Artificial Intelligence and Machine Learning project proposal

Matteo Cerutti (s265476), Antonio Santoro (s264014), Marco Testa (s265861)

Computer Engineering - Politecnico di Torino

1 The problem

Nowadays people need to have the possibility to select music and make playlists based on their mood. Many music platforms sponsor different music playlists made by hand that include the most popular songs only with the intention to maximise ratings. One of the most used feature on these platforms is to make playlists similar to other personal ones, the point is that all the songs included into the new created playlists are selected based on the "similarity". Our proposal focuses on making classifications of music from a different point of view, in fact our aim is to exploit a visual representation of an audio file and try to catch features on that. Our interest is to train a neural network on different audio speeches files that represents different human emotions. The model will then be applied on a different domain to see whether the learned peculiarities can be matched on the music or not.

2 Data and readings

For this purpose we took inspiration from these readings:

- http://deepsound.io/music_genre_recognition.html
- http://cs231n.stanford.edu/reports/2017/pdfs/22.pdf

All datasets we are going to use have to be modified in order to get the spectrogram from the audio files. We are going to train our network using this pure vocal melody as sample, with "emotions" as labels:

- https://www.kaggle.com/uwrfkaggler/ravdess-emotional-song-audio

So, this dataset will be transformed in order to obtain a (spectrogram - emotion) pair and this will be our starting point for the training phase. At test time, we will evaluate our model on other datasets such as Cal-500 and Cal-500exp after transforming these two as well.

3 Methods and algorithms

The algorithms we intend to use are three different neural networks: GoogLeNet, ResNet and VGG. All of them have a convolutional part composed by different layers (22 for GoogLeNet, 147 for ResNet, between 16 and 19 for VGG), useful for the features extraction from spectrogram's images. ResNet and VGG have also a small fully connected part used for the classification objective. All the neural networks will be modified in order to have as output the number of classes we will consider.

4 Results evaluation

All the results will be qualitatively evaluated by means of plots that measure the accuracy on the test set, moreover we are going to perform a statistical test by matching the opinions of a sample of people and the outputs of our model.