K-means clustering analysis of Chinese traditional folk music based on midi music textualization

Zhang Liumei*
School of Computer Science
Xi'an Shiyou University
Xi'an, China
zhangliumei@xsyu.edu.cn

Jiang Fanzhi
School of Computer Science
Xi'an Shiyou University
Xi'an, China
19211060559@stumail.xsyu.edu.cn

Li Jiao
School of Computer Science
Xi'an Shiyou University
Xi'an, China
lijiao@xsyu.edu.cn

Ma Gang
School of Computer Science
Xi'an Shiyou University
Xi'an, China
jhlmg@xsyu.edu.cn

Liu Tianshi
School of Computer Science
Xi'an Shiyou University
Xi'an, China
liutianshi@xsyu.edu.cn

***Abstract*—The current mainstream feature extraction of music information retrieval (MIR) is based on acoustics, such as frequency, loudness, zero-crossing rate. while it is rare to perform feature extraction and music analysis directly on symbolic music. This article seeks to introduce the idea of text clustering in natural language processing into the field of symbolic music style analysis. From this, this work got inspiration to turn midi music into text data and transform it into weighted structured data through tf-idf, and then use the K-Means clustering algorithm to perform cluster analysis and comparison on the traditional Chinese folk music dataset we crawled, and finally use The T-SNE algorithm performs dimensionality reduction and visualization of high-dimensional data. After a series of objective indicators evaluation, it is proved that the clustering algorithm has achieved a good clustering effect on the midi note dataset we extracted; through the clustering results, comprehensive professional music theory knowledge and the historical development characteristics of traditional Chinese folk music, Reverse verification of the 1300 midi music data sets has distinct modal characteristics of traditional Chinese folk music.**

Keywords-music information retrieval, feature extraction, traditional Chinese music, text clustering, midi dataset.

I. INTRODUCTION

Music information retrieval is an interdisciplinary science that combines music, psychology, signal processing, informatics, and computing science [18]. After more than 20 years of development, the industry is divided into many sub-fields of research, such as: Music recommendation (Aaron van den Oord [17]); music sentiment analysis based on music content (Antonio Roda, Sergio Canazza, and Giovanni De Poli [16]) explored cluster analysis of emotional factors in classical music beyond the constraints of valence and arousal; music style classification, this article focuses on this field. At present, the main research objects of music style recognition are audio signals and waveform data. Many technologies have been developed to extract interesting information from music residing in digital audio signals, such as melody content (Akeroyd and Moore[1]; Durey and Clements[5]), the instruments involved (Herrera and X. Amatriain, E. Batlle, and X. Serra [7]; Eronen [6]), genre (Tzanetakis and Cook [17]; Xu [20]) and artist or singer (Kim and Whitman [10]; Liu and Huang [12]).

(Sergio Oramas[13]) emphasized a method of clustering and scoring relationships in the relationship extraction pipeline to build a music knowledge base system. (Yu Qi [14]) proposed a hierarchical clustering algorithm based on one-sided continuous matching to be applied to music retrieval, which uses wave files to cut and extract features. (Changsheng Xu, Namunu C. Maddage, and Xi Shao [20]) After distinguishing pure music from vocal music using support vector machines, clustering algorithm is used to construct music content, which is also based on acoustic files for feature extraction. Feature selection is for pure music , Uses power-related features, such as Mel frequency perception coefficient, amplitude envelope and power spectrum; for vocal music, uses voice-related functions, such as cepstrum coefficient derived from LPC, zero-crossing rate, spectral flux and Cepstrum flux. (Dong-Moon Kim [10]) Extract the properties of music from the sound waves of music, and use STFT (shortest time Fourier form) to analyze the properties of music. Then, in order to analyze user preferences, a dynamic K-means clustering algorithm is proposed. (Wei-Ho Tsai, Dwight Rodgers, and Hsin-Min Wang [16]) studied the feasibility of unsupervised clustering of music data based on related singers. From music, the voice of the singer can be extracted by sound segment detection and sound signal modeling. Feature, but because its data set is not large enough, the robustness of its model cannot be guaranteed. The tree-based compression method is used in the clustering of midi music works (Rudi Cilibrasi, Pual Vitanyi and Ronald de Wolf [4]), which uses each song as a unit to encode strings, and finally achieves the goal of distinguishing by various music styles Unfortunately, its performance on a large data set of thousands of songs is not satisfactory.

Chinese traditional folk music refers to traditional Chinese music played in solo or ensemble with traditional Chinese instruments. Chinese traditional folk music has various artistic genres, but the modes have common points. As one of the birthplaces of the five-degree phase-generation law (similar to the three-point profit and loss law) (Zhang Xun [23]), the earliest records of this law appear in "Lü Shi Chunqiu. Musical Chapter" and "Guanzi. Diyuan Chapter" In ancient books (Zhu Yijun [24]), it described the pentatonic scale: Gong Shang Jiao Zhengyu's generation algorithm, which is equivalent to today's stave: do, re, mi, sol, la. This law avoids the minor second, etc. The generation of harmonious intervals reflects the wisdom of the ancients. Then, on the basis of the five-tone mode, the six-tone and seven-tone national mode scales were developed, so in ancient times there were: "Nine songs, eight winds, seven tones, and six rhythms, with five tones." (Cao Xiaofeng [19])

We performed K-Means clustering after textualization of note data based on symbolic music (ie MIDI data). Objective evaluation indicators show that clustering produces good results. According to the clustering results, we can analyze the 1300 captured by the network. The first midi music data set conforms to the composition law of traditional Chinese folk music in the original scale composition. As far as we know, this is the first paper that starts with the textualization of symbolic music data, uses clustering methods in machine learning and analyzes music styles based on the historical development and grammar of regional ethnic music.

In this paper, we perform text processing on the note data based on symbolic music (ie MIDI data), and then use the tf-idf algorithm to weight the text data to make data for the use of K-Means clustering algorithm Pre-processing work. The reason why we choose the clustering algorithm is because we want to adopt automatic cluster division method for note clusters when we don't know which cluster the note data belongs to. The purpose is to observe that these data are relative to the traditional pentatonic scale. The degree of fit. Then we used a clustering algorithm to perform cluster analysis on 1300 midi data that we crawled. According to objective clustering index evaluation, the clustering showed a good clustering effect in comparison. On the other hand, for the visualization part, we adopted the T-SNE algorithm to visualize the clustering results and verify our analysis based on our professional music theory knowledge and the mode characteristics of traditional Chinese folk music. That is, the 1300 midi music data sets are in the original The composition of the scale conforms to the rules of traditional Chinese folk music. As far as we know, this is the first paper that starts with the textualization of symbolic music data, uses clustering methods in machine learning and analyzes music styles based on the historical development and grammar of regional ethnic music.

Figure 1. The corresponding notes of the pentatonic scale on the staff

II. CLUSTERING MODEL

A. Basic Principles

This section mainly introduces the background of the clustering model and the principle of the algorithm.

Assuming that a given data sample X contains objects $X=\{X_1, X_2, X_3, ..., X_n\}$, each of which has m properties of dimensions. The goal of K-means algorithm is to gather k objects into specified clusters based on the similarity between n objects. Each object belongs to and belongs to only one cluster with the smallest distance from the center of the cluster [9]. For K-means, first need to initialize cluster k centers $\{C_1, C_2, ..., C_n\}$, $1<k\leq n$, and then calculate the Euclidean distance from each object to each cluster center, as shown in the following formula:

$$dis(X_i, C_j)=\sqrt{\sum_{t=1}^{m}(X_{it}-C_{jt})^2} \quad (1)$$

In the above formula, X_i represents the object i, $1<i\leq \mathrm{n}$, C_j represents the attribute of the cluster center j, $1<j\leq k$, X_{it} represents the attribute t of the object i, $1<t\leq m$, C_{jt} represents the attribute t of the cluster center j. Compare the distance of each object to each cluster center in turn, assign the objects to the clusters of the nearest cluster center, and obtain k clusters $\{S_1, S_2, S_3, ..., S_k\}$.

