Visualizing Symbolic Music via Textualization: An Empirical Study on Chinese Traditional Folk Music

Liumei Zhang(✉) and Fanzhi Jiang

Xi'an Shiyou University, Xi'an 710065, Shannxi, China
zhangliumei@xsyu.edu.com, 19211060559@stumail.xsyu.edu.com

Abstract. The continuous integration of computer technology and art has promoted the development of digital music. In recent years, the massive music data has promoted the problems of music inquiry, music classification, music content understanding and so on, making a new subject develop continuously, that is music information retrieval. The current mainstream feature extraction of music is based on acoustics, such as pitch, timbre, loudness, zero-crossing rate, etc. Direct symbolic music feature extraction and music analysis are relatively rare. This paper aims to present an empirical study on text clustering ideas in natural language processing into the field of symbolic music style analysis. Firstly, Symbolic music is textualized and proposed by us as an inspiration. To be precise, textualization of symbolic music is converted into weighted structured data through tf-idf algorithm. In the following step, three different types of mainstream clustering algorithms, K-Means, OPTICS, and Birch are used to perform cluster analysis and comparison on the traditional Chinese folk music dataset we crawled, T-SNE algorithm is used to visualize the dimensionality reduction of high-dimensional data. Finally, a series of objective evaluation indicators of clustering are used to evaluate the three clustering algorithms. Through comprehensive evaluation of indicators, it is proved that the clustering algorithm has achieved an excellent clustering effect on the midi note dataset we extracted. As a result of the clustering, the professional music theory knowledge and the historical development characteristics of traditional Chinese folk music are comprehensively integrated, which reversely verifies that the 1300 midi music data sets have distinct modal characteristics of traditional Chinese folk music.

Keywords: Music information retrieval · Midi datasets · Traditional Chinese music · Text clustering · Feature extraction

Supported in part by the scholarship from National Natural Science Foundation of China (No. 61802301) and (No. 211817019), Shaanxi Natural Science Foundation of China (No. 2019JQ-056).

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 647–662, 2021.
https://doi.org/10.1007/978-3-030-89814-4_47

1 Introduction

The combination of music and technology has developed for a long time. As a branch of sound and music computing, music information retrieval has also experienced vigorous development in recent years [19]. The industry has conducted deep exploration of music information retrieval, which has derived many research directions. Music recommendation, for example, uses collected user information to process and infer users' specific explicit or hidden preferences for music to achieve accurate music recommendation [18]. There is also the calculation and identification of music emotion through acoustic characteristics [16]. They generally start from acoustics, and then use the proven Russell Ring two-dimensional emotional model to analyze music emotion. There is also research on content-based music style. At present, the research based on music content mainly takes audio signals as the research object. For example, the calculation method is used to analyze and understand the content of digital sound and music [1,6], as well as the research on music produced by instruments [7,8], as well as the theoretical limitations of various aspects of acoustics in the in-depth study of musical elements and musical structures [17,21]. As well as studies on information computation of human-generated music and songs [11,13].

Sergio Oramas [14] proposed a clustering scoring method for building music databases. Yu Qi [15] uses one-sided continuous matching similarity algorithm to cluster the feature database, and marks the cluster center, and then uses linear alignment matching (LAM) algorithm to accurately match each cluster center and its elements in the cluster, which ensures the accuracy of the retrieval. Changsheng Xu [21] proposed a new method to distinguish music styles. Firstly, the support vector machine was used to classify pure music and vocal music supervisedly and automatically. Secondly, the unique music features were extracted from the objective characteristics of the two kinds of music. Finally, the clustering method was used to reconstruct the music content, which used a large number of acoustic and energy features for feature extraction and analysis. Dong-Moon Kim [10] conducted short-time Fourier transform on waveform corpus to extract features from this corpus from the perspective of providing music customization, and then proposed a dynamic K-Means algorithm to recommend corpus fragments in music corpus from different schools and styles. Wei-HoTsai [17] different from the traditional clustering method, studied the unsupervised clustering of vocal music. It still used Fourier transform and other means to cut the vocal music, and then extracted the vocal characteristics with the waveform vocal music fragment. However, the scale of its vocal music corpus was small, which did not prove its universality. The research based on waveform music has common characteristics, that is, the research is difficult and the experimental accuracy is low. Rudi Cilibrasi [5] classified the corpus fragments based on the compression method. In order to visually display the information in the distance matrix, hierarchical clustering of the corpus was adopted. However, the test on large datasets was different from that on small datasets, and the results were not ideal.

