

# Opera Clustering: K-means on librettos datasets

정 하 림<sup>1\*</sup>      유 주 헌<sup>2</sup>  
Harim Jeong      Joo Hun Yoo

## ABSTRACT

With the development of artificial intelligence analysis methods, especially machine learning, various fields are widely expanding their application ranges. However, in the case of classical music, there still remain some difficulties in applying machine learning techniques. Genre classification or music recommendation systems generated by deep learning algorithms are actively used in general music, but not in classical music. In this paper, we attempted to classify opera among classical music. To this end, an experiment was conducted to determine which criteria are most suitable among, composer, period of composition, and emotional atmosphere, which are the basic features of music. To generate emotional labels, we adopted zero-shot classification with four basic emotions, 'happiness', 'sadness', 'anger', and 'fear.' After embedding the opera libretto with the doc2vec processing model, the optimal number of clusters is computed based on the result of the elbow method. Decided four centroids are then adopted in k-means clustering to classify unsupervised libretto datasets. We were able to get optimized clustering based on the result of adjusted rand index scores. With these results, we compared them with notated variables of music. As a result, it was confirmed that the four clusterings calculated by machine after training were most similar to the grouping result by period. Additionally, we were able to verify that the emotional similarity between composer and period did not appear significantly. At the end of the study, by knowing the period is the right criteria, we hope that it makes easier for music listeners to find music that suits their tastes.

☞ keyword : Music Analysis, Music Information Retrieval, Natural Language Processing, Embedding, Classification, K-means Clustering

## 1. Introduction

Machine learning has made tremendous progress and is being applied in a variety of fields. Especially in the area of machine learning for music, researchers continued to expand the area with Music Information Retrieval (MIR) and music genre classification. This kind of music classification algorithm is utilized in a music recommendation system and contributes to more effectively finding music that users want [1]. However, there is a huddle to apply those machine learning techniques on classical music.

In order to generate an optimal machine learning algorithm, the proper amount of dataset and the following significant data variables are required. Generally, music contains composer, published year, audio signal, genre, and

lyrics as primary data. However, in the case of classical music, there is no specific genre such as rock or ballade and not many classical music have lyrics. Also, the audio signal of classical music is incredibly long and hard to compress due to the complexity. If we exclude those data, only composer and published year left as a variable of classical music.

Therefore, in this study, we aim to classify classical music with more data for better accuracy. We also determine what criteria are the most appropriate to classify classical music. To include various data as much as possible, we narrowed classical music to opera that contains text data called *librettos* in musical terms. Also, regarding there is no genre for classical music on opera, we obtain emotional data as part of genre referring to the method from [2]. With these generated data variables, we will examine to find the proper criteria for classifying classical music through the clustering method by vectorizing the text data of operas and comparing it to notated variables of music.

According to a review by GW Milligan, clustering, the wide range of unsupervised classification, is a suitable

1 Dept. of Interaction Science, Sungkyunkwan University, Seoul, 03063, Korea.

2 Dept. of Artificial Intelligence, Sungkyunkwan University, Suwon, 16419, Korea.

\* Corresponding author (harim4110@gmail.com)

[Received 08 October 2021, Reviewed 01 December 2021, Accepted 05 February 2022]

method for exploratory factor analysis [3]. Thus, for classical music, which is challenging to apply confirmatory factor analysis with clear indicators for classification, we decided to analyze the characteristics of each clustered group. Getting clustering by measuring the vector distance between data points, we compare each clustered label with other data variables that the music initially has. Through this process, we expect to be able to suggest the right method to classify operas.

We can contemplate four main factors to classify classical music. The ‘composer’ who composed the music, the ‘period’ in which the music was composed, the ‘instrumentation’ that music includes, and the ‘emotional atmosphere’ of the music. Since we considered only operas in the study, the instrumentation will be all similar. Accordingly, excluding the instrumentation factor, we will compare our clustering results with composer, period, and emotions.

It is a challenge to define the emotion or atmosphere of the music. However, one of the previous studies related to our study, *Study on sentiment analysis for opera* proposed a method of sentiment analysis of the opera by computational method with *librettos* [4]. The study suggested labeling the emotional scores for each opera through the zero-shot classification. We apply the same method to get the emotion variables of the opera. Nevertheless, different hyperparameters on preprocessing and embedding model generating steps are added for better and more accurate results.

