Journal of Engineering Innovations & Technology (2024)

Journal of Engineering Innovations & Technology

journal homepage: www.journal-eit.online

Paper Type: Original Article

Application of Python-Based Music Spectrum Analysis and Machine Learning Regression Models in Music Classification

Dinghao Mo[1,*]

1.Shanghai Fu Dan High School

Abstract

This paper aims to investigate the analysis of audio waveforms, spectrograms, amplitude spectra after Fast Fourier Transform (FFT), and Mel spectrograms of different genres of music in the Python programming environment. By quantifying music, the paper seeks to achieve classification of various music genres. Regression models in machine learning are employed to analyze the amplitude spectra and classify different genres of music. The paper trains the model using several common music genres, validates its feasibility through testing, and concludes by applying the insights from the analysis to guide the composition of music and classify musical compositions.

Keywords: Mel spectrogram, Fourier Transform, machine learning, regression model, music classification

1. Introduction

Sound, as a physical phenomenon generated by vibrations, can largely be described by its frequency and amplitude. The loudness and pitch of a sound are determined by variations in its frequency and amplitude. As illustrated in Figure 1, the waveform of a classic cough sound recorded throughout the entire process can be quantified into sinusoidal functions. By analyzing the amplitude and frequency of each segment of the wave, the three stages of a cough can be discerned.

In the field of music, these sound characteristics constitute the unique musical language of different works, reflecting the emotions, styles, and creative expressions of their composers. To deeply understand and quantify these acoustic properties in music, spectrum analysis becomes an important tool. By analyzing the spectrum of music, we can obtain information such as audio waveforms, spectrograms, amplitude spectra after Fourier Transform, and Mel-spectrograms. These spectral representations can reveal important information about the vibrational frequencies, energy distribution, harmonics, and resonances contained in the audio signals. Therefore, by analyzing the music spectrum, we can gain a deeper understanding of the structure and expression forms of music.

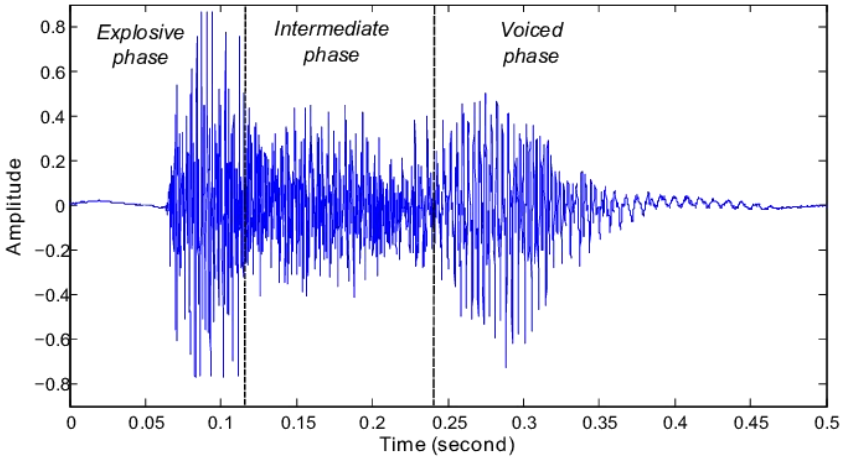

Figure 1 Waveform of Cough Sound[1]

However, as the volume of music increases, relying solely on manual analysis and classification becomes increasingly difficult. To achieve more efficient classification of different types of music, machine learning methods have emerged. In this paper, we will use regression models, a tool in machine learning, to analyze the amplitude spectrum of music and establish classifiers to automatically identify different music genres. By combining music spectrum analysis and machine learning regression models, we aim to deepen our understanding of music and provide a novel and efficient method for music classification. This will bring new possibilities to music research, creation, and industry development, promoting continuous innovation in music science and technology.

2. Music Spectrum Analysis

2.1 Basics of Spectrum Analysis

Spectrum analysis tools, utilizing the open-source programming language Python and the Librosa music analysis library for Python, were employed for analysis, plotting, and data processing using numpy and matplotlib.pyplot libraries. In this analysis, we introduce the audio waveform, audio spectrogram, amplitude spectrum after Fast Fourier Transform (FFT), and the Mel-spectrogram mentioned earlier.

