Scholar Publishing Group
International Journal of Art Innovation and Development
https://doi.org/10.38007/IJAID.2023.040107
ISSN 2789-7192 Vol. 4, Issue 1: 94-101

Intelligent Music Chorus Recognition and Evaluation System Based on Deep Learning

Xi Wang[1, 2*]

[1]*Jinzhong Normal Junior College, Jinzhong, China*

[2]*Philippine Christian University, Manila, Philippine*

hcyzh1111@163.com

[*]*corresponding author*

Keywords: Deep Learning, Timbre Analysis, Neural Network, Feature Extraction

Abstract: The key to the chorus recognition is to find the effective representation of the chorus timbre. Through the analysis of the temporal frequency domain features, inverted spectrum features, sparse features and probability features, it shows that the chorus can be effectively identified by using the temporal and frequency domain features. Based on this result, using the layer-by-layer abstraction feature of deep learning, the advanced time-frequency representation of the chorus timbre is extracted for the chorus recognition. This paper aims to study the design of an intelligent music chorus recognition and evaluation system based on deep learning. Considering the problem of high classification error, using the time domain frequency domain characteristics, clutter characteristics, sparse characteristics and probability features to identify the adverse effects of impact instrument, put forward a low noise instrument recognition model is a use of cochlear model harmonious decomposition music, including time and frequency information auditory spectrum, similar to the human hearing. In order to combine the feature expression ability of the serial-level noise cancellation encoder and the abstract feature ability of the deep belief network, the five-layer deep hybrid network is constructed for the deep learning framework with the above two basic modules. It is proved that the evaluation error of chorus recognition and professional judges is less than 5%. And it greatly improves the evaluation efficiency.

1. Introduction

With the development of technology, information sharing is an important means of machine interaction. Due to the large amount of data, voice signal is becoming more and more important in

human society, becoming the most important information provider in human-computer interaction. Voice signal has become important information for human beings. Therefore, as the first step of human-computer interaction, due to its complex application scenarios and computing limitations, the application of speech recognition technology may face various challenges, and become the focus and difficulty of research [1-2].

In the design of intelligent deep learning music chorus recognition evaluation system, many scholars study, and achieve good results, for example: Abdalla A N made an experiment, experiment through different voice signal through six different band-pass filter, each filter in deflection magnetic field is controlled by different controller, finally by observing the degree of deflection between signal and trajectory to distinguish the difference between the sound signal [3]. Wu J, et al., applied the deep residual network to the keyword recognition task for the first time, and achieved a good effect of [4] by adjusting the depth and width of the model.

This paper starts from the two aspects of acoustic model and acoustic features, and studies the application of deep neural network in chorus recognition and evaluation: it completes the work of feature extraction of deep automatic encoder based on supervised learning method. Mainly for feature preprocessing mode, hidden layer structure determination, for the training network data quantity and initial parameter selection of the deep neural network construction, training and performance in feature extraction based on DNN-HMM to build the overall acoustic model, through the experimental results to analysis. Comparing DNN-HMM and GMM-HMM in training methods and network structure use DNN to replace GMM to describe HMM state output probability distribution. Meanwhile, use Kaldi open source speech recognition platform for related experimental work, and finally verify the effectiveness and feasibility of DNN-HMM model as an acoustic model.

