研究員の眼

大規模言語モデルの裏付け理論

大きいモデルほど高性能 !?

保険研究部 主席研究員　篠原　拓也
(03)3512-1823　tshino@nli-research.co.jp

AI(人工知能)という言葉が日常で用いられるようになって久しい。総務省の「情報通信白書」によると、現在は第三次人工知能ブームなのだという。

このブームは日本では、2000 年代に始まっている。ビッグデータと呼ばれる大量のデータを用いることで AI 自身が知識を獲得する機械学習が実用化された。さらに、知識などの対象を認識する際に注目すべき特徴を定量的に表すことを AI が自ら行って、知識を習得していくディープラーニング(深層学習)が始まった。

そして、2020 年代に入り、大規模言語モデル(LLM)が登場した。これは、ディープラーニングの技術をもとに作られた言語モデルで、人間同士が会話をしたりチャットのやり取りをしたりするのに近いような、流暢な言語処理ができることが特徴となっている。そのためには、相手から発せられた言葉(データ)を解釈して、それに応じて話のテーマの予測を行い、適切に応答することが必要となる。OpenAI 社により 2022 年に導入された ChatGPT などの生成 AI の登場につながっている。

この LLM は、2020 年にジョンズ・ホプキンス大学と OpenAI 社の研究者によって公表された一編の論文(*)にもとづいている。今回は、その論文について少し見ていきたい。

(*) “Scaling Laws for Neural Language Models” Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, Dario Amodei (arXiv:2001.08361 [cs.LG], https://doi.org/10.48550/arXiv.2001.08361)

◇「計算量」、「データ量」、「モデルの大きさ」が大きいほど、自然言語モデルの誤差は小さくなる

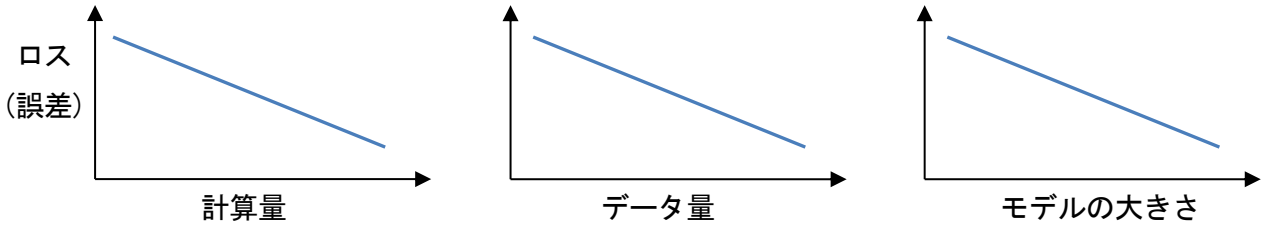
まず、論文のタイトルを見ると、“Scaling Laws for Neural Language Models”となっている。「ニ

ューラル言語モデルのスケーリング則」という訳があてはまるだろう。ニューラルは、人間の脳神経網を模したネットワークを指す。スケーリング則というのは、べき乗則とも言われるもので、さまざまな自然現象に見られるものだ。

例えば、1992 年刊行のベストセラー「ゾウの時間・ネズミの時間」（本川達雄著，中公新書）によると、「いろいろな哺乳類で体重と時間とを測ってみると、（中略）時間は体重の 1/4 乗に比例するのである。（中略）体重が 10 倍になると、時間は 1.8（$10^{1/4}$）倍になる」とのことだ。こうした生物の体重と時間（たとえば寿命）の関係は、べき乗則の一つといえる。

日本語や英語などを用いる自然言語モデルにも、こうしたべき乗則が存在する。「計算量」、「データ量」、「モデルの大きさ」が大きいほど、自然言語モデルの誤差は小さくなる、ということの発見である。それが、この論文のテーマとなっている。

論文では、3 つのグラフが示されている。厳密な単位や尺度等を省略して、筆者がイメージ化して示すと、つぎのような感じになる。



注意が必要なのは、3 つのグラフの縦軸と横軸がいずれも対数スケールであることだ。自然言語モデルの誤差は、計算量、データ量、モデルの大きさが大きいほど、べき乗的に小さくなる。つまり、モデルのディープラーニングにおいて、計算量を増やし、学習データを増やし、モデルのパラメータ数を増やすと、モデルは改善されることとなる。

◇ 大きいモデルほど性能がよくなる、との結論は常識を覆すものとなった

少し、この 3 つのグラフについて考えてみよう。計算量が大きいほど、モデルの性能が高まるというのは納得できる話だろう。たくさん計算するほど、誤差が減るというのは、実感に合っている。

データ量が大きいほど、モデルが改善されるというのも理解できる内容だ。ビッグデータなどの大量のデータを用いて学習するほど、誤差が減少するというのは、「まあそうだろう」との納得感がある。

問題は、モデルの大きさが大きくなるほど、モデルの性能が高まるという点だ。ここでいうモデルの性能というのは、会話のテーマを“予測”する性能だ。ディープラーニングの際に用いた学習デー