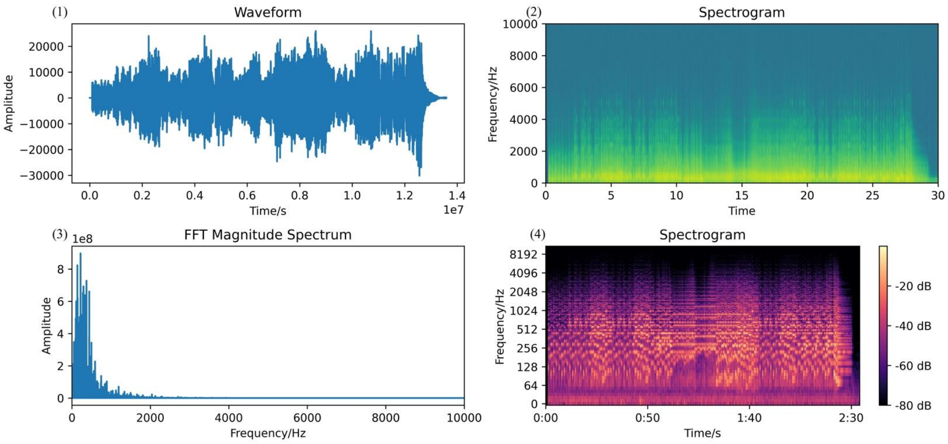

Figure 2 Quantitative analysis of Piano Music of the Night

(1) Waveform (2) Spectrogram (3) Amplitude Spectrum after FFT (4) Mel-Spectrogram

First, we extracted the waveform of "Piano Song at Night," characterized by a low pitch, slow rhythm, and soothing style. Figure 2(a) shows the waveform of this music, with the horizontal axis representing time and the vertical axis representing amplitude. From the waveform, it can be seen that the music's loudness shows obvious sections, with more regions of larger amplitude and less variation compared to regions of smaller amplitude.

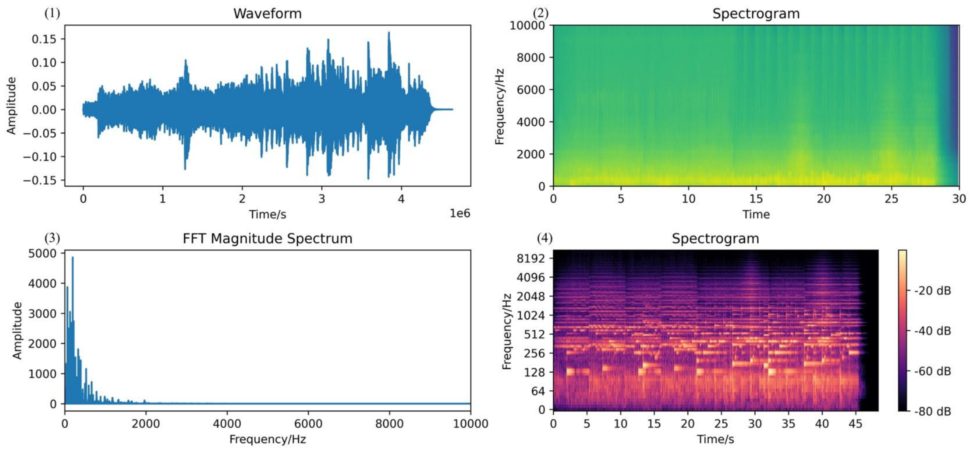

Figure 3 Quantitative analysis of "Ski Dream"

(1) Waveform (2) Spectrogram (3) Amplitude Spectrum after FFT (4) Mel-Spectrogram

Similarly, we conducted a simple arrangement based on the waveform of "Piano Song at Night," resulting in a song titled "Skiing Dream." This piece, similar in melody to "Piano Song at Night," features smooth and quiet music. From the waveform, it can be seen that the amplitude remains relatively stable, with only a slight increase in the latter part.

2.2 Spectrum Analysis in Python and Use of the Librosa Library

Figures 2 and 3 were plotted and analyzed in the Python environment. This section continues to introduce two other spectrograms, specifically the spectrograms (2) and (4) in Figures 2 and 3, with (4) being the Mel-spectrogram. The general spectrogram's horizontal axis represents time, and the vertical axis represents frequency, i.e., the pitch of the music.

