

Part 1: RNN Acceptor
1.1: Understanding the Challenge
Lior Shimon 341348498
Lev Levin 342480456

We **can not** distinguish the two languages using a bag-of-words approach.

As we saw in lecture 7, CBOW allows encoding arbitrary length sequences, but loses all order information.

Since the difference between the positives and negatives examples rely on the order of arrival between the sequence of b's and the sequence of c's, using CBOW approach does not enable to distinguish the two languages.

Using a bigram or trigram based approach, the two languages **can not** be distinguished.

We know any n-gram based approach don't give importance to the order of the overall sentence, and focus only on each sequence of n- "words" (here words mean strings), the local order.

Since the difference of the two languages reside in the arrival order between the sequence of b and the sequence of c in the overall sentence, and the length of each sequence in the sentence is not pre-define, this approach is not relevant.

using a convolutional neural network, the two languages **can not** be distinguished.

Once again, the difference of the two languages reside in the arrival order between the sequence of b and the sequence of c in the overall sentence.

Since the convolutional neural network focus on local features, the filter will "scan" each sentence with a predefine filter-length and a predefine stride before sending the information collected into a MLP.

Even though in a conv-net we can learn local order information in a sequence using a right filter and a right paddle, we know each sequence in the sentence can be of any length, and their length can be different between all sentences as well. Therefore, we need to keep track of long-term dependencies, and LSTM is the more accurate kind of approach for this training task. Eventually, the Conv net will reach some accuracies over "simple" examples, who are structurally similar one another, but over all the sentences and their wide range of different length, a CNN will reach its limit early during the training.

For this Specific language, we could eventually reach good result using Hierarchical Neural Net as seen in lecture, with each filter feeding a higher ranked filter. with this structure, the "Root filter" actually get information on the overall sentence. But since filter's length have predefine length, we suppose this model will reach result more hardly less accurate than the LSTM Model.

