

世界最大規模の高品質データセットの構築およびそれを用いた大規模言語モデルの開発

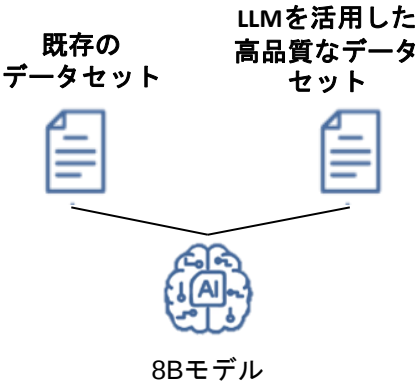
実施者

(株)Preferred Networks/(株)Preferred Elements

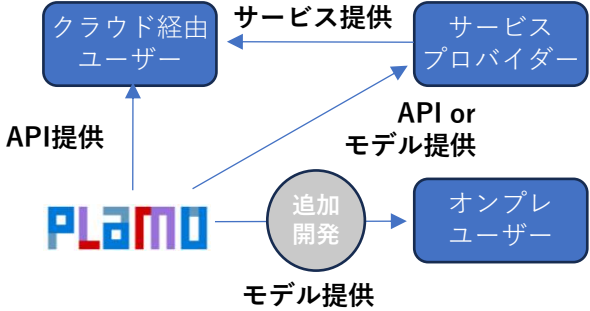
事業概要

LLMを利用し世界最大規模の高品質な学習データを構築し、このデータを含めてフルスクラッチの事前、事後学習を行い、数倍から10倍のサイズのモデルと同レベルの精度を達成する。

高品質なデータ不足に対して、LLMを利用し、世界最大規模の高品質な学習データを構築する。このデータと既存のデータを組み合わせて、フルスクラッチの事前学習を実施し、その後事後学習を行う。このような高品質なデータを大量に利用することで、既存の10倍近く大きなモデルと比較して同レベルの精度を達成する。



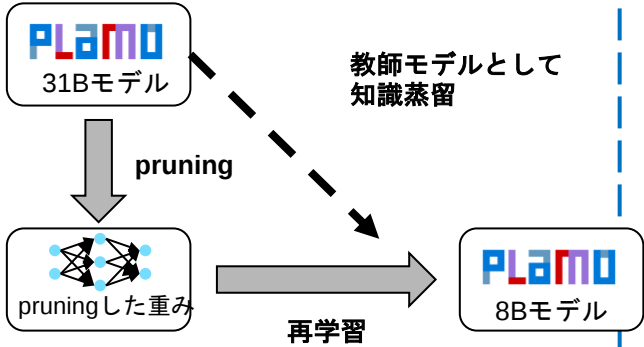
社会実装イメージ



国産かつ柔軟なカスタマイズ能力と提供方法を持つLLMのニーズに対応しつつ、API提供、モデル提供、開発案件を継続的に実施。これまでの複数の外部サービスに連携・公開済みであり、オンプレモデル提供や個別開発に関するPoCも多数実施中

事業成果

- 事前学習済みモデルでState Space ModelとSliding Window Attention を採用したことにより、推論時のメモリ増加を抑えられ、128k tokenの生成が可能
- weight reusingを活用して小さいモデルの重みを次のサイズの重みの初期値として活用して少ない計算リソースで高い精度を達成。
- これに加えて、大きいモデルを活用するpruningと蒸留を利用して、より高い精度のモデルの開発に成功



ベンチマーク	PLaMo 2	比較対象のスコア（モデル名）
JMMLU (5-shot)	0.63 (事前学習済みモデル8B)	0.57 (PLaMo 100B)
JHumanEval (0-shot)	0.70 (事前学習済みモデル8B, デバッカあり)	0.6 (Llama 3 Swallow 70B)
Jaster (0-shot)	0.57(事後学習済みモデル8B)	0.57 (PLaMo-Prime, 100Bモデル)
Jaster (4-shot)	0.63(事後学習済みモデル8B)	0.61(PLaMo-Prime, 100Bモデル)
Japanese MT Bench	7.0(事後学習済みモデル8B)	6.4(PLaMo-Prime, 100Bモデル)