Music Genre Classification

A Project Work Synopsis

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE WITH SPECIALIZATION IN

BIG DATA ANALYTICS

Submitted by:

20BCS3923 Anup

Under the Supervision of:

Mr. Pulkit Dwivedi (Assistant Professor)

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,

PUNJAB

Abstract

Keywords: Music Genre Classification Model is a model to classify songs or an audio music based on variety of features of it into the corresponding genre. As the lifestyle of the people in the world is more depending on the music, the technology and the internet becoming cheaper to the end users, there is a much need of developing an efficient and more accurate model for this classification. So this project aims at developing a model that can classify and organise the audio music to its belonged genre out of many genre. Machine Learning techniques are employed to do this task. The Classification is performed based on acoustic features of the audio music and the spectrogram of the audio. This model is designed and developed to produce a better accuracy for Music Genre Classification.

Table of Contents

1. INTRODUCTION

1.1 Problem Definition

Audio processing is one of the most complex tasks in data science as compared to image processing and other classification techniques. One such application is music genre classification which aims to classify the audio files in certain categories of sound to which they belong. The application is very important and requires automation to reduce the manual error and time because if we have to classify the music manually then one has to listen out each file for the complete duration. So To automate the process we use Machine learning and deep learning algorithms and this is what we will implement in this article.

1.2 Problem Overview

In short, we can define our project problem statement as like given multiple audio files, and the task is to categorize each audio file in a certain category like audio belongs to Disco, hip-hop, etc. The music genre classification can be built using different approaches in which the top 4 approaches that are mostly used are listed below.

Multiclass support vector machine

K-Nearest Neighbors

K-means clustering algorithm

Support Vector Machine(SVM)

We will use K-Nearest Neighbors algorithm because various researches prove it is one of the best algorithms to give good performance and till time along with optimized models organizations uses this algorithm in recommendation systems as support.

1.3 Hardware Specification

The minimum hardware requirements for trying out our deep learning will vary depending on the specific tasks you want to perform. However, our project purpose system with the following specifications should be sufficient for most basic deep-learning tasks:

- CPU: Quad-core Intel Core i5 or higher
- RAM: 8GB
- GPU: NVIDIA GeForce GTX 1060 or higher
- Storage: 256GB SSD

1.4 Software Specification

1. Operating System: Most machine learning frameworks are compatible with popular operating systems such as Windows, macOS, and Linux.

2. Python: Python is the most commonly used programming language for machine learning, and most machine learning frameworks are built using Python.

3. Machine Learning Frameworks: There are many machine learning frameworks available, such as TensorFlow, PyTorch, Keras, and scikit-learn. These frameworks provide tools and APIs for building, training, and deploying machine learning models.

4. Development Tools: IDEs such as PyCharm, Jupyter Notebook, and Visual Studio Code are commonly used for machine learning development.

5. Libraries: Libraries such as NumPy, Pandas, and Matplotlib are commonly used for data manipulation, analysis, and visualization.

2. LITERATURE SURVEY

2.1 Existing System

Generalized audio classification has more far-reaching applications such as speech detection, separating background noise from primary speech, sentiment analysis, speaker identification for voice verification. I will look at one such approach of building a music genre classifier model using an open-source dataset GTZAN.

About GTZAN Music Genre Dataset

This GTZAN Music Genre Dataset contains 1,000 song samples, each 30 seconds long, belonging to a total of 10 conventional music genres. The samples are classified into blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. The dataset also contains an alternate representation as images of Mel Spectrograms.

Thus, with the GTZAN dataset, we can perform music genre classification using CNN or other computer vision models or use time-series data like MFCCs in training an LSTM or RNN.

2.2 Proposed System

Here I will be using Mel-Frequency Cepstral Coefficients(MFCC) from the audio samples. A deciding factor in what kind of predictions our model gives methods like KNN, Random Forest,

SVM. The MFCC summarises the frequency distribution across the window size, so it is possible to analyse both the frequency and time characteristics of the sound. These audio representations will allow us to identify features for classification.

Testing Some Test Audio Data

Steps

- Preprocess the new audio data
- predict the classes
- Invers transform your Predicted Label

2.3 Literature Review Summary

Year& Citation	Article/ Author	Tools/ Software	Technique	Source	Evaluation Parameter
2020	Manish Agrawal and Abhilash Nandy	VS, G-Collab, Tenser-Flow, Pandas	CNN,HAN	[Bahdanau et al.(2014)Bahdanau, Cho, and Bengio]	Logloss and test accuracies
2021	Safaa Allamy and Alessandro Lameiras Koerich	Jupyter NoteBook, Numpy, Scikit-learn	1D CNN	A. Krizhevsky, I. Sutskever, and G. E. Hinton	Short-term Fourier Transform (STFT) spectrogram
2021	Peace Busola Falola and Solomon Olalekan Akinola	TenserFlow, Pandas, VS Code	1D CNN	Weibin Zhang, Wenkang Lei, Xiangmin Xu, Xiaofeng Xing (2016)	Genre classification results on GTZAN dataset
2017	Lin Feng, Shenlan Liu, and Jianing Yao	Py-Torch, Pandas, Tenser-Flow	MEL,SVM	G. Tzanetakis and P. Cook	Feature Fusion and Classifier Block
2018	Deepanway Ghosal and MF Kolekar	WEKA,Numpy, Matplot lib	KNN,DT,CNN	Z. Fu, G. Lu, K. M. Ting, and D. Zhang	Improved performance with RNN for different CNNs

2021	Athulya KM et al.	Anaconda, Tenser-Flow, Pandas	Decision Tree, KNN	B. K. Baniya, D. Ghimire, and J. Lee	Bidirectional Gated Recurrent Units
2021	Macharla Vaibhavi P Radha Krishna.	H2O, Scikit-learn, Pandas	CNN, RNN	G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew	End-to-End at scale cnn architecture
2021	Dhevan S Lau and Ritesh Ajoodha	G- Collab, Tenser-Flow	Cat Boost, CNN	L. Breiman, “Bagging predictors	Average accuracy and standard deviation achieved by the 1d CNN architectures

.

