



ある(Aru)'sテクログ

主にプログラミング・AIについて発信するブログ

プログラミング

機械学習

確率・統計

その他



🕒 2024.06.16 📁 機械学習 記事内に商品プロモーションを含む場合があります

ディープラーニング Colab Pro活用術|学習はハイメモリが良い理由

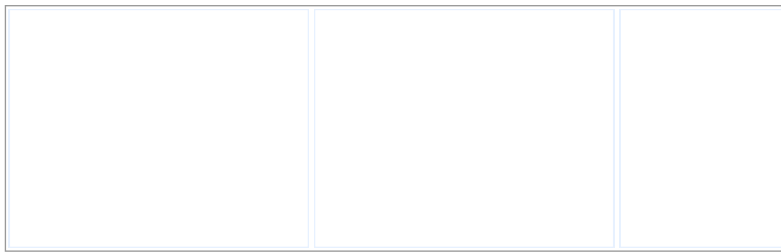
この記事の結論は「Q ディープラーニングではハイメモリを使おう」です。なぜ、ハイメモリが良いのかに興味がある方は記事を読んでみてください。Colab Proのコンピューティングユニットを節約できる情報です。

Contents

CLOSE

- 01 はじめに
- 02 ランタイムのタイプの違い
 - ランタイムとは
 - ランタイムのタイプの違い
- 03 ランタイムはハイメモリがおすすめ
 - ハイメモリはCPUが多い
 - YOLOv8の学習での比較
 - YOLO以外も効果がある
 - TPUのCPU数40も活用できそう
- 04 まとめ

オンラインでディープラーニ...



はじめに

Google Colab Pro, Colab Pro+を契約すると、Q ランタイムを選択することができるようになります。この記事では、実際にQ ディープラーニングに活用する場合のQ ランタイムの選択について説明します

有料契約すると、24時間までノートブックを実行できるようになりますし、切断されにくくなります。また、Colab Pro+ではバックグラウンド処理をサポートしているので、ノートブックを閉じてでも実行を続けてくれます。

私の場合、kaggleのコンペで活用していますが、ハイメモリのQ ランタイムがかなり有効なことに気づきました。この記事では、ハイメモリがなぜ有効なのかについて、実験結果を交えて解説します。

処理の完了と同時に切断したい場合は、こちらの記事を参考にしてください



Google Colabのランタイムをコードで終了させる方法



解約時に余った期間やコンピューティングユニットがどうなるかはこちら



Google Colab Proを解約した場合はどうなるのか？ 加入と解約の手順も解説



ランタイムのタイプの違い

Q ランタイムとは

Colab Pro、またはColab Pro+を契約すると、Q ランタイム環境を選択することができるようになります。

Q ランタイム環境とは、クラウド上でアプリケーションやサービスが実行されることです。Colabでは、Q CPUQ ランタイム、Q GPUランタイム、T...
た構成の異なるQ ランタイムが提供されています。

オンラインでディープラーニ...

ランタイムのタイプの違い

具体的には、Colab Pro（Colab Pro+）では、以下のようなランタイムを選択することができます。

以下の表は、Q ランタイムのメモリ・ディスク・Q CPU数です。



実際に実行して割り当てを確認した結果ですが、公式にスペックを言っているわけではないので、変更される可能性はあります。

	標準（ <u>Q GPU</u> なし）	ハイメモリ（ <u>Q GPU</u> なし）	標準（T4）	ハイメモリ（T4）
1時間あたりのコンピューティングユニット数	0.08	0.12	1.96	2.05
メモリ	12.7GB	51.0GB	12.7GB	51.0GB
ディスク	225.8GB	225.8GB	166.8GB	116.8GB
<u>Q CPU</u> 数	2	8	2	8
<u>Q GPU</u>	-	-	TeslaT4(15GB)	TeslaT4(15GB)

	標準（TPU）	ハイメモリ（TPU）	標準（V100）	ハイメモリ4（V100）	A100
1時間あたりのコンピューティングユニット数	1.96	2.05	5.36	5.45	13.08
メモリ	12.7GB	32.5GB	12.7GB	51.0GB	83.5GB
ディスク	225.8GB	225.8GB	166.8GB	116.8GB	166.8GB
<u>Q CPU</u> 数	2	40	2	8	12
<u>Q GPU</u>	TPU	TPU	V100(16GB)	V100(16GB)	A100(40GB)



記事作成時の価格です（2023年10月）。価格は変動しています。現在の単価とは異なることがあります。

コンピューティングユニット数とは

Pro, Pro+を契約すると、それぞれ100, 500コンピューティングユニットが割り当てられます。およそ、1コンピューティングユニット＝10円だと考えれば良いです。

Q ランタイム環境は、ハイスペックなものを選べば選ぶほど1時間あたりのコンピューティングユニットの消費が大きくなります。現状最もQ GPUの性能が高いA100の場合、1時間で130円くらいのコストがかかります。

オンラインでディープラーニ...

