Beat and Downbeat Tracking of Symbolic Music Data Using Deep Recurrent Neural Networks

Yi-Chin Chuang, Li Su

Music & Culture Technology Lab, Institute of Information Science, Academia

APSIPA ASC 2020, New Zealand

Outline

❶ Problem definition

- What is beat/downbeat tracking?
- What is symbolic music?
- Why symbolic music beat tracking?

❷ Proposed method

- Data representation
- Network architecture

❸ Experiment and discussion

- Dataset
- Baseline
- Results
- Conclusions

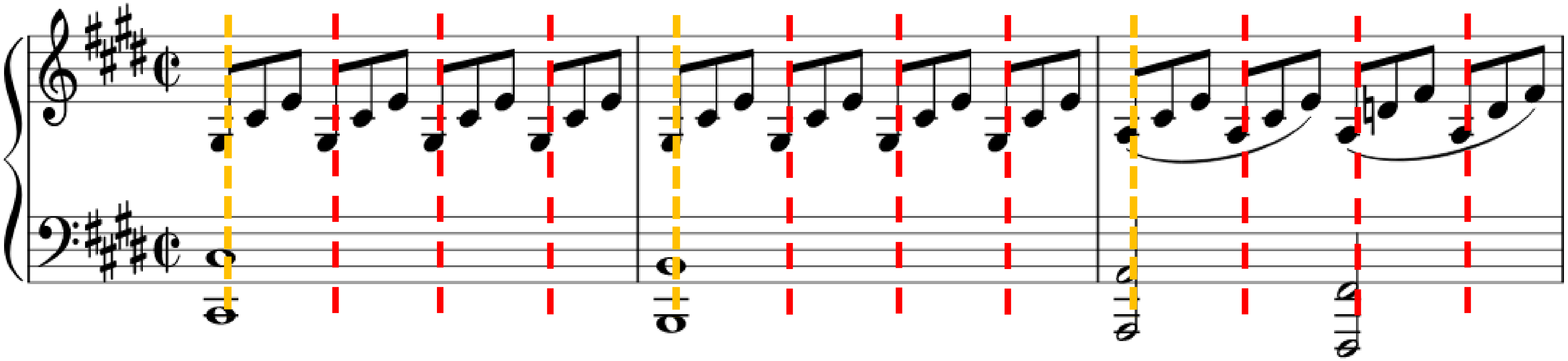
Problem definition – What is beat/downbeat tracking?

Beat

The beat is defined as the rhythm listeners would tap their toes to when listening to a piece of music

