March 21, 2022

STANFORD CS224N: NLP WITH DEEP LEARNING

! self-attention has ability to make models for problems we care about.

Deep learning is all about representation learning, so we need to build the right tools - be it RNNs, GRUs etc.

PROBLEM — But RNNs aren't super-efficient b/c its sequential (not parallel), & it doesnt let us model hierarchy. In the realm of large datasets, RNNs just aren't good enough

SOLUTION — Why not use attention for representations? Attention between encoder & decoder.

(Diagram of transformer w/ self-attention)

Attention is also "cheaper" than RNN & convolutional neural networks

→ self-attention $O(\text{length}^2 \cdot \text{dim})$

→ RNN $O(\text{length} \cdot \text{dim}^2)$

→ CNN $O(\text{length} \cdot \text{dim}^2 \cdot \text{kernal-width})$

APPLIC.
- → image generation
- → remove noise in images, eg astronomy, reit
- → music generation (relational info) (motifs)
- → language translation