2. Design and Research of the Intelligent Music Chorus Recognition and Evaluation System Based on Deep Learning

2.1. Intelligent Music Chorus Recognition and Evaluation System Based on Deep Learning

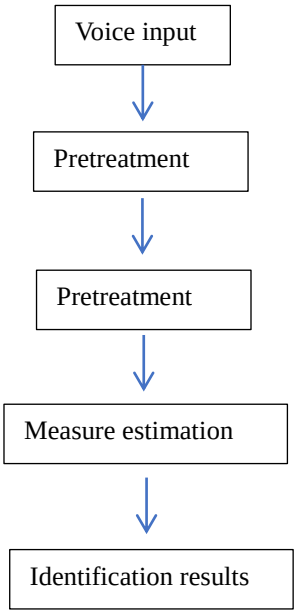

Figure 1. Identification system block diagram

With the support of the deep autoencoder, this paper can not only obtain the efficient encoding method, but also obtain the representation of the raw data, which is the feature extraction. It is able to test with the raw input as the standard when the training sample category is unknown, using unsupervised feature learning methods to process a series of unlabeled data information. The most representative autoencoder, composed of the input layer, the hidden layer, and the output layer, can be called a depth autoencoder if the number of hidden layers exceeds 1. If the number of hidden layer nodes is fewer than the input layer, the main purpose of such encoders is feature compression; otherwise, the features are the maps to the high-dimensional space [5-6]. The schematic diagram of the recognition system is shown in Figure 1. The detailed software design process is shown in Figure 2.

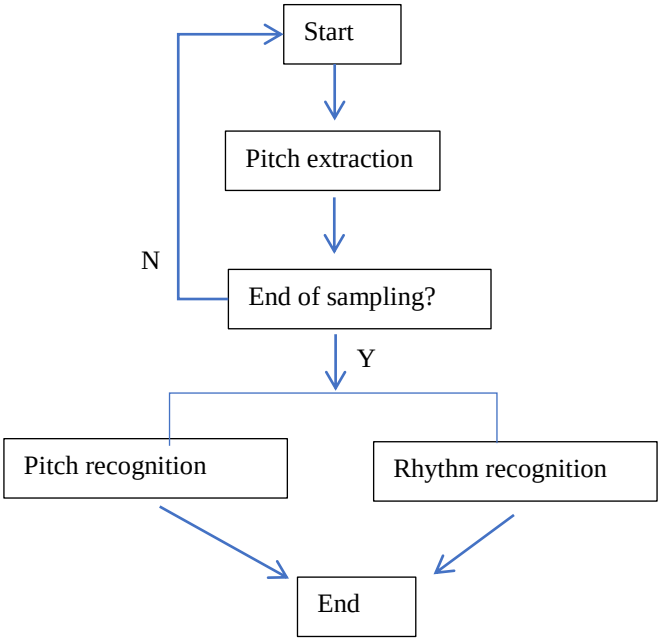

Figure 2. System flow chart

2.2. Defects of Speech Recognition Technology

Although the traditional speech recognition technology, including HTK, has long been very mature, and the development of speech recognition technology has also accumulated rich achievements, but the existing technology is not enough to achieve a complete free interaction between man and machine. Therefore, in-depth research in this area is necessary. With the advent of the era of big data, the development of this technology faces greater challenges. First of all, the rapid development of mobile Internet and the increasing improvement of multimedia technology greatly reduce the difficulty of obtaining voice data, but the lack of annotated data cannot be directly[7-8]provided as input to traditional speech recognition systems; Secondly, Moore's Law observed the computer hardware industry for a long time shows that every year to one and a half years, the storage capacity will decrease by 50%. Cloud computing, computing cluster and GPU parallel computing framework are widely used in the Internet industry. How to give full play to the performance of the above technologies in data mining in the field of pattern recognition is worth more thorough research. The traditional machine learning algorithms do not provide us with answers. From the technical level, the problems facing traditional speech recognition technology are mainly reflected in these aspects of [9-10]:

(1) In terms of feature extraction. Traditional speech feature extraction methods require experts to study the nature of acoustic features. It usually requires a lot of experiments, which takes a long time to study. However, in the new information age background, people's needs change very fast. If the method of directly processing a large amount of data and obtaining features under specific constraints can be invented, the performance of the speech recognition system will achieve a leap forward again[11].

(2) In the aspect of acoustic model creation. At present, most speech recognition systems on the market are applied to the GMM-HMM model, but the GMM model belongs to the shallow model, and is not suitable for complex data scenarios. The method to increase the model accuracy by increasing the mixed number of models is not a long-term solution, and this method can play a very limited role, especially as the current voice data increases at an unprecedented rate, and the data becomes more diverse and complex. Therefore, to fundamentally solve these problems, we need to find the acoustic modeling methods with a strong versatility, and can more completely and accurately characterize the data. Deep neural networks may be the answer we need.