## 2. Related Works

### 2.1 Music Classification

Machine learning methods have recently been applied for the music classification task. With extracted features from signal characteristics and frequencies of general music, it was possible to perform genre classification analysis [2, 5, 6] or recommendation algorithms [7, 8]. Related studies have been conducted in two main directions: an analysis of the domain value of music signal data, and an analysis method of spectrogram image data.

In the field of music analysis, time domain values and

space domain values can be extracted from the waveform of the audio signal, which is considered one of the features representing the music and used for classification model generation. Carlos proposed an automatic music genre classification task on the Latin music dataset in [2]. By applying time and space decomposition to machine learning methods, it was possible to classify 3,160 music pieces into 10 music genres.

There are also some other researches that apply deep neural network algorithms to music genre classification tasks. For example, Bahuleyan applied the Convolutional Neural Network model on an ensemble dataset in [9]. He selected Audioset data, which contains more than two million sound clips, to classify into seven music genres. A pre-trained VGG16 model was mainly used to train the significant features of extracted MEL spectrogram images. To improve the performance, the gradient boosting algorithm is ensemble with the VGG-16 model, and finally achieved about 65% of test accuracy. These studies applied whether extracted time and space domain decomposition features or spectrogram images from music data. However, classification on one extracted musical feature did not show enough performance.

To this end, Costa applied the ensemble method on both hand-crafted features and spectrogram images extracted from Western music collections, Latin American music, and ethnic African music datasets [10]. Support Vector Machine classifier was mainly selected for the hand-crafted feature data and Convolutional Neural Network on sliced spectrogram image data. Each classifier shows high performance on music genre classification, but the accuracy outperforms when the ensemble method is applied on both classifier results.

From the above studies, general music can be classified with time domain, space domain, and spectrogram features from the music dataset. However, in the case of classical music, it is hardly possible to apply these algorithms, due to the length and dynamic characteristics of classical music.

### 2.2 Classical Music Classification

Even with these hurdles, there have been several attempts to classify classical music. For example, in the study by

Weiss, the author classifies classical music based on tonal complexity [11]. Machine learning algorithms, especially Support Vector Machine (SVM) and Gaussian Mixture Model (GMM) were applied to audio data, recording by piano or orchestra. They showed that the general classification method can also be applied to classical music. However, when complex music and standard music were combined, the accuracy was 67% in orchestra music. Therefore, it is difficult to classify the various types of classical music at one with audio format data.

In another approach, Jiang attempted to classify classical music using spectral contrast in [12]. The author collected the first 10s clips from the music and analyze them with Mel-Frequency Cepstral Coefficients (MFCC) that are widely used in audio classification. This study also can be an example of applying the classification method to classical music. However, in the research, they include Baroque, Romantic, into classical music and other music were pop and rock which are clearly different. If all different kinds of music are mixed it may be possible to classify classical music from other types of music. However, it is hard to conclude that using audio data for the analysis is also suitable to classify only in classical music.

Regarding these related works, we added a newly generated emotional data variable for the analysis. Furthermore, with clustered classical music results, we will find which criteria are acceptable.

## 3. Methodology

### 3.1 Datasets

For the analysis, we attempted to utilize as many diverse opera libretto datasets as possible. We referred to *The book of 101 opera librettos* and collected a total of 80 librettos [13]. Few copyright-restricted operas were excluded from the data collection. The dataset covers operas from all different periods, Baroque, Classical, Romantic, and Contemporary, by thirty-four different composer. All librettos were translated if it is not originally written in English, and scraped from [www.opera-arias.com](http://www.opera-arias.com).

## 3.2 Classification

### 3.2.1 Data Preprocessing

As in all machine learning tasks, how to preprocess data in Natural Language Processing (NLP) tasks is directly related to the accuracy of the learning model. In this study, a total of 80 opera librettos were grouped into one paragraph, preprocessed, and then the data structure was changed to have one row for each libretto.

In the preprocessing model, simple processes such as punctuation removal, stopwords removal, lower casing, and tokenization were added for each libretto to reduce data size. In other words, label generation by zero-shot classification was applied with meaningful words left to determine the scores for the four candidate emotion labels that we aim for.

### 3.2.2 Libretto Classification

Zero-shot classification [14] is one of the representative unsupervised Natural Language Inference (NLI) analysis methods. When users set the list of candidate labels that they want to know in advance through the input data script, it analyzes data features' word frequency and bag-of-words to generate the score results for given candidate labels.

We adopted this on our opera libretto classification to get labeled variable because the genre cannot be distinguished. With these limitations, pre-trained zero-shot classification models from the transformer pipeline are recommended.