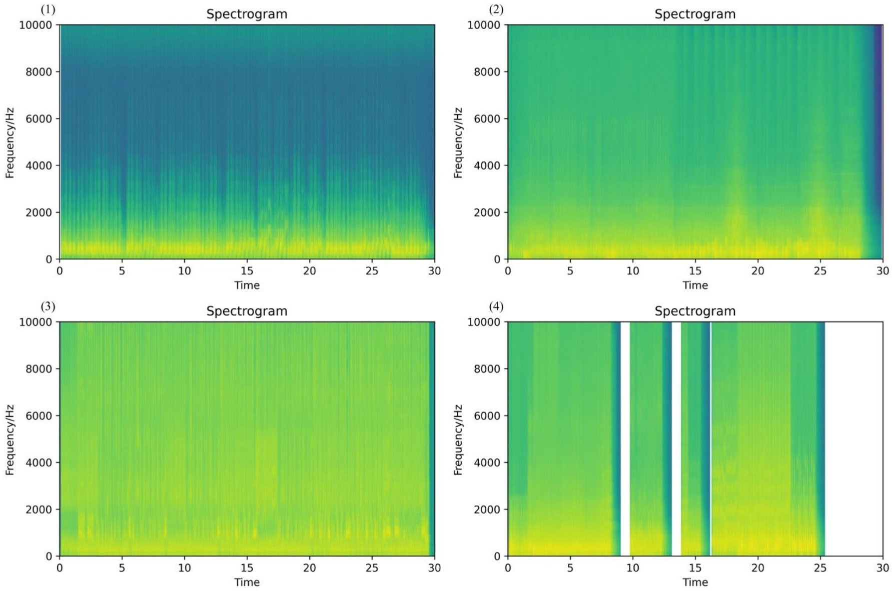

Figure 4 Music spectrum diagram

(1) "Piano Song at Night" (2) "Skiing Dream" (3) "No Place to Hide" (4) "In the Sea"

Figure 4 shows the spectrograms of four different pieces of music. "Skiing Dream" and "In the Sea" are self-composed songs, while (1) and (2) represent soothing songs of the same style. These pieces' frequencies mainly concentrate in the low-frequency range. In contrast, (3) and (4) exhibit faster rhythms and higher pitches, showing more high-frequency components in their spectrograms. The combination of spectrograms and waveforms vividly displays the changes in pitch and loudness over time. Notably, the song "In the Sea" was composed under the guidance of the spectrogram of "No Place to Hide."

The Mel-spectrogram, extracted using Python's Librosa library as shown in Figure 5, and the general spectrogram play different roles in audio signal analysis. The general spectrogram, obtained via Fourier Transform, displays the intensity distribution in the frequency domain, with the horizontal axis representing time and the vertical axis representing frequency. The Mel-spectrogram, on the other hand, applies a Mel filter bank to the frequency axis, making its vertical axis represent Mel frequency, aligning better with human auditory perception of pitch.

