

Colab基本操作筆記

 mattwang44.github.io/en/articles/colab

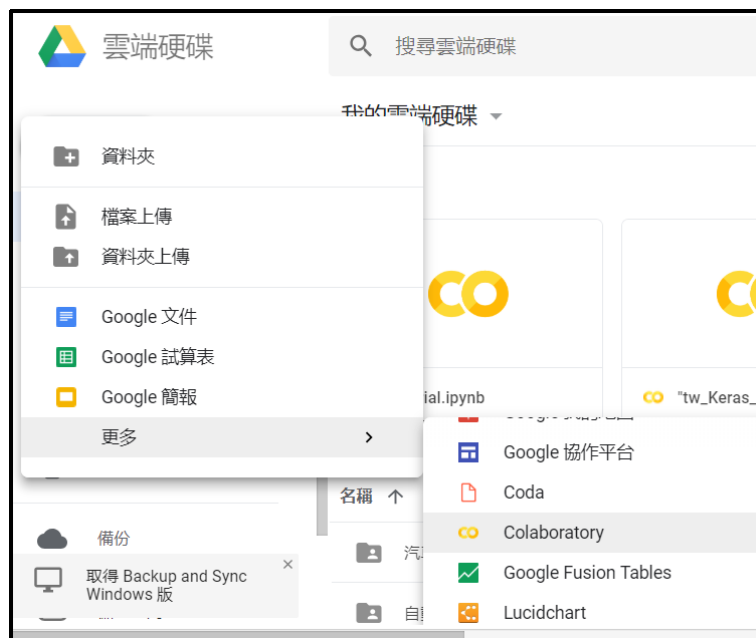
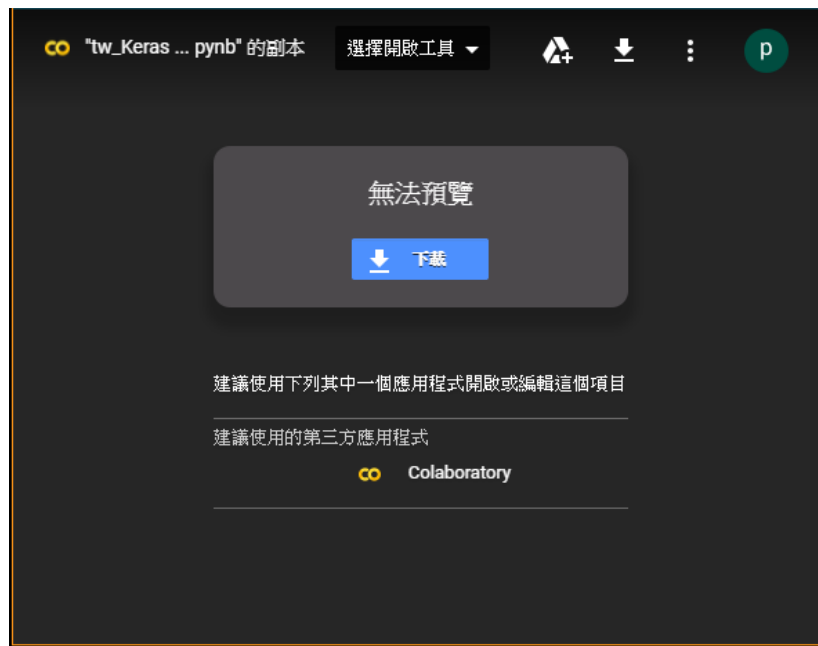
2018年6月25日

是Google提供的雲端Jupyter Notebook開發環境，主要作為Python/機器學習/深度學習教學的輔助工具，最大賣點是提供了**免費**的GPU(型號為Tesla K80 GPU)，雖然虛擬化過後運行速度比直接裝一顆K80還慢(IO時間長)，但是已經比大多數配有中低階GPU的電腦還來的快了(筆者筆電有顆GTX950M，運行上感覺Colab在某些專案還會快上一些)。不過每次開啟有12小時的連續使用時間限制，12小時過後虛擬機需要重新開啟才能繼續運行，也就是說不能在Colab上運行一個需要跑超過12小時的程式。

