

CS 410 Final Documentation

Applying BERT on Facebook Children's Book Test

evanluo2 (Evan Luo), chiachi5 (Chia-Chi, Chen), tiehchu2 (Tieh Chu)

1) An overview of the function of the code (i.e., what it does and what it can be used for).

Our task is to apply the BERT model on Facebook Children's Book Test Dataset. In the dataset, each question is constructed by taking 20 consecutive sentences from the book text and leaving the 21st as the query statement. A word from the query is selected and masked, and the model is tasked with selecting which word from the text (of the chosen type) should be used to fill this placeholder in the query.

The python code will be able to ingest a file with a set of questions. After parsing all the questions correctly, it would output the best word as the answer for each question.

2) Documentation of how the software is implemented with sufficient detail so that others can have a basic understanding of your code for future extension or any further improvement.

First, the software loads the tokenizer and tokenizes the input. Then it masks a token that we will try to predict back. After that, we convert tokens to vocabulary indices then to PyTorch tensors. Finally, these tensors are fed into the model for predictions using the BERT model we trained.

3) Documentation of the usage of the software including either documentation of usages of APIs or detailed instructions on how to install and run the software, whichever is applicable.

First, clone our repository from github.

```
git clone git@github.com:daisy91530/CourseProject.git
```

Second, install the required packages. (pytorch, tqdm, boto3, requests, nltk)

```
pip install -r requirements.txt
```

Third, run our code with the following command.

```
python final_project.py
```

4) Brief description of the contribution of each team member in the case of a multi-person team.

The following are the contributions of each team member.

Task	Responsible Member
Parse and clean raw data	Evan Luo, Tieh Chu
Train and tune BERT model	Evan Luo, Chia-Chi
CBT answering model implementation	Chia-Chi, Chen, Tieh Chu
Fine-tune and demonstration	Evan Luo, Chia-Chi, Chen, Tieh Chu
Finish all documentation	Evan Luo, Chia-Chi, Chen, Tieh Chu