

Introduction

In this section you will be building a sequence to sequence (seq2seq) model. Seq2Seq models are used for tasks that involve reading in a sequence of text and generating an output text sequence based on the input.

A. Sequence to sequence

One of the most important frameworks in NLP is the sequence to sequence (seq2seq) framework. This framework encompasses any task that involves taking in some text and returning some generated text. Dialog systems (e.g. chatbots), text summarization, and machine translation are all seq2seq applications.



A translation from English to French using Google Translate. This is an example of machine translation, a seq2seq task.

In the past, these types of seq2seq tasks were performed by regular statistical models, many of which utilized [Bayesian statistics](#). However, with the rise of neural networks in the past decade, we've been able to apply deep learning to all these seq2seq tasks with great success.

In fact, there is an extremely powerful model called the *encoder-decoder*, which is specifically designed for seq2seq applications.

The encoder-decoder is named for its two parts: the encoder and the decoder. The encoder and the decoder are both language models (like the LSTM models

from previous sections).

First, an input sequence is fed to the encoder. The output from the last layer of the encoder becomes the input for the first layer of the decoder. The decoder transforms that input back into a text sequence.

In the following chapters you'll learn more about the encoder-decoder, as well as write code to build one yourself.