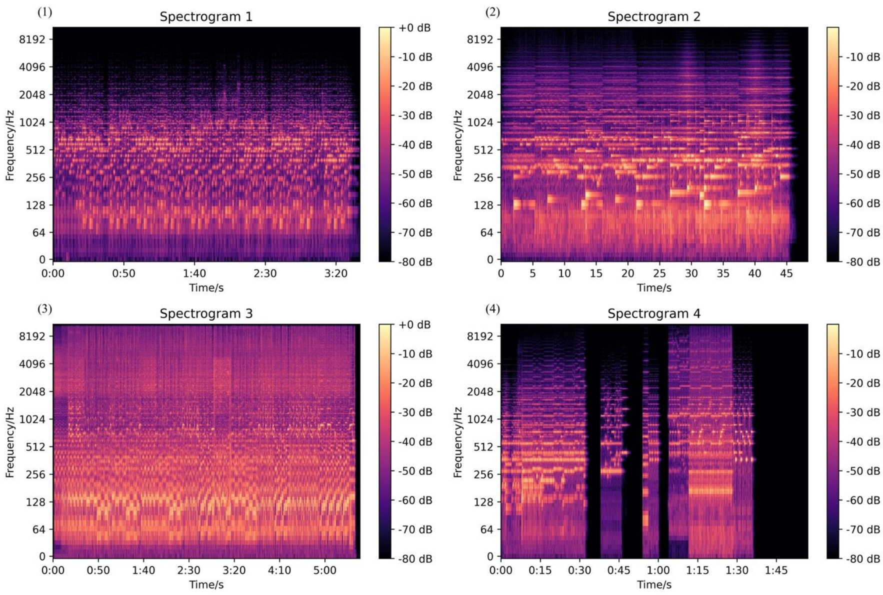

Figure 5. Meir language spectrum

(1) "Piano Song at Night" (2) "Skiing Dream" (3) "No Place to Hide" (4) "In the Sea"

2.3 Spectral Features and the Role of Fast Fourier Transform in Music Classification

Although we obtain the waveform, spectrogram, and Mel-spectrogram of a piece of music, and these diagrams can provide much information for arranging music, distinguishing and comparing songs with subtle differences from complex spectra remain challenging. Therefore, this paper utilizes the Fast Fourier Transform (FFT) from electromagnetic wave knowledge to obtain the amplitude spectrum of the music over the entire time period. Although some information is lost, it more clearly reveals the similarities and differences between different pieces of music for distinction.

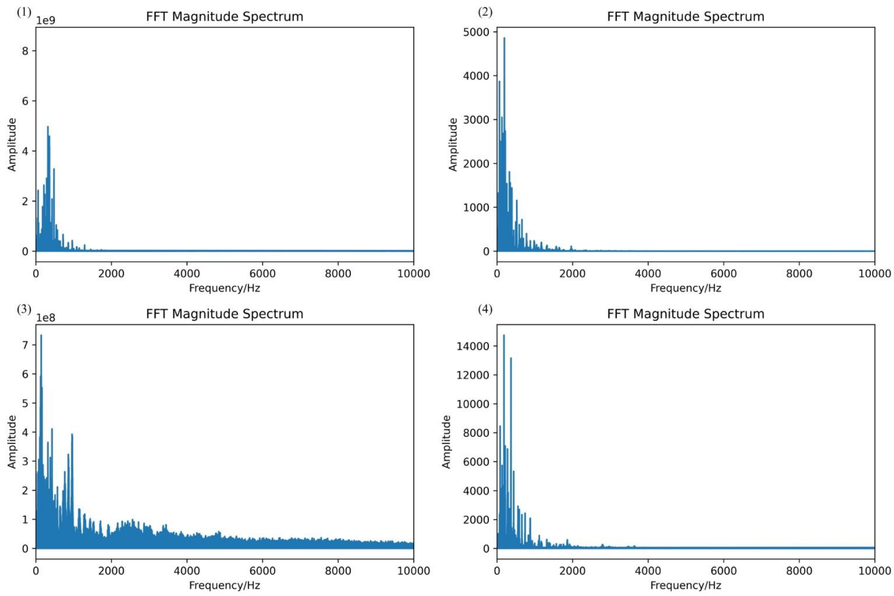

Figure 5 Amplitude spectrum of fast Fourier transform

(1) "Piano Song at Night" (2) "Skiing Dream" (3) "No Place to Hide" (4) "In the Sea"

From the amplitude spectrum, it can be clearly seen that the frequencies of "Piano Song at Night" and "Skiing Dream" mainly range between 300-800Hz, characterized by a lower and slower sound. In contrast, the frequency distribution of the song "No Place to Hide" by Black Panther Band is very wide, with significant components even in the extremely high frequency range of up to 10000Hz. Although "In the Sea" appears similar to "No Place to Hide" in the spectrogram, the amplitude spectrum after FFT shows that "In the Sea," despite having some components in the 2000-4000Hz range, lacks the extremely high-frequency components (4000-10000Hz) seen in "No Place to Hide."

3. Machine Learning Regression Models

3.1 Logistic Regression Model

For a simple binary classification problem, we need to input an independent variable, also called a feature value x, which can quantify the problem well. After inputting x, a y value of 0 or 1 can be output, ultimately achieving classification.[5]

$$z = \omega^T x + b \qquad (1)$$

This function is a linear function and cannot output 0 or 1, so an additional function called the Sigmoid logistic function is introduced.

