

report

June 4, 2023

1 Names and Net IDs

- Kyle Hwang (ksh6947)
- Michael Lin (qlb968)
- Dylan Wu (dwg0364)

2 Abstract

Our final project seeks to use `pytorch` to replicate the “Attention is All You Need” paper, which introduced the Transformer as a way to improve upon existing sequence transduction language models. We will attempt to implement the model’s architecture and train the model on a subset of the WMT 2014 English-French dataset. We will then perform analyses on the training results, and, where appropriate, compare the model performance against results from the paper.

3 Goals and Discussion

3.1 Essential Goals

- Generally, as we train the model for longer, the BLEU score should be increasing (we will include a chart to show this).

Discuss what happened

3.2 Desired Goals

- Achieve results of translation that is not simply mappings between the two language vocabularies but rather encompasses the context and attention mappings of the whole sentence. This could be achieved by inputting examples where gendered noun and adjective would be correctly translated (i.e., the Transformer can differentiate between genders).

– “I am a tall man” vs. “I am a tall woman”

Discuss what happened

- Have consistency across languages when training under the same model in terms of BLEU score. Testing with another language dataset (most likely English to French) and have a similar BLEU score performance under the same training settings and time.

Discuss what happened

3.3 Stretch Goals

- Based on the findings from the paper “Reformer: The Efficient Transformer”, we would try to implement and quantify the impact of the suggested changes from that paper compared to the Transformer.

Discuss what happened

- Try pretraining the model on English-to-French, then fine tune the model to translate English-to-Spanish (the motivation is that since both French and Spanish are Romance languages, the pretrained model could have already learned important parts of the mapping from English to any Romance language).

Discuss what happened

3.4 Other Challenges

- Challenge 1
 - Why was it important?
 - What was so challenging/interesting about this?
- Challenge 2
 - Why was it important?
 - What was so challenging/interesting about this?

4 Code and Documentation

4.1 `data.py`

What does this contain and why it’s relevant?

4.2 `decoder.py`

What does this contain and why it’s relevant?

4.3 `encoder.py`

What does this contain and why it’s relevant?

4.4 `multihead_attention.py`

What does this contain and why it’s relevant?

4.5 `optimizer.py`

What does this contain and why it’s relevant?

4.6 `position_wise_feed_forward.py`

What does this contain and why it’s relevant?

4.7 positional_encoder.py

What does this contain and why it's relevant?

4.8 transformer_runner.py

What does this contain and why it's relevant?

4.9 transformer_trainer.py

What does this contain and why it's relevant?

4.10 transformer.py

What does this contain and why it's relevant?

5 Reflections

5.1 What was interesting?

5.2 What was difficult?

5.3 What's left to do?