

Seq2seq

Technique

Seq2seq **turns one sequence into another sequence** (sequence transformation). It does so by use of a recurrent neural network (**RNN**) or more often LSTM or GRU to avoid the problem of vanishing gradient.

The context for each item is the output from the previous step.

The primary components are one **encoder** and one **decoder** network.

The encoder turns each item into a corresponding hidden vector containing the item and its context.

The decoder reverses the process, turning the vector into an output item, using the previous output as the input context.

The optimization include

- **Attention**: the input to the decoder is a single vector which stores the **entire context**.
Attention allows the decoder to look at the input sequence selectively.
- **Beam search**: Instead of picking the single output (word) as the output, **multiple highly probable choices are retained**, structured **as a tree** (using a Softmax on the set of attention scores). Average the encoder states weighted by the attention distribution.
- **Bucketing**: Variable-length sequences are possible because of padding with 0s, which may be done to both input and output. However, if the sequence length is 100 and the input is just 3 items long, **expensive space is wasted**. Buckets can be of varying sizes and specify both input and output lengths.