Qeios

SONGS CLASSIFICATION PROBLEM RESEARCH BY GENRE BASED ON NEURAL NETWORK

Roman Lynnyk[a], Victoria Vysotska[b]
Lviv Polytechnic National University, Lviv, Ukraine

[a] ORCID iD: 0009-0007-0948-4338

[b] ORCID iD: 0000-0001-6417-3689

Summary. This article explores the use of neural networks for classifying musical compositions by genre based on their audio characteristics. With the increasing availability of large musical datasets and advancements in computational technologies, this topic has gained significant relevance. The article emphasizes the importance of selecting and processing audio features, particularly focusing on Mel-Frequency Cepstral Coefficients (MFCC), which reflect the key sound characteristics essential for genre identification. The approach to creating a neural network capable of effectively processing these characteristics and classifying songs is examined within the research. It outlines the main challenges associated with classification accuracy, error calculation, and the need to adapt to rapidly changing musical trends. This article also reviews and constructs a test neural network for classifying songs into three genres, creates diagrams to depict classification accuracy, and shows the distribution of each class by frequency characteristics. Additionally, it analyzes recent research in this field and highlights the importance of using MFCC. The article illuminates not only the technical aspects of creating and training neural networks for song classification but also the significance of such systems in the context of the widespread availability of musical content. It underscores the role of these systems in enhancing the quality of recommendation services, providing a more comprehensive musical experience to users, and supporting artists and musical styles. In conclusion, the article contributes valuable insights into the potential of neural networks in song genre classification, pointing to future research and development directions in this area.

Keywords: Neural Networks, Song Classification, Audio Features, MFCC (Mel-Frequency Cepstral Coefficients), Genre, Machine Learning, Audio Data Analysis

Introduction

In today's world, where digital technology penetrates every aspect of our lives, music is no exception. Today, we have access to millions of tracks thanks to streaming platforms such as Spotify, Apple Music, Deezer, and others. This vast amount of musical content creates a challenge for users in selecting songs that match their musical tastes. This, in turn, highlights the need for automated systems for classifying songs by genre.

Classifying songs by genre is not only an academic question but also a practical task that directly impacts user experience. Proper classification can provide users with relevant music recommendations, enhancing their satisfaction with listening to music and increasing the time spent on the platform.

On the other hand, incorrect or insufficiently accurate classification can drive users away from the platform as they do not receive the appropriate content. This is especially relevant for new artists or less popular genres, which may go unnoticed due to incorrect classification.

It is also essential to consider the globalization of the music industry. Artists from different countries and cultures now have the opportunity to share their music with a global audience. This expands the boundaries of genres, creating hybrids and new styles. Traditional classification methods may be ineffective for such diverse content, thus there is a need for more flexible and adaptive systems capable of accounting for these nuances.

https://doi.org/10.32388/59LGK6

Machine learning and neural networks have proven to be extremely useful in solving this problem. They can learn and adapt to new musical trends, identifying the characteristics of songs based on their audio features and other data sources.

In summary, the relevance of the topic of classifying songs by genre in the modern world is undeniable. With the growth of the digital music industry, the need for automated, efficient, and accurate classification systems continues to increase, making this research important and relevant for a broad audience.

Formulation of the problem

In the field of audio data processing, one of the key tasks is the classification of songs by genre. The challenge lies in developing a system capable of automatically determining the genre of a song based on its audio recording. This involves analyzing the audio signal, extracting relevant features, and applying machine learning algorithms for accurate classification. When addressing this problem, the following stages can be identified: audio signal representation, feature extraction, and classification using a neural network such as described below.

1. The audio signal can be represented as a time function $S(t)$.
2. For digital processing, the audio signal is converted into a discrete form: $S[n] = S(nT_s)$, where T_s is the sampling interval.
3. The use of Mel-Frequency Cepstral Coefficients (MFCC), which can be defined by the formula: $C_i = \sum_{n=1}^{N} \log(S_n) * \cos\left[i(n-0.5)\frac{\pi}{N}\right], i = 1, \dots, M$, where M is is the number of coefficients, N is the number of filters in the Mel scale and S_n are the spectral components of the signal.
4. The final stage, after data preparation, involves the challenge of building a neural network, presenting the classifier, and calculating the error.