In other words, the zero-shot classification method was set for learning opera librettos and outputting the scores for four emotions of 'happiness', 'sadness', 'fear', 'anger,' The Big-Six emotions, 'happiness', 'sadness', 'fear', 'anger', 'disgust', and 'surprise,' were most widely accepted. On the other hand, recent studies of human emotion recognition narrowed six emotions into four emotions that can more clearly explain human expressions [15]. By referring to these facts, we applied four emotions instead of Big-six emotions.

## 3.3 Clustering

### 3.3.1 Data Preprocessing

Since the clustering method calculates the distance between each data feature, the step of embedding libretto

configured in word format is essential. Each opera libretto consists of about 7,000 words. After applying the same preprocess from the classification, document embedding is required to perform clustering which is based on the measuring distance between extracted data features. Understanding our libretto dataset, the process that converts whole text script into meaning vector format is needed, we choose doc2vec model. Verifying the probability distribution of the words, the doc2vec embedding technique [16] calculates the vector distances between each document. To extract the representative features of each data, therefore, we adopted a doc2vec embedding technique that converts documents into a vector format.

### 3.3.2 Libretto Clustering

The first thing we considered before the clustering method is dimensionality reduction. Each opera libretto was converted into an assigned maximum vector size in the embedding model generation step, but its dimension is still too large to calculate and visualize the similarities between each data point. In addition, high-dimensional input data normally causes the curse of dimensionality issues, which seriously degrades machine learning performances. To solve the problem, the Principal Component Analysis (PCA) technique was utilized before k-means clustering in the first place. Under the condition of  $n$  selected centroids, k-means clustering is a proper unsupervised learning method to separate high-dimensional data. It divides the overall dataset into  $k$  clusters by minimizing the variance distance between them. We determine the optimal number of clusters with the results from the elbow method and silhouette method.

## 4. Experiment Results

### 4.1 Clustering

Attempting the doc2vec embedding model for our datasets, the distributed bag of words (PV-DBOW) method was basically adopted with hyperparameters including maximum vector size, minimum word count, and model training epochs. Hyperparameter selection was conducted by changing the maximum vector size to 100, 300, 500, 1,000,

and 1,500, respectively. The minimum word count, which removes that do not appear as much as the frequency of a certain threshold, was tested with values 1, 3, 5, and 10. We then train each model for 200 epochs.

Optimal hyperparameters for doc2vec embedding were checked by comparing clustered results with the labels from the classification approach and noted music variables. Since the most accurate clustering result was shown in the environment of 100 maximum vector sizes with 10 minimum word counts, they were designated as the right hyperparameters for the clustering method. Overall experiment was implemented with Pytorch 1.9.1., Scikit-Learn 1.0., on Intel Core i7-10700 processors with NVIDIA GeForce RTX 3070.

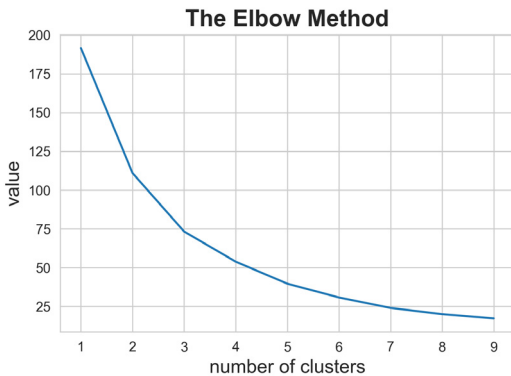
We then tried to reduce high-dimensional vector data into two or three dimensions, which are most commonly used to visualize the results of dimensional reduction. However, we are concerned about significant data feature loss from the dimensionality reduction process, which reduces 100 data variables into minimal dimensions. To keep enough data features, the dimensional reduction was also attempted in 10 dimensions. After that, we also tried to find the most suitable dimension to represent more than 95.0% variance of the whole dataset. Table 1 shows the result of cumulative variance explained by the following  $n$ -principal components.

(Table 1) Cumulative Variance with PCA

Reduced Dimension	Cumulative Variance
2	7.62%
3	10.49%
10	24.98%
66	95.74%

The cumulative variance result of 2, 3, and 10 dimension PCA is less than 30%, which is not efficiently representing our dataset features. On the other hand, to get a value of more than 95% should contain at least 66 dimensions of the data. Based on these results, to have a sufficient representation of data without feature loss, the preprocessed vector format from the doc2vec embedding step is directly

applied to the clustering analysis without dimensionality reduction. Although there is an issue with performance degradation caused by using high-dimensional data, it is more acceptable than the important feature loss from excessive dimensionality reduction.



