

Lab 2 – MATH 240 – Computational Statistics

Avery Johnson
Colgate University
Department of Mathematics
aqjohnson@colgate.edu

Abstract

This lab aimed to automate the creation of a batch file to process music tracks. The process involved building a batch file that generates commands for each song in the directory, allowing for automated analysis. The task provided practice with installing, loading, and learning to use libraries, working with character objects, coding for loops, and accessing elements of vectors and lists.

Keywords: libraries; character objects; `for()` loops; and vectors/lists.

1 Introduction

In 2018, the Front Bottoms and Manchester Orchestra collaborated on a track called "Allen Town." To explore which band contributed most to the song, a data analysis was conducted on their previous tracks. The goal was to analyze the sound features on each track using Essentia, an open-source tool for music analysis. Given the large number of songs to process, the task was to automate the data collection process using R, enabling a more efficient workflow for handling 181 tracks. In this lab, we will be working with smaller set of .WAV files to complete this same task. This lab aims to build a batch file to facilitate the analysis of each track's audio data and automate the command line process for each song.

2 Methods

The data consists of audio files in the .WAV format stored in a nested directory structure. The first directory level represents artists, and the second represents albums. The analysis involves:

2.1 Task 1: Build a Batch File

Task 1 focuses on creating a batch file that automates the execution of Essentia for each audio track. The key steps in this process include:

1. Extracting subdirectories for each album
2. Filtering and counting .WAV files in each album's subdirectory
3. Constructing commands that execute the Essentia program for each track
4. Writing these commands to a batch file for execution

We used the `stringr` package for R (Wickham, 2023) to manipulate and analyze file paths and the `list.files()` function to retrieve directory contents. The `for()` loop structure facilitated processing each track.

2.2 Task 2: Process JSON Output

Task 2 focuses on processing the JSON output. After the batch file runs and generates the JSON files, the next step is to process these JSON files and analyze the data to determine the musical contributions of each band. In this lab, we focused just on the .JSON output for the song Au Revoir on the Talon of the Hawk album by The Front Bottoms. The steps involved in this task are:

1. Parsing the file names to extract the artist, album, and track information
2. Loading the JSON data into R
3. Extracting key audio features

We used the `stringr` package for R (Wickham, 2023) to handle string splitting in the extracting process. The `jsonlite` package for R (Ooms, 2014) was utilized, specifically the `fromJSON()` function to convert the JSON files into R list objects for easier extraction of the relevant features.

3 Results

3.1 Task 1 Results

The R script successfully identified album subdirectories and filtered .WAV files. It then generated batch commands for each track, which were saved in a text file named `batfile.txt`. These commands generated the Essentia program for each track, saving the corresponding output as JSON files. The process was automated, enabling batch processing for all audio tracks.

3.2 Task 2 Results

Task 2 involved processing the JSON output for a single track. The artist, album, and track name were extracted from the file name, and relevant audio features were successfully extracted. Although only one track was processed, the methods are scalable for multiple files.

4 Discussion

This lab successfully automated the process of generating batch file commands and processing JSON output to extract audio features for analysis. Task 1 demonstrated the efficiency of automating file handling and command generation, while Task 2 allowed for the extraction of key features from JSON files. Although only one track was processed in this lab, the methods are scalable for analyzing larger datasets.

References

- Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv:1403.2805 [stat.CO]*.
Wickham, H. (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1.