Proceedings of the Conference on Computational Engineering and Science JSCES
Vol. 29, Merch. 2024

Development of a system for classifying Japanese and British music genres using music2vec

Nanase Kishi[1)] , Ryuji Shioya[2)] and Yasushi Nakabayashi[3)]

1) Graduated Student, Toyo University (2100 Kujirai Kawagoe Saiamta 3508585 Japan, s3b102300052@toyo.jp)
2) Dr. Eng. Professor, Toyo University (shioya@toyo.jp)
3) Dr. Eng. Professor, Toyo University (nakabayashi@toyo.jp)

This research aims to use AI and machine learning to classify music genres and assist artists. By objectively analyzing their own music, artists can better understand the characteristics of their music and incorporate popular trends. They can also evaluate their own work, set goals, and drive continuous improvement. In this work, we use Music2vec as a method to convert music into vector representations for classification. The study plans to classify music by genre, country of origin of the artist, and songs that are popular in each country. Experiments using Music2vec and the GTZAN dataset achieved a classification accuracy of 62%, demonstrating the feasibility of classifying music genres. However, further research is needed to explore the subdivided genre taxonomy and the impact of the artist's country of origin on the taxonomy.

Key Words : *Music2vec, FMA, Mel Spectrogram,Soundnet, GTZAN*

1. INTRODUCTION

(1) Background

In recent years, there has been remarkable progress in AI, with increasing opportunities for its application becoming commonplace. Examples include ChatGPT by OpenAI [1] and Stable Diffusion by Stability AI [2]. ChatGPT is capable of generating text, while Stability AI can produce images and artistic works. Moreover, various products utilizing AI for music have emerged. For instance, Ozone 11 by Izotope [3] features AI-driven suggestions for mastering tasks. Additionally, Mubert by Mubert [4] offers a tool for automatic music generation from text. Thus, AI is believed to provide diverse approaches from a third-party perspective in supporting creative works.

However, tools for music analysis using AI have yet to be widely adopted. Analyzing one's own compositions with AI to determine differences compared to currently popular songs could provide valuable support to artists.

(2) Objectives

The objective of this research is to leverage AI and machine learning for music classification to provide support for artists. Music genres are defined by humans and categorized based on their distinct characteristics [5]. Music classification by AI offers artists an objective perspective on their own compositions, facilitating the incorporation of features and characteristics from popular songs into their own work. Understanding the differences between composed songs and popular music enables artists to evaluate themselves, set improvement goals, and facilitate continuous growth. Further subdivision of categorized music genres allows for distinguishing the specific characteristics of each song.

2. PREVIOUS STUDIES

(1) Music2vec [6]

One of the methods for music classification using AI is Music2vec. Music2vec, developed by Rajat Hebbar, converts music into vector representations. This technique applies features from text-based information retrieval models. The dataset utilized is the Free Music Archive (FMA), consisting of a large collection of 106,574 tracks from 16,341 artists and 14,854 albums, organized into a hierarchical classification of 161 genres. Mel spectrograms are used as input information. While spectrograms are commonly used for visualizing audio in a 2D representation of time and frequency axes, mel spectrograms are employed for faster computational processing. Figure 1 depicts a mel spectrogram of electronic music used in the study. By converting this mel spectrogram into vector representations and conducting learning, the test set achieved an accuracy of 43%.

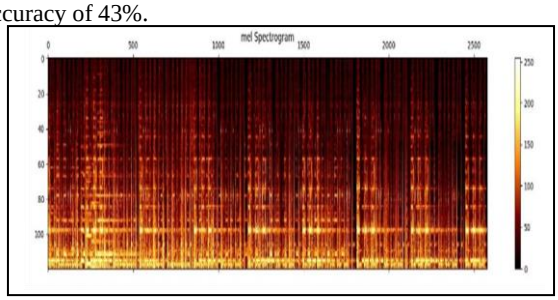

Fig. 1 Mel Spectrogram of Electronic Music

(2) Soundnet

Soundnet is a deep learning model for audio recognition in machine learning [7]. It utilizes Convolutional Neural Networks (CNNs), a type of deep learning, to learn features of audio. According to research, Soundnet learns from large amounts of unlabeled video data and performs classification of acoustic scenes and objects. Video data contains both visual and audio information, and Soundnet learns audio features using visual recognition models. Figure 2 displays labels recognized by AI from audio, including objects and landscapes, as reported in the study.

