Art and Performance Letters (2021)
ISSN 2523-5877 Vol. 2 Num. 5

DOI: 10.23977/artpl.2021.020509
Clausius Scientific Press, Canada

Research on the Evolution of Music Schools

Chengsi Wang[1], Jiusi Yu[2]

[1]*School of Statistics, Beijing Normal University, Beijing, 100875*

[2]*Department of Physics, Beijing Normal University, Beijing, 100875*

Keywords: Cluster analysis, Model for music similarity evaluation, Genre differences

Abstract: In order to understand the music more deeply, we need to develop a way to quantify the evolution of music. To solve the problems mentioned above, this paper classifies and weights the twelve features through cluster analysis, and establishes two model development parameters for judging music influence and music similarity to describe music influence. Reduce the five-dimensional parameters to the point that each artist was a five-dimensional vector. The artist style of a particular year is analyzed, and the measurement criteria are developed to define revolutionaries in two ways.

1. Introduction

Music, an important cultural heritage, has played a great role in the long history of mankind. The study of music and music genres, together with the analysis of its evolution are of great significance.

Musicto some extent, is the picture of life to the crowds. In order to measure the influence of music better, our team look at the relationship between the musicians and genres, using chordal graph to intuitively show the relationship between influencers and followers within the genre. We analyze the influence of the music on the base of the four characteristics we select as well as a new definition of new research conclusion.

2. Model Building

To simplify our model, we make some general assumptions. The assumptions together with corresponding arguments are as follows:1. Assuming that the data given by ICM is complete. 2. Each artist belongs to and only belongs to one genre, and the data of non-genre artists is not taken into consideration.

2.1 Data Preprocessing

Music genres being highly contagious as viruses, the influence of the same art work varies from person to person. Our existing data are made by musicians listing a dozen or more artists who have influenced their own music works, each of which is highly subjective in its analysis. So we add up the data, merge the influencers and followers of the same genre, and calculate the influence. Here we directly use the values given in the data to represent the quantified influence, that is:

$$influ_i = num_i \tag{1}$$

$influ_i$: The influence of the i-genre after quantification

num_i: The number of followers of the i-genre

By counting the genres of influencers and followers, we arrive at a table of influence between genres. We only extract data from the affected population of the same genre greater than 20 to simplify the calculation.

The parameters built above to describe *music influence* succinctly illustrates the proportion of people between genres. This parameter can also be used to quantify the influence of genres for subsequent analysis.

At the same time, using the **Principal Component Analysis (PCA)** and **cluster analysis** to classify these features, we extract four of the least relevant values: *duration_ms*, *speechiness*, *mode*, *key* and weight the remained seven values to produce a new variable, named *new*:

$$new_i = 0.001 * tempo_i + danceability_i + energy_i + valence_i + 0.01 * loudness_i + acousticness_i + instrumentalness_i \tag{2}$$

We only selected five eigenvalues for analysis, so all samples can be expressed as:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{15} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{n5} \end{bmatrix} \tag{3}$$

We create a measurement model of musical similarity based on SAW (Simple Additive Weighting) [1] which can be expressed as:

$$Sim_{ij} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}k_j \tag{4}$$

We assign a weight to each input value, denoted as k_j,j=1,2…n, where n is the total and x_{ij} is the input value of the *j-th* criterion of the *i-th* DMU(Desicion-making Unit). To eliminate the effect of dimensionality, we normalized each musical feature that is considered:

$$Sim_{ij} = \frac{1}{n}\sum_{j=1}^{n} x'_{ij}k_j \tag{5}$$

x'_{ij} is the normalized value of x_{ij}.

In order to intuitively see the difference of each eigenvalue among different genres, we calculate the average and variance of each column.we calculate the variance of the five eigenvalues of artists of the same genre, and eliminate the dimensional influence with the formula of $Sim_{i,j}$:

$$Sim_{i,j} = \frac{\sum_{0<m\leq n} a_{i,m} * a_{j,m}}{\sum_{0<m\leq n}(a_{i,m})^2 * \sum_{0<m\leq n}(a_{j,m})^2} \tag{6}$$

There are several indicators we need to look at before we compare the degree of dispersion within and between genres：

1. Comedy consists of types of entertainment in order to make people laugh. [2]

This nature of genre gives Comedy/Spoken a much higher value in *speechiness* than any other.

2. Due to the particularity of *mode* (concluding only two indexes: 0,1), there will be large variance between genres.

Excluding the above conditions, we can generally see that the degree of dispersion between schools is higher than within genres.

3. Solution Method for Model

3.1 Analysis Model of Genre Change over Time

12 eigenvalues calculation together being too complex, we use the evaluation model mentioned above to simplify the data with the five characteristics of genre list (delete the sample which has less samples), to make the radar map.

The speechiness of comedy/spoken is much higher than that of others. Therefore, in order to facilitate the analysis, we screen out the data of comedy/spoken.

