Comparison Between Two Algorithms in Music Genre Classification

Runming Weng
Faculty of Science, McGill Montreal, Quebec, H3A 033600, Canada

Keywords: Music Genre, Classification, Comparison, Nearest Neighbor, KNN.

Abstract: Through people's thousands of years of effort, a huge amount of music genres were created. Therefore, finding algorithms that can automatically classify the genres of music has become an essential problem in contributing modern digital music industry. Also, finding out which algorithm can complete the task more accurately can dramatically improve the efficiency in real applications like sending music that users are interested in based on the music users hear most. This study compares several algorithms in the use of music genre classification and convinces the importance of music genre classification in modern digital applications, certain the advantages and disadvantages of different algorithms. The research is mainly focused on K-Nearest Neighbors (KNN) and Convolutional Neural Networks (CNN) using the GTZAN dataset The study discusses the capability of CNN in capturing complex temporal and spectral patterns, and KNN's effectiveness in genre identification based on feature proximity. The result proves KNN's reliability, accuracy, and adaptability. Offering insight into the realistic usage of the algorithms in the technology-driven music industry.

1 INTRODUCTION

Music plays an important role in modern society. It is not only a good way for people to relax, but also sometimes inspires people from frustrations. The reason why music is so useful in various aspects is that the type of music can be diverse. Music is usually compounded by several instruments and vocals. As a result, the music genre comes out to distinguish between different feelings music can provide. There are approximately 1300 kinds of music genres nowadays, some of them are well-known like Blues, Classic, Jazz, Rock, and Country (Steve 2023). Therefore, the classification of music genres is becoming more and more important in the evolving landscape of digital music. Which can be the base for constructing contemporary music recommendation systems, digital libraries, and streaming services. Hence the study of music genre classification not only contributes to the academic field of musicology but also holds significant practical relevance in the technology-driven music industry.

Classifying modern music genres accurately and effectively can be a hard task due to the vast and diverse music repositories nowadays. Also, music genres are often subjective and can overlap, making automated classification even more challenging One of the key advantages of using deep learning in music genre classification is its ability to learn data representations directly from the audio, images, and text, without requiring extensive manual feature engineering. This learning process involves multimodal approaches that combine audio, visual, and textual data, providing a more holistic view of music and improving classification performance (He 2022).

The research analyzes various algorithms used to classify different music genres. The first significance of this study is that it contributes to a better understanding of how different algorithms perform in the context of music genre classification, which can be helpful in building applications in the modern music industry. Second, it identifies potential gaps and areas for improvement in current classification methodologies, paving the way for future research and development.

By exploring and comparing the effectiveness of these algorithms, the research can conclude the accuracy, efficiency, and adaptability of different algorithms. In practice, the result provides convictive evidence for modern music applications to improve their recommend system therefore improving users' satisfaction.

In conclusion, this research stands at the intersection of musicology, computer science, and information technology, offering valuable

Weng, R.
Comparison Between Two Algorithms in Music Genre Classification.
DOI: 10.5220/0012829700004547

In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 137-141
ISBN: 978-989-758-690-3

contributions to each of these fields. By optimizing the understanding of music genre classification algorithms, the way people enjoy and explore music can be greatly enhanced.

2 METHODOLOGY

For studying different methods to classify different types of music genres. GTZAN dataset can be helpful (Andrada 2020). The GTZAN dataset, a cornerstone in the field of music genre classification, consists of 1000 audio tracks evenly distributed across 10 genres, each track being 30 seconds long. This dataset is widely recognized for its diversity and has been a benchmark in numerous studies, providing a reliable basis for evaluating classification algorithms. The reason why two methods were chosen, Convolutional Neural Networks (CNN) and K-Nearest Neighbors (KNN), is predicated on their contrasting natures: CNN's capability in handling complex patterns and KNN's effectiveness in feature-based classification. In applying CNN, a multi-layer architecture to process the extracted features was designed. The model included convolutional layers for feature detection, pooling layers for dimensionality reduction, and fully connected layers for classification. Aiming to capture both the spectral and temporal features inherent in the music tracks. Conversely, the KNN algorithm was used to explore a more straightforward, distance-based approach. By computing the distance between feature vectors of different tracks, KNN aimed to classify genres based on similarity in their feature space.

3 EXPERIMENTAL SETUP AND EVALUATION

KNN is a neighbor-based classification algorithm that is both simple and effective in identifying musical genres. It will compare the unknown songs from existing songs and find out which genre is the most relevant. To complete the whole process of classifying data. The first step should be down is Feature extraction,

Mel-Frequency Cepstral Coefficients (MFCCs): MFCCs are a feature representation in the field of audio signal processing and speech recognition, often used for identifying musical genres, speaker identification, and other audio analysis tasks. They are derived from the real cepstral representation of a windowed short-time signal derived from the fast Fourier transform of a signal. The Mel scale is applied to the power spectrum of this signal, followed by taking the log of the powers at each of the Mel frequencies. Finally, the discrete cosine transform (DCT) is applied to these log Mel spectrum values to yield the MFCCs. This process emphasizes the perceptually relevant aspects of the spectrum, making MFCCs particularly useful in audio-related machine-learning tasks (Logan 2000).

