Study on Speech Recognition Method of Artificial Intelligence Deep Learning

Study on Speech Recognition Method of Artificial Intelligence Deep Learning

Zhang Leini[1], Sun Xiaolei[2]

[1]qingdao institute of technology

[2]jiaozhou city vocational education center

Abstract: This paper briefly describes the basic principles and related theories of speech recognition system, points out the existing problems of speech recognition technology of artificial intelligence deep learning, analyzes the speech recognition methods of artificial intelligence deep learning, puts forward the optimization methods of speech recognition technology of artificial intelligence deep learning from the aspects of strengthening targeted feature recognition of speech system, repeatedly carrying out simulation training of speech recognition and combining acoustic features with sports features, and forecasts the future prospects of speech recognition methods of artificial intelligence deep learning.

1. Introduction

The continuous development of information technology has further deepened the research of artificial intelligence and brought many changes to people's work and life. Although speech recognition based on artificial intelligence deep learning has made great progress, there are still some shortcomings, and the accuracy of speech recognition needs to be improved[1]. Therefore, the author puts forward the following opinions based on his own experience, so as to promote the sound development of speech recognition in artificial intelligence deep learning.

2. Basic Principles and Related Theories of Speech Recognition System

Speech recognition refers to the transformation of speech information involved in speech into data that can be recognized by computer system by means of scientific methods, and then provides a variety of services for machines or people. Speech recognition system usually consists of the following modules: acoustic related model, language related model, decoder and acoustic extraction processing module. The working principle of the speech recognition system is to collect the characteristic information in the speech information model, build the acoustic model with the help of training or other methods, make it match the speech model, and then use scientific algorithms to decode this kind of information to get the same data information as the original information.

HMM is also called hidden Markov acoustic model, which is used to express the relationship between sequence and hidden state in speech. Nearly half of speech recognition systems will apply HMM, and then establish acoustic model. Its structure diagram is shown in Figure 1. In essence, HMM is a probabilistic model, which uses parameters to represent random statistical states and characteristics, and is mainly composed of random function HMM and fixed state number HMM. Because HMM can analyze the local smooth characteristics of speech and the overall stable characteristics of speech, it can create corresponding sound models according to the time series signals of speech, so HMM is widely used in acoustic modeling [2].

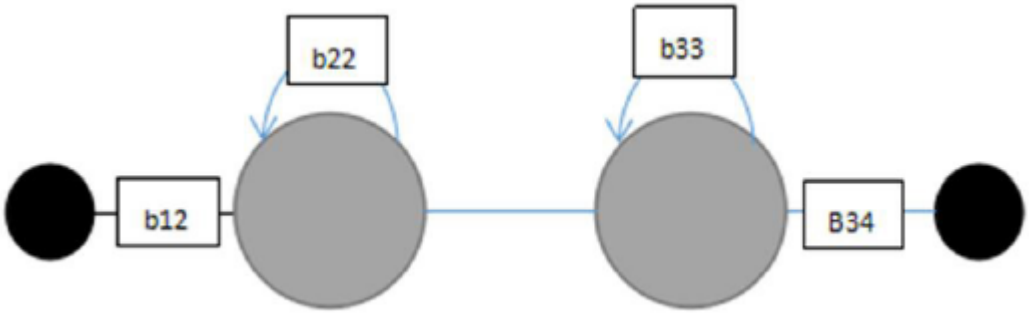

Figure 1 HMM Structure Diagram

3. Problems in Speech Recognition Technology of Artificial Intelligence Deep Learning

3.1