Framework for Building Knowledge Map of Ethnic Music Based on Big Data

Jinglin Peng¹,a,*, Jianfeng Zhou²,b

¹College of Humanities and Arts, Nanyang Agricultural Vocational College, Nanyang, Henan, China
²Nanyang Lida Optic-electronics Co., Ltd., Nanyang, Henan, China

¹pengjinglin330@sina.com, ²330869112@qq.com

*Corresponding author

Keywords: Knowledge Map, Big Data Analysis, National Music, Co-word Analysis

Abstract: National music is a cultural treasure with unique charm and charm in the traditional Chinese culture, which has high research value and broad social influence. This paper aimed to explore the construction of ethnic music knowledge map based on big data analysis. This paper proposed the knowledge map of ethnic music and big data clustering analysis, and studied the experimental results of constructing the knowledge map of ethnic music based on this research. The experimental results of this paper showed that the knowledge map provided a scientific framework and method for the exploration of ethnic music knowledge. It made a reasonable explanation and evaluation for researchers in the field of ethnic music and pointed out the direction for the development of ethnic music. This paper identified 24 high-frequency keywords, which are the basis of co-word analysis. Among them, “national music” appeared 1940 times, and “ethnomusicology” appeared 465 times. “Folk music” appeared 415 times, and “music tradition” appeared 276 times. “Chinese music” appeared 270 times. Multivariate statistical methods are often used in co-word analysis. These are the central links in co word analysis. Clustering analysis was used to classify keywords in ethnic music, thus revealing the current hot topics in ethnic music. In a word, the construction framework of ethnic music knowledge map based on big data analysis is conducive to the development of ethnic music.

1. Introduction

National music is one of the key objects to protect the excellent traditional cultural resources of Chinese minorities, and it plays an important role in the traditional Chinese culture. China has a long history and culture, which is the common wealth created by all ethnic groups. In order to better protect the diversity of ethnic cultures, ethnic music should be included in the World Intangible Cultural Heritage List and continue to grow. Chinese folk music has a history of thousands of years. It is not only a cultural form of creation and dissemination, but also a humanistic phenomenon. It is a form of expression and art of the collection of ethnic groups as a unit with a strong national style and characteristics of the times. With the constant changes of the times, national music plays an increasingly important role in people’s lives. Its rich and unique content and value still have great
practical significance in today’s society. Therefore, strengthening the study of national music has become one of the new hot issues in the academic research.

According to the existing research progress, different researchers also have corresponding cooperative research on knowledge mapping. Hogan Aidan gave a comprehensive introduction to knowledge mapping. In the case of dynamic and large-scale data collection, knowledge mapping has become an important research content in the industry and academia [1]. The concept based on knowledge map represents the structural relationship between entities, which is an important research field of human cognition and intelligence. Ji Shaoxiong mainly introduced the basic content of knowledge map, including: learning of knowledge map; knowledge acquisition and completion. The application of knowledge was learned, and the recent progress and prospects were reviewed to provide reference for future research [2]. In order to further enhance the practicability of the knowledge map, people proposed many improved algorithms to try to reason the missing knowledge, add it to the map, or identify it. Paulheim Heiko studied the method of refining such knowledge map and the evaluation method adopted [3]. Al-Saleem Jacob established the biomedical knowledge map and identified 1350 host proteins that may be used to treat new coronaviruses and small molecules in the process of disease [4]. However, these scholars lack some technical argumentation on knowledge map exploration, and there are better findings on knowledge map research based on big data analysis, so this paper reviewed the relevant literature on big data analysis.

Some scholars also have some research on big data analysis. Ma Li aimed to discuss the inheritance and development of Chinese traditional music in the era of big data, examined the existence from the perspective of data analysis and ecology, and conducted in-depth discussions on it, so as to establish the cultural resource library in the era of big data [5]. Hamilton Craig outlined the relationship between streaming media services, online channel data collection and automatic recommendation systems, and how they together cause people to participate in music receiving behavior [6]. However, these scholars did not explore the construction of ethnic music knowledge map based on big data analysis, but only unilaterally discussed its significance.

