The Analysis of Chinese Listeners' Preference for European and American Pop Music

Yang Zhang[1,a] , Hui Zhang[1,b], Juanjuan Cai[2,c] , Qi Shen[1,d]

School of Science, Communication University of China, Beijing, 100024,China

Key Laboratory of Media Audio & Video, Ministry of Education, Communication University of China, Beijing, 100024,China,

e-mail: [a]yang.zhang.vera@outlook.com, [b]zhanghui0931@cuc.edu.cn, [c]caijuanjuan@cuc.edu.cn, [d]2890866362@qq.com

Keywords: pop music; music listeners; preference analysis

Abstract. This paper focuses on the music field in the cultural industry and analyzes the preference of the listeners of western pop music in China. This paper captures the data such as the characteristics of singers and audience ratings. Then the clustering of the artists is conducted according to the chosen indicators and the artists are divided into four clusters in which the representatives are Justin Timberlake, Sam Smith, Westlife and M2M respectively. The results will help users find their favorite songs more quickly and accurately, and help QQ music and other music platform provide more accurate and reasonable recommendations for the audience.

1. Introduction

Music is a part of culture, and it is an important way for people to relax and communicate with each other. Personal status and emotion, geographical features, aesthetic preferences will be reflected in it, and each age has its own unique era mark. The McLuhan said, "The Media is The Message"[1], meaning that the media form itself is rich in information. The innovations of communication technology bring great changes for human life and promote the development of social culture. The development of new media represented by the Internet has brought an interactive and personalized way of communication. In this case, in the consumption and use of cultural products, people gradually become looked for what they need actively from passive acceptance. The users mainly take the information retrieval from the perspective of singers or musicians, so this paper makes the analyses from the perspective. Europe and the United States pop music is the trend of the vane of the world music. This paper focuses on popular music in Europe and America, and analyzes the audience's song preferences by the cluster analysis in order to help users locate their favorite songs quickly and accurately. From the music platform and the music industry point of view, entertainment spending is becoming an important component of household consumption expenditure, and the song copyright issues are getting more and more attention, so reasonable recommendation will undoubtedly bring good economic benefits..

2. Data Sources

The data of this paper comes from the QQ music official website, Douban music official website, and Kugou music PC version, with respect to relevant indicators of the Europe and the United States top pop music singers. The three music service platforms are characterized by rich product, widely accepted by users and they are representative, so it can be used as the reliable source of data.

For each singer sample, there are 13 indicators. The indicators can be divided into three categories. The first category includes the first five indicators, such as singer name, singer type, genre style, location, age, which are the basic situation of the singer. Above are the objective characteristics of indicators.

The sixth to the ninth indicators are the second category, including the number of followers of the singers in Kugou music and QQ music, the last three albums' average score of kugou music, the recent album score of Douban music, which reflect the audiences' attitude and preference to the singers. To a certain extent, the two indicators of concern are good reflection of the singers' heat in China and the size of the audience, which are very important core indicators. The last three albums' average score of Kugou music reflects the singers' audience evaluation in a long time. The recent album score of Douban music reflects the popularity. The two audience ratings also reflect the quality of the singer's work. Above are the subjective indicators.

The other four indicators are the third category, such as the release date of the latest album, the number of mvs, the number of singles and the number of albums, reflecting the professional performance of the singer, mainly the degree of activity and the size of the works released.

This paper bases on the index system constituted by the above indicators. Variable classification is summarized in Table 1.

TABLE I. VARIABLE CLASSIFICATION TABLE

Classification	Singer Type	Style	Location	Age	Concerns	Album Time	Singles/MVs
1	Male singer	Popular	America	<30	0-100,000	<100days	0-100
2	Female singer	R&B	Britain	30-40	100,000-3,000,000	Within a year	100-200
3	Combination	Rap	Canada	40-50	≥3,000,000	Within two years	200-300
4		Rock	Others	≥50		More	300-400
5		Electron					≥400
6		Others					

The singer type is divided into male singer, female singer and band combination. It can be seen from the Figure 1 that the number of male singer, female singer and the combination is similar. The audience attention is not much difference. Female singer's average number of concerns is the most, but the combination's concern is the least. This shows that the audience most favored in Europe and the United States female singer, the combination concerned at least.