(3) In the aspect of language model modeling. Today, statistical language models have the problem of sparse data. Although we can alleviate this problem through smoothing methods such as fallback and interpolation, the language model based on statistical data is imperfect from a theoretical point of view. For example, the memory length of some words is different, simply counting the probability between N words directly next to to estimate the probability of word appearance. There is a certain difference between this method and the actual situation. It is fully proved that for the problem of language model modeling, the academia should strengthen the research on the nature of language phenomena, and propose a more scientific and effective modeling method, [12-13].

2.3. Algorithm Selection

In the process of normalization the dimension of hyperframe features, the Gauss-Bernoulli RBM model is selected, where the energy function is:

$$E(v,h)=\sum_{i\in V}\frac{(v_i-a_i)^2}{2\sigma_i^2}-b^Th-\sum_{i\in V,j\in H}\frac{v_i}{\sigma_i}h_jw_{ij} \tag{1}$$

After CMVN processing, the input data distribution in equation (1) meets ai = 0 and, and the energy function is equivalent to [14-15]:

$$\sigma_i=1 \tag{2}$$

$$E(v,h)=\sum_{i\in V}\frac{v_i^2}{2}-b^Th-\sum_{i\in V,j\in H}v_ih_jw_{ij} \tag{3}$$

3. Design, Research and Design Experiment of Intelligent Music Chorus Recognition and Evaluation System Based on Deep Learning

3.1. Pre-Training

Compared with deep neural networks, the pre-training mode is no different. The autoencoder also defines the RBM as the basic modeling unit of the network, which is trained hierarchical to complete the initialization of various parameters (such as connection weight and node offset). Although the gradient of each training sample can be found to update the weight, after the training

set is subdivided into several training subsets, the efficiency of defining the subset defined as unit update weights can be greatly improved. In the final analysis, the latter method can reflect the characteristics of GPU or Matlab efficient matrix operation. Regarding the structure of deep autoencoder in the experiment, two different types of RBM models, Gauss-Bernoulli RBM and Bernoulli-Bernoulli RBM, were trained to realize parameter initialization.

3.2. Experimental Design

This paper is based on deep learning intelligent music chorus recognition evaluation system design experiment, first is the accuracy experiment, music recognition accuracy of the system for the corresponding experiment, in the analysis in different groups and different scale factors and accuracy, the second is judged by professional judges to the accuracy of the chorus staff recognition evaluation.

4. Design Research and Experimental Analysis of the Intelligent Music Chorus Recognition and Evaluation System Based on Deep Learning

4.1. Accuracy Experiment

This paper analyzes the accuracy of the deep learning-based intelligent music chorus recognition and evaluation system. When the scaling factor is unchanged, by increasing the number of groupings to judge the accuracy and the accuracy changes of the different scaling factors under the same grouping, the experimental data are shown in Table 1.

Table 1. Accuracy of different groups g

	g=1	g=2	g=3	g=4	g=8
S=1	92.12	94.12	90.13	92.76	92.37
S=2	93.76	94.89	94.51	94.32	94.17

