Learning Hierarchical Metrical Structure Beyond Measures

ISMIR 2022 Poster Presentation

Junyan Jiang[12] , Daniel Chin[12], Yixiao Zhang [13] , Gus Xia[12]

jj2731@nyu.edu, daniel.chin@nyu.edu, yixiao.zhang@qmul.ac.uk, gxia@nyu.edu

[1] Music X Lab, NYU Shanghai [2] Mohamed bin Zayed University of Artificial Intelligence [3] Centre for Digital Music, QMUL

Project Repository

https://github.com/music-x-lab/Hierarchical-Metrical-Structure

All annotations, codes & pre-trained models are available!

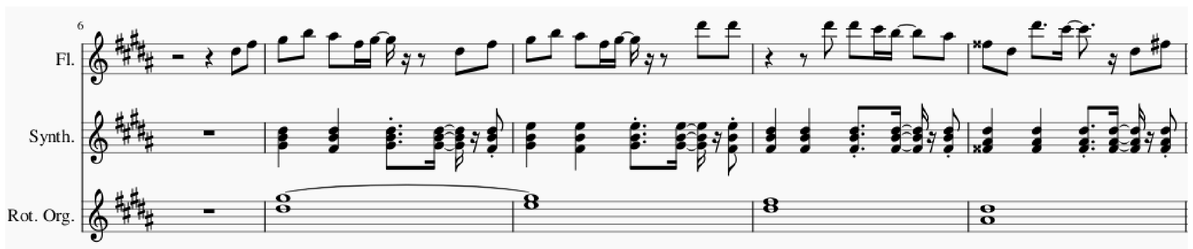
Introduction

- Music contains strong-weak rhythmic pulses above measures
- **Generative Theory of Tonal Music (GTTM)** formally discussed the notation of hierarchical metrical structure using different number of dots to indicate metrical boundaries of different levels
- **Metrical structure analysis** is important for down-stream MIR, computer musicology and music generation tasks
- Previous **hierarchical metrical structure analyzers** are typically
  - Designed for *low-level* metrical layers
  - Designed for *monophonic*, *short* snippets (e.g., 8 bars)
  - Focusing on low-level syntax, ignoring high-level semantic properties of music

Task Formulation

Input: multi-voiced full-song score (multi-track MIDI file), with downbeat annotation

Output: 4-layer metrical structure labeling above measures



Output:

