DNN in the AcousticBrainz Genre Task 2017

Nicolas Dauban

IRIT, Université de Toulouse, CNRS, Toulouse, France

nicolas.dauban@irit.fr

ABSTRACT

This working note describes the participation of the SAM-IRIT team for the Mediaeval AcousticBrainz Genre Task 2017.

1 INTRODUCTION

The AcousticBrainz Genre Task 2017 is a music genre recognition (MGR) task organised by Mediaeval [1] where participants had to make predictions on genre and subgenres, based on audio features extracted from Essentia [2]. The target labels are provided by 4 different sources : Discogs, Allmusic, Tagtraum and Lastfm.

2 RELATED WORK

I participated to this challenge during my internship on content-based music recommendation. During this internship, I worked on music genre recognition using music features with deep neural network [3]. This approach has been tested on the GTZAN [4] corpus then on the MagnaTagATune [5] corpus. The results obtained were about 90% on the GTZAN corpus and about 80% with MagnaTagA-Tune. As these results were satisfying, we decided to take part of the challenge to try the neural network approach.

3 APPROACH

As the time evolution of the music features was not supplied, the convolutional approach - which needs correlations between successive frames of the input image - has been quickly aborted. Thus, we used a classic deep neural network instead of a convolutional one. The neural network has been implemented using the Theano framework « Lasagne » [1]. We tried to use different features and different architectures of neural networks.

3.1 Features

During the development phase, the following features have been tried : Mel bands, spectral rolloff, zero crossing rate, spectral entropy, Harmonic Pitch Class Profile (HPCP), tempo, danceability, key strength, dissonance, tuning diatonic strength. The choice of those features has been made by relying on their definitions and by choosing those which seem to be the more relevant to characterize the different music genres. In our preliminary experiments, Mel bands yielded the best results, thus, only Mel bands were used as input for our final submission.

---

1. https://lasagne.readthedocs.io/en/latest/

3.2 Architecture of the network

To establish the final topology of the neural network, different architectures have been tried. The final architecture comprises 8 full connected layers of 2000 neurons with ReLUs as activation function (Figure 1). The input vector had size of all the data used (for the 9 statistics on the 40 MEL bands, the input was a vector of size 360), and an output with the total number of genre and sub-genres (e.g. 315 for the Discogs subset) with a sigmoid activation function. With the sigmoid output function, it was common that the network did not predict any genre for some tracks, because not a single output neuron had a value superior to 0.5 for those tracks. In order to have at least one genre prediction per track if no genre was predicted, the genre with the maximum output value is chosen as the predicted one. For the submission, the network was trained with 40 epochs.
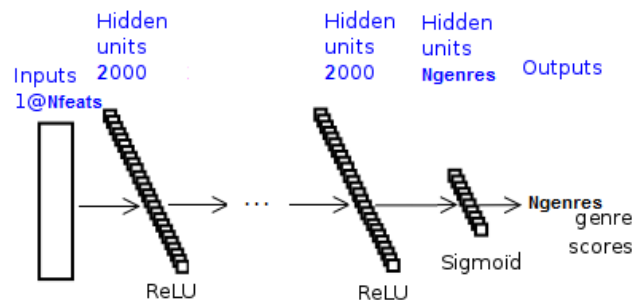

FIGURE 1: DNN architecture

4 RESULTS AND ANALYSIS

During the development phase, the network obtained a 60% F-Score on genre-only on Discogs (trained with 80% of the data and tested on the 20% remaining). However, the best score obtained on genre-only of our final submission was 35% maybe due to a lack of generalization power of our network.

5 DISCUSSION

We plan to explore our initial idea to use combinations of different acoustic feature types as we previously showed that these were as successful as using Mel bands. Furthermore, some modifications on the network can be done : adding batch-normalization after each dense layer, residual blocks, and also try some different initialization and decay strategies for the learning rate.

RÉFÉRENCES

[1] Dmitry Bogdanov, Alastair Porter, Julián Urbano, and Hendrik Schreiber. The mediaeval 2017 acousticbrainz genre task : Content-based music genre recognition from multiple sources. 2017.

[2] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, Oscar Mayor, Gerard Roma, Justin Salamon, José R Zapata, Xavier Serra, et al. Essentia : An audio analysis library for music information retrieval. In *ISMIR*, pages 493–498, 2013.

[3] Christine Senac, Thomas Pellegrini, Florian Mouret, and Julien Pinquier. Music feature maps with convolutional neural networks for music genre classification. In *Proceedings of ACM Workshop on Content Based Multimedia Indexing (CBMI)*. ACM, June 19-21, 2017.

[4] George Tzanetakis and P Cook. Gtzan genre collection. *Music Analysis, Retrieval and Synthesis for Audio Signals*, 2002.

[5] Edith Law, Kris West, Michael I Mandel, Mert Bay, and J Stephen Downie. Evaluation of algorithms using games : The case of music tagging. In *ISMIR*, pages 387–392, 2009.