Preprocessing and nuances

- Both german and english sentences have to be appended with eos tokens at the end

  eos = end_of_sentence

- Since this is English -> German, german sentences have to be appended with sos tokens at the beginning

  sos = start_of_sentence

- The german and english sentences have to be made of the same length else the entire thing breaks

- We do this by appending ‘padding’ tokens after eos

Inference

- The transformer is what in NLP is called a “ Language model “
- The job of a language model is to predict the next word conditioned on the previous words it has seen

- **Inference** in Machine Learning is when you use the model learned (so far) to predict (usually on test data).
- This is important when using things like BatchNorm, Dropout, etc
- There is a train( ) mode and eval mode in Pytorch( )

Linear and Softmax Layer in Decoder

Will project onto the vocabulary of the target language thereby giving a probability distribution over the candidate next words.

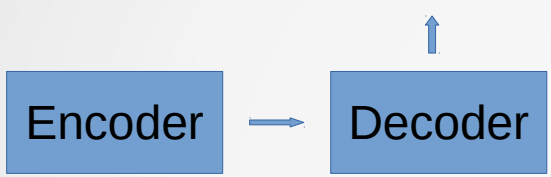
Two ways of choosing words

- Greedy Approach
- Beam Search

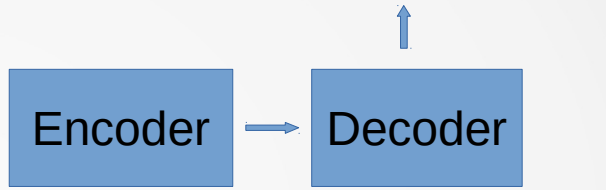
Inference

ran here indicates some word that we don't care about

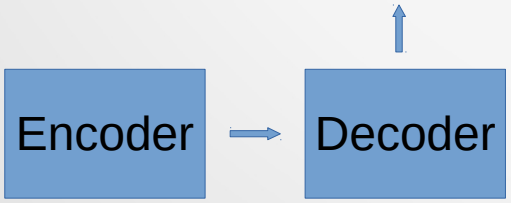
1)



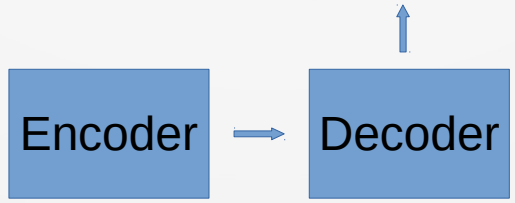
2)



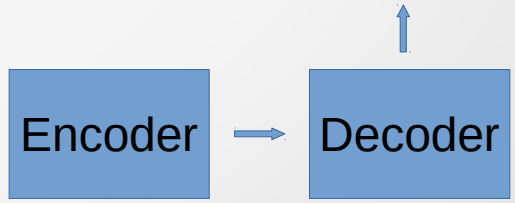
3)



4)



5)



Future Work

5.2 Hardware and Schedule

We trained our models on one machine with 8 NVIDIA P100 GPUs. For our base models using the hyperparameters described throughout the paper, each training step took about 0.4 seconds. We trained the base models for a total of 100,000 steps or 12 hours. For our big models,(described on the bottom line of table 3), step time was 1.0 seconds. The big models were trained for 300,000 steps (3.5 days).

~ Attention Is All You Need

- Not possible to train
- Documentation update

Proof Of Work

https://github.com/shouvikcirca/Transformer

References

Attention Is All You Need
Google
https://arxiv.org/abs/1706.03762

The Illustrated Transformer
http://jalammar.github.io/illustrated-transformer/

Attention Is All You Need | AISC Foundational
https://www.youtube.com/watch?v=S0KakHcj_rs