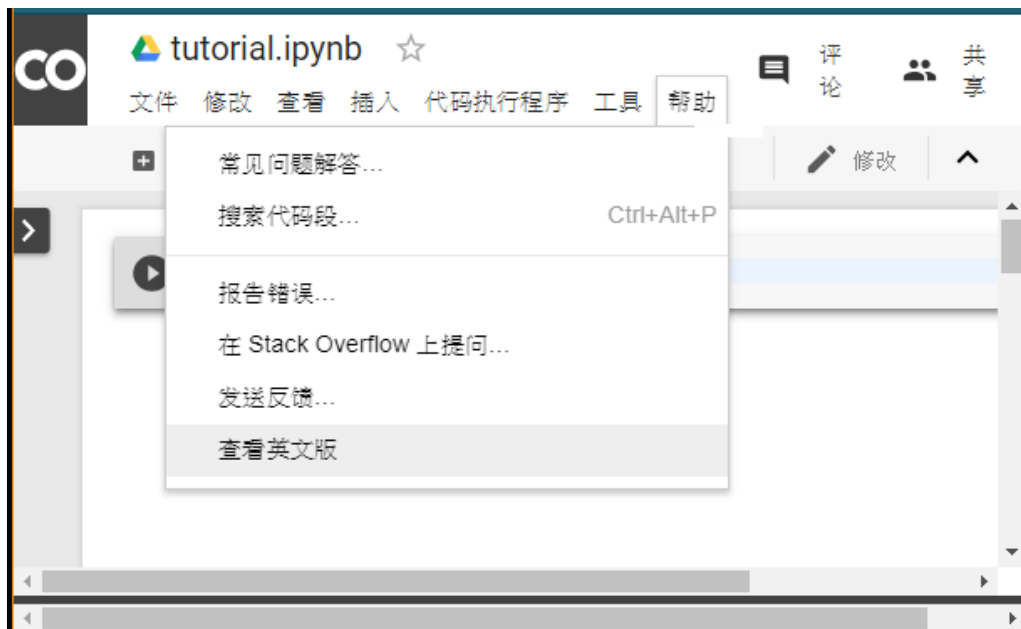
本文只是網路上數篇英文操作教學的翻譯加上幾篇中文教學的總整理。

1. 第一次開啟Colab檔案

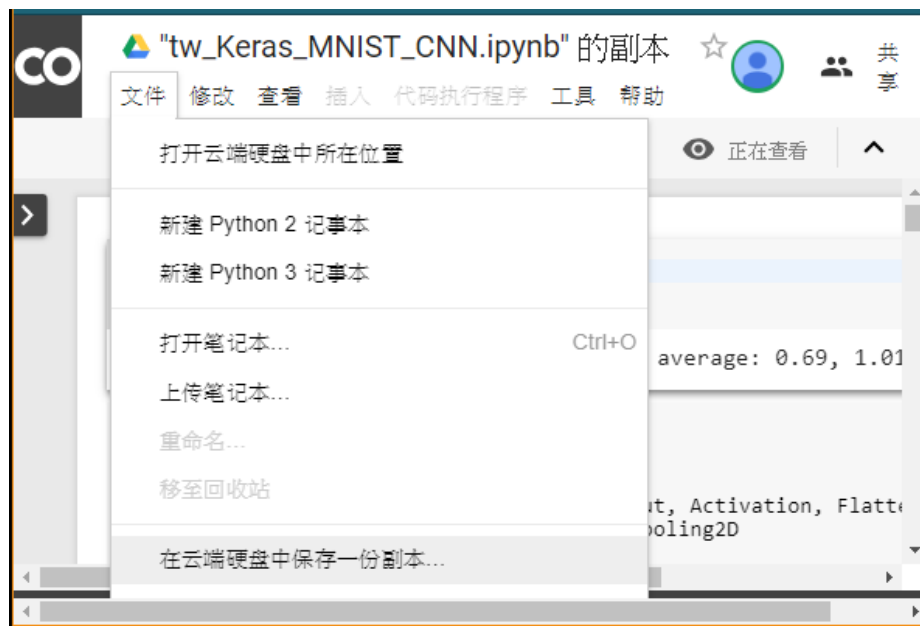
若是開啟Google Drive上分享的檔案，大概都會引導你使用Colab功能(左圖中的建議第三方應用程式)，或是也可以像右圖中從你的Google Drive中開啟一份全新檔案。



