Investigating Unsupervised Constituency Parsing for Music Harmony

Ningxiang Xie, Eamonn Bell, Robert Lieck

ICCCM2025, Aalborg

Durham University

Aim & Motivations

Music harmony exhibits a hierarchical, recursive structure and has been formulated with context-free grammars [1]. **Can such a structure be induced without labels** - via unsupervised constituency parsing (UCP)? In natural language, UCP has successful models (e.g., DIORA [2]). Based on the similarity between language and harmony, we evaluate DIORA for harmony.

Conclusions & Future Work

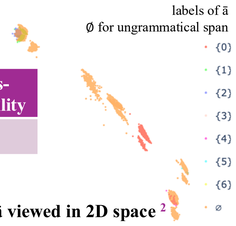
- There is a solution in DIORA's parameter space which satisfies both ideal reconstruction and correct parsing, but without supervision, DIORA's parsing is incorrect due to its flexibility
- When supervised to score gold trees (by a large margin)[8], the grammar DIOR learned is consistent with the generating one
- We explored multiple constraints without success; more direct methods to impose constraints are planned for future work.
- We built analysis tools (e.g., class-separability and rule extraction), providing insights into DIORA's behaviour

Main Results

- **Reconstruction only**

  ① Incorrect parses despite effective optimisation of the reconstruction objective ② Low parse self-similarity across training runs under identical conditions ③ Low consistency between the inside vector ($\bar{a}$) and ground-truth nonterminal set (from the PCFG inside–outside algorithm) per span

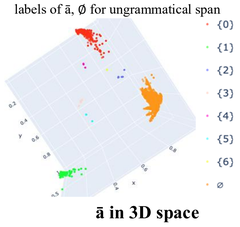
PARSEVAL (initiated [1])	PARSEVAL (trained)	Parse Self-Similarity	$\bar{a}$ Class-Separability
0.13	0.12	0.23	0.51



ā viewed in 2D space [2]

- **+ Parse supervision**

  ① Exact parses (PARSEVAL=1.0) ② Consistency between $\bar{\mathbf{a}}$ vectors and the ground-truth nonterminal sets per span ($\bar{\mathbf{a}}$ class-separability > 0.99) ③ DIORA-extracted rules matching the generating grammar



ā in 3D space

Extracted rules - (in-grammar, T/F)
In-grammar rules are ranked highest

Idx.	Rule	Score	T/F	Idx.	Rule	Score	T/F
1	6 5 → 5	1.85	T	6	1 0 → 0	1.20	T
2	4 3 → 3	1.82	T	7	1 4 → ∅	0.74	F
3	3 2 → 2	1.80	T	8	6 1 → 1	0.65	F
4	2 1 → 1	1.71	T		…	…	
5	5 4 → 4	1.30	T	71	∅ 0 → 4	-1.67	F

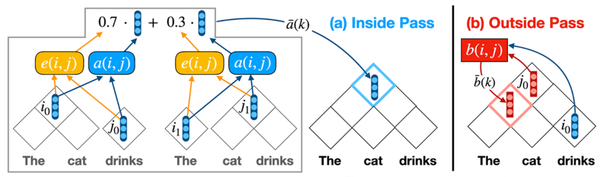
- **+ Constraints**: still incorrect parse trees. Detailed results omitted from the poster

Notes

1. PARSEVAL initialised and trained are mean values for 10 training runs
2. In the reconstruction-only condition, $\bar{\mathbf{a}}$ is plotted in 2D for better visualisation
3. We are aware that PCA is the more principled method for the purpose.
4. The span labels in the plots were produced with the 1-NN classifier.
5. In practice, we implemented this only when $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$ are consistent with ground-truth sets of nonterminals
6. We extracted rules for both the inside and outside passes. For brevity, only describe the inside
7. We experimented with the author's reconstruction objective, which is based on negative sampling, but mainly used another we designed based on ground-truth terminals
8. In practice, we not only applied a margin but ensured a large margin. If not, we can have parsing correct, but did not get good class-separability for $\bar{\mathbf{a}}$

Methods

DIORA



This illustration is extracted from the original work.

