Multimedia Systems

Assignment GA2. Low-Level features and timbre characterization (Classification)

FEUP

---

1. Goal

The goal of this assignment is to accomplish a step-by-step audio classification system. While pursuing this goal, we will be able to understand, implement and evaluate a simple set of low-level audio descriptors and analyse their distribution over a collection of sounds.
The chosen sounds are samples of isolated notes from musical instruments, and we will be evaluating the classification over 2 dimensions:

- Instrument excitation: percussive vs non-percussive. These are found by the existence of *pizz* on the filename;
- Instrument type: the name of the instrument (e.g. accordion, flute, cello, etc.). The first letters of the filename give the name of the instrument.

Due date: 26/Nov

2. Resources

Available base implementations:

- (Python) Librosa + Base Code
- MIR.EDU Vamp Plugins for feature extraction (https://github.com/justinsalamon/miredu)

Sound material:

- Samples (isolated notes) from different instruments. ("InstrumentalSounds.zip")

3. Tasks

Part I

Task 1 – Sound Descriptors

Please review the paper by Peeters (Peeters, 2004)*"A large set of audio features for sound description (similarity and classification) in the cuidado project"*, to make sure that you understand the following descriptors:

Time-domain:
Instantaneous
1. RMS/Energy; 2. Zero Crossing Rate
Global
3. Log-attack time; 4. Temporal centroid; 5. Effective duration

Frequency-domain:
Instantaneous
6. Spectral centroid; 7. Spectral spread; 8. Spectral variation / spectral flux; 9. Spectral flatness

Please pick 2 descriptors by group (one from time-domain and another from frequency-domain), depart from the formula and explain the expected values for a sinusoid and white noise (this is a theoretical question).

Task 2 – Exploratory Data Analysis

Implement a function to obtain the following for a given audio file:

- Instantaneous descriptors (1, 2, 6-9)
- Global descriptors (3, 4, 5)
- Statistics for the aforementioned instantaneous descriptors (1, 2, 6-9), such as the mean, standard deviation, minimum, and maximum.

1.1. Create plots to visualize the extracted instantaneous low-level descriptors and examine their evolution across a small set of instrumental samples (for example, percussive, string, and wind instruments).

1.2. Analyze the values of these descriptors for the given instrumental samples and assess how they characterize aspects such as percussive vs. non-percussive sounds, sustained vs. non-sustained tones, low pitch vs. high pitch, and instrument type. To facilitate this, construct 2-D plots that visualize the values of two descriptors for the different samples, for instance:

- Spectral Flux mean vs. Spectral Spread mean
- Spectral Flux mean vs. Spectral Flatness
- Spectral Centroid mean vs. Zero Crossing Rate mean
- Temporal Centroid vs. Log Attack Time.

1.3. Do you think these descriptors are enough to accomplish both classifications or do you feel the need for more descriptors? You may revisit this question after attempting the classification.

Task 3 - Applications

Describe in a short paragraph (max 4/5 lines) a sound-based multimedia application that could make use of this set of sound descriptors.

Part II

Task 4 - Classification - Percussive / Non-percussive

Implement the type of excitation classifiers with simple rules (i.e. without any machine-learning algorithm): percussive vs non-percussive. These are found by the existence of *pizz* on the filename.

This is a binary classification, as there are only two types of excitation: percussive or non-percussive.

Present the classification results in terms of:

- accuracy
- precision
- recall
- F1-Score
- Confusion Matrix

Task 5 - Classification – Instrument Recognition

Try to redo task 4 for the second type of classification:

- Instrument type: the name of the instrument (e.g. accordion, flute, cello, etc.). The first letters of the filename give the name of the instrument.

Is it possible to solve this problem using such sound descriptors and such rules?

Note:
One of the difficulties you will find is that this is a multi-class type of classification. However, algorithms that are designed for binary classification can be adapted for use for multi-class problems.

This involves using a strategy of fitting multiple binary classification models for each class vs. all other classes (called one-vs-rest) or one model for each pair of classes (called one-vs-one).

- One-vs-Rest: Fit one binary classification model for each class vs. all other classes.
- One-vs-One: Fit one binary classification model for each pair of classes.

Further Info: https://machinelearningmastery.com/types-of-classification-inmachine-learning/

Task 6 – Machine Learning

Can you solve tasks 4-6 using Machine Learning? If so, show me and comment your results.

4. Delivery

Deliver your working code in a zip.

Send me a single zip with both your report and code zip by email (named **GA2_GXX.zip –** *example: for Group 1, the file should be named GA2_G01.zip*).

5. References

Bogdanov, D., Wack, N., Emilia, G., Gulati, S., Herrera, P., Mayor, O., Roma, G., & Salamon, J. (2013). Essentia: An Audio Analysis Library for Music Information Retrieval. *ISMIR 2013*, 2–7.

Lartillot, O., & Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction from Audio. *Proc of the 10th International Conference on Digital Audio Effects DAFx07*, 1–8. http://dafx.labri.fr/main/papers/p237.pdf

Peeters, G. (2004). *A large set of Audio features for sound description (similarity and classification) in the CUIDADO project*.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, *130*(5), 2902–2916. https://doi.org/10.1121/1.3642604