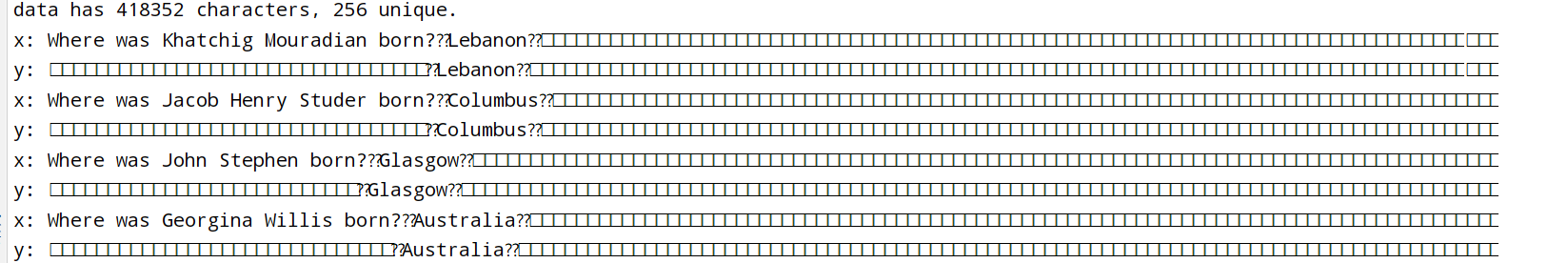
Self-Attention, Transformers, and Pretraining作业

2. Pretrained Transformer models and knowledge access

(a) Check out the demo.

(b) Read through NameDataset, our dataset for reading name-birthplace pairs.

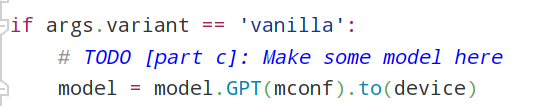
python src/dataset.py namedata

Result: 

(c) Implement finetuning (without pretraining).

Ⅰ：run.py文件57行：

model = model.GPT(mconf).to(device)



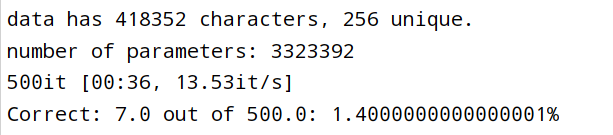
Ⅱ：增加预训练模型vanilla的超参数，run.py代码134-173

(d) Make predictions (without pretraining)

run the following three commands:

Result:

1 dev



2 test\_input

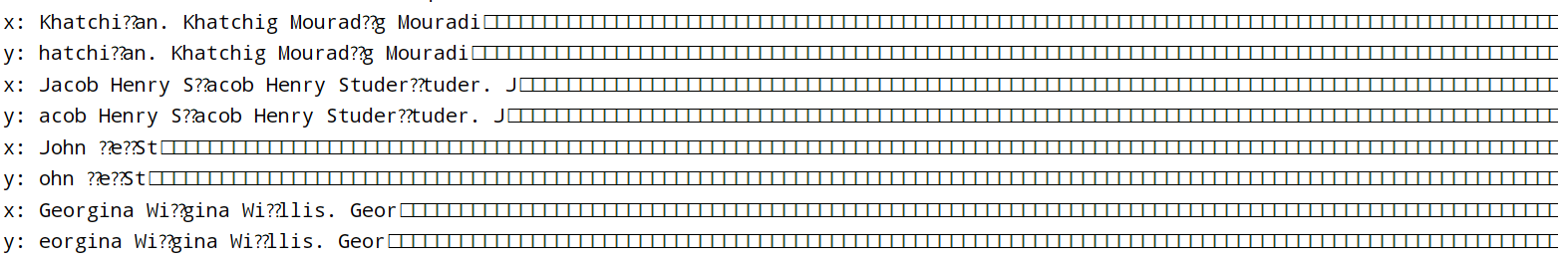


Changes:与c相同

(e) Define a span corruption function for pretraining.

