Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
D		2123447

Music Evolution: A Network Model of Influence

Summary

Music is an important part of the cultural heritage of human beings, and quantifying the history of music development can help us better understand the context of the development of music today and the role it plays in human society. In this paper, we rely on BP neural network to conduct machine learning to study the relationship between music, and draw influence network to spy on the relationship between music influencers. Principal component analysis is used to find the time point of music revolution. In this way, we can have a glimpse of the history of modern music.

1. For the first question, we used Gephi software to plot the network of 80 highest influencers, and found that these influencers had extremely close relationships with each other. Then, we use MPAI website machine learning to draw a targeted network of music influence in a random sample. Based on the new evaluation index calculation method of scientific research influence, we combine 14 music indexes into music influence parameters. It shows how much influence a musician has over other people.

2. For the second question, we use cosine similarity as the musical similarity measure. Again we put the string data into digital music genre and genre as dependent variable, song characteristic parameters as independent variables, through three models: the decision tree, bp neural network and LightingGBM models rely on the part of the training data set, from another part of the data set to test: two matching different songs from all genres, cosine similarity calculation. After testing, we still found that songs of the same genre do have significant similarities.

3. For the third question, we use the neural network model of the second question to analyze, which can calculate the similarity of different genres and judge the types of songs. We give the relation of different characteristics of classical music with time.

4. For the fourth question, we defined the inheritance ratio of each genre, calculated the contagiousness coefficient of each genre and ranked it accordingly. We screened the first five genres and analyzed their characteristics: accessibility, inclusiveness, and uniqueness

5. For the fifth question, we used principal component analysis (PCA) to reduce the dimension of music features and get the analysis factors. By analyzing the change of factors over time, we find out the time node when the music has a significant change, and study the historical environment behind it.

6. In the sixth question, we selected the data of the top six musicians in the Pop genre and analyzed the curve of the number of their followers over time.

7. In the last question, we analyzed the change curve of the two characteristics of Pop Music over time, and found out the background factors that may affect them.

After all, we send a letter to ICM to describe the results of our research.

Key words:Music influence, PB neural network, decision tree, Light TGBM model, cosine similarity,

Contents

1 Introduction

1.1 Restatement of the problem

To quantify musical evolution, which is based on the influence that musical artists have on each other, we hope to develop a model that measures musical influence, to measure and examine evolutionary and revolutionary trends of artists and genres, to explore the musical evolution.

1 Use influencedata to create multiple directed network of music influence, capture "music influence" in this network, and find it reveal meaning.

2 Use summary data sets of music charecters develop mesaures of music similarity, to mesaure the similarity of music from different artists in same genre.

3 Compare similarities between different genres, dectet the impact between themselves, and their changes overtime,

4 Indicate the ture relationship between the influencer and their followers. Are some music really have the "contagious", or some artists good at learn widely from others?

5 Identify if there are characteristics might signify revolutions in music evolution? What artists repersent revolutionaries?

6 Analysis the influence progress of musical evolution over time, identify indicators, and explain the change of the genres and artists over time.

7 Express information of culture influence in time, dentiify the different environment effects in the network.

2 Assumptions

1. Our data are detailed and reliable
2. The data given are more comprehensive within this time frame

3 The Model

3.1 Question 1

For problem 1, we built the network graph with Gephi software. We use the top 80 influencers with the highest influence to build the influence network, to study their influence relationship. The graph of the network we built is as follows

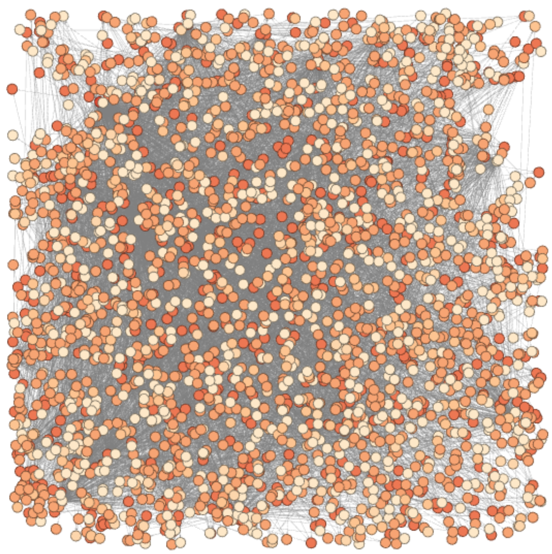

Figure 1: Top 80th Influencer Network

In this figure, each dot represents an influencer. The closer the color of the dot is to red, the larger the number of influencers at the next level/the higher level. In other words, the red dots represent the influencer has a wider number of influencers or the follower who are good at learning from many others. The connection line between them represents the relationship of influence between them. It can be observed from this figure that the lines of interaction among the 80 influencers are very dense, indicating that the influential people are all related to each other to some extent.

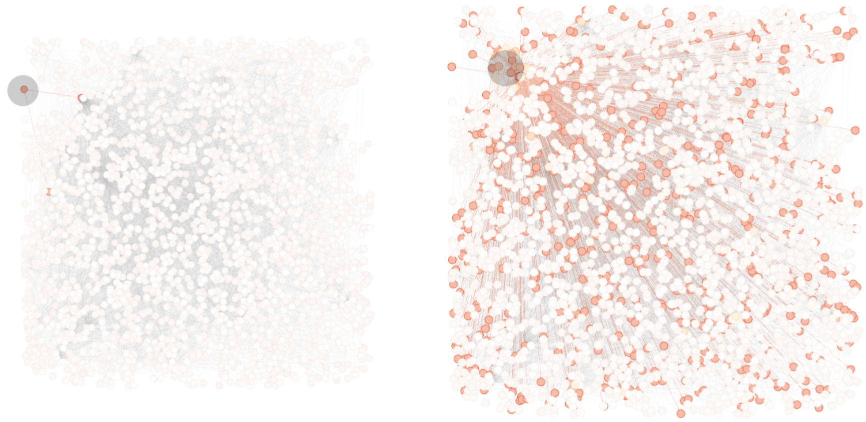

Figure 2: Top 80th Influencer Network details

The two graphs are the detailed illustration of Figure 1. The left graph is an illustration of Figure 1, showing the relationship between one node in the network and two other nodes (influencers or followers).
Only the red dots remain on the right-hand graphic, representing the most influential people.

In the first Figure (Figure 1), the number of people connected to the black dot is about less than 5, while the number of people connected to the orange dot is 5/6, and the number of people connected to the red dot is basically 20 to 30.

