ISSN 2618-1568 Vol. 3, Issue 2: 10-14, DOI: 10.25236/FAR.2021.030203

Research on the influence of Music revealed by big data based on the Analysis of Network Theory

Zimai Dong, Cailian Xie, Yukun Xiao, Yuxuan Hou, Zhixiong Tao

Dalian Maritime University, Liaoning, Dalian, 116026

Abstract: *Music is the result of human's conscious cultural creation. The creation of some artists has been deeply influenced and inspired by other artists and works of art, thus showing similarities in certain features of songs. There are also revolutionary changes in music, such as the creation of brand-new musical styles, the emergence of new genres, re-creation of genres and so on. Through supervised learning, the paper directly use genre as the label of classification. However, due to the failure of the test, the paper reclassified the artist and music attributes in the only way based on the data. In the process of classification, firstly, the data are normalized, then the principal component analysis ((PCA)) is used to reduce the dimension of the attributes, and then the K-means clustering algorithm is used to cluster the samples to get new classification results. The paper creates the concept of degree coefficient to determine whether the new category reflects the original genre. The paper also introduces the concept of information entropy, which is based on artist data and music data to quantify the similarity between and within genres. As a result, the purpose of developing the measure of music similarity is achieved. The paper uses the degree coefficient to eliminate the general level of music attributes in the environment of influencers and followers and highlight the differences between them and the general population. Therefore, it can truly reflect whether the influencer has influence on his followers, as well as the size of the influence.*

Keywords: *Information entropy, Degree coefficients, Two-way pointer search, Fitting*

1. Introduction

Music is the result of human's conscious cultural creation. Artists in the creation of music, the background of The Times, their own originality, personal experience and other factors affect their works [1]. Based on the dataset, the paper evaluates the influence of music by creating a targeted network of music influence. The paper also developed a music similarity measurement model, analyzed the influence of music, and studied the evolution of music. In this paper, the paper preprocesses the data first. This paper creates a directional network of music influence, and quantifies the artist's music influence with the central concept. Develop music similarity to reclassify artists and music, and data is the only basis for classification. The concept of information entropy is introduced to quantify the similarity between and within schools. Create the concept of degree coefficient to visualize the influence of the influencer on the follower. This paper designs a two-way pointer search method to find out the revolutionary music characteristics and its influence on the development of music. Then the fitting method is used to analyze the evolution of music with time [2].

2. Developing measures of music similarity

Genre is the classical idea that people used to distinguish music styles. Through supervised learning, the paper directly take genre as the label of classification. On the one hand, the paper can preliminarily explore whether genre can reflect the difference of music. On the other hand, if the model can be tested, it will greatly reduce the workload of follow-up research [3].

First, the paper use the logistic regression distribution function:

$$F(x) = P(X \le x) = \frac{1}{1+e^{-(x-u)/y}} \tag{1}$$

Among them, X is the column vector of xi, F(x) is the discriminant function, u is the position parameter, and y is the shape parameter. Since the logistic regression distribution function has parameters

ISSN 2618-1568 Vol. 3, Issue 2: 10-14, DOI: 10.25236/FAR.2021.030203

u, y unknown, the paper need to estimate the parameters. The parameter estimation adopts maximum likelihood estimation [4]. The paper can solve the parameter through cost function and logarithmic transformation. Logarithmic transformation calculation formula:

$$L(\omega)=\sum\left[y_i(\omega_i\cdot x_i)-1n\left(1+e^{\omega_i\cdot x_i}\right)\right] \tag{2}$$

3. The new classification

The paper connects the artist id and genre data from data by artist and the influence data out of the database to get the new data. The new data are: all artists, artist genres, and all music attributes of the artist [5].

(1) Missing value processing: find the tuple with missing value in the genre attribute in the new data, and replace all of them with unknown.

(2) Averaging: Averaging all music attributes of each genre to get the data the paper need.