(Figure 1) Elbow method for clustering

Checking the optimal number of clusters for the opera libretto used in the experiment using the elbow method in Figure 1, it was given that 4 clusters were recommended.

## 4.2 Analysis

After obtaining the most suitable four clusterings as a result, we explored the clustering by comparing them with other data variables from the music. ‘Fear’, the highest emotion in all librettos, we compared the operas that showed ‘fear’ as the highest emotional score by different periods. Consequently, we can summarize that the data with similar emotional moods do not belong to the same clustering as shown in Table 2.

By confirming that the results are not clustered with emotions, it was able to find that it is consistent with the periods. It can be also inferred that it is clustered by the composer. However, when we evaluate the similarity index with the composer label, it showed a lower score than the similarity with the period label. In addition, after reviewing the composer as a standard and linking the relationships, we found out that one composer does not necessarily compose music with a similar emotional atmosphere as in Table 3.

To calculate the accuracy with a computational evaluation method, an adjusted rand index score from scikit-learn is adopted. Comparing pair similarities between clustered target variables, the score provides a quantified accuracy value. The result shows the highest value with the period variable, while the lowest similarity result comes from the composer variable.

## 5. Conclusion

In this paper, we have shown that classical music, especially opera, can be clustered based on its periodic features. Since classical music has no evident genre or specific signal distinctions, analyzing opera libretto with an unsupervised k-means clustering was considered. By comparing the accuracy of predicted labels with similarities among composer, period, and emotional atmosphere of the opera, we conclude that the embedded libretto datasets contain significant features from the periodic backgrounds. About the method, doc2vec embedding model with 100 vector size, 10 minimum word count, and 200 training epochs were optimal hyperparameter settings for opera libretto datasets.

In fact, we were already taking into account the periodic of composition when classifying classical music. However, we believe that this study was worthwhile because it numerically confirmed that the period is the most appropriate among other classification criteria.

This study has limitations in that the emotional label was based on the pre-trained engine, zero-shot classification and 80 limited amount of libretto datasets. Also, the analysis was limited to only opera was covered, not entire classical music. However, we still strongly believe that it is valuable because opera is one of the most popular classical music. In this respect, it will be more meaningful to find the most suitable classification criteria for all classical music in future research. And finally, it is hoped that a music recommendation algorithm can be also efficiently developed for classical music.

(Table 2) Predicted Cluster values with same emotion

Composer_Opera	Emotion	Period	Predicted Clusters
Cimarosa_Il maestro di cappella	Fear	Baroque	C
Mozart_Don Giovanni	Fear	Classical	D
Mascagni_Iris	Fear	Romantic	A
Bellini_Norma	Fear	Romantic	A
Bizet_Carmen	Fear	Romantic	A
Borodin_Prince	Fear	Contemporary	B

(Table 3) Predicted Cluster values with same period variable

Composer_Opera	Emotion	Period	Predicted Clusters
Massenet_Cendrillon	Anger	Romantic	A
Massenet_Don Quichotte	Sadness	Romantic	A
Massenet_Herodiade	Anger	Romantic	A
Massenet_Sapho	Anger	Romantic	A
Massenet_Thais	Sadness	Romantic	A
Massenet_Werther	Fear	Romantic	A
Verdi_Aida	Fear	Romantic	A
Verdi_La forza del destino	Sadness	Romantic	A
Verdi_Don Carlos	Sadness	Romantic	A
Verdi_Ernani	Sadness	Romantic	A
Verdi_Falstaff	Anger	Romantic	A
Verdi_La Traviata	Happiness	Romantic	A
Verdi_Un ballo in maschera	Sadness	Romantic	A
Verdi_Luisa Miller	Fear	Romantic	D
Verdi_Otello	Happiness	Romantic	A
Verdi_Rigoletto	Happiness	Romantic	D
Verdi_Il trovatore	Sadness	Romantic	A

