Machine Translation English to German

By Yue Zhao and Cindy Lin

INTRODUCTION

Task: Machine Translation from English to German

Experimental setup:

Data Selection: WMT14 English - German

- Size: 1.4GB
- Training set: 800 thousands bilingual sentence pairs
- Test set: 2000 examples
- Validation set : 2000 examples

Preprocessing:

- First model: Pre-trained BERT tokenizer + target tokenizer in hw5
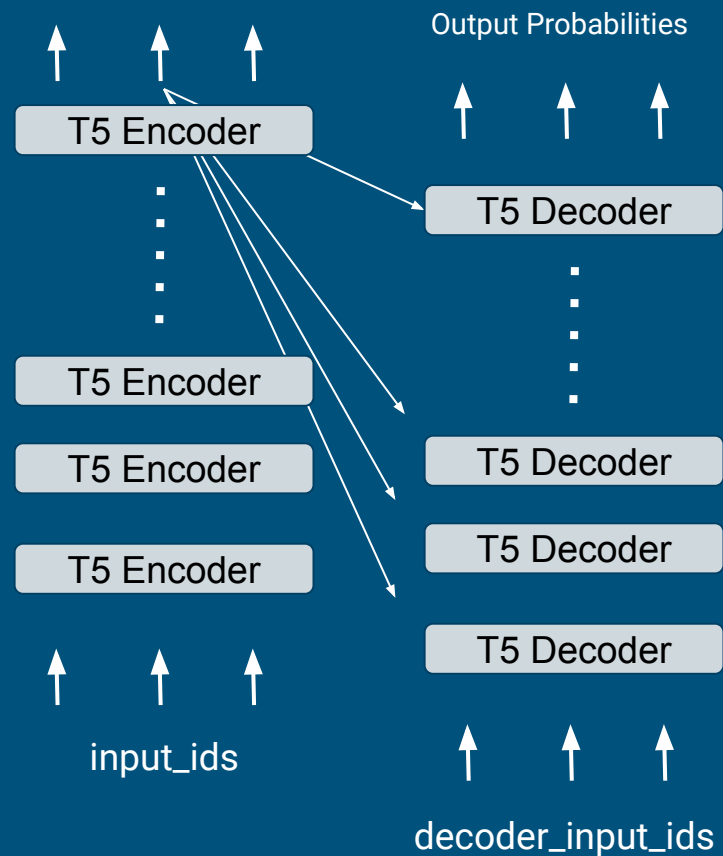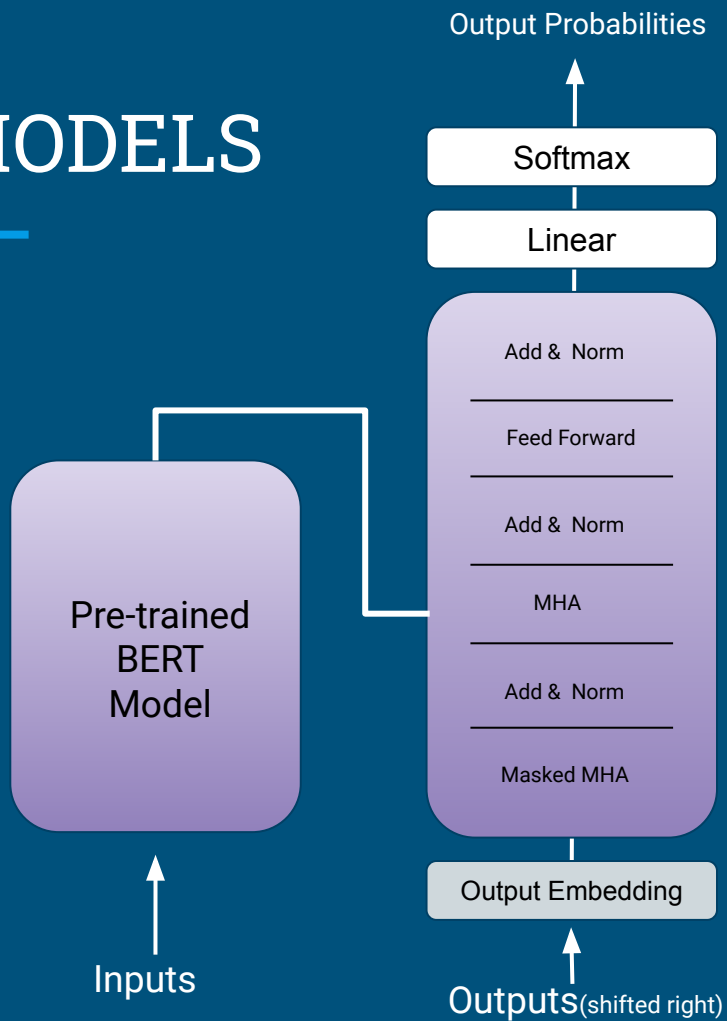- Second model: pre-trained T5 tokenizer

Architecture:

- First model: Pre-trained BERT encoder + transformer decoder
- Second model: Fully pre-trained T5-base model + fine tuning

Metric:

Bleu, No human evaluation

Input and Output example: I don't speak German—> Ich spreche kein Deutsch

MODELS

RESULTS

After **318** runs on GCP, we got these results: