

複数Colabでtraining

- アイデア
 - 通常の Colab training では learner (GPU) と複数の actor (CPU) が同一マシンで動く。learner と actor の間の通信は GRPC Server/Client (localhost:port) で行われる。GRPC で通信しているだけなので actor は別のマシンで動いていても良いので別の Colab で動かそうというアイデア。
 - 通常 Colab 同士は直接通信できないが ngrok というサービスを使い actor Colab から ssh で learner Colab に ssh portforwarding を介して接続する
- 上記のアイデアは Learner Colab + actor Colab x 4 で動いている。
- 準備
 - [Google Colaboratoryにsshログインをしてお手軽GPU実験環境を作ってみた - Qiita](#) を参考に ngrok にアカウントを作りなんとなくシステムを理解する
 - training したい ckpt 等が入ったディレクトリを用意する
- Learner
 - 共有した GRPC_learner.ipynb を upload する
 - 以前と同じように training したいディレクトリ名から restore_id の部分を自分の環境に合わせる
 - 制限 現時点では actor Colab を ぴったり3台起動する前提で

```
--num_actors=32
```

と設定している。32 = 8 * 3 + 8 である。actor を増やしたい場合は 事前に 32 の部分を書き換えること。これは learner 実装の制限だと思う。深追いはしていない。
 - 実行する
 - 1個目のセルの実行で auth token を入力するがなぜかエラーになる。これは無視する。
 - 2個目のセルで auth token を入れると

```
Root password: CzbdvAKsSv5ioamvviEI
tcp://0.tcp.ngrok.io:11844
```

 - のような表示がある。この password, 0.tcp.ngrok.io の部分。そして : 以降の 11844 が actor からの接続に必要なになる。
- 複数 Actor
 - 前提知識: すべての actor には一意な番号がついています。なので各 actor Colab 毎に自分の actor 番号を設定する必要があります。
 - Actor の数だけ以下のことをやる。
 - GRPC_based_actors.ipynb
 - 2個目のセルに learner から以下の情報をコピーする。これは learner に ssh するための情報。

```
!kill -f ssh
password='QyTADPhHsEZfe72xA0cA'
port=17026
host='4.tcp.ngrok.io'
```

- 前提知識で述べたように actor に番号をつけるために

```
num=$(( actor + 8 ))
```

の 8 の部分を書き換える。1 台目の actor なら 8、n 台目の actor なら 8 x n。

- デバッグ
 - ssh portforwarding が失敗することがあります。nohup.out というファイルにログにログが吐かれています。
 - stderr[0-9].log はそれぞれの actor の標準エラー出力です。

- すべての actors が起動したことを確認して、learner の log を見ると

```
Actor ids needing reset: [3 7 1 ... 13 9 10]
```

のように 32 個の actor 分の初期化が learner で行われたことが分かる。

- これで設定終了