This paper made the following conclusions by studying and analyzing the experimental results of constructing the knowledge map of ethnic music: the construction framework of ethnic music knowledge map based on big data analysis provided a scientific theoretical framework and method for the mining and excavation of ethnic music knowledge, made a reasonable interpretation and evaluation for researchers in the ethnic music industry, and pointed out the direction for the development of ethnic music.

2. Construction Method of Ethnic Music Knowledge Map Based on Big Data

2.1 Knowledge Map of National Music

Knowledge map is an advanced information processing technology, which can provide an accurate, intuitive and effective tool for knowledge structure construction [7-8]. When a knowledge thing spreads from a simple node to its periphery, it can form a knowledge network. When the knowledge map contains more information, the network boundary can be clearer and the distance between nodes can be farther. This kind of network is called knowledge network, and the knowledge map is a network model used to describe various knowledge structures [9-10]. Because knowledge is composed of different structures, when a knowledge point or knowledge base contains a lot of knowledge, it has a complete structure. This structure is composed of the conceptual structure, attribute structure and network structure represented by each node, and the knowledge map is a knowledge network structure composed of these [11-12]. In order to understand the knowledge structure and information integration, it is necessary to provide a very important but
complex and flexible tool to deal with massive knowledge and build knowledge networks and models. The establishment of national music knowledge map can provide guidance for solving the problem of music resources [13-14]. Therefore, it is necessary and meaningful to build a knowledge map of national music through knowledge map technology and use knowledge to predict the future development trend and direction. Figure 1 shows the classification of national music. The knowledge atlas of ethnic music has also attracted the attention of researchers, which has been widely used in the field of music research.

![Figure 1: Classification of folk music](image1)

2.2 Cluster Analysis of Big Data

With the continuous deepening of big data analysis, people find that traditional analysis methods have limited ability to analyze the music knowledge map, so data mining clustering and prediction methods are needed. Big data analysis music knowledge map can conduct multi-dimensional statistics on the collected music files based on user behavior data, knowledge structure and network structure. Through different data mining clustering algorithms, more music knowledge map data can be analyzed and obtained. Figure 2 shows the analysis of music knowledge map based on big data.

![Figure 2: Analysis of music knowledge map based on big data](image2)

The music knowledge map can be used not only for knowledge mining in the music industry, but also for prediction of the knowledge map [15]. In terms of prediction, the traditional music knowledge map usually has some big defects, mainly reflected in the fact that some key nodes cannot be accurately predicted in the construction process. Of course, after the prediction, it still needs some time to revise and improve, ensuring that the whole music world can be more perfect.

The purpose of clustering is to divide a group of data without classification marks into several classes or clusters, so as to reveal the internal structure hidden in the data. So far, there is no unified
concept of clustering. Most researchers use the method of similarity to segment data, so that different data objects are similar in different categories (clusters).

Assuming that the dataset to be clustered is $C = \{c_1, \cdots, c_M\}$, and the purpose of clustering is to divide dataset $C$ into $l$ classes $V_1, \cdots, V_l$, where $V_o \subseteq C$ meets the following conditions:

$$\begin{cases}
    \bigcup_{o=1}^{l} V_o = C \\
    V_o \cap V_k = \emptyset, \forall o \neq k \\
    V_o \neq \emptyset
\end{cases} \tag{1}$$

The above cluster is used to describe hard partitions. Corresponding to hard partition, fuzzy segmentation refers to dividing each data object into several classes according to a certain degree of dependency. In the past decades, the research of clustering has made great progress, and various types of clustering algorithms emerge in endlessly. The differences are mainly manifested in similarity calculation and mathematical models.

Similarity refers to the degree of similarity or distance measurement defined between data objects. It is an important issue in cluster analysis, which directly affects the quality of clustering. In clustering algorithm, Minkowski distance, Mahalanobis distance and cosine similarity are generally used to calculate the similarity.