Downbeat

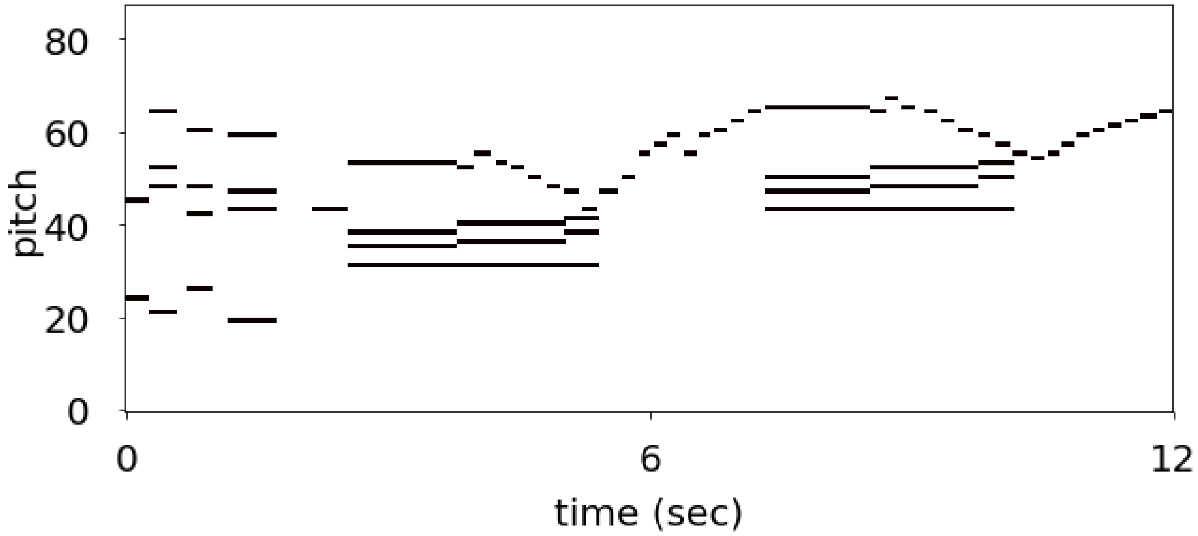
The first beat of bar



Problem definition – What is symbolic music?

- Any kinds of score representation with an explicit encoding of musical entities (notes, chords, intervals, …)
- Example: MIDI, MusicXML, Piano roll format

```
<note>
  <pitch>
    <step>E</step>
    <alter>-1</alter>
    <octave>4</octave>
  </pitch>
  <duration>2</duration>
  <type>half</type>
</note>
```

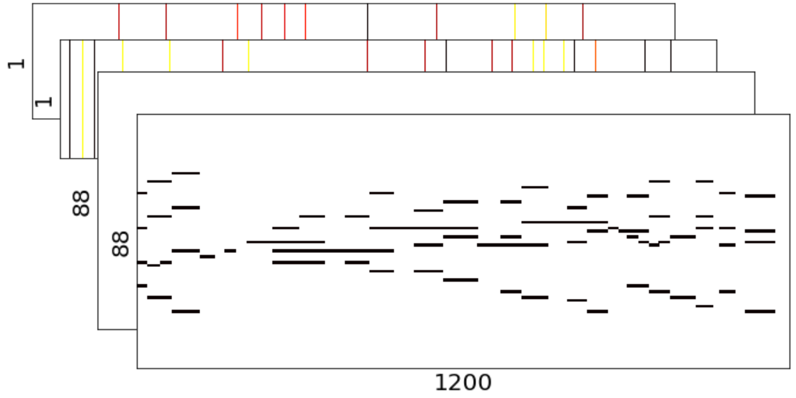


Problem definition – Why symbolic music beat tracking?

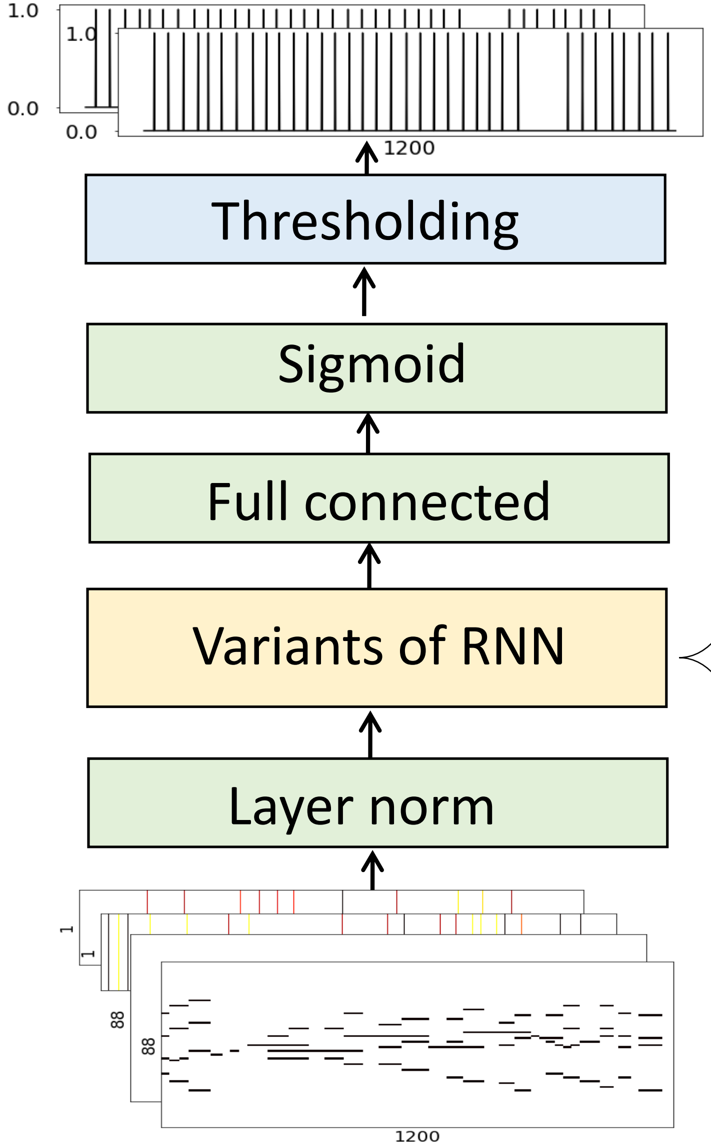
- Beat represents a key portion of music language modeling (infer the temporal structure)
- Beat tracking has various practical application (score parsing, automatic accompaniment, music generation…)
- However, most of the beat tracking methods are designed for audio signals only
- Challenge: lack of information on music accents and dynamics