In order to make a more accurate quantitative analysis of similarity, we use **cosine similarity** [3] to calculate. Given two genres {I, J}. The a_i, n represents the n-th feature of the genres. The similarity between the two genres is defined as follows:

$$Sim_{i,j} = \frac{\sum_{0<m\le n} a_{i,m} * a_{j,m}}{\sum_{0<m\le n} (a_{i,m})^2 * \sum_{0<m\le n} (a_{j,m})^2} \tag{7}$$

Using digital processing technology, the number of songs of each genre in every five years is counted and calculated, and a line chart is made. The change trend we need can be intuitively obtained from the figure, which will not be described here.

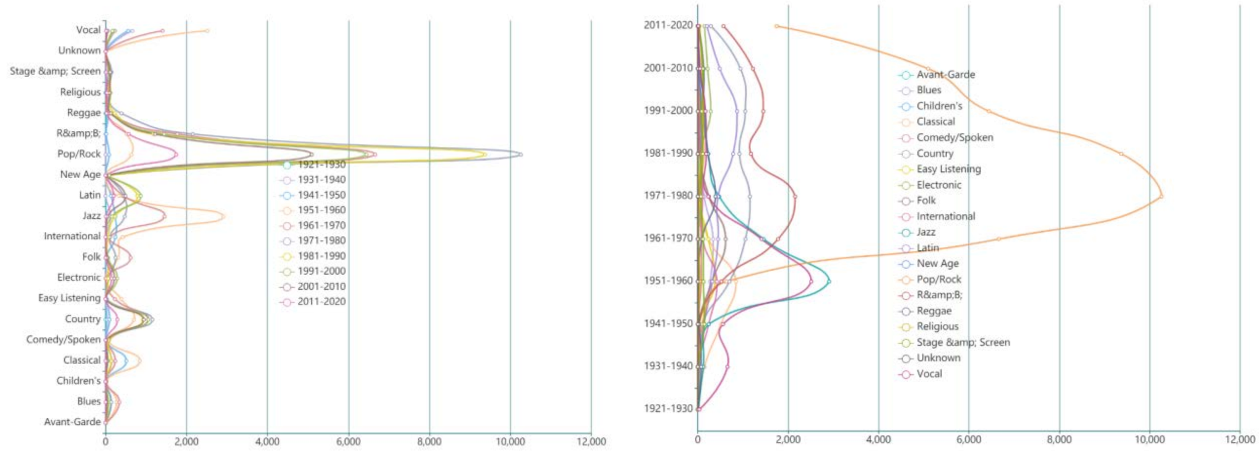

Figure 1: Genres-time-amount

3.2 Testing the Model

In this part, we will analyze the sensitivity of the parameter we set.

$$new_i = k_1 * tempo_i + k_3 * danceability_i + energy_i + valence_i + k_2 * loudness_i + acousticness_i + instrumentalness_i$$

k_1, k_2, k_3 respectively represents *tempo*, *loudness* and *dance-ability*. We enter different k_1, k_2 and k_3 into program, and get different *new*. Here is the k_1 result:

Table 1: Sensitivity analysis on k_1

k_1	0.001	0.0011	0.0012	0.0009	0.0008
new	2.259223	2.270893	2.282564	2.247552	2.235881
sensitivity			0.010331745		

The sensitivity index shown above means that when k_1 increase or decrease by 20%, the value of *new's* change is limited to 1.033%. So the variation of this index has minor influence on the result.These analysis prove its stability.

After data analysis, we can have a clearer observation of the genre effects, from which we can go a step further and make in-depth analysis of the traceability between music genres. It will also be a better starting point to explain the characteristics of the future analysis so as to make the explanation more convincing, and some problems that are not easy to be discovered by the public before will also be found.

4. Summary

Our model is Easy to use and can intuitively and easily eliminate the differences brought about by dimension. Can compare eigenvalues in different units and different directions. Weakened the defect of SAW method, that is, each condition must be positive or negative in the same direction. However our model still has some limitations: Inconvenient to deal with extreme values. Types of external influences cannot be clearly distinguished. There is some subjectivity in the definition of data sources.

We only have 12 eigenvalues during analysis, but in reality, the number of eigenvalues will increase along with the improvement of people's cognition. Cluster analysis method is used again to select classes. I values are selected for analysis according to the actual situation, and the data is appropriately weighted for easy analysis. We can draw the change of orientation and direction of each genre under the influence of human society or analyze specific works to get its genre. More information can be obtained by analyzing the characteristics of specific composers, such as the internal and external influences of composers. The more data items we have, the more categories we will select for cluster analysis, and the higher the accuracy of analysis.

Our solution mainly measures music characteristics by listing the five parameters selected. First, the data analysis shows the evolution direction of genres, and the mutations combined with the historical events can analyze mass psychology as well as culture. What's more, the analysis of us can also summarize the common points, similarities and differences between various genres.

References

[1] Zhou Benjin, Tao Yizheng, Ji Bin, et al. K-means initial clustering center optimization method to minimize the error square sum [J]. Computer Engineering and Applications, 2018, 54(15): 48-52.

[2] http://dict.youdao.com/w/comedy

[3] Wang Wei, Wang Hongwei, Meng Yuan. Collaborative filtering recommendation algorithm research: Considering online review sentiment tendency [J]. System engineering theory and practice, 2014,34 (12): 3238-3249