Fig. 2 Image Displaying Labels Recognized by Soundnet for Objects and Landscapes from Audio

(3) CRNN (Convolutional Recurrent Neural Network)

CRNN combines CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network). It is used to model continuous data such as audio signals and word sequences. CNN, known as Convolutional Neural Network, is utilized for processing mel spectrogram images. Then, the music's time-series data is handled by RNN, a type of recurrent neural network. According to the research by Keunwoo Choi et al. [8] which utilized CRNN for music classification, it reported higher effectiveness in music tagging compared to three other CNN models.

3. RESEARCH METHODOLOGY

(1) Three Types of Classification

There are three types of classification to be conducted. Firstly, classification by genre, which solely categorizes music genres such as rock and jazz to examine the extent to which genres can be further subdivided. Even within rock, there exist closely related genres like metal and punk, which can be further differentiated. This aims to investigate whether AI can classify genres that are difficult even for humans to distinguish. Secondly, classification by country, which categorizes songs based on the nationality of the composing artist. This investigates whether classification is possible based on the artist's country of origin, even for music of the same genre. Lastly, classification of popular songs in each country. Although popular songs vary by country, this classification aims to train the model on the hit charts of each country to classify popular songs.

(2) Definition of Music Genres

Defining each music genre is necessary when preparing the dataset. However, to date, there is no universal definition of music genres. For example, the genre of rock generally emphasizes electric guitar sounds, and it may include instruments such as bass and drums in a band setup. Similarly, the genre of metal primarily features distorted electric guitars, bass, and drums, which are also present in rock. The difference between rock and metal is believed to lie in metal's emphasis on low-frequency sounds, although there isn't a clear distinction in low frequencies. Furthermore, there are rock songs that emphasize low frequencies as well. In many cases, music genres lack clear classification boundaries. However, by considering the following two items, it is possible to differentiate unclear music genres when defining music genres [9]. Firstly, classification based on performance forms and musical structures. For example, in classical music, there are symphonies and concertos, while in popular music, there are bands, solo performances, and instrumentals. Secondly, classification based on historical descriptions, such as the classification of classical music into Baroque, Classical, Romantic, and Contemporary periods. By considering these two items, it is possible to prepare a dataset for music genre classification.

4. EXPERIMENT

(1) Genre of Analyzed Songs

In this study, the analyzed songs focus on the genre of Hardcore, one of the subgenres of Electronic Dance Music (EDM). Hardcore emerged in the early 1990s in countries such as the Netherlands, Belgium, and Germany [10]. In Japan, unique derived genres like Jcore and JHardcore have been established, while in the UK, it is known as UKHardcore. The definition of Hardcore for the two countries is "Hardcore in general released by labels and track makers within each country" [11].

(2) Experimental Method

The experiment involves converting music into vector representations and conducting classification using Music2vec and Soundnet. Referring to the model by Rajat Hebbar and KMASAHIRO [12]. As input, 30-second music data and corresponding country vectors were loaded for training. The output was assigning country labels to data every 30 seconds. Python was used as the programming language. The file format was WAV (Waveform Audio File Format), with a playback time of 30 seconds and a sampling rate of 22050Hz. When preparing the model, a function to construct Soundnet using pre-trained weights was used.

(3) Training Data

Two types of input data were prepared: three songs and ten songs for each country. Since each song is from two different countries, there are a total of six songs and twenty songs, respectively. Each song was divided into 30-second segments to increase the amount of data. Two methods of segmentation are shown in Figure 3. One method divides the song into 30-second segments without striding. The other method strides within one song to divide it into 30-second segments, aiming to increase the training data. It is ensured that one piece of data does not span multiple songs.

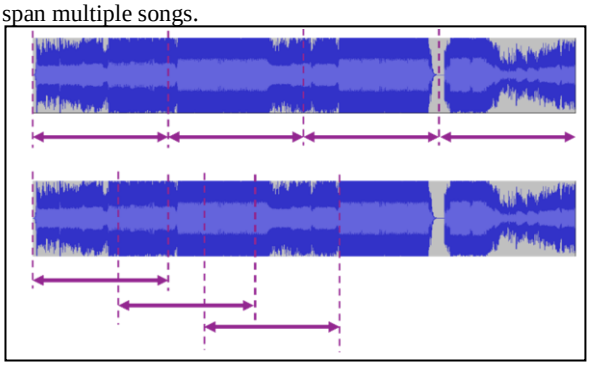

Fig. 3 Above: Segmentation without striding Below: Segmentation with striding.