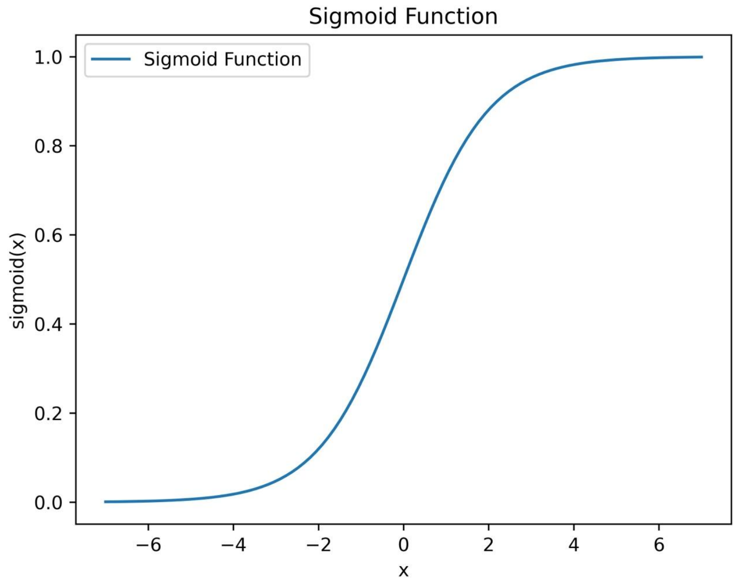

Figure 6 Sigmoid Function Graph

Using z as the horizontal coordinate of the Sigmoid function can achieve an output of 0 or 1. Let:

$$y=\frac{1}{1+e^{-z}}=\frac{1}{1+e^{-\left(\omega^{T}x+b\right)}} \qquad (2)$$

Taking the logarithm of both sides, we get:

$$Ln\frac{y}{1-y}=\omega^{T}x+b \qquad (3)$$

If y is the positive probability, then 1-y is the negative probability, thus obtaining the conditional probability.

$$J(w)=\min\left(-\frac{1}{n}\sum_{i=1}^{n}\left[y_{i}\left(w^{T}x+b\right)-\ln\left(e^{w^{T}x+b}+1\right)\right]\right) \qquad (4)$$

Given a loss function, the corresponding value can be found when the loss function is minimized. Solving for the minimum of this function uses high school calculus knowledge of derivatives, and this algorithm is called the gradient descent algorithm.

The above implementation of a general binary classification problem uses logistic regression. For multi-classification problems, the Sigmoid function cannot meet the requirements, so a function called Softmax[6] is introduced to solve multi-classification problems. The function is given as:

for a given n-dimensional vector, the Softmax function maps it to a probability distribution. The standard Softmax function is defined by the following formula:

$$\sigma(X)_i = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)} \quad (5)$$

Where n is the number of classifications, this function can realize the multi-classification problem of logistic regression.

4. Experimental Testing and Case Study

4.1 Data Selection and Description

This study extracts certain feature values to construct the regression model, as shown in Figure 7, selecting common music classifications such as Classical, Jazz, Country, Pop, Rock, and Metal, attempting to analyze their features.

