SPH 6004 Advanced Statistical Learning

Week11
NLP with Transformer
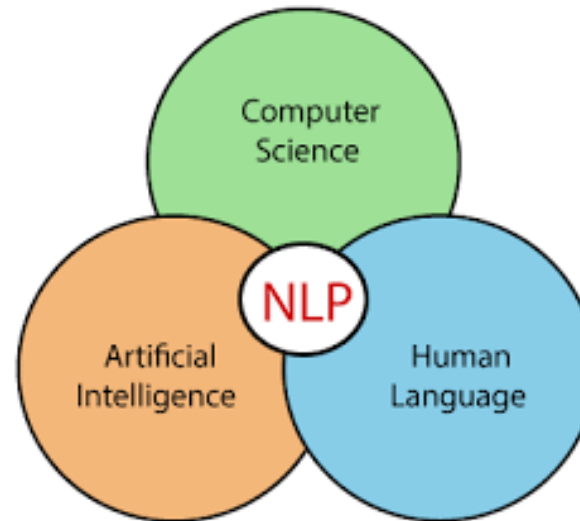
Dr. Hongfei YANG
hfyang@nus.edu.sg

Asst Prof Mengling 'Mornin' FENG
ephfm@nus.edu.sg

Turning Discovery into Healthier Communities

Learning Objectives

- Natural Language Processing
  - What is NLP?
  - Examples of NLP
  - Long distance dependency
- Transformer
  - Encoder-decoder structure
  - Self-attention
  - Positional embedding

Part I: Natural Language Processing

Natural Language Processing (NLP)

NLP concerns with developing algorithms to understand text and spoken words in much the same way human beings can.

- Verbal understanding
- Textual understanding

Natural Language Processing (NLP)

Example 1: language translation

Natural Language Processing (NLP)

Example 2: next word prediction

Neural network is a powerful tool for designing and analyzing complex systems. Thanks to a paradigm shift over the past two decades, the notion of designing and optimizing complex adaptive systems with this tool has greatly advanced our understanding of their operation, enabling novel scientific studies and applications. This background, combined with a strong desire to support and share the achievements of the community, has generated an ambitious and fascinating open source project.

https://app.inferkit.com/demo

Natural Language Processing (NLP)

Long distance dependency between words

"The animal didn't cross the street because it was too tired"

Part II: Transformer

Transformer

Transformer is the state-of-the-art AI model for NLP modeling

- Encoder--decoder structure
  - Encoder: project words to high dimensional representations/embeddings
  - Decoder: decode high dimensional representations to target words
- Self-attention
  - Attention: learn long distance dependency
  - Multi-headed attention: learn multiple dependency
- Positional encoding
  - Sin/cos functions: represent position of words

