**SYSC4415: Introduction to Machine Learning**

**Assignment 3**: A music genre classifier problem

**Group 6**:

Keertikumar Kubareea, 101167431

Aryan Laxman Sirohi, 101199497

**Submission Date**:

March 29th, 2024

**INTRODUCTION**

The title of today’s exploration is a ‘Musical classifier’. This proposal outlines an initial exploration of the GTZAN dataset [1] and a plan to develop a potentially robust musical genre classifier using transfer learning techniques. This proposal touches on the following areas: the dataset, model architecture, references for the research done and some dataset exploration in the appendix. We wish you a good read.

**THE DATASET**

The dataset that will be used is the GTZAN dataset [1], loaded onto Google Drive (PATH=’drive/genres\_original’). This dataset comprises audio files across multiple genres (rock, blues, classical, country, disco, hip-hop, jazz, metal, pop, and reggae). Through our initial exploration of the dataset in a Jupyter Notebook, we have collectively concluded that the dataset is balanced [1, Appendix]. Each genre folder inside the dataset contains 100 samples of WAV sound signals.

**MODEL ARCHITECTURE**

* The data has been preprocessed using SciPy for feature extraction. This preprocessing is temporary for part 1 to only create a data frame to ease the visualization of the data and check whether the sample counts are balanced from each genre.
* Future preprocessing will meet the input requirements of YAMNet from the TensorFlow framework. The baseline YAMNet model will be used to extract embeddings from the audio waveforms.
* The panda data frame has already been created in this initial exploration (please view the uploaded Jupyter notebook). This data frame will be partitioned into training, validation, and test sets accordingly. We will be doing a 70-15-15 split. 70 samples will used for training, 15 for validation and 15 in the test set.
* To perform transfer learning, based on research [2], we found that the YAMNet model is pre-trained on a large audio dataset. This will be used as a feature extractor. The embeddings will be used as input features for a shallow classifier.
* A simple classifier model will be trained and fine-tuned on top of the YAMNet baseline. The training will be done on the embeddings for genre classification.
* Hyperparameters like learning rate and batch size will be defined and optimized using techniques like grid search or random search on the validation set.
* A categorical cross-entropy loss function (or another) may be defined.
* Regarding evaluation, model performance will be assessed using metrics such as accuracy, precision, F1-Score, and recall. Tuning may be carried out to optimize performance further.

As compared to a pre-trained YAMNet model, a fine-tuned classifier will be expected to have enhanced performance tailored to the musical genre classification.

**REFERENCE**

[1] GTZAN dataset. Found Online on Kaggle. Online: <https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification/code>

[2] TensorFlow: Transfer learning with YAMNet for sound classification. Online: <https://www.tensorflow.org/tutorials/audio/transfer_learning_audio>

**APPENDIX**

1. Chart from the Jupyter Notebook (using matplot library) showing that the sample count is **balanced** across the following genres: rock, blues, classical, country, disco, hip-hop, jazz, metal, pop, and reggae.

A chart of different colors

Description automatically generated