

報道関係者各位

2024年3月26日
株式会社 Ridge-i

AI・ディープラーニング技術開発のリッジアイ、 目的特化型の日本語 LLM 開発において Llama2 に次ぐ精度の検証結果を達成

AI・ディープラーニング技術のコンサルティングと開発を行う株式会社 Ridge-i（本社：東京都千代田区、代表取締役社長：柳原 尚史、以下「リッジアイ」）は、さくらインターネット株式会社の国産 GPU 開発環境で、目的特化型の日本語 LLM（大規模言語モデル）開発を行い、主要な日本語ベンチマーク(JCommonsenseQA)で Llama2 に次ぐ精度の検証結果を達成しました。

リッジアイは、情報セキュリティ面で安心して使用できる国産の LLM の開発をめざして、2023年7月～2024年1月の期間で開発検証を行いました。研究開発にあたり、必要となる画像処理半導体（GPU）は、国内事業者としてクラウドインフラに強みを持つさくらインターネット株式会社の GPU 環境を採用しました。また、研究開発チームには、自然言語処理分野におけるデータ構築に関しては国内トップの豊富な知見を持つ株式会社バオバブおよびカーネギーメロン大学准教授の Graham Neubig 氏が参画しました。

検証の結果、軽量モデル（モデルサイズ：1.3B、2.7B）で、主要日本語ベンチマークの一つである JCommonsenseQA において、国内の主要 LLM の精度を 20 ポイントほど上回り、Llama2 に次ぐ精度であることを確認しました。（2024年1月時点。検証結果は別紙参照）

今回の検証を通じて、日本語の学習データの質と配分に独自の工夫を行うことによる精度向上のノウハウと、さくらインターネット株式会社の GPU 環境下での LLM 開発ノウハウを得ることができました。今後リッジアイでは、これらのノウハウを活かし、目的特化型 LLM の開発サービスおよびファインチューニングのサービスの提供開始に向けて準備を進めていきます。

株式会社バオバブ 代表取締役社長 相良美織 氏のコメント

あらためてタスクに応じたデータの重要性を確認できた今回の検証結果を踏まえ、ますます高まりを見せるファインチューニングのニーズに、バオバブも国内最多の実績と万全の体制でお応えします。

さくらインターネット株式会社 代表取締役社長 田中邦裕 氏のコメント

リッジアイ社の目的特化型の日本語 LLM が、さくらインターネットの GPU 開発環境にて開発されたことを非常にうれしく思います。さくらインターネットは今後も、さらに高まる AI 需要に応えられるよう、コンピューティングリソースを安定供給確保し、DX プラットフォーマーとしてデジタル社会の継続的な発展へ寄与してまいります。

参考情報

2023 年 7 月 19 日付プレスリリース：[「安心して使える」国産の生成 AI・大規模言語モデル\(LLM\) の開発を開始](#)

以上

株式会社 Ridge-i (リッジアイ) について

リッジアイは、AI・ディープラーニング技術を活用したソリューションにより、経営・社会課題の解決に挑むテックイノベーションファームです。特に、画像解析ディープラーニング、センサーによる異常検知 AI、最適化 AI など様々なデータに対応する AI を組み合わせた「マルチモーダル AI」に強みを持ちます。戦略策定から要件定義フェーズに始まり、現場のコンサルテーションから開発・運用保守まで、投資対効果を実感するまで一気通貫で伴走し、多くの実績をあげています。また社会課題にも積極的に取り組んでおり、SDGs 課題と環境変化を衛星画像から発見する「GRASP EARTH」などで第 4 回、第 5 回、第 6 回宇宙開発利用大賞を 3 連続受賞しました。今後とも技術とビジネスの高みを追求し、社会・顧客が持続的に効果を実感できる最高峰のソリューションを提供します。

会社名：株式会社 Ridge-i

設立：2016 年 7 月

所在地：東京都千代田区大手町 1-6-1 大手町ビル 438

代表：代表取締役社長 柳原 尚史

資本金：10,000,000 円

事業内容：

- ① AI・ディープラーニング技術のコンサルティングおよび開発
- ② 共同事業、ライセンス、保守モデル、自社開発等によるプロダクトの提供
- ③ 人工衛星データ AI 分析サービスの提供