The algorithm uses the center to define the prototype of the cluster. The center of the cluster is the average value of all objects in the cluster in each dimension. The calculation formula is as follows:

$$c_t=\frac{\sum_{X_i\in S_l}X_i}{|S_l|} \quad (2)$$

In the formula, C_l represents the center l of the cluster, $1 \le l \le k$, $|S_l|$ represents the number of objects in clusters l , X_i represents the object i in clusters $l, 1 \le i \le |S_l|$.

B. Algorithm

Algorithm: K-Means

Input: Sample set $D = \{x_1, x_2, x_3, ..., x_m\}$; Clusters k

Output: Clustering $C = \{C_1, C_2, ..., C_k\}$

Randomly select k samples from D as the initial mean vector $\{\mu_1, \mu_2, \mu_3, ..., \mu_k\}$

repeat

$C_i = \varnothing (1 \le i \le k)$

for $j = 1, 2, ..., m$ ***do***

Calculate the distance between sample x_j and each mean vector $\mu_i (1 \le i \le k) : d_{ji} = \|x_j - \mu_i\|_2$;

Determine the cluster label of x_j according to the nearest mean vector: $\lambda_j = \arg\min_{i \in \{1,2,3,...,k\}} d_{ji}$

Divide sample x_j into corresponding cluster $C_{\lambda j} = C_{\lambda j} \bigcup \{x_j\}$;

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 $\mu_i^{'} \ne \mu_i$ ***then***

Update the current mean variable μ_i to $\mu_i^{'}$;

else

Keep the current mean unchanged

end

end

III. Implementation

A. Dataset

The musical note text dataset used in this article is obtained by processing the midi file data set we grabbed. This midi data set is created by ourselves through 1300 Chinese traditional folk music grabbed on the Internet. The average length is about 12 seconds. We use the python package music21 to parse and process midi files. First, convert the midi file into a score object, and then extract the notes in the piano part of the instrument, because the note data has different scales, and the scales are in the range of one group of small letters to three groups of small letters, so the scale mark is discarded and the pitch is extracted Part, and then textualize it to get the original text data set based on symbol music.

B. Data Preprocessing

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

Because this article clusters symbolic music textually, the current mainstream clustering algorithms mainly operate on structured data, so this article first needs to structure irregular text data. Here we use The classic weighting method tf-idf, used for text mining and information retrieval, represents the product of term frequency (TF) and inverse document frequency (IDF). The specific calculation formula is as follows:

$$tf - idf = tf * idf \tag{3}$$

$$f_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{4}$$

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

$$idf_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \tag{5}$$

In the formula, $|D|$ is 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 $n_{i,j} \ne 0$).

Through tf-idf calculation, the note elements are weighted. The first step is to extract the complete notes for textualization, which means that they have different scales. After processing, the number of features extracted is 19, which are distributed in the piano keyboard. In the interval from one group of small characters to three groups of small characters; then in order to simplify the data and facilitate clustering statistics, the pitch of its notes is extracted, and the scale is discarded. After vectorization, the number of features is extracted to nine, namely C, D, E, F, F#, G, A, B, Bb, and then converted to weight matrix calculation, get 87000*9-dimensional tensor.