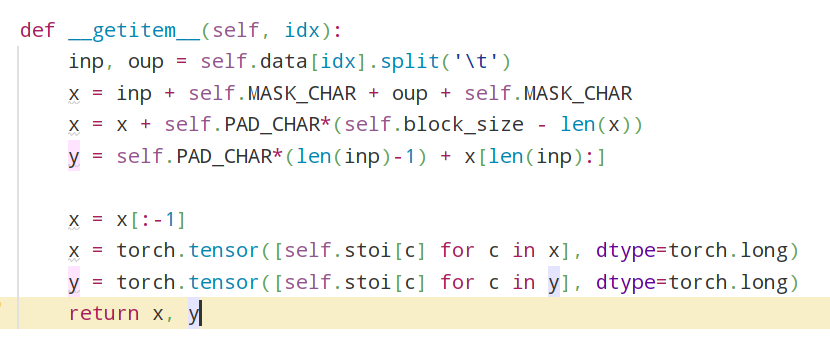
python src/dataset.py charcorruption

Result:



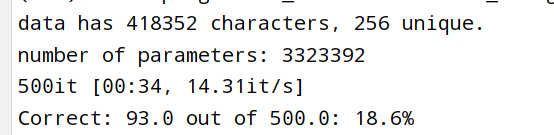
Changes:

在dadaset.py文件中重写NameDataset类，添加\_\_getitem\_\_方法。



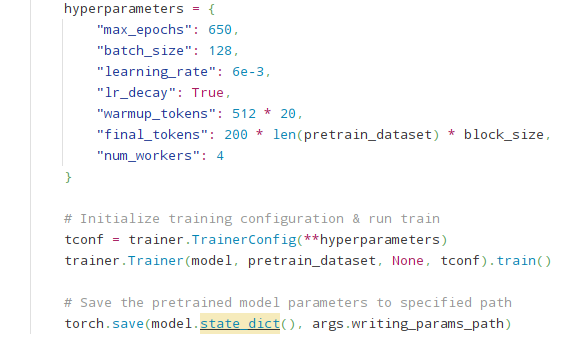
(f) Pretrain, finetune, and make predictions. Budget 2 hours for training

Result:



Changes:

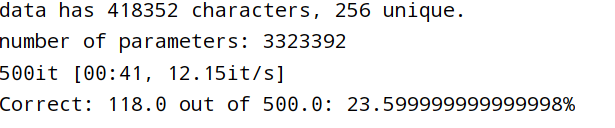
在run.py中添加训练模型超参数



(g) Research! Write and try out the synthesizer variant (Budget 2 hours for

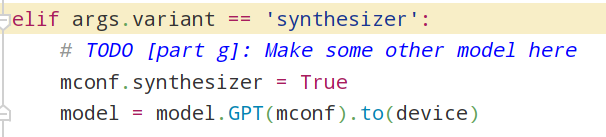
pretraining!)

Result:



Changes:

添加模型synthesizer：



3. Considerations in pretrained knowledge (5 points)

Answer:

1. The pretrained model was able to capture the contextual dependencies between words and understand how different parts of a sentence are related to each other. On the other hand, the finetuned model was specifically trained to extract specific parts of the sentence without having a deep understanding of their relationship to the overall input. Additionally, the larger pretraining dataset allowed the pretrained model to potentially memorize and encode the relevant patterns and information present in the questions.
2. 1. This model behavior may inadvertently lead users to incorporate incorrect information into their work. For instance, when referencing the birthplace of a well-known individual, users may unknowingly include a false birthplace if the model fails to provide accurate information. This has the potential to compromise the accuracy and reliability of their work.

2. The inability to distinguish between retrieved and fabricated information by the model can result in the dissemination of false facts by users. For instance, if users rely on the model's output to learn about famous individuals and unknowingly encounter false birthplace information, they may inadvertently propagate erroneous details about these individuals. This can create confusion and disagreements within society, leading to a distorted understanding of historical events or public figures.

1. The model employs a strategy of maximizing relevance based on its parameters when predicting the birthplace for a given name. This strategy involves searching for information related to individuals with similar names to the provided name. While this functionality can be beneficial in cases where there is a typo in the name, it becomes problematic when a new name is only similar to one of the names the model has been trained on. In such cases, the retrieved information may be inaccurate and can potentially give rise to concerns regarding the quality of the predictions and their impact on social contexts, as mentioned in 3b.