Applying Data Mining in Music influences analysis

Siqi Lai [1]

[1]Zhejiang University of Finance & Economics Hangzhou Zhejiang, China

Abstract. Nowadays, music has become an indispensable part of our life. Studying the influence and evolution process of music will help to promote the progress of human civilization. The purpose of this paper is to understand and measure the impact of previously produced music on new music and music artists, and to develop a model for measuring the impact of music. Taking music genre as the main body, this paper constructs a directional network of music influence and a sub-network reflecting the relationship between "influencer and follower". In addition, this paper also defines the influence of music, and puts forward the "contagiousness index" to reveal a person's actual influence. Finally, using the parallel coordinate characteristics, the paper analyzes the influence of music on culture in time or environment.

1 Introduction

Have you ever had a moment when you thought there were a few songs that sounded so similar to each other? Some people even think that all pop music sounds similar. This phenomenon may be due to the mutual influence between music artists and music genres, or the influence of the social environment.

The development of the whole society is a complex process of interweaving and interaction of various factors. Various kinds of music are by no means isolated in the development of social culture. [1]The formation and development of new music is inevitably influenced by many aspects, such influences can be the creative characteristics and life experiences of musicians, as well as the influences brought by politics, economy and culture in the social period. Therefore, by studying the influence of songs, it will help us to understand the influence of different musical artists on each other. In addition, further quantifying the evolution of music will enable us to more intuitively understand the evolution and revolutionary trend of music artists and various music genres.

2 Literature Review

To solve the problem of music similarity, some scholars have published academic articles on the establishment of music similarity model. In 2003, Matthew Hoffman established a non-parametric Bayesian model (HDP) to represent songs by the mixture of multivariate Gaussian distribution, and to calculate the similarity of songs according to the different mixing weights of each song. [2]In 2005, Mandel described a new system, tested on the task of artist identification, that used support vector machines to classify songs based on features calculated over their entire lengths.[3]

In 2015, Xu Gang adopted audio feature extraction and neighborhood calculation to improve the G1 algorithm of classical music similarity measure, and proposed the ML-GS music similarity measure. [4] In 2017, Daniel Mullensiefen and Marcus Pearce extended the auditory expectation information theory model (IDYOM) to calculate the compressed distances of different symmetry and planning degrees, taking into account the characteristics of song pitch and the rhythm structure. The model was trained on one song and used to predict the notes of another song. The higher the prediction rate, the higher the similarity. [5]

3 Construction of the Directed Networks

Figure 1 The Influence of Each Genre on Pop/Rock Per Decade

Firstly, the goal of this question is to quantify the interaction between various music artists. After analyzing the data, it is found that all known artist influence/following relationships directly reflect the relationship between artists' musical ideas and creative inspiration from their respective genres. Such a one-way relationship is eventually reflected in the various genres and becomes the musical influence among the genres. Therefore, by exploring the influence of each genre on the output and input through the music achieved by the artist,

895673457@qq. com

the characteristics of the music influence relationship between them can be obtained. Taking Pop/Rock as example to show the output of these two music genres and the influence of each genre in each decade in Figure1.

In my approach, the above input and output behaviors generated by each genre across the entire timeline are used to construct targeted network relationships of musical influence. Focusing on the input and output behavior of a genre and further exploration, the sub-network relationship can be get directly related to the genre, as shown in Figure2.

Figure 2 The Sub-network of Artists Interacting

4 Determination of the Musical Influence

The "music influence" parameter is used to describe the strength of the interaction between music. Music influence is defined as the ratio of the number of new artists in a genre to the number of artists in that genre in a given time interval. For a genre that has a positive influence on artists, it will receive a high "musical impact" score. A genre with a high "musical influence" score will be represented by a large amount of musical influences output in its related sub-network structure. Take Figure3 for a specific analysis of music influence parameters. In the figure, B learns from A and B learns from N people in total, so it contributes 1/ N /1 of music influence to A (the next generation). C learns from B, and C learns from M individuals in total, so he contributes 1/ M /1 of his musical influences to B (every other generation) and 1/ M/N /2 of his musical influences to A (every other generation). That is:

$$weight = 1/total\_number\_of\_musicians\_followed\_by\_this\_follower/depths \quad (1)$$

Figure 3 The Example

In addition, each music artist is ranked according to the calculated music influence, and the top 20 artists are shown in Figure 4.

Figure 4 Artists Ranking

5 Evaluation of the Actual Influence

First, normalize the characteristics, analyze the characteristic of all music, and use the popularity weighting to figure out the standard deviation corresponding to each characteristic, labeled as STD_BEFORE – that is, the dispersion degree of each characteristic distribution when the follow relationship is not clear.

Then analyze all the identified influencers to obtain the characteristics of all the followers of each influencer. And also to obtain the average dispersion degree STD_AFTER, which is generated

from the variation of followers' characteristics on the characteristic of their corresponding influencer. That is to say, STD_AFTER is the distribution dispersion degree of each characteristic of all the followers with respect to their corresponding influencers.

5.1 Contagiousness indicator

Contagiousness is defined as the reciprocal of the ratio of STD_AFTER and STD_BEFORE, indicating the percentage of the influencers promoting the characteristic certainty in the network relationship above. The detailed calculation process is as follows:

$$mean = \sum_{i=0}^{N}(S_i * P_i) / \sum_{i=0}^{N} P_i) \quad (2)$$

$$std\_before = \sqrt{\sum_{j=0}^{N}(A_j - mean)^2 / N} \quad (3)$$

$$std\_after = \sum_{j=0}^{N}(\sqrt{(\sum_{j2}^{n\_follows}((A_j - F_{j2}))/n\_follows)}/N \quad (4)$$

$$contagiousness = (std\_after/std\_before)^{-1} \quad (5)$$

Contagiousness over 1 stands for a more significant role in influencing a particular artist's music. After the calculation of all characteristics is completed, the results are shown in Table1.

