Research and analysis of music development based on k-means and PCA algorithm

Zimo Cai¹,², Luqi Fu², b and Wenchao Li², c, *

¹ School of QingHu, Nanchang University, Nanchang 330031, China
² School of Mechanical and Electrical Engineering, Nanchang University, Nanchang 330031, China

a zimoC0104@163.com, b a272696448@outlook.com,
*, c Corresponding author e-mail: chadli@ncu.edu.cn

Abstract. The purpose of this article is to establish an algorithm model that can measure the influence of music, capture the evaluation index reflecting the influence of music, and extend the model to other fields such as politics, culture, and society. We have established a music influence-oriented network algorithm model based on influencers and followers, where each artist is a node, and each follower is a connection between artists. We define relative interaction strength indicators to help understand the entire network algorithm. In addition, we also used time, genre and other scales to further optimize the network algorithm. We first use the PCA algorithm to determine indicators that reflect music similarity, such as vitality, activity, popularity, overall loudness, etc. On this basis, an evaluation algorithm model based on cosine similarity is established to calculate music similarity values of different genres. In addition, we use the K-MEANS algorithm to normalize each feature index and sum its variance. Finally, we noticed that the similarity of artists within genres is higher than the similarity of artists between genres. We further analyzed the differences and influences within and between genres. Taking time as a distinction, a relative heat map of the interactive influence of genres is drawn. It is understood that certain genres will obviously have a certain influence over time. We summarize this model as an impact correlation analysis model. First, we choose a representative influencer. Then, based on the cosine similarity, we obtained the music similarity with the fans in batches, thus more intuitively concluded that the Internet celebrities did affect the respective artists. In addition, we combined the calculation of SPSS variance and selected different indicators to visualize the radar chart to understand the attractiveness differences of certain music features. We first select the musical characteristics with obvious changing trends, then locate the position of the changer in the music evolution process through the time distribution diagram of the corresponding work, and finally select the representative changer. We analyzed the change history of each indicator in the selected genre over time, and finally got the global directed network diagram. Based on the network algorithm model established in the previous question, we analyzed the background of the times and found that there is an interaction between music and the cultural environment. Finally, we also analyzed the advantages and disadvantages of the algorithm model, and discussed the application of the method in other fields.
Keywords: K-means algorithm, PCA algorithm, quantitative index, correlation, model optimization.

1. Introduction

1.1. Background

Since ancient times, music has been a part of human society, and it continues to develop with the changes of the times. In the process of music evolution, many factors such as talent creation, personal experience, the emergence of new musical instruments, or current social or political events will affect the artist's creation of new music. Under the influence of these factors, music sometimes undergoes revolutionary changes, new genres appear, or existing genres are reinvented. Among them, previously produced music will also have a greater impact on new music and music artists. To measure this kind of musical influence, it is necessary to use the similarity between the network of songs and their musical characteristics (for example: structure, rhythm or lyrics) to capture the mutual influence between music artists in order to better understand how music changes evolving with social changes.

2. The Evolution Law Model of Music Genre

2.1. Model introduction

Algorithm 1: PCA, also known as principal component analysis technology, aims to use the idea of dimensionality reduction to convert multiple indicators into a few comprehensive indicators. The specific PCA transformation steps are as follows:

1. Calculate the covariance matrix $S$ of the matrix $X$ sample (this is non-standard PCA, standard PCA calculates the correlation coefficient matrix $C$);
2. Calculate the eigenvectors $e_1, e_2, ..., e_N$ and the eigenvalues $t=1, 2, ..., N$ of the covariance matrix $S$ (or $C$);
3. Project the data into the space formed by the feature vector and use the formula

$$newBV_{i,p} = \sum_{k=1}^{n} e_i BV_{i,k}$$

The BV value is the value of the corresponding dimension in the original sample.