It is also necessary to consider the wide variety of genres, as they may have similar audio characteristics, complicating the classification process.

Therefore, the development of such a system has significant practical and theoretical importance, enabling the automation of music classification processes, improving recommendation algorithms on music platforms, and contributing to scientific research in the field of audio data processing.

Analysis of Recent Research and Publications

In the contemporary scientific environment, there is a substantial amount of research dedicated to the analysis of audio data and the classification of musical compositions. One of the most important aspects of this field is the identification of a song's genre, which has significant applications in recommendation systems, streaming audio services, and music libraries.

Most research focuses on using audio features such as MFCC. Tzanetakis and Cook (2002) were among the pioneers in using MFCC for classifying music tracks by genre, demonstrating the effectiveness of this approach [1].

With the development of deep learning, many researchers, including LeCun et al. (2015), have turned their attention to neural networks as a means of processing audio signals. Their ability to detect complex patterns in data makes them particularly effective for music analysis [3].

Librosa, mentioned by McFee et al. (2015), has become a key tool in audio analysis, providing researchers with means for efficient feature extraction [4].

All these studies emphasize the combination of theoretical knowledge and practical methods to achieve high classification accuracy and understanding of the features of musical compositions.

Analysis of Known Software Systems

The analysis of software and hardware systems for classifying songs by genre reveals several notable solutions:

1. **Shazam**: A well-known system for song recognition, Shazam uses its proprietary algorithm for audio fingerprinting to identify songs in real time. Although its primary function is song recognition, it can also determine the genre of a song based on its audio characteristics.
2. **Spotify**: Spotify employs advanced machine-learning techniques for song recommendations to users. Their system analyzes genres, tempo, harmony, and other song characteristics to classify and recommend music.
3. **Musixmatch**: This platform, primarily known for recognizing song lyrics, also determines the genre of a song based on its lyrics and audio features.
4. **Echonest** (now part of Spotify): Echonest is a music service API that uses advanced analysis algorithms to determine the tempo, energy, danceability, and genre of songs.
5. **Moodagent**: This system identifies the mood and nuances of musical tracks using their audio characteristics. It can aid in genre classification based on the mood of the tracks.

Most of these systems use a combination of audio features (such as MFCC, chroma, tempo, etc.) and machine learning algorithms to classify songs by genre. They may also utilize other sources of information, such as song lyrics, metadata, or user ratings, to enhance the accuracy of their recommendations and classifications.

Formulation of the article goal

The goal of this research is to develop and validate an effective method for classifying songs by genre based on their audio characteristics using neural networks. This study aims to contribute to the development of automated recommendation systems and improve the user experience on music platforms.

The object of the study is musical compositions, their audio characteristics such as MFCC, and the processes involved in processing this data. Specifically, the focus is on examining how different audio features can be used for accurate genre classification of songs and how modern machine learning methods can optimize this process.

For classifying songs, we need the features of the songs and their corresponding genres. We can consider features such as tempo, loudness, duration, etc. Since we will use random data, we will generate it for the example.

Main idea and algorithm

Description of the Neural Network Operation

Neural networks are a powerful tool for solving classification tasks. In the context of our task, we use them to classify songs based on their audio characteristics. A neural network consists of inputs, weights, biases, and activation functions. During the training process, weights and biases are adjusted to minimize prediction error.

The main idea is to feed input data into the network, compute the output (prediction), and compare it with the actual value. Based on this comparison, the error is calculated, which is then used to adjust the weights and biases using the backpropagation algorithm.

MFCC (Mel-Frequency Cepstral Coefficients) are key features used as input to the neural network. They reflect the spectral characteristics of sound and are very useful for analyzing musical signals [2]. Training a model on such data can achieve high accuracy in classifying songs by genre.

For the classification task, we used a model with one input layer, one hidden layer, and one output layer. The activation function "relu" was used in the hidden layer, while "softmax" was applied in the output layer to obtain classification probabilities.

As a result, we obtained a model capable of classifying songs by genre based on their audio characteristics with high accuracy.

Therefore, after conducting the research, the following algorithm was used:

1. Initialization of the neural network's weights and biases.
2. Feeding input data into the network and calculating the output using the activation function.

