

Proposal title: Classifying composers for Western classical music with MusicNet

Team members: Peter Wang, Allen Zhang

Description of the problem and the data:

Western classical music has a rich history and is a marvelous treasure to all humanity and the composers who wrote them were respected to a great extent. Throughout Western classical music's history, there have been many stories of music unearthed whose composers were not known or misattributed, and although historians are usually able to make educated guesses, it is often difficult to truly ascertain the creators of such mysterious works.

We chose to investigate the MusicNet dataset, a collection of carefully labeled classical music. It contains 330 public-domain classical music recordings from some of the most famous Western classical composers, labeled with 1+ million annotations of various measures, such as the instruments being played and the timing of each note in the recordings. The dataset can be accessed via [this link](#) on Kaggle.

Although the MusicNet dataset is arguably quite limited considering the vast amount of classical music and classical composers that exists, it contains extremely detailed information about its 330 pieces to a degree unmatched by other music datasets. Therefore, if classification is effective by training on the MusicNet dataset, it stands to reason that similar detailed approaches like that of MusicNet could shed light on future music research.

Description of what you plan to do:

We aim to develop a machine learning package that could identify the most possible composers for the input Western classical music piece by learning from the MusicNet dataset.

We will use K-NN and decision trees as classification models, which will definitely involve some manual preprocessing to improve the decision trees' algorithm of locating the best feature to split. However, we also plan to try other models to determine the best method of classification.

We are also interested in digging into the style change of composers and determining the time period the composer creates the music piece if possible.

The most basic experiments we plan to run include simply inputting a classical music piece and getting an output of what the model predicts the composer to be. Given the specificity and limited scope of the MusicNet dataset, we will likely have to split it into training and testing data. K-fold and Monte Carlo cross-validation will also be performed to test our model's accuracy.

Further investigations will likely relate to the different styles of the various composers. Western classical music is generally divided into three eras: Baroque, Classical, and Romantic (Impressionist and 20th-century styles follow after, but their distinctions are much harder to define). The most famous classical composers are often divided into these categories—their styles of music were in large part defined by the time they lived in, which offers an interesting

opportunity to test the composer classification model. For instance, Ludwig van Beethoven and Franz Schubert were both late Classical composers—will the model be able to distinguish between these two, or, say, if one was excluded from the training data, will the model be able to identify the other as a close contemporary? Additionally, certain composers like Beethoven are known for having a shift in composing style over their lifetime—will the model be able to capture both sides of a composer’s compositions? Given the variety of music by such prolific composers, there is no shortage of interesting questions that can be asked, especially knowing the history and context of such music.

References:

Google’s music identifying AI Hum to Search:

<https://ai.googleblog.com/2020/11/the-machine-learning-behind-hum-to.html>

Basic Composer Predictor from Kaggle:

<https://www.kaggle.com/code/joshuadubbsnadeau/capstone-composer-predictor-public>

Large-Scale MIDI-Based Composer Classification:

<https://arxiv.org/abs/2010.14805>