Inference of grammar complexity by Bayes factors using sequential Monte Carlo

Caroline Lawless, Judith Rousseau , Robin Ryder
University of Oxford, Department of Statistics

UNIVERSITY OF OXFORD

Abstract

The class of context-free grammars is believed to be too restrictive to fully describe all features of natural language. The class of context-sensitive grammars, on the other hand, is too complex: modelling with them would require an unrealistic amount of computational time. Various mildly context-free grammar formalisms, which may be placed between context-free grammars and context-sensitive grammars in terms of complexity, have thus been proposed in the last few decades. We will be interested in the class of 2-multiple context-free grammars (2-MCFGs) [2], which properly include the class of context-free grammars. We propose a Bayesian non-parametric model for 2-MCFGs within which a model for context-free grammars is naturally embedded. We carry out model choice by Bayes factors using sequential Monte Carlo in Birch probabilistic programming language.

Motivation

- Given a set of sentences, how can we characterize the complexity of the underlying language?
- Muriqui monkey vocalizations dataset.

Formal grammars

- Chomsky (1956) [1] fully characterized a formal grammar, $\mathcal{G} = (\mathcal{A}, \mathcal{B}, \mathcal{S}, \mathcal{R})$ by the following components.
  - A finite set of terminal symbols $\mathcal{A} = \{a_1, \ldots, a_K\}$
  - A finite set of nonterminal symbols $\mathcal{B} = \{B_0, \ldots, B_J\}$
  - A start symbol (a distinguished nonterminal symbol), $\mathcal{S}$
  - A finite set of rules $\mathcal{R}$, each of the form

$$(\mathcal{A} \cup \mathcal{B})^\star \mathcal{B} (\mathcal{A} \cup \mathcal{B})^\star \to (\mathcal{A} \cup \mathcal{B})^\star,$$

    where * is the Kleene star, and $\cup$ denotes set union.
- A probabilistic grammar is a formal grammar which, additionally, associates to each rule some probability.
- A probabilistic grammar may be represented as a random tree, with
  - $\mathcal{S}$ at the root
  - $B_i \in \mathcal{B}$ at internal nodes
  - $a' \in \mathcal{A}$ at leaves.

Complexity of grammars

- The complexity of a formal grammar is defined by the structure of the type of rules it allows.
- The Chomsky containment hierarchy for complexity of grammars is as follows.

$$\text{regular} \subset \text{context-free} \subset \text{context-sensitive} \subset \text{unrestricted}$$

- The classes are defined by the rules which they allow.
- There now exist extensions to the Chomsky hierarchy of grammars, for example

$$\text{context-free} \subset \textbf{2-multiple context-free} \subset \text{context-sensitive} \subset \ldots.$$

  - Context-free grammars in Chomsky normal form

$$B \to B'_1 B'_2 \text{ or } B \to a', \qquad a' \in \mathcal{A} \quad B, B'_1, B'_2 \in \mathcal{B}$$

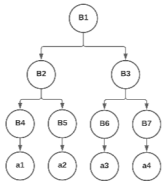
  - 2-multiple context-free grammars

$$B \to f'[B'_1, \ldots, B'_n] \text{ or } B \to (a'_1, a'_2)$$
$$a'_1, a'_2 \in \mathcal{A}, \quad B, B'_1, \ldots B'_n \in \mathcal{B}, \quad f' \in \mathcal{F}$$

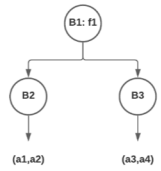
    where $\mathcal{F}$ is a particular class of functions that takes $2n$ arguments from $\mathcal{A}^\star$, and which has its values in $\mathcal{A}^\star \times \mathcal{A}^\star$.

Example trees from random grammars

- Below is a realization from a context-free grammar. Its associated sentence is $a_1 a_2 a_3 a_4$.



- Let f_1 be the function defined by $f_1(x_1^{(1)}, x_1^{(2)}, x_2^{(1)}, x_2^{(2)}) = x_1^{(1)} a_5 x_1^{(2)} x_2^{(2)} x_2^{(1)}$. Below is a realization from a 2-multiple context-free grammar. Its associated sentence is $a_1 a_5 a_2 a_4 a_3$.



Model and inference goals

- We aim to preform model choice between
  - **context-free grammars** and
  - **2-multiple context-free grammars.**
- We propose a model for 2-mulitple context-free grammars.
  - Context-free grammars are natually embedded within our model.
- Our model is an extention of the model of Ryder et. al (in progress) for context-free grammars.
- **Hierarchical Dirichlet process** [3] prior on rules.
  - Flexible models, unbounded number of possible rules.
  - Clustering within and between groups.
- We aim to carry out model choice using Bayes factors, using sequential Monte Carlo in Birch programming language.

References

References

[1] Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on information theory 2*(3), 113–124.

[2] Seki, H., T. Matsumura, M. Fujii, and T. Kasami (1991). On multiple context-free grammars. *Theoretical Computer Science 88*(2), 191–229.

[3] Teh, Y. W., M. I. Jordan, M. J. Beal, and D. M. Blei (2006). Hierarchical dirichlet processes. *Journal of the american statistical association 101*(476), 1566–1581.