Unsupervised Induction of Harmonic Syntax

Ruben Cartuyvels[0000-0003-1063-4659], John Koslovsky[0000-0001-9910-8158], and Marie-Francine Moens[0000-0002-3732-9323]

ruben.cartuyvels@kuleuven.be
Department of Computer Science, KU Leuven, Belgium

Keywords: Parsing · Deep learning · Music theory

1 Introduction

Hierarchical structures describing a syntax of harmony have long been studied and proposed by music theorists, based on musical relations like *prolongation* or *preparation* [15,9,14,10,11]. Algorithms that model these structures exist but they either require costly expert annotations for training [3], or are based on music theorists' predispositions about harmonic syntax [7]. Figure 1 shows such an example tree resulting from the parsing of a chord sequence according to syntax rules from a context-free grammar. These hierarchical representations of harmony can aid the analysis of music, similarly to Schenkerian analysis where foreground notes are related to the *Ursatz* – a deeper structure [13].

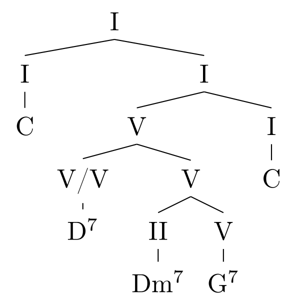

Fig. 1: A syntax tree representing the harmony of a part of 'Take the "A" train' by Duke Ellington and Billy Strayhorn.

We propose to use neural networks to exploit parameter sharing when estimating rule probabilities for **probabilistic context-free grammars** (PCFG's), to induce a grammar for chord sequences from jazz pieces. For the first time, we do this in an entirely *unsupervised* manner, i.e., entirely from raw textually encoded sequences of chord symbols, without access to annotated parse trees (except for evaluation on sequences not seen during training) *and* while adding minimal music theoretical knowledge. This allows us to train on more data: datasets with tree annotations contain little more than 100 samples [5,6], while datasets with raw chord sequences exist with up to 20K samples [2].

2 Methods

A PCFG consists of rules in Chomsky normal form, like $S \rightarrow A$, $A \rightarrow B_1\ B_2$, and $P \rightarrow c$. S is the start symbol, A, B_1, and B_2 are nonterminal symbols (representing groups of chords), P is a preterminal symbol (representing a single chord), and c is a chord symbol. Each rule r is associated with a probability

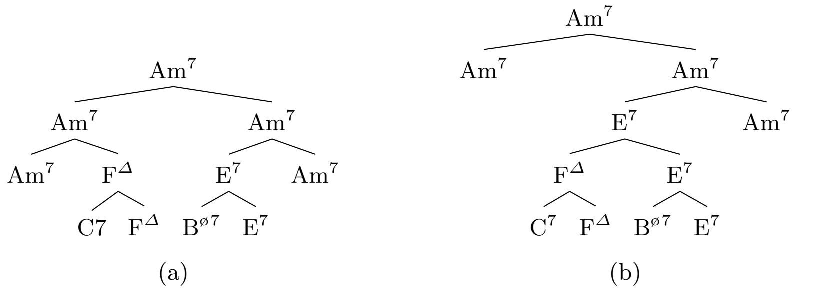

Fig. 2: Beginning of Sunny by Bobby Hebb: (a) predicted and (b) annotated tree. N-PCFG correctly identifies B^{ø7} - E^7 - Am^7 as a ii - V - I progression.

π_r. Starting with the root symbol, the rules can be recursively applied to arrive at a binary tree with only chord symbols as leaves. The probability of a parse tree t is given by the product of the probabilities of the rules that t consists of: $p(t) = \prod_{r \in t} \pi_r$. Assuming that all sequences are generated by a PCFG, and that any sequence s might have several (exponentially in the sequence length L) parse trees of which the leaves form s, we get a probability distribution over sequences: $p(s) = \sum_{t \in \mathcal{T}(s)} p(t)$, where $\mathcal{T}(s)$ is the set of parse trees of s. We use the neural parameterization of [8], in which rule probabilities π_r are computed by MLP's from embeddings representing each root, non- or preterminal symbol and each chord symbol. During training, we simply maximize the likelihood of sequences under $p(s)$. The sum over exponentially many latent trees is computed with the inside algorithm [1]. At inference, we find the optimal tree using Viterbi or Minimum Bayes Risk decoding [12,4].