- Self-supervised: train with a reconstruction objective to learn span features directly used for parsing without supervision
- Inside-outside pass: same as PCFG inside–outside algorithm regarding how spans are indexed in chart parsing, differs significantly in ① training objective (reconstruction: outside leaf spans match corresponding inside leaf spans) ② span contents (sub-symbolic vectors $\bar{\mathbf{a}}$ for inside and $\bar{\mathbf{b}}$ for outside), ③ weights for split aggregation (i.e., e for inside and f for outside, only in DIORA)

Data

- By an unambiguous toy harmonic PCFG data:
  $S \rightarrow N_0$ [p = 1]
  $\forall\, n \leq 6$: $N_n \rightarrow N_{n+1}\ N_n$ [p = 0.5] | n [p = 0.5]; $N_7 \rightarrow 7$ [p = 1]
- Size: Train: 16000 (len ∈ [9, 24]); Test: 3000 (len ∈ [25, 27])
- Content: sequences, gold trees, and per-span ground-truth non-terminal sets from the PCFG inside–outside algorithm as span labels
- Sequence example and its gold tree

(1 (((3 2) (2 1)) 0))

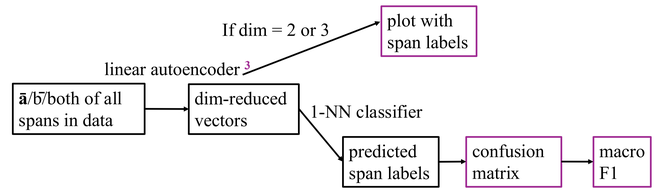
Evaluation Metrics

- Full-tree correctness: Exact-match of DIORA's predicted trees to gold trees (unlabelled)
- PARSEVAL(micro/corpus-level): Unlabelled, against gold trees, skip leaf and root spans. Measured for both randomly initialised and trained models
- Self-similarity: Compute PARSEVAL between predicted trees from every pair of training runs under identical conditions. Then take the mean over all pairs (no gold trees)

Analysis Tools

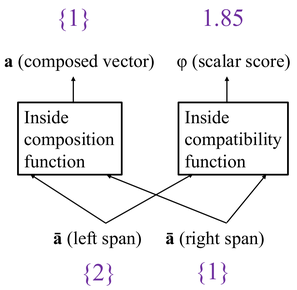
Class-Separability Analysis

- What & Why: Given the parallel between DIORA and PCFG, we ask if DIORA's $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$ span vectors encode the same sets of nonterminals as the PCFG inside-outside algorithm
- How: We analyse three conditions: $\bar{\mathbf{a}}$, $\bar{\mathbf{b}}$ and joint. In each condition, vectors for all spans in the dataset are linearly reduced in dim. When reduced to 2 or 3, we plot the vectors with their span labels[4] to assess separability qualitatively. For dim ≥ 3, we quantify separability via multi-class classification performance (confusion matrix and macro F1) using a 1-NN classifier



Rule Extraction

- What & Why: Extracting rules from DIORA[5] facilitates understanding what structure it learned to reconstruct the data, further sheds light on its parsing behaviour
- When: $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$ are distributed in a small number of clusters, so it's possible to label them with symbols[5]
- How[6]: ① Label the clusters of $\bar{\mathbf{a}}$ ② "Symbolise" the inside composition function, yielding rule in the form X Y → Z ③ Record the compatibility score of the current usage of the rule ④ Extract all calls of the composition and compatibility functions when DIORA processes the dataset. For each rule, collect its scores and number of usages, and rank rules by their mean scores



Training Conditions

Reconstruction-only

- Outside-pass leaf-span representations are encouraged to align with the corresponding inside-pass representations at the same step[7]

+ Parse Supervision

- By a margin loss that raises the score of the ground-truth tree above the scores of other candidates[8]

+ Constraints

Add Noise

- Method: Add Gaussian Noise to the inside and outside vectors. Observe the parsing performance at different noise levels

Extend Reconstruction to Non-Leaf Spans

- Method: "All-span" reconstruction objective by [3]

Lower the Temperature of the Inside Softmax Function

- Method: Set T = 0.01 for the inside softmax to approximate argmax

+ Span Classification with Separability

- Method: Train two classifiers (inside/outside) to predict nonterminal sets from span vectors ($\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$); use ArcFace loss to encourage class-separability

+ Span Grammaticity

- Method: Apply a margin loss to raise the validity scores ($\bar{\mathbf{e}}$) of valid spans and penalise those of invalid spans

Reference

[1] M. Rohrmeier, "Towards a generative syntax of tonal harmony," Journal of Mathematics and Music, vol. 5, no. 1, pp. 35–53, Mar. 2011
[2] A. Drozdov, P. Verga, M. Yadav, M. Iyyer, and A. McCallum, 'Unsupervised Latent Tree Induction with Deep Inside-Outside Recursive Auto-Encoders', in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long andShort Papers), 2019, pp. 1129–1141
[3] R. Hong, J. Cai, and K. Tu, 'Deep Inside-outside Recursive Autoencoder with All-span Objective', in Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 3610–3615.