The attributes of music works are normalized as follows:

$$x_{new}=\frac{x-\mu}{\sigma} \tag{3}$$

In this formula, the value of any attribute is the difference bet. The paper its original attribute value and its mean value divided by the standard deviation. Normalization enables different attributes and dimensions of values to be analyzed in the same method. Pairs of covariances of the 14 attributes the paper calculated, and the covariances the paper formulated as follows:

$$\operatorname{cov}(X,Y)=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{n-1} \tag{4}$$

Eigenvector is a non-zero vector that makes a matrix satisfy the following formula:

$$A\vec{v}=\lambda\vec{v} \tag{5}$$

After PCA dimensionality reduction of 14 attributes was obtained by Python software, 6 principal components with pheromone index greater than 95% are selected as new attributes. 6.2.4 K-means clustering the paper input the six new attribute values of all the artists into Python software for a 6dimensional k-means clustering. The main function of this algorithm is to automatically group similar samples into a category. Clustering attempts to divide the samples in the data set into several subsets, which are usually disjoint, and output these subsets. In the process of clustering, the paper test the number of clustering classes by trial calculation. First, the number of clustering classes is equal to 20, and then the deviation values of all clustering are calculated. The calculation formula of deviation value is as follows:

$$p=\sum_{i=1}^{n}\sum_{j=1}^{14}\left(p_{ij}-p_{0j}\right)^2 \tag{6}$$

Secondly, the cluster number is equal to 19, 21, 18, 22, etc. (in accordance with the fluctuation law), and the deviation value is obtained. Finally, select the cluster number with the smallest deviation value. In the end, the paper came up with 20 new categories.

4. Results music within and between genres

When comparing music within and between genres, the data used for K-means clustering is different from the previous article. The paper connects the artist id and 17 kinds of music attribute data in full music data with the artist id and genre data in influence data to obtain new data. The new data are: all artists, artist genres, and various attributes of all musical works.

ISSN 2618-1568 Vol. 3, Issue 2: 10-14, DOI: 10.25236/FAR.2021.030203

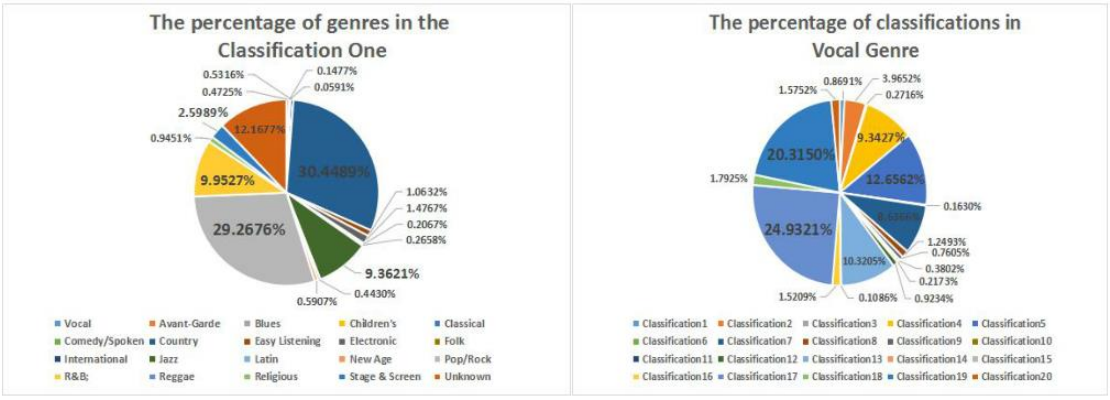

Figure 1: Pie chart

Repeat the above steps. The PCA method reduces the dimensions of 17 attributes into 9 new attributes, K-means clustering clusters the samples into 20 categories, and finally calculates the information entropy.

The paper does average processing in the new data to get the music attributes of each genre. Extract the Electronic genre, and do a unary regression analysis of its various music attributes over time. As shown in Figure 2, the higher the correlation coefficient R, the more obvious the linear relationship of the attribute value over time. Observing the scatter plot, in the explicit plot, there was a sudden increase in the value around 2000. The paper thinks this phenomenon may be related to factors such as technological progress and Internet development. In summary, the paper conclude that genres have a stronger or weaker linear relationship with time, but it is also due to consideration of the revolutionary changes in the background of the times as time advances.

