Attention and Pattern Recognition

Michael Wollowski
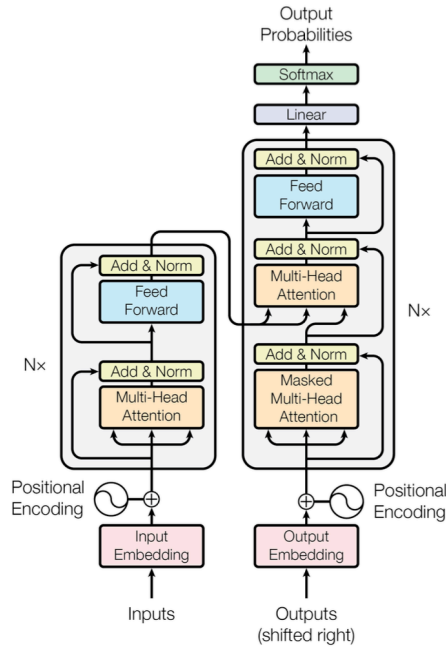
Some pertinent information from the paper:

https://arxiv.org/abs/1706.03762
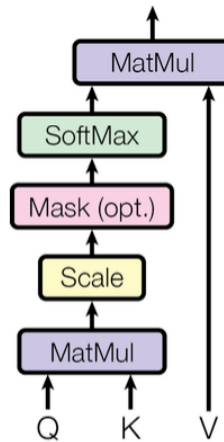
Brief Summary

- We discussed the materials from the paper over this past week.
- As a sanity check, in what follows are the components we discussed.
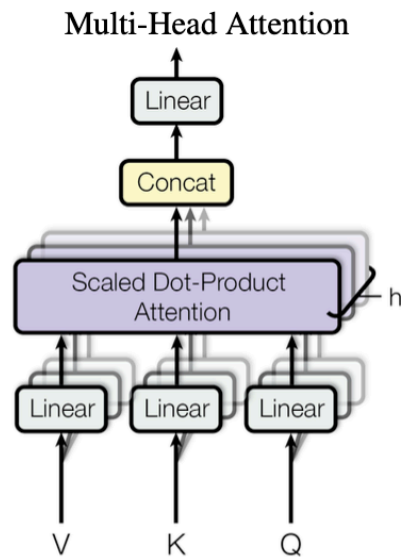
Overall Architecture

A Single Attention Block
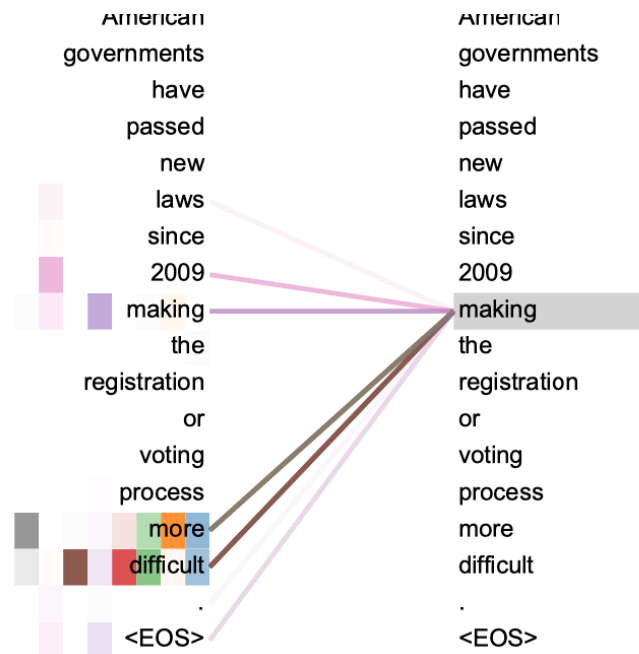
Scaled Dot-Product Attention
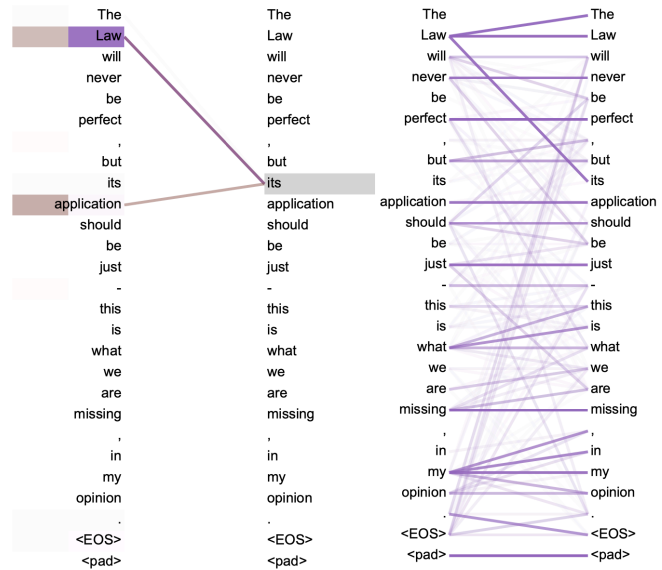
Multi-head Attention

Multi-Head Attention

Multiple Attention Heads pay attention to long-distance dependencies

- Different colors represent different heads.
- Attention is only shown for the word "making"

Anaphora Resolution

- Far right: Full attentions for head 5.
- RIght: Isolated attentions from just the word 'its' for attention heads 5

Attention Heads Related to Structure of Sentence