Chinese traditional music refers to the music created by the Chinese using their own national inherent methods and taking their own national inherent forms with their own national inherent morphological characteristics. It includes not only ancient works that have been produced and circulated in history, but also contemporary works. It is an essential part of Chinese national music. Chinese traditional folk music has its inherent mode characteristics. Thousands of years ago, the law of five-degree intergrowth emerged simultaneously in the Chinese cultural region and the Greek cultural region. Although the law of three-point profit and loss and the law of five-degree intergrowth have small differences, they have very high similarities in the principle and method of the law of life [23]. As the ancestor of the Chinese music system, the records of the three-point profit and loss method first appeared in the Spring and Autumn Period's *Guanzi Diyuan*. At this time, the three-point profit and loss method is related to the records of Gong, Shang, Jiao, Zhi, Yu. After the *Lv's Spring and Autumn Annals -Rhythm* was completed, the three-point profit and loss law began to be linked to the rules on the length of twelve laws such as Huang Zhong and Lin Zhong [24]. On the basis of the five-tone scale, later gradually appeared other partial tones, evolved into seven national tone scale, but the main tone is still five-tone scale [20].

In this work, we first crawled the music corpus that established Chinese traditional music and extracted the corpus data in midi format for text representation of music symbols. Then, the tf-idf algorithm is used to weight the textualized note data and process them into vectors. In the next step, K-Means, OPTICS, and Birch clustering methods are used to cluster and compare the weighted data, and different clustering indicators are selected for evaluation. This paper uses the T-distributed Stochastic Neighbor Embedding (T-SNE) algorithm to reduce the dimension and visualize the clustering. According to the clustering results, combined with the music theory of traditional Chinese national music, this corpus has distinctive tonal characteristics of traditional Chinese national music. The main ideas of our research are as flowcharts Fig. 2 (Fig. 1).

Fig. 1. The Pentatonic scale of Gong Shang Jiao Zhi Yu.

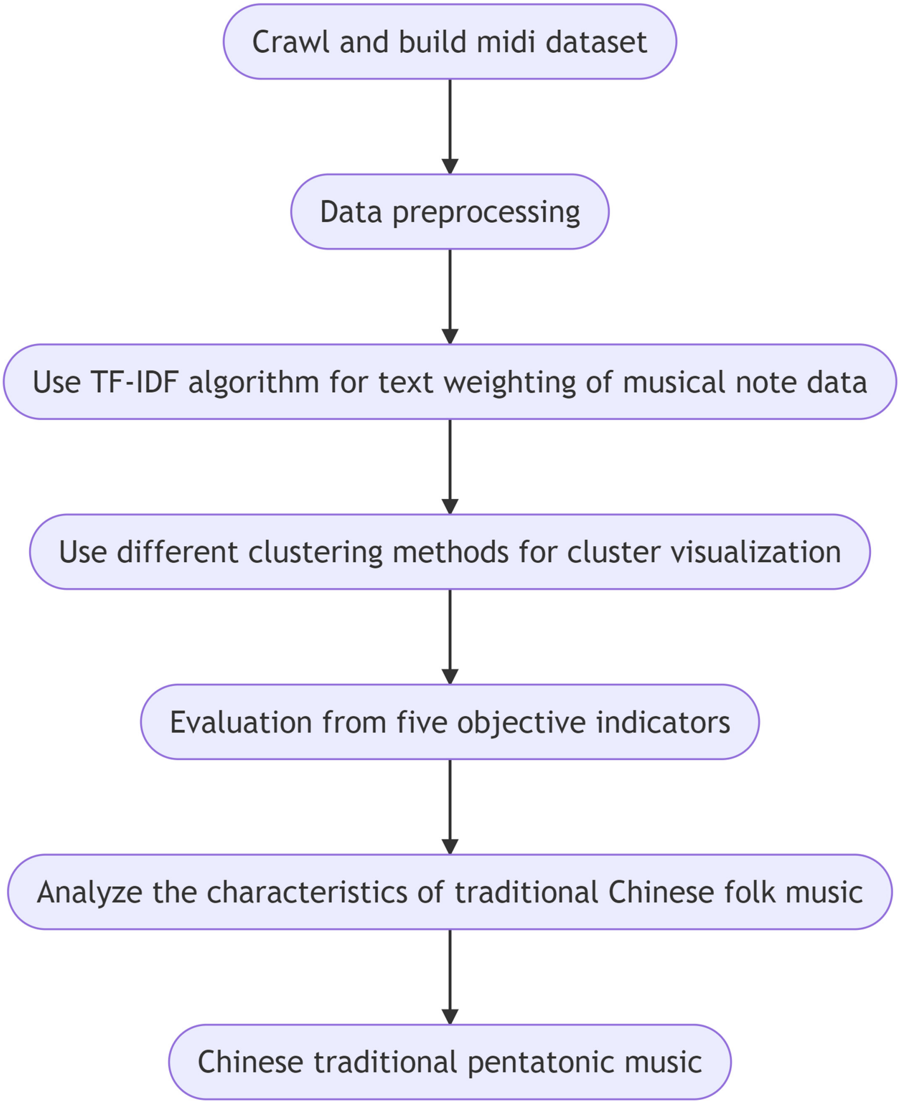

Fig. 2. Flow chart of main study idea.