The goal of PCA is to find $r (r < n)$ new variables, make them reflect the main characteristics of things, compress the original data matrix size, reduce the dimension of the feature vector, and select the least dimension to summarize the most important features. Each new variable is a linear combination of the original variables, reflects the comprehensive effect of the original variables, and has a certain practical meaning. These $r$ new variables are called "principal components", they can reflect the influence of the original $n$ variables to a large extent, and these new variables are not correlated with each other and are orthogonal. Through principal component analysis, the data space is compressed, and the characteristics of the multivariate data are visually expressed in the low-dimensional space [2].

Algorithm 2: k-means, k-means algorithm is a very widely used algorithm in spatial clustering algorithms, and it also plays an important role in cluster analysis. It is a centroid-based division technique and one of the most commonly used clustering methods in the field of data mining. It originally originated in the field of signal processing. Its goal is to divide the entire sample space into several subspaces, each of which has the smallest average distance from the center of the space.

The k-means algorithm accepts the input quantity $k$, and then divides the $n$ data objects into $k$ clusters in order to satisfy the obtained clusters. The similarity of objects in the same cluster is high, while the similarity of objects in different clusters is small. Cluster similarity is calculated by using the mean value of the objects in each cluster to obtain a "central object" (center of gravity) [3].

In general, the working process of the k-means algorithm is explained as follows:
1. Select k objects from n data objects as the initial cluster centers. For the remaining objects, assign them to the most similar clusters according to their similarity with these cluster centers.
2. Calculate the cluster centers of each new cluster obtained.
3. Repeat this process until the standard measure function starts to converge.

Generally, the mean square error is used as the standard measurement function. The k clusters have the following characteristics: each cluster itself is as compact as possible, and each cluster is separated as much as possible.

Suppose that the data set D contains n objects in Euclidean space. The partition method assigns the objects in D to K clusters C_1,...C_j, so that the objects 1 ≦ i, j ≦ k, C_i∈D and C_i∩C_j=Ø, an objective function is used to evaluate the quality of the division, so that the clusters Inner objects are similar to each other, but different from objects in other clusters. In other words, the objective function aims at high similarity within clusters and low similarity between clusters.

The centroid-based division C_i technology uses the centroid of the cluster to represent the cluster. The difference between the object s∈C_i and the representative c_i of the cluster is measured by dist(s, c_i), where
\[ d(s, c_i) = \sqrt{\sum (x_i1 - y_i2)^2} \]

The quality of cluster C_i can be measured by intra-cluster variation, which is the sum of squared errors between all objects in C_i and centroid [c.i], then:
\[ \alpha = \sum_{i=2}^{k} \sum_{p \in c_i} d(s, c_i)^2 \]

Among them, \( \alpha \) is the sum of squares of the errors of all objects in the data set; s is a point in space, which represents a given data object; c_i is the centroid of cluster C_i.

2.2. Model establishment

We use the three comprehensive feature indicators derived from the PCA algorithm to draw a three-dimensional map, and use the k-means algorithm to perform cluster analysis on artists of different genres. Because the data display of the three-dimensional map is not clear and intuitive, we select the first two comprehensive feature indicators to draw the plan, where the dots with different colors represent different genres, and the number of dots represents the number of artists in that genre. Then, based on the three comprehensive feature indicators obtained by dimensionality reduction, we draw the true plane distribution map of each genre artist.

Figures 1 reflect that the distribution of artists in real situations is relatively scattered, which is in line with the reality. However, the floor plan processed by the k-means algorithm shows that the distribution of artists of different genres is regular and concentrated, which does not match the reality, indicating that the model has shortcomings. The k-means method is not fully applicable and cannot fully explain the real situation. The model needs to be further improved.
2.3. Music and cultural environment interact

The evolution of music is the result of the interaction of various cultural environment factors, and music also reacts to various cultural environments. The two influence each other, which is a coexistent relationship. According to the previously established model, we can analyze the changes in the number of followers and influencers of each music genre over time, and then introduce the development status of the genre in this era and the general trend of music evolution. You can find it in the context of the times. To the cultural environment factors that influence this change in the genre. We can also analyze the most influential music genre at that time based on the model network and infer the influence of music on the cultural environment of that era based on the representative music characteristics of that genre. Generally, these cultural environmental factors include: the local and surrounding cultural environment, business environment, media environment, educational environment, social cultural psychological environment, social group participation environment, cultural environment under the influence of government regulation, and formed due to the impact of major social events Cultural environment, etc. [4].

