Pre-training music transformer with masked-language model

Yingfeng Fu[1], Yusuke Tanimura[2,1], Hidemoto Nakada[2,1]   1.UNiversity of Tsukuba 2.AIST

Abstract

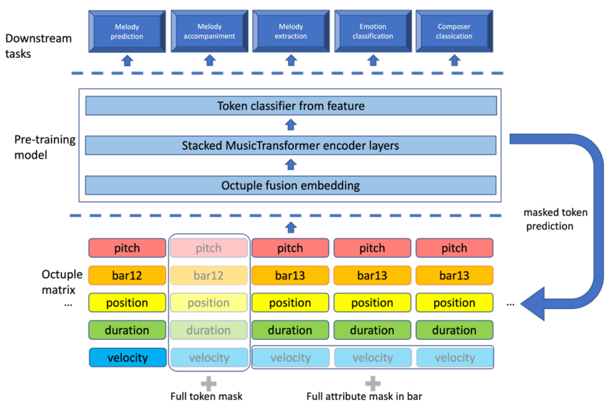
- Works in pre-training for symbolic music are not general enough
- We pre-trained our model with MusicTransformer and compared with previous works under the same condition
- We added three downstream tasks to evaluate our work
- In most of downstream tasks our model works better then the previous works

Backgrond

- PiRhDy: Word2vec like pre-training model
- MusicBERT: Pre-training model using stacked Transformer and masked language model(MLM) strategy
- MusicTransformer: A improved Transformer structure with optimized memory usage, which could capture dependency for extremely long sequence and also focus on the relative relationship among the tokens

Method

- Implement our model using stacked MusicTransformer
- Pre-training PiRhDy, MusicBERT and our model under the same condition(MAESTRO dataset), with MLM
- Finetune and evaluate on the downstream tasks

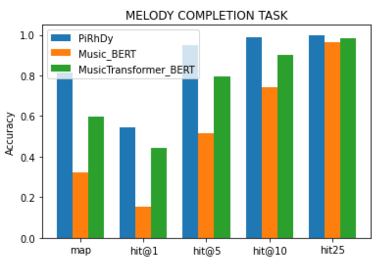


Results on downstream tasks

A.PiRhdy downstream tasks

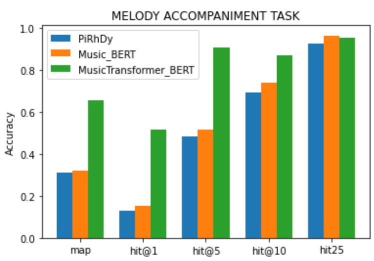
1) Melody completion

- Phrase level task
- Melody : notes from the highest octave
- predict [former melody, latter melody] pair



2) Melody accompaniment

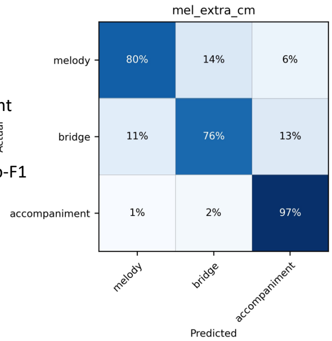
- Phrase level task
- Harmony: rests notes besides melody
- predict [melody, harmony] pair, one melody to multiple harmony



B. Added downstream tasks

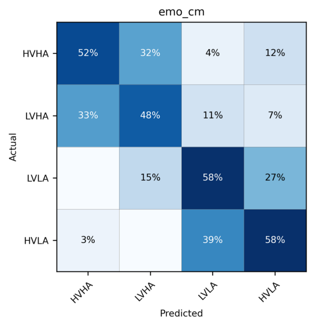
1) Melody extraction:

- Note level task
- Notes in three classes
- Melody, bridge, accompaniment
- Incomparible with PiRhDy, MusicBERT has 0.479 Marco-F1 While our model has 0.849 Marco-F1



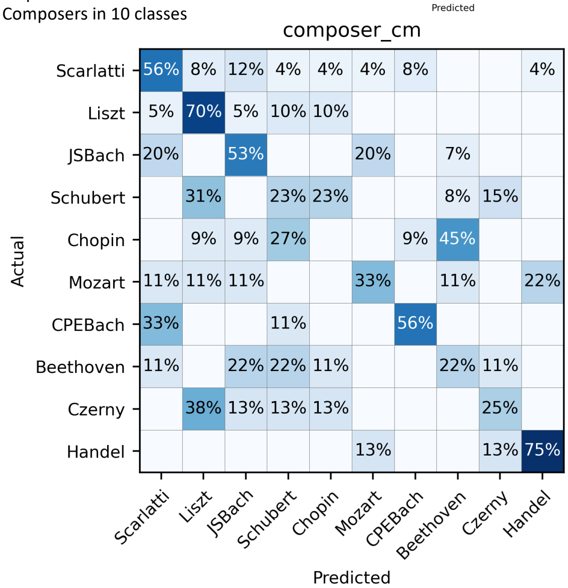
2) Emotion classification:

- Sequence level task
- Emotion in four classes
- V for Valence A for Arousal
- HVHA, LVHA, LVLA, HVLA
- Incomparible with PiRhDy, MusicBERT has 0.466 Marco-F1 While our model has 0.519 Marco-F1



3) Composer classification:

- Sequence level task
- Composers in 10 classes



Conclusion and future work

- We reproduced the MusicBERT model and modified it into the MusicTransformerBERT
- We pre- trained the models with the same number of epochs and then compared the model's results on the existing downstream tasks and our complementary downstream tasks.
- Our model has a better performance in most cases

➢ Fulfill the pre-training tasks
➢ Not only discriminative tasks but also generation tasks
➢ introducing other modalities

Referenece

[1]Yingfeng Fu, Yusuke Tanimura, Hidemoto Nakada, "Improve symbolic music pre-training model using MusicTransformer structure", *IEEE IMCOM2023*.

[2]M. Zeng, X. Tan, R. Wang, Z. Ju, T. Qin, and T.-Y. Liu, "Musicbert: Symbolic music understanding with large-scale pre-training," *arXiv preprint arXiv:2106.05630*, 2021.

[3]H. Liang, W. Lei, P. Y. Chan, Z. Yang, M. Sun, and T. Chua, "Pirhdy: Learning pitch-, rhythm-, and dynamics-aware embeddings for symbolic music," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 574–582.

ACKNOWLEDGMENT: This work was supported by JST SPRING, Grant Number JPMJSP2124 and KAKEN JP19K11994.