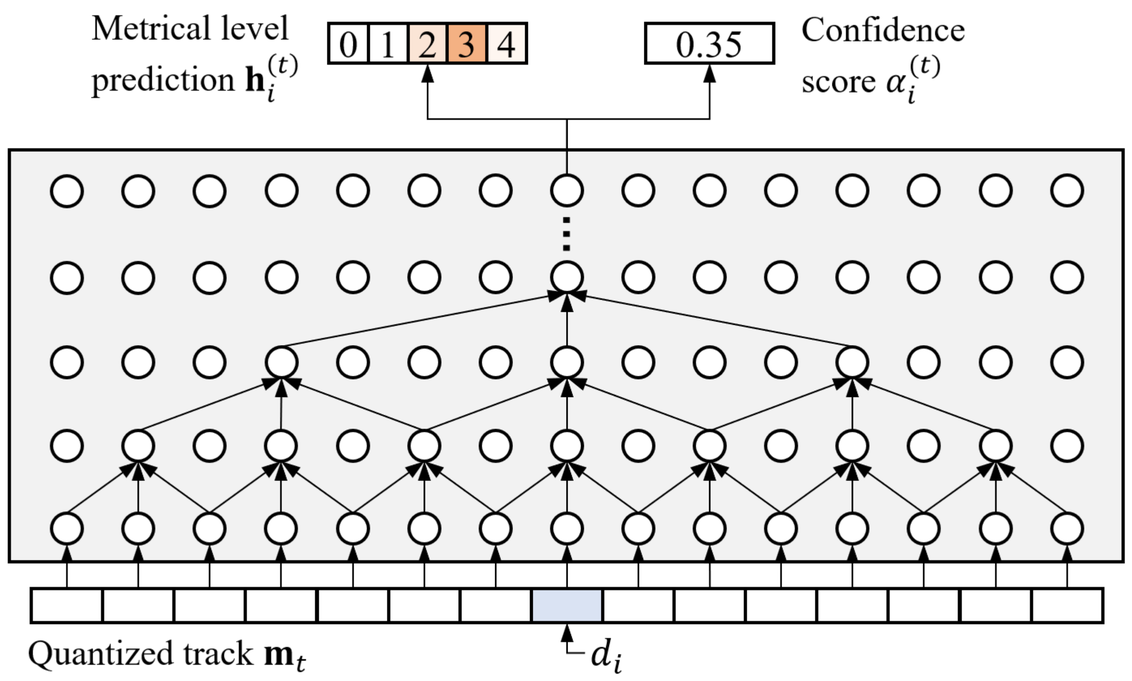
Binary Regularity

- Higher-level metrical structures are ambiguous and less stable
  - Hypermeter changes occurs much more often than meter changes
- Still, **binary regularity** is a very important inductive bias

16 bars															
8 bars								8 bars							
4 bars				4 bars				4 bars				4 bars			
2 bars		2 bars		2 bars		2 bars		2 bars		2 bars		2 bars		2 bars	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

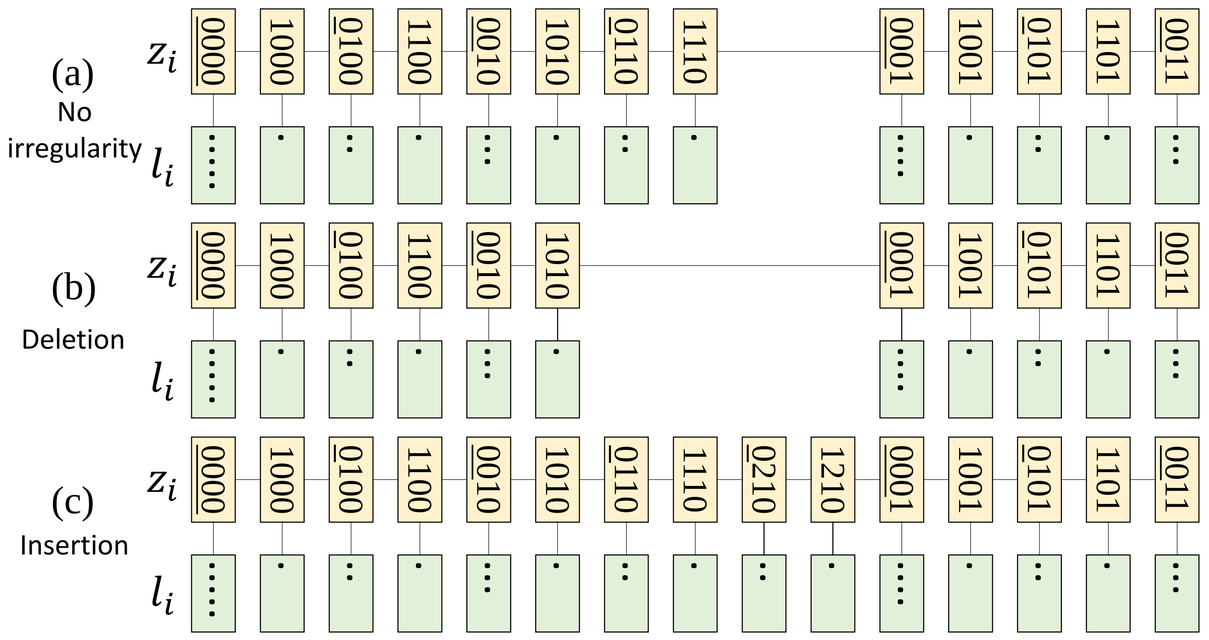
Method: Temporal Convolutional Network

- Dilations: 1, 2, 4, …, 64 (strongly encourages information exchanges in the context of **binary regularity**)