3. Calculating the error between the actual and predicted labels.
4. Applying the backpropagation algorithm to adjust the weights and biases.
5. Repeating steps 2-4 for a certain number of epochs or until the desired accuracy is achieved.

Formulas used [4]:

1. The output of the neuron:

$$y = f(\sum_i w_i x_i + b) \tag{1.1}$$

2. Error calculation:

$$E = \frac{1}{2}\sum_i (target_i - output_i)^2 \tag{1.2}$$

3. Backpropagation for Weights and Biases

$$\Delta w = -\eta \frac{\partial E}{\partial w} \tag{1.3}$$

$$\Delta b = -\eta \frac{\partial E}{\partial b} \tag{1.4}$$

where η is the learning rate.

For testing and training the network, there was prepared data representing songs with genres of rock, pop, and jazz that were used. A diagram based on tempo and loudness metrics was constructed to visualize the results.

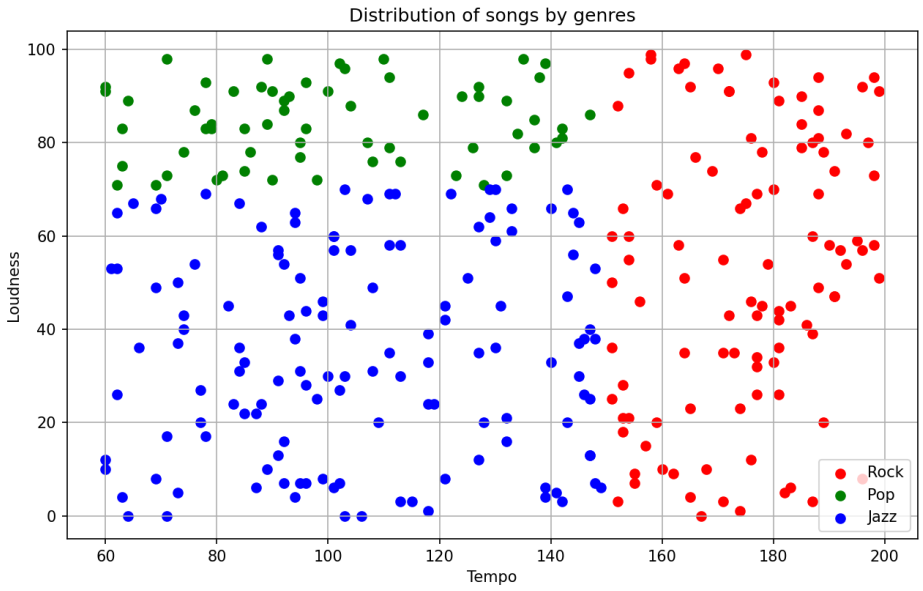

Fig. 1. Distribution of Songs by Genre

A neural network was created using these data for training the classification of songs. The predicted genres were visualized and compared with the actual genres in Figure 2.

