Composer Style Classification of Piano Sheet Music Images Using Language Model Pretraining

TJ Tsai, Kevin Ji

Harvey Mudd College, Claremont, CA USA

Abstract

We propose a way to predict the composer of a page of piano sheet music by converting the sheet image into a sequence of "words" and using a text classification approach. By pretraining a language model on a large amount of unlabeled data, we are able to substantially improve performance on the classification task.

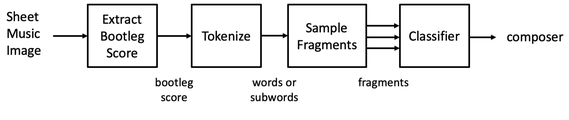
The Problem

Composer style classification

- predict the composer of a piano sheet music image
- classification task with K composers

The Approach

Model architecture:

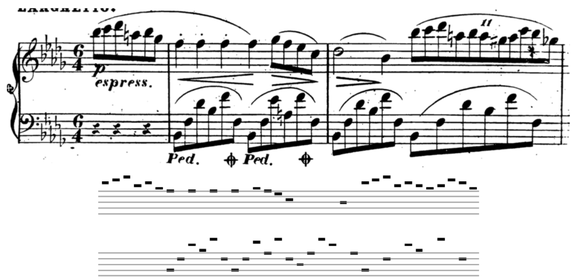


Two key principles:

- use as much data as possible (IMSLP)
- utilize unlabeled data (LM pretraining)

Model Description

1) Extract Bootleg Score



- describes notehead locations in sheet music
- $62 \times N$ binary matrix

2) Tokenize

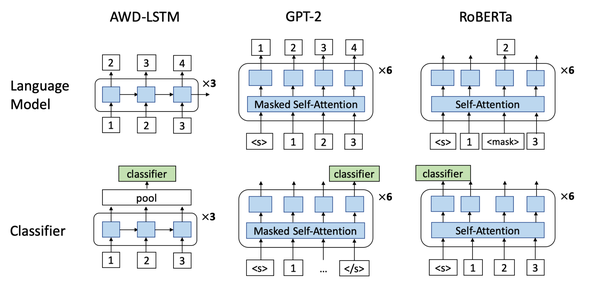
- treat each bootleg score column as a word (word-based models)
- convert each column into sequence of bytes (characters) and perform Byte Pair Encoding to get subwords (subword-based models)

3) Sample Fragments

- Given a (sub)word sequence $x_1, x_2, \cdots, x_N$, sample fragments of fixed length L
- Benefits: significantly augments data, can achieve balanced classes

Model Description (cont'd)

4) Classifiers



- AWD-LSTM, GPT-2, RoBERTa
- pretrained on language modeling task
- at test time, average predictions on all fragments from a single page

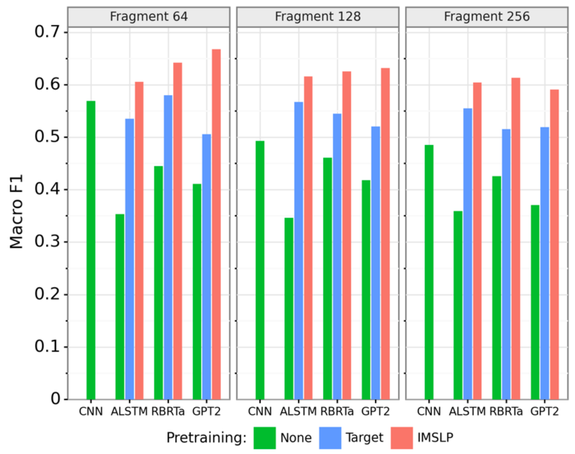
Pretraining conditions:

- no pretraining
- target pretraining: on labeled data only
- IMSLP pretraining: on unlabeled data, finetune on labeled data

Experimental Setup

- Labeled data: 9 composers, 7.1k pages
- Unlabeled: all IMSLP piano scores, 255k pages

Results



- pretraining helps a lot
- Transformer-based models perform best

Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Large-scale computations on IMSLP data were performed with XSEDE Bridges at the Pittsburgh Supercomputing Center through allocation TG-IRI190019. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the GPU used for training the models.