Classification of Lung Sounds Using CNN-Attention

Chenghan Li, Huaichang Du and Bing Zhu

October 10, 2020

Classification of lung sounds using CNN-Attention

Chenghan Li ,Huaichang Du ,Bing Zhu
School of Information and Communication Engineering,Communication University of China,Beijing,China
944653839@qq.com, duhuaichang@cuc.edu.cn, zhubing@cuc.edu.cn
Corresponding Author: Chenghan Li Email:944653839@qq.com

Abstract—**Respiratory disease is a kind of disease which causes a high mortality rate, whether in rural areas or in cities. It is necessary to detect respiratory diseases in advance, and with the rapid development of deep learning in recent years, it will become a new trend of disease detection to analyze and calculate the respiratory sound through the computational power of computer.**

Attention mechanism has recently been proposed to deal with sequence problems (such as audio, text, etc.),It is claimed to be a kind of network that can replace RNN, LSTM and GRU. By imitating human visual attention mechanism, it can learn a weight distribution of data features, and then apply this weight distribution to the original features, so that the task mainly focuses on some key features, neglects unimportant features, and improves task efficiency.On the basis of the ICBHI benchmark data set, we propose a network, which combines CNN and attention mechanism to classify audio data and distinguish abnormal breath sounds from normal respiratory sounds. And we have carried on the experiment to test the robustness of the model.

Keywords—deep learning ; CNN ; Attention mechanism

I. INTRODUCTION

Respiratory system disease is a kind of disease that can not be ignored in today's world, and occupies a large proportion in human diseases. And respiratory sound is an important reference index of respiratory system health. For example, wheezing is a common phenomenon in patients with asthma or chronic obstructive pulmonary disease (COPD)[1]. Although there are many detection methods for various specific respiratory diseases in medicine, with the development of big data and the increasing computing power of computers, deep learning technology is used to train a good classifier to classify the respiratory sounds of patients and judge whether there is abnormality. This method may be more cost-effective than traditional medical diagnosis. If the trained model has good robustness, it may even have higher accuracy than traditional medical diagnosis.

In this paper, we propose a cnn-attention classification model, since the attention mechanism is widely used to deal with sequence problems, and introduce a new data preprocessing method, which has better effect than previous experimental results.As a complex information system, large-scale network has complex system properties, usually unable to implement the overall survivability strategy and unified management. For example, the backbone of Internet has no global policies in consideration, the reason is that it does not exist a global management. Therefore, simple network survivability structure does not apply to open complex systems. Large-scale network survivability mainly considers the system as a whole to provide the critical services survivability. The theories and methods of large-scale network survivability should be investigated and proposed from the essential characteristics of large-scale network as an open complex system.

Related work: Koki Minami, Huimin Lu et al., classify large respiratory sound data sets based on convolutional neural network[2]; Lin Li, Wenhao Xu et al., used deep neural network to classify normal and irregular lung sounds[3]; Murat aykanat et al., also studied lung sounds classification based on convolutional neural network[4]; Valentyn vaityshyn et al. Proposed a convolutional neural network system for classification of bronchial diseases by lung sounds[5]; Diego Perna, Andrea tagarelli et al. Proposed to predict respiratory abnormalities and diseases by recurrent neural network[6]; Kirill kochetov, Evgeny Putin et al. Proposed a noise masking recurrent neural network for the classification of respiratory sounds[7]; Valentyn vaityshyn, Hanna porieva et al. Proposed a pretrained convolutional neural network to classify lung sounds[8].