Proposed method – Data representation

- Directly derived from MIDI data: onset, duration, pitch
- A music clip contains 4 parts: pitch profile, onset profile, spectral flux, inter-onset interval (IOI)
  - $[p_t,\ o_t,\ s_t,\ i_t] \in R^{178}$, frame rate = 100 Hz
- Segment music clip into overlapped packed sequence
  - sequence length = 12 (sec)
  - overlapped length = 6 (sec)



Proposed method – Network architecture



Output data: beat (1 x 1200), downbeat (1 x 1200)

- Bidirectional LSTM (BLSTM)
- BLSTM + Attention mechanism (BLSTM-Attn)
- Hierarchical Multiscale LSTM (HM-LSTM)

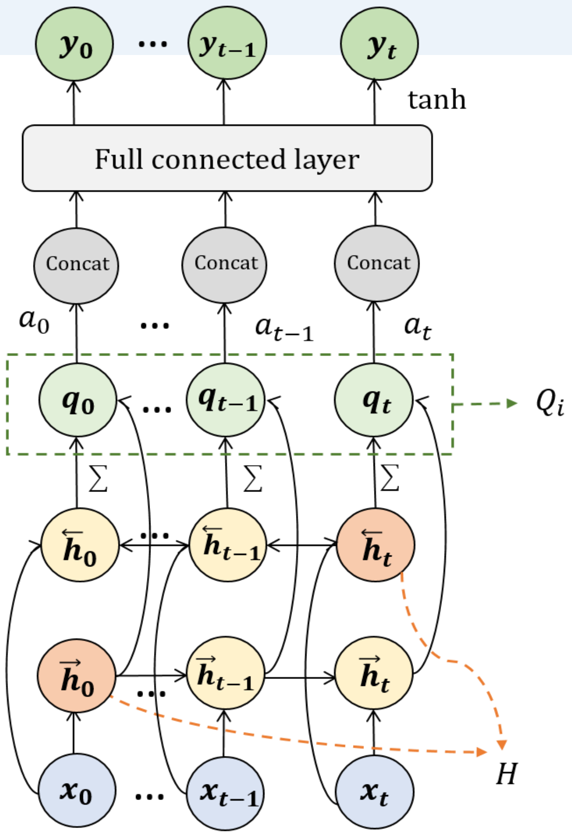
Input data (178 x 1200)

Assumptions – Why (we think) would it work?

Beat/downbeat tracking is...

Sequence-to-sequence task, possess hierarchical structure in music theory

- Bidirectional LSTM (BLSTM): classic neural network that is used to analyze sequential data prediction
- Attention mechanism: focus at each time step on certain elements of the sequence data



Assumptions – Why (we think) would it work?