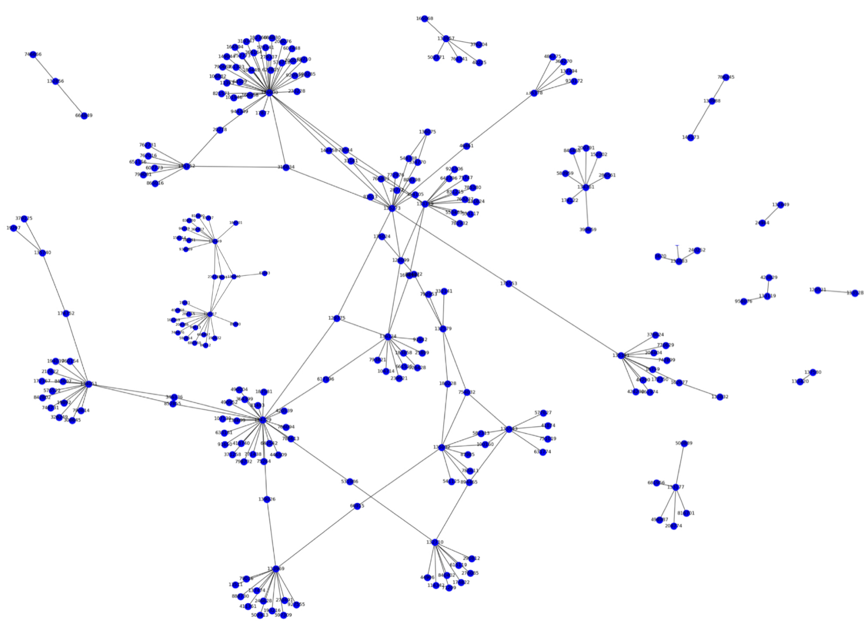

Figure 3: Music Network

Based on MPai website,First, we will randomly select data to affect the number of other people in descending order, respectively, the top 1% of data, the top 5%, the top 10%, the top 25%, the top 50%

and the bottom 50% of the data to give different weights from 6 to 1. Then draw the following network diagram with the artist's ID as the node and the influence relationship as the edge.Now we have now drawn the influencer network diagram. Observing the network, we find that most of the network is influenced by a group, while a few are isolated and "independent" from the mainstream.

This network has some similarities to the distribution of human clusters, which may be potential evidence of a mapping between the physical distribution of human beings and the mental distribution of human populations.

In order to find a reasonable index of music influence, we draw on the impact factor evaluation system in the field of scientific research.

3.1.1 The Parameter of Music Influence-Use Percentile Rank

The National Science Foundation of the United States uses the Percentile Rank as the evaluation index of scientific engineering[1] . Percentile Rank ranks the citation frequency of journal papers in the discipline into different levels, and gives corresponding weights to each level to get the weight sum of all papers in the journal, and then calculates the average weight value to evaluate the influence of journals.

Now we use it to evulate the parameter of Music Influence(MI). We caculate every single genre, and in each genre, we divid the influencer of the same discipline into six grades, namely, Top1 % , Top5%, Top10%, Top25%, Top50% and Bottom50%, and then assigned the weight of the first grade influencer to 6, the second grade influencer to 5, and so on, and finally calculated the average weight.[2] The parameter of Music Influence(MI) is calculated by the following formula

$$\mathrm{MI} = \sum_i x_i * f(x_i) / N \tag{1}$$

In Equation (1), $x_i$represents the weight of Grade i, $f(x_i)$ represents the number of Grade i influencer's followers, and N represents the total number of all followers in total.

Now we find the most famous influencers in each genre:

Avant-Garde		Blues		Classical		Comedy/Spoken		Country		Easy Listening	
750519	1.721518987	608701	0.54545455	183867	183867	744548	0.7894737	549797	0.33444411	678009	1.25
496525	0.607594937	276085	0.39823009	37658	37658	639005	0.3947368	816890	0.20357467	249175	0.4210526
186927	0.303797468	815039	0.24939662	849672	849672	12654	0.2631579	410219	0.1817631	311074	0.2763158
Electronic		Folk		International		Jazz		Latin		New Age	
104714	0.912676056	577531	0.47524752	785283	0.41158537	423829	0.353461	781837	0.32850242	489520	0.8510638
2411	0.23943662	266160	0.23267327	138833	0.35060976	175553	0.2606775	186312	0.16908213	675495	0.3829787
378288	0.225352113	170815	0.18316832	404463	0.32012195	211758	0.2098675	607283	0.16908213	945269	0.3829787
Pop/Rock		R&B;		Reggae		Religious		Stage & Screen		Vocal	
754032	0.152851995	316834	0.18336347	785380	0.3573201	668448	0.4129555	798662	0.81967213	79016	0.4497878
66915	0.096681993	128099	0.16708861	295276	0.21712159	24944	0.3238866	327765	0.3715847	184502	0.2652051
894465	0.079284205	622805	0.14972875	328014	0.21712159	31263	0.3238866	760240	0.34972678	792507	0.2510608

Figure 4: Top Influenceers

[1]Analysis of New Scientific Research Influence Evaluation Index

[2]If one genre do not have enough musicans, we cut the first grade, the new first grade begin as Top %5, 5

3.2 Question 2

For the second question, we use cosine similarity as the music similarity measure, and calculate the similarity through pairwise pairing. One part of the data is randomly selected as the training set and the other part as the test set. Machine learning is carried out based on the neural network, and then the test is carried out.

Firstly, we preprocessed the data. We cut out music with multiple genres/authors (two or more). Because the exact classification of these music is uncertain, it is unlikely that the authors took into account specific genres and styles when composing. We also excluded outliers and invalid data, as they did not make any positive contribution to learning outcomes.

Because the neural network can't recognize words (strings), we also convert the Music genre name to the corresponding numeric value (e.g., Pop Music to 10).

We use cosine similarity as a similarity measure because it is not affected by zero transactions and is therefore crucial in measuring the correlation between data in large transactional databases.

3.2.1 The Cossine Similarity

Cosine similarity is a method of calculating similarity. The method firstly maps the individual index data into the vector space, and then measures the similarity between two individual vectors, which by measuring the cosine of the Angle between them in the inner product space. The closer the Angle between two individual vectors is to 0°, as the greater cosine value of the Angle, meaning the higher the similarity between two individuals is. The closer the Angle between two individuals is to 180°, as smaller cosine value of the Angle between two individuals, indicating the lower the similarity. As can be seen from Figure 11, vector **A** and vector **C** are more similar than vector **B**.