II. THE DATABASE AND DATA PREPROCESSING：

A. Lung sound database Database：

In our experiment, we used the ICBHI dataset, which was built in the context of a challenge on respiratory data analysis organized in conjunction with the 2017 Int. Conf. on Biomedical Health Informatics (ICBHI).The audio samples of breath data ICHBI challenge database were collected by two laboratories in two different countries, including a total of 5.5 hours of records, including 6898 respiratory cycles, of which 1864 contained cracks, 886 contained wheezes, and 506 contained both cracks and wheezes. These samples were from 126 subjects[9]. The ICBHI provides the starting position of each respiratory cycle in each respiratory sound frequency, in which the audio of each breathing cycle is marked as normal, crackle, wheeze, or both. Corresponding to our project, the data is actually divided into four categories, in which the ICBHI dataset provides a label for each piece of data. Crackle is a discontinuous and explosive lung sound, which can be divided into two types according to its duration: coarse crack greater than 10ms, fine crack less than 10ms. The frequency range of cracks is 60-2000 Hz, and the most informative frequency is up to 1200 Hz . On the contrary, wheezing is a kind of high continuous and irregular lung sound, which is generally characterized by 400Hz main frequency and sinusoidal waveform. The standard definition of continuous sound is that the duration is more than 250 ms, but the wheezing sound is not necessarily. It is usually longer than 80-100 Ms. Severe obstruction of lower respiratory tract or upper respiratory tract in thoracic cavity may be accompanied by inspiratory wheezing. Patients with asthma and chronic obstructive pulmonary disease (COPD) develop systemic airway obstruction. However, healthy people can even detect wheezing at the end of expiratory period after forced exhalation[10] .

B. Data preprocessing:

For We propose a new model for lung sound classification. Based on our model, we also propose a novel data preprocessing method to deal with the uniqueness of ICBHI data.

We extract the MFCC feature of audio as the input of the model. MFCC is the abbreviation of Mel frequency cepstrum coefficient. It is based on the auditory characteristics of human ears, and it has a nonlinear relationship with Hz frequency. And mel frequency cepstrum coefficient (MFCC) is the frequency spectrum characteristics calculated by using the relationship between them. MFCC has been widely used in speech recognition. Because of the nonlinear relationship between Mel frequency and Hz frequency, the accuracy of MFCC decreases with the increase of frequency. Therefore, in application, only low frequency MFCC is used, while medium and high frequency MFCC are discarded.

Because the length of each cycle of icbhi data is not equal and there is a certain duration of respiratory cycle, MFCC features extracted by the same window length and frame shift have different dimensions. In the experiment, the feature vector extracted from each breathing cycle audio is (x, 20) dimension, where x depends on the duration of each breathing cycle audio. In order to avoid different dimensions of input vectors, we flatten the extracted features into one dimension, and then fill the data with 0 to a custom length. We fill each cycle data to 20000 sample points. Before inputting the model, reshape the data into the shape required by the first layer network. Please refer to figure 1 for the specific process.

We believe that the same laws of data reconstruction will not break the internal rules of the data itself. For the end-to-end deep learning method, the data reconstruction of this rule has little effect on the model training.

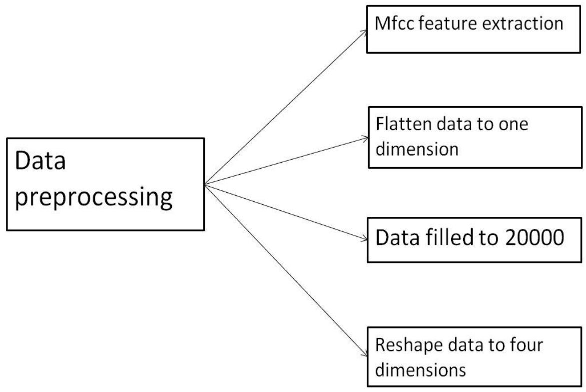

Fig. 1. The framework of data preprocessing

III. THE METHOD：

When Nowadays, recurrent neural network（RNN） is widely used in natural language processing, which has its own advantages in dealing with sequence problems.But the disadvantage of recurrent neural network is that it will be affected by short-term memory. If a sequence is long enough, it will be difficult for them to transfer information from earlier time step to later time step[11].For our task, the time-frequency structure of abnormal lung sounds is a large whole. However, due to the short-term memory ability of RNN, the model may only learn the local characteristics of lung sounds, but not have a good description of the whole, which may bring bad interference to the final experimental results.

