

Hierarchical Generalisation in Recurrent Neural Networks

Saujas Srinivasa Vaduguru, Ujwal Narayan

Computational Linguistics 1 Project

For the course project, we plan to replicate parts of the study performed by McCoy *et al.* [3] about how recurrent neural networks perform on tasks that require hierarchical generalisation.

To evaluate this, the authors choose a simple task – forming questions from declarative sentences in English. The process for forming a question from an English declarative sentence is moving the main verb’s auxiliary to the front of the sentence. For simple sentences like

The bird *can* fly.

the question can be formed by simply moving the first auxiliary to the front to give

Can the bird fly?

But, for more complex sentences, the first auxiliary may not be the one that is to be moved. Consider

The bird that *will* eat ***can*** fly.

Here, moving the first auxiliary will result in the incorrect question

* *Will* the bird that eat ***can*** fly?

Instead, it is the hierarchical rule of moving the main verb’s auxiliary that produces the correct question:

Can the bird that *will* eat fly?

We plan to use such examples to train sequence-to-sequence (seq2seq) neural network models [4] to transform the declarative sentence to the question. The details of the experiments follow.

To evaluate how the NN models perform on this task, we will evaluate their performance on two variants of the task. One variant is with auxiliaries that do not exhibit agreement, like *can*, *could*, *will*, and ones that do agree in number, like *do*, and *don’t*. This will also allow us to see how syntactic features affect the performance of the NN model.

These sentences will be generated using a basic, limited context free grammar, and divided into sets based on the position of the relative clause (none, subject, object), and whether the task is to generate the same sentence again, or the question. One of these sets will be held out during training to see how the NN model generalises.

While McCoy *et al.* [3] train multiple architectures and compare their performance, we will only be studying the performance of one of the architectures – LSTM with attention [1] [2]. McCoy *et al.* [3] find that the initialisation of parameters affects performance, so we will be training multiple networks with different random initialisations.

Finally, if time permits, we will analyse the representations of the sentence learnt by the NN models to see what features they learn. This analysis will be performed using linear classifiers that predict certain specific words that capture information that would indicate

that the network is acquiring some hierarchical information (main auxiliary, fourth word, subject noun).

This project relates to a deep question in linguistics about the acquisition of language. Chomsky’s famous argument – the argument from the poverty of the stimulus – states that there must be some hierarchical bias innately present in the mind. This is because children rarely encounter complex sentences with relative clauses, but manage to learn the correct rule for tasks like question formation anyway. The authors claim that is a sequence based model can learn hierarchical features, then there need not be an innate hierarchical bias in children. To lend more substance to this argument, the authors compare the errors made by the neural network to those made by children who are acquiring language. The authors do refrain from making a hasty generalisation, indicating that much more elaborate experimentation would be required to draw that conclusion.

References

- [1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014.
- [2] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997.
- [3] R. Thomas McCoy, Robert Frank, and Tal Linzen. Revisiting the poverty of the stimulus: hierarchical generalization without a hierarchical bias in recurrent neural networks. *CoRR*, abs/1802.09091, 2018.
- [4] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. *CoRR*, abs/1409.3215, 2014.