目前中文介面似乎只有簡體中文版本而沒有繁體中文，習慣英文操作介面者可以如下圖中操作，或是直接在網址後面加上 `?hl=en` 便會自動改成英文介面。

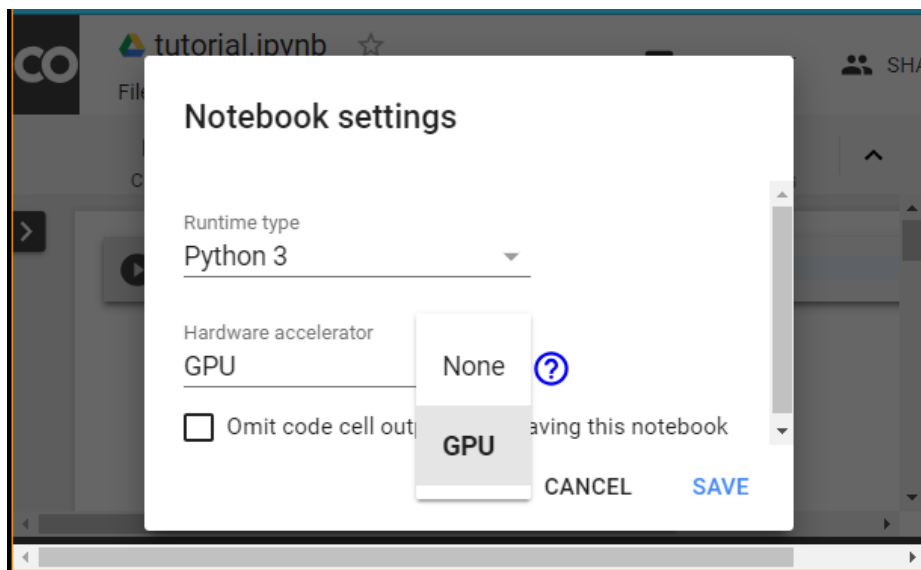
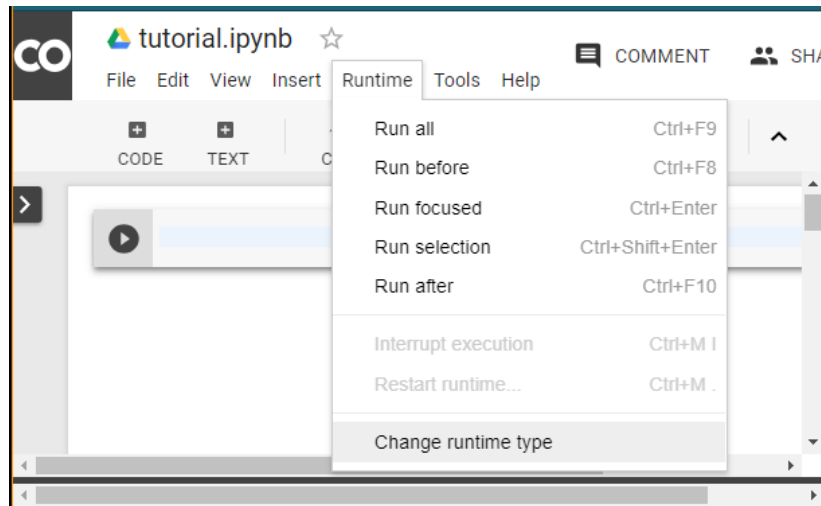


其實檔案操作方法就很像Google Drive其他類型的檔案，擁有閱讀權限者可以看到上面做的所有程式碼編輯，但若直接運行未開放編輯權限的檔案會看到如左下圖中的警告標語。這時可以如右下圖操作把這份檔案複製一份到自己的Drive中，便可自由運行。