Transformer

Encoder
Map words to embeddings

Decoder
Use embeddings to predict words in target language

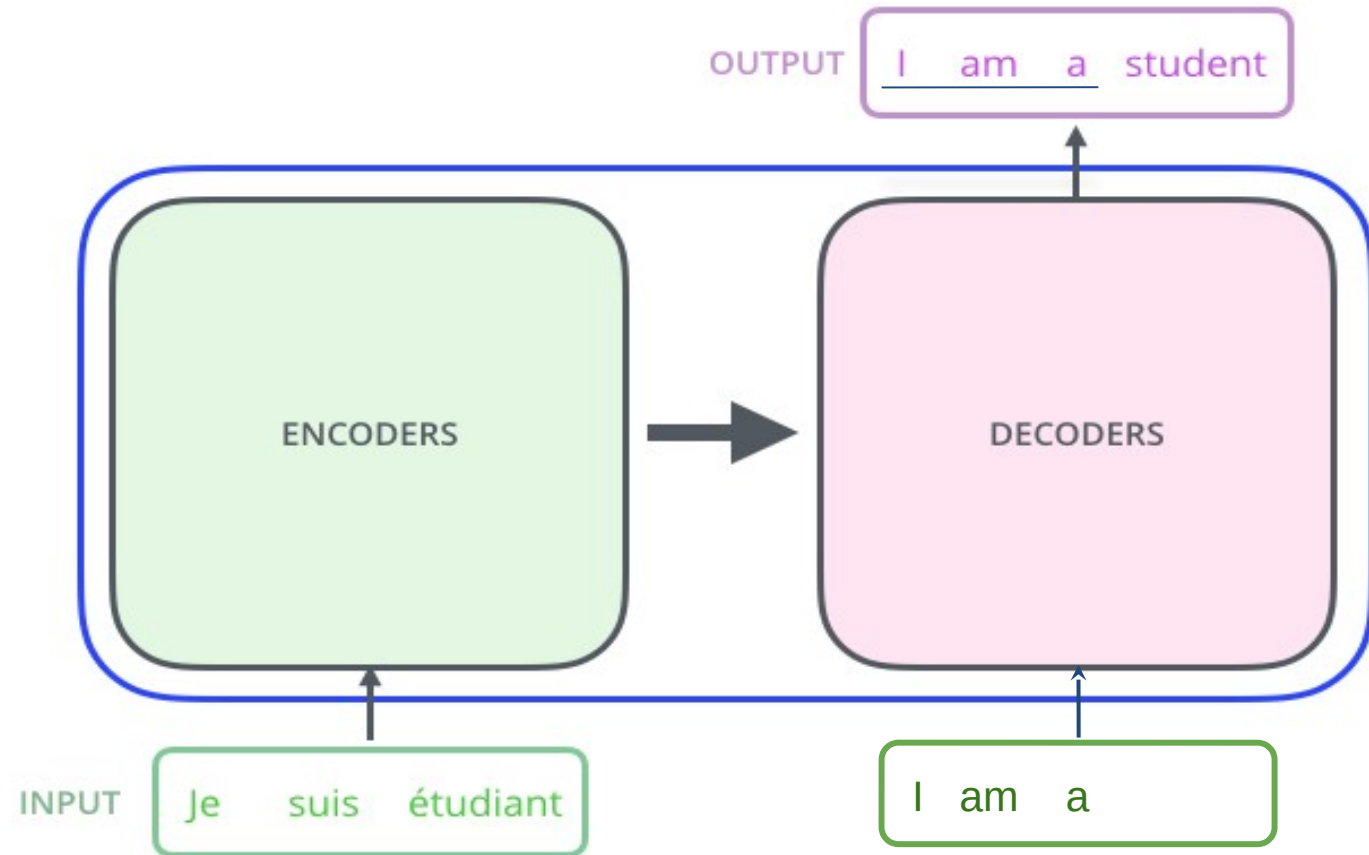Decoder use both encoder outputs and words already predicted to predict the next word
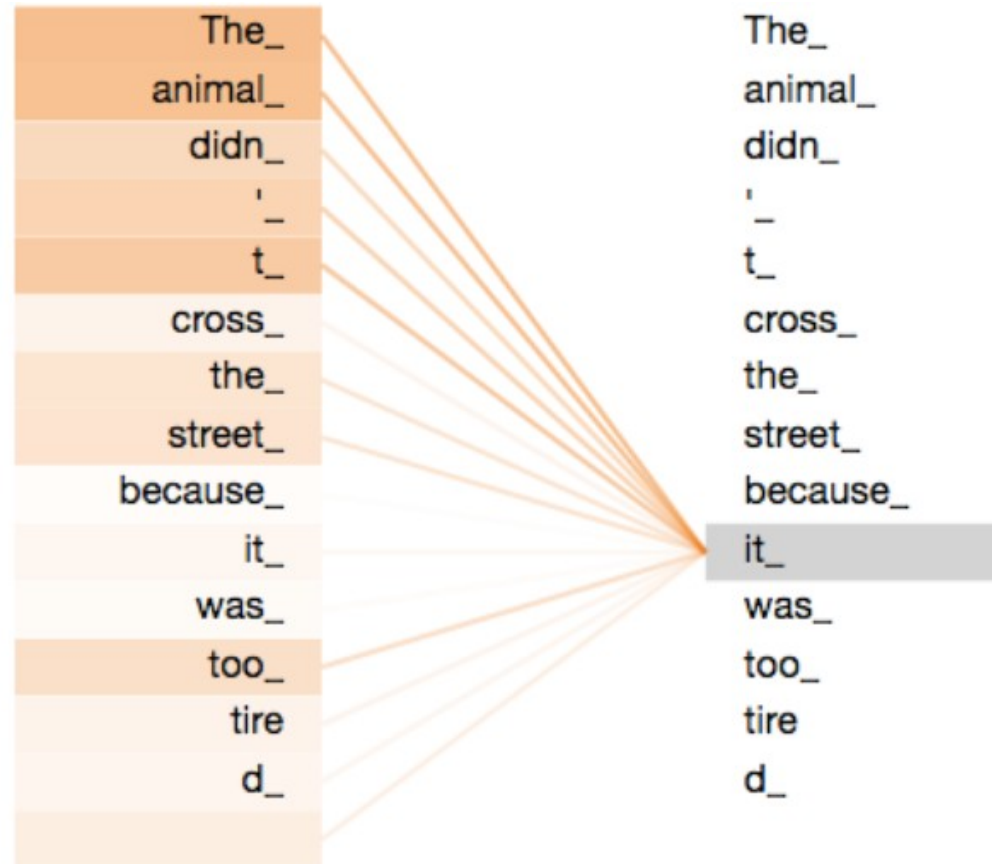

