

Self -Attention NLP

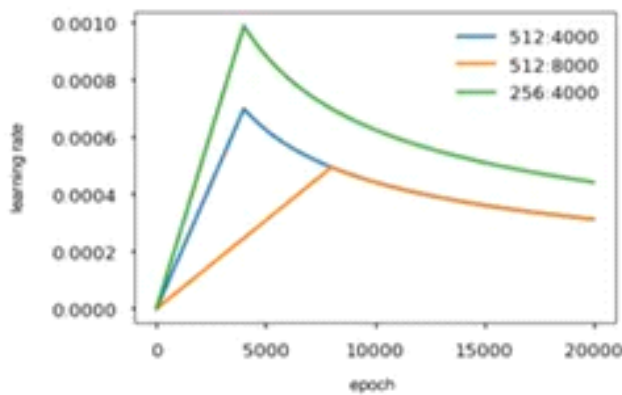
Thursday, October 11, 2018 9:43 AM

https://www.youtube.com/watch?v=OYygPG4d9H0&list=PLGVZCDnMOq0rjkF7p_F4qtaVJQnjK1oKT

Working well

There are A LOT of details in the implementation

For training you have to use learning rate warm-up



$$lrate = d_{model}^{-0.5} \cdot \min(step_num^{-0.5}, step_num \cdot warmup_steps^{-1.5})$$

<http://nlp.seas.harvard.edu/2018/04/03/attention.html>

Don't start from scratch

Tensorflow

<https://github.com/tensorflow/tensor2tensor>

mxnet

<https://github.com/aws-labs/sockeye>

PyTorch

<https://github.com/jadore801120/attention-is-all-you-need-pytorch>

Resources:

<http://nlp.seas.harvard.edu/2018/04/03/attention.html>

<https://mchroliak.github.io/articles/2017/Sep/12/Transformer-Attention-is-all-you-need/>