Concert Program Note Extraction

Alex Miesbauer

Motivation

Music recommendations often rely on a user's listening preferences. If we could determine the emotional qualities of music automatically, we could generate playlists that guide the user through a range of emotions. Such playlists would be beneficial to fields like music therapy.

Currently, methods to determine emotion in music require large datasets of songs annotated with emotional qualities. A handful of these datasets exist, but they are either small human-annotated datasets or generated using crowd-sourced tags from platforms like `Last.fm`. These datasets only feature recent popular music as well. A dataset for emotion in classical music does not exist. The overall goal of our project is to build a novel dataset of annotations of emotions for classical music.

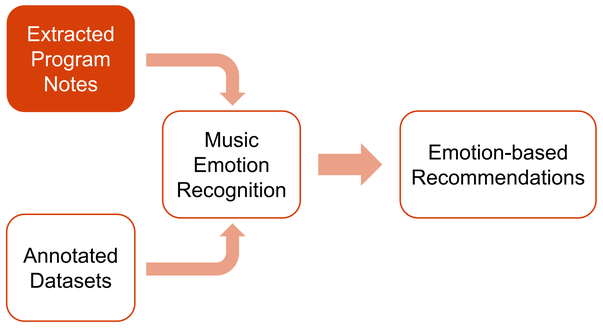


Figure 1: Overview of music emotion recognition.

Data

The New York Philharmonic maintains an extensive digital archive of concert programs dating back to 1842. In total, it contains:

- 14,647 printed concert programs
- Scanned text for over 6 million pages

Extraction Process

These programs often contain large amounts of emotionally descriptive text for classical pieces. However, given the wide variance in formatting and layout, it is necessary to isolate and extract these descriptions. To this end, we developed an automated method using a two-step process.

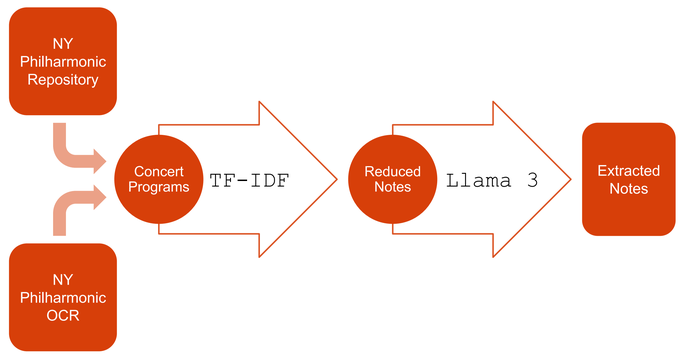


Figure 2: Two-step process for extracting concert program notes.

Step 1: Use term frequency-inverse document frequency (`tf-idf`) to identify and remove irrelevant text – reducing the size of the text.

Step 2: Use language model `Llama 3` to isolate and extract descriptions from the shorter text.

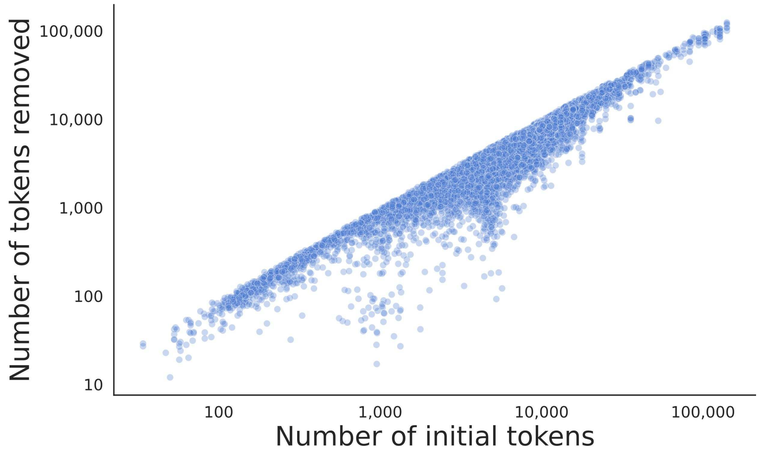


Figure 3: Results of `tf-idf` showing number of tokens removed.

Reduction of Output

- `Llama 3` has a context of only 8,000 tokens.
- Our results confirm the feasibility of our approach in reducing the size of the text input.
- Over the set of concert programs, `tf-idf` only failed to produce results 6% of the time.
- Over a small sample of tests, we confirmed that `Llama 3` identified and extracted a text descriptions related to a given piece of music.

	Words	Tokens	Paragraphs
Before	6,337	8,428	438
After	1,652	2,197	50

Table 1: `tf-idf` resulted in a 73.9% reduction in average words and tokens, and an 88.6% reduction in paragraphs.

Future Steps

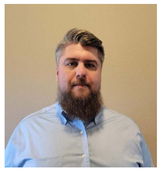
- Identify additional relevant keywords from other sources (e.g., Wikipedia, New Grove).
- Eliminate certain key words to improve focus.
- Determine a comprehensive set of metrics to measure the success of automated extraction.
- Extract descriptions from the entire dataset with goal of reducing the number of programs that need to be processed or verified by hand.

Team



Alex Miesbauer

Project Mentor:
Patrick Donnelly



soundbendor lab