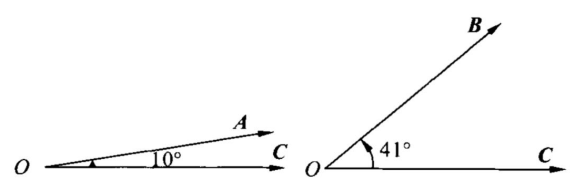

Figure 5: The explanation of cosine similarity

The cosine value between two vectors is calculated by Euclidean dot product formula:

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| \times |\mathbf{b}| \cos \theta, \theta \in [0, 2\pi] \tag{2}$$

$\cos \theta$ is defined as the similarity between two vectors, and the value range is [1, -1]. The cosine similarity of the two individuals is calculated by the following formula

$$\cos \theta = \frac{\mathbf{Y}_i \cdot \mathbf{Y}_s}{|\mathbf{Y}_i||\mathbf{Y}_s|} = \frac{(y_{i1}, y_{i2}, \cdots y_{in}) \cdot (y_{s1}, y_{s2}, \cdots y_{sn})}{\sqrt{\sum_{j=1}^{n} (y_{ij})^2} \times \sqrt{\sum_{s=1}^{n} (y_{sj})^2}} \tag{3}$$

In this formula, $\mathbf{Y}_i$ and $\mathbf{Y}_s$ respectively represent i and s different vectors.

3.2.2 Mesaure the Similarity of Music

There are many evaluation indexes for music. In order to comprehensively consider the characteristics of all aspects of music, we set and standardize 13 indexes of music measurement: danceability-x_1, energy-x_2, valence-x_3, tempo-x_4, loudness-x_5, mode-x_6, key-x_7, acousticness-x_8, instrumentalness-x_9, liveness-x_{10}, speechiness-x_{11}, explicit-x_{12}, duration-x_{13}, popularity-x_{14}.

However, the simple cosine similarity can only distinguish the difference between vectors from the direction, but is not sensitive to the vector amplitude. In order to correct the insensitivity of the simple cosine similarity algorithm to vector modulus, it is necessary to standardize the original data, and then use the cosine similarity algorithm to calculate.The standardized formula is as follows

$$y_i = \frac{x_i - min_{1\leq j\leq n}\left\{x_i\right\}}{max_{1\leq j\leq n}\left\{x_j\right\} - min_{1\leq j\leq n}\left\{x_j\right\}}$$

From this transformation, we get a new sequence y_i.y_i is a dimensionless sequence of normalized parameters.

The answer as follow.

3.2.3 Question 2 solution

1. First of all, we take out the previously processed data for the following calculation.
2. We calculate the similarity between the songs written by a singer and the rest of his songs
3. Calculate the similarity between the songs written by him and other music of the same genre
4. We also compared his songs with works of different schools to calculate the similarity.

Our final comparison table of similarity of the same genre is as follows, which is the average value of 170,000 data of each genre.

Electronic	Religious	Blues	R&B;	Vocal	Country	Pop/Rock
0.857778	0.876166	0.8847905	0.87961	0.892105	0.913821	0.876418

Figure 6: The same genre similarity

Ultimately, we found that the music produced by musicians of the same genre was still more similar

3.3 Question 3

3.3.1 Three Machine Leaining models

BP neural network

BP neural network is developed in the 1980s, the network structure is simple, easy to use. BP neural network adopts the training algorithm of acyclic multi-level network, which has wide applicability. After it was proposed in 1986, it soon became the most widely used multi-level network algorithm. R Hecht-Nielsen has proved that a 3-layer BP neural network can meet the requirements of general function mapping, and the BP neural network with a finite hidden layer can approximate any multivariable function with arbitrary accuracy. BP neural network is composed of input layer, hidden layer and output layer. The way of training is to use a tutor to learn. In general, the input sample set is X, Y : X is the input vector and Y is the ideal output vector corresponding to X. The neuron of BP neural network that receives sample data is the input layer, the output layer that outputs the final result is the output layer, and the hidden layer between the input layer and the output layer is called the hidden layer. The tutor learning process of BP neural network is the training process of it. In other words, after the sample set constituted by sample vectors is input into the network, the connection weight between neurons is adjusted in a certain way, so that the network can store the connotation of the sample set in the form of the connection weight matrix, so that the network can give appropriate output when receiving input. The method of the fastest descent is used for network training optimization. For each neuron, the input is:

$$net = XW$$

$$X = (x_1, x_2, \ldots x_n)\,; W = (w_1, w_2, \ldots, w_n)$$

In this equation, X is the input vector; W is the join weight vector. Activation function (output) of single neuron (node) in hidden layer adopts ReLU function:

$$O = f\,(net) = \begin{cases} net & if\ net > 0 \\ 0 & if\ net \leq 0 \end{cases}$$

The training process of network mainly includes two repeated processes: information forward propagation and error back propagation. In the forward propagation process, the input of the previous layer is weighted as the input of the later layer, i.e. XW. Error back propagation learning refers to adjusting the weight matrix by minimizing the error according to the difference between the actual output O_p and the ideal output X_p, and finally controlling the error within a certain required range. The p sample error can be described as:

$$E_p = \frac{1}{2}\sum_{j=1}^{m}\left(y_{pj} - O_{pj}\right)^2$$

The total error of the sample set is $E = \sum E_p$, until the samples that meet the requirements are found.[3]

decision-making tree

The basic algorithm of the decision tree is shown below

[3]Kang Mengyu, Shao Baorong, Zhu Yueqin, Chen Chen, Wang Tao. *Study on Prediction Method of Landslide Sliding Distance Based on Multiple Nonlinear Regression and BP Neural Network* [J/OL]. Geological conditions: 1-12 [2021-02-09]. HTTP: http://kns.cnki.net/kcms/detail/11.4648.P.20210108.1348.002.html.

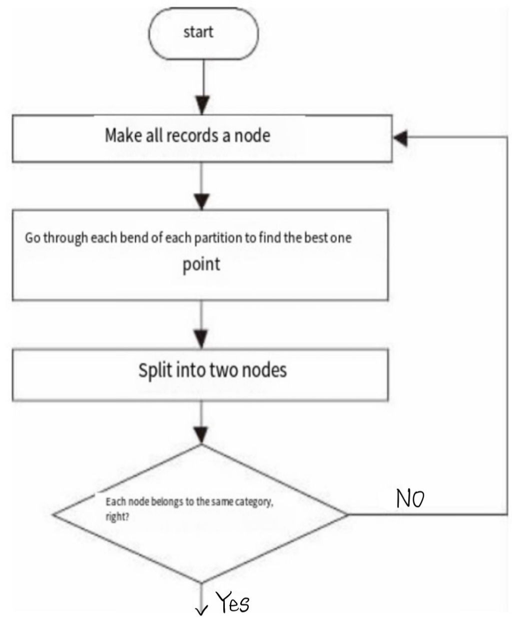

Figure 7: The decision-making tree

LightTGBM

LightTGBM (Light Gradient Boosting Machine) is a framework to implement GBDT algorithm, which supports efficient parallel training, faster training speed, lower memory consumption, better accuracy, and distribution support for fast processing of massive data, etc. LightTGBM has carried out the following optimizations on the traditional GBDT algorithm: decision tree algorithm based on Histogram; Using Gradient-based One-Side Sampling(GOSS) can reduce a large number of data instances with only small gradients and save time and space costs. Exclusive Feature Bundling(EFB) can be used to bind many mutually Exclusive features into a single Feature to achieve the purpose of dimensionality reduction. Leaf growth strategy with depth limitation in leaf-wise; Directly support Categorical Feature; Supported efficient parallelism and Cache hit ratio optimization.[4]

