MUSICAL GENRE CLASSIFICATION USING A MFCC-FED ARTIFICIAL NEURAL NETWORK

Román Via-Dufresne Saus, II "Bachillerato" at La Farga, rvds910@gmail.com
Mentor: Barbara Hajdarević, Faculty of Electrical Engineering, University of Belgrade, barbi.hajdarevic@gmail.com

INTRODUCTION

Music genre classification using machine or deep learning is present in many popular apps and platforms such as Spotify and Pandora. There are different methods and algorithms that can be used to achieve so. In this paper, an **artificial neural network** (ANN) has been used.

The aim of this project has been to develop a music genre classifying algorithm and to optimize it to obtain the most accurate classification possible. To do so, **MFCC** features have been extracted from 30s audio files and fed into a neural network. Several feature extraction methods and artificial neural network architectures have been evaluated to achieve optimal results.

METHODS

ANN

Artificial neural networks are machine learning algorithms that mimic human brain **neurons** architecture. Each neuron and each connection between neurons are represented with a numerical value. Applying each value to the input data results in the output data. Neural networks can be designed by changing the number of **layers** of neurons and the number of neurons in each layer. Once a neural network is built, it is **trained** by changing the value of each neuron and connection to make the output as accurate as possible. They have been built and trained using **MATLAB**'s patternet.

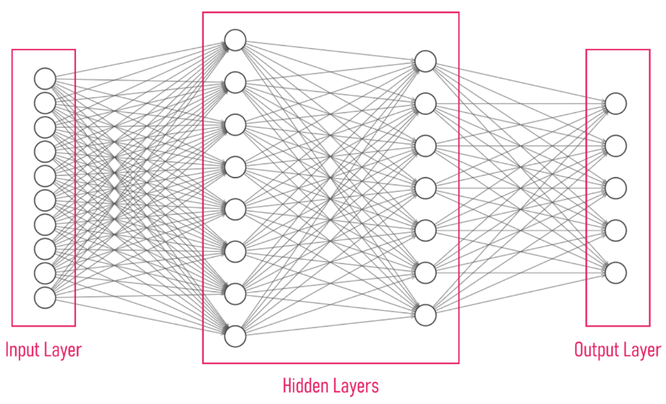

Figure 1: Artificial Neural Network

MFCC

Mel Frequency Cepstral Coefficients have been used as features to input into the artificial neural network. These are based on **frequency** properties of the signal. However, they differentiate from standard frequency spectral coefficients because they are obtained with a method that relies on the Mel scale. The Mel scale is a scale that takes into account **human perception** of frequency changes. They have been calculated using **MATLAB**.

DATABASE

The GTZAN database has been used for this project. **GTZAN** is a free database that contains 1000 songs of 10 different genres (**100 songs per genre**). GTZAN songs are 30 seconds long ".au" files with a 22050 Hz sample rate. However, only **5 genres** have been used in this project: jazz, classical, pop, rock and reggae.

METHOD OPTIMISATION

In order to achieve as high accuracy of classification as possible, the following has been varied:

- Representing input data as an array of MFCC or as a mean and covariance vector of these MFCC.
- Using 13 MFCC or just 5 MFCC.
- Splitting the samples into smaller clips - 5s or 3.5s - to have more data samples or to use the full 30s samples.
- Using a Mel filterbank - to calculate the MFCC - that extend from 300 Hz to 8000 Hz or from 20 Hz to 8000 Hz.
- Varying the number of layers of the artificial neural network.
- Varying the number of neurons in each layer of the artificial neural network.

RESULTS

OPTIMAL CONFIGURATION

Mean & Covariance	**13 MFCC**	**Full 30s Samples**
20 Hz - 8000 Hz	**1 Hidden Layer ANN**	**10 Neurons per Hidden Layer**

CLASSIFICATION ACCURACY

Using the above configuration, the artificial neural network has been trained and tested several times. A mean **accuracy of 87.6%** ± 3.7% has been obtained. The figure below shows one confusion matrix obtained:

PREDICTED GENRE \ ACTUAL GENRE	Reggae	Pop	Jazz	Rock	Classical	
Reggae	20	1	0	1	0	90.9%
Pop	1	8	0	1	0	80.0%
Jazz	0	0	15	0	0	100.0%
Rock	1	2	0	14	2	73.7%
Classical	0	0	0	0	9	100.0%
	90.9%	72.7%	100.0%	87.5%	81.8%	88.0%

Figure 2: Confusion matrix

CONCLUSION

The best way to obtain the MFCC is to use the full 30s clips of the database and to extract the first 13 MFCC with a filterbank that extends from 20 Hz to 8000 Hz. To feed the artificial neural network, the mean and covariance method performs well. The architecture for the MATLAB's patternet artificial neural network that worked the best consists in a single hidden layer of 10 neurons.

By using the configuration mentioned above, an accuracy of 87.6% ± 3.7% has been achieved. This represents a good performance that can be classified as **over average** compared to some similar papers (see Literature section).

LITERATURE

- McKinney, M., & Breebaart, J. (2003). Features for audio and music classification.
- Haggblade, M., Hong, Y., & Kao, K. (2011). Music genre classification. Department of Computer Science, Stanford University, 131, 132.
- Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on speech and audio processing, 10(5), 293-302.
- Feng, T. (2014). Deep learning for music genre classification.