2 Clustering Model

2.1 Basic Principles

In this part, we mainly describe the three clustering algorithms in depth from the background, motivation and calculation theory of the algorithm, paving the way for the processing and application of music data in the next part.

K-Means. Setting an initial value of k, representing k class clusters, aggregate all data points to the center of the closest class cluster, and keep iterating in accordance with this idea. Calculate the average value from the point to the center of the class cluster, and update the center of the class cluster until the iteration is stopped. This is the core idea of k-means algorithm.

Setting $Y = \{Y_1, Y_2, Y_3, ..., Y_n\}$. Each element in Y has m dimensions. Firstly, we initialize the center of the cluster according to the core idea of k-means and initialize k clustering centers $N_1, N_2, ..., N_n$, and then calculate the Euclidean distance from each element to each clustering center.

$$\operatorname{dis}(Y_i, N_j) = \sqrt{\sum_{t=1}^{m} (Y_{it} - N_{jt})^2} \tag{1}$$

The mean value of each dimension data of all elements in a specific region can be calculated as the center of the class cluster. The specific calculation equation is as follows:

$$N_l = \frac{\sum_{Y_i \in M_j} Y_i}{|M_l|} \tag{2}$$

The l cluster center is represented by N_l, where $1 \le l \le k$, $|M_l|$, the number of elements in the l cluster is represented by Y_i, and the i element in the l cluster is represented by Y_i, $1 \le i \le |M_l|$.

OPTICS. OPTICS is improved on the basis of the DBSCAN algorithm and is also a density-based clustering algorithm [2]. Choosing the appropriate two initial parameters and is the key to the DBSCAN algorithm, because it is very sensitive to the initial parameters, and whether the selection is appropriate will lead to huge differences in the calculation results. OPTICS came into being on this basis. Since the OPTICS algorithm is an improvement of the DBSCAN algorithm, some concepts are shared, such as: -Neighborhood, core object, direct density, density reachability, density connection, etc. The following is the definition related to OPTICS (assuming my sample Set is):

ϵ-Neighborhood: For $x_i \in X$, its ϵ-neighborhood contains the sub-sample set X whose distance between x_j and in the sample set is not greater than ϵ. ϵ-Neighborhood is a set, expressed as follows, the number of this set is recorded as $|N_\epsilon(x_j)|$.

$$N_\epsilon(x_j) = \{x_i \in X \mid \text{ distance } (x_i, x_j) \le \epsilon\} \tag{3}$$

Core object: For any sample $x_j \in X$, if its ϵ-neighbourhood corresponds to $N_\epsilon(x_j)$ at least contains $MinPts$ samples, that is, if $|N_\epsilon(x_j)| \ge MinPts$, then x_j is the core object.

Density direct: If x_i is located in the ϵ-neighborhood of x_j and x_j is the core object, it is said x_i to be direct by density x_j. The opposite is not necessarily true, that is, it cannot be said x_i to be directly reached by density X_j at this time, unless x_i is also the core object, that is, direct density does not satisfy symmetry.

Density is reachable: for x_i and x_j, if there is a sample sequence $p_1, p_2, ..., p_T$ that satisfies $p_1 = x_i$, $p_T = x_j$, and p_{t+1} is directly reached by the density p_t, it is said that the density is reachable. In other words, the density can be reached to meet the transitivity. At this time, the transferred samples in the sequence are all core objects, because only core objects can make other sample densities reachable directly. The density is reachable and does not satisfy the symmetry, which can be derived from the direct asymmetry of the density.

Density connected: For x_i and x_j, if there is a core object sample x_k, so that both x_i and x_j are reachable by the density x_k, then it is said that x_i and x_j the density are connected. The density connection relationship satisfies symmetry.

On the basis of the above definition of DBSCAN, OPTICS has introduced two definitions required by the algorithm:

Core distance: For a sample $x \in X$, for a given ϵ and $MinPts$, the smallest neighborhood radius that makes x a core point is called the core distance of x. It is mathematical expression is as follows, which $N_\epsilon^i(x)$ represents the node i closest to the node x in the set $N_\epsilon(x)$, such as $N_\epsilon^1(x)$ in $N_\epsilon(x)$ and Nearest node x:

$$cd(x) = \begin{cases} undefined\,|N_\epsilon(x)| < MinPts \\ d\left(x, N_\epsilon^{MinPts}(x)\right)|N_\epsilon(x)| >= MinPts \end{cases} \tag{4}$$