We use BP neural network to calculate the similarity. BP neural network sets four training layers, and its hidden layer parameters are set as 150,100,100,150 respectively. The maximum number of iterations is set to 200. The final training results are shown in the figure

[4]Fang Tingting, Li Quan, Qin Mingyuan.(2020). *Enterprise Credit Assessment and Forecasting Based on LightGBM.* Information Technology and Informatization (12),17-19. DOI: 10.3969/j.issn.1672-9528.2020.12.004

3.3.2 question result

predict outcome	Y genre	energy	speechines	danceabilit	valence	loudness	key	acousticne	year	popularity	tempo	duration_n	instrument	liveness	mode	explicit		
4	4	0.875	0.0378	0.672	0.676	-3.728	0	0.0136	2012	60	118.014	232320	0	0.279	1	0		
4	1	0.735	0.04	0.619	0.95	-7.072	5	0.406	1964	15	125.017	155693	1.39E-05	0.292	1	0		
4	4	0.983	0.0674	0.533	0.382	-2.964	2	0.00118	1998	40	94.938	188973	3.13E-05	0.378	1	0		
4	4	0.43	0.0303	0.598	0.46	-16.621	4	0.132	1993	51	100.012	202840	0.274	0.151	0	0		
4	6	0.794	0.0668	0.641	0.784	-5.586	2	0.465	2016	55	95.032	173613	0	0.269	1	0		
4	4	0.923	0.0689	0.439	0.673	-6.007	0	0.00425	1993	34	92.293	303017	0.0512	0.0974	1	0		
4	7	0.225	0.0915	0.63	0.662	-12.19	10	0.908	1934	0	146.572	211933	0.634	0.103	1	0		
4	4	0.866	0.0351	0.659	0.765	-5.789	8	0.165	1999	42	100.997	233600	3.65E-05	0.0988	1	0		
4	1	0.533	0.0366	0.43	0.574	-10.6	4	0.372	1968	32	186.059	149413	0.000932	0.284	1	0		
4	4	0.0936	0.0503	0.564	0.412	-18.218	7	0.877	1968	22	99.862	82720	0	0.135	1	0		
0.0059341	0.10575984	0.02126843	0.00334799	0.6725122	0.00748941	0.0951237	0.00327911	0.0067403	0.00800376	0.065443	0.00089349	0.00039887	0.00187209	0.00029947	3.38E-05	0.00056682	0.00099632	3.73E-05
0.01562969	0.10980489	0.07821763	0.00489074	0.46317749	0.01750576	0.11656336	0.03479357	0.0233793	0.0276941	0.06408256	0.00762305	0.00364216	0.00975381	0.008913	0.00052974	0.00284907	0.01035884	0.00059122
0.00945342	0.11523692	0.0443849	0.00455067	0.56668508	0.01112803	0.13298227	0.01161928	0.00973611	0.01325761	0.0680304	0.00313709	0.0011529	0.00304567	0.00191581	0.00013919	0.00077173	0.00264828	0.00012465
0.0081541	0.10235019	0.04278852	0.00355735	0.59843292	0.00916955	0.12053761	0.01212803	0.00924958	0.01455742	0.06634404	0.00310493	0.00101695	0.00286373	0.00218063	0.00011693	0.00068126	0.00266167	0.00010459
0.00989658	0.11108809	0.03067899	0.00585679	0.59542902	0.01214352	0.11199031	0.00502074	0.01150542	0.01327519	0.08202249	0.00185958	0.00095206	0.00396483	0.00065027	0.00011205	0.00133676	0.00209218	0.00012514
0.00815173	0.11047464	0.05354469	0.00147451	0.62458946	0.00867174	0.07710333	0.03143406	0.01033056	0.01267109	0.04377861	0.00272914	0.00103739	0.00296767	0.00483811	6.71E-05	0.0008455	0.00522779	6.28E-05
0.01930228	0.10842308	0.09991071	0.00454113	0.42721876	0.02317614	0.08521121	0.07016152	0.02056561	0.0197646	0.05656498	0.0093202	0.00467984	0.00850499	0.01978943	0.00051315	0.00438304	0.01750878	0.00046057
0.00650267	0.09635268	0.03747924	0.00251522	0.63973751	0.00705674	0.10833202	0.01099952	0.00772412	0.01314489	0.06020168	0.0024166	0.00072189	0.00221329	0.00182384	7.18E-05	0.00048084	0.00216189	6.36E-05
0.01199184	0.11581142	0.05319416	0.00613197	0.52006207	0.01435053	0.14158071	0.01442983	0.01269723	0.01655326	0.07411315	0.00449697	0.00179684	0.00444783	0.002827	0.00025809	0.00123058	0.00378746	0.00023908
0.02547991	0.09930894	0.11087576	0.00870522	0.33158683	0.0311362	0.10557618	0.06097074	0.03151813	0.03081663	0.06694499	0.01534434	0.00883009	0.01633964	0.02305087	0.00164407	0.00782888	0.02233617	0.0017063

Figure 8: BP text set

The data segmentation was set at 0.75, that is, 3/4 data in each data set were used for training and the rest were used for verification. Finally, an evaluation model was obtained, which could better explain the genre of a music. We also cross-verified the decision tree and LightTGBM, and the confusion matrix is as follows

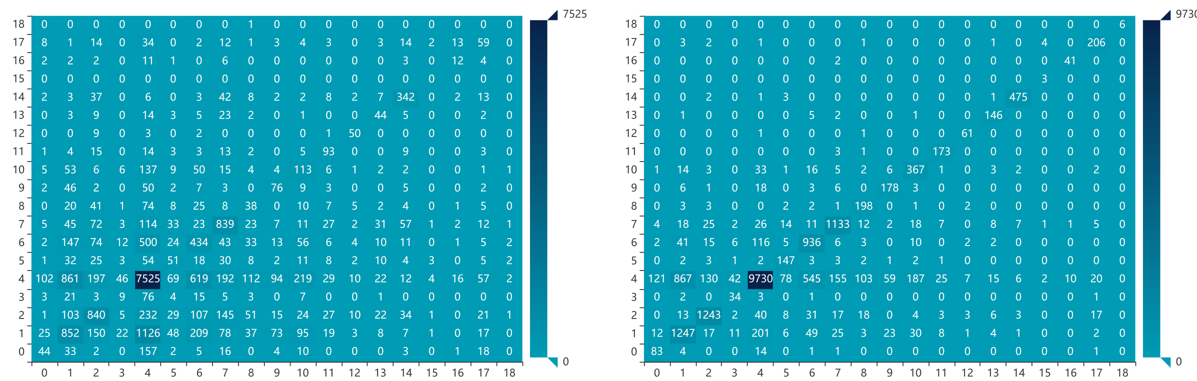

Figure 9: confusion matrix

Through the neural network model, we can identify the similarities between different genres and judge the types of songs