Zero Crossing Rate: a measure used in audio signal processing and speech recognition to quantify the smoothness of a signal. It calculates the rate at which the signal changes from positive to zero to negative or vice versa within a specific time frame. Essentially, it counts the number of times the audio waveform crosses the zero-amplitude axis. This measure is useful for distinguishing between different types of sounds in a signal, as smooth sounds like voiced speech have fewer zero crossings compared to rougher sounds like unvoiced fricatives (Bäckström 2024).

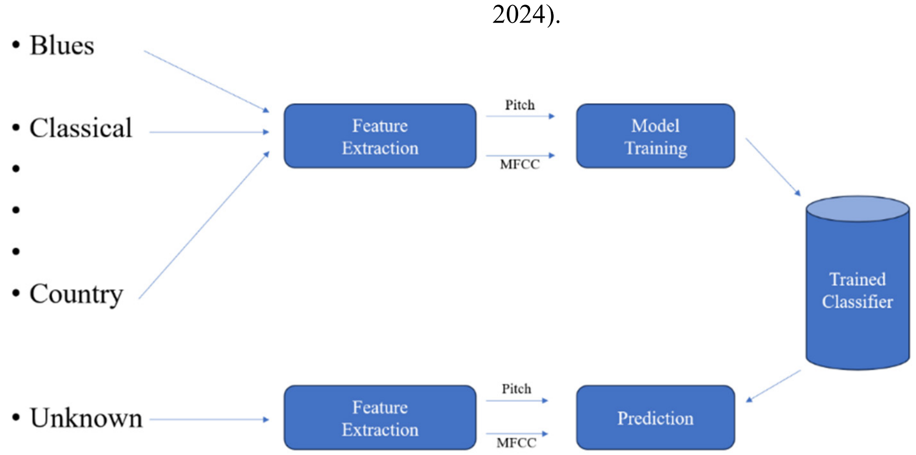

Figure 1. Working Principle of a Classification System (Data Flair2020).

Table 1: The accuracy of using KNN.

Algorithm	Accuracy	Algorithm	Accuracy
Decision Trees	0.63	Neural Network	0.68
KNN	0.81	Random Forest	0.81
Logistic Regression	0.69	Support Vector Machine	0.75
Naïve Bayes	0.51	XGBoost	0.91

Tempo (Beats Per Minute - BPM): tempo is a significant musical element that influences emotional processes in listeners. Tempo, often measured in beats per minute (bpm), can evoke different emotional responses (Liu et al 2018).

Spectral Centroid: an important concept in the analysis and description of musical timbre. It can be understood as the 'center of gravity' of the spectrum of a sound, representing the center frequency of the sound's spectral energy. This parameter is often associated with the perceived brightness of a sound. In a more technical sense, the spectral centroid is determined by calculating the weighted mean of the frequencies present in the signal, with their magnitudes as the weights (Sköld 2022).

The second step is to use Python, NumPy, and the Librosa library, which is adept at music and audio analysis, to build a model that can use those features to predict the kind of music by using the features mentioned before, the whole process showing in Figure 1 (Data Flair 2020). The result of the accuracy by using KNN to predict the music genres using a similar feature extraction approach, in this approach, 75% of data is used to train the model and the other 25% is to do some tests and find out the accuracy. The result is shown in Table 1 (Ghildiyal et al 2020).

CNNs have achieved significant success in the realm of image processing, which translates effectively into the audio field. In this context, audio features can be viewed as sequences of temporal images. This perspective allows the convolutional layers of CNNs to capture local patterns and spectral features within audio efficiently, aiding in distinguishing between various music genres.

Furthermore, the inherent nature of convolutional layers, characterized by parameter sharing and local receptive fields, equips CNNs with an inherent ability to handle translation invariance. In audio processing, this means that CNNs are capable of recognizing the same audio features, despite temporal shifts along the time axis. This attribute is particularly beneficial for music genre classification, as it ensures consistent identification of genre characteristics across different segments of a song.

The architecture of CNNs, with multiple convolutional and pooling layers, facilitates a gradual abstraction and combination of higher-level features. This hierarchical approach to feature learning allows networks to autonomously develop abstract representations that are crucial for music genre classification, thereby reducing the reliance on manual feature engineering.

Finally, the integration of regularization techniques, such as batch normalization and dropout, plays a vital role in enhancing the network's ability to generalize and prevents overfitting. This aspect is of paramount importance in music genre classification, where distinct music genres may exhibit overlapping audio features, necessitating a network with exceptional generalization capabilities.