Reachability-distance: Let $x, y \in X$, for a given ϵ and $MinPts$, the reachability-distance of y about x is defined as:

$$rd(y, x) = \begin{cases} \text{undefined } \ |N_e(x)| < MinPts \\ \max\{cd(x), d(x, y)\}|N, (x)| >= MinPts \end{cases} \tag{5}$$

In particular, when x is the core point (the corresponding parameters are and), it can be understood $rd(y, x)$ according to the following formula:

$$rd(y, x) = \min\left\{\eta : y \in N_n(x) \& \left|N_\eta(x)\right| \geq \text{Min}\,Pts\right\} \tag{6}$$

Birch. The Birch algorithm builds a dendrogram called the cluster feature tree (CF tree). The CF tree can be constructed by scanning the data set in an incremental and dynamic manner. Therefore, it does not require the entire data set in advance [22].

It has two main stages: first scan the database to build a memory tree, and then apply the algorithm to the cluster leaf nodes. The CF tree is a highly balanced tree based on two parameters: branching factor B and threshold T. The CF tree is constructed when scanning data. When a data point is encountered, the CF tree will be traversed, starting from the root and selecting the nearest node at each level. If the closest leaf cluster of the current data point is finally determined, a test is performed to see whether the data point belongs to the candidate cluster or does not belong to the candidate cluster. Otherwise, a new cluster with a diameter larger than the given T will be created. Some other scans. It can also deal with noise effectively. However, when the clusters are not spherical, Birch may not work properly because it uses the concept of radius or diameter to control the boundaries of the clusters. In addition, it is sequence-sensitive and may generate different clusters for different sequence software input data.

3 Implementation

3.1 Dataset

This work is based on music information mining of symbolic music. Unlike waveform music, we have captured as many as 1,300 symbolic music of traditional Chinese people on the Internet and established a corpus of midi music. Since we didn 't make a detailed distinction when we caught it, its beats weren 't fixed, meaning their spectrum and beat numbers weren 't certain. But their average time is about ten seconds. py-midi and music21 toolkits are used to parse the music data set and rough pre-processing work. They separate the music according to the head and read the corpus with the music score object, and observe its main melody is in the piano track. So we extract the data of its piano track, and the basic scale spans the range of a group of small words to three groups of small words. Finally, a new data set based on text is established.

3.2 Data Preprocessing

Similar to text clustering, we first need to transform the music symbol information of this traditional music dataset document into mathematical information, so as to form high-dimensional space points and then calculate the similarity distance between the symbol element and the symbol element, so as to aggregate the symbol cluster. We use the classical statistical method if-idf algorithm to process the mathematical word vector of textualized note data, and give different weights. The specific algorithm of Ti-idf is as follows:

$$tf - idf = tf^{*}idf \tag{7}$$

$$f_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{8}$$

In the formula, $n_{i,j}$ is expressed as times the word appears in the document d_j, and the denominator is expressed as the sum of the times of all words in the document d_j.

$$idf_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \tag{9}$$

In the formula, $|D|$ is expressed as the total number of files in the corpus, and $|j : t_i \in d_j|$ represents the number of files containing words t_i (the number of files with $n_{i,j} \neq 0$). Tf-idf is used for the first mathematical processing of previously processed textualized music documents, including vectorizing notes data and giving specific weights based on the frequency of occurrence. Then, the structured data that can be used by the traditional clustering method are converted to music clustering. According to the music score object, the 19 extracted features can be observed across three octaves. In the piano, they are mainly distributed in the interval from a small word group to a small word group. The vectorized music data are the tensor of 87000 * 19.

3.3 Model Using

In order to prove the wide applicability of different clustering models in this dataset and the objectivity of clustering, we used three different clustering methods to cluster music data based on division, density and hierarchy. Among them, the k cluster determined by K-means clustering based on partition reaches the least square error, and the clustering results are dense, and the difference between classes is significant, as is shown in Algorithm 1. However, it may be difficult to select the k value or sensitive to noise points and outliers, resulting in large number of iterations and long time consuming. OPTICS does not explicitly generate data clustering, it only sorts the objects in the data object set and then calculates the sequence table, which contains a lot of information for extracting clustering, as is shown in Algorithm 2. It is not sensitive to the transformation of parameters in the process of clustering, but the clustering results are not as excellent as K-means and Birch. Birch has high efficiency in the calculation process, saves memory, and the clustering effect is better than OPTICS, as is shown in Algorithm 3. The time cost of the three clustering methods was statistically analyzed, as is shown in Table 1.

Table 1. Efficiency comparison of three clustering methods

Clustering Algorithm	K-Means	OPTICS	Birch
Clustering time (sec)	6.5	2.2	2.0

3.4 Algorithm

4 Objective Metrics for Evaluation

This paper evaluates the clustering effect from two kinds of metrics, and uses seven clustering methods. One of the measurement indexes is called internal evaluation method, which means to evaluate the algorithm by a single quantitative score, and the other is called external evaluation method, which compares the clustering results with the existing real classification results. Their classification is shown in Fig. 3.