Attention mechanism can solve this problem very well. The design concept of attention mechanism is to design a scoring function. For each attention vector (each input sequence), a score is calculated, and the scoring basis is the correlation degree of each different attention vector. The more relevant, the greater the value. In this way, the correlation degree of each input vector is calculated in parallel, which avoids the defects of RNN.

Specifically, each input vector is copied into three vectors called query, key and value, abbreviated as Q, K, V. each vector makes dot product of its own Q vector and K vector of other input vectors respectively, and the resulting fraction represents the correlation between vectors. Then, make a linear transformation on this result vector and do a dot product with its own V vector. Finally, the attention between two vectors can be obtained through a softmax layer.The formula is as follows:

$$Attention(Q,K,V) = soft\max(\frac{QK^T}{\sqrt{d_k}})V \qquad (1)$$

Where d_k is the dimension of the K vector[12].

In addition, convolutional neural network (CNN) can also be used in audio processing. Convolutional neural network (CNN) is a kind of feedforward neural network.

Its artificial neurons respond to a portion of the surrounding units within the coverage. It is first used for image processing, and then used for natural language. The most important parts of CNN are convolution layer and pool layer. There is a convolution kernel in the convolution layer, which is calculated by sliding windows one by one on the upper input layer. Each parameter in the convolution kernel is equivalent to the weight parameter of the depth neural network, and is connected with the corresponding local pixels. Multiply the sum of the convolution kernel parameters with the corresponding local pixel values, and then add them to get the convolution layer result[13].

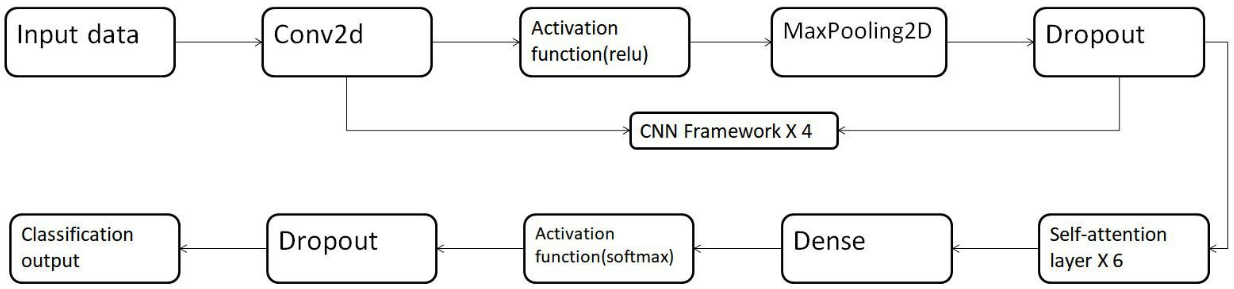

Fig. 2. The model of the neural network

The purpose of pooling is to reduce the size of the image through a down sampling process after convolution of the original image. Pool operation is divided into mean pooling, which is to calculate the average value of all points; max pooling, which is to calculate the maximum value of all points.

We The model uses four layer convolution neural network to extract the local features of audio MFCC and down samples the data. We use six attention layers to learn a set of phoneme features in time series. Finally, we input a linear classification layer and output the probability of four categories. The advantage of this modeling is to ensure that the local features of the audio spectrum are learned, and then a better learning representation of the overall feature structure is obtained. At the same time, the convolution neural network downsampling the audio information, so that the final modul training results also reduce some unnecessary noise interference, and finally achieve a good effect. The model structure is shown in Figure 2.

IV. EXPERIMENTAL RESULTS：

We chose Adam as the optimizer of the network and categorical_crossentropy was used as the loss function,

$$L_{\log}(Y,P) = -\log\Pr(Y \mid P) = -\frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{K-1} y_{i,k} \log p_{i,k} \tag{2}$$

where N−total number of samples, K−number of tags, pi,k−the probability that the i-th sample is predicted to be the k-th tag value, yi,k−prediction value of tag probability.

