

人工知能アプリケーションの信頼性・安全性 2

報告書番号：R24JDG20113

利用分野：研究開発

URL：https://www.jss.jaxa.jp/ar/j2024/27136/

● 責任者

清水太郎, 研究開発部門第三研究ユニット

● 問い合わせ先

金子泰之(kaneko.yasuyuki@jaxa.jp)

● メンバ

金子 泰之・伊藤光一郎・石濱直樹

● 事業概要

JAXA 内のドキュメント自動解析に向けて、画像と言語が同時入力となる VLM (Vision-Language Model) の開発を行い、特に図に対する対話が可能なローカルモデルの開発を進めています。実験では、図認識に関する新たな指示学習方法の検討を行い、その初期成果を人工知能学会に投稿しました。

また、本事業では、社内でのローカル LLM (大規模言語モデル) 活用のため、スーパーコンピュータを使用して学習を行い、社内での LLM の有用性を検討しました。大規模な学習データを使用し、効率的なモデル学習を進め、学習したモデルのパフォーマンスを評価しました。今回の学習結果と評価を通じて、次年度に向けた課題を抽出し、今後の研究に向けた重要な基盤を築くことができました。

● JAXA スーパーコンピュータを使用する理由と利点

JAXA のスーパーコンピュータを採用する主な理由は、極めて大規模なデータセットを処理できる計算能力と、機密性の高いデータを処理できる能力を備えていることです。大規模で機密性の高いデータセットを使用しても効率的な学習が行えるほか、JAXA のスーパーコンピュータは、大規模な計算を長時間にわたって実行する際に安定した性能を発揮できます。

● 今年度の成果

本年度は JAXA スーパーコンピュータ (JSS) を活用してローカル LLM のトレーニング・学習を実施し、クローズなモデルと比較した際にローカル LLM の有効性が確認された。この過程で、LLM 学習に必要なライブラリの仕様やコーディング、さらにローカル LLM の学習時の挙動などについて開発担当者がノウハウを蓄積できたことは、次年度以降の開発において大きな強みとなる。また、モデルを迅速に更新できることがわかったため、社内の業務改善に大いに寄与するだけでなく、今後の研究開発に向けた重要な礎となりうることも示唆された。

● 成果の公表

なし

● JSS 利用状況

● 計算情報

プロセス並列手法	非該当
スレッド並列手法	非該当
プロセス並列数	1
1 ケースあたりの経過時間	8 時間

● JSS3 利用量

総資源に占める利用割合^{※1} (%) : 0.00

内訳

計算資源		
計算システム名	CPU 利用量(コア・時)	資源の利用割合 ^{※2} (%)
TOKI-SORA	0.44	0.00
TOKI-ST	445.27	0.00
TOKI-GP	568.63	0.01
TOKI-XM	0.00	0.00
TOKI-LM	212.47	0.02
TOKI-TST	0.00	0.00
TOKI-TGP	0.00	0.00
TOKI-TLM	0.00	0.00

ファイルシステム資源		
ファイルシステム名	ストレージ割当量(GiB)	資源の利用割合 ^{※2} (%)
/home	245.00	0.17
/data 及び/data2	15,310.00	0.07
/ssd	2,510.00	0.13

アーカイバ資源		
アーカイバシステム名	利用量(TiB)	資源の利用割合 ^{※2} (%)
J-SPACE	0.00	0.00

※1 総資源に占める利用割合：3つの資源(計算,ファイルシステム,アーカイバ)の利用割合の加重平均

※2 資源の利用割合：対象資源一年間の総利用量に対する利用割合

● ISV 利用量

ISV ソフトウェア資源		
	利用量(時)	資源の利用割合※2 (%)
ISV ソフトウェア(合計)	0.00	0.00

※2 資源の利用割合：対象資源一年間の総利用量に対する利用割合