Algorithm 1: K-means

Input: Noteset $D = \{x_1, x_2, x_3, ..., x_m\}$; Clusters k
Output: NoteCluster $C = C_1, C_2, ..., C_k$
k is selected Randomly from D:$\{\mu_1, \mu_2, \mu_3, ..., \mu_k\}$;
repeat
  $C = \varnothing(1 \leq i \leq k)$
until *something happens*;
for $j = 1, 2, 3, ..., m$ **do**
  Calculate the distance between sample x_j and each mean vector $\mu(1 \leq i \leq k) : d_{ji} = \| x_j - \mu_i \|_2$;
  Determine the cluster label of x_j according to the nearest mean vector:$\lambda = argmin_{i \in \{1,2,3,...,k\}} d_{ji}$;
  Corresponding cluster $C_{\lambda j} = C_{\lambda j} \cup \{x_j\}$ is divided into x_j ;
end
for $i = 1, 2, ..., k$ **do**
  Calculate the new mean vector $\mu_i^{'} = \frac{1}{|C_i|} \sum_{x \in C_i} x$;
  if $\mu_i^{'} \neq \mu_i$ **then**
    Update μ_i to $\mu_i^{'}$;
  else
    Keep the current mean unchanged ;
  end
end

The value of the ARI belongs to [−1, 1]. A higher value means that the more the clustering result matches the real situation. In a broad sense, ARI measures how well the two data distributions fit together. Homogeneity means that each cluster contains only members of a single class. Completeness refers to the extent to which all members of a given class are assigned to the same cluster. V measure score is symmetrical, it can be used to evaluate the consistency of two independent assignments on the same dataset. These indicators all use conditional entropy analysis to define some intuitive measures. The range of Mutual Information scores values is [−1, 1], and their larger values mean that the more the clustering results match the real situation. The range of Silhouette Coefficient is [−1, 1], and the closer the same category sample is, the farther away the sample of different categories is, the higher the score. A higher Carlinski-Harabasz score indicates that the model of the cluster is better. The index is the ratio of the difference between clusters and the degree of discreteness between clusters.

5 Experiment and Results

The experiments in this article are all carried out on computers equipped with Intel Core i7-9700 (3.00 GHz) CPU, 16 GB RAM and Microsoft Windows 10 operating system, and the development environment is Python 3.7. Figure 2

Algorithm 2: OPTICS

Input: Given parameter $\varepsilon, M, N_\varepsilon(i)$and $c_i, i = 1, 2, ..., N$.
Output: p $= \{p_i\}_{i=1}^N$

$k = 1; v_i = 0, i = 1, 2, ..., N;$;
$r_i = UNDEFINED, i = 1, 2, ..., N; I = 1, 2, ..., N;$;
while $(I \neq \varnothing)$ **do**
  Get an element i from I ,and let $I := I/i$;
  if $(v_i = 0)$ **then**
    $v_i = 1;$;
    $p_k = i, k = k + 1;$;
    if $(N_\varepsilon(i) \geq M)$ **then**
      //Insert the unvisited nodes in $N_\varepsilon(i)$ into *queueseedlist* according to the reachable distance ;
      $insertlist(N_\varepsilon(i), \{v_t\}_{t=1}^N, \{r_t\}_{t=1}^N, c_i, seedlist)$;
      while $(seedlistNOTEMPTY)$ **do**
        Get the first element j from *seedlist* ;
        $v_j = 1$;
        $p_k = j, k = k + 1$;
        if $(|N_\varepsilon(j) \geq M|)$ **then**
          break
        end
        //Insert the unvisited nodes in $N_\varepsilon(i)$ into *queueseedlist* according to the reachable distance ;
        $insertlist(N_\varepsilon(i), \{v_t\}_{t=1}^N, \{r_t\}_{t=1}^N, c_i, seedlist)$;
      end
    end
  end
end

Algorithm 3: Birch

Input: The dataset, threshold T, the maximum diameter (or radius) of a cluster R, and branching factor B.
Output: Compute CF points, where $CF =$ (step of points in a cluster N , Linear sum of the points in the cluster LS, the square sum of N data SS).