We also choose accuracy as the standard of model robustness,

$$\text{acc} = \frac{tp+tn}{tp+tn+fp+fn} \tag{3}$$

where tp−true positive=correctly identified, fp−false, positive=incorrectly identified, tn−true negative =correctly rejected, fn−false negative =incorrectly rejected.

We divided the experimental results into three parts. We first evaluated the effect of the number of attention layers on the experimental results, and found the most appropriate number of layers for this dataset. Then we compared the robustness of the attention with Gru, bigru, LSTM and bilstm based on the method of down sampling audio data using CNN. The results show that attention is better than others. Finally, we compared the whole model with some classical models, and the results show that for this ICBHI dataset, the classification results of this model are better.

A. Optimal construction of the model

In the field of deep learning, over fitting and under fitting are always existing problems. In this part, we test how many layers of attention is the optimal value for ICBHI dataset. We set the attention of layer (3, 4, 5, 6, 7, 8) as (S3, S4, S5, S6, S7, S8). The experimental results are shown in Figure 3.

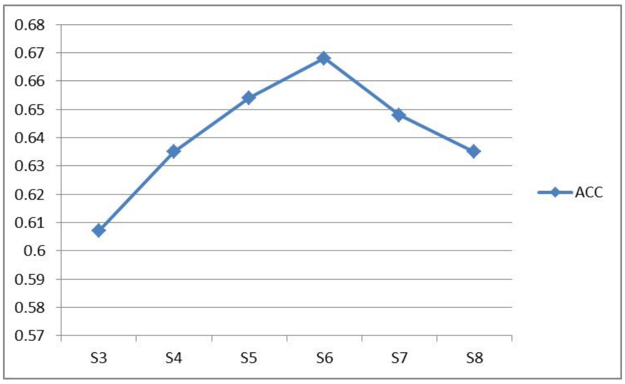

Fig. 3. The influence of the number of attention layers on the accuracy

B. Test the robustness of attention

Once proposed, attention mechanism has been widely used in various research fields, and is regarded as one of the most advanced models to deal with sequence problems. In this part, we compare attention with RNN, BIRNN, GRU, BIGRU and so on. The classification accuracy of attention is better than other models.Our experimental results are based on the premise of using CNN to train audio data.The experimental results are shown in Figure 4.

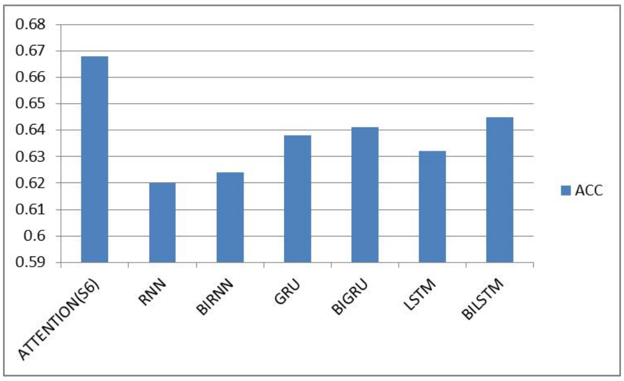

Fig. 4. The attention part compares with other sequence models

C. Compare the whole model with other classical models

In this part, we compare the experimental results of the whole model with other related classical models, and the proposed model has better robustness. The experimental results are shown in Figure 5.

