MIE324 Project Proposal

Music Genre Classification

Robert Adragna

Yuan Hong (Bill) Sun

Word count: 1140

1. Introduction

A music genre classifier is a software program that predicts the genre of a piece of music in audio format. These devices are useful for a variety of tasks such as automatically tagging music for distributors such as Spotify, Billboard, or other music services, and automatically determining appropriate background music for events.

Currently, genre classification is performed manually by individuals through their personal knowledge and understanding of music genres. This is because it is tedious to perform this task through conventional algorithmic approaches since the distinctions between music genres are relatively subjective and ill-defined. However, this lack of distinction makes machine intelligence well-suited for genre classification. Given enough data, of which plenty of songs exist, machine learning can recognize these ill-defined patterns in music, and make predictions based on these observations.

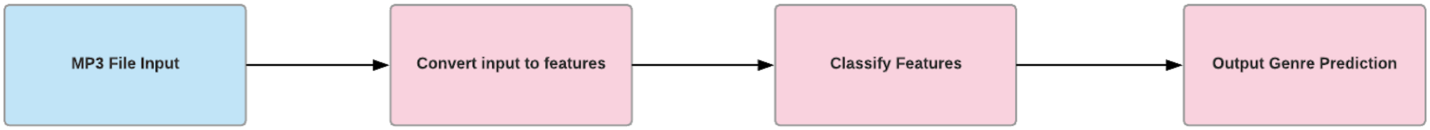
The goal of this project is to build a music genre classifier using a deep learning approach that can correctly predict the genre of any piece of Western music.

2. Source of Data

The main dataset used for this project is the Million Song Dataset (MSD), which is commonly used for studying musical properties through data science. Every record of the dataset includes the full audio recording of the song along with its relevant metadata such as title, artist, year, and genre.

We will extensively process this data before it is used as input to a neural network based model. First, we will strip the dataset of all metadata such that for each song only the genre and audio data remain. Then, to make the project's scope more manageable, we will remove and amalgamate genres such that only several categories remain from the 25 possible genres in the MSD. Next, since processing the full audio of each song (usually several minutes in length) can be too computationally expensive, we will automatically select a short, representative sample of each song that will be used for training instead. Thirdly, we will encode the audio data into a format that is feasible for use as input into a neural network. Inspired by the successful work of [1], we plan to encode the data as a time series of the song's Fourier decomposition. If this is not sufficient, we plan to extract a time series of the song's relevant aural features as successfully attempted in [2].

3. Overall Structure of Software



- Converting input to features:
    - Filters out useless data and condenses music genres (labels) from 25 to less than 10 (exact number will be determined)
    - Takes a 10-second sample from each audio file at a specific location
    - Converts raw audio files into time series matrix of Fourier decompositions or other features for input into the neural network
    - Splits data into training and validation sets

- Classifying Features:
    - Main neural network infrastructure
    - Training, validation, and evaluation loops
    - Generates a model / classifier

- Output Genre Prediction:
    - Evaluates the model on test data
    - Outputs the predicted music genres of the test data

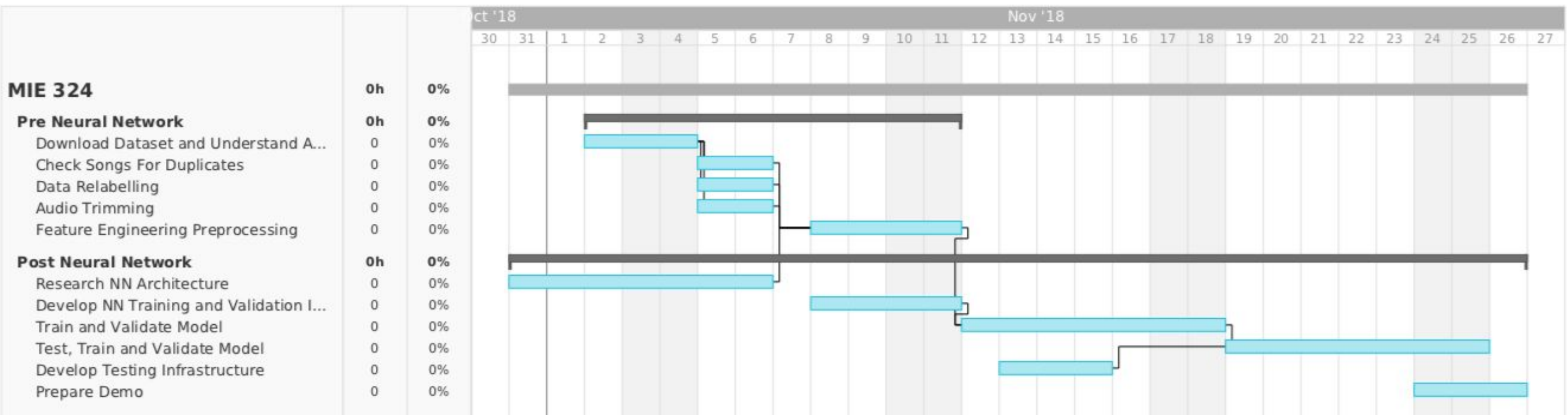
4. Plan

List of sub-tasks:

