Music Genre Classification

Fan Wenyi
1023041004
Nanjing University of Posts and Telecommunications
Nanjing, China

***Abstract*—Music genre classification is a significant task in the field of audio signal processing and machine learning, aiming to automatically categorize music into predefined genres based on its acoustic features. This paper explores the application of various classification algorithms, including logistic regression and k-nearest neighbors, on a dataset containing diverse musical genres such as classical, jazz, country, pop, rock, and metal. We conduct extensive experiments to evaluate the performance of these algorithms, considering factors like accuracy, precision, and confusion matrices. Additionally, the study investigates the impact of different feature representations, such as Fast Fourier Transform (FFT) coefficients, on the classification accuracy. The results reveal the strengths and limitations of the selected algorithms in the context of music genre classification, providing insights into the optimal approaches for accurate genre prediction. This research contributes to the broader field of audio classification and offers practical implications for automated music organization and recommendation systems.**

***Index Terms*—music, classification, k-nearest, genre, logic regression**

I. Introduction

The dynamic landscape of digital music, characterized by an ever-expanding array of genres, necessitates sophisticated methods for effective music classification. As the volume of musical content continues to grow, the demand for automated systems capable of accurately categorizing songs becomes increasingly critical. Music genre classification, a discipline within the domains of audio signal processing and machine learning, assumes a central role in navigating and organizing this vast musical ecosystem.

Our study embarks on an exploration of the intricate domain of music genre classification, focusing on a diverse dataset that spans genres such as classical, jazz, country, pop, rock, and metal. The overarching goal is to harness the power of machine learning techniques, particularly logistic regression and k-nearest neighbors, to discern the efficacy of these algorithms in accurately categorizing music across diverse genres.

In addition to algorithmic scrutiny, we delve into the realm of feature representations, particularly emphasizing the influence of Fast Fourier Transform (FFT) coefficients. Understanding how different feature representations impact overall classification performance is pivotal for refining algorithms and enhancing the robustness of music classification systems.

The study not only seeks to unravel the technical intricacies of classification algorithms but also aims to shed light on the challenges and opportunities inherent in music genre classification. Through a comprehensive analysis, we aspire to offer valuable insights that contribute to the continual improvement of classification algorithms. Beyond the technical nuances, the research endeavors to enhance automated music recommendation systems, fostering a more personalized and enjoyable music exploration experience for users.

As we navigate through this multifaceted exploration, our research addresses fundamental questions surrounding the adaptability of machine learning algorithms to the intricacies of musical genres. By doing so, we aim to equip researchers, practitioners, and enthusiasts with a deeper understanding of the evolving landscape of music genre classification, fostering advancements that resonate with the diverse tastes of music enthusiasts globally.

II. Related Works

Music genre classification has been a prominent research area within the domains of audio signal processing and machine learning, spurred by the increasing availability of large-scale music dataset. Numerous studies have contributed to advancing the understanding and effectiveness of algorithms in this domain.

Early approaches often relied on handcrafted features derived from signal processing techniques, such as Mel-frequency cepstral coefficients (MFCCs)[3] and chroma features[6]. These features aimed to capture essential characteristics of the audio signal, enabling effective discrimination between genres. Noteworthy contributions include Tzanetakis and Cook's seminal work[7], which explored the application of statistical and spectral features for genre classification.

As the field progressed, machine learning techniques gained prominence, allowing for a more automated and data-driven approach to genre classification. Supervised learning algorithms, including support vector machines (SVMs)[2] and k-nearest neighbors (KNN)[4], have demonstrated success in discerning genre patterns within music data. Additionally, deep learning models, particularly convolutional neural networks (CNNs)[5] and recurrent neural networks (RNNs)[1], have shown promise in learning hierarchical representations from raw audio data, bypassing the need for handcrafted features.

Moreover, research has extended beyond traditional audio features, incorporating additional modalities such as lyrics, album metadata, and user listening patterns to enhance classification accuracy. The integration of multiple modalities has become a burgeoning area, reflecting the multidimensional nature of music classification.

Despite these advancements, challenges persist, including the inherent subjectivity of genre labels, the dynamic nature of musical styles, and the need for robustness to variations in recording quality. This section provides an overview of the diverse approaches and methodologies employed in music genre classification, laying the foundation for our exploration of novel strategies and insights in the subsequent sections of this paper.

III. Problem Statement

Music genre classification is inherently a multifaceted problem, as it involves categorizing songs into predefined genres based on their inherent audio characteristics. Formally, the task can be defined as follows.