3 Results

Neural PCFG's (N-PCFG) learns viable structures that overlap with annotations (compare to Random in table 1). Training on more data (... + ChoCo [2]) and with an extra loss that incentivizes chord groups based on the musical 5$^{\text{th}}$ relation (... + Prog. loss) both help. There remains a considerable gap with supervised prediction (MuDeP [3]) and annotations. N-PCFG offers alternative viable explanations: F$^\Delta$ as degree VI of the Am key in fig. 2a versus as tritone substitution for E^7's relative dominant B^7 in fig. 2b according to annotation.

Model	**Train data**	**F1**
N-PCFG	JHT	.387
N-PCFG	... + ChoCo	.455
... + Prog. loss		.477
MuDeP	JHT	.623
Random		.178

Table 1: Test F1 of unsupervised N-PCFG, supervised MuDeP and random predictions on JHT corpus [6].

Conclusion. Unsupervised induction and parsing of harmonic syntax trees from chord sequences is viable but hard.

Acknowledgments. This work is part of the CALCULUS project, which is funded by the ERC Advanced Grant H2020-ERC-2017 ADG 788506.[1]

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Baker, J.K.: Trainable grammars for speech recognition. The Journal of the Acoustical Society of America **65**(S1), S132–S132 (1979)
2. de Berardinis, J., Meroño-Peñuela, A., Poltronieri, A., Presutti, V.: Choco: a chord corpus and a data transformation workflow for musical harmony knowledge graphs. Scientific Data **10**(1), 641 (2023)
3. Foscarin, F., Harasim, D., Widmer, G.: Predicting music hierarchies with a graph-based neural decoder. In: Proceedings of the 24th International Society for Music Information Retrieval Conference, ISMIR. pp. 425–432 (2023). https://doi.org/10.5281/ZENODO.10265315
4. Goodman, J.: Parsing algorithms and metrics. In: 34th Annual Meeting of the Association for Computational Linguistics. pp. 177–183. Morgan Kaufmann Publishers / ACL (1996). https://doi.org/10.3115/981863.981887
5. Hamanaka, M., Hirata, K., Tojo, S.: Musical structural analysis database based on gttm. In: Proceedings of the 15th Conference of the International Society for Music Information Retrieval (ISMIR). pp. 325–330. ISMIR (2014)
6. Harasim, D., Finkensiep, C., Ericson, P., O'Donnell, T.J., Rohrmeier, M.: The jazz harmony treebank. In: Proceedings of the 21th International Society for Music Information Retrieval Conference, ISMIR. pp. 207–215 (2020)
7. Harasim, D., Rohrmeier, M., O'Donnell, T.J.: A generalized parsing framework for generative models of harmonic syntax. In: Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR. pp. 152–159 (2018)
8. Kim, Y., Dyer, C., Rush, A.M.: Compound probabilistic context-free grammars for grammar induction. In: Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL, Volume 1: Long Papers. pp. 2369–2385. ACL (2019). https://doi.org/10.18653/V1/P19-1228
9. Lerdahl, F., Jackendoff, R.S.: A Generative Theory of Tonal Music. MIT press (1983)
10. Rohrmeier, M.: Towards a generative syntax of tonal harmony. Journal of Mathematics and Music **5**(1), 35–53 (2011)
11. Rohrmeier, M.: The syntax of jazz harmony: Diatonic tonality, phrase structure, and form. Music Theory and Analysis (MTA) **7**(1), 1–63 (2020)
12. Sakai, I.: Syntax in universal translation. In: Proceedings of the International Conference on Machine Translation and Applied Language Analysis (1961)
13. Schenker, H.: Der freie satz. Universal Edition, Vienna (1956)
14. Steedman, M.J.: A generative grammar for jazz chord sequences. Music Perception **2**(1), 52–77 (1984)
15. Winograd, T.: Linguistics and the computer analysis of tonal harmony. Journal of Music Theory **12**(1), 2–49 (1968)

[1] https://calculus-project.eu/