SYMBOLIC MUSIC REPRESENTATIONS FOR CLASSIFICATION TASKS: A SYSTEMATIC EVALUATION

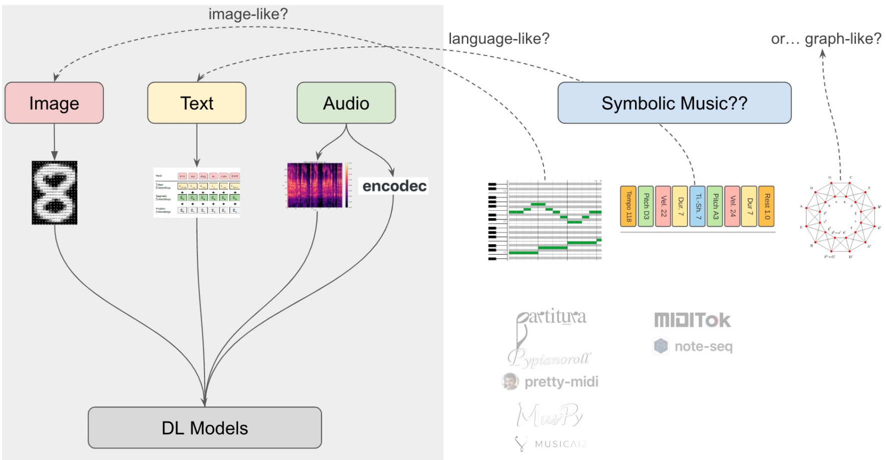
Huan Zhang Emmanouil Karystinaios Simon Dixon Gerhard Widmer Carlos Eduardo Cancino-Chacón

{huan.zhang, s.e.dixon}@qmul.ac.uk
{emmanouil.karystinaios, gerhard.widmer, carlos_eduardo.cancino_chacon}@jku.at

Contributions

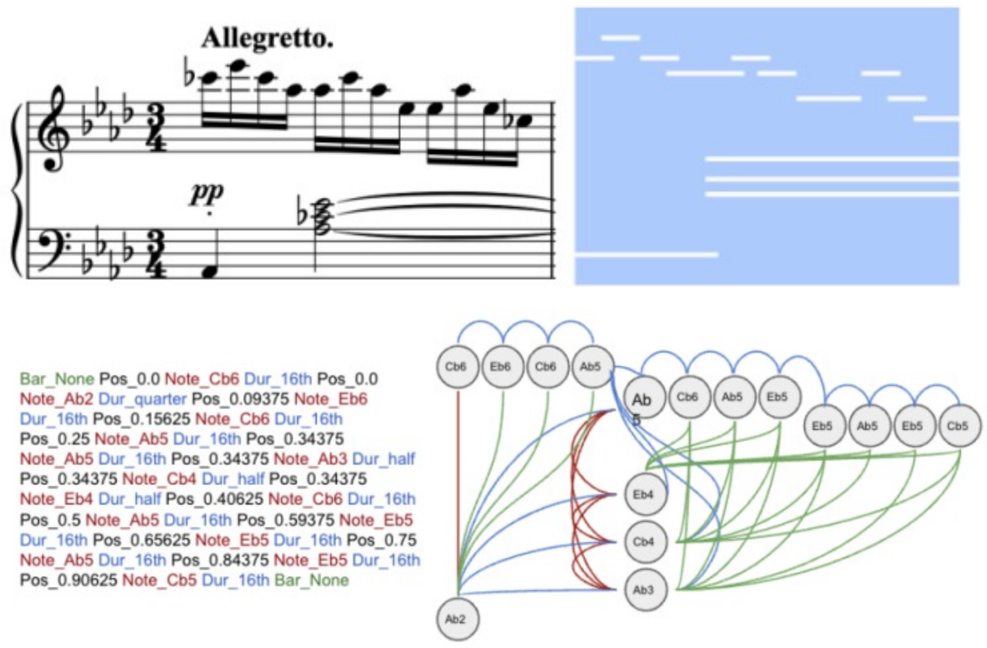
- We investigate the performance and complexity of matrix, sequence and graph input representations, and their corresponding neural architectures (CNN, Transformer, GCN)
- We compare the impact that the different information contained in symbolic scores and performances has on different piece-level classification tasks.
- We introduce a new graph representation for symbolic performances and explore the capability of graph representations in classification tasks.

