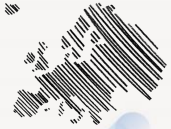
LINKSFOUNDATION.COM

FONDAZIONE links
PASSION FOR INNOVATION

Music Emotion Recognition

Politecnico di Torino
1859

Applied Data Science Project



ellis
European Laboratory for Learning and Intelligent Systems

Turin

October 19, 2021

PROJECT DESCRIPTION

In the context of audio analysis, mood classification is a task that is slowly gaining more and more importance.

It can be done not only by using the information given by the instruments (for example, the timbre or the tone) but also by the lyrics present in the songs.

The purpose of this project is to create a multi-label model capable of classifying moods within a piece of music.

VALUE DRIVEN PROJECT

The analysis of the mood of a song can be used in different ways.

The class can be used as a seed for a generative task (be it audio or image).

Or, we can also use the extracted mood as information for recommendation systems (for example, we can recommend songs with similar moods to Spotify users).

Data - Is there any dataset?

Jamendo is a dataset released in 2019 and available at the following link:
https://github.com/MTG/mtg-jamendo-dataset

Information on Jamendo:

- full songs (at least 30s duration)
  - 3.777 hours overall
- encoding as 320kbps MP3
- annotated by 195 tags
  - 41 instrument
  - 59 mood/theme
  - 95 genre

Data - how is it created and maintained?

The idea on which the creation of this dataset is based relies on the poor quality of the other auto-tagging existing datasets like MTAT (MagnaTagATune), MSD (Million Song Dataset) or FMA (For Music Analysis).
These present some problems such as:

- little variety of tracks and artists
- lack of complete tracks
- poor audio quality
- noisy annotations

The dataset was created using the tracks uploaded on the Jamendo platform. The annotations were then taken and uploaded by the artists who post their tracks on the platform.

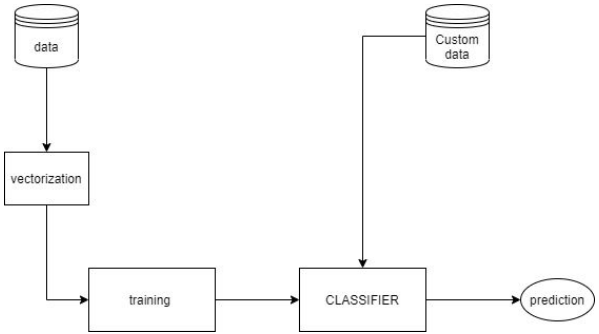
Data - Is there any quality process running?

The Dataset was used for the Emotion and Theme Recognition in Music, in 2019, 2020, and 2021.
https://multimediaeval.github.io/2021-Emotion-and-Theme-Recognition-in-Music-Task/

TASK DESCRIPTION

The task consists in creating a multi-label classifier capable of recognizing moods. To do this, we can divide the task into the following points:

- Analysis of the dataset downloadable at the following link [mettere dimensione].
- Vectorization of the audio tracks using VGGish
- Creation of the multi-label classifier
- Creation of an input block on Colab where you can load an audio file to make inference with, and view the output results of the model



MANAGE THE HIGH SIZE OF THE DATASET

In order to use VGGish it is necessary to use WAV format files. However, the entire conversion of the dataset is very costly in terms of resources, so it is recommended to proceed as follows:

- Take an .mp3 file
- Convert it into .wav using the pydub python library
- Using the .wav file to generate the embeddings
- Save the embeddings

```
from pydub import AudioSegment

# files
src = "test.mp3"
dst = "test.wav"

# convert wav to mp3
sound = AudioSegment.from_mp3(src)
sound.export(dst, format="wav")
```

Mentors

alberto.benincasa@linksfoundation.com

giuseppe.rizzo@linksfoundation.com

FONDAZIONE links

PASSION FOR INNOVATION

Thank you for your attention.

Questions?

FONDAZIONE links

PASSION FOR INNOVATION

CONTACTS

Alberto Benincasa

Researcher

e. alberto.benincasa@linksfoundation.com

FONDAZIONE LINKS
Via Pier Carlo Boggio 61 | 10138 Torino
P. +39 011 22 76 150
LINKSFOUNDATION.COM