A dataset consisting of audio samples from various songs. Each audio sample is associated with a predefined genre label. The task faces amount of challenges. High Dimensionality Challenge: Audio data is often represented in high-dimensional feature spaces, posing challenges in identifying relevant patterns and avoiding the curse of dimensionality.Subjectivity and Ambiguity Challenge: Genre classification is subjective and may vary among listeners. Some songs may exhibit characteristics of multiple genres, introducing ambiguity into the classification task.

Despite all the challenges, we have good hopes for his future.The model should be robust to variations in recording quality, background noise, and diverse musical styles to ensure real-world applicability. Capturing temporal dynamics in music, such as rhythm changes and transitions, is crucial for a holistic understanding of genre characteristics.The primary goal is to design a data mining solution that achieves high accuracy in predicting the genre labels of songs, while also addressing the aforementioned challenges. The solution should leverage advanced techniques in machine learning and signal processing to extract meaningful representations from the audio data.This formalization sets the stage for exploring innovative methodologies and algorithms to tackle the intricacies of music genre classification within the realm of data mining.

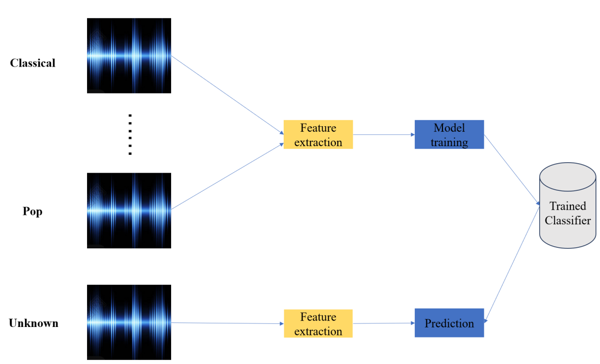

Fig. 1. Music classify pipeline

Algorithms

Addressing the intricacies of music genre classification involves the orchestration of advanced algorithms and innovative solutions. In our pursuit of an accurate and adaptable model, we employ a combination of feature extraction, machine learning, and signal processing techniques.

A. Feature Extraction

To distill meaningful information from the complex audio data, we employ feature extraction methods. These include the extraction of spectral features such as Mel-Frequency Cepstral Coefficients (MFCCs), Fast Fourier Transform (FFT), and Pitch. By capturing key characteristics of the audio signal, these features serve as the foundation for subsequent analysis.

The sound signal is a complex signal that consists of a number of sound waveforms of different frequencies that propagate together by causing perturbations to the medium (changing pressure) to form the sound we hear. When recording sound, we simply capture the synthesized amplitudes of these waveforms . The Fourier transform is a mathematical formula that allows us to break down a signal into individual frequencies and frequency amplitudes. In other words, it converts a signal from the time domain to the frequency domain, and the result is called a spectrum. This is possible because every signal can be broken down into a set of sine and cosine waves that add up to equal the original signal. This is a well-known theorem known as Fourier's theorem. The Fourier Transform is one of the most commonly used acoustic features and is an operation that maps an audio signal from the time domain to the frequency domain. In this paper, we mainly use FFT to extract the audio features (as shown in the visualization of Fig. 2).FFT is equivalent to transforming the signal from the temporal space, which has time as the coordinate, to the frequency space, which has FFT basis as the coordinate.The FFT algorithm is simple and suitable for application scenarios where CPU resources are tight, and the frequency components are better preserved, especially in the high-frequency part.

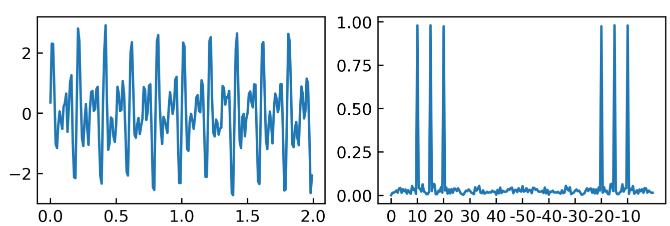

Fig. 2. Fast Fourier Transform feature

B. Machine Learning Models

Our approach embraces the power of machine learning models for classification tasks. Logistic Regression and k-Nearest Neighbors (KNN) models are utilized to discern patterns within the extracted features. Logistic Regression provides a probabilistic framework for multi-class classification, while KNN leverages proximity-based relationships to assign genres based on similarity to training instances.