The singer area is divided into the United States, Britain, Canada and others. The Figure 2 shows that the United States has the largest number of singers, followed by other countries and regions, and then the United Kingdom, Canada singer samples at least. Canadian singers' average number of concerns is the most, which is twice that of the United States and three times that of the UK. That is to say, the singer in North America is more popular.

The singer's style is divided into popular, rhythmic blues, rap, rock, electronics and others. As can be seen from the Figure 3, the number of singers is sorted by: popular, others, rock, electronic, R&B and rap. By the average number of concerns to see the degree of popularity, the most popular is rock and roll. The popularity of R&B, popular, electronic and rap is relatively low.

The singer's age is divided into within 30 years old, 30-40 years old, 40-50 years old and above 50 years old. As shown in the Figure 4, the number of singers is sorted by: 30-40 years old, less than 30 years old, 40-50 years old, above 50 years old. The popularity of the singer decreases with the increase of the age of singers. People who are less than 30 years old get the most attention.
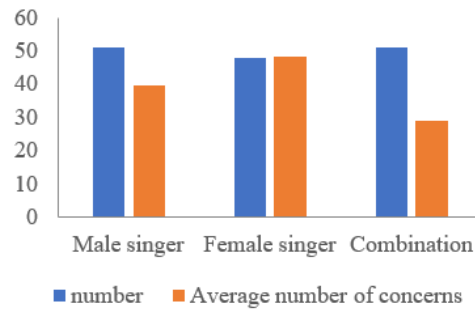
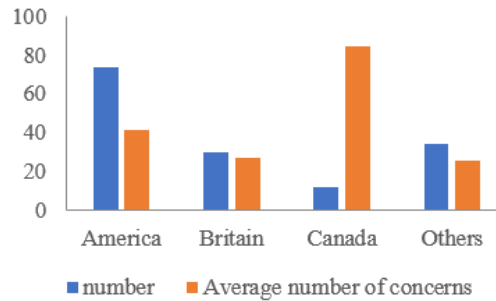Figure 1.   Singer type and average number of concerns

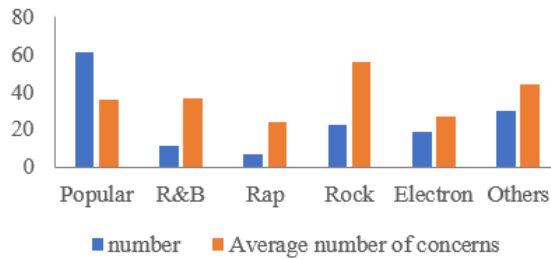Figure 2.   Location and average number of concerns
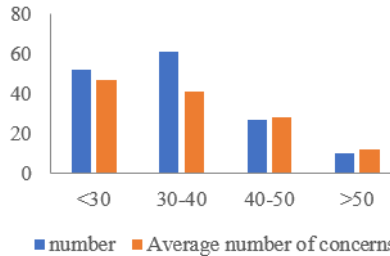

Figure 3.   Style and average number of concerns

Figure 4.   Age and average number of concerns

3. Model Building

Cluster analysis is one of the important methods of data analysis, which is helpful for solving practical problems. For example, solving the problem of customer segmentation is a good application of this method.

This paper applies clustering analysis to the classification of selected singers. Artificial classification has a strong subjectivity, but clustering analysis bases on multiple observations and the data is divided

into meaningful or useful groups. The greater the similarity in the group, the greater the difference between groups, the better the clustering effect [2].

Samples (or indicators) are sorted according to the data matrix in the table below. x_{ij} represents the value of the j-th index of the i-th sample. Each sample has p indicators, so each sample can be seen as a point in the p-dimensional space. N samples make up n points in p-dimensional space. Use distance to measure the similarity between samples.

x_{ij} represents the value of the j-th index of the i-th sample; the mean and standard deviation of the j-th index are denoted by $\bar{x}_j$ and S_j, d_{ij} means the distance between the i-th sample and the j-th sample.