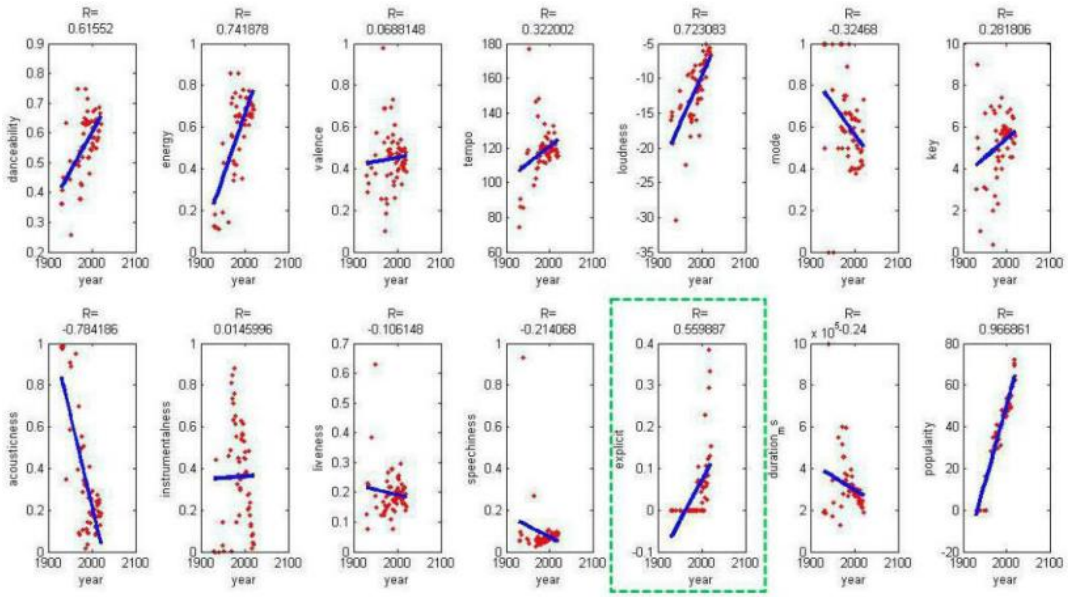

Figure 2: Linear regression graph

The paper calculated the coefficient of the average level of influencers and artists and the coefficient of the average level of followers and artists. The statistical results are shown in Figure 3. The main purpose of this calculation is to observe the respective attribute levels of influencers and followers on the average level of all artists. The degree coefficient indicates that influencers and followers reflect their differences from the general public when the universal music attributes are eliminated. It is these differences that can very well reflect whether the creation of followers is really affected by the influencers.

ISSN 2618-1568 Vol. 3, Issue 2: 10-14, DOI: 10.25236/FAR.2021.030203

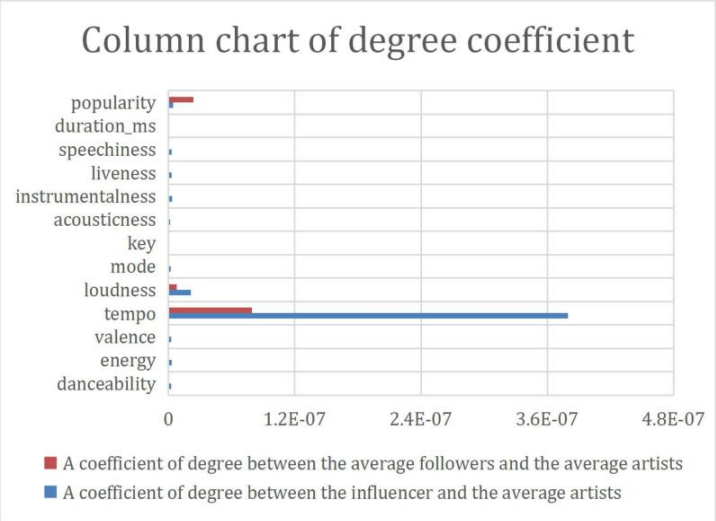

Figure 3: Histogram of degree coefficient

Observing the histogram, the paper finds that the overall change trends of the two-degree coefficients are roughly the same, and there is no reverse change phenomenon. And the degree coefficient of tempo attribute influencer has a greater influence on the degree coefficient of followers. So, the paper can conclude that influential people will influence the music created by followers. The musical feature of tempo is more infectious than other features.