2. 選取GPU作為運算單元

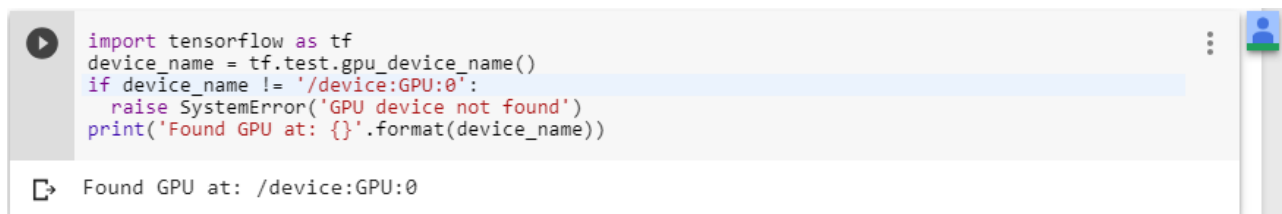
開啟檔案後預設的運行單元為CPU，若要改成GPU則可以按下面兩圖操作。



更改後可以用運行這一段程式碼確認GPU運行正常

```
import tensorflow as tf
device_name = tf.test.gpu_device_name()
if device_name != '/device:GPU:0':
    raise SystemError('GPU device not found')
print('Found GPU at: {}'.format(device_name))
```

若系統有抓到GPU，則會顯示：

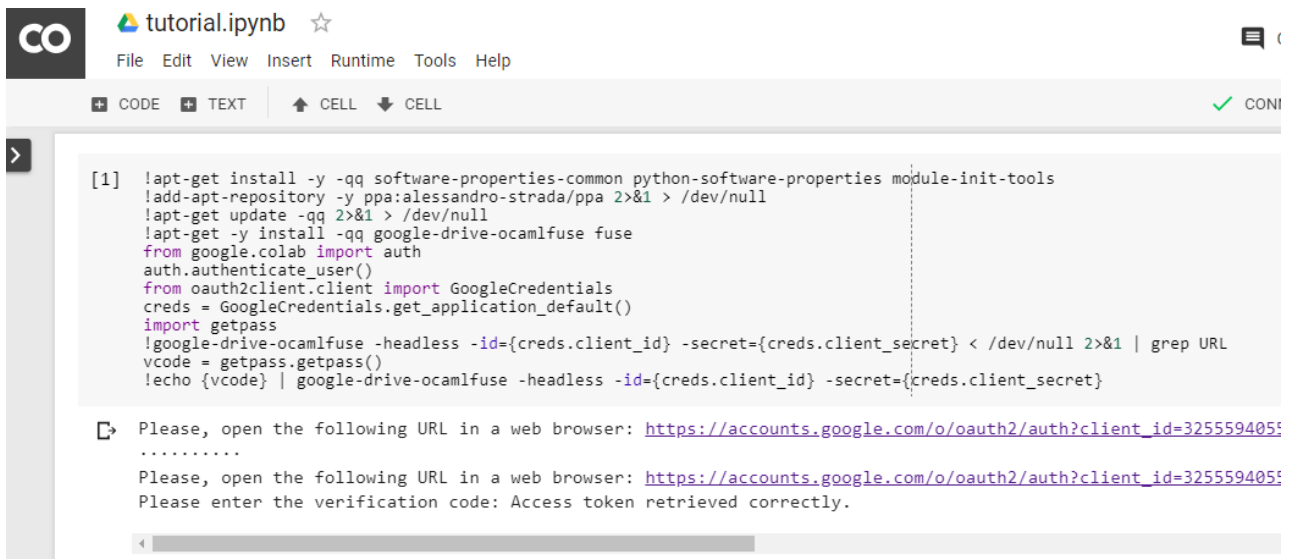


3. 上傳/下載檔案

3.1 開啟權限

開始執行新的一份Notebook建議先運行下面這段程式碼，可以幫助安裝一些需要的package和開啟和Google Drive連動的權限。執行程式碼後會需要你開啟提示網站，複製上面的金鑰並貼回Notebook的Cell中再按enter。完成後會顯示成功開啟權限(如圖)。

```
!apt-get install -y -qq software-properties-common python-software-properties module-init-tools
!add-apt-repository -y ppa:alessandro-strada/ppa 2>&1 > /dev/null
!apt-get update -qq 2>&1 > /dev/null
!apt-get -y install -qq google-drive-ocamlfuse fuse
from google.colab import auth
auth.authenticate_user()
from oauth2client.client import GoogleCredentials
creds = GoogleCredentials.get_application_default()
import getpass
!google-drive-ocamlfuse -headless -id={creds.client_id} -secret={creds.client_secret} < /dev/null 2>&1 | grep URL
vcode = getpass.getpass()
!echo {vcode} | google-drive-ocamlfuse -headless -id={creds.client_id} -secret={creds.client_secret}
```