Problem definition

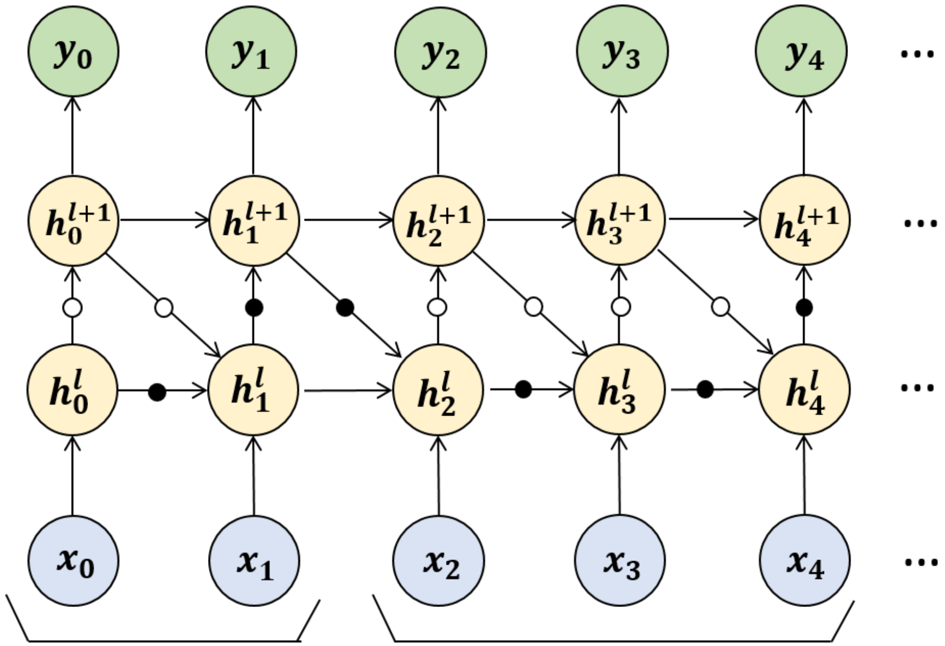
Proposed method

Experiment and discussion

Beat/downbeat tracking is…

Sequence-to-sequence task, possess hierarchical structure in music theory

- Hierarchical Multiscale LSTM (HM-LSTM): capture the hierarchical structure with different time-scale in sequential data



Dataset – MusicNet

- A collection of 330 freely-licensed classical music recordings, with over 1M annotated labels for each note in every recording.

	Training set	Validation set	Testing set
Number of songs	111 songs	12 songs	31 songs
Length	12h 23m	1h 27m	3h 16m

Baseline

- madmom (audio-based): predicts the beat positions with RNNs and a dynamic Bayesian network approximated by a Hidden Markov Model
- librosa (audio-based): computes onset envelop, then uses dynamic programming algorithm to perform beat tracking
- pretty_midi (symbolic-based): predicts the beat positions according the MIDI tempo changes

Experiment results – Beat tracking

Method		Precision	Recall	F1-score
Proposed	BLSTM	0.520	**0.724**	**0.605**
	BLSTM-Attn	**0.522**	0.715	0.603
	HM-LSTM	0.513	0.675	0.583
Baseline	madmom (syn)	0.497	0.641	0.560
	madmom (real)	0.427	0.547	0.480
	librosa (syn)	0.388	0.600	0.471
	librosa (real)	0.277	0.394	0.325
	pretty_midi	0.207	0.303	0.246

Experiment results – Downbeat tracking

Method		Precision	Recall	F1-score
Proposed	BLSTM	0.262	0.448	0.331
	BLSTM-Attn	0.264	0.466	**0.337**
	HM-LSTM	0.198	**0.643**	0.303
Baseline	madmom (syn)	**0.319**	0.641	0.190
	madmom (real)	0.286	0.547	0.186
	pretty_midi	0.067	0.303	0.072

Experiment results (1)