Logistic regression was originally designed to deal with binary classification problems, but through clever morphing, it can also be applied to multi-category problems. For problems with multiple categories, logistic regression transforms the problem into multiple binary classification problems. Each category is treated as a binary classification task, and the goal of the model is to distinguish the current category from all other categories. For each category, a separate logistic regression model is trained. During the training of each model, samples from the target category are labeled as positive categories, while samples from all other categories are labeled as negative categories. In the prediction phase, input samples are fed into all trained logistic regression models. Each model outputs a probability value indicating the probability that the sample belongs to the current category. Eventually, the category with the highest probability is selected as the predicted category for the sample. The idea of this multi-categorization approach is similar to a hierarchical strategy, where each logistic regression model focuses on distinguishing one specific category, while all other categories are considered as a whole. This approach allows logistic regression to be easily extended to deal with multicategorization problems without the need to redesign the model structure.

The K-nearest neighbor method is an instance-based learning method that can be applied in multiple classification tasks as well. For each sample to be classified, the K nearest neighbors to that sample are found in the feature space. The distance is usually measured using Euclidean distance or other similarity measures. The number of samples belonging to each category in these K neighbors is counted. The majority voting principle is used to determine the category to which the samples to be classified belong, i.e., the category with the most votes is chosen as the prediction. In some variants of KNN, the weights of neighbors can be considered. Closer neighbors may be given higher weights to more strongly influence the voting result.The decision boundaries of KNN are nonlinear and adaptable to complex data distributions.The advantage of KNN is that it is simple and intuitive and does not require an explicit training process.

C. Model Training and Evaluation

In the intricate process of model training and evaluation, our focus extends beyond mere algorithmic implementation. The foundation of our study lies in the meticulous curation of a labeled dataset that faithfully captures the rich diversity of music genres. This dataset becomes the crucible in which our models are forged, offering them the opportunity to glean insights into the nuanced acoustic characteristics that distinguish various musical genres.

As our models traverse the intricacies of training, the learning process is akin to tuning an instrument, with each epoch refining the model's understanding of the intricate interplay of features within the music data. The goal is not just classification but a nuanced comprehension that transcends raw data – an understanding of the rhythmic cadence of jazz, the melodic richness of classical compositions, or the energetic pulse of rock.

To measure the efficacy of our models, a robust evaluation framework is imperative. The partitioning of the dataset into training and test subsets ensures a rigorous examination of the model's generalizability. Through this separation, we gauge the model's ability to extrapolate its learning beyond the confines of the training set, a pivotal characteristic for real-world applicability.

Performance metrics serve as our compass in navigating the evaluation landscape. Accuracy, a fundamental metric, quantifies the overall correctness of our model's predictions. Precision delves into the model's capability to avoid false positives, ensuring that when it asserts a genre classification, it does so with confidence.

As we venture into the realm of evaluation, our study aims not only to quantify performance but also to unravel the subtleties that underpin the success or challenges encountered. This involves a granular examination of the confusion matrix, unraveling instances where our models excel and areas where refinement may be required.

Through this holistic approach to model training and evaluation, our research seeks not just to build accurate classifiers but to contribute to the broader discourse on the application of machine learning in the domain of music genre classification. By delving into the intricacies of algorithmic decision-making, we strive to provide a comprehensive understanding that transcends numerical metrics and resonates with the nuanced artistry embedded in musical genres.

D. Model Adaptability

Given the dynamic nature of music genres and the constant changes in listener preferences, our model can be highly adaptive. In the future, regular updating and retraining of the model with new data can ensure that the model always adapts to emerging music genres and changes in musical styles.

In the symphony of algorithms and solutions, our methodology strives for a harmonious blend of precision, adaptability, and interpretability, paving the way for an enriched music genre classification experience.

IV. Evaluation

A. Dataset

The efficacy of our music genre classification system is contingent on the diversity and representativeness of the dataset employed. We curated a comprehensive dataset spanning genres such as classical, jazz, country, pop, rock, and metal. Each genre comprises 100 audio samples, ensuring a robust and balanced representation across the spectrum of musical styles.

The dataset encompasses variations in tempo, instrumentation, and tonality, mirroring the heterogeneity inherent in real-world musical collections. This diversity is crucial for training a model capable of generalizing well to unseen instances and addressing the challenge of genre ambiguity.

B. Experimental Results

We conducted experiments on music classification models using two algorithms, logistic regression and K-nearest neighbor method, and evaluated them on the same test set. The experimental results show that the model accuracy of the logistic regression algorithm is around 50 percent, while the K-nearest neighbor method achieves a model accuracy of around 67 percent. Clearly, the K-nearest neighbor method outperforms the logistic regression algorithm on this task.

