

# 最適化設定：バッチ学習

- 事前学習の安定性とスループットを向上させるため、バッチサイズは2048や4096など大きな数に設定するのが一般的である
- GPT-3やPaLMは学習中にバッチサイズを動的に増加させる戦略をとる
- GPT3のバッチサイズは、3200から320万トークンへ徐々に増加する
- バッチサイズを動的とすることは、学習プロセスを効果的に安定させることが示されている

# 最適化設定：学習速度

- 学習初期には最大値にいたるまで learning rate を線形で徐々に増加させる (Warmup)
- その後、コサイン関数に従って徐々に学習率を低下させていく (Cosine Decay)