(1) Minkowski distance

It is supposed that any data object $c_o = \{c_{o1}, \cdots, c_{of}\}$ in C is composed of the characteristics of f, and the Minkowski distance between two arbitrary objects $c_o$ and $c_k$ in C is defined as:

$$f_z(c_o, c_k) = \left(\sum_{l=1}^{f} |c_{ol} - c_{kl}|^z\right)^{1/z} \quad (z > 0) \tag{2}$$

In the above formula, $z$ takes different values, and Minkowski distance also corresponds to different forms. When $z$ is taken as 1, it corresponds to the Manhattan distance in the following formula:

$$f_1(c_o, c_k) = \sum_{l=1}^{f} |c_{ol} - c_{kl}| \tag{3}$$

The distance from Manhattan is also called $L_1$, or “urban distance”. When Manhattan distance is used to calculate the distance between objects, it can move forward along the grid lines, just like walking on the streets of the city.

When $z$ is taken as 2, the Minkowski distance corresponds to the Euclid distance, as shown in the following formula:

$$f_2(c_o, c_k) = \sqrt{\left(\sum_{l=1}^{f} |c_{ol} - c_{kl}|^2\right)} \tag{4}$$

Euclid distance, also known as $L_2$ type, is the most familiar and widely used method for measuring distance, namely, the “distance” between two points in practice. Among many clustering and classification algorithms, Euclid distance is the most common method to calculate similarity.

When $z$ tends to infinity, it is type $L_\infty$. At this time, only the characteristic quantity of the maximum distance can really play a role. Therefore, the Minkowski distance is the largest one of the $l$ characteristic quantities, namely the Chebyshev distance, as shown in the following formula.

$$f_\infty(c_o, c_k) = \max_{1 \leq l \leq f} |c_{ol} - c_{kl}| \tag{5}$$

Minkowski distance is a relatively easy to understand method, but its disadvantage is that the scale or measurement unit of different attribute dimensions can greatly affect the clustering effect. Therefore, before clustering, it is often necessary to standardize each attribute dimension. In addition, when Minkowski distance is used as the similarity measure, the correlation between
attributes is not taken into account.

(2) Mahalanobis distance
Mahalanobis distance has different characteristic dimension units on the dataset, and there is a certain correlation in these characteristic dimensions. Generally, the Mahalanobis distance between object $c_o$ and object $c_k$ is defined as follows:

$$f_Z(c_o,c_k) = \left( (c_o - c_k)^T D^{-1} (c_o - c_k) \right)^{1/2}$$

(6)

Among them, D is the covariance matrix of the data similarity matrix. Mahalanobis distance is actually an improved form of Minkowski distance, which can not only make up for the limitation of Minkowski distance in scale, but also partially eliminate the influence of this connection.

(3) Cosine similarity
In clustering algorithms, cosine similarity is also a common similarity index. It regards data objects as vectors in f-dimensional space and obtains the similarity between two vectors by calculating the angle cosine between them. The smaller the angle between two vectors, the larger the cosine of the included angle, which indicates that the similarity between data objects is higher, and vice versa. Cosine similarity is defined as follows:

$$d(c_o,c_k) = \frac{c_o \cdot c_k}{\|c_o\| \|c_k\|}$$

(7)

The most common cosine similarity application scenario is to classify two texts, construct two vectors according to semantics, and then determine the angle cosine value between the two vectors according to the word frequency, so as to obtain the similarity of the two texts.

(4) Similarity and distance measurement
In general, the greater the distance between two data objects, the lower the similarity between two data points, and vice versa. Similarity and distance measurement can be converted. The common method for converting distance measurement to similarity is to define a non negative decreasing function with distance measurement as a variable. The general conversion framework is:

$$d(c_o,c_k) = G(f(c_o,c_k))$$

(8)

Among them, function $G(c)$ is a non negative decreasing function about c and satisfies $0 \leq G(c) \leq 1$. $f(c_o,c_k)$ is a certain distance between data objects $c_o$ and $c_k$, and the calculated $d(c_o,c_k)$ is the similarity between these two data points. Thurstone-Shepard model is a common method to convert distance measurement into similarity, as shown in the following formula:

$$d(c_o,c_k) = \exp(-\delta f(c_o,c_k)^{\tau})$$

(9)

Among them, $\delta$ and $\tau$ are positive real numbers.

