

Transformer language models: architectures, word sense disambiguation, and targeted syntactic testing

9.19(0) Fall 2023, Instructor: Roger Levy

6 November 2023

BERT versus GPT

The prerecorded lecture videos introduced the Transformer neural network architecture. The videos focused on the GPT architecture, but there was also brief discussion of BERT. What is the fundamental *architectural* difference between BERT and GPT?

Emergent word sense disambiguation

Earlier in the semester we looked at **word embedding** vector representations. One of the bugbears for word embedding models such as word2vec and GLoVe is **word sense ambiguity**: for example, *bank* can mean where you put your money, or the land at the edge of a river.

One of the powerful effects of deep neural networks that process multi-word string input is that word sense disambiguation emerges (at least to some extent) in the process of computing **contextualized embeddings** for the tokens in the string. In this in-class exercise, your job is to come up with cases that evaluate how well this works in practice. The Colab notebook for this exercise is here:

<https://colab.research.google.com/drive/1jYnOyLVAqphV3pDtkdDztz-USfhzxua4?usp=sharing>

Controlled syntactic testing of neural language models

Finally, we will develop some controlled syntactic tests of an autoregressive language model, GPT-2. (The same principles would apply to more recent “large” language models like GPT-4 or LLaMa, but those are either not open or too large to easily run on Colab.) The Colab notebook for this exercise is here:

https://colab.research.google.com/drive/1OQ0UgQVCMK9kwRweGFcmMj8JpW7hQB_7?usp=sharing