Minkowski distance:.

$$d_{ij}(q) = [\sum_{k=1}^{p} |x_{ik} - x_{jk}|^q]^{1/q}$$

(1)

When q tends to infinity, the distance is called the Chebyshev distance.
Chebyshev distance:

$$d_{ij}(\infty) = \max_{1 \le k \le p} |x_{ik} - x_{jk}|$$

(2)

TABLE II.   DATA MATRIX

No.	X_1	X_2	...	X_p
1	X_{11}	X_{12}	...	X_{1p}
2	X_{21}	X_{22}	...	X_{2p}
⋮	⋮	⋮		⋮
n	X_{n1}	X_{n2}	...	X_{np}

The K-means clustering algorithm used in this paper is called fast clustering method, which directly reflects the important characteristics of the method, which belongs to the non-pedigree clustering method and is based on the division method. K-means clustering is divided into the following three steps:

(1)Select the initial K centroid from the data set, which is the number of classes we want to get. Then calculate the distance from the remaining elements to the center. Each point will be assigned to the nearest center of mass;

(2) Recalculate the centroid of each class until the criterion function used in the clustering algorithm converges and reassemble all the elements according to the new center;

(3) Repeat step (2) before the clustering results do not change.

Distance formula:

$$d_{ij}(2) = [\sum_{k=1}^{p} (x_{ik} - x_{jk})^2]^{1/2}$$

（□ ）

Each sample has p indicators, means the first k indicators of the i-th sample, means the distance between the i-th sample and the j-th sample[3].

4. The Classification of the Singers

This paper uses K-means clustering method to cluster analysis of all observations, integrate singer attributes and audience attitude indicators, and classify singers so that there are many similarities between singers who are assigned to the same class. The purpose of the singer classification is that each type of singer has its own specific audience. In the audience's music recommendation, you can get a certain application.

In the K-means clustering, the number of classes K can be set in advance, and can be identified in the clustering process. In this paper, let K take values from 2 to 6. Then take fast clustering and come to

conclusion. The ratio of the sum of group distance squares and the sum of the overall distance squares is determined by the change of k value. Determine the value of K (the number of classes) and the clustering results[4].
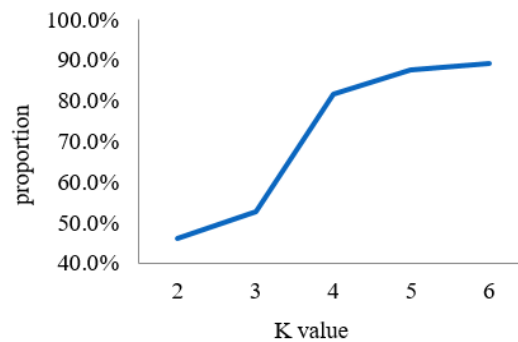

Figure 5.   Ratio changes with the value of K

Figure 5 shows the ratio of the sum of squares of the distance between groups and the sum of the squares of the total distance. With the increase of the K value, the ratio also increases gradually.

When the value of K is 4, the ratio of square sum is larger and the growth rate is the most obvious. The ratio is 81.6%, which indicates that the distance between each cluster is larger. The increase is 55.13%, which shows that the clustering effect is improved remarkably.

When K is greater than 4, the ratio of squared sum will increase by 14.6%, 7.48% and 1.6% respectively, and the clustering effect may be better, but the increase is not obvious. Taking into account the problem of excessive number of categories will exist, the final determination of the K value is the abscissa at the inflection point in the figure. The best result is all the samples into four categories.

When the K value is 4, the proportion of the whole group is about 24.67%, 59.33%, 15.33% and 0.67% respectively. The number of singers in the second category is the largest. The fourth class singer is the least, only one bit, but by K-means outlier detection, the point is not outlier.

The results are as follows：

The first category includes：Justin Bieber, Pitbull, Eminem, Wiz Khalifa, Usher, Justin Timberlake, Chris Brown, Flo Rida, Akon, David Guetta, Jason Mraz, Drake, Jason Chen, Lil Wayne, Rihanna, Taylor Swift, Kelly Clarkson, Katy Perry, Mariah Carey, Beyonce, Britney Spears, Shakira, Lana Del Rey, Nicki Minaj, Madonna, Jennifer Lopez, Kylie Minogue, Linkin Park, Coldplay, Two Steps From Hell, Queen, the Rolling Stones, Bon Jovi, Green Day.