The following photographies is the changing describition of different characteristiss of Classic music:

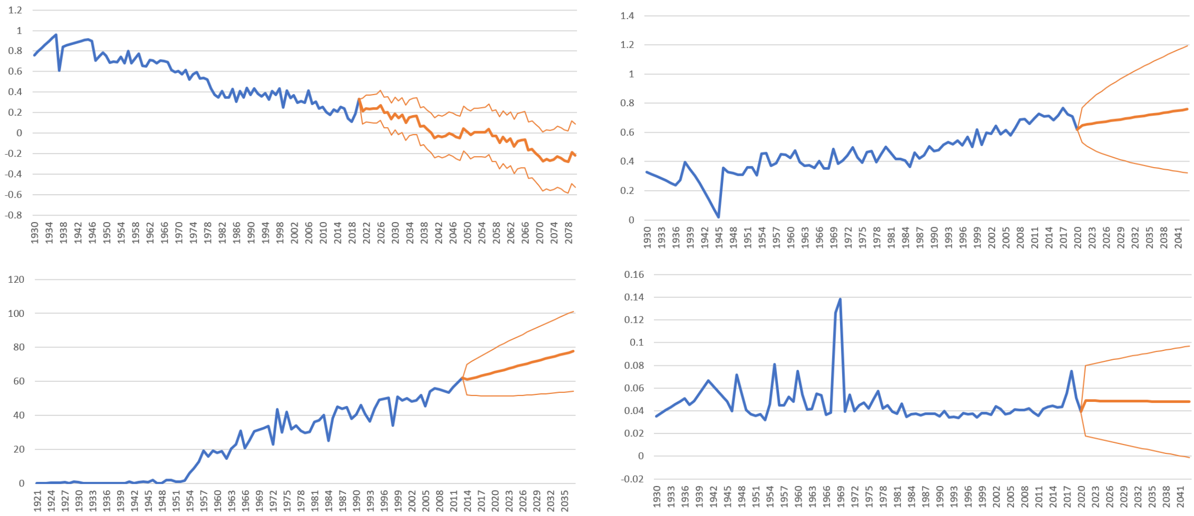

Figure 10: Classical Music

For the acousticness graphic, it is clear that the trend of acousticness is remain decline over decades. And we can find that the energy of the vlassical music is remain increase, after the shake may caused by the WWII. Popularity is increasing among the fluctuation, and the speechiness is fluctuating among the horizontal line.

3.4 Question 4

In order to evaluate the most impressive music characters, we first lists the number of each genre influencers-NI, and the number of followers in their same genre-NF. We introduce the concept of "following proportion" which is a quotient of the two Numbers, and the actual meaning of this concept is music genre of ability, namely the successors of the amplification multiple.

(To our surprise, the number is consistently less than 1, which mean that many artists are good at absorbing the best from several predecessors, or the artists from different genres borrow from each other, with a large group of artists preferring their own original music.)

Then, we introduce a concept of "counting ratio", that is a ratio of the number of influencers in one genre to all influencers, which can represent the group size of influencers for a genre to a large extent. In the end, we calculated the “promotion ratio” by subtracting the counting ratio from the following ratio to measure the inheritance influence of the genre. We then came up with the top 5 genres of influence across all genres:

Reggae, Country, Spoken, Latin and Religious.

· These music has a rural flavor and is close to the daily life of the people in their respective areas.
· They are inclusive, or blend of several other musical genres.
· Distinctive and impressive

3.5 Question 5

For question 5, we look at major turning points in musical characteristics over time. Principal component analysis(PCA) was used to first reduce the dimensionality of the data, reducing several music feature data to one dimension, which was the final analysis factor. This makes it easier to analyze how music changes over time. The PCA method is described as follows:

PCA, as one of the most commonly used data dimensionality reduction algorithms, can also be regarded as a multivariate statistical analysis method to grasp the main contradiction of things, and is the most commonly used feature extraction method, which has been paid attention to and studied by people. It greatly simplifies the difficulty and complexity of problem processing by processing the original data. From the perspective of probability and statistics, the variance of a random variable represents the information it contains, and the larger the variance value, the more information it contains. If the variance of a variable is zero (constant), the variable does not contain any information.

3.5.1 Algorithm steps of principal component analysis

1. Calculate the mean vector μ of the samples in the sample data set X, $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
2. The mean of each sample is de-averaged, that is, the sample data is centralized, that's $\widetilde{X} = X - \mu$
3. Construct the covariance matrix V,$V = \frac{1}{n}\widetilde{X}\widetilde{X}^T$ of data matrix $\widetilde{X}$
4. Carry out eigendecomposition of matrix V, obtain eigenvalue λ_i and corresponding eigenvector w_i, and arrange eigenvalue λ_i in descending order
5. According to the contribution rate, take the first d eigenvalues $\Lambda = diag\,[\lambda_1, \lambda_2, \ldots, \lambda_d]$ and the corresponding eigenvector $W_d = [w_1, w_2, \ldots, w_d]$ as the basis of the subspace, then d principal components to be extracted are $F = W_d^T \widetilde{X}$
6. Reconstruct the original data $X = WF + \mu$ from the extracted principal components

3.5.2 Answer and Explain

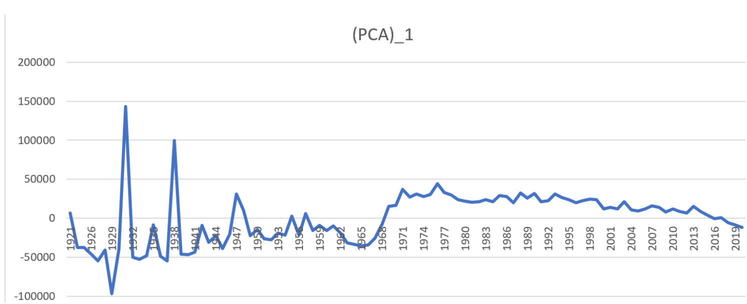

Figure 11: The explanation of cosine similarity

By analyzing the graph of the change of the data over time, we can find that there was a significant turning point in 1929,1939,1965. This was a period of dramatic changes in the social environment. People expressed their resistance and dissatisfaction through music, which led to the revolutionary change of music style.