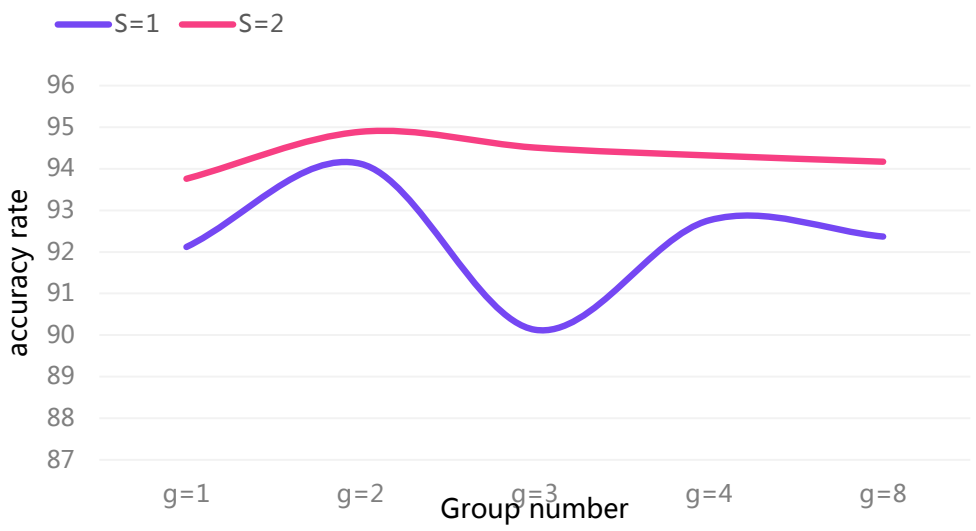

Figure 3. Changes in accuracy with different scaling appearances

As can be seen from Figure 3, when the scale factor s is increased, the recognition accuracy is improved under different group number g values, but the number of parameters increases accordingly. Adjust the size of the model by adjusting the size of s. Generally, the higher the model size, increase the recognition accuracy, therefore, which proves the effectiveness of group convolution and channel rearrangement under a certain number of parameters.

4.2. Evaluation Accuracy

The recognition accuracy in the chorus recognition system is an important criterion. This paper invites several professional music judges to score the chorus members separately, and then lets five members use the chorus at the same time. The chorus recognition and evaluation system constructed in this paper gives a unified score. The experimental data are presented in Table 2.

Table 2. Error comparison of the professional evaluation and the systematic evaluation in this paper

	1	2	3	4	5
Professional evaluation	7.9	7.6	8.3	8.6	8.2
Systems assessment	7.9	7.7	8.3	8.5	8.3

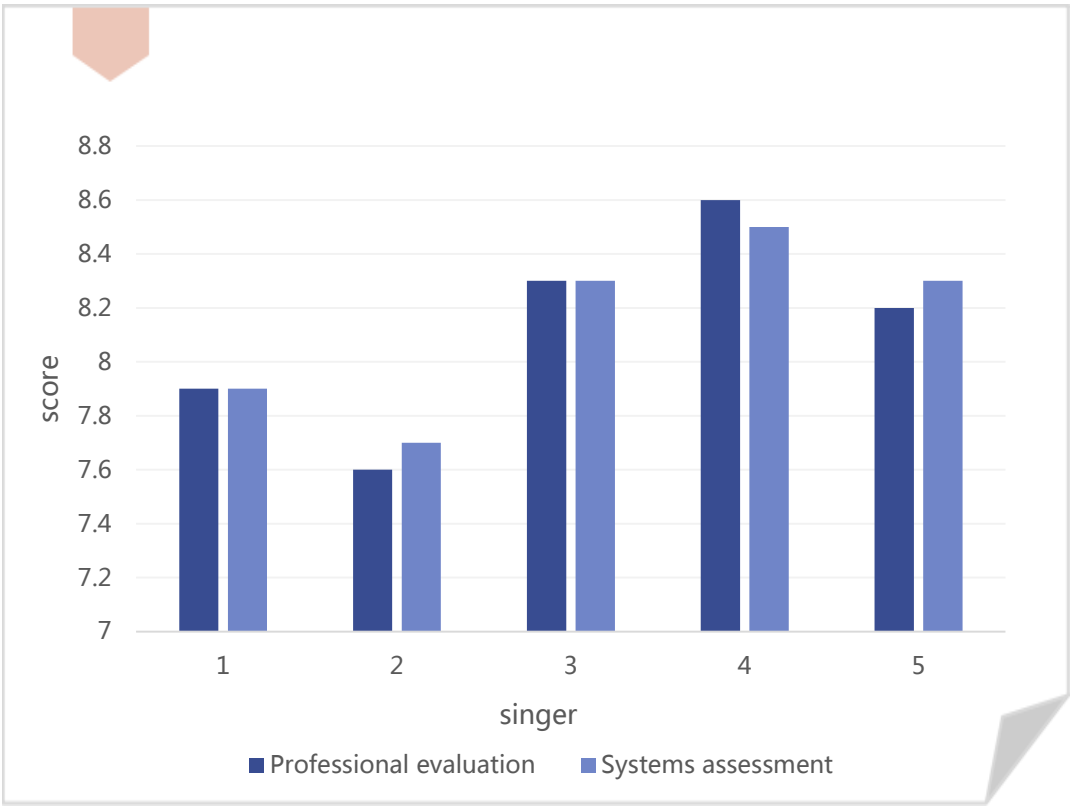

Figure 4. Different scores of the professional evaluation and systematic evaluation of the five players