For example, during the outbreak of the World War, social turmoil occurred and people's social psychology changed fundamentally, resulting in negative psychology such as tension, anxiety, and depression. The war environment has prompted a new evolution in music. We can deeply feel this in the musical works created by artists such as Bazak, Schoenberg, and Berger. In this environment, the music they create no longer shows. It is a peaceful pastoral life. It is no longer an exaggerated fantasy inner world, nor does it pursue the momentary erratic feeling like impressionism but shows the ideological realm of the suffering people longing for peace [5]. These musical works also had a certain impact on society at that time. They were inspiring, gave spiritual support, inspired the people to resist bravely, and promoted the arrival of the era of peace. According to the model established by the previous questions, it can be seen that the followers of the Pop/Rock genre with loud and energetic musical style were born during World War II. Reversely, the Pop/Rock genre emerged during World War II. Due to the unique music style of this genre, it can inspire people, give people the confidence of victory at that time, and accelerate the dawn of peace.

Music as one of the manifestations of art, its evolution will also be affected by modern science and technology. Every step of the development of music is carried out under the premise of social change and scientific and technological development. From the making of chimes in the Bronze Age of mankind to the production of various exquisite musical instruments after the outbreak of the Industrial Revolution, from the dissemination of music on phonographs and records to the dissemination of digital media and Internet art at this stage, it all demonstrates that the development of music art is the result of technological advancement. For example, music in the current Internet age not only overcomes the shortcomings of the natural communication era of music (that is, the original communication form), such as the small transmission range and the vague transmission effect, but also overcomes the traditional technology transmission era that is separated by both parties and the recipient is passive. Disadvantages, compared with the previous music creation, great changes have taken place in terms of communication, creation, style, and aesthetic value orientation [6]. However, due to the emergence of music in the Internet era, it can be seen in the network diagram drawn that many music genres gradually declined in 2000.

3. Conclusions

In our work, we have established multiple models to quantify the influence of music evolution. Firstly, by constructing a music influence directional network, the influencers and followers are connected, the relative interaction intensity index is introduced, and the musical influence of each genre is shown in the form of a heat map. Next, an evaluation model based on cosine similarity is established, and the music similarity values of different genres are calculated. It is found that the music similarity of artists within the genre is higher than that of artists between genres. Then, using time as a distinction, the PCA and k-means algorithm are used to establish related models, and it is concluded that the following influence relationship among the genres will gradually weaken over time. Then we borrowed the variance to
calculate the similarity radar chart of the music characteristics of each genre. The results show that influencers influence the music creation of followers mainly because some of their musical characteristics are more "infectious." In addition, we selected representative music changers based on relevant models and established a global directed network. Finally, combining the background of the times and the network model built, we can learn that there is a mutual influence between music and the cultural environment.

References
[1] [Machine Learning] Analysis of several similarity algorithms: https://blog.csdn.net/weixin_39050022/article/details/80732249
[2] PCA technology: https://blog.csdn.net/weixin_43758551/article/details/87351073
[3] Wang Yuanru. (2015). Overview of k-means algorithms in cluster analysis. 3-7
[4] Wang Siqi. The Influence of Contemporary Chinese Social and Cultural Environment on the Formation of the Development Characteristics of Pop Music [J]. Music Studies (4):11-17.
[5] Wang Jianxun. The change of music in the 20th century and its causes [J]. Journal of Beihua University: Social Science Edition, 1995.
[6] Xu Xiaoting. Research on the Changes and Causes of Music in the Internet Era [D]. Jinan: Shandong University, 2009,(5).
[7] The cosine similarity of similarity algorithm: https://blog.csdn.net/zz_dd_yy/article/details/51926305
[8] Wang Yuanru. (2015). Overview of k-means algorithms in cluster analysis. 3-7.