(4) Test Data

Test data consisted of five Hardcore songs produced by Japanese artists and five produced by British artists. Each song was divided into 50 datasets, resulting in a total of 100 datasets. These 100 datasets were evaluated using Music2vec to determine the accuracy. Additionally, the test songs were not included in the training data.

(5) Results of Training Data with 40 Datasets (3 Songs)

Training was conducted with six songs divided into 40 datasets without striding, and the results of testing 100 datasets are shown in Table 1. The accuracy was 50%. As the predictions are column-wise, all 100 datasets were classified as Japanese Hardcore.

Table 1: Results of Training with 40 datasets (3 songs)

	Jcore	UKHardcore
Jcore	50	0
UKHardcore	50	0

(6) Results with 120 datasets (20 songs) for Training Data

Next, training was conducted on twenty songs divided into 120 datasets without striding, and the results of testing are shown in Table 2. The accuracy was 74%, which was higher than the accuracy achieved with six songs.

Table 2: Results of Training with 120 datasets (20 songs)

	Jcore	UKHardcore
Jcore	45	5
UKHardcore	21	29

(7) Results with 360 datasets (3 songs) for Training Data

Furthermore, training was conducted on six songs divided into 360 datasets with striding, and the results of testing are shown in Table 3. The test result was 78%.

Table 3: Results of Training with 360 datasets (3 songs)

	Jcore	UKHardcore
Jcore	40	10
UKHardcore	12	38

(8) Results with 360 datasets (20 songs) for Training Data

Similarly, training was conducted on twenty songs divided into 360 datasets, and the results of testing are shown in Table 4. The test resulted in 77% accuracy. All datasets without striding achieved higher accuracy than those without striding.

Table 4: Results of Training with 360 datasets (20 songs)

	Jcore	UKHardcore
Jcore	47	3
UKHardcore	20	30

5. CONSIDERATION

In this experiment, we observed differences in AI classification accuracy based on the quantity of training data. Increasing the number of input songs or expanding the volume of training data can lead to improved accuracy. Additionally, considering the maximum accuracy of 78%, it can be inferred that the AI has grasped the characteristics of Japanese and British Hardcore. Below, we discuss the differences between Japanese and British Hardcore concerning the key and BPM (Beats Per Minute) used in the training data, as summarized in Table 5, and illustrated in Figure 4.

According to Table 5, Japanese Hardcore has an average BPM of 186.7, with 4 out of 10 songs in major keys, indicating a faster tempo and brighter compositions compared to British Hardcore. Conversely, British Hardcore has an average BPM of 165.5, with 7 out of 10 songs in minor keys, suggesting a slower tempo and darker compositions compared to Japanese Hardcore. These differences are likely influenced by each country's market. Japanese Hardcore is primarily produced for Japan's music gaming market. This can be attributed to the establishment of Jcore in the early 2000s [13], coinciding with the late 1998 music gaming boom, which introduced Hardcore into the music

gaming market. Consequently, Jcore is expected to be sold at events related to music gaming, leading to an increase in compositions targeting music gaming enthusiasts. The higher average BPM in Japanese Hardcore may be due to setting higher difficulty levels for fingertip-based gaming rather than dance-based music typically played in clubs, compared to British Hardcore. On the other hand, British Hardcore targets primarily overseas club markets. This can be traced back to Hardcore Techno, the predecessor of UKHardcore, being used as EDM in clubs and festivals. As a result, UKHardcore is expected to be sold as dance music, leading to an increase in compositions used by DJs and performance artists. Considering these observations, the reason Music2vec demonstrated high accuracy may be attributed to considering differences in BPM and key signatures of the compositions.