Beethoven’s String Quartet in A major No. 5, Op. 18 , II. Menuetto

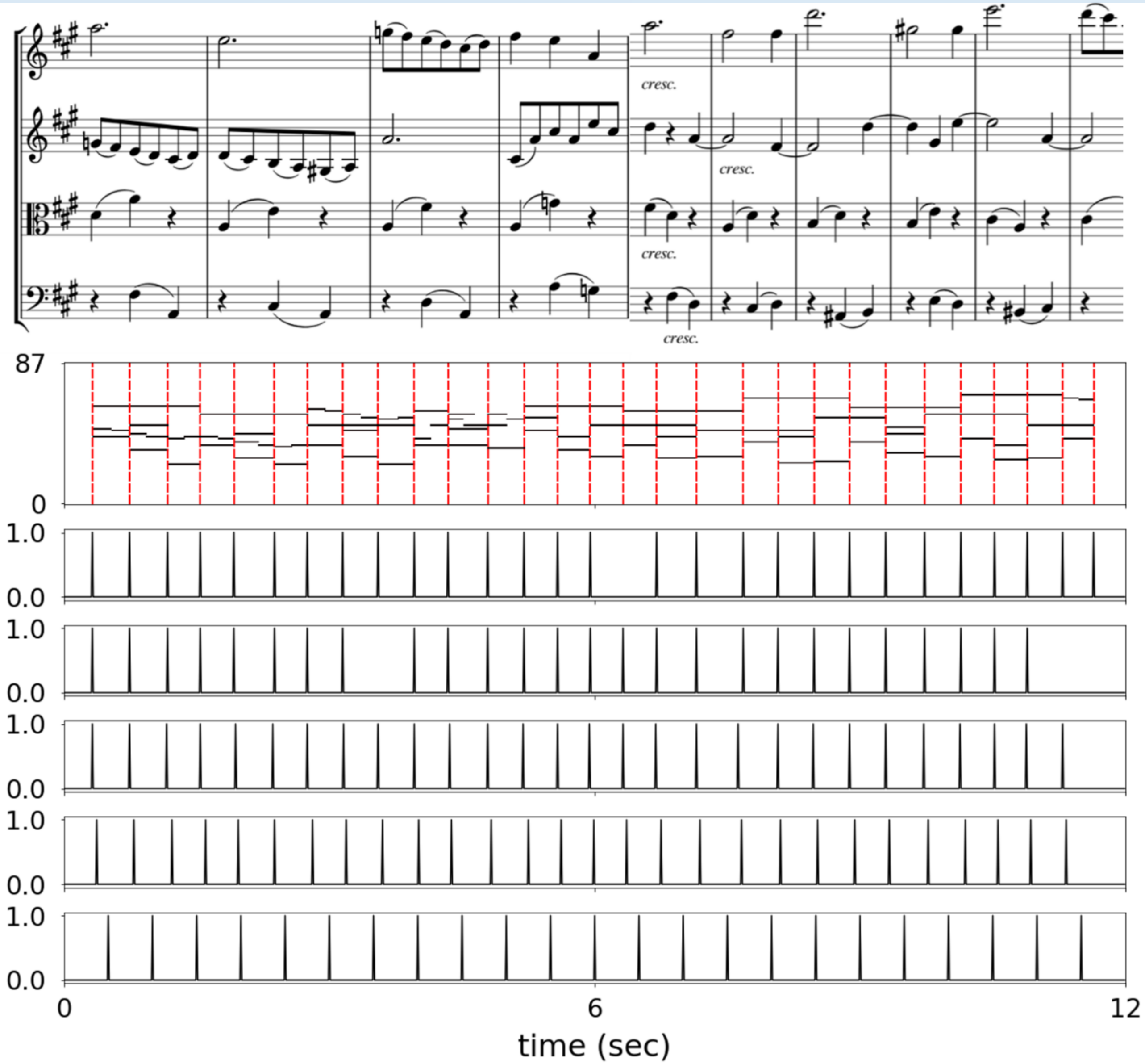
BLSTM

HM-LSTM

madmom (syn)

librosa (syn)

pretty_midi



Experiment results (2)

Beethoven's Piano Sonata No. 14 (Moonlight) in C-sharp minor, Op. 27, I. Adagio sostenuto

