課題名　トランザクショナル AI プラットフォーム

課題名　Transactional AI Platform

川島英之
慶應義塾大学環境情報学部

１．研究目的

本研究の目的は、ミッションクリティカルな業務における意志決定を自動化するため、高頻度なデータ更新とリアルタイムな推論を両立する「トランザクショナル AI プラットフォーム」の基盤理論を構築することにあった。従来のシステムでは、データベース（DB）の整合性確保と AI の計算負荷が切り離されていたが、本研究ではこれらを統合し、動的な環境変化に即応することを目指す。AI とトランザクションシステムの統合にあたり、AI の学習フェーズも研究スコープに含めた。学習フェーズで重要な役割を果たす最適化アルゴリズムの高性能化を試み、収束速度を向上させる手法を探求した。

２．研究成果の内容

本研究では最適化アルゴリズム Muon の高性能化に取り組んだ。Muon は、大規模言語モデルの学習加速を目的として開発された。Muon は、従来の AdamW に代表される 1 次モーメントに基づく最適化手法の限界を打破するため、行列の幾何学的構造を直接的に利用して収束を早めるアプローチを採っている。Muon の技術的な核は、Newton-Schulz 反復を用いた勾配行列の直交化プロセスである。これは、更新前の勾配に対してプレコンディショニングを施すことで、パラメータ間の冗長な相関を排除し、勾配を行列空間において直交化する。このプロセスにより、形式的には 1 次最適化の枠組みでありながら、実質的にはヘシアン等の 2 次情報に近い幾何学的特性を近似的に利用することが可能となり、損失関数の曲面に対して効率的な更新方向を選択できる。Muon は GPU クラスタを用いた大規模な分散学習環境に最適化されている。プレコンディショニングに伴う計算負荷を各ノードへ戦略的に分散させることで、通信オーバーヘッドを最小限に抑えつつ、ステップあたりの計算時間を AdamW と同等にする。

Muon は、Newton-Schulz 反復を用いた行列の直交化によって高い収束性能を発揮するが、モデルサイズや計算ノード数が拡大するにつれ、2 つのシステム的ボトルネックに直面する。第一に、直交化に必要となる密行列演算の計算コストが、行列サイズに対して超線形的に増大する。第二に、FSDP（Fully Sharded Data Parallel）等を

用いた大規模分散学習で、各ノードに分割された重みを直交化のために一度集約する必要があり、これが通信オーバーヘッドを引き起こす。

これら 2 点の問題を解決する手法を探求した。現在、論文投稿準備中であるため、詳細記述を控える。

３．学際共同利用プログラムが果たした役割と意義

本研究を遂行するには GPU が必要不可欠である。GPU がなければ研究を遂行できない。そのため、学際共同利用プログラムを利用できたことは、本当に大きな喜びであり、本件に関してご支援を下さったあらゆる人に心から感謝する。

４．今後の展望

期間中に本研究を完了できなかったため、2026 年 4 月以降は一般利用でマシンを利用させて頂いている。成果をまとめて国際会議に投稿予定である。

５．成果発表

(1) 学術論文

(2) 学会発表

(3) その他

使用計算機	使用計算機に○	配分リソース※		
		当初配分	移行*	一般利用による追加
Pegasus	○	800		
Miyabi-G	○	2700		
Miyabi-C	×	0		
	※配分リソースについてはノード時間積をご記入ください。 ＊バジェット移行を行った場合、「+2000」「-1000」のように記入			