However, despite the better performance of the K-nearest neighbor method, we also found that there is room for improvement in its model. To gain a deeper understanding of the classification results, we further calculated the confusion matrix (shown in Fig. 3). Observing the confusion matrix, we found that the music classification model of the logistic regression algorithm is prone to misclassification errors on the three categories of country, pop, and rock, incorrectly predicting them as other categories. In contrast, the model of the K-nearest neighbor method makes classification errors on the three categories jazz, country, and pop, incorrectly predicting them as other categories.

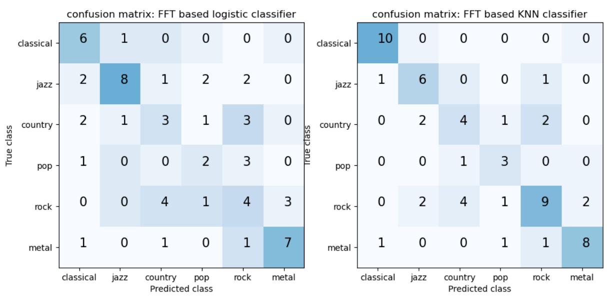

Fig. 3. Confusion Matrix

These observations provide valuable clues for further optimization and improvement of our music classification model, especially for the case of classification errors of both algorithms on specific categories.

V. Conclusion

In conclusion, the exploration of music genre classification within the context of data mining unveils a myriad of challenges and opportunities. As we traverse the landscape of audio data, the intricate nature of musical content, subjectivity in genre perception, and the dynamic interplay of various musical elements necessitate sophisticated data mining approaches.

Throughout this study, we have observed the significance of leveraging advanced techniques in machine learning, signal processing, and feature extraction to unravel patterns within high-dimensional audio data. The quest for a robust and accurate music genre classification model has driven the development of innovative methodologies capable of addressing challenges such as high dimensionality, subjectivity, and temporal dynamics.

However, it is crucial to acknowledge the evolving nature of music and the perpetual emergence of new genres. As the landscape of musical expression evolves, so too must our data mining models adapt to encompass a broader spectrum of genres and account for evolving listener preferences.

In the future, further research and development in music genre classification can benefit from interdisciplinary collaborations, incorporating insights from music theory, psychology, and human-computer interaction. The pursuit of real-world applicability demands continuous refinement of models, ensuring their adaptability to diverse musical styles, recording conditions, and evolving listener perceptions.

As we navigate the harmonious intersection of music and data mining, the journey unfolds not only as a technical pursuit but as a continuous exploration of the multifaceted nature of musical expression, enriching our understanding of both the art and science encapsulated within each note and genre label.

References

[1] Anirudh Ghildiyal, Komal Singh, and Sachin Sharma. "Music Genre Classification using Machine Learning". In: *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. 2020, pp. 1368–1372. DOI: 10.1109/ICECA49313.2020.9297444.

[2] De Rosal Ignatius Moses Setiadi et al. "Effect of Feature Selection on The Accuracy of Music Genre Classification using SVM Classifier". In: *2020 International Seminar on Application for Technology of Information and Communication (iSemantic)*. 2020, pp. 7–11. DOI: 10.1109/iSemantic50169.2020.9234222.

[3] D Pradeep Kumar et al. "A comparative study of classifiers for music genre classification based on feature extractors". In: *2016 IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*. 2016, pp. 190–194. DOI: 10.1109/DISCOVER.2016.7806258.

[4] Krittika Leartpantulak and Yuttana Kitjaidure. "Music Genre Classification of audio signals Using Particle Swarm Optimization and Stacking Ensemble". In: *2019 7th International Electrical Engineering Congress (iEECON)*. 2019, pp. 1–4. DOI: 10.1109/iEECON45304.2019.8938995.

[5] Mitt Shah et al. "Music Genre Classification using Deep Learning". In: *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*. 2022, pp. 974–978. DOI: 10.1109/ICCMC53470.2022.9753953.

[6] Leisi Shi, Chen Li, and Lihua Tian. "Music Genre Classification Based on Chroma Features and Deep Learning". In: *2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP)*. 2019, pp. 81–86. DOI: 10.1109/ICICIP47338.2019.9012215.

[7] G. Tzanetakis and P. Cook. "Musical genre classification of audio signals". In: *IEEE Transactions on Speech and Audio Processing* 10.5 (2002), pp. 293–302. DOI: 10.1109/TSA.2002.800560.