1 (Load data into memory) An initial in-memory CF-tree is constructed with one scan of the data. Subsequent phases become fast, accurate and less order sensitive ;
2 (Condense data) Rebuild the CF-tree with a larger T ;
3 (Global clustering) Use the existing clustering algorithm on CF leaves ;
4 (Cluster refining) Do additional passes over the dataset and reassign data points to the closest centroid from step3 ;

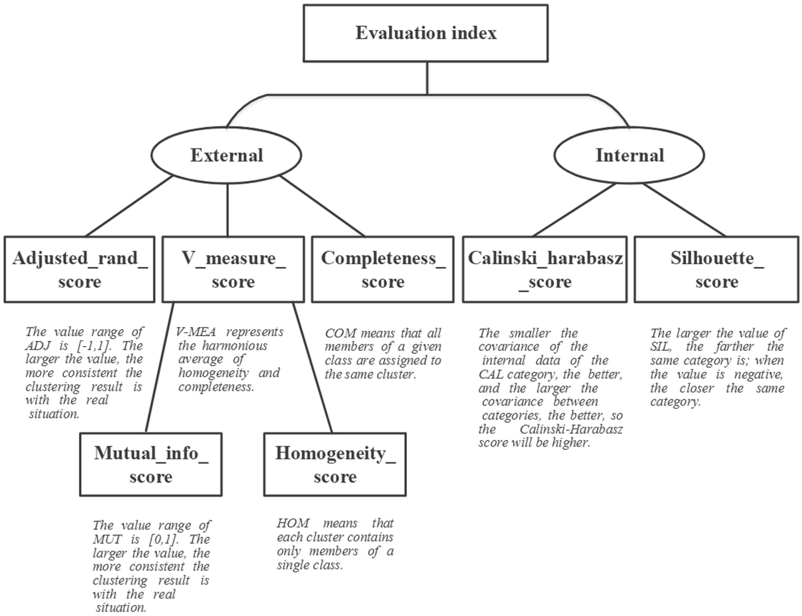

Fig. 3. Objective evaluation indicators for clustering effect

shows the main research route and ideas of this paper. Table 2 shows the comparison of the evaluation index results of K-Means, OPTICS and Birch.

Table 2. Comparison of objective indicators of clustering algorithm clustering effect

Algorithm	Silhouette	Calinski	Homogeneity	Completeness	V-measure	Adjusted rand	Mut information
K-Means	0.930	293355.316	0.934	1.000	0.966	0.949	0.966
OPTICS	0.908	165675.291	0.949	1.000	0.974	0.972	0.974
Birch	0.861	158523.624	0.859	1.000	0.924	0.878	0.924

T-distribution and Stochastic Neighbour Embedding algorithm is used for data dimension reduction and visualization of music data clustering. It is currently a very popular algorithm for dimension reduction of high-dimensional data. It is nonlinear and can adapt to the underlying data, support the optimization of parameters and confusion, and balance the local and global attention of data. We use it to reduce the data to three-dimensional data, and then visualize it. The elbow method is used to determine the optimal number of clusters. The reason why the elbow method is effective is based on the following observation: increasing the number of clusters helps to reduce the sum of intra-cluster variances of each cluster and calculate the sum of intra-cluster variances, the best value of k is 9, as is shown in Fig. 4.

After the clustering, after the clustering results, we use the reverse data to trace the clustered clusters, and trace back to the top seven largest number of note elements in each cluster, which corresponds to the five main voices and four partial voices of traditional Chinese pentatonic folk music: gong, Shang, Jiao, Zhi, Yu, Qingjiao, Bianzhi, Biangong, Run, as is shown in Fig. 5. The letter system is expressed as C, D, E, G, A, F, F#, B, Bb. C is used as the singing name of do, and the number of notes in various clusters is expressed as shown in the Table 3.

Table 3. Cluster tonic and other scattered notes in the cluster.

Cluster	Main note	First note	Second note	Third note	Fourth note	Fifth note	Sixth note
Cluster1	si fall	si	mi	sol	fa	re	do
Cluster2	la	sol	fa	mi	re	do	si
Cluster3	re	sol	fa	mi	do	si	la
Cluster4	mi	sol	fa	re	do	la	si
Cluster5	sol	fa sharp	fa	mi	re	do	la
Cluster6	fa sharp	do	sol	fa	mi	re	si
Cluster7	fa	fa sharp	sol	mi	re	do	si
Cluster8	si	sol	fa	mi	re	do	la
Cluster9	do	sol	fa	mi	re	si	la