When the similarity $d$ is found and there is $d \leq 1$, the similarity can also be transformed from a decreasing function into a distance measure.

For the above similarity calculation method, one can be selected from different applications and needs for clustering.

3. Experimental Results of Constructing the Knowledge Map of National Music

3.1 Questionnaire and Data
This paper adopted the method of random sampling to conduct a questionnaire survey on the sophomores and juniors in four universities. There were 200 questionnaires for University A, with 75 for music majors and 125 for non music majors. The other three universities, B, C and D, respectively, distributed 150 questionnaires, including 75 questionnaires for music and non music,
as shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>A University</th>
<th>B University</th>
<th>C University</th>
<th>D University</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issued questionnaires</td>
<td>200</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>650</td>
</tr>
<tr>
<td>Professional valid questionnaire</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>300</td>
</tr>
<tr>
<td>Non professional valid questionnaire</td>
<td>125</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>350</td>
</tr>
</tbody>
</table>

(1) Favorite music style

In terms of the absolute number and relative proportion, both professional and non professional students are more inclined to like modern lyrics and popular songs. As shown in Figure 3, Figure 3 (a) is for music major, and Figure 3 (b) is for non music major. Among music majors, 58.33% people like pop music. 16.67% people like folk music and 25% people like serious music. Among non music majors, 63.71% people like pop music, 22.29% people like folk music, and 14% people like serious music.

![Figure 3: Favorite music style](image)

From the survey results, the proportion of national folk music loved by the two groups of students is low, reflecting the lack of college students’ recognition of national culture and the absence of colleges and universities in national folk music and cultural education. As for music majors, they prefer serious music to some extent, which is related to their profound connotation, philosophy and aesthetic requirements and certain musical culture.

(2) Whether to take the initiative to contact with folk music in various parts of China except for courses

According to the results of the questionnaire, Figure 4 shows whether people take the initiative to contact folk music in various parts of China. Figure 4 (a) is for music majors, and Figure 4 (b) is for non music majors. Among music majors, 54.67% of students said they would, and 45.33% said they would not. Among non music majors, 44.86% of students said they would, and 55.14% said they would not.
Figure 4: Whether to take the initiative to contact with folk music in various parts of China

On this issue, the views of the two groups are roughly the same, which can be said to be mixed. This not only reflects the inner contradictions of college students, but also reflects the recognition and pride of a considerable number of students on Chinese traditional culture. This questionnaire is: in addition to learning, whether they can actively contact with folk music in various parts of China, causing people to think. Maybe there is more room for improvement in the teaching content, mode and method of colleges and universities, especially the choice of professional students on this issue can explain the problem.

(3) Whether to have contacted with local folk music (Folk songs, operas, religious music, folk art)

As shown in Figure 5, whether to have contacted with local folk music, where Figure 5 (a) is music major, and Figure 5 (b) is non music major. The number and relative ratio of professional and non professional students are also roughly the same. This shows that college students have a high degree of recognition of the richness, fun and humor of local folk music. At the same time, it reflects that the inheritance and promotion of local folk music still has a certain sense of identity and foundation among college students, which creates opportunities for the development of university education.

Figure 5: Whether to have contacted with local folk music
The study found that among the students who chose the “yes” answer, especially the students majoring in music, almost all had activities related to local folk music. In recent years, local governments have paid more attention to the development of folk music, and a large number of teachers and students have actively participated in relevant competitions organized by incentive mechanisms. This is a good platform for inheriting and carrying forward national folk music.

