Computational Expressivity of Neural Language Models

Alexandra Butoi, Ryan Cotterell and Anej Svete

Language models (LMs) are currently at the forefront of NLP research due to their remarkable versatility across diverse tasks. However, a large gap exists between their observed capabilities and the explanations proposed by established formal machinery. To motivate a better theoretical characterization of LMs' abilities and limitations, this tutorial aims to provide a comprehensive introduction to a specific framework for formal analysis of modern LMs using tools from formal language theory (FLT). We present how tools from FLT can be useful in understanding the inner workings and predicting the capabilities of modern neural LM architectures. We will cover recent results using FLT to make precise and practically relevant statements about LMs based on recurrent neural networks and transformers by relating them to formal devices such as finite-state automata, Turing machines, and analog circuits. Altogether, the results covered in this tutorial will allow us to make precise statements and explanations about the observed as well as predicted behaviors of LMs, as well as provide theoretically motivated suggestions on the aspects of the architectures that could be improved.

Alexandra Butoi, Ph.D. Candidate at ETH Zürich.
email: `alexandra.butoi@inf.ethz.ch`
website: `https://rycolab.io/authors/alexandra/`
Her current interests include formalisms for mildly context-sensitive languages and parsing.

Ryan Cotterell, Assistant professor at ETH Zürich in the Institute for Machine Learning.
email: `ryan.cotterell@inf.ethz.ch`
website: `https://rycolab.io/authors/ryan/`
His research focuses on a wide range of topics, including information-theoretic linguistics, parsing, computational typology and morphology, and bias and fairness in NLP systems.

Anej Svete, Ph.D. Candidate at ETH AI Centre at ETH Zürich.
email: `anej.svete@inf.ethz.ch`
website: `https://anejsvete.github.io/`
His main research interests lie at the intersection of formal language theory and LMs, where he is working on improving our understanding of the formal properties of modern architectures.

1