DNN in the AcousticBrainz Genre Task 2017

Nicolas Dauban
IRIT, Université de Toulouse, CNRS, Toulouse, France
nicolas.dauban@irit.fr

ABSTRACT

This paper presents a method of genre classification using deep neural networks for the AcousticBrainz genre classification task of MediaEval 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 in 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 and then on the MagnaTagATune [5] corpus. The accuracy results obtained were about 90% on the GTZAN corpus and about 80% with MagnaTagATune. 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.

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.

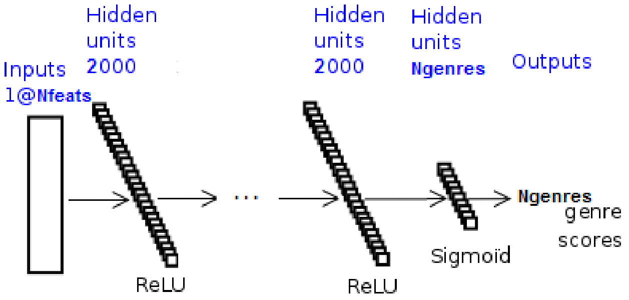

Figure 1: DNN architecture

4 RESULTS AND ANALYSIS

Only one submission was made for each dataset, for the task 1 only.

The Table 1 and Table 2 contain the results per tracks for all labels and for genre labels:

Table 1: Results per track, for all labels

	Precision	Recall	F-score
Allmusic	0.3751	0.1455	0.1813
Discogs	0.3043	0.1243	0.1664
LastFM	0.1845	0.0813	0.1073
Tagtraum	0.2892	0.1214	0.1649

The results for sub-genres and per label are not shown here because they were all between 0% and 10%. All the results are provided on the AcousticBrainz Genre Task results page [2].

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

[2] https://multimediaeval.github.io/2017-AcousticBrainz-Genre-Task/results/

MediaEval'17, 13-15 September 2017, Dublin, Ireland

Table 2: Results per track, for all labels

	Precision	Recall	F-score
Allmusic	0.3962	0.3334	0.3513
Discogs	0.3221	0.2674	0.2812
LastFM	0.1937	0.1764	0.1813
Tagtraum	0.3231	0.3055	0.311

During the development phase, the network obtained a 60% F-Score on genre-only on Discogs. During this phase, the data has been split using the python script provided by organizers to ensure album filtering: 80% to train the network and 20% to test it. However, the best F-score obtained on genre labels of our final submission was 35% per track 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, try some different initialization and decay strategies for the learning rate, and also lower the threshold of the output sigmoid in order to give more predictions.

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. In *Working Notes Proceedings of the MediaEval Workshop*, Dublin, Ireland, September 13-15, 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.