Image source: https://jalammar.github.io/illustrated-transformer/

Transformer

Self-attention

Image source: https://jalammar.github.io/illustrated-transformer/

Transformer

Self-attention

with multi-heads

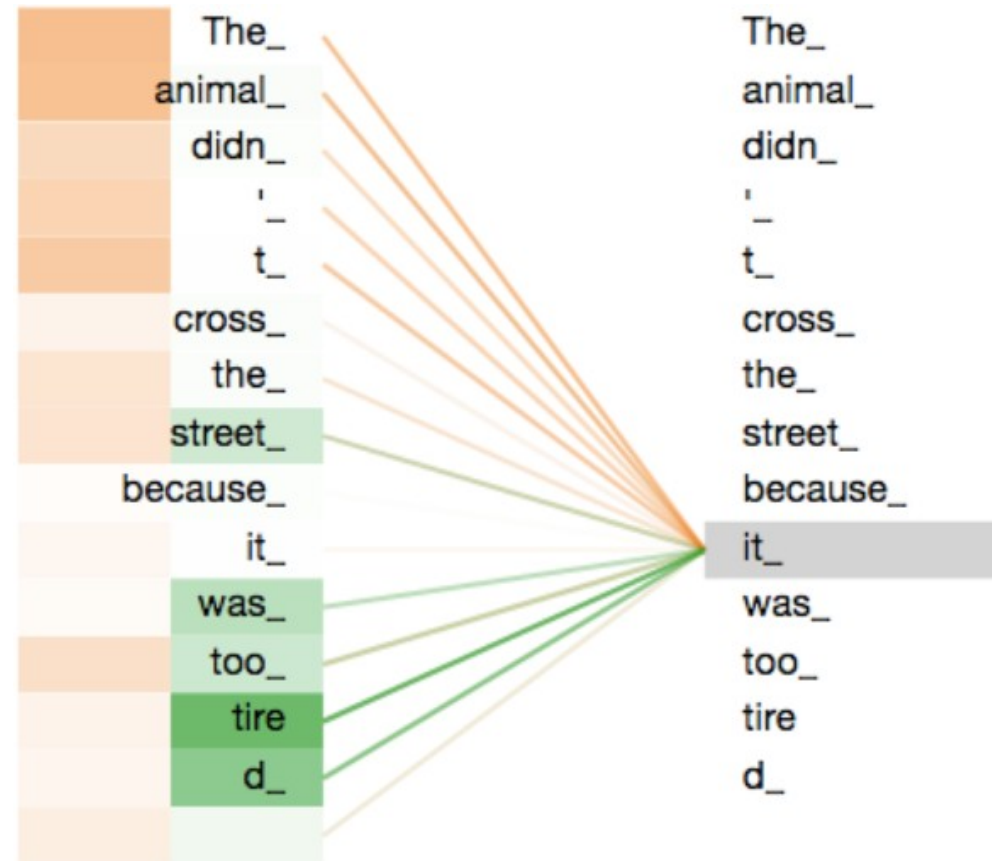Image source: https://jalammar.github.io/illustrated-transformer/