The second category includes: Bruno Mars, Adam Lambert, Calvin Harris, Sam Smith, the Weeknd, Ed Sheeran, Shayne Ward, Darin, James Blunt, Avicii, Charlie Puth, Enrique Iglesias, Alan Walker, Tez Cadey, Troye Sivan, Zedd, Christopher, Timbaland, Shawn Mendes, Frankmusik, Daniel Powter, Chris Medina, Zayn, John Legend, Martin Garrix, Adele, Carly Rae Jepsen, Lenka, Ellie Goulding, Ariana Grande, Sia, Jessie J, Colbie Caillat, Ke$ha, Meghan Trainor, Miley Cyrus, Demi Lovato, P!nk, Selena Gomez, Rachel Platten, Birdy, Iggy Azalea, Leona Lewis, Enya, Lorde, Bebe Rexha, Rita Ora, Halsey, Tove Lo, Fifth Harmony, Take That, Imagine Dragons, Major Lazer, Far East Movement, Twenty One Pilots, Black Eyed Peas, Rudimental, Clean Bandit, Michael Learns To Rock, Passenger, Ylvis, Simple Plan, Walk Off the Earth, Lukas Graham, Little Mix, DNCE, Sum 41, Eagles, Magic!, the Vamps.

The third category includes: Kenny G, Tobu, Mark Ronson, John Denver, Avril Lavigne, Maria Arredondo, Dido, Lene Marlin, Westlife, Groove Coverage, Backstreet Boys, Blue, Nightwish, Rixton, Carpenters, the Script, the Wanted, Fun, Nickelback, t.a.t.u, the Cranberries.

The fourth category only includes M2M.

5. Conclusion

In the first category, American singers and pop singers accounted for the majority. It includes a number of singers who have debuted for many years with a high reputation and particularly large audience. As long as they have new songs issued, the new songs will be quickly spread through the Internet and other channels, and will be on the list of music list and even become the top spot.These singers include Justin Bieber, Taylor Swift, Madonna, Linkin Park and so on.

The second type of singer is the largest number. The number of male singers, female singers and the combinations is almost the same. They all have issued a new album within two years. But most of the singers do not have much work with small scale. Many of the singers are young and be well-known in recent years, such as Sam Smith, Zayn, Charlie Puth, Adele and so on. Charlie Puth sang the theme song of movie "Speed and Passion 7" in 2015, and became an overnight star. 29-year-old Adele won five awards for the Grammy Awards this year.

The third category contains fewer singers. They are mostly old singers and band combinations, once all the rage. They are less active in the near future, with no new work issued within a year. But it does not mean that such singers have been out of the market. They have their own listeners and fans, and is very valuable in the similar singers' recommendation, such as Westlife, Backstreet Boys and so on.

The fourth category contains only one singer M2M. This is a combination of women's bands from Norway. They had high popularity and high degree of activity, but has long been dissolved without issuing new works.

For the audience, they may have a preference for a class of singers, and when there is a choice, songs will be compared in each category. The songs copyright issues get more and more attention today, so many songs need to pay or join the VIP to download their work. In view of the conclusion drawn from this study, this paper gives some reasonable suggestions from the perspective of audience and music service provider.

First, the audience can learn from the analysis of the audience preferences, access to meet their own unique needs of a reasonable recommendation, greatly reducing the time cost of searching songs and singers.

Second, for the music service provider, people often pay attention to their recommended modules, so they should focus on the audience preferences according to the status quo for the audience preferences.

Acknowledgement

References

[1] Marshall McLuhan: Understanding Media: The Extensions of Man. McGraw-Hill.(1964)

[2] WeiZhou: Research on Customer Group Consumption Behavior Based on Clustering and Association Rules.Chengdu.(2015)

[3] Xiaoqun He: Multivariate Statistical Analysis. Chinese Renmin University Press.(2008)

[4] Liuwei Xing: Application Research of K-Means Algorithm in Customer Segmentation. Chengdu. (2007)