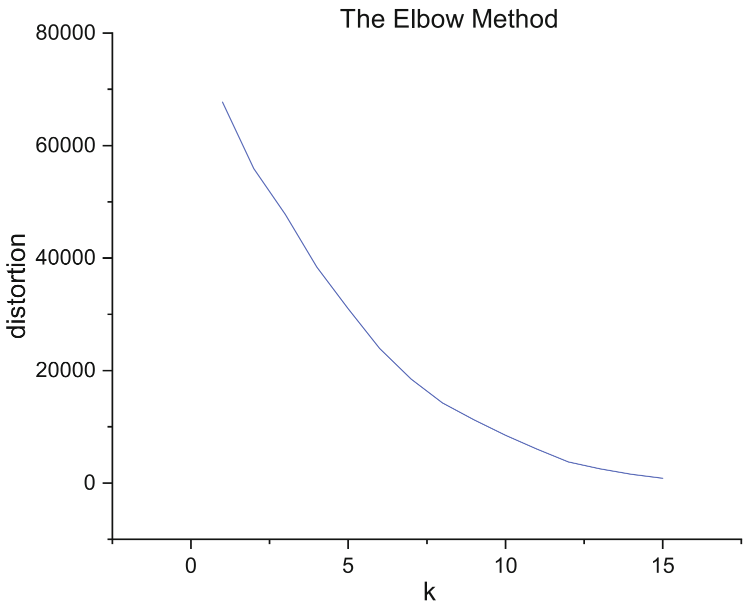

Fig. 4. The evaluation result of the error variance within the cluster to the K value.

Fig. 5. The pentatonic scale and the remaining four partial tones of the national seven-tone mode.

Chinese traditional pentatonic scale folk music mode, after a long development, mainly appeared Qingyue, Yanyue, and Yayue. There are three sources of Yayue in each dynasty. One is to inherit the palace music works of the Zhou Dynasty, the second is to reconstruct according to the music theory of the Zhou Dynasty, and the third is to make new music according to new sounds and customs. Its composition is Gong, Shang, Jiao, Bianzhi, Zhi, Yu, Biangong [25]. Qingyue refers to Qingshang music, also known as Qingshang music, is a traditional music rising in the Three Kingdoms, Jin Dynasty, Northern and Southern Dynasties and dominated in music life at that time. Its composition is the Gong, Shang, Jiao, Qingjiao, Zhi, Yu, and Biangong [12]. Yanyue is a very artistic song and dance music that provides entertainment and appreciation during the court dinner from Sui and Tang Dynasty to Song Dynasty. The palace yan music in Sui and Tang Dynasties reflects the highest achievement of music culture in this period. It originates from the accumulation of traditional music of Han nationality and the large-scale input of foreign music since Han and Wei Dynasties. Its composition is Gong, Shang, Jiao, Qingjiao, Zhi, Yu, Run [3]. Among them, cluster 1 has the structural characteristics of Yanyue mode, clusters 2, 3, 4, 8, and 9 have the structural characteristics of Qingyue mode, and clusters 5, 6, and 7 have the structural characteristics of Yayue mode. Figure 6, Fig. 7, and Fig. 8 show the clustering 3D visualization results under K-means, OPTICS, and Birch. It can be seen that the graph corresponds to the objective evaluation indicators in Table 2. K-Means and OPTICS show relatively better aggregation. The similar effect indicates that the note text data is more suitable for partition-based and density-based clustering methods.

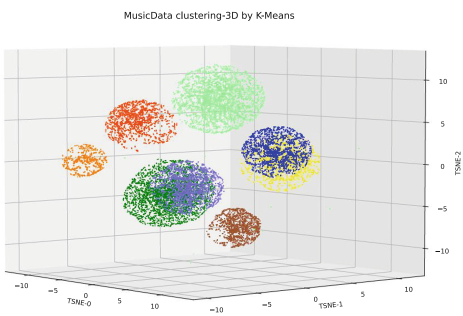

Fig. 6. Clustering 3D visualization results performed by K-means.

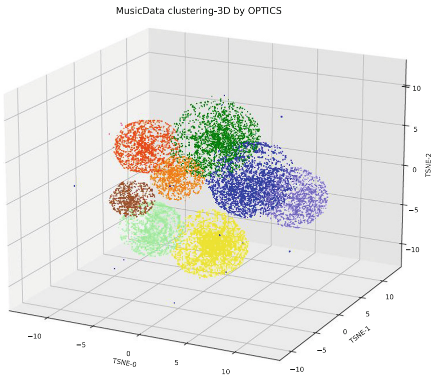

Fig. 7. Clustering 3D visualization results performed by OPTICS.

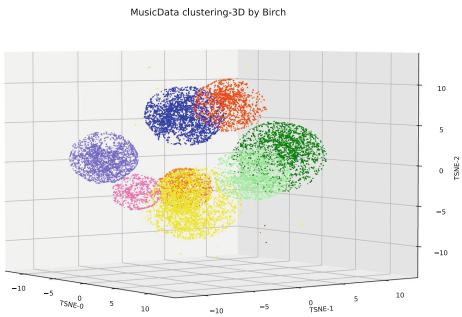

Fig. 8. Clustering 3D visualization results performed by Birch.

