Music Genre Classification

Areeb Khan Shabih A20469525 and Carmen Acero Vivas A20472656

ashabih@hawk.iit.edu, cacerovivas@hawk.iit.edu

1 Problem Description

With the exponential growth of online music streaming services, it has become a challenging task both for users and the service providers to maintain music library manually. Accurate and automated music classification to different genres using machine learning algorithms can not only solve this problem for music streaming companies like Spotify and iTunes and improve user experience as well.

Musical Information Retrieval (MIR) is the science of retrieving information from music. The information can be used for different research topics like genre classification, recommender systems, instrumental recognition or sentiment analysis. In this project we will focus on music genre classification, a field where it can be found rich and varied literature. Different approach has been proposed to solved this task, going from classic algorithms to deep learning models.

Nearly 40,000 tracks get added to Spotify along everyday. A robust classifier is essential to characterize the music and tag unlabelled music. At the same time, lack of standardization and ambiguous boundaries between music genres make this a very challenging task. Moreover, human perception change depending on their experiences and opinion and this idea stands out especially when talking about art, particularly in music. This leads to an absence of a global data-set that can be used to obtain reproducible results. Each data-set has its unique structure, presenting different number of genres and descriptors.

Through our project, we will dive and explore the shallow learning and deep learning approaches to perform classification. We will understand, implement and leverage the work already done for feature extraction and lay majority of the focus towards the actual classification problem. Starting from diverse shallow learning algorithms like Naive Bayes, Logistic Regression, SVC, Random Forest, Decision Trees, Passive Aggressive Classifier , Perceptron and KNN , we'll go on to explore the deep learning techniques like CNN and RNN for music classification. The implementation and design of these models will be driven by the type of data available, classification goal and our understanding of the problem. Towards the end of our project we'll do a comparative analysis and see which algorithm performs the best job.

2 Background

The study on this field started in 2002 with Tzanetakis and Cook [1], where the authors extracted 30 features from the audio files to predict 10 different music genres. They also build the GTZAN dataset [2], which has become the most well known dataset. However, this dataset present inconsistencies and is relatively small, with only 1000 songs. Other available and public datasets such as FMA dataset [3] counts with 0.1 million songs, which enables more complex models, that requires a greater amount of data, to be trained. We can also find MillionSong Dataset (MSD) [4] or AudioSet [5]. Historically, this challenge started getting addressed by employing classical methods. The most commonly used classical methods included decision trees, an appealing approach as this is a classification task and trees have proved to show good results [1], probabilistic classifiers, such as Naïve Bayes Classifier [1] and Support Vector Machine (SVM) [1]. These models relied basically on three categories of features; rhythm, pitch and temporal structure. However, recently using audio spectrogram has become popular for music genre classification. Spectogram encodes time and frequency information of the music track in its entirety [2]. While it is true that the classical algorithms have shown good results, over the course of time, deep learning techniques have demonstrated outstanding results, close to human level of accuracy. Within this field, there exist literature using Fully Connected Neural Networks which potential relies when the audio file is introduced as input with some extra information so the problem becomes richer and more suitable for this type of deep models. Convolutional Neural Networks [2] using spectrogram results as an input have achieved great results. Since the beat, rhythm, frequencies and tone of the music changes with time, Recurrent Neural Networks, have been recently used to model temporal information and sequences. The most popular architecture that has been used in this field is Long-short term memory (LSTM) [6].

3 Proposed work

In this project we want to compare different architectures, classical algorithms and deep learning techniques like CNN and RNN, with a rich data-set and compare them. There will be many challenges along the way. The first and foremost challenge will be the dimensionality of the music signal in time domain which can make the classification task computationally expensive. Beginning with data pre-processing, we will discover innovative ways to utilize the temporal spectogram features while addressing the dimensionality problem without compromising the model quality. Selecting an optimal time window to achieve dimensionality reduction and sample the song data will also be a challenge since want to predict the genre on a real-time basis and discover how the model's prediction changes as the song progresses. This will allows us to study which parts of the songs are more representative. Some studies argue that the central part of the song is the one that defines the genre [5]. We will use the standard set of raw and derived features with the shallow learning techniques like,

- Naive Bayes
- Logistic Regression
- SVM
- Perceptron
- Passive Aggressive Classifier
- Decision trees and Random Forest

The analysis will eventually expand to deep learning and we will look to implement following algorithms,

- CNN
- RNN
- C-RNN

We expect to get much better results from deep learning techniques. Specially since RNN and C-RNN [6] have the capability to master time series prediction and sequence classification so we will be looking forward to implement them and observe our results.

Towards the end of our analysis we will compare the performance of all the algorithms using different standard metrics and identify the optimal one. We will also be reflecting on few data insights and highlight the challenges which still remain after using the best classification algorithm.

4 Milestones

The following table shows the initial plan, milestones and timelines.

Table 1. Milestones

Date	**Task**
01-March to 08-March	Choose dataset
08-March to 15-March	Model development
15-March to 22-March	Train and evaluate
22-March to 05-April	Intermediate project, model improvement
05-April to 12-April	Training and testing
12-April to 19-April	Report writing, presentation preparation, final model details
19-April to 26-April	Final presentation
07-May	Final project due date

References

[1] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10:293 – 302, 08 2002. doi:10.1109/TSA.2002.800560.

[2] Bob L. Sturm. The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use. *CoRR*, abs/1306.1461, 2013. URL `http://arxiv.org/abs/1306.1461`.

[3] Michaël Defferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. Fma: A dataset for music analysis, 2017.

[4] Thierry Bertin-Mahieux, Daniel Ellis, Brian Whitman, and Paul Lamere. The million song dataset. pages 591–596, 01 2011.

[5] J. R. Castillo and M. J. Flores. Web-based music genre classification for timeline song visualization and analysis. *IEEE Access*, 9:18801–18816, 2021. doi:10.1109/ACCESS.2021.3053864.

[6] Chi Zhang, Yue Zhang, and Chen Chen. Songnet: Real-time music classification. *Standford*, 2018.