CNN is constructed using Keras, featuring an input layer followed by five convolutional blocks. Each block included a convolutional layer with a 3x3 filter, 1x1 stride, and mirrored padding, a ReLU activation function, max pooling with a 2x2 window size and stride, and dropout regularization with a 0.2 probability. The filter sizes of these blocks were 16, 32, 64, 128, and 256, respectively. Following these blocks, the 2D matrix was flattened into a 1D array, followed by a regularization dropout with a probability of 0.5. The network concluded with a dense fully-connected layer using a SoftMax activation function. Which is also based on the GTZAN dataset. The result is shown in the table 2 (Lau and Ajoodha 2022).

Table 2: The result of CNN to classify the music genres.

Classifier	Epochs	Test Loss	Test Accuracy
CNN (30-Sec Features)	30	1.609	53.5%
CNN (3-sec Features)	50	0.873	72.4%
CNN (Spectrograms	120	2.254	66.5%

4 RESULT COMPARING

By comparing the accuracy shown before, CNN can provide 72.4% accuracy in predicting the music genres however KNN's accuracy can reach 81% by using the same dataset, which provides evidence that KNN performs better in predicting various music genres. KNN's good performance can be attributed to its efficacy in handling the specific characteristics of the GTZAN dataset. This demonstrates that the feature space of this dataset is well-suited for KNN's distance-based classification approach. Also shows that simpler algorithms like KNN can be more effective than their more complex counterparts in dealing with datasets where genres are well-separated in the feature space.

However, it's important to note that while KNN showed higher accuracy, it was not without its limitations. The algorithm struggled with genres that had subtle differences, a common issue in genre classification due to the subjective nature of music. Despite this, the overall performance of KNN was notably robust across the diverse genres present in the GTZAN dataset.

Although the performance of CNN is lower than KNN in this instance, was still noteworthy. CNN's ability to extract layered and complex features from the music tracks was evident, though it did not translate into superior accuracy in this particular study. This suggests that while CNNs are powerful tools for pattern recognition, their effectiveness can vary depending on the dataset and the specific characteristics of the task at hand.

5 COMPARATIVE ANALYSIS

The comparative analysis between KNN and CNN in this study offers valuable insights into the applicability of these algorithms in music genre classification. KNN's success indicates that for certain datasets, simpler algorithms can not only compete with but also surpass more complex models like CNN in terms of accuracy.

However, CNN's lower performance in this context does not diminish its potential in other scenarios. CNNs are known for handling complex patterns and large datasets, making them suitable for tasks where the feature space is not as clearly defined or where the data is more complex.

In conclusion, this study highlights that the choice between KNN and CNN for music genre classification should not be based on the complexity of the algorithm alone. Instead, it should be informed by the characteristics of the dataset and the specific requirements of the classification task. The paper's findings suggest that in scenarios where the feature space is well-structured and genres are distinctly separable, simpler algorithms like KNN can provide superior performance.

6 CONCLUSION

This paper uses the GTZAN dataset to test the accuracy between two common algorithms, KNN and CNN, used in automatic music genre classification. The result is that KNN performs better. Therefore, sometimes simple methods can be more effective compared with difficult methods.

Music always plays an important part in people's daily lives, and using machine learning to classify music genres automatically can be important to change the way people appreciate music. Also, with the great improvement nowadays in machine learning and music databases, music genres can be more and more advanced in the future.

The study underscored the potential of machine learning algorithms in music genre classification, with KNN showing promising results. This proves that the modern music industry can use KNN to build an auto music genre classification application. However, it also highlighted the need for more nuanced approaches to address the inherent complexity and subjectivity in music genres.

REFERENCES

P. M. H. Steve, What is a Music Genre? Definition, Explanation & Examples, 2023 available at https://promusicianhub.com/what-is-music-genre/

Q. He, Mathematical Problems in Engineering, (2022).

Andrada, GTZAN Dataset - Music Genre Classification, 2020, https://www.kaggle.com/andradaolteanu/ gtzan-dataset-music-genre-classification

Data Flair, Python Project – Music Genre Classification, 2020, available at https://data-flair.training/blogs/ python-project-music-genre-classification/

B. Logan, Ismir. 270 (1): 11, (2000).

T. Bäckström, Introduction to Speech Processing. Aalto University, (2024).

Y. Liu, G. Liu, D. Wei, et al, Frontiers in Psychology, 9, 2118, (2018).

M. Sköld, "The Visual Representation of Timbre," Organised Sound, vol. 27, no. 3, pp. 387–400, 2022.

A. Ghildiyal, K. Singh, S. Sharma, "Music genre classification using machine learning". In *2020 4th*

International Conference on Electronics, Communication and Aerospace Technology (ICECA), (2020), pp. 1368-1372.

D. S. Lau, R. Ajoodha, “Music genre classification: A comparative study between deep learning and traditional machine learning approaches”. In Proceedings of *Sixth International Congress on Information and Communication Technology: ICICT 2021* Volume 4 (Springer, Singapore, 2022), pp. 239-24.