

Finding grammar in neural language models

Project Description

A statistical language model defines a probability distribution over sequences of words (in other words, they can be used to compute the likelihood of sentences). Having a way to estimate the (relative) probability of sentences is useful in a number of natural language processing applications, such as speech processing, machine translation and part of speech tagging, but is also interesting from a cognitive perspective, considering the fact that also humans have an internal language model. Arguably, understanding whether a sentence is probable or not requires some understanding of its syntactic structure, but language models are usually not provided explicit syntactic information during training. In this project we will study to what extent a neural language model is sensitive to certain syntactic phenomena, aiming to get a better insight at what these models actually encode. You will first be asked to replicate a result from an existing paper, that tests the sensitivity of a neural language model to subject-verb agreement. In the next step, you are asked to investigate a question of your own.

Requirements

As a final product, you will be asked to write a report with your findings, which should at least contain:

- A background section, in which you explain what is language modelling, what kind of language model you studied, and what is the problem you focus on. To write this background section, you can use the suggested papers, but also add papers;
- A summary of [1] the paper of which you are asked to reproduce part of their results;
- A description of your replication;
- A report of the investigation of your own question about grammar in the provided neural language model, this can be very similar to the one asked in [1], but can also be something very different;
- A conclusion/summary.

Getting Started

For this project, there are a couple important concepts, before diving into coding, you can start by familiarising yourself with these concepts.

(Neural) language modelling Language modelling was covered during the lectures, make sure that you understand it. The most simple version of a *neural* language model, which is not discussed during the lectures, was presented by Elman in 1990. The model that we use is more complex than this one, but the basic principle is the same. A good overview of language models (neural and more traditional) can also be found in the book *Speech and Language Processing* (Chapters 4 and 8), a pdf can be found here: <https://web.stanford.edu/~jurafsky/slp3/>.

Long distance dependencies An aspect that is particularly difficult for language models, and considered very important by linguists are long distance dependencies. In the paper you are asked to reproduce [1], you can find both an apt description of the problem, as well as a worked out experiments to investigate the sensitivity of language models to long distance dependencies. This paper also contains many references for future reading.

Running the language model To do your experiments, you will be provided a neural language model that is trained using pytorch, a description can be found at https://github.com/pytorch/examples/tree/master/word_language_model. We will train this model for you. You will soon receive a notebook that contains instructions on how to download and run the model.

Further reading Additional papers will be provided during the course of the project, feel free to propose anything that you think is related or find interesting!

References

- [1] Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. Assessing the ability of lstms to learn syntax-sensitive dependencies. *arXiv preprint arXiv:1611.01368*, 2016.