3. PROBLEM FORMULATION

The Genre of the song is one of the main category to classify millions of songs. There are handful of genres that categorise the songs. To the current generation and to the upcoming, a futuristic model that brings out better classification of the songs in music industry has to be built with the most efficient techniques and algorithms we have in today's world. To produce a model based on machine Learning that solves the challenge of automatic classification of songs into it's corresponding genre. To bring a better accuracy rate in the produced model/project as compared to the early attempts and preexisting models. With the fast growing technologies of Machine Learning, a useful model to the Entertainment industry has to build.

4. OBJECTIVES

To develop a model that classifies the input audio sample or song into the music genre which it belong.

To reduce or replace the involvement of manual tasks in the music genre classification.

To produce a model with better accuracy rate so that it can take part in real-time applications.

To develop a model that is better than many early research.

5. METHODOLOGY

Various kinds of time domain and frequency domains features are extracted from our dataset. We will choose the best features out of all extracted features. All the features which are finalised will be appended to a CSV file which will be used by our machine learning classifiers to classify different genres. The GTZAN dataset contains the audio file in .wav format. To train our machine learning models there is a need to extract features from that data. For features extraction and required initial preprocessing a python library is used for audio analysis call Librosa. After which, various techniques are used to clean and pre-process our data to make it suitable for ML classifiers. Finally, we optimise our model's performance by selecting the beat hyper-parameters. This is extremely important and a deciding factor in what kind of predictions our model gives.

METHODS

K-Nearest Neighbour (K-NN)

K-NN is a supervised Machine Learning Algorithm. K-NN reads the similarity between new case and the

preexisting case and categorise the new case into the most similar category in the available cases.

Support Vector Machine (SVM)

SVM is another supervised machine learning algorithm that classifies linearly based on margin maximisation principle. SVM creates a hyperplane that distinguish between two different classes. The algorithm improves the complexity of the classifier by performing structural risk minimisation to achieve good generalisation performance.

Random Forests

Random forest is a Machine Learning Algorithm that is used in classification and to solve regression problems. The algorithm works based on decision trees. Random forest creates uncorrelated forest of trees whose prediction is more accurate than a that of a single tree.

6.EXPERIMENTAL SETUP

In this section we look at the results we get from different levels of cross validation. In each case we obtain a confusion matrix. In the confusion matrix, the rows signify the actual genre while the columns are the genres in which the songs were classified into. So in each confusion matrix we're looking at the diagonal elements of the matrix to get the accuracy of the method. For the SVM and Random Forests classifier we used an open source software which gave us the confusion matrix along with the accuracy.

Random Projections with Random Forests

Fluctuation Patterns with k-NN

Fluctuation Patterns with Random Forests & many more

7.CONCLUSION

We will started the project with the initial setup and used MFCC to extract features from audio files. After that, we have to built a KNN classifier from scratch that finds K number of nearest neighbour based on features and maximum neighbour belonging to particular class gives as an output. We will cover approximately 80 per cent accuracy on the model. The main thing to identify and divide the audio into different features is amplitude and frequency that changes within a short span of time.

8.REFERENCES

[1] A Porter, D Bogdanov, R Kaye, R Tsukanov, and X Serra. Acousticbrainz: a community platform for gathering music information obtained from audio. In ISMIR, 2015.

[2] Michael I. Mandel and Daniel P.W. Ellis, Song-level Features and Support Vector Machines for Music Classification, Queen Mary, University of London, 2005.

[3] Jean-Julien Aucouturier and Francois Pachet. Improving timbre similarity : How high's the sky. Journal of Negative Results in Speech and Audio Sciences, 1(1), 2004

[4] Prajwal R, Shubham Sharma, Prasanna Naik, Mrs. Sugna Mk. A paper on MUSIC GENRE CLASSIFICATION USING MACHINE LEARNING (IRJMETS), July-2021.

[5] Dan Ellis, Adam Berenzweig, and Brian Whitman. The "uspop2002" pop music data set, 2005.

http://labrosa.ee.columbia.edu/projects/musicsim/uspop2002.html.

[6] Beth Logan and Ariel Salomon. A music similarity function based on signal analysis. In ICME 2001, Tokyo, Japan, 2001.

[7] K. Choi, G. Fazekas, and M. Sandler, “Automatic tagging using deep convolutional neural networks,” arXiv preprint arXiv:1606.00298, 2016.

[8] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in Proceedings of the 27th international conference on

machine learning (ICML-10), pp. 807–814, 2010.

[9] K. Cho, B. Van Merri¨enboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” arXiv

preprint arXiv:1409.1259, 2014.

[10] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] U. Marchand and G. Peeters, “The extended ballroom dataset,” 2016.

[12] F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer, “Evaluating rhythmic descriptors for musical genre classification,” 2004.