Investigating Unsupervised Constituency Parsing for Music Harmony

Ningxiang Xie, Eamonn Bell, Robert Lieck
ningxiang.xie@durham.ac.uk, eamonn.bell@durham.ac.uk, robert.lieck@durham.ac.uk
Durham University

Abstract

Musical harmony and language both exhibit hierarchical and recursive latent structures (Rohrmeier et al., 2015). Deep inside–outside recursive autoencoders (DIORA) (Drozdov et al., 2019) have demonstrated superior performance in unsupervised constituency parsing of natural language. However, unsupervised constituency parsing of harmony remains underexplored. We therefore investigated whether DIORA can perform unsupervised constituency parsing on artificial music harmonic data produced by a context-free grammar. Interestingly, our experiments show that DIORA *fails* to learn correct parses *despite* effectively optimising its reconstruction objective. This suggests that harmonic grammars might have some property that differs from natural language and makes them more difficult to learn for DIORA.

We therefore performed further analyses on musical harmony data. First, we confirmed through supervised training that DIORA's reconstruction objective does not conflict with correct parses, that is, the objective's value does not change but DIORA learns to produce correct parses. This suggests that it might be possible to solve the unsupervised case by adding an additional inductive bias to the model.

We therefore experimented with several alternative configurations of DIORA, each aiming to induce correct parsing, including adding noise to create an information bottleneck and enforcing DIORA's internal span representations to match the set of ground-truth nonterminals. However, none of these interventions yielded correct parses. These unsuccessful attempts suggest an alternative hypothesis: the harmonic language generated by the true CFG may admit multiple equivalent CFG formulations beyond the one used for data generation. We theoretically confirmed for simple examples that this can actually occur, but it remains unclear whether this is the reason for DIORA's inability to identify the ground truth CFG.

Our work points the way to further research bridging data-driven models of natural language, music, and the theory of formal languages, shedding light on the complexities of the music harmony domain.

Reference list

Drozdov, A., Verga, P., Chen, Y.-P., Iyyer, M., & McCallum, A. (2019). Unsupervised Labeled Parsing with Deep Inside-Outside Recursive Autoencoders. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 1507–1512). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1161

Rohrmeier M, Zuidema W, Wiggins GA, Scharff C (2015) Principles of structure building in music, language and animal song. Philosophical Transactions of the Royal Society B: Biological Sciences 370:20140097. https://doi.org/10.1098/rstb.2014.0097

Statement as to whether you prefer to present orally or as a poster: We prefer to give an oral presentation but poster is also fine.