(4) Whether people are interested in folk music

As shown in Figure 6, the statistical results of whether people are interested in folk music are shown. Figure 6 (a) is for music majors, and Figure 6 (b) is for non music majors. In terms of the absolute number and relative ratio of students who choose to be interested, students from non music departments are significantly more than those from music departments. It is undeniable that, in addition to loving local folk songs, students who are not music majors are more like a hobby because they do not have a lifelong career. As a music major, students should have a strong interest in folk music from their extensive and profound professional knowledge structure, from the cultural resources of national folk music to absorb creative materials, from personal career planning and other aspects. The answer tells people that for music majors, teachers should strengthen their importance and necessity in music creation and career development, and expand students’ horizons, namely, “professional education” and guidance.

3.2 Construction of Knowledge Map of National Music Culture

In order to better develop national music, this paper started from the research idea of knowledge map and applied the knowledge map construction framework to the field of national music. In the field of traditional national music, the core idea is to build the knowledge map of national music with music knowledge as the core. It provided visual services for music lovers by building a map of national music knowledge. The knowledge map of ethnic music brings the ethnic knowledge points involved in traditional ethnic music into the system for research and understanding, so as to grasp the laws, development status and trends of ethnic music as a whole. This knowledge map provides a scientific framework and method for the mining of national music knowledge, helps researchers in the field of national music to make a reasonable interpretation and evaluation of the theoretical system of national culture, and guides and makes decisions for the development direction of national music.
(1) Keyword acquisition

In this paper, 4659 articles with keywords were counted by using the method of programming. The keywords were ranked from high to low according to their frequency, and the keywords whose cumulative number of keywords exceeds 100 were taken as indicators to reflect the current hot topics of ethnic research. 24 keywords with high frequency were selected as the basis of co-word analysis, as shown in Table 2.

Table 2: High-frequency keyword list

<table>
<thead>
<tr>
<th>Serial No</th>
<th>Key word</th>
<th>Serial No</th>
<th>Key word</th>
<th>Serial No</th>
<th>Key word</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>National Music</td>
<td>9</td>
<td>Central Conservatory of Music</td>
<td>17</td>
<td>Musicians</td>
</tr>
<tr>
<td>2</td>
<td>Ethnomusicology</td>
<td>10</td>
<td>Music education</td>
<td>18</td>
<td>Musical works</td>
</tr>
<tr>
<td>3</td>
<td>Folk music</td>
<td>11</td>
<td>National instrumental music</td>
<td>19</td>
<td>Chinese musicians association</td>
</tr>
<tr>
<td>4</td>
<td>Musical tradition</td>
<td>12</td>
<td>China Conservatory of Music</td>
<td>20</td>
<td>National style</td>
</tr>
<tr>
<td>5</td>
<td>Chinese music</td>
<td>13</td>
<td>Ethnic music</td>
<td>21</td>
<td>Chinese folk music</td>
</tr>
<tr>
<td>6</td>
<td>National Folk Music</td>
<td>14</td>
<td>National Music Culture</td>
<td>22</td>
<td>Traditional Chinese Music</td>
</tr>
<tr>
<td>7</td>
<td>Music culture</td>
<td>15</td>
<td>Musical composition</td>
<td>23</td>
<td>Folk songs</td>
</tr>
<tr>
<td>8</td>
<td>Music research</td>
<td>16</td>
<td>National Instruments</td>
<td>24</td>
<td>Western music</td>
</tr>
</tbody>
</table>

Figure 7 shows the frequency statistics results of high-frequency keywords. Among them, “national music” appeared 1940 times, and “ethnomusicology” appeared 465 times. “Folk music” appeared 415 times. “Music tradition” appeared 276 times, and “Chinese music” appeared 270 times.

(2) Cluster analysis

Based on the keyword theory, the selected high-frequency keywords were further preprocessed by using the method of codeword analysis, and the high-frequency keywords were clustered and visualized. By using the method of cluster analysis, this paper classified the important words in folk music, thus revealing the hot topics in the current study of folk music.

