

Week 5 : POS tagging as Seq2Seq

Sequence to sequence (aka Seq2Seq) is an NLP paradigm (a way to solve tasks) that became popular a few years ago along side the encoder-decoder architecture. The core idea behind Seq2Seq is as its name implies, to map input sequences to output sequences. The most stereotypical Seq2Seq task is translation where you want to generate a sentence in a target language given a sentence in a source language. But the truth is that many tasks can be cast in a way or another in the Seq2Seq format.

POS tagging is a sequence labeling task, but it can readily be framed as a Seq2Seq problem where given an input sentence one wants to produce the corresponding sequence of POS tags.

1. Use your own CoNLL-U parser, accuracy measuring methods and so on.
2. Create a simple sentence encoder (use Torch if you have it). You'll need an embedding layer for your words and some form of sentence encoding mechanism, likely an RNN, though a CNN could work as well.
3. From the encoded sentence, predict the sequence of POS tags. Here again, you'll likely want to use an RNN. Also, don't forget to think of a way to stop generating POS.
4. On top of your existing encoder-decoder model, add an attention head.
5. Apply this new model both to mono-lingual and cross/multi-lingual settings.