- Download dataset from the internet and understand its API.
- Take the internet dataset and convert into a form that we can use for NN.
    - “Standard processing”: Filtering unwanted data from the raw data
    - “Advanced processing”: Convert raw data into features or a fourier decomposition time series that can be used as NN input.
- Design neural network architecture
    - Research best practices for music classification
- Build neural network and training, evaluation and testing infrastructure
- Train the neural network to achieve certain ’threshold’ validation accuracy
    - Along the way, iteratively adjust architecture and tune hyperparameters.

- Once ‘threshold’ validation accuracy achieved, evaluate on test set to test performance.
    - Return to adjusting model if poor test results.
- Prepare demo

For greater detail, see our project Gantt Chart:



5. Risks

1. Our dataset might not contain examples that are fully representative of all western music. For example, the MSD only contains songs that are instrumental or sung in English. This might cause the model to perform poorly when classifying other songs sung in different languages. We will address this, depending on time constraints, by either changing the scope of our project to classifying only the types of music which are present in the MSD or by training on additional datasets with more diverse music.

2. While it should be straightforward to find the fourier decomposition of songs, it might be difficult to decide which musical features should be used as inputs to the neural network. Since we are not experts in music theory, we will likely not have enough time to fully learn about all possible characteristics of music and deduce for ourselves which ones might be best for genre classification. To mitigate this risk, we can consult past feature-based approaches to music classification and use features that worked well in these scenarios. It will also be helpful to find 3rd party libraries for feature extraction that abstract away the need for detailed knowledge of music theory.

3. It is possible that the model performs well on data from our validation set, but poorly on data from our test set. In this case, we will have to retrain the model with hyperparameter and architecture adjustments until both validation and test accuracy can be improved.

6. Things to Learn

- **Sound Processing**:
  - We must learn how to convert audio mp3 files into a discrete time series of its fourier components.
  - Should the fourier analysis approach be insufficient, we will perform feature engineering on the input audio data. We must learn more about both the musical theory underlying the features of music we may wish to use, such as tonality and chromatics, along with how these features are typically represented in computers.
- **Databases** - The Million Song Dataset is stored online in a SQL relational database. We must learn how to extract this information so that it can be easily inputted into our neural network.
- **Best Practices -** Research is needed to discover what neural network architectures and techniques work best for genre classification.

7. Ethical Issues

A possible ethical issue resulting from automatic genre classification is that these systems may be used to regulate or censor different types of music. Music from various genres (ie: hip-hop, jazz, punk rock) has been used throughout history as a medium for publicly sharing socio-political-cultural messages. These messages may go against the interests of those in power. With our genre classification system, music that is likely to contain offending messages can be easily identified and this information can be used to suppress or encourage its popularity and exposure according to the interests of those in power. This threatens the democratic right to freedom of expression.

We are also aware of the potential of our automatic classifier to replace jobs in the music industry which involve music classification (for example, playlist curators at Spotify). However, given the small number of workers in the music industry [3], we do not believe that this technology will meaningfully change employment patterns.

References

[1] Hamel, P. and Eck, D. (2018). LEARNING FEATURES FROM MUSIC AUDIO WITH DEEP BELIEF NETWORKS. [online] DIRO, Universite de Montreal CIRMMT. Available at: http://ismir2010.ismir.net/proceedings/ismir2010-58.pdf [Accessed 29 Oct. 2018].

[2] Dieleman, S., Brakel, P. and Schrauwen, B. (2018). AUDIO-BASED MUSIC CLASSIFICATION WITH A PRETRAINED CONVOLUTIONAL NETWORK. [online] Electronics and Information Systems department, Ghent University. Available at: https://biblio.ugent.be/publication/1989534/file/6741798.pdf [Accessed 29 Oct. 2018].

[3] Statista. (2018). Number of employees in the U.S. music industry 2015 | Statistic. [online] Available at: https://www.statista.com/statistics/184577/number-of-employees-in-the-us-music-industry-by-sector/ [Accessed 29 Oct. 2018].