Phase 2 Model Training Documentation

ST1506: DSDA Final Year Project

Company Profiling Text Mining

August 2021

SINGAPORE POLYTECHNIC | SP

Synopsis

The purpose of this documentation is to demonstrate the steps a user should take to successfully run the ipython file for model training.

This documentation is for data scientists who want to see what the process of training the model is, and possibly to use their own data to train the model.

Table of Contents

Running The Notebook - GPU

1. Upload “NLP_Bert.ipynb” to Google Collaboratory (i.e. Google Colab).
2. Under the Runtime tab, change runtime type to “GPU” and click “SAVE”. This will allow for faster running times when running the notebook so that time can be saved.

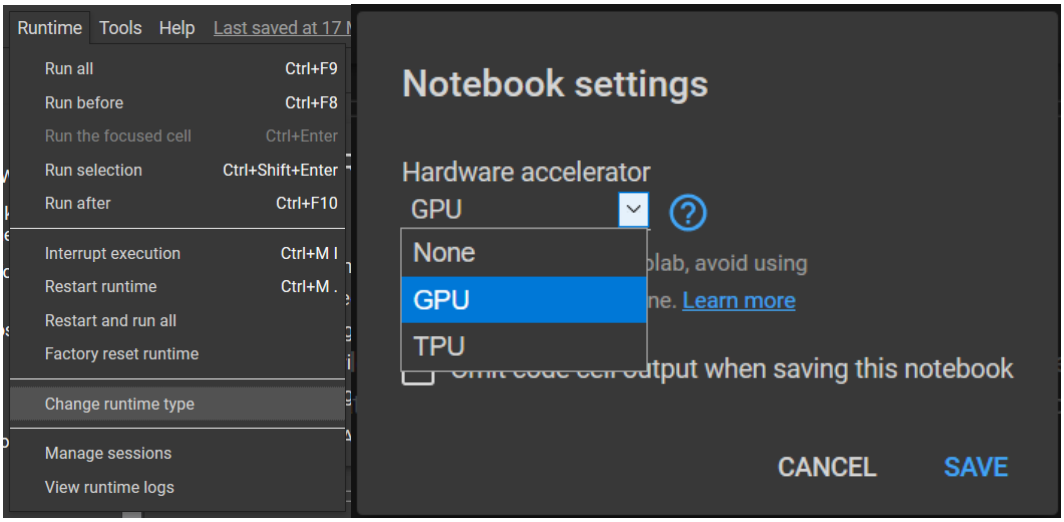

Figure 1: Changing Runtime to GPU

3. Connect to Colab runtime by clicking on the connect button found on the top right bar if not already converted. You should see the following if the connection is successful.

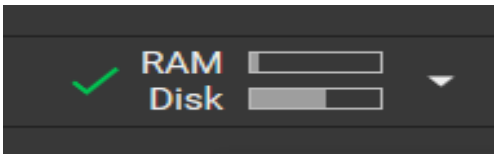

Figure 2:Ensuring runtime is connected

4. On the left tab, click on the icon of a folder and upload “clean_dataset.xlsx”, “sector_master_definition.xlsx” and “val_dataset.xlsx”. Once uploaded you should see that the files are in the temporary working directory in Google Colab.

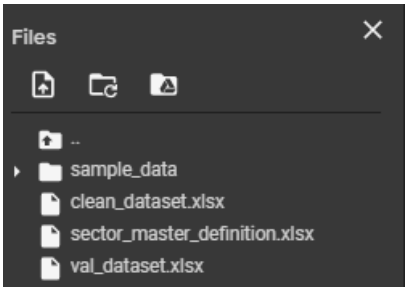

Figure 3: Uploaded excel data files

5. Now, the notebook is ready to run. Run the notebook cells one at a time starting from the start of Section 1, labelled as “Data Importing” until the end of Section 4, labelled as “BERT Model”.

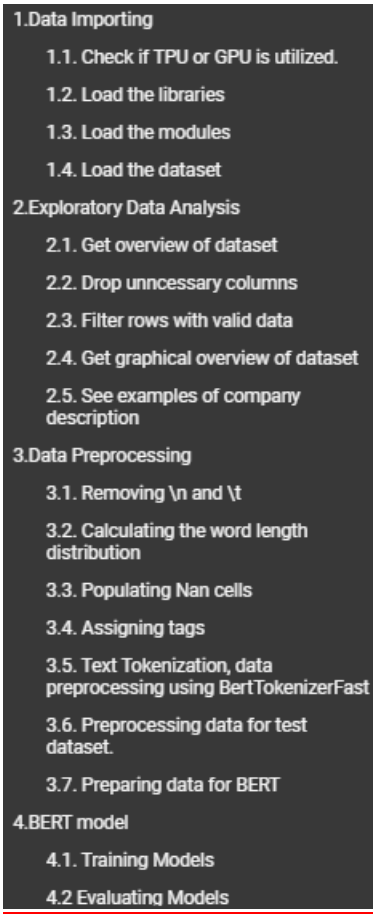

Figure 4: List of Contents to run

While running cell 1.1., labelled "Check if TPU or GPU is utilized", please ensure the cell output is as of below. Otherwise, check the Hardware accelerator used in the current session.