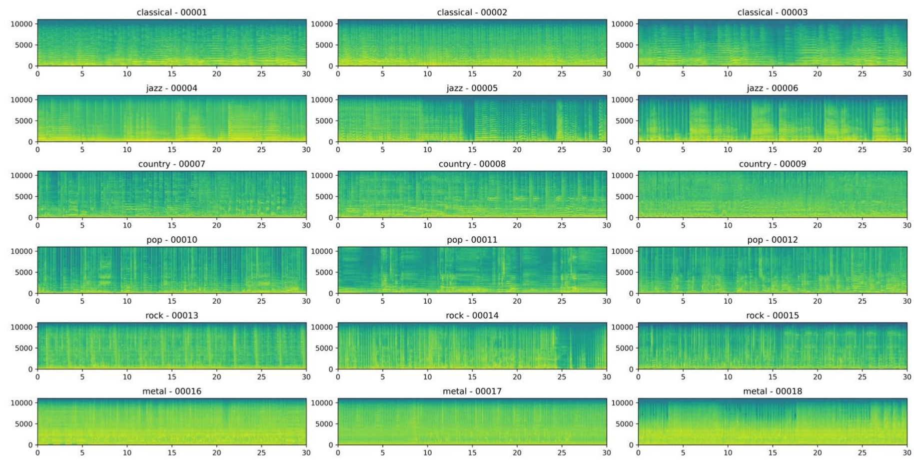

Figure 7 Spectrum Diagram of Different Types of Music

Classical music mainly has its frequencies distributed in the mid-to-low frequency range and is relatively smooth and monotonous. Jazz music, despite not having high frequencies, has fast frequency changes, characterized by quick variations. Compared to Classical music, Country music has lower frequencies and is more soothing. Pop music has the fastest frequency changes. Rock music is evenly distributed from low to high frequencies, while Metal mainly concentrates in the mid-to-high frequency range. The extraction of spectral features is of some reference value, but due to the complexity of the spectrum, Fourier transform magnitude spectrum analysis is used.

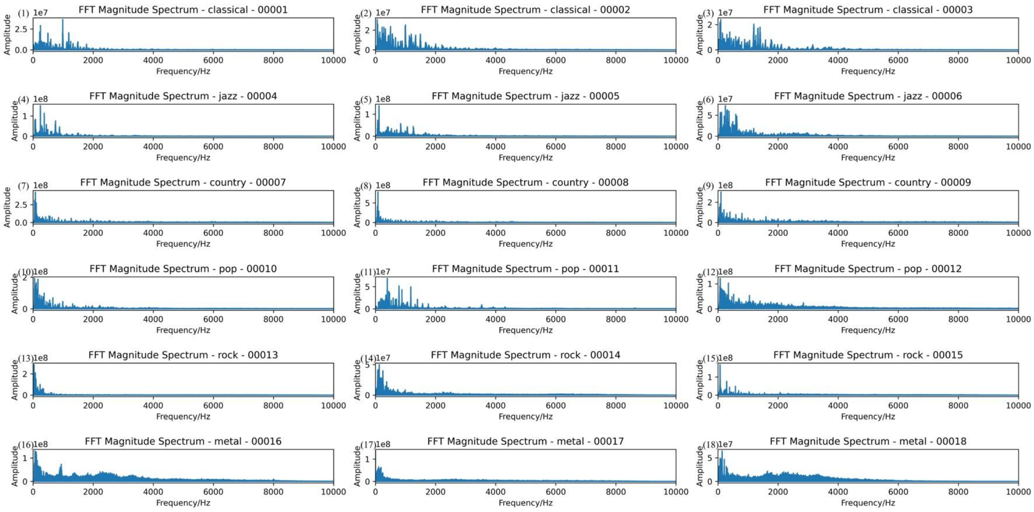

Figure 8 Spectrum Diagram of Different Types of Music
(1)-(18) Classical, Jazz, Country, Pop, Rock, Metal

In the Fourier transform magnitude spectrum, the characteristics of various music types can be clearly seen. Classical music, being classical, mainly has frequencies distributed between 0-2000Hz with concentrated distribution; Jazz music is mainly distributed between 0-1000Hz; Country music has the lowest frequencies, mainly between 0-500Hz; Pop music frequency distribution varies with different songs due to the diverse nature of pop music; Rock music has a broad frequency distribution but mainly concentrates around 300Hz; Metal music has the highest frequency range, mainly between 0-8000Hz. These Fourier transform magnitude spectra, with their distinct features, are used as feature values to construct the regression model.

4.2 Experimental Testing

After determining the feature values, we can construct our regression model. Due to copyright issues, we downloaded an open-source public music set, which includes 6 categories with 100 songs each. Fourier transform was performed on the spectrums of these songs to obtain the fft data files, which were used as the feature value training and testing sets for the model.

"from sklearn.linear_model import LogisticRegression"

Before this, a training set needs to be constructed, with the music divided into six categories: "Country, Classical, Rock, Jazz, Pop, Metal."