IV. Objective Metrics for Evaluation

In order to evaluate the effect of the clustering algorithm on the note text data set, we selected seven objective indicators to evaluate from the outside and inside of the clustering model. The so-called internal refers to measuring the tightness between the sample points of a cluster. Compared with measuring the distance between the sample and other clusters, external measurement refers to comparison with a reference model. The index classification is shown in Figure 3:

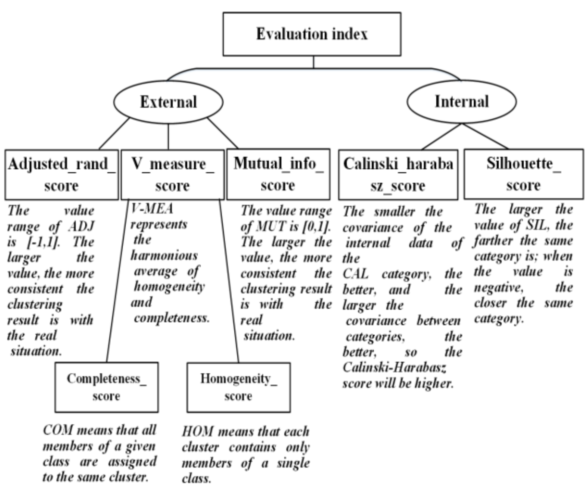

Figure 3. Evaluation indicators based on internal and external clustering

V. Experiment and Results

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

TABLE Ⅰ shows the evaluation index results of K-Means.

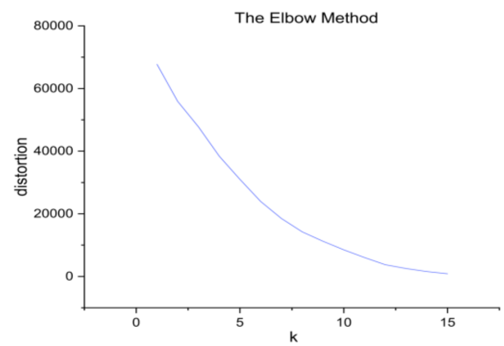

Figure 2. The evaluation result of the error variance within the cluster to the K value

TABLE I. Comparison of Objective Indicators of Clustering Algorithm Clustering Effect

K-Means	Table of Clustering indicators		
	sil	***cal***	***hom***
	0.930	293355.316	0.934
com	***v_me***	***adj***	***mut***
1.000	0.966	0.949	0.966

We use T distribution and Stochastic Neighbour Embedding (T-SNE) to visualize the dimensionality reduction of high-dimensional note text data. K is evaluated and selected by the error variance within the cluster, which is commonly known as the elbow method. When the K value is 9, the number of clusters is 9, as shown in Figure 2.

After clustering, the standardized data in the clusters were converted into original data for observation. The experiment found that the largest proportions of notes in the 9 clusters were C, D, E, F, F#, G, A, B, Bb. This is the pentatonic scale in traditional Chinese folk music: Gong, Shang, Jiao, Zhi, Yu, and the other four partial tones that evolved in the seven-tonic scale during its historical development: Qingjiao, Bianzhi, Biangong, and Run, as shown in Figure 4.

Figure 4. The pentatonic scale and the remaining four partial tones of the national seven-tone mode

In addition to the notes with the highest proportions, there are a small number of other notes in each part of the clusters. The other few notes in each cluster are shown in TABLE Ⅱ.

TABLE II. Cluster Tonic and Other Scattered Notes in the Cluster

	main note	**1**	**2**	**3**	**4**	**5**	**6**
Cluster1	**Bb**	B	E	G	F	D	C
Cluster2	**A**	G	F	E	D	C	B
Cluster3	**D**	G	F	E	C	B	A
Cluster4	**E**	G	F	D	C	B	A
Cluster5	**G**	F#	F	E	D	C	B
Cluster6	**F#**	C	G	F	E	D	B

Cluster7	**F**	F#	G	E	D	C	B
Cluster8	**B**	G	F	E	D	C	A
Cluster9	**C**	G	F	E	D	B	A