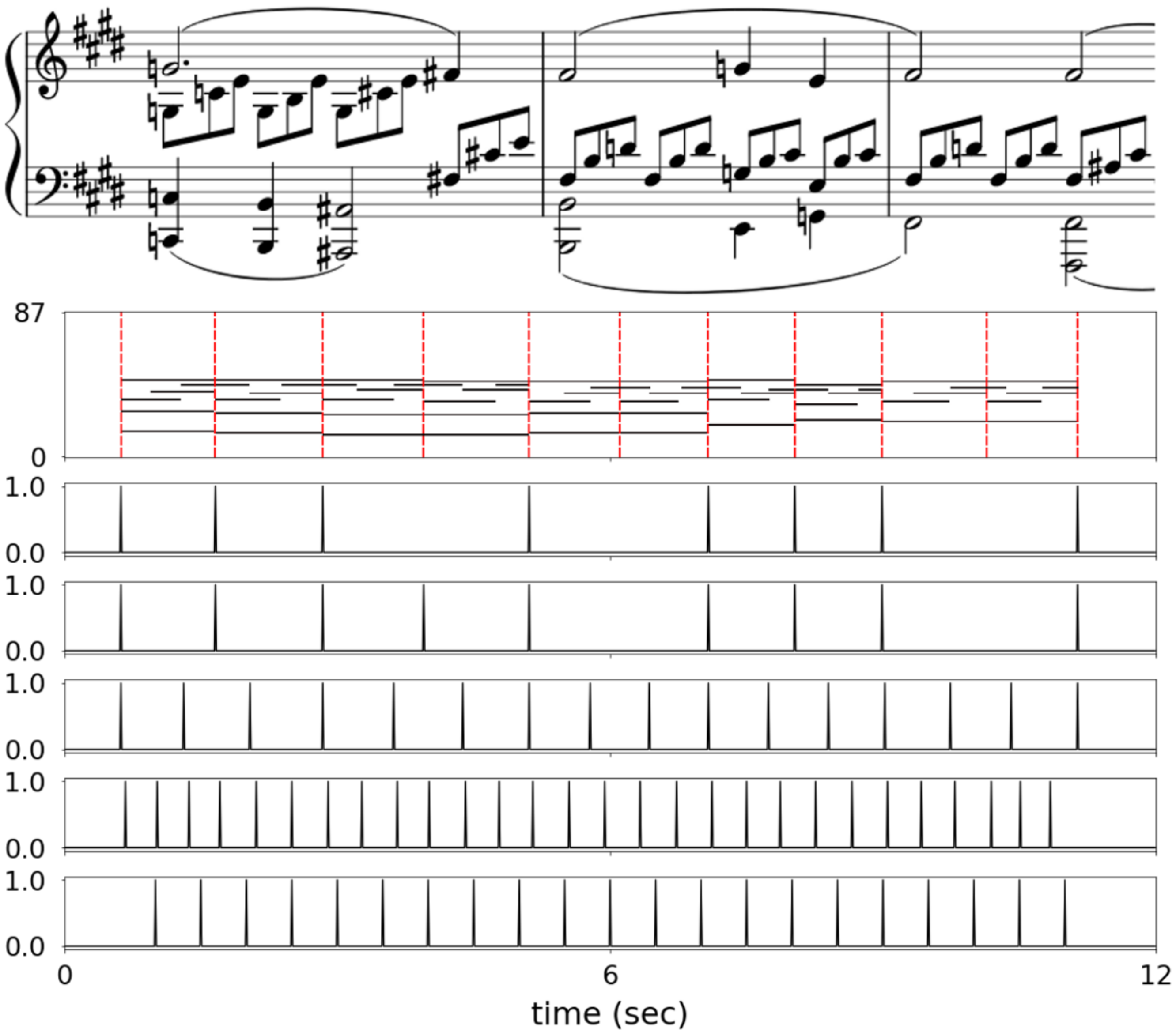
BLSTM

HM-LSTM

madmom (syn)

librosa (syn)