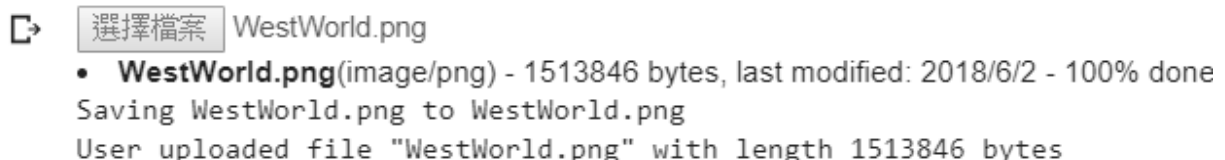


3.2 上傳/下載至Google Drive

以下兩段程式碼功能分別為上傳和下載檔案。

第一段程式碼執行後會在該cell輸出中顯示按鈕，就可以從本機上選擇檔案，上傳後是傳到虛擬機中。

```
from import files
uploaded = files.upload()
for fn in uploaded.keys():
    ('User uploaded file "{name}" with length {length} bytes'.format(
        name=fn, length=len(uploaded[fn])))
```



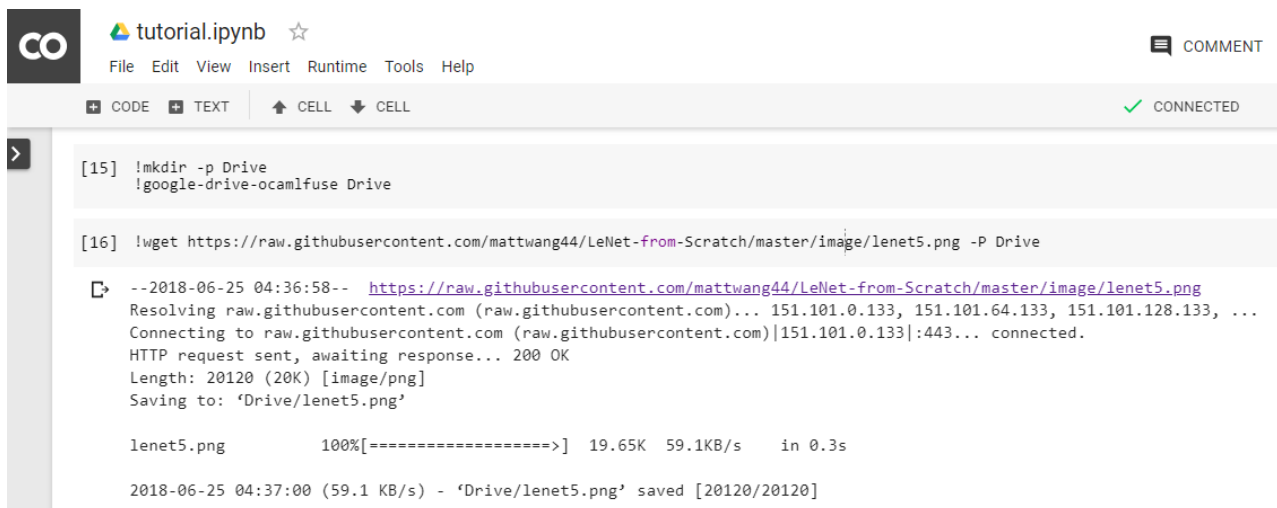
而透過 `files.download()` 便可從虛擬機上下載檔案。

```
files.download('WestWorld.png')
```

而以下程式碼則可以幫我們把虛擬機和自己的Google Drive連接，執行後虛擬機目錄下會多出一個名為Drive(可自行更改)的資料夾，內容即為自己的雲端資料。

```
!mkdir -p Drive
!google-drive-ocamlfuse Drive
```