In 1929, the United States had an economic crisis. This disaster, which lasted several years, later had a profound impact on every corner of the world. In 1930, with people's disappointment in reality and yearning for spiritual life, our factor value reached a wave peak.

In 1939, World War II broke out. With the gradual warming of the war, the value of the impact factor climbed to one of the peaks

The 1960s and 1970s were the most tumultuous and chaotic period in the United States. This country has witnessed many movements, such as the black civil rights movement, the new left movement, the counterculture movement, the feminist movement, the anti-war movement... The form of American music in the 1960s and 1970s was a prominent musical form after the Second World War in the United States. The form of American music in this period had an important milestone significance in the development history of western music, surpassing its own musical form and having social significance. For example, rock music is no longer a simple form of music, but a voice of resistance and Shouting to express people's dissatisfaction with the society in this background period. In the development process of American music, due to the unique regional characteristics, ethnic characteristics and many other factors, there will be a lot of music styles unique to the United States in the development process. Combined with the special background of this period, many musical forms with special characteristics and special significance developed. As a result, American music in this period was deeply loved by the public and played a great role in promoting the development of American pop music and even world music.

3.6 Question 6

To solve this problem, we chose the Pop/Rock genre to analyze, and selected the top 6 artists of the music influence of the genre as the representative figures of the genre. The artist IDs of these six artists are [754032], [66915], [894465], [531986], [139026], and [354105].

Applying the file influence data, we screened out all the followers of these artists and classified them according to the time when these followers started to create. The number of followers in each decade was counted, and the following figure is obtained. And the data simply represent the influence of influencer.

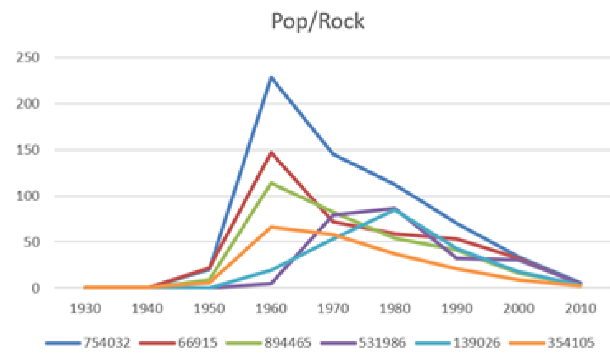

Figure 12: Classical Music-energy

The full music data file was applied to screen out all the songs created by these artists (excluding the influence of co-creation on the music style). For this file, we also conducted the following operations:

1. Classify the music creation year and count the number of songs created by artists in each decade (for example, songs created in 1965 are classified as 1960 and songs created in 1988 are classified as 1980 for corresponding data in influence data)

2. Draw and analyze the images of various musical features that change over time, check whether there are musical features that change significantly over time, and focus on the features that change significantly. Through the above operation, we find several types of features closely related to the change of artists' influence, which can be divided into one main feature and two secondary features. Key Features: Number of songs released by the artist We find that the changing trend of artists' influence is similar to the changing trend of the number of songs released

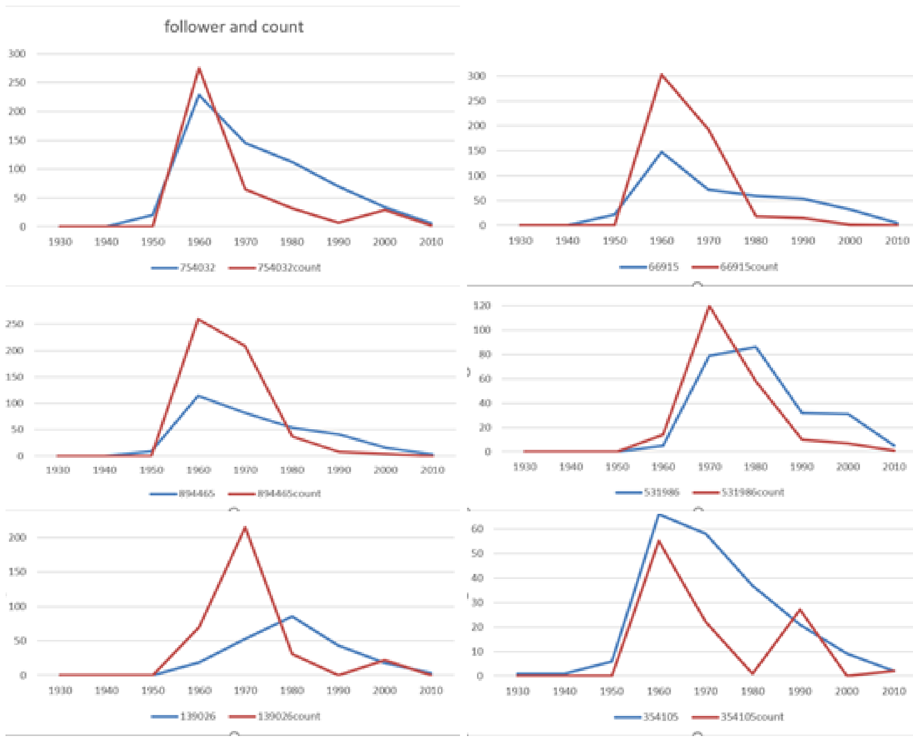

Figure 13: Classical Music

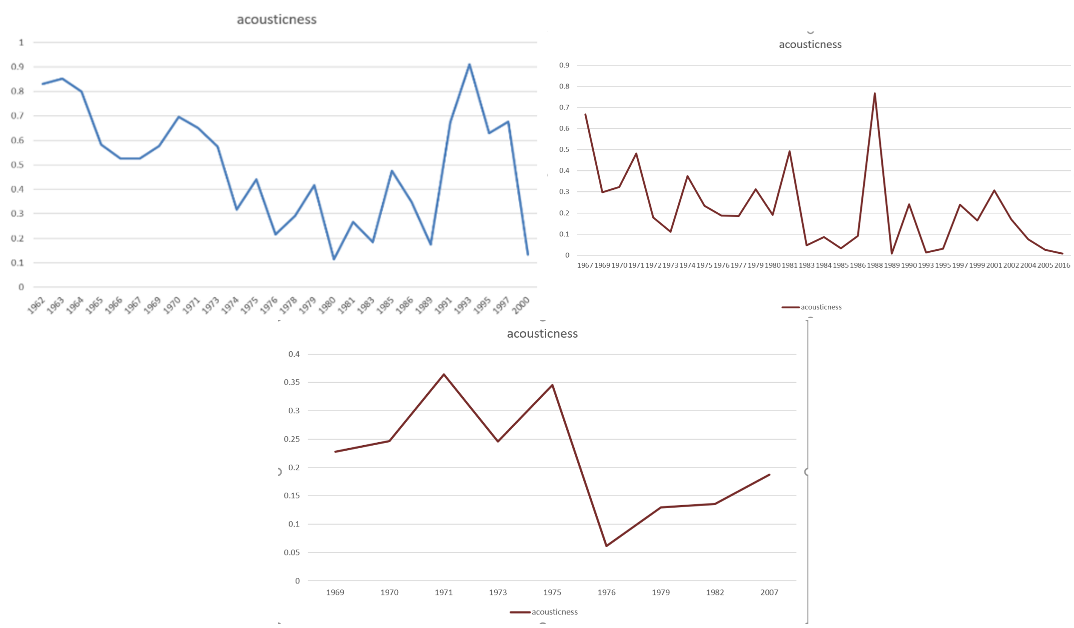

Figure 14: acousticness1-3,[66919],[531986],[139026]