From Figure 4, it is obvious that the chorus recognition and evaluation system constructed in this paper has less evaluation error compared with professional judges. And this paper can accurately

score each person in the unified chorus, which is far lower than the professional jury in terms of time and efficiency. It can effectively improve the effectiveness of recognition and evaluation during chorus.

5. Conclusion

The acoustic feature extraction is the research object based on the deep autoencoder model. Deep autoencoder is a multi-layer network model, which is implemented based on unsupervised training, and is widely used in the fields such as data dimension reduction and feature extraction. In this paper, this deep learning model is analyzed on feature data preprocessing, model structure and network training parameters. The speech data used in this paper is limited, which can be verified by collecting more abundant and comprehensive data.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1]Premalatha G, Bai VT. Design and implementation of intelligent patient in-house monitoring system based on efficient XGBoost-CNN approach. Cognitive Neurodynamics. (2022) 16(5): 1135-1149.

[2]Abdulkadhim EG .Design and Develop an Attendance System Based on Fingerprint and Arduino Board. Journal of Physics Conference Series. (2021) 1804(1): 012011.

[3]Abdalla AN, Nazir MS, Tao H, et al. Integration of energy storage system and renewable energy sources based on artificial intelligence: An overview. The Journal of Energy Storage. (2021) 40(2021): 102811.

[4]Wu J. Intelligent Building Design Based on Internet of Things. Journal of Physics: Conference Series. (2021) 2143(1): 012032-.

[5]Guan H, Shizhong HE, Qiuqiu LI, et al. A Review of Convolutional Neural Networks in Equipment Wear Particle Recognition. Tribology. (2021) 42(2): 426-445.

[6]Qi Y, Mu S, Wang J, et al. Intelligent Recognition of Transmission Line Inspection Image Based on Deep Learning. Journal of Physics: Conference Series. (2021) 1757(1): 012056 (6pp).

[7]Wei J, He J, Meng W, et al. Application of Running Speed Coordination on the Analysis of Road Alignment Intelligent Recognition. Journal of Physics: Conference Series. (2021) 1982(1): 012053 (5pp).

[8]Wang F, Zhang Y, Sun A, et al. Research on Music Fountain Design based on Intelligent Water System. IOP Conference Series Earth and Environmental Science. (2021) 690(1): 012021.

[9]Guo C. Research on pre-competition emotion recognition of student athletes based on improved machine learning. Journal of Intelligent and Fuzzy Systems. (2020) 39(4): 5687-5698.

[10]Ma C, Yang P. Research on Classroom Teaching Behavior Analysis and Evaluation System Based on Deep Learning Face Recognition Technology. Journal of Physics: Conference Series. (2021) 1992(3): 032040 (7pp).

[11]Hao K, Yang G, Zhang L, et al. Study on model selection and evaluation method of weapon system based on ISM and GAHP. Journal of Northwestern Polytechnical University. (2022) 40(4): 926-934.

[12]Pinto D, Arnau JM, A Gonz á lez. Design and Evaluation of an Ultra-Low-power Human-quality Speech Recognition System. ACM Transactions on Architecture and Code Optimization. (2020) 17(4):1-19.

[13] Zhang W, Su C, He C. Rehabilitation Exercise Recognition and Evaluation Based on Smart Sensors With Deep Learning Framework. IEEE Access, 2020, PP(99):1-1.

[14] Nie X. Intelligent analysis of classroom student state based on neural network algorithm and emotional feature recognition. Journal of Intelligent and Fuzzy Systems, 2020, 40(1):1-12.

[15] Sun H, Liu Z, Wang G, et al. Intelligent analysis of medical big data based on deep learning. IEEE Access, 2019, PP(99):1-1.