Thoughts on Attention Is All You Need

Owen Lockwood

August 5, 2020

Although a well established work in the field of NLP (and computer vision to a lesser degree), these are not my fields and as such I had never actually read the "transformer paper". However, this series seemed like a good time to change that and upon reading it I am very glad I did. It is an extremely concise paper, although it is very dense with material. I understand the motivation and results, granted I have done very little work with seq2seq myself. This work is certainly one that I will try to implement. A common criticism I have for ML papers is not going through the training process in any useful detail. It is certainly well explained from a general design point of view, but I found the practical training algorithms to be lackluster (that being said I may find them in the SI). Although attention existed before this work, there also a largely absent theoretical basis for this work and although I appreciate the lack of equation packing, more in depth mathematical background would've been nice. One certainly does not need to say that this work is important, that has clearly been established by now, but I think this work would've benefited from a longer paper. I am excited to implement it and think about its RL applications.