Secondary features: 1 acousticness

With the development of time, this music feature showed an obvious decline around 1978, which indicated that with the development of time, the intensity of electronic processing of the music soundtrack was strengthened around 1978 [5]

2instrumentalness

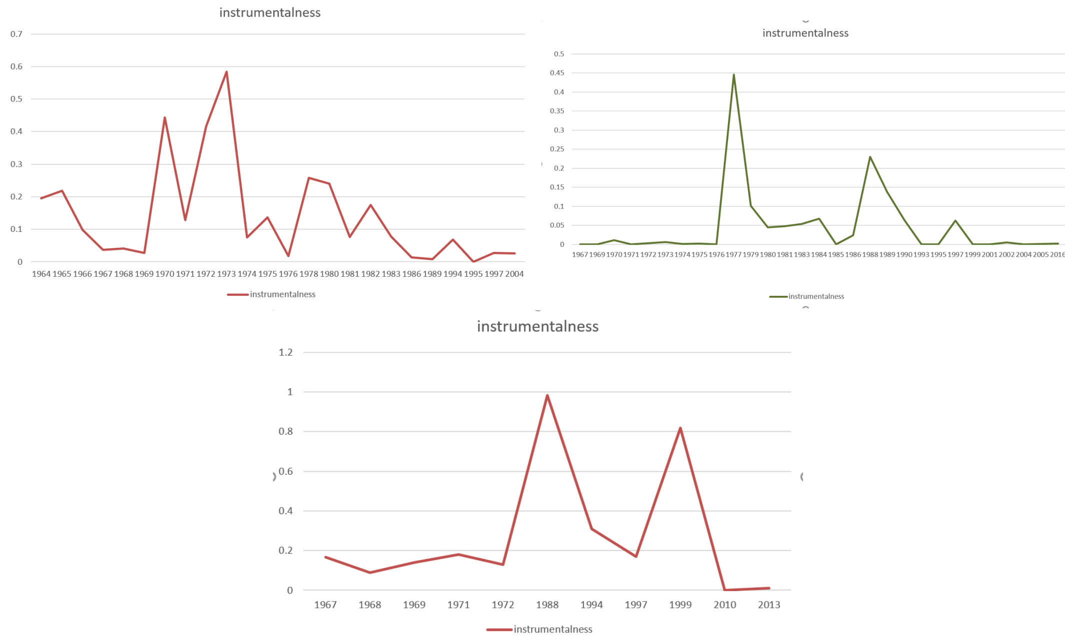

Figure 15: instrumentalness1-3,[894465],[531986],[354105]

This music feature is characterized by low data statistics at the beginning and the end of statistics, and high data statistics at the middle, which is roughly similar to the changing trend of the number of influential people, which reflects the high frequency of the use of Musical Instruments in the music of these artists in the period of their greater influence

3.7 Question 7

We can analyze the relationship between the changes of some characteristics of certain genres of music over time, and look for the time when these characteristics mutated. We believe that these mutated times may have been significantly affected by the external environment. We can then look specifically for historical events that had a significant impact on the music at that point in time. Let's take the Poprock school as an example to show what we think.

Through the study of the changes of the characteristics of popular music over time, we found that the curve showed a jump in 1965. We verified the data and found that the social changes in this era were as follows:

[5]which may be related to the development of electronic music.

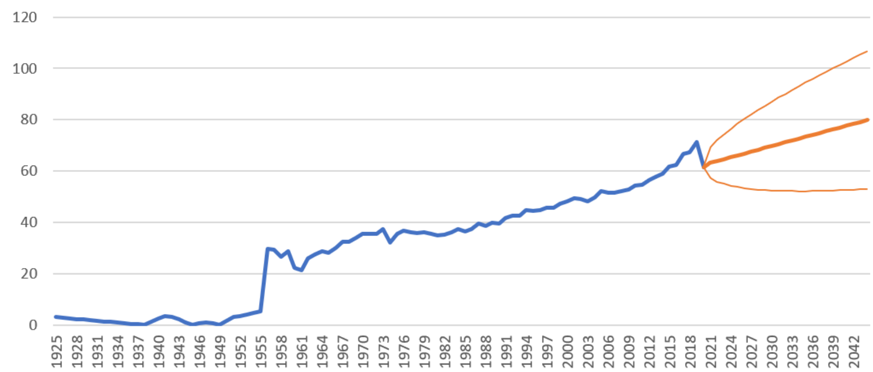

Figure 16: Pop Music-Populrity

With the recovery of the American economy after World War II, the productive forces of science and technology developed greatly, and the material life was greatly enriched. This was in sharp contrast to the oppressive political and cultural atmosphere during the Cold War between the United States and the Soviet Union. For the young generation of the middle class who have individual pursuit, carry good ideals and long for a comprehensive and free way of life, it is intolerable and acceptable, should be resolutely resisted. So in the 1960s there was the counterculture movement. Many young people choose to stay away from politics and place their spiritual sustentation elsewhere, putting their feelings of pessimism and dissatisfaction into words and even action. They abandon the traditional life credo with a cynical attitude to life and live a disinhibited life. Pop rock music also emerged from this time.

