

Week 3 – sequence models notes

-The translation model is very similar to the language model, except that instead of always starting along with the vector of all zeros, it instead has an encoded network that figures out some representation for the input sentence, and it takes that input sentence and starts off the decoded network with representation of the input sentence rather than with the representation of all zeros.

-why not just use greedy search?

greedy search is an algorithm from computer science which says to generate the first word just pick whatever is the most likely first word according to your conditional language models. Going to your machine translation model and then after having picked the first word, you then pick whatever is the second word that seems most likely, then pick the third word that seems most likely.

-drawback: this is an example of a broader phenomenon, where if you want to find the sequence of words, y_1, y_2 , all the way up to the final word that together maximize the probability, it's not always optimal to just pick one word at a time. also, it might choose a word that that is common but not Grammatically correct.

-Machine translation can be posed as a conditional language modeling problem. But one major difference between this and the earlier language modeling problems is rather than wanting to generate a sentence at random, you may want to try to find the most likely English sentence, most likely English translation. But the set of all English sentences of a certain length is too large to exhaustively enumerate. So, we have to resort to a search algorithm. So, with that, let's go onto the next video where you'll learn about beam search algorithm.

-Length normalization is a small change to the beam search algorithm that can help you get much better results.

- What the BLEU score does is given a machine generated translation, it allows you to automatically compute a score that measures how good is that machine translation.

- An LSTM has both a hidden state and cell state

-unigram: meaning we're looking at single words in isolation.

-Machine translation models can be used to map from one sequence to another. They are useful not just for translating human languages (like French->English) but also for tasks like date format translation. An attention mechanism allows a network to focus on the most relevant parts of the input when producing a specific part of the output. A network using an attention mechanism can translate from inputs of length T_x to outputs of length T_y , where T_x and T_y can be different. You can visualize attention weights $\alpha(t, t')$ to see what the network is paying attention to while generating each output.

-Data synthesis is an effective way to create a large training set for speech problems, specifically trigger word detection.

Using a spectrogram and optionally a 1D conv layer is a common pre-processing step prior to passing audio data to an RNN, GRU or LSTM. An end-to-end deep learning approach can be used to build a very effective trigger word detection system.