Transformer

- Positional encoding -- order of words matters!
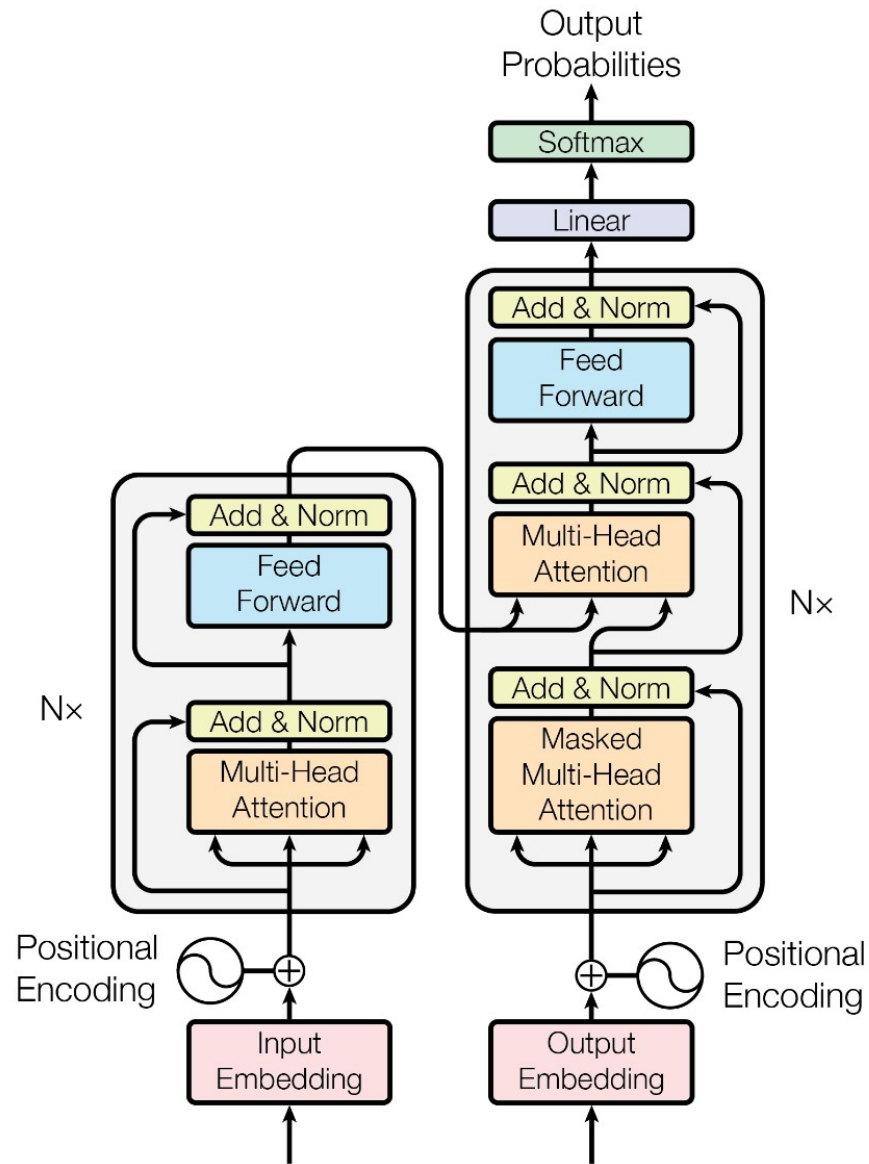   - Encoding positions by using sin/cos functions with various freq.

Image source: https://jalammar.github.io/illustrated-transformer/

Transformer

Image source: https://arxiv.org/pdf/1706.03762.pdf

Useful Resources

1. Attention Is All You Need (transformer original paper)
2. The Illustrated Transformer (nice visualizations!)
3. Stanford CS224n
4. The Annotated Transformer (Explanation with codes)

More Questions?

mornin@gmail.com

www.mornin-feng.com