## References

- [ 1 ] Meyers O.C. "A Mood-Based Music Classification and Exploration System," Massachusetts Institute of Technology, 2007.  
<https://dspace.mit.edu/handle/1721.1/39337>
- [ 2 ] Silla Jr., Carlos N., Alessandro L. Koerich, and Celso A. Kaestner. "A Machine Learning Approach to Automatic Music Genre Classification," *Journal of the Brazilian Computer Society*, vol.14, no.3, pp.7-18, 2008.  
<https://doi.org/10.1590/s0104-65002008000300002>
- [ 3 ] Milligan "Clustering and Classification Methods." *Handbook of Psychology*, 2003.  
<https://doi.org/10.1002/0471264385.wei0207>.
- [ 4 ] H Jeong, "Study on sentiment analysis for Opera." in *Proc. of APIC-IST 2021*, 89-91, 2021.
- [ 5 ] Haggblade, Michael, Yang Hong, and Kenny Kao. "Music genre classification." *Department of Computer Science, Stanford University*, 2011.
- [ 6 ] Nanni, Loris, et al. "Combining visual and acoustic features for music genre classification." *Expert Systems with Applications*, Vol.45, pp.108-117, 2016.  
<https://doi.org/10.1016/j.eswa.2015.09.018>
- [ 7 ] Van Den Oord, Aäron, Sander Dieleman, and Benjamin Schrauwen. "Deep content-based music recommendation." *Neural Information Processing Systems Conference (NIPS 2013)*. Vol. 26. Neural Information Processing Systems Foundation (NIPS), 2013.
- [ 8 ] Wang, Xinxin, et al. "Exploration in interactive personalized music recommendation: a reinforcement learning approach." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol.11, No.1, pp.1-22, 2014.  
<https://doi.org/10.1145/2623372>
- [ 9 ] Bahuleyan, Hareesh. "Music genre classification using machine learning techniques." *arXiv preprint arXiv:1804.01149*, 2018.  
<https://arxiv.org/abs/1804.01149>
- [10] Costa, Yandre MG, Luiz S. Oliveira, and Carlos N. Silla Jr. "An evaluation of convolutional neural networks for music classification using spectrograms." *Applied soft computing*, vol.52, pp.28-38, 2017.  
<https://doi.org/10.1016/j.asoc.2016.12.024>
- [11] Weiss, Christof, and Meinard Müller. "Tonal complexity features for style classification of classical music." *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015.  
<https://doi.org/10.1109/ICASSP.2015.7178057>
- [12] Jiang, D. N., Lu, L., Zhang, H. J., Tao, J. H., and Cai, L. H. "Music type classification by spectral contrast feature." *IEEE International Conference on Multimedia and Expo*, pp.113-116, 2002.  
<https://doi.org/10.1109/ICME.2002.1035731>
- [13] MacMurray, Jessica M., and Allison B. F. "The Book of 101 Opera Librettos: Complete Original Language Texts with English Translations," New York: Black Dog & Leventhal Publishers, 1996.
- [14] Ye, Meng, and Yuhong G. "Zero-Shot Classification with Discriminative Semantic Representation Learning." *CVF Open Access*, January 1, 1970.  
<https://doi.org/10.1109/cvpr.2017.542>
- [15] Jack, Rachael E., Wei S, Ioannis D, Oliver G. Garrod, and Philippe G. Schyns. "Four Not Six: Revealing Culturally Common Facial Expressions of Emotion." *Journal of Experimental Psychology: General*, 145(6), pp.708-730, 2016.  
<https://doi.org/10.1037/xge0000162>
- [16] Le, Quoc, and Tomas M. "Distributed Representations of Sentences and Documents." *PMLR.*, June 18, 2014. <http://proceedings.mlr.press/v32/le14.html>

## ● Authors ●



### 정 하 립(Harim Jeong)

2017 Bachelor of Music, University of Cincinnati, Ohio, United States.

2019 Master of Music, New England Conservatory of Music, Massachusetts, United States.

2022 Master of Cultural Management, Sungkyunkwan University, Seoul, Korea.

2022~Present: Ph.D. student in Dept. of Interaction Science, Sungkyunkwan University, Seoul, Korea.

Research Interests: HCI, Computational Method, Music, UX, Machine Learning.

E-mail : harim4110@gmail.com



### 유 주 현(Joo Hun Yoo)

2020 B.S. in Applied Statistics, Chung-Ang University, Seoul, Korea.

2022 M.S. in Artificial Intelligence, Sungkyunkwan University, Suwon, Korea.

2022~Present: Ph.D. student in Dept. of Artificial Intelligence, Sungkyunkwan University, Suwon, Korea.

Research Interests: Machine Learning, Deep Learning, Federated Learning, Medical AI, Big Data Analysis.

E-mail : andrewyoo@g.skku.edu