Q ディープラーニングで学習を行う場合、数時間かかることは普通なのでA100を使うとあっという間に割り当てがなくなってしまいます。

とはいえ、生成AIのようにGPUメモリが必要なモデルの学習はA100を使わざるおえないこともあります（A100はVRAMが40GB）。

とはいえ、普段使う場合は、GPUはT4になるのではないかと思います。

ハイメモリはメモリの割り当てが多いランタイムですが、調べてみるとCPU数も増えていることがわかりました。また、GPUではなくTPUのランタイムでは、CPU数が40と突出して多いこともわかりました。

ちなみに、ランタイムの変更は、Colabの「ランタイム」メニューから「ランタイムのタイプを変更」を選択して表示される下記のメニューから行います。



ランタイムのタイプを変更

ランタイムのタイプ

Python 3 ▼

ハードウェア アクセラレータ ?

☐ CPU ☐ A100 GPU ☐ V100 GPU ☒ T4 GPU

☐ TPU

形状

☒ ハイメモリ

キャンセル 保存

ランタイムはハイメモリがおすすめ

ここから本題です。

結論から言えば、「メモリがなくてもランタイムはハイメモリがおすすめ」と言うことです。

ハイメモリはCPUが多い

理由は、CPU数の増加です。

たとえば、T4のランタイムではCPU数が2から8にアップします。

ディープラーニングはGPUを使うのでCPUはそれほど関係ないだろうかもしれませんが、それは間違いです。

オンラインでディープラーニ...

たとえば、Pytorchのデータローダー（DataLoader）では、CPU処理の並列化に対応しています（num_workersで設定）。データローダーを並列化すると、データを読み込んで加工する処理が並列化され高速化します。

この差がどれくらいあるかをYOLOv8の学習で比較してみました。

YOLOv8の学習での比較

実際に、標準（T4）と、ハイメモリ（T4）を比較して学習を行わせてみました。学習は以前紹介した記事のもので、これをColab上で実行した形になります。

あわせて読みたい



YOLO v8/YOLO v9で物体検出 | 独自（カスタム）データの学習と推論を実践



以下が実行時間です。

標準（T4）	14min 41s
ハイメモリ（T4）	6min 42s

ハイメモリの方は半分の時間で終わっていることがわかんと思います（**ハイメモリが約2倍高速**）。

1時間あたりのコンピューティングユニットが、それぞれ1.96と2.05とハイメモリの方が高いですが実行時間が2倍以上短くなっているんで、結果としてハイメモリの方が費用がかかっていません。

	標準（T4）	ハイメモリ（T4）	差分
1時間あたりのコンピューティングユニット	1.96	2.05	+0.09（約4.5%高い）
実行時間	14min 41s	6min 42s	-7min59s（約2.19倍）
実行にかかった費用（コンピューティングユニット）	0.48	0.23	-0.25（約50%ダウン）



実際の計測結果はこちら

標準 CPU times: user 7min 14s, sys: 31.2 s, total: 7min 45s
Wall time: 14min 41s
ハイメモリ CPU times: user 6min 10s, sys: 31.4 s, total: 6min 42s

YOLO以外も効果がある

YOLOv8だけにハイメモリの効果があるわけではありません。

実は、ハイメモリの方がコスパが良いことに気づいたのは、BERTの学習です。

オンラインでディープラーニング

「**CPU数が多いから、もしかしてハイメモリの方が速い？**」と思って試しに**ランタイム**を変更して実行してみると、**12時間以上かかっていた学習が8時間ほどに短縮**されました。YOLOv8ほどではないですが、時間は短くなっているし、必要とするコンピューティングユニットも減りました。

BERTのコードもデータロード中の前処理が少しあったのですが、データの前処理などがある場合は、[Q CPU数の増加による恩恵を得られやすいの](#)かもしれません。

TPUの[Q CPU数40も活用できそう](#)

表形式のデータを[Q 機械学習](#)する場合などは、CPU数が多いと高速に処理できるのでTPUの[Q ランタイムのCPU数が40](#)と言うのは意外と活用できるかもしれません。ただ、価格差を考えると普通のハイメモリ（0.12）の方が良いかもしれません。

まとめ

[Q ディープラーニング](#)を行う場合は、[Q CPUの数](#)も速度に影響があったりします。ハイメモリと標準の価格差がわずかなので、**ハイメモリを選択すると「時間も短くなって、コストも抑えられる」**といいことづくめです。

Colab Proで[Q ディープラーニング](#)を行なっている人は、ハイメモリを活用してはいかがでしょうか。

#Colab Pro #ディープラーニング #ハイメモリ

おすすめ書籍

[最短コースでわかる PyTorch & 深層学習プログラミング](#)

[Amazonで見る 楽天市場で見る Yahoo!ショッピングで見る](#)

この記事にコメントする

ハンドルネーム

メールアドレス（公開されません）

オンラインでディープラーニ…