In the national seven-tone mode of our country, the scale composition is mainly composed of five-tone mode Gong, Shang, Jiao, Zhi, and Yu based on adding two of ICS. the other four partial tones [3]. Divided into three main modes: Qingyue mode, Yanyue mode, and Yayue mode. The composition of Qingyue modes is: Gong, Shang, Jiao, Qingjiao, Zhi, Yu, and Biangong [11]. The composition of Yanyue scale is: Gong, Shang, Jiao, Qingjiao, Zhi, Yu, Run, [2] Yayue scale. The composition is: Gong, Shang, Jiao, Bianzhi, Zhi, Yu, Biangong [22]. 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 5 show the clustering 3D visualization results under K-means. It can be seen that the graph corresponds to the objective evaluation indicators in TABLE Ⅰ. K-Means show relative good aggregation.

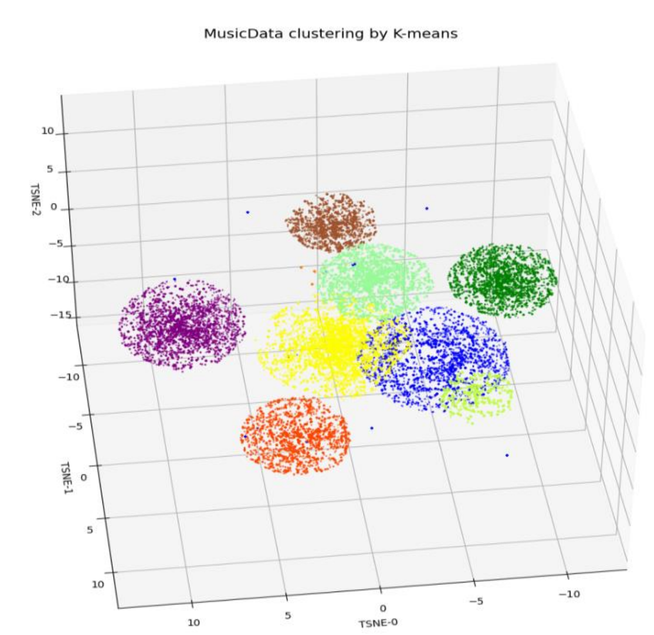

Figure 5. Clustering 3D visualization results performed by K-means

VI. CONCLUSION

In this work, we first crawled and established a MIDI data set of traditional Chinese folk music; then proposed the idea of textualizing symbolic music data for cluster analysis; and then combined professional knowledge of music theory with traditional Chinese folk music. Cultural history conducted a textual clustering analysis on the data set we built. According to the seven objective evaluation indicators of the clustering algorithm, K-Means produced a good clustering effect, combining professional music theory knowledge with the historical development of traditional Chinese folk music. The characteristics show that this data set has distinct modal characteristics of traditional Chinese national music; and this experiment has produced a good combination of music grammar. We believe that the closer integration of music grammar and artificial intelligence is the future trend of computational music development; in the next step, we think we can extract more symbolic music feature texts, and combine with more professional music grammar for deeper exploration.

ACKNOWLEDGMENT

We would like to thank Ms. Luo Luo from the Music Department of Xi’an Shiyou University for helpful discussions and advice. This work is 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), and Xi’an Shiyou University Graduate Innovation and Practical Ability Training Project (No. YCS20213205).

REFERENCES

[1] M. A. Akeroyd, B. C. Moore, and G. A. Moore. Melody recognition using three types of dichotic-pitch stimulus. The Journal of the Acoustical Society of America, 110(3):1498–1504, 2001.

[2] L. Bing. On yanyue scale. Chinese Musicology, (2):60–65,

[3] L. Chongguang. Fundamentals of music theory. [M]. Beijing: People’s Music Publishing House, 1962.

[4] R. Cilibrasi, P. Vitanyi, and R. d. Wolf. Algorithmic clustering of music based on string compression. Computer Music Journal, 28(4):49–67, 2004. 2

[5] A. S. Durey and M. A. Clements. Features for melody spotting using hidden markov models. In 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 2, pages II–1765. IEEE

[6] A. Eronen. 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., volume 2, pages 133–136. IEEE, 2003.