タだけではなく、モデルが見たことのない未知データに対して行う予測の性能だ。

従来の考え方では、大きいモデル、つまり大量のパラメータを用いたモデルでは、未知データに対して行う予測の精度は下がってしまうとされてきた。この現象は、「過学習」と言われる。

これを人間の勉強に例えれば、試験前の丸暗記に相当する。数学の試験の前日に、問題を解くための定理や技法ではなく、試験範囲の問題と答えをひたすら丸暗記したとする。試験で、暗記したものとまったく同じものが出題されれば解答はできる。だが、数値や条件などを少し変えた問題が出題されたら、お手上げになってしまうだろう。つまり、未知の問題に対応するための応用力がないわけだ。

こうした過学習を避けるために、モデルの大きさは適切な規模にすべき、との考え方が従来は一般的であった。しかし、この論文は、こうした常識を覆すものとなった。(なお、このべき乗則の成立にはいくつかの条件が必要とされている。べき乗則は、AI の機械学習全般に当てはまるものではなく、トランスフォーマー構造といわれる言語モデルのディープラーニングを条件としている。)

この論文を裏付けとして、OpenAI 社の ChatGPT、Google 社の Bard (2024 年 2 月に Gemini に改称)、Meta 社の LLama など、大手 IT 企業による生成 AI の開発競争が隆盛となっている。

◇ 生成 AI はまだ改良の途上 ― さまざまな問題を抱えている

現在は、生成 AI の性能向上が日進月歩の勢いで進められている。ただ、ここで忘れてはならないのは、生成 AI には、ハルシネーション(幻覚)や、個人情報や機密情報の取り扱いの不備、差別や偏見や攻撃性などの倫理面の問題があることだ。

このうち、ハルシネーションとは、生成 AI がユーザーの質問に対して、事実とは異なる回答を生成することを指す。その内容がもっともらしいために、ユーザーが回答の真偽を確かめにくい。このため、生成 AI の回答を鵜呑みにすると、ユーザーや社会全体に誤解や混乱を巻き起こしてしまう恐れがあるとされる。

ハルシネーションが発生する原因として、学習データが古い情報であったり、偏った情報であったりすることや、学習プロセスに問題があったりすることが考えられる。ただ、その改善は簡単ではないとされる。

AI が人間社会で重要なツールとなっていくためには、予測や判断の性能を向上させるとともに、これらの問題への対応が欠かせない。

◇ 問題への対応のために“ラベラー”による修正が行われている

現在は、生成AIの応答を矯正する方法として、“ラベラー”と呼ばれる人による修正が行われている。ラベラーは、生成AIの応答の偏見や攻撃性の有無を判断したり、人間としての適切な振舞いに照らして生成AIの応答の評価を行う人だ。一般の人が行うこともあるが、これらに精通した専門家が担うことが多い。生成AIが引き起こす問題への対応には、まだ人の手が多くかかるわけだ。

ここからは、筆者の想像する将来の話。多少、妄想も含まれているので、話半分で聞いていただきたい。

将来、ラベラーの役割を肩代わりする“ラベラーAI”が開発されるかもしれない。そうなれば、生成AIの改良に加速度がつくこととなろう。

生成AI(Generative Artificial Intelligence, GAI)の次に、間もなく登場するものとして、汎用AI(Artificial Gereral Intelligence, AGI)が注目されている。GAIの次に、AGIが登場するという流れだ。略語が似ていて、当初は少し戸惑うかもしれない。

汎用AIは、生成AIが進化したもので、さまざまな仕事(タスク)を人間と同等か、それ以上のレベルで実現できるーそれが、汎用AIの特徴とされている。AIやロボット工学の専門家からは、汎用AIは、人々にとって、同僚や相談相手のような存在になるだろうと言われている。

ここまで考えると、ラベラーAIによる生成AIの応答の矯正は、どのように変化していくのかが気になってくる。やはり矯正は、引き続き人間が行うのか。だがそれでは、いつまでも人の手を離れないことになる。

それとも、「ラベラーAIによる生成AIの応答の矯正」を矯正する“ラベラー・ラベラーAI”が開発されるのか。そうなると、ラベラー・ラベラーAIの応答の矯正は、一体誰が行うのか...。

このようなとりとめもないことを考えつつ、生成AIの開発や進化に関する日々のニュースを見ていくのも、なかなか面白いと思われるが、いかがだろうか。

(参考文献)

“Scaling Laws for Neural Language Models” Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, Dario Amodei

(arXiv:2001.08361 [cs.LG], https://doi.org/10.48550/arXiv.2001.08361)

「ゾウの時間・ネズミの時間」(本川達雄著, 中公新書, 1992 年)

「大規模言語モデルは新たな知能かーChatGPT が変えた世界」(岡野原大輔著, 岩波科学ライブラリー 319, 岩波書店, 2023 年)

「平成 28 年 情報通信白書」(総務省)

「大規模言語モデル」「ChatGPT」「ハルシネーション」(ナレッジインサイト 用語解説, 野村総合研究所サイト)

