

A summary of
*‘What do RNN Language Models Learn about Filler-Gap
Dependencies?’*
(Wilcox et al. 2018)

Tanise Ceron & Bogdan Kostić

September 30, 2019

1 Introduction

Recurrent Neural Networks (RNNs) have achieved impressive results on NLP tasks. Long Short Term Memory (LSTM), for instance, is a type of RNN model performing well in tasks such as machine translation, language modeling and syntactic parsing. In this study, Wilcox et al. (2018) investigated whether LSTMs have acquired knowledge of filler-gap dependencies.

Filler-gap dependencies consist of a filler and a gap. The former refers to a wh-complementizer, such as ‘what’ and ‘who’, and the latter is an empty syntactic position licensed (‘allowed’) by the filler. Nonetheless, filler-gap dependencies are not observable in all natural language constructions. This is called island constraint.

2 Methods

2.1 Language Models

Two models were tested and compared in this paper. One of them is the Google model, which was trained on the One Billion Word Benchmark and consists of two hidden layers with 8196 units each. The other model, called Gulordava model, was trained on 90 million tokens of English Wikipedia containing two hidden layers with 650 units each. As a baseline, an n-gram model was trained on the One Billion Word Benchmark in order to compare its capabilities in detecting filler-gap dependencies to the other two LSTMs.

2.2 Dependent variable: Surprisal

For assessing the performance of the models in detecting filler-gap dependencies, a measure called surprisal was applied. The surprisal value provides information about how unexpected a word or a sentence is under the language model’s probability distribution. It is computed as following:

$$S(x_i) = -\log_2 p(x_i|h_{i-1})$$

The degree of surprisal should be higher when the model comes across a gap without the existence of a filler.

2.3 Experimental design

A 2x2 interaction between the presence of a gap and the presence of a wh-licensor was used to indicate the surprisal reduction caused by the wh-licensor linked to the gap. This is called wh-licensing interaction.

To determine whether the models have also acquired knowledge about the island constraints, the authors looked at interactions between the wh-licensing interaction and other factors, such as the possibility of wh-licensing interaction decreasing when a gap would be grammatical ('syntactic licit position') or ungrammatical ('syntactic island position').

The experimental sentences were created by the researchers themselves. They made sure to locate the gap in an obligatory argument position and to embed the phrase with the gap inside a complement clause. The surprisal is measured at the word immediately following the gap and also summed over all words from the gap to the end of the embedded clause.

Wilcox et al. (2018) formulated two hypotheses. The first refers to the expectation of a higher surprisal in syntactic positions where a gap is likely to occur in sentences containing a wh-licensor but no gap. The second concerns the expectation of a higher surprisal in the presence of a gap and the absence of a wh-licensor compared to when a wh-licensor is present.

3 Representation of filler-gap dependencies

This research analysed whether the LSTM models complied with three basic characteristics of filler-gap dependency. The first characteristic of filler-gap dependency is flexibility, which means being able to place the wh-complementizer in various syntactic positions. The second one is robustness to intervening material, meaning that the dependency is still possible even with a long distance between filler and gap. The last one concerns the one-to-one relationship between a wh-phrase and a gap.

Wilcox et al. (2008) showed that while both the Google model and the Gulordava model managed to detect filler-gap dependencies with their characteristics, the n-gram model failed to do so.

4 Syntactic islands

There are some limitations to filler-gap dependencies related to syntactic positions in which gaps are not allowed. These positions are called syntactic islands. This study aims to point out whether LSTM language models have learned these constraints. In total four constraints were tested, the wh-island constraint, the adjunct island constraint, the complex NP constraint and the subject constraint.

5 Conclusion

Finally, this study has demonstrated that LSTM language models are capable of learning to represent filler-gap dependencies with their characteristics and some of their limitations. Whereas both models managed to learn most of the constraints, neither the Google model nor the Gulordava model was able to learn the subject constraint. In addition to that,

the Google model was unsuccessful in learning the that-headed complex NP island and the Gulordava model to learn the wh-island.

References

Wilcox, E., Levy, R. P., Morita, T., & Futrell, R. (2018). What do RNN Language Models Learn about Filler–Gap Dependencies? In *Proceedings of the Workshop on Analyzing and Interpreting Neural Networks for NLP*.