URL：<https://ridge-i.com>

【本件に関するお問い合わせ先】

株式会社 Ridge-i 豊吉、青江

e-mail：pr@ridge-i.com

共同ピーアール株式会社（広報代行）

担当／電話 本田 江代（070-4303-7350）、竹村 良子（080-8870-0347）、児玉 千尋（070-4303-7256）

e-mail：Ridge-i-pr@kyodo-pr.co.jp

<別紙> 今回の検証概要結果について

作成モデルの概要

言語モデルには、因果的言語モデル（causal language model: CLM）を採用しました。CLM は、過去の単語が全て与えられたときに次に来る単語を予測するモデルで、主に文の生成に使用され、代表的なモデルとしては GPT が挙げられます。プロンプトをはじめとした近年の言語モデル関係の技術は CLM の上に構築されています。言語モデルの内部構造には Transformer を採用しました。Transformer は self-attention と feedforward network という 2 種類の基本構造の積み重ねによるモデルであり、近年の CLM の事実上の標準の構造となっています。Transformer 方式の CLM には、NVIDIA が提案する Megatron 系のモデル構造を採用しました。モデルの学習には、Megatron 系モデルを学習するツールとして Eluther AI が公開している GPT-NeoX を使用しました。

学習データの収集と選抜

学習に用いたのは、基幹となる大規模クロールデータ（mC4）に加えて、Wikipedia やニュースサイト、国会議事録などのデータを API 経由で取得しました。本実験では、学習する言語モデルのターゲットタスクをニュース（時事情報）の読解とし、常識や事実関係にやや特化した性能を持つモデルを作成することとしました。学習データは、自然言語 200B トークン、コード 200B トークン程度であり、自然言語データのみを用いた場合でも 10B パラメータ級のモデルの学習に共用可能な量となっています。こうして集めたデータについては、独自の前処理を行い、また様々な配分でデータを統合させています。

トークナイザの学習

トークナイザには SentencePiece を使用し、語彙サイズ 5 万のモデルを学習させました。トークナイザの学習データには、言語モデルの学習データから 3 億文字相当の文をランダムサンプリングしたものを作成しました。

言語モデルの学習

作成する言語モデルのモデルサイズは、1.3B、2.7B、7B、13B としました。このうち CoT などの現代的な LLM の応用が可能となるのは、おおよそ 13B からであり、それより小さなモデルに関しては embedding の取得や簡単な判定問題などへの応用が可能です。

本実験では、最終的に供用された計算機の量の関係から、1.3B モデルと、2.7B モデルの学習が可能でした。学習には、NVIDIA A100 40GB が 4 枚搭載されたマシンをモデルごとに 1 台ずつ占有の上、それぞれ独立した学習器を動作させました。

検証結果

言語モデルの日本語推論の評価に標準的に用いられる JGLUE に含まれる 4 つのベンチマーク（JCommonsenseQA、JNLI、MARC-ja、JSQuAD）で検証した結果、JCommonsenseQA において、両モデルともに精度 60 ポイント以上を安定して達成しました。

▼JCommonsenseQAの入出力例

Question: オーストラリアの隣にある島国は？

Choices:

南極: -14.660587310791016

南太平洋: -13.955972671508789

東南アジア: -19.789459228515625

* ニュージーランド: -12.631352424621582

珊瑚礁: -15.375509262084961

AI 回答: ニュージーランド

* は正解の選択肢、スコアはシステムの出力（0に近いほど高いスコア）

これはリーダーボード（注1）で公開されている国内の主要モデル（Cyber2.7B、rinna1.3など）と比較して20~30ポイント高い水準であり（注2、注3）、Llama 2-13B（精度74.89ポイント）に次ぐ精度であることを確認できました。データ量・パラメータサイズを比較すると、高い精度を達成した理由として、データの配分における独自の工夫が奏功したと考えられます。

検証結果については、さらにチューニングを行い、また研究発表などで共有することも検討しています。

注1 リーダーボード (<https://github.com/Stability-AI/lm-evaluation-harness/tree/jp-stable?tab=readme-ov-file>)

注2 CyberAgent-open-calm-3B (<https://huggingface.co/cyberagent/open-calm-3b>)
パラメータ数 2.7B、JCommonsenseQA 精度 27.79%

注3 rinna-japanese-gpt-1b <https://huggingface.co/rinna/japanese-gpt-1b>
パラメータ数 1.3B JCommonsenseQA 精度 34.76%