Mbay Audio Genre Classification

Magloire LOUDEGUI DJIMDOU

E-mail: djimdouloudegui@yahoo.fr

Github: https://github.com/Djimdou/Side-Projects/tree/master/Mbay%20Audio%20Classification

Introduction

The Mbay live in southern Chad. They are Sara and the largest ethnic group in the Bahr-Sara department (capital Moïssala). The Mbay contributed to spreading the legend of Su, the civilization's founder, among the Sara.

Objective

The aim of this work is to classify Mbay audio files according to their genres. 4 genres are considered: *bang* and *klag* are dancing songs, the *bordero* genre contains religious songs, *terta* audios are stories. The audio files are available at shorturl.at/hzBT5. Figures 1 - 4 display waveplots of sample of each audio genre.

Waveplots

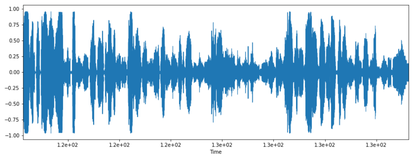

Figure 1: Waves of a *bang* song

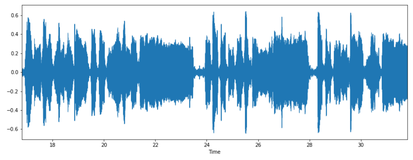

Figure 3: Waves of a *bordero* song

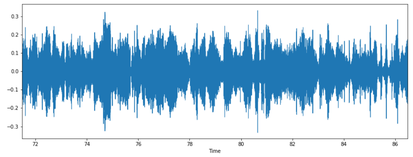

Figure 2: Waves of a *klag* song

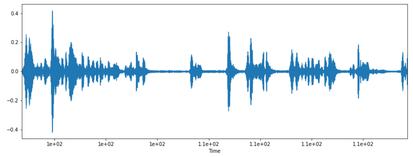

Figure 4: Waves of a *terta* audio file

Spectograms

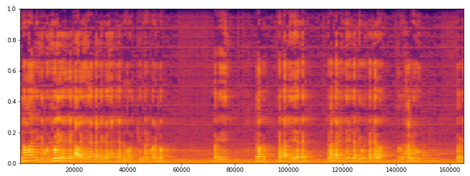

Figure 5: Spectogram of a *terta* audio

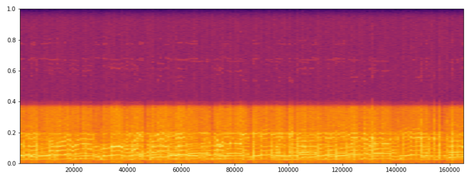

Figure 7: Spectogram of a *klag* song

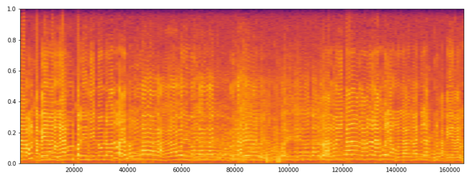

Figure 6: Spectogram of a *bang* song

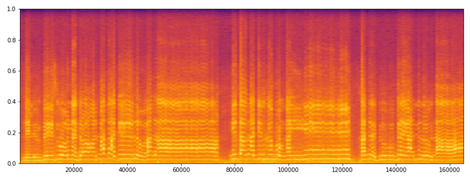

Figure 8: Spectogram of a *bordero* song

Methods

We use the following methods to classify the files:

• Multi-Layer Perceptron Classifier (MLP) • Artificial Neural Networks (ANN) • Multinomial Regression (MNR) • K-Nearest Neighbors (KNN) • Support Vector Machines (SVM) • Random Forests (RF) • Naive Bayes (NB) • Linear Discriminant Analysis (LDA) • Convolutional Neural Networks (CNN)

196 audio files were used for the classification. From each audio file, I sampled 15 seconds, starting from the middle. The samples were divided in two parts: a training set (70% of the files, for building the models) and a test set (30% of the files, for evaluating model performances).

CNN use spectogrammes (Figures 5 - 8) as input. The other methods use features extracted from the spectogrammes (such as mel-frequency cepstral coefficients, spectral centroid, zero crossing rate...).

genre	train	test
bang	46	21
bordero	52	23
klag	18	8
terta	19	9
Total	135	61

Table 1: Data for CNN

genre	train	test
bang	45	22
bordero	52	23
klag	19	7
terta	21	7
Total	137	59

Table 2: Data for other methods

Results

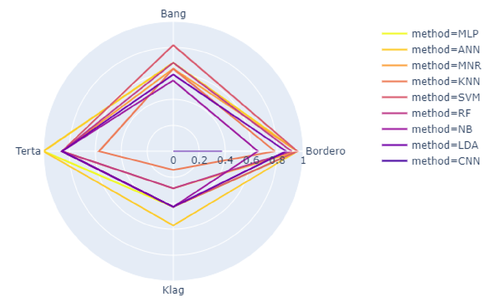

Figure 9: Precision

Bordero songs are the easiest to classify: all methods recognize them correctly in more than 60% of the cases (Figure 9).

Klag audio files are the hardest to correctly identify by the classification methods (Figure 9). This is probably due to the variability within the genre.

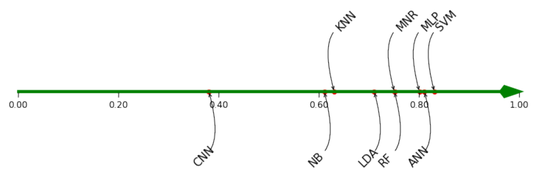

Figure 10: Accuracy

For the overall performance, CNN have the worse performance: this method correctly classifies only 38% (Figure 10) of the test files, whereas the other techniques have an accuracy of more than 60%. Maybe because the sample size (196 files) is too small for the CNN.

SVM have the highest accuracy: they correctly classify 83% of the test files (Figure 10).