"

```
genre_list = ["Country", "Classical", "Rock", "Pop", "Jazz", "Metal"]
X = []
Y = []
for g in genre_list:
    for n in range(100):#Six categories of music, 100 songs each
        rad="/Users/a/Downloads/trainset/"+g+"."+str(n).zfill(5)+ ".fft"+".npy"
        fft_features = np.load(rad)
        X.append(fft_features)
```

Y.append(genre_list.index(g))

X = np.array(X)#Convert list to array

Y = np.array(Y)

"

Randomly select half of the music as the training set and the other half as the testing set. Construct the logistic regression model:

model = LogisticRegression()#Create logistic regression object

model.fit(X, Y)

Finally, import the test music for testing.

4.3 Results Analysis and Discussion

sample_rate, test = wavfile.read("/Users/a/Music/Music/Media.localized/Music/Unknown Artist/Unknown Album/Ski Dream.wav")

test=np.reshape(test,(1,-1))[0]

testdata_fft_features = abs(fft(test))[:1000]

print(sample_rate, testdata_fft_features, len(testdata_fft_features))

type_index = model.predict([testdata_fft_features])[0]

A total of 10 songs were tested. The self-composed song "Ski Dream" and the song "Night Piano" both belong to Country music, which is consistent with the initial analysis of this paper. "Ski Dream" was composed with the spectral guidance of "Night Piano." "Restless" and "In the Sea" are classified as Pop songs after testing due to their varying tones and rhythms, which basically meets expectations.

```
/var/folders/64/hpr1dhyx2r94zd3nxm7z081w0000gn/T/ipykerr
ional header parsing as it was created on Python 2. Save
  fft_features = np.load(file_path)
Logistic Regression Accuracy: 0.48333333333333334
Starting read wavfile...
Predicted Genre: classical
/Library/Frameworks/Python.framework/Versions/3.11/lib/p
lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale th
    https://scikit-learn.org/stable/modules/preprocessir
Please also refer to the documentation for alternative s
    https://scikit-learn.org/stable/modules/linear_model
  n_iter_i = _check_optimize_result(
```

Figure 9 Regression Model Prediction Results and Accuracy

We tested the accuracy of the model, and after actual testing, the accuracy reached 0.48333.

Subsequently, we constructed our own "Chinese Modern Music Dataset," consisting of some open-source Chinese classical music. Due to the limited amount of music, only three categories were selected: "Shaanxi Folk Songs," "Chinese Rock," and "Popular Ancient Style."

genre_list = ["Shaanxi Folk Songs", "Chinese Rock", "Popular Ancient Style"]

X = []

Y = []

for g in genre_list:

for n in range(20):#3 categories, 20 songs each

rad="/Users/a/Downloads/trainset1/"+g+"."+str(n).zfill(5)+ ".fft"+".npy"

fft_features = np.load(rad)

X.append(fft_features)
Y.append(genre_list.index(g))
X = np.array(X)#Convert list to array
Y = np.array(Y)

After testing, the music classification was basically achieved, but due to the small sample size, some music was misclassified between "Popular Ancient Style" and "Shaanxi Folk Songs," necessitating further sample size increase to improve the model's accuracy.

4.4 Comparison between KNN Model and Regression Model

Based on the regression model, this paper also uses the common KNN model to compare prediction results. The self-composed song "Ski Dream" is still selected, and Figure 10 shows the prediction results and accuracy of the KNN model and regression model.

```
/var/folders/64/hpr1dhyx2r94zd3nxm7z081w0000gn/T/ip
tional header parsing as it was created on Python 2
  fft_features = np.load(file_path)
/Library/Frameworks/Python.framework/Versions/3.11/
lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or sca
    https://scikit-learn.org/stable/modules/preproc
Please also refer to the documentation for alternat
    https://scikit-learn.org/stable/modules/linear_
  n_iter_i = _check_optimize_result(
Logistic Regression Accuracy: 0.48333333333333334
KNN Accuracy: 0.5333333333333333
Starting read wavfile...
Predicted Genre (Logistic Regression): country
Predicted Genre (KNN): country
```

Figure 10 Prediction Results and Accuracy of Regression Model and KNN Model

The results show that both models predict "Ski Dream" as Country music. The regression model has a prediction accuracy of 0.48333 on a large testing set, while the KNN model achieves an accuracy of 0.5333, slightly better than the regression model.