Through the study of the changes of the characteristics of popular music over time, we found that the curve showed a jump in 1970. We verified the data and found that the social changes in this era were as follows:

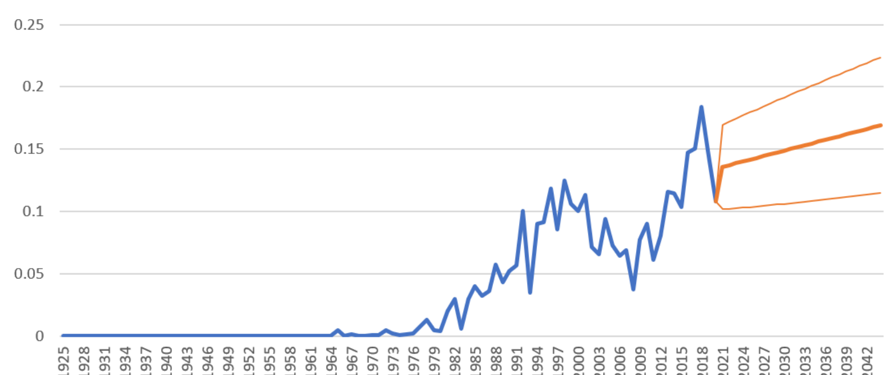

Figure 17: Pop Music-Explicit

The punk movement of the 1970s was also based on young people's rejection of old social norms, but this time it wasn't just shoulder pads or grease heads. The new punk bands were in the vanguard against convention: they not only railed against what they saw as the false values of a rich but dishonest society, but they overtly disrupted them through behavior, dress, noise, and offensive language. One or two of them had the firepower to challenge the established norms of the 1970s, the most important being the Sex Pistols.

4 coclusion

Based on a given data set, our team has mapped the network between influencers and followers, which can intuitively reflect the connection between influencers and followers. A music influence index is developed, which can eliminate the influence of genre size and show the influence (status) of influencers in their genres.
By using cosine similarity as an evaluation index, we compare the music similarity between different genres and between agreed genres. The results show that the music similarity between the same genre is higher than that between different genres, and the music similarity between the same genre is more similar.
Through BP neural network, decision tree and LightGBM network model, we have carried out three different types of machine learning. These three models support each other and successfully distinguish the music characteristics of different genres. It is worth mentioning that through machine learning, we can also classify the genre of songs. At the same time, we also made a detailed analysis of the musical characteristics that changed over time in each genre.
We introduce the concept of probability of followings into the model. Based on this parameter, we find that influencers do influence the music creation of their followers. For some music genres with high probability of followings, we believe that influencers have stronger appeal.
By introducing principal component analysis (PCA) into the model, we successfully discovered the time point when music changed, and searched for relevant literature to explain it.
Instruments and Instrumentalness are the factors that influence the number of fans in Pop/Rock genres, according to our team's analysis of the music produced by the genre's major influencers.
Finally, we extract some musical features that vary significantly at certain points in time and analyze the social and environmental causes behind them. Using this as an example, we demonstrate the methods our model uses to explore the social and historical factors that influence the development of music.

5 Strengths and Weeknesses

Straight: 1.the composite model constructed by our team can analyze the history of modern music in a comprehensive and specific way
2.We can automatically categorize songs into genres

Weekness: the principal component analysis used in the third question reduces the data from 14 dimensions to 1 dimension, which may lead to loss of information

6 Document to ICM

Document to ICM

In the model established by our team, we first built a network that connects influencers to followers. Through this network, we can intuitively explore the relationship between influences and followers among artists as well as the relationship between secondary influences and followers. Then we construct the evaluation index of music influence, which can clearly reflect the influence status of artists in their genres. The more influential the music, the higher its status in the genre.

In our model, we use cosine similarity to calculate the similarity between music within the same genre and between different genres, and find that artists within the same genre are more similar than artists at the top of the genre.

Our team also in the model using the BP neural network classifier, LightGBM classification and three kinds of classification decision tree classification, through the way of machine learning, identify the characteristics of all kinds of schools, because we have identified the various genres of music characteristic, the model has been able to our team to a song have been confirmed music features for genre classification.

We classified and summarized the full_music_data.csv file to extract the trend of the development of various musical characteristics of each genre of music over time, and made careful analysis to determine the change rule of the genre over time

By counting the proportion of influencers in each genre to the total number of influencers and the proportion of followers influenced by the same genre to all followers of the genre, our team determined the uniform measure of growth ratio, and found that the uniform measure of growth ratio was generally larger. Based on this, we think that influencers actually influence the music that followers make.

Our team through principal component analysis (pca), a variety of music characteristic variable dimension for one dimension, to establish the relationship between principal component variables change over time, and drawing on this and the timing of the larger changes in data, we think it is a great leap in the process of evolution of music, our team for this type of time to consult the relevant historical documents, analyzes the causes of these changes.

Our team choose Pop/Rock this music genre, and select the genre of several main effect, characteristics of music and the adjustment of the influence of the change of the correlation analysis, sums up the main factors: the influence of a number of songs are released and two secondary factors: acousticness and instrumentalness, and the reasons of the effects are analyzed in the appropriate. It explains how the influence of artists changes over time.

Our team has also created a way to identify environmental factors that have a strong influence on music. That is, by looking for the changes of certain musical features with the development of time, and for the time node where the mutation occurs, searching for the changes in the external environment at a specific time node, so as to find the environmental factors affecting the development of music

References

[1] CHEN Da-li, SHEN Yan-tao, XIE Bing-zhu, MA Ying—yi. *A Measure Model of Similarity for Finding the Best Coach.* [J]Journal of Northeastern University(Nature Science), 2014.12

[2] Kang M Y, Shao B R, Zhu Y Q, Chen C, Wang T.*Comparison of multiple nonlinear regression model and BP neural network model for landslide slip distance prediction.* [J/OL]. Geological Bulletin of China:1-12[2021-02-09]. http://kns.cnki.net/kcms/detail/11.4648.P.20210108.1348.002.html.

[3] FANG Ting-ting ,LI Quan & QIN Ming-yuan.(2020). *Enterprise credit evaluation and prediction based on LightGBM.* Information Technology and Informatization (12),17-19. DOI: 10.3969/j.issn.1672-9528.2020.12.004

[4] Chen Pei.(2014). *Research on Principal Component Analysis and Its Application in Feature Extraction* (Shaanxi Normal University)https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201501&filename=1014400060.nh

[5] Yu Mingchuan.(2019). *Research on American musical morphology in the* 1960*s and* 1970*s* (Northeastern Normal University). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201902&filename=1019841140.nh

[6] Zhou Hailong.(2015). *The Punk Storm in Popular Music* (SOOCHOW UNIVERSITY). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201601&filename=1015400996.nh

[7] Lu Xiuyuan. (2010). *BOB DYLAN ROCK 'ROLL ARTISTIC VALUE EXTRICATED* (Zhejiang Normal University). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD2011&filename=2010241285.nh

[8] CHEN Da li,SHEN Yan-tao,XIE Bing-zhu,MA Ying-yi.(2014). *A Measure Model for Similarity analysis of the Best coachmen.* Journal of Northeastern China (Natural Science) (12),1697-1700.doi: 10.3969 / j.issn.1005 – 3026.2014.12.006.

[9] Chen Fuyou & Yang Liying.(2014). *Analysis on a New Scientific Evaluation Indicator.* JOURNAL OF INTELLIGENCE (07),81-85+62. doi: 10. 3969 / j. issn. 1002-1965. 2014. 07. 015.