News Release

RICOH
imagine. change.

2024 年 9 月 30 日

リコー、モデルマージによって GPT-4 と同等の高性能な日本語 LLM（700 億パラメータ）を開発

～お客様のオンプレミス環境でのプライベート LLM 導入を加速～

株式会社リコー（社長執行役員：大山 晃）は、米 Meta Platforms 社が提供する「Meta-Llama-3-70B」の日本語性能を向上させた「Llama-3-Swallow-70B*1」をベースモデルに、同社の Instruct モデルからベクトル抽出した Chat Vector*2 とリコー製の Chat Vector*3 をリコー独自のノウハウでマージすることで、高性能な日本語大規模言語モデル（LLM*4）を新たに開発しました。これにより、リコーが開発・提供する LLM のラインナップに、米 OpenAI が開発した GPT-4 と同等レベルの高性能モデルが追加されました。

生成 AI の広がりにより、企業が業務で活用できる高性能な LLM のニーズが高まっています。しかし、LLM の追加学習は、コストが高く、時間もかかるという課題があります。その課題に対して、複数のモデルを組み合わせて、より高性能なモデルをつくる「モデルマージ*5」は効率的な開発手法として注目されています。

リコーは、モデルマージのノウハウと、LLM 開発の知見に基づき、今回、新たな LLM を開発しました。本技術は、企業独自のプライベート LLM や特定業務向けの高性能な LLM の開発の効率化につながるものです。

リコーは、自社製 LLM の開発だけではなく、お客様の用途や環境に合わせて、最適な LLM を低コスト・短納期でご提供するために、多様で効率的な手法・技術の研究開発を推進してまいります。

【評価結果*6（ELYZA-tasks-100）】

複雑な指示・タスクを含む代表的な日本語のベンチマーク「ELYZA-tasks-100」において、今回リコーがモデルマージの手法で開発した LLM は GPT-4 と同等レベルの高いスコアを示しました。また、比較した他の LLM はタスクによって英語で回答するケースが見られましたが、全てのタスクに対して日本語で回答して高い安定性を示しました。

企業/組織	モデル	ELYZA-tasks-100			
		「GPT-4」スコア	「GPT-4o」スコア	平均スコア	英語で回答されたタスクの割合 [%]
OpenAI	gpt-4-0613	4.45	3.92	4.19	-
rinna	llama-3-youko-70b-instruct	4.11	3.54	3.83	3
Tokyotech	Llama-3-Swallow-70B-Instruct-v0.1	3.88	3.31	3.60	7
Ricoh	Llama-3-Ricoh-70B-Instruct	4.02	3.37	3.70	0
Ricoh	Llama-3-Ricoh-70B-Merge-Instruct-v0.1	4.43	3.97	4.20	0

ベンチマークツール（ELYZA-tasks-100）における他モデルとの比較結果（リコーは最下段）

株式会社リコー https://jp.ricoh.com/

報道関係のお問い合わせ先　広報室　TEL：050-3814-2806（直通）　E-mail：koho@ricoh.co.jp

お客様の問い合わせ先　仕事のAI お問合せフォーム
https://www.secure.rc-club.ricoh.co.jp/shigoto-no-ai_inq?

【リコーの LLM 開発の背景】

労働人口減少や高齢化を背景に、AI を活用した生産性向上や付加価値の高い働き方が企業成長の課題となっており、その課題解決の手段として、多くの企業が AI の業務活用に注目しています。しかし、AI を実際の業務に適用するためには、企業固有の用語や言い回しなどを含む大量のテキストデータを LLM に学習させ、その企業独自の AI モデル(プライベート LLM)を作成する必要があります。

リコーは国内でもトップクラスの LLM の開発・学習技術をベースに、企業向けプライベート LLM の提供や、社内文書の活用を後押しする RAG の導入支援等、様々な AI ソリューションの提案が可能です。

*1 Llama-3-Swallow-70B:東京工業大学情報理工学院 情報工学系の岡崎直観教授と横田理央教授らの研究チームと国立研究開発法人 産業技術総合研究所によって開発された日本語 LLM モデル。

*2 Chat Vector:指示追従能力を持つモデルからベースモデルのウェイトを差し引き、指示追従能力のみを抽出したベクトル。

*3 リコー製の Chat Vector: Meta 社のベースモデル「Meta-Llama-3-70B」に対し、リコー独自開発を含む約 1 万 6 千件のインストラクションチューニングデータで追加学習した Instruct モデルから抽出した Chat Vector。

*4 Large Language Model (大規模言語モデル):人間が話したり書いたりする言葉(自然言語)に存在する曖昧性やゆらぎを、文章の中で離れた単語間の関係までを把握し「文脈」を考慮した処理を可能にしているのが特徴。「自然文の質問への回答」や「文書の要約」といった処理を人間並みの精度で実行でき、学習も容易にできる技術。

*5 モデルマージ:複数の学習済みの LLM モデルを組み合わせて、より性能の高いモデルを作る新たな方法のこと。GPU のような大規模な計算リソースが不要で、より手軽にモデル開発ができるとして、近年注目されています。

*6 2024 年 9 月 24 日時点の評価結果。「スコア」の算出に際して、生成文の評価には「GPT-4」(gpt-4-0613)と「GPT-4o」(gpt-4o-2024-05-13)を使用し、英語での回答による減点は行っていない。「英語で回答されたタスクの割合」は 100 タスクのうち英語で回答されたものの割合。

■関連ニュース

リコー、日英中 3 言語に対応した 700 億パラメータの大規模言語モデル(LLM)を開発、お客様のプライベート LLM 構築支援を強化
https://jp.ricoh.com/release/2024/0821_1
インストラクションチューニング済みの 130 億パラメータの日本語 LLM を開発
https://jp.ricoh.com/release/2024/0603_1
日本語精度が高い 130 億パラメータの大規模言語モデル(LLM)を開発
https://jp.ricoh.com/release/2024/0131_1

| リコーグループについて |

リコーグループは、お客様のDXを支援し、そのビジネスを成功に導くデジタルサービス、印刷および画像ソリューションなどを世界約200の国と地域で提供しています(2024年3月期グループ連結売上高2兆3,489億円)。

“はたらく”に歓びを 創業以来85年以上にわたり、お客様の“はたらく”に寄り添ってきた私たちは、これからもリーディングカンパニーとして、“はたらく”の未来を想像し、ワークプレイスの変革を通じて、人ならではの創造力の発揮を支え、さらには持続可能な社会の実現に貢献してまいります。

詳しい情報は、こちらをご覧ください。https://jp.ricoh.com/