而大部分Linux中用到的terminal指令(`mv` , `rm` , `wget` , `apt-get` ...etc)，只需要在前面加上 **!** 就可以直接在cell中運行，如這裡展示用來下載網路檔案的wget指令，用後綴 `-P` 指定存放路徑，下圖為下載我GitHub上的一張圖到Drive中，Drive便會即時顯示這份檔案。



```
[15] !mkdir -p Drive
!google-drive-ocamlfuse Drive

[16] !wget https://raw.githubusercontent.com/mattwang44/LeNet-from-Scratch/master/image/lenet5.png -P Drive

--2018-06-25 04:36:58-- https://raw.githubusercontent.com/mattwang44/LeNet-from-Scratch/master/image/lenet5.png
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 151.101.0.133, 151.101.64.133, 151.101.128.133, ...
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|151.101.0.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 20120 (20K) [image/png]
Saving to: 'Drive/lenet5.png'

lenet5.png      100%[=====] 19.65K  59.1KB/s   in 0.3s

2018-06-25 04:37:00 (59.1 KB/s) - 'Drive/lenet5.png' saved [20120/20120]
```



3.3 從GitHub下載

若要開啟的是一份存放在GitHub上的Notebook，可以直接更改網址便能自動以Colab開啟。例如Notebook連結為：

<https://github.com/<一大串東西>.ipynb>

能用以下網址開啟：

而若是要複製整個repo，也可在Colab上使用 `!git clone` 指令，執行後也是存放到虛擬機上。不過要注意，若將東西從虛擬機移到自己雲端所要花費的時間頗長。

```
[32] # Clone a GitHub Repo to virtual machine
!git clone https://github.com/wxs/keras-mnist-tutorial.git

Cloning into 'keras-mnist-tutorial'...
remote: Counting objects: 26, done.
remote: Total 26 (delta 0), reused 0 (delta 0), pack-reused 26
Unpacking objects: 100% (26/26), done.

[44] !ls

datalab  Drive  keras-mnist-tutorial  WestWorld.png
```

3.4 查看目錄方法

`!ls` 可以查看目前所在資料夾的目錄，若要查看特定資料夾之目錄可以用 `!ls <資料夾路徑>`，或是先用Python code先移動到特定資料夾再下 `!ls` 指令：

```
import os
os.chdir("<資料夾路徑>")
```

```
!ls
```

4. 錯誤處理

若是出現 `Warning: apt-key output should not be parsed (stdout is not a terminal)` 警告標語，代表和自己雲端連接的權限已經到期，重新執行這段code就可以重新連接：

```
!mkdir -p drive
!google-drive-ocamlfuse drive
```

出現其他錯誤，可以執行下面這一行，斷開連結，重新連接虛擬機(主要使用時機為清除佔據記憶體數的變數)：

```
!kill -9 -1
```

5. 安裝package

Keras

```
!pip install -q keras
import keras
```

PyTorch


```
from os import path
from wheel.pep425tags import get_abbr_impl, get_impl_ver, get_abi_tag
platform = '{}{}-{}'.format(get_abbr_impl(), get_impl_ver(), get_abi_tag())
accelerator = 'cu80' if path.exists('/opt/bin/nvidia-smi') else 'cpu'
!pip install -q http://download.pytorch.org/whl/{accelerator}/torch-0.3.0.post4-{platform}-
linux_x86_64.whl torchvision
import torch
```

或 `!pip3 install torch torchvision`

MxNet

```
!apt install libnvidia-ml-dev
!pip install mxnet-cu80
import mxnet as mx
```

OpenCV

```
!apt-get -qq install -y libsm6 libxext6 && pip install -q -U opencv-python
import cv2
```

XGBoost

```
!pip install -q xgboost==0.4a30
import xgboost
```

GraphViz

```
!apt-get -qq install -y graphviz && pip install -q pydot
import pydot
```

6. Reference



Wei-Hsiang Wang

Grad. student, Dept. of Mechanical Engineering, National Taiwan University, Taiwan

Updated: June 25, 2018

Comments