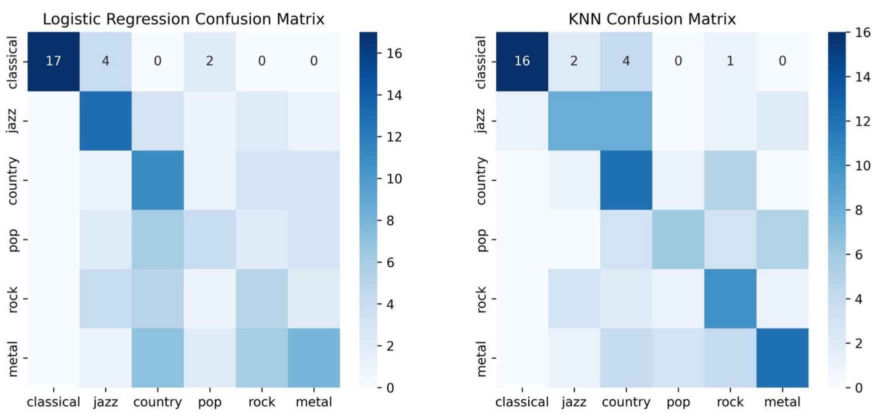

Figure 11 Heat Map of Prediction Results of Regression Model and KNN Model

Figure 11 shows the accuracy of different models in predicting different categories of music in the testing set. Darker color blocks indicate accurate predictions, while white color blocks indicate errors. Both models predict classical music very accurately, while the regression model is more accurate in predicting Jazz music. The prediction for Rock music is inaccurate. The KNN model accurately predicts Country and Metal music.

5. Conclusion

This paper uses Python for music data analysis and machine learning modeling, deeply studying the spectrograms, Mel spectrograms, and waveforms of different music types, further conducting creative practice and constructing music classifiers. The entire process covers music analysis, creation, and the building of machine learning models, involving the physics of sound, frequency, and mechanical waves, as well as mathematical regression models and conditional probabilities. It provides a rich and in-depth attempt at interdisciplinary research in the music field. Two music classifiers were successfully constructed using regression models. The first classifier includes six main music categories, covering a wide range of music types such as "Rock," "Metal," "Pop," "Country," "Classical," and "Jazz." The second classifier focuses on three Chinese music categories: "Shaanxi Folk Songs," "Popular Ancient Style," and "Chinese Rock." This classification system, using machine learning methods, provides a novel and efficient way for the automatic classification of music styles, offering potential for optimizing music library management and recommendation systems. Through this research, we not only explored the combination of music creation and data science but also revealed the broad application potential of machine learning in music classification. Future work can further explore the analysis of more music features, introducing more complex models to improve classifier accuracy. Additionally, deep collaboration with music professionals and timely collection of user feedback will help continuously optimize and improve music classifiers, better serving music creators and enthusiasts. This research framework provides an encouraging practical example of introducing data science into the music field, paving the way for future interdisciplinary research.

References

[1] Drugman T, Urbain J, Bauwens N, et al. Audio and contact microphones for cough detection[J]. arXiv preprint arXiv:2005.05313, 2020.

[2] Yang Y H, Lin Y C, Su Y F, et al. Music emotion classification: A regression approach[C]//2007 IEEE International Conference on Multimedia and Expo. IEEE, 2007: 208-211.

[3] Kumar D P, Sowmya B J, Srinivasa K G. A comparative study of classifiers for music genre classification based on feature extractors[C]//2016 IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER). IEEE, 2016: 190-194.

[4] Zhao Kaihua, Luo Weiyin. New Concept Physics Teaching, Mechanics [M]. Higher Education Press,1995. (in Chinese)

[5] Logistic Regression (Logistic Regression). [EB/OL]. https://blog.csdn.net/weixin_55073640/article/details/124683459, 2023-06-04. (in Chinese)

[6] Luyouqi11. Machine Learning (10) Logistic regression Multivariate classification (Multi - class classification). [EB/OL]. https://blog.csdn.net/luyouqi11/article/details/132080943, 2023-08-03. (in Chinese)