1. Motivation



2. Methodology

- Representation configuration:
  - Matrix:
    - Resolution and channels
  - Sequence:
    - Encodings
    - Byte pair encoding
  - Graph:
    - Bi-directions
    - Edge relationships
- Information level:
  - Basic: Pitch, onset, duration
  - Advanced: Voicing, markings (score), velocity (perf)
- Architecture
  - Frontend:
    - Matrix - ResNet family
    - Sequence - Transformer
    - Graph - GCN from GraphSAGE blocks
  - Backend: Multihead attention block
- Dataset: ATEPP / ASAP
  - Performance MIDI & Score MusicXML
  - Classification tasks: Composer, Performer, Difficulty



Excerpt of Schubert's *Impromptu Op. 90 No.4* and its input visualizations (from left to right): generic matrix, sequence (REMI-like) and graph.

3. Results

Composer classification results for all representations, on all target subsets of our datasets on the composer classification task using only basic level features.

		ASAP-performance		ASAP-score		ATEPP-performance		ATEPP-score	
		ACC	*F1*	*ACC*	*F1*	*ACC*	*F1*	*ACC*	*F1*
Matrix									
Resl	Chnl								
400	On+Fm	0.59±0.04	0.18±0.02	0.59±0.03	0.18±0.01	0.24±0.05	0.20±0.04	**0.25±0.02**	0.16±0.03
600	On+Fm	0.62±0.06	0.21±0.03	**0.61±0.07**	**0.19±0.02**	0.28±0.01	**0.22±0.03**	0.24±0.02	0.16±0.04
800	Fm	0.62±0.04	**0.21±0.02**	0.58±0.06	0.18±0.03	0.22±0.03	0.17±0.01	0.22±0.02	**0.18±0.03**
800	On+Fm	**0.63±0.04**	0.20±0.01	0.57±0.04	0.18±0.03	**0.28±0.02**	0.22±0.01	0.22±0.04	0.14±0.02
Sequence									
Tokn	BPE								
MidiLike	×	**0.53±0.05**	**0.16±0.02**	N/A	N/A	0.18±0.04	0.10±0.02	N/A	N/A
REMI	×	0.51±0.04	0.15±0.02	0.43±0.04	**0.14±0.01**	**0.23±0.04**	0.10±0.02	0.23±0.04	**0.13±0.02**
CP	×	0.48±0.02	0.09±0.05	**0.45±0.05**	0.10±0.01	0.11±0.02	0.09±0.01	0.17±0.06	0.11±0.04
MidiLike	4	0.52±0.04	0.15±0.02	N/A	N/A	0.17±0.03	0.12±0.01	N/A	N/A
REMI	4	0.51±0.02	0.15±0.01	0.43±0.03	0.13±0.01	0.21±0.01	**0.13±0.03**	**0.23±0.03**	0.13±0.01
Graph									
Bi-dir	Multi-rel								
×	×	0.56±0.01	0.17±0.02	0.51±0.05	0.16±0.02	0.22±0.02	0.10±0.03	0.23±0.03	0.21±0.05
×	✓	0.58±0.03	0.19±0.01	**0.54±0.05**	**0.17±0.02**	**0.27±0.03**	0.13±0.02	**0.29±0.10**	0.18±0.06
✓	✓	**0.62±0.02**	**0.21±0.01**	0.50±0.04	0.17±0.01	0.23±0.04	**0.16±0.03**	0.27±0.06	**0.22±0.03**

4. Conclusion & Takeaways

- Performance:
  - Matrix ≈ Graph > Sequence, but overall achieves similar level of acc
  - Matrix approach trains more robustly, while graph approach the least
  - Graph structures benefit the most from voicing information
- Model complexity:
  - Sequence (12.8M) >> Matrix (4.3M) > Graph (1.3M) (Minimal model that achieve the same result)
- Transformer vs. GNN: Are we learning the same set of musical edges?
  - Not entirely, but we observed some structural similarities
- The *Album Effect*:
  - Multiple interpretations of the same composition may cause information leakage. Happens in existing literature already! (~30% acc boost)

`https://github.com/anusfoil/SymRep`

This work was supported UKRI Centre for Doctoral Training in Artificial Intelligence and Music, funded by UK Research and Innovation [grant number EP/S022694/1], as well as European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme, grant agreement No. 101019375 (Whither Music?).

Engineering and Physical Sciences Research Council

JOHANNES KEPLER UNIVERSITÄT LINZ

Queen Mary University of London