README *for* Beyond Embedding: Finetuning Transformer Models in Psychological Research*

Yu Wang and Wen Qu

Institute for Advanced Study in Social Sciences, Fudan University, Shanghai, China

This README file contains three sections. The first section, *General Description*, describes the contents in this replication folder. The second section, *Hardware Requirements*, describes what hardware is used for running the scripts. The third section, *Instructions*, reports on how to run the scripts.

1 General Description

The replication package contains the following files:

- README.pdf
- BERT_Table_3.R: this file generates the results for the *Text* package for BERT results for Table 3.
- RoBERTa_Table_3.R: this file generates the results for the *Text* package for RoBERTa results for Table 3.
- BERT_Table_4.R: this file generates the results for the *Text* package for BERT results for Table 4.
- RoBERTa_Table_4.R: this file generates the results for the *Text* package for RoBERTa results for Table 4.
- BERT_PM_A_V1.ipynb: this file generates the finetuning results using the BERT model for Table 3.

*To whom correspondence should be addressed: Wen Qu. E-mail: wqu@fudan.edu.cn.

- RoBERTa_PM_A_V1.ipynb: this file generates the finetuning results using the RoBERTa model for Table 3.
- BERT_PM_B_V1.ipynb: this file generates the finetuning results using the BERT model for Table 4.
- RoBERTa_PM_B_V1.ipynb: this file generates the finetuning results using the RoBERTa model for Table 4.
- Binary: this folder contains the raw data and the intermediate results (e.g. embeddings) for Table 3.
- Regression-Anxiety: this folder contains the raw data and the intermediate results (e.g. embeddings) for Table 4.

2 Hardware Requirements

The R scripts are tested on a MacBook Pro with an Apple M1 Pro chip and 16GM of memory. The Jupyter Notebooks are tested for running on A100 GPUs on Google Colab.[1]

Exact performance metrics would vary to some extent depending on the type of GPUs and CPUs used, but they would not affect the key results. Exact performance metrics should be replicable on the same type of GPUs.

3 Instructions

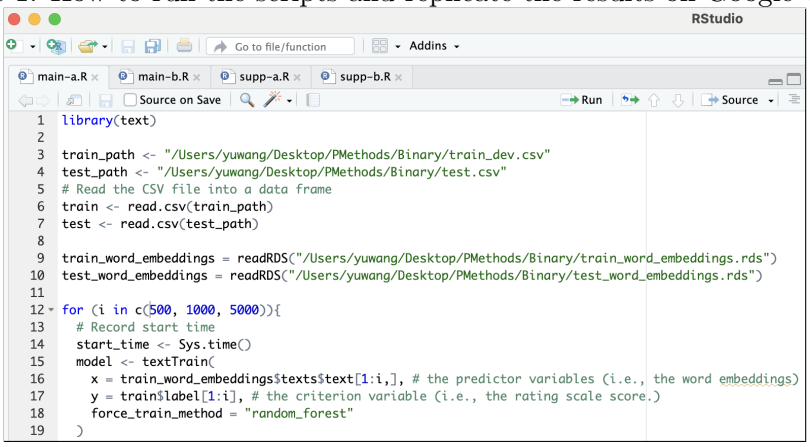
R Execution

Each of R script can be run independently. In RStudio, we only need to hit the **Run all** button under **Runtime** (Figure 1).[2] All the key results will be printed out during execution.

[1] Note that usage of A100 GPUs on Google Colab is not free. Usage of some other types GPUs (e.g. T4) is free.

[2] Note that we can also run the scripts interactively if readers so prefer.

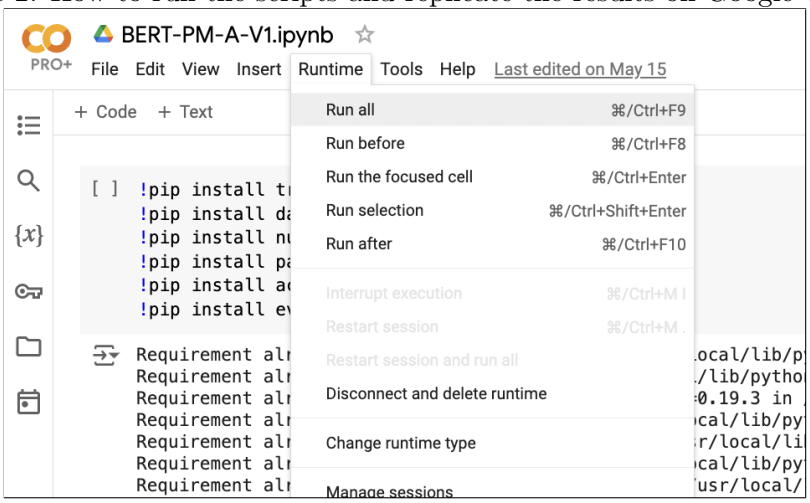
Figure 1: How to run the scripts and replicate the results on Google Colab.



Python Execution

Each of Jupyter Notebook above can be run independently. In Google Colab, we only need to hit the **Run all** button under **Runtime** (Figure 2).[3] All the key results will be printed out during execution.

Figure 2: How to run the scripts and replicate the results on Google Colab.



[3] We can also run the scripts interactively if readers so prefer.