6 Conclusion

In this work, we first crawled and established a MIDI data set of traditional Chinese folk music, and proposed the idea of textualizing symbolic music data for cluster analysis. Then we combined professional knowledge of music theory with traditional Chinese folk music, cultural history conducted a textual clustering analysis on our own data set. According to the seven objective evaluation indicators of the clustering algorithm, K-Means and OPTICS produced relatively better results, combining professional music theory knowledge with traditional Chinese folk music. The historical development characteristics show that this dataset has distinct modal characteristics of traditional Chinese national music, and this experiment has produced an excellent combination of music grammar. We believe that the closer integration of music grammar and artificial intelligence is the future trend of computational music development. Regarding the next step, we think that we can extract more symbolic music characteristic texts and combine them with more professional music grammar for deeper exploration.

References

1. Akeroyd, M.A., Moore, B.C., Moore, G.A.: Melodyrecognition using three types of dichotic-pitch stimulus. J. Acoust. Soc. Am. **110(**3), 1498–1504 (2001)
2. Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J.: Optics: ordering points to identify the clustering structure. ACM SIGMOD Rec. **28**(2), 49–60 (1999)
3. Bing, L.: On Yanyue scale. Chin. Musicol. (2), 60–65 (1986)
4. Chongguang, L.: Fundamentals of Music Theory. People's Music Publishing House, Beijing (1962)
5. Cilibrasi, R., Wolf, R.d.: Algorithmic clustering of music based on string compression. Comput. Music J. **28**(4), 49–67 (2004)
6. Durey, A.S., Clements, M.A.: Features for melody spot-ting using hidden Markov models. In: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. II–1765. IEEE (2002)
7. Eronen, A.: Musical instrument recognition using ICA-based transform of features and discriminatively trained HMMs. In: Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings, vol. 2, pp. 133–136. IEEE (2003)
8. Herrera, P., Amatriain, X., Batlle, E., Serra, X.: Towards instrument segmentation for music content description: a critical review of instrument classification techniques. In: International Symposium on Music Information Retrieval, vol. 9, p. 2 (2000)
9. Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern Recognit. Lett. **31**(8), 651–666 (2010)
10. Kim, D., Kim, K.-s., Park, K.-H., Lee, J.-H., Lee, K.M.: A music recommendation system with a dynamic k-means clustering algorithm. In: Sixth International Conference on Machine Learning and Applications (ICMLA 2007), pp. 399–403. IEEE (2007)
11. Kim, Y.E., Whitman, B.: Singer identification in popular music recordings using voice coding features. In: Proceedings of the 3rd International Conference on Music Information Retrieval, vol. 13, p. 17 (2002)
12. Na, L.: On the difference between the European major scale and the national seven-tone unvoiced scale. GeHai (2), 78–80 (2010)
13. Liu, C.-C., Huang, C.-S.: A singer identification technique for content-based classification of MP3 music objects. In: Proceedings of the Eleventh International Conference on Information and Knowledge Management, pp. 438–445 (2002)
14. Oramas, S., Espinosa-Anke, L., Sordo, M., Saggion, H., Serra, X.: Information extraction for knowledge base construction in the music domain. Data Knowl. Eng. **106**, 70–83 (2016)
15. Qi, Y., Yongping, J., Du, X., Chuanze, L.: Application of a hierarchical clustering method in music retrieval. Comput. Eng. Appl. **47**(30), 113–115 (2011)
16. Roda, A., Canazza, S., De Poli, G.: Clustering affective qualities of classical music: Beyond the valence-arousalplane. IEEE Trans. Affect. Comput. **5**(4), 364–376 (2014)
17. Tsai, W.-H., Rodgers, D., Wang, H.-M.: Blind clustering of popular music recordings based on singer voice characteristics. Comput. Music J. **28**(3), 68–78 (2004)
18. Van den Oord, A., Dieleman, S., Schrauwen, B.: Deep content-based music recommendation. In: Advances in Neural Information Processing Systems, pp. 2643–2651 (2013)

19. Wei, L., Zijin, L., Yongwei, G.: Understanding digital music—summarization of music information retrieval technology. Fudan J. (Nat. Sci. Edit.) **57**(3), 271–313 (2018)
20. Xiaofeng, C.: Five-degree mutual generation and five-tone mode. Today Sci. Court (5), 64–64 (2006)
21. Xu, C., Maddage, N.C., Shao, X.: Automatic music classification and summarization. IEEE Trans. Speech Audio Process. **13**(3), 441–450 (2005)
22. Zhang, T., Ramakrishnan, R., Livny, M.: BIRCH: an efficient data clustering method for very large databases. ACM SIGMOD Rec. **25**(2), 103–114 (1996)
23. Xun, Z.: Interpretation of the five-degree interaction Law in Chinese music system. Contemp. Music (8) (2020)
24. Yijun, Z.: Analysis of the relationship between traditional Chinese musicology and ethnomusicology. Northern Music **02**(01), 39–40 (2020)
25. Zuxiang, Z.: Seven Tones of Yayue. Ph.D thesis (1987)