CHINESE OPERA GENRE INVESTIGATION BY CONVOLUTIONAL NEURAL NETWORK

Huan Zhang[1] **Yiliang Jiang**[2] **Weifeng Zhao**[2] **Tao Jiang**[2] **Peng Hu**[2]
[1] School of Music, Carnegie Mellon University, Pittsburgh, US
[2] Tencent Music Entertainment, Shenzhen, China
huanz@andrew.cmu.edu yiljiang@tencent.com

ABSTRACT

In this study, we explore the stylistic differences of regional Chinese opera genres with convolutional neural networks in audio domain. In the classification experiment, we report an F1 value of **0.88** for the classification of 18 Chinese opera genres. Besides, we also performed clustering of the learnt embeddings to investigate the similarity between genres. Finally, a positive correlation between music embedding distance and geographical distance of these regional genres was found.

1. INTRODUCTION

Chinese opera, or *Xiqu*, is a form of theatrical art with more than a thousand years of history. It contains many regional branches such as *Yu Ju*, *Hu Ju*, and *Jing Ju* (Peking Opera), which eventually evolve into opera genres with large stylistic differences. Most of the genres are named after their birthplaces, and closely associate with regional style and traditions, especially dialects.

As the most popular and representative Chinese opera genre, Peking Opera has received lots of attention in the MIR field. In works like [1–3], music structure and singing styles of Peking Opera were analyzed. However, few works has conducted over other Chinese opera genres. To our knowledge, only [4] used traditional audio descriptors to discriminate genres with feature fusion techniques. How much stylistic difference exists among Chinese Opera genres? In what ways do the genres influence each other? With these questions in mind, this extended abstract investigates the genre differences among Chinese operas through genre classification and clustering analysis tasks performed by convolutional neural network.

2. METHODOLOGY

2.1 Data

The dataset contains 4070 aria recordings from 18 genres: *Huangmei Xi* (黄梅戏), *Jing Ju* (京剧), *Yue Ju* (越剧), *Yu Ju* (豫剧), *Yue Ju* (粤剧), *Chao Ju* (潮剧), *Hu Ju* (沪剧), *Jin Ju*(晋剧), *Ping Ju*(评剧), *Huagu Xi*(花鼓戏), *Qu Ju*(曲剧), *Qin Qiang*(秦腔), *Lu Ju*(吕剧), *Kun Qu*(昆曲), *Bangzi*(梆子), *Chuan Ju*(川剧), *Xi Ju*(锡剧), *Huai Ju*(淮剧), where the genres are balanced with roughly 240 pieces for each. The arias are sung by Chinese opera masters and are collected by Tencent Music Entertainment. To eliminate the effect of individual voice on the classification, the recordings were obtained from a wide range of singers. We split the training and testing by $9:1$.

The model discussed in next section will take in time-frequency representation of *3s* audio segment from the tracks, and we utilized both vocal and instrumental regions of the track, as they both encodes stylistic information regarding the specific genre. For the classification of the testing track, we averaged over the softmax predictions on all segments to achieve the final label.

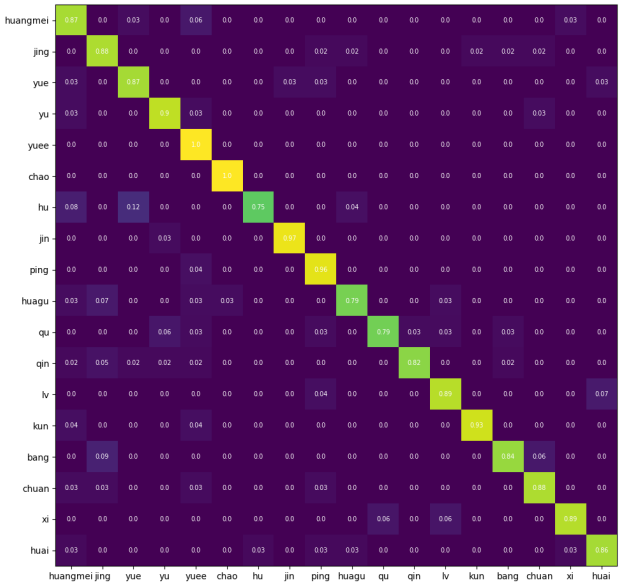

Figure 1. Confusion matrix among 18 opera genres on the testing set.

2.2 Model

We adopted the Musicnn [5] model for our genre classification task. Designed for music auto-tagging, the model contains horizontal (temporal) and vertical (spectral) convolutional filters that was perceptually and musically mo-

tivated. We chose a batch size of 64 and used Adam as the optimizer, with a learning rate of 10^{-5}.

3. ANALYSIS

3.1 Classification Result

Classification on the 18 genres reached an accuracy of **87.9%** and f1 score of **88.1%** on the testing set. the confusion matrix is shown in Figure 1. We can observe that most genres are correctly identified with accuracy > 0.8, whereas the most easily mistaken genres are *Hu Ju* and *Yue Ju*, which both features *Wu Chinese* in their arias, a dialect originated from Shanghai and Zhejiang.

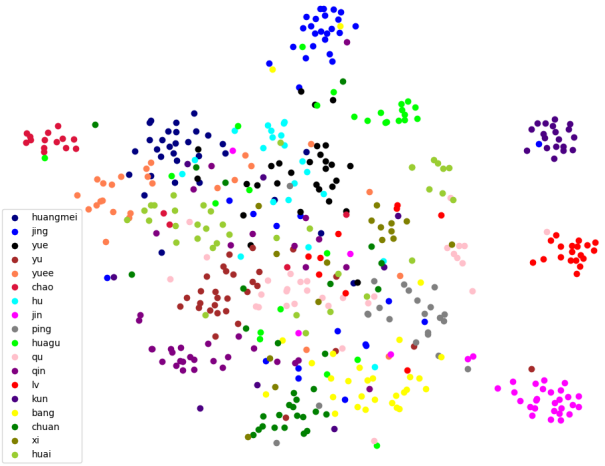

Figure 2. TSNE embedding projection of a subset of clips from the testing set, with colors distinguishing genres.