The paper use artist id and 17 kinds of music attribute data in full music data to connect to the data of artist id and genre in influence data to connect to the database, and filter out the data with the genre Electronic from the new data, and then average all music attributes for each year. Then use MATLAB to analyze linear regression of each attribute, as shown in Figure:

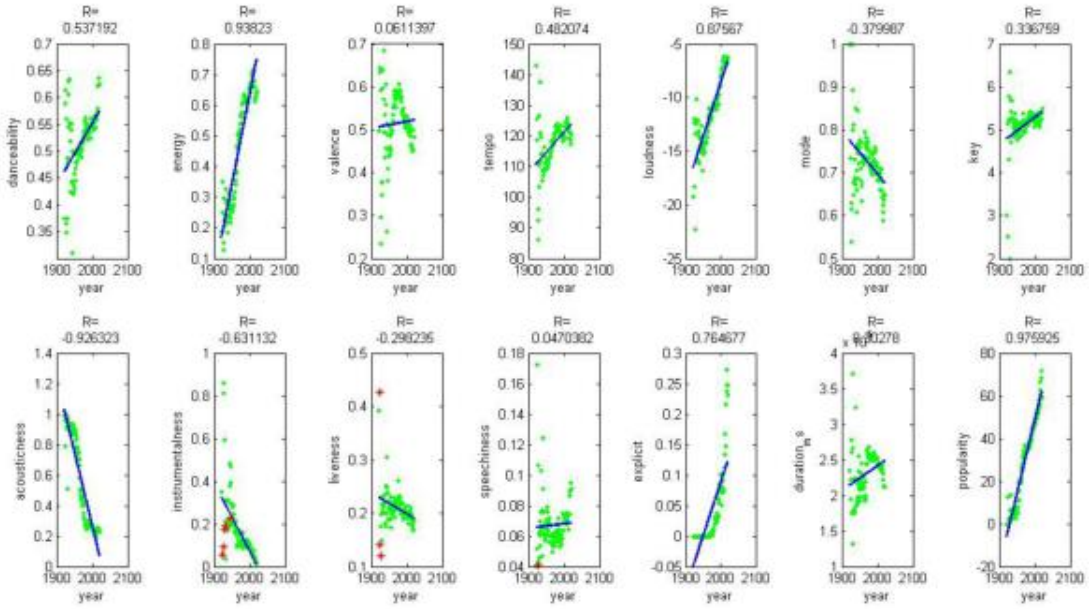

Figure 4: Linear regression graph

According to the above figure, the paper believes that the attributes with the absolute value of the correlation coefficient R greater than 0.5 can be analyzed by linear regression. Therefore, the paper performs a univariate regression on the attributes of danceability, energy, loudness, instrument as well as, explicit, and popularity. The regression equation explains the changes of these attributes of the Electronic genre over time. The regression equation is shown in Table 13, where a can reveal indicators of dynamic influencers.

ISSN 2618-1568 Vol. 3, Issue 2: 10-14, DOI: 10.25236/FAR.2021.030203

5. Conclusions

The paper can see that by changing the threshold Z, the two sets of years extracted by the program are completely consistent. Therefore, it can be concluded that the model is basically stable by comparing the threshold change and the model output change. Fuzzy mathematics is introduced to express similarity. Looking for the weight of each attribute in the existing research, the paper can carry out the maximum expected (EM) clustering for the Gaussian mixture model. (GMM) The exploration of genres is the center of the whole work. The formation of music genres did not happen overnight. It originated from the ancient people's understanding of the world of music and the yearning for their love. Musicians are happy to engage in the enjoyment of music belonging to the genre, but big data tells us: This may not be the case. This is not a bad thing, it greatly reflects the fusion of music. New music is more popular among the masses. Diversified music is inevitable in the information age. This result is drawn from linear regression prediction: irreversible!

References

[1] Y F Huang, S M Lin, H Y Wu, Y S Li. Music genre classification based on local feature selection using a self-adaptive harmony search algorithm. Data & Knowledge Engineering 92 (2014) 60–76.
[2] C M Wu, Z Cao. Noise distance driven fuzzy clustering based on adaptive the paper lighted local information and entropy-like divergence kernel for robust image segmentation. Digital Signal Processing 111 (2021) 102963.
[3] Henk Jan Honing. Evolving Musicality. Massachusetts: The MIT Press, 2019.
[4] Biggs, Norman, E. K Lloyd, and Robin J. Wilson. Graph Theory, 1736-1936. Oxford University Press, 1986. Oja, E., 1982. A simplified neuron model as a principal component analyzer. J. Math. Biol., 267-273.
[5] S Wang, S X Sa, Introduction to Database System. Beijing: Higher Education Press. 2014.9