Method: Conditional Random Field

- We allows binary irregularity (b, c) with a certain degree of penalty



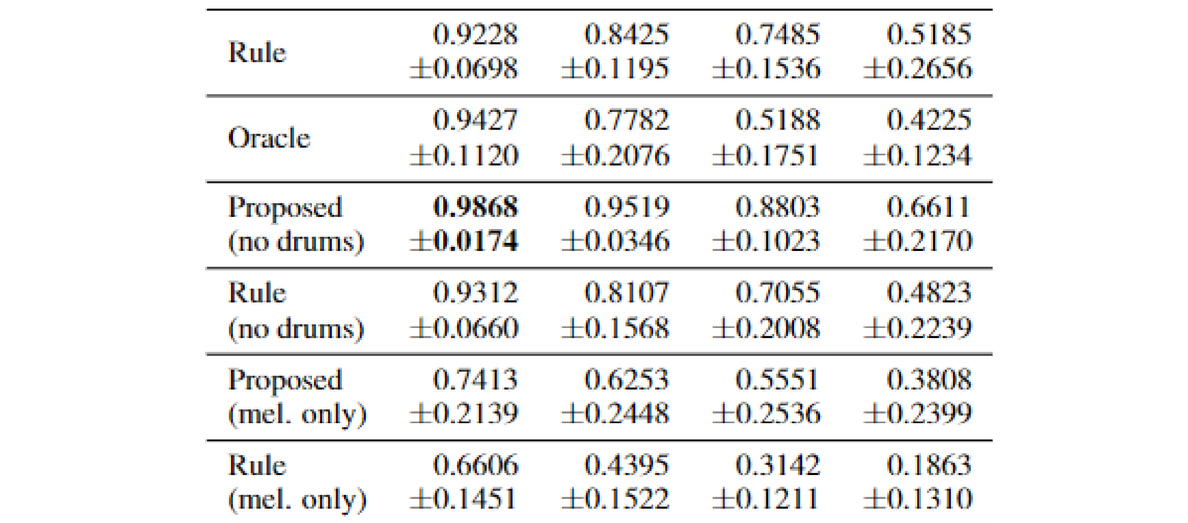
Experiments & Main Results

- Dataset: 70 songs in RWC Pop (annotations available in the repo)
- Good results on full scores for pop songs
- Not satisfactory under simplified instrumentations (e.g., melody only, piano rearrangement)
- Lack of annotated training data limits the model's performance

Model	Level 1	Level 2	Level 3	Level 4
Proposed	0.9848 ±0.0215	**0.9559 ±0.0386**	**0.8880 ±0.0889**	**0.6849 ±0.1900**
Proposed w/o CRF	0.9338 ±0.0390	0.8528 ±0.0937	0.7971 ±0.1276	0.6646 ±0.0844
Rule	0.9228 ±0.0698	0.8425 ±0.1195	0.7485 ±0.1536	0.5185 ±0.2656
Oracle	0.9427 ±0.1120	0.7782 ±0.2076	0.5188 ±0.1751	0.4225 ±0.1234
Proposed (no drums)	**0.9868 ±0.0174**	0.9519 ±0.0346	0.8803 ±0.1023	0.6611 ±0.2170
Rule (no drums)	0.9312 ±0.0660	0.8107 ±0.1568	0.7055 ±0.2008	0.4823 ±0.2239
Proposed (mel. only)	0.7413 ±0.2139	0.6253 ±0.2448	0.5551 ±0.2536	0.3808 ±0.2399
Rule (mel. only)	0.6606 ±0.1451	0.4395 ±0.1522	0.3142 ±0.1211	0.1863 ±0.1310

Binary Irregularity Detection

											16 bars															
10 bars										8 bars								8 bars								
4 bars				6 bars						4 bars				4 bars				4 bars				4 bars			
2 bars		2 bars		2 bars		2 bars		2 bars		2 bars		2 bars		2 bars		2 bars		2 bars		2 bars		2 bars		2 bars	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1



Future Work

- Semi-supervised learning / self-supervised learning
- Audio metrical structure analysis
- End to end analysis with performance to score / audio beat tracking