

Formal languages and neural models for learning on sequences

William Merrill

New York University, USA

Editors: François Coste, Faissal Ouardi and Guillaume Rabusseau

Abstract

The empirical success of deep learning in NLP and related fields motivates understanding the model of grammar implicit within neural networks on a theoretical level. In this tutorial, I will overview recent empirical and theoretical insights on the power of neural networks as formal language recognizers. We will cover the classical proof that infinite-precision RNNs are Turing-complete, formal analysis and experiments comparing the relative power of different finite-precision RNN architectures, and recent work characterizing transformers as language recognizers using circuits and logic. We may also cover applications of this work, including the extraction of discrete models from neural networks. Hopefully, the tutorial will synthesize different analysis frameworks and findings about neural networks into a coherent narrative, and provide a call to action for the ICGI community to engage with exciting open questions.