pretty_midi



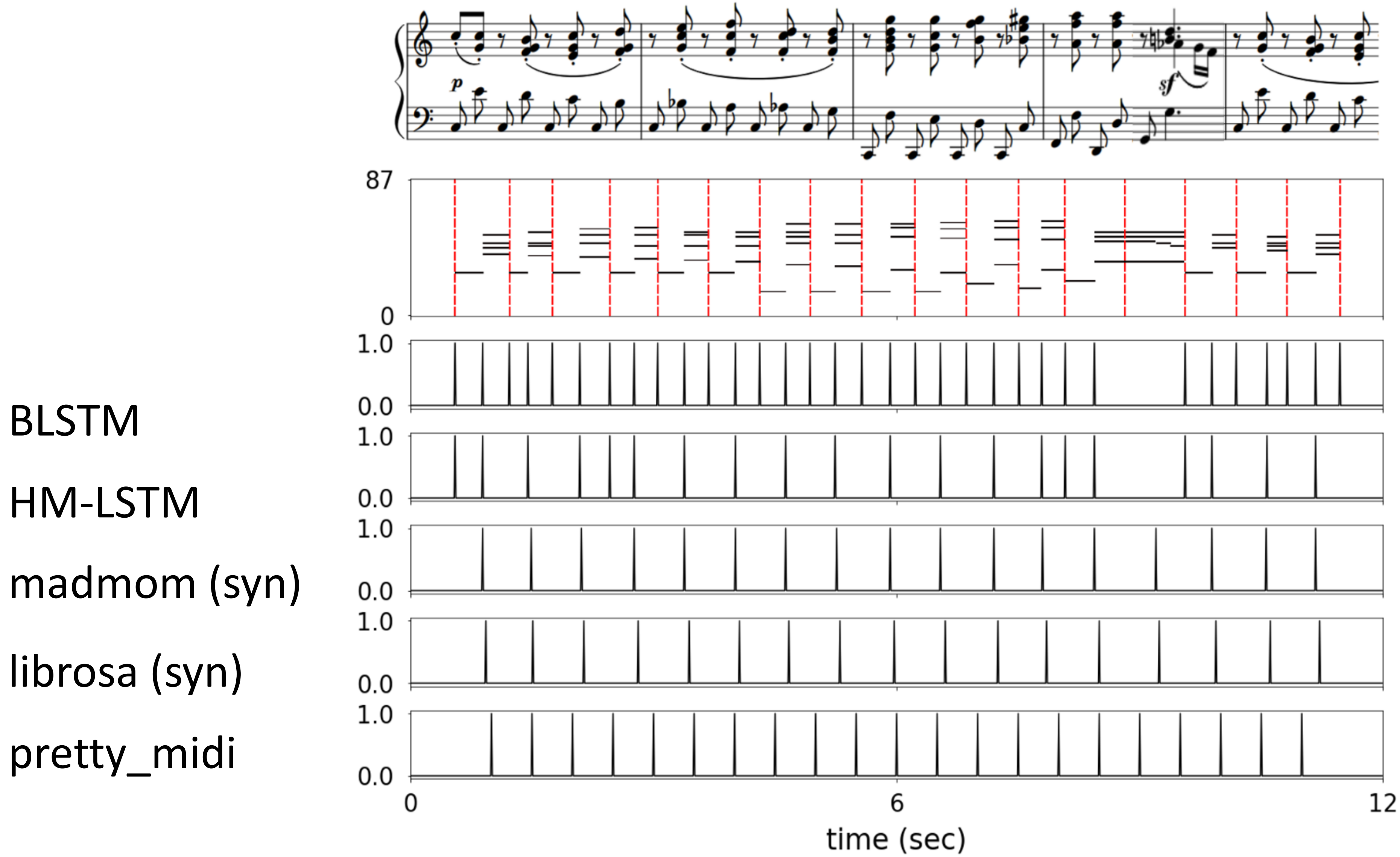
Experiment results (3)

Beethoven’s Piano Sonata No. 10 in G Major, Op. 14, II. Andante

Problem definition

Proposed method

Experiment and discussion



Conclusions

- Symbolic beat tracking task should be independent from audio beat tracking task
  - Construct a specific model upon the symbolic data is necessary
- BLSTM-based methods can be adopted in the application
- Future work: redesign the input data representation, construct advance sequence-to-sequence models

Reference

- S. Bock, F. Krebs, and G. Widmer, “Joint beat and downbeat tracking with recurrent neural networks,” in ISMIR. New York City, 2016, pp. 255–261.

- D. Ellis, “Beat tracking by dynamic programming,” Journal of New Music Research, vol. 36, no. 1, pp. 51–60, 2007.

- J. Chung, S. Ahn, and Y. Bengio, “Hierarchical multiscale recurrent neural networks,” in 5th International Conference on Learning Representations (ICLR), 2017.