3.2 Cluster Analysis

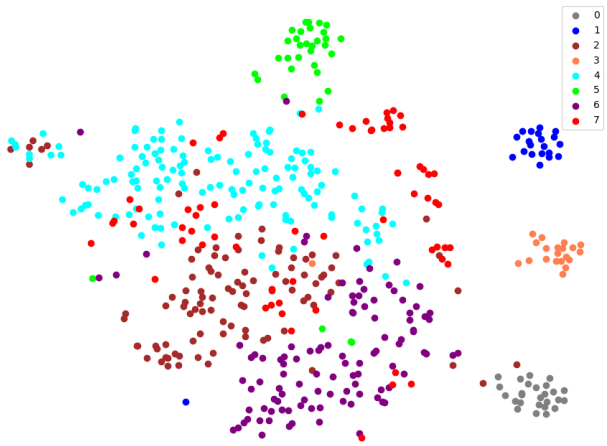

Figure 3. KMeans clustering result of the testing set clips, projected by TSNE.

With the dimension reduction algorithm T-SNE, we plotted the last layer output learnt by the model, where each segment is encoded as a 200-dimensional vector. As shown in Figure 2, several prominent clusters are already formed in the visualization of embedding space. Among the 18 genres, *Jing Ju*, *Kun Qu*, *Chao Ju*, *Jin Ju*, *Lv Ju* are the ones that stands out distinctively from the data.

What does the clustering of these features tell us if we don't provide labels? KMeans clustering divides the data in 200-dimension space into 8 clusters, and then projected into 2D space by TSNE as shown in Figure 3. Given the clustering result and comparing with 2, we can observe three prominent clusters besides the distinguishing genres: Southern genres (*Huangmei Xi - Yue Ju - Hu Ju - Huai Ju*), Yellow River region genres (*Qin Qiang - Yu Ju*), and Northern genres (*Ping Ju - Bangzi*) . It's worth noticing that the clusters exhibits close relation with geographical proximity.

3.3 Regional Influence

How much does regional influence affect the evolution of Chinese opera style? Does adjacent regions have their opera style similar to each other? We approximate the geographical location of each genre as their originate city according to [6]. With the help of Geopy [7], the physical distance between these cities are calculated. We evaluate the musical similarity using the cosine distance between each pair of genres ($18C_2 = 153$ pairs), which is calculated as the mean of all data points' last-layer embedding. As shown in Figure 4, the two variables demonstrate a weak positive correlation, which is confirmed by the Spearman's correlation coefficient ($r = 0.162$ and $p < 0.05$). Thus, music content embeddings supports the hypothesis that Chinese opera styles were influenced and mixed across regions.

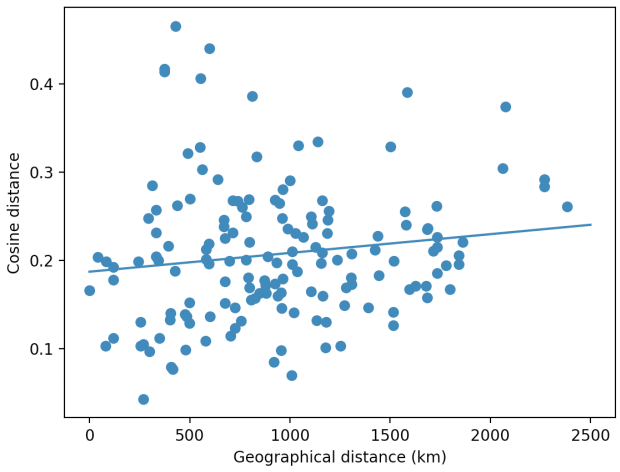

Figure 4. Geographical distance vs. cosine distance between each pair of the genres.

4. CONCLUSION

In this work, we first classified 18 regional genres of Chinese operas through Musicnn model and achieved 88% accuracy. Then, clustering and similarity analysis were performed on the last layer vectors from the learnt model.

5. ACKNOWLEDGEMENT

The research work in this paper was supported by Tencent Music Entertainment, with great help from the Audio-Music Team at WeSing.

6. REFERENCES

[1] R. Gong, Y. Yang, and X. Serra, "Pitch contour segmentation for computer-aided jingju singing training," in *13th Sound and Music Computing Conference (SMC)*, 08 2016.

[2] R. C. Repetto, R. Gong, N. Kroher, and X. Serra, "Comparison of the singing style of two jingju schools," in *16th International Society for Music Information Retrieval Conference*, 2015.

[3] Y. Yang, "Structure analysis of beijing opera arias," Master's thesis, Universitat Pompeu Fabra, 2016.

[4] J. Wang, C. Wang, J. Wei, and J. Dang, "Chinese opera genre classification based on multi-feature fusion and extreme learning machine," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2015, pp. 811–814.

[5] J. Pons, O. Nieto, M. Prockup, E. M. Schmidt, A. F. Ehmann, and X. Serra, "End-to-end learning for music audio tagging at scale," *Proceedings ofthe International Society for Music Information Retrieval Conference (ISMIR)*, vol. abs/1711.02520, 2018.

[6] "Chinese opera web," http://www.xi-qu.com/, accessed: 2021-07-14.

[7] K. Esmukov, "Geopy." [Online]. Available: https://pypi.org/project/geopy/