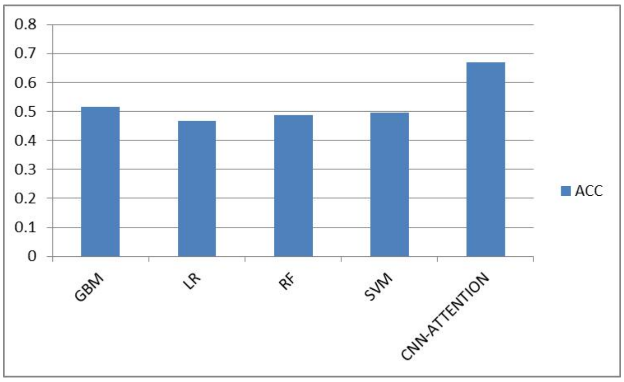

Fig. 5. Compared with other classical machine learning models

V. CONCLUSION:

This paper proposes a network model based on CNN model and attention mechanism for classification tasks. Attention is one of the best models to deal with sequence problems. Based on the icbhi data set, this paper proposes a new data and processing method for this data set. The lung sound audio is divided into four categories, and a better classification accuracy is obtained.

REFERENCES

[1] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano and S. Kido, "Automatic Classification of Large-Scale Respiratory Sound Dataset Based on Convolutional Neural Network," 2019 19th International Conference on Control, Automation and Systems (ICCAS), Jeju, Korea (South), 2019, pp. 804-807, doi: 10.23919/ICCAS47443.2019.8971689.

[2] L. Li, W. Xu, Q. Hong, F. Tong and J. Wu, "Classification between normal and adventitious lung sounds using deep neural network," 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP), Tianjin, 2016, pp. 1-5, doi: 10.1109/ISCSLP.2016.7918407.

[3] KV. Vaityshyn, M. Chekhovych and A. Poreva, "Convolutional Neural Networks for the Classification of Bronchopulmonary System Diseases with the Use of Lung Sounds," 2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO), Kiev, 2018, pp. 383-386, doi: 10.1109/ELNANO.2018.8477483.

[4] K.Q. Zhao, A.L. Xuan, "Set Pair Theory - A New Theory Method of Non-Define and Its Applications," System Engineering, vol. 14, issue 1, pp. 18-23, 1996.

[5] D. Perna and A. Tagarelli, "Deep Auscultation: Predicting Respiratory Anomalies and Diseases via Recurrent Neural Networks," 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain, 2019, pp. 50-55, doi: 10.1109/CBMS.2019.00020.

[6] Kochetov K., Putin E., Balashov M., Filchenkov A., Shalyto A. (2018) Noise Masking Recurrent Neural Network for Respiratory Sound Classification. In: Kůrková V., Manolopoulos Y., Hammer B., Iliadis L., Maglogiannis I. (eds) Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018. Lecture Notes in Computer Science, vol 11141. Springer, Cham. https://doi.org/10.1007/978-3-030-01424-7_21

[7] V. Vaityshyn, H. Porieva and A. Makarenkova, "Pre-trained Convolutional Neural Networks for the Lung Sounds Classification," 2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 2019, pp. 522-525, doi: 10.1109/ELNANO.2019.8783850.

[8] World Health Organization, The top 10 causes of death, https://www.who.int/en/news-room/factsheets/detail/the-top-10-causes-of-death (Accessed: 22 May 2019)

[9] International Conference in Biomedical and Health Informatics, ICBHI 2017 Challenge, https://bhichallenge.med.auth.gr (Accessed: 14 May 2019)

[10] D S Morillo, S A Moreno, M A F Granero, et al. Computerizedanalysis of sounds during COPD exacerbations [J]. Computers inbiology and medicine, 43(7): 914-921, 2013.

[11] C. Laurent, G. Pereyra, P. Brakel, Y. Zhang, and Y. Bengio, "Batch normalized recurrent neural networks," in Procs IEEE Int. Conf. On Acoustics, Speech and Signal Processing, 2016, pp. 2657–2661.

[12] C. Wen, M. Hong, X. Yang and J. Jia, "Pulmonary Nodule Detection Based on Convolutional Block Attention Module," 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 8583-8587, doi: 10.23919/ChiCC.2019.8865792.

[13] D. Perna, " Convolutional neural networks learning from respiratory data," in Proc. IEEE Int. Conf. on Bioinformatics and Biomedicine, 2018, pp. 2109–2113.