Table 5: Key and BPM

BPM	Jcore	UKHardcore
Average	186.7	165.5
Minimum	165	153
Maximum	222	170
Major Key	4tracks	3tracks
Minor Key	6tracks	7tracks

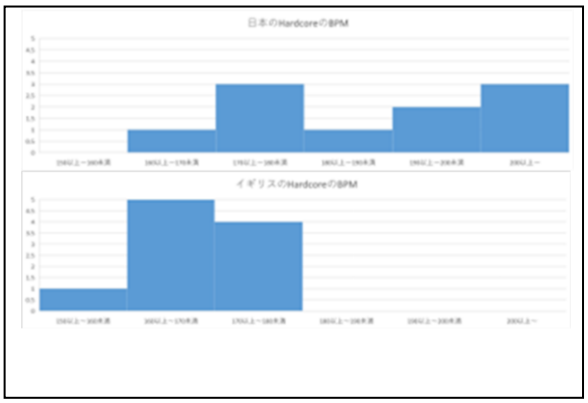

Fig. 4 Top: BPM of Japanese Hardcore, Bottom: BPM of British Hardcore

6. CONCLUSION

In this experiment, we utilized Music2vec for music classification. The results showed that when training with 40 datasets without sliding 6 songs, the accuracy was 50%. When training with 120 datasets without sliding 20 songs, the accuracy improved to 74%. Training with 360 datasets by sliding 6 songs resulted in an accuracy of 78%, and training with 360 datasets by sliding 20 songs yielded an accuracy of 77%. These findings indicate that increasing the original number of songs in the training data and augmenting the data by splitting songs contribute to improving the accuracy of Music2vec.

Furthermore, the high accuracy can be attributed to the characteristics of the Japanese and British music markets. The Japanese Hardcore market includes compositions targeted at music gaming enthusiasts due to the prevalence of music gaming-related compositions and events, while the British Hardcore market targets events such as clubs and festivals, owing to the history of Hardcore Techno being used in DJ-centric contexts.

Future research directions include standardizing the data extracted from the songs used for training and testing. Currently, each data point is being classified based on the country due to splitting each song into 30-second segments, necessitating a mechanism to assign a single label per song. Additionally, classifying songs based on their popularity presents another research avenue, aiming to clarify the preferences of Japanese and British music consumers. Metrics such as YouTube views can be used for popularity-based classification, aiding in predicting the popularity of songs.

The potential of the system developed in this study lies in its ability to provide artists with a multifaceted perspective by classifying their compositions using AI, thereby contributing to the demand for internationalized content.

REFERENCES

[1] OpenAI,ChatGPT,2024. https://openai.com/
[2] Stable AI ,Stable Diffusion,2024. https://ja.stability.ai/stable-diffusion
[3] Izotope,Ozone11,2024. https://www.izotope.jp/jp/products/ozone-11/#
[4] Mubert,Mubert,2024. https://mubert.com/
[5] Chillara, Snigdha, et al. "Music genre classification using machine learning algorithms: a comparison." Int Res J Eng Technol 6.5 (2019): 851-858.
[6] Ashish Bharadwaj Srinivasa, Rajat Hebbar “music2vec: Generating Vector Embeddings for Genre-Classification Task”(2017), https://medium.com/@rajatheb/music2vec-generating-vector-embedding-for-genre-classification-task-411187a20820
[7] Aytar, Yusuf, Carl Vondrick, and Antonio Torralba. "Soundnet: Learning sound representations from unlabeled video." Advances in neural information processing systems 29 (2016).
[8] Choi, Keunwoo, et al. "Convolutional recurrent neural networks for music classification." 2017 IEEE International conference on acoustics, speech and signal processing (ICASSP). IEEE, 2017.
[9] What is the genre of music?,JUN,2021. https://note.com/junjunjunpiano/n/nc57d89fed91f
[10] Aggressive Dance Music. Masterpieces of Hardcore Techno, RAG MUSIC Editorial Department,2023. https://techno.studiorag.com/hardcore-techno-songs?disp=more
[11] Ny4nne's Thoughts on J-CORE Genre Definitions, Ny4nne, 2022. https://note.com/nakisunachan/n/n4d2e9341e469
[12] Music Genre Classification Model Using music2vec, KMASAHIRO,2021.https://qiita.com/KMASAHIRO/items/cae4dfb0657eec4a2dca
[13] J-CORE Cultural Revolution, Rough Sketch,2007. https://data.technorch.com/data/page/gbn93.html