Table 1 Calculation of Contagiousness

Characteristic	Mean	std_before	std_after	contagiousness
danceability	0.5490	0.1671	0.3878	0.4309
energy	0.5841	0.2691	0.3885	0.6926
valence	0.5358	0.2585	0.3854	0.6707
tempo	0.4929	0.1227	0.3681	0.3333
loudness	0.7886	0.0811	0.4636	0.1749
mode	**0.7141**	**0.4480**	**0.4196**	**1.0678**
key	0.4738	0.3188	0.3532	0.9024
acousticness	**0.3387**	**0.3644**	**0.2828**	**1.2884**
instrumentalness	**0.0971**	**0.2742**	**0.1245**	**2.2020**
liveness	0.2012	0.1861	0.2255	0.8255
speechiness	0.0659	0.0789	0.1341	0.5855
explicit	**0.0538**	**0.1897**	**0.0297**	**6.3719**
duration_ms	0.0620	0.0285	0.1305	0.2188

6 Influence over Time and Circumstance

6.1 Insights of Data

Before 1956, the annual data volume was very small, the statistical results fluctuated too much, and the statistical error was high. Therefore, this question began to analyze in 1956. Using the characteristic parallel coordinate defined in task 5, the two characteristics of validity and explicit are extracted, as shown in Figure5.

Figure 5: Characteristic Parallel Coordinate about Valence and Explicit

6.2 Analyze influence from four periods

Through the analysis of the figure, four interest points marked in Figure 6 are found and studied.

Figure 6: Characteristic Parallel Coordinate about Valence and Explicit 2

6.2.1 The valance continued to rise from the middle and late 1950s to the early 1960s. (Mark A)

From the statistical results, this period is the rising period of country, Latin, jazz, pop / rock has just begun its rapid growth, and gradually become the main part of pop music. In social and political aspects, the United States and Europe have basically completed the recovery after the war, and social development has entered a rising period. Culture has also ushered in a renaissance. The theme of music is more and more positive, which is an affirmation of these phenomena.

6.2.2 In the middle and late 1970s, explicit appeared and rose for the first time. (Mark B)

After oil crisis, energy crisis and the 1973-1975 recession, the United States entered the post Watergate period, the European countries' economic downturn and social problems began to highlight. Pop / rock has become the mainstream of music in this period. Each genre produced a large number of variations and derived a large number of branches. Among them, punk branch, which belongs to pop / rock, causes punk subculture in society. At the same time, it is found for the first time that the explicit value is greater than 0, which means that the limiting factors begin to appear in music creation.

6.2.3 From the 1980s, the value began to decline. (Mark C)

The impact of 70s further led to the recovery of global economy. The 80s also brought about many political changes, and the threat of war increased. The reform of popular culture represented by the launch of MTV has affected the business model of music. The overall popularity of music continues to rise, but the emotions expressed in music also begin to develop negatively.

6.2.4 After entering the 2010s, the number of explicit increased significantly. (Mark D)

With the development of digital media, digital music distribution has become the primary form of music consumption. The share of record companies has fallen sharply. The rise of social media has also redefined the mode of music transmission, changing the past radio format. Pop music shows strong adaptability and has achieved notable commercial success in these changes. Genres such as hip hop and rap also ushered in a period of rapid growth, and more new genres with regional characteristics also appeared in the public view. To sum up, in this period, music became more diversified and complex, the original constraints were reduced, and the creation of artist was more free.

7 Conclusion

This paper holds that the essence of the mutual influence among artists is the mutual influence among their corresponding genres. Therefore, based on the input and output of music genres, a directed network of music influence and a sub-network reflecting the "influencer - follower" relationship are constructed. In addition, the music influence is defined and the quantitative results of this index are displayed.

In order to reveal the actual influence, the dispersion of each feature is calculated, and whether the index is reduced after binding the network is checked. The definition of contagiousness index reveals this relationship. If the index exceeds 1, the impact is realistic and effective. The results show that some features are more easily affected, and the "explicit" feature is the most obvious.

In order to analyze the influence of music on culture in time or environment, the characteristics of the parallel coordinates were used to check the change of characteristics 'valence' and 'explicit'. Four interest points and the corresponding analysis were shown in the paper.

My research findings show the value of understanding the influence of music through networks, and will contribute to the future musical research.

References

1. Wang Siqi. (2005). A study on the Development of Chine City Popular Music between 1978- 2003 and its Interactive relation with the Social-cultural Environment[D](Doctoral dissertation, Fujian Normal University).

2. Hoffman, M. D, Blei, D. M. , &Cook, P. R. (2008, September). Content-Based Musical Similarity Computation using the Hierarchical Dirichlet Process. In ISMIR (pp. 349-354).

3. M. Mandel, D. Ellis (2005). Song-Level Features and Support Vector Machines for Music Classification Proc. Int. Conf. on Music Info. Retrieval ISMIR-05, London, September 2005.

4. Xu Gang. (2015). Research on Audio Similarity and their Applications in Music Retrieval(Master's thesis, University of Electronic Science and Technology of China).

5. Pearce, M. &Müllensiefen, D. (2017). Compression-based modelling of musical similarity perception. Journal of New Music Research, 46(2), 135-155.