[7] P. Herrera, X. Amatriain, E. Batlle, and X. Serra. Towards instrument segmentation for music content description: a critical review of instrument classification techniques. In International symposium on music information retrieval, volume 9, 2000. 2

[8] A. K. Jain. Data clustering: 50 years beyond k-means. Pattern recognition letters, 31(8):651–666, 2010. 2

[9] D. Kim, K.-s. Kim, K.-H. Park, J.-H. Lee, and K. M. Lee. A music recommendation system with a dynamic k-means clustering algorithm. In Sixth International Conference on Machine Learning and Applications (ICMLA 2007), pages 399–403. IEEE, 2007. 2

[10] Y. E. Kim and B. Whitman. Singer identification in popular music recordings using voice coding features. In Proceedings of the 3rd international conference on music information retrieval, volume 13, page 17, 2002.

[11] LiNa. On the difference between the european major scale and the national seven-tone unvoiced scale. GeHai, (2):78–80, 2010.

[12] C.-C. Liu and C.-S. Huang. A singer identification technique for content-based classification of mp3 music objects. In Proceedings of the eleventh international conference on Information and knowledge management, pages 438–445, 2002.

[13] S. Oramas, L. Espinosa-Anke, M. Sordo, H. Saggion, and X. Serra. Information extraction for knowledge base construction in the music domain. Data & Knowledge Engineering, 106:70–83, 2016.

[14] Y. Qi, J. Yongping, X. Du, and L. Chuanze. Application of a hierarchical clustering method in music retrieval. Computer Engineering and Application, 47(30):113–115, 2011.

[15] A. Roda, S. Canazza, and G. De Poli. Clustering affective qualities of classical music: Beyond the valence-arousal plane. IEEE Transactions on Affective Computing, 5(4):364–376, 2014.

[16] W.-H. Tsai, D. Rodgers, and H.-M. Wang. Blind clustering of popular music recordings based on singer voice characteristics. Computer Music Journal, 28(3):68–78, 2004.

[17] A. Van den Oord, S. Dieleman, and B. Schrauwen. Deep content-based music recommendation. In Advances in neural information processing systems, pages 2643–2651, 2013.

[18] L. Wei, L. Zijin, and G. Yongwei. Understanding digital music summarization of music information retrieval technology. Fudan Journal (Natural Science Edition), 57(3):271r313, 2018.

[19] C. Xiaofeng. Five-degree mutual generation and five-tone mode. today science court, (5):64–64, 2006.

[20] C. Xu, N. C. Maddage, and X. Shao. Automatic music classification and summarization. IEEE transactions on speech and audio processing, 13(3):441–450, 2005.

[21] ZhangXun. Interpretation of the five-degree interaction law in chinese music system. contemporary music, (8), 2020.

[22] ZhuYijun. Analysis of the relationship between traditional chinese musicology and ethnomusicology. Northern Music, 02(01):39–40, 2020.

[23] Z. Zuxiang. Seven Tones of Yayue. PhD thesis, 1987.

Please fill in the Authors' background:

Position can be chosen from:

Prof. / Assoc. Prof. / Asst. Prof. / Lect. / Dr. / Ph. D Candidate / Postgraduate / Ms.

Full Name	Email address	Position	Research Interests	Personal website (if any)
Zhang Liumei	zhangliumei@xsyu.edu.cn	Assoc.Prof.	Machine learning	
Jiang Fanzhi	19211060559@stumail.xsyu.edu.cn	Ms.	Music computing	
Li Jiao	lijiao@xsyu.edu.cn	Assoc.Prof.	Blockchain	
Ma Gang	jhlmg@xsyu.edu.cn	Assoc.Prof.	Data Mining	
Liu Tianshi	liutianshi@xsyu.edu.cn	Prof.	Computer application	

*This form helps us to understand your paper better; the form itself will not be published.