On the basis of cluster analysis, combined with relevant literature, this paper analyzes the current hot issues in the study of ethnic music.

The theoretical study of national music. The theme includes ethnomusicology, comparative
musicology, music research, ethnic music culture, and Chinese ethnic music. Some are from the perspective of musical instruments, and others are compared in terms of language, music, dance, etc. The theory of national music is based on practice and reality, so its research approach and field investigation are necessary.

Ethnic music institutions. The content includes the Central Conservatory of Music, China Conservatory of Music, and China Music Association. These associations facilitate academic exchanges and discussions. Whether in theory or in practice, innovations are constantly being carried out, making the development of national music more scientific and management more sound.

National musical instruments and national music. The content includes national instruments, musicians, national music and western music. Every nation has its own favorite musical instruments, so the playing methods are also very different. The names of various musical instruments are also very different, and each nation has its own unique style and strong local characteristics. In terms of the types of musical instruments, different ethnic groups have different situations. Some ethnic groups have 10 to 20 kinds of musical instruments.

Culture and education of folk music. This theme is reflected in music culture, music education, traditional music, Chinese music, music creation, music works, etc. Music creation is the basis of music development, and its development is related to people’s daily life, material culture, thoughts and feelings, regions and customs.

Ballad. The content covers Chinese traditional music, music types, folk music, music and dance, and Chinese folk music. Folk music has gone deep into the life and production of all ethnic groups, and to a certain extent, it affects the production and life of all ethnic groups. Chinese folk music refers to music related to opera, folk art and folk songs.

(3) Analysis of knowledge graph based on keyword co occurrence

Keyword co occurrence is the most widely used co word analysis method at present. It uses the frequency of simultaneous occurrence of keywords in the same literature to detect the relevance of keywords. Through clustering analysis and other statistical means, closely related keywords are gathered to form a category.

By analyzing the frequency of co words and using the method of cluster statistics, the complex co word relationship is simplified, and a relatively simple class relationship is obtained, thus further discussing the composition, evolution and disappearance of each class.

In the field of national music, the analysis of the common words of high-frequency keywords shows that the frequency of Chinese music, national folk music, music culture, music research, etc. is high, with the highest frequency reaching 1940 times, which is the focus of national music.

4. Conclusions

With the development of the times, the contents and forms of national music have changed, or some of them even died out. Although many national music is still spreading and used, many music forms can no longer represent the style and characteristics of the original songs, which are not preserved and protected as cultural heritage. As an important part of the traditional culture of the Chinese nation, the inheritance and protection of national music is one of the important contents of the cultural protection of the whole nation, which must be given high attention. In the future development of music in the new era, people should adhere to the principle of equal emphasis on protection and development. The protection of national music cultural heritage is not only a matter of protecting inheritors, but also a responsibility and obligation of music inheritors. The spiritual value and significance contained in the traditional national music culture should be deeply explored, and the ways and means of communication should be innovated, so as to improve the international influence and reputation of national music, as well as the recognition and innovation consciousness of national music culture. In a sense, cultural self-confidence is an important manifestation of a country’s comprehensive national strength and core competitiveness. The country should take the
lead in culture. People should attach importance to the differential competition and mutual reference between national music heritage and other cultures. For the minority music culture, the diversity of business under the multicultural background should be respected to form the unique national music culture. It is also necessary to maintain the unique charm of national music culture and get social recognition, which is also an affirmation and recognition of the overall cultural quality and national identity of the country. It should make people all over the country understand and love ethnic music works, and make more excellent ethnic music works participate in international competitions or other similar projects to compete for honor, so more people can enjoy the achievements of national music through art. It can also be promoted to a higher level to the mass field and other ways to enhance its sustainable development. Finally, while the overall artistic level of the country is improving, the traditional music culture of ethnic minorities should be protected and inherited. These excellent traditional cultures should be carried forward and developed together to provide more spiritual food for mankind. All ethnic groups should unite more closely to realize the Chinese Dream of the great rejuvenation of the Chinese nation.

References