なオープンソースの情報データに頼って仕事をしている知識労働者やストラテジストにとって、大きな意味を持つだろう。

AI 研究コミュニティは今、企業の複雑なニーズを満たし、一般ユーザーにも使いやすく機能的なツールとなる実用的で信頼性の高い AI を配備する様々な方法論をめぐって岐路に立たされている。一方では、アクセラレータ、データセット、計算システムなど、AI を実現しているあらゆる技術の規模拡張を含む、総当たりのアプローチが AI に大きな進歩をもたらすかもしれない。更に一方で、最先端の AI モデルは、次なる飛躍を実現するために新たな技術パラダイムを必要とする段階に達しているのかもしれない。もしそうなら、型破りなモデルや新たな計算法が AI の破壊的な可能性を引き出し、核となるパフォーマンス障害を解決する鍵になるだろう。いずれの道をとにしても、技術開発者にとっては大きな挑戦となる。AI モデルとデータセットの規模拡張を行えば学習時間が長くなり、大規模モデルを効率的に学習・実行する、強力な専用 AI チップやハードウェアが必要になる。また、新たな AI モデルや計算システムを追求するなら、技術開発者は民間企業にその採用を働きかけねばならず、実用化のハードルは相当に上がってくる。実用的な AI の需要はほぼすべての業界で高まっており、有望な解決策が必ずや出てくるはずだ。

SoC1249

本トピックスに関連する Signals of Change

SoC1243 パーソナルAI
SoC1232 遍在するセンシングとインテリジェンス
SoC1119 AIをトレーニングするAI

関連する Patterns

P1667 AIシステムの制限の緩和
P1631 AIによる競争環境の劇的変化の予兆
P1628 コグニティブなAI を可能に