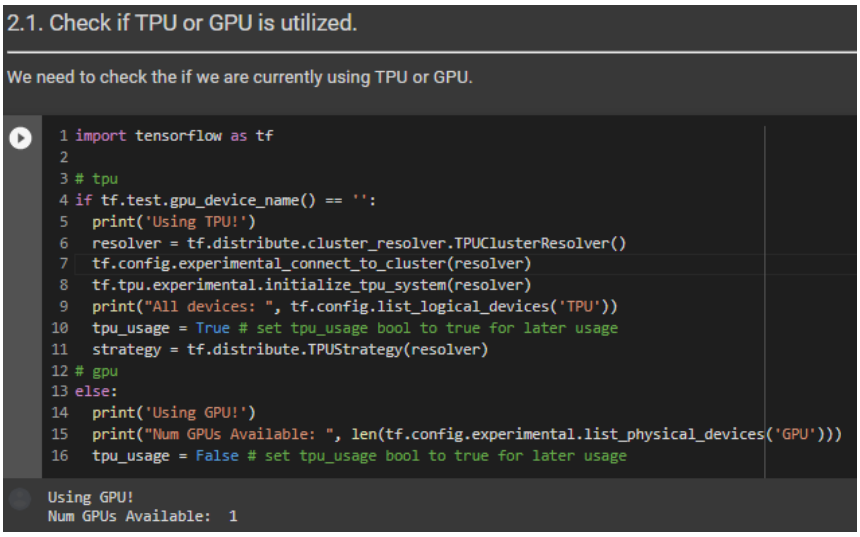

Figure 5: Sample Cell Output for Section 1.1.

6. To save the model, run *Section 5: Saving Model* of the notebook.

Running The Notebook – TPU

1. Similar to that of Section 1 of this documentation, upload “NLP_Bert.ipynb” to Google Collaboratory (i.e. Google Colab).
2. Under the Runtime tab, change runtime type to “TPU” and click “SAVE”. This will allow for even faster running times when running the notebook so that time can be saved.

Note: While using TPU may deliver the training result faster than GPU, it is impossible to export the trained model using TPU.

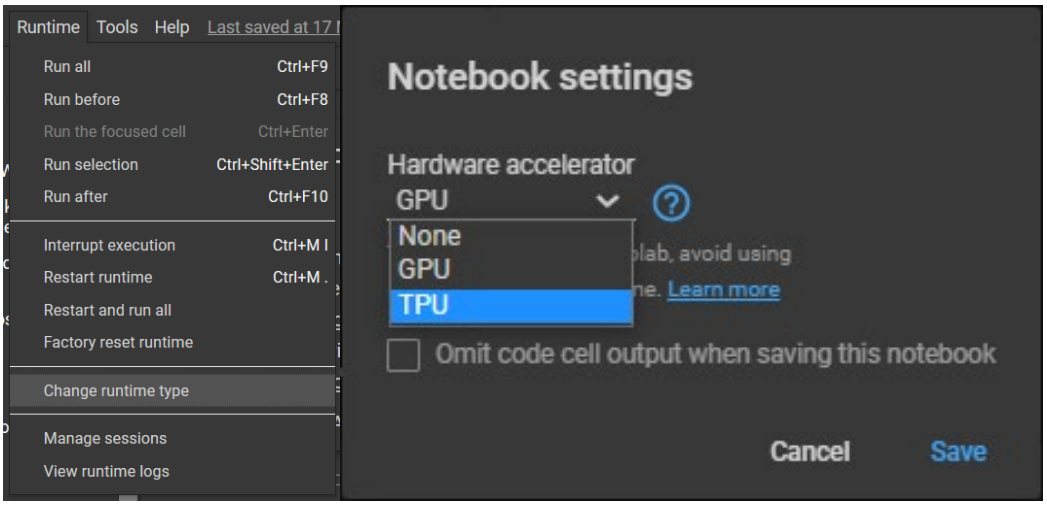

Figure 6:Changing Runtime to TPU

3. Repeat step 3 to 5 in *Section 1: Running The Notebook – GPU.*
4. While running cell 1.1., labelled “Check if TPU or GPU is utilized”, please ensure the cell output is as of below. Otherwise, check the Hardware accelerator used in the current session.

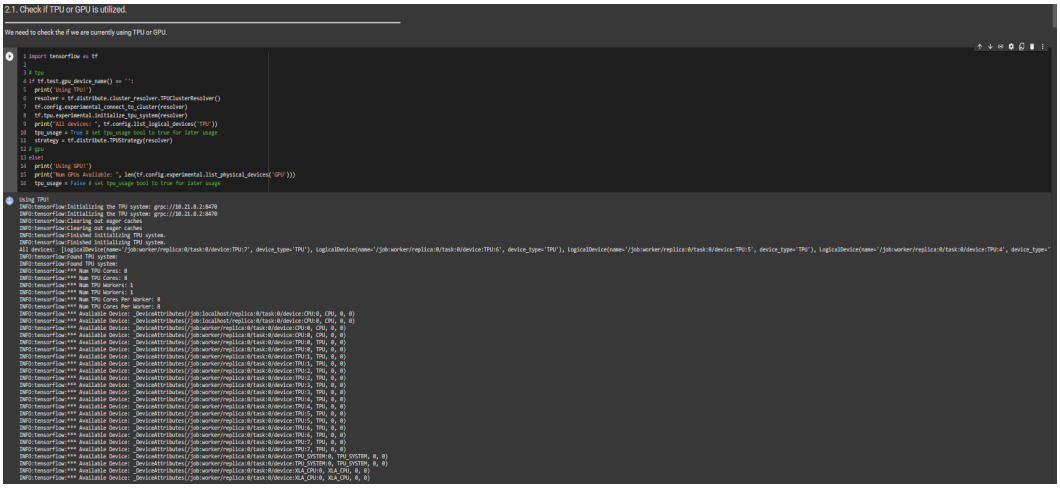

Figure 7:Sample Cell Output for Section 1.1.