Workshop 9

COMP90051 Statistical Machine Learning

Semester 2, 2024

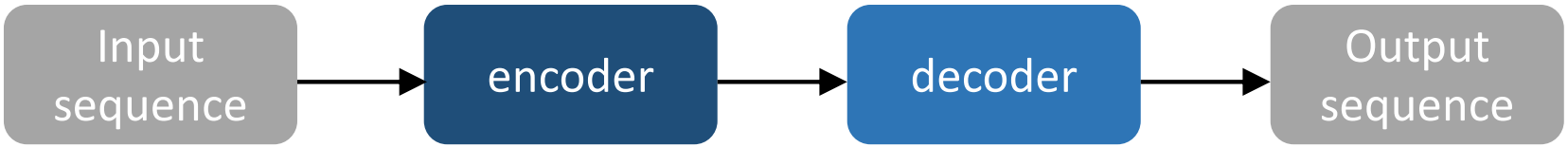
Learning Outcomes

By the end of this workshop you should be able to:

1. explain sequence-to-sequence task, familiar with encoder-decoder architecture
2. Be able to implement Transformer encoder and decoder
3. Know how to train a translation model from scratch using Transformer

Seq2seq task

- Sequence-to-Sequence (Seq2seq) is mapping an input sequence to an output sequence.
- It has been widely used in NLP tasks, such as machine translation, chatbots, text summarization.
- Encoder-decoder is most popular architecture for seq2seq task
  - Both can be implemented using RNN, GRU, LSTM, or Transformer



Transformer

- The Transformer is a NN architecture introduced by paper "Attention is All You Need"
- Similar to the traditional Seq2Seq model, it also comprises an encoder and a decoder
- By leveraging a self-attention mechanism, model to capture long-term dependencies more effectively than RNN

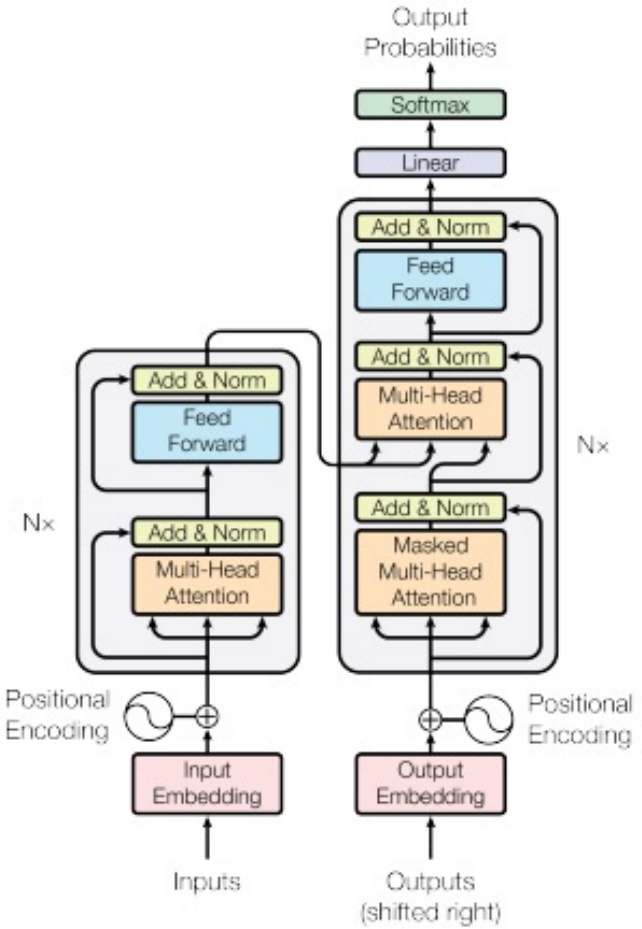


Figure 1: The Transformer - model architecture.

Transformer architecture from "Attention is all you need"

Useful Reading Materials

- https://towardsdatascience.com/build-your-own-transformer-from-scratch-using-pytorch-84c850470dcb
- https://jalammar.github.io/illustrated-transformer/
- https://www.youtube.com/watch?v=4Bdc55j80l8

Capture multiple

drop out : randomly → 变0 防止 数据过度依赖某些 neurons

训练过程中

Worksheet 9