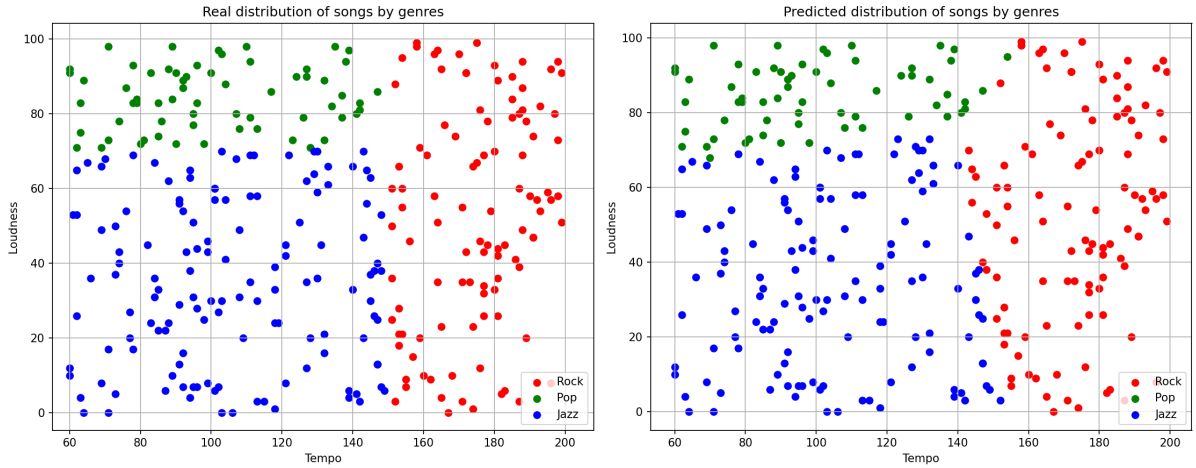

Fig. 2. Visualization of Song Classification Results

The displayed diagrams show the distribution of songs by genre: the actual distribution on the left and the predicted distribution by the neural network on the right. From this, we can conclude that the neural network classifies songs by genre with a high degree of accuracy.

Also, it would be important to mention that an error accuracy of approximately 94% in Figure 3. Sure it is not the most accurate one but pretty good to start with.

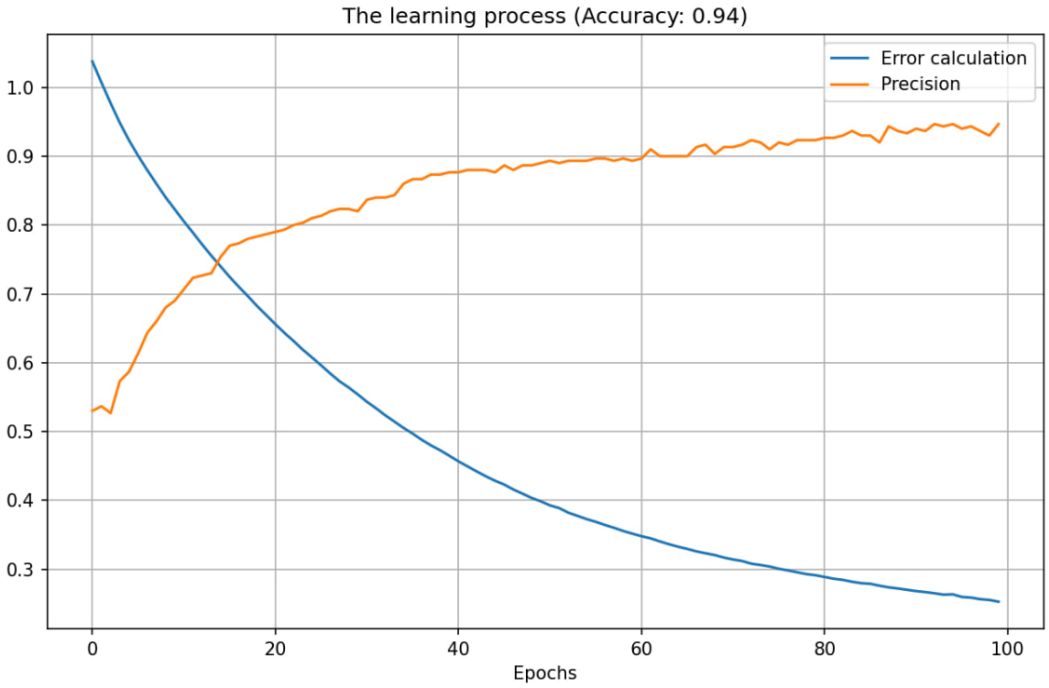

Fig. 3. The learning process with provided accuracy

Conclusions

The study explored the possibility of using neural networks for classifying songs by genre based on their audio characteristics, specifically MFCC. It was found that this approach allows for fairly accurate differentiation of musical compositions by genre, opening up broad possibilities for its implementation in recommendation systems on music platforms and other applications. It is important to note that the quality of classification significantly depends on the quantity and quality of input data, as well as the architecture and parameters of the neural network used.

Optimal selection of these parameters can greatly improve classification results. Overall, the results of the study indicate the great potential of using neural networks in audio data processing and analysis tasks. Machine learning-based methods can significantly contribute to the development of the music industry, making it more automated and accurate in meeting user needs.

References

1. Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293-302.
2. Cho, T., & Bello, J. P. (2011). On the relative importance of individual components of the Mel-frequency cepstral coefficients for automatic speech recognition. Proceedings of Interspeech, 2820-2823.
3. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
4. McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and music signal analysis in Python. Proceedings of the 14th Python in Science Conference, 18-24.