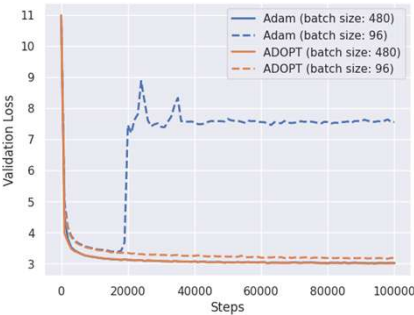


ABCI 3.0開発加速利用（2024年度）成果概要（公開用）

課題名： 深層学習における最適化アルゴリズムの理論と応用	実施時期：2024年4月～2025年3月 所属機関名：東京大学 代表者氏名：谷口 尚平																																													
成果概要： 深層学習の最適化アルゴリズムであるAdamの改良手法ADOPTを開発し、ABCIを利用して、GPT-2スケールの言語モデルでの性能検証を行った。検証の結果、ADOPTがAdamと比較して、性能向上に寄与することが確認され、また学習の安定化にも繋がることを実験的に観測した。																																														
成果のポイント： 確率的最適化アルゴリズムは、深層学習のコアとなる技術要素であり、学習後のモデルの性能に大きく影響を与える。深層学習の領域では、Adamと呼ばれるアルゴリズムが広く用いられてきたが、先行研究において、Adamは理論的に収束を保証できないことが知られており、理論と実践においてギャップが生じていた。本研究では、Adamに軽微な修正を加えることで、その収束を保証できることを理論的に示し、また実践的な設定においても、提案した手法（ADOPT）がAdamよりも良い性能を発揮することを実験的に確認した。特に、Transformerベースの言語モデルの学習において、Adamはロスがスパイクしてしまう現象がしばしば観測されるが、ADOPTは常に安定して学習できることが確認され、近年の大規模言語モデルのような巨大なスケールのモデルの学習においても、ADOPTが最適化アルゴリズムとして有用であることが示唆される成果が得られた。																																														
<table><caption>Approximate data points from the GPT-2 training loss comparison graph</caption><thead><tr><th>Steps</th><th>Adam (batch size 480)</th><th>Adam (batch size 96)</th><th>ADOPT (batch size 480)</th><th>ADOPT (batch size 96)</th></tr></thead><tbody><tr><td>0</td><td>11.0</td><td>11.0</td><td>11.0</td><td>11.0</td></tr><tr><td>10,000</td><td>4.0</td><td>4.0</td><td>3.5</td><td>3.5</td></tr><tr><td>20,000</td><td>7.5</td><td>7.5</td><td>3.2</td><td>3.2</td></tr><tr><td>30,000</td><td>8.5</td><td>8.5</td><td>3.1</td><td>3.1</td></tr><tr><td>40,000</td><td>7.5</td><td>7.5</td><td>3.1</td><td>3.1</td></tr><tr><td>60,000</td><td>7.5</td><td>7.5</td><td>3.1</td><td>3.1</td></tr><tr><td>80,000</td><td>7.5</td><td>7.5</td><td>3.1</td><td>3.1</td></tr><tr><td>100,000</td><td>7.5</td><td>7.5</td><td>3.1</td><td>3.1</td></tr></tbody></table> GPT-2の訓練におけるAdamとADOPTの学習曲線の比較		Steps	Adam (batch size 480)	Adam (batch size 96)	ADOPT (batch size 480)	ADOPT (batch size 96)	0	11.0	11.0	11.0	11.0	10,000	4.0	4.0	3.5	3.5	20,000	7.5	7.5	3.2	3.2	30,000	8.5	8.5	3.1	3.1	40,000	7.5	7.5	3.1	3.1	60,000	7.5	7.5	3.1	3.1	80,000	7.5	7.5	3.1	3.1	100,000	7.5	7.5	3.1	3.1
Steps	Adam (batch size 480)	Adam (batch size 96)	ADOPT (batch size 480)	ADOPT (batch size 96)																																										
0	11.0	11.0	11.0	11.0																																										
10,000	4.0	4.0	3.5	3.5																																										
20,000	7.5	7.5	3.2	3.2																																										
30,000	8.5	8.5	3.1	3.1																																										
40,000	7.5	7.5	3.1	3.1																																										
60,000	7.5	7.5	3.1	3.1																																										
80,000	7.5	7.5	3.1	3.1																																										
100,000	7.5	7.5	3.1	3.1																																										
成果についてより詳細な情報を提供しているWebページ、発表論文などの情報： 論文： https://arxiv.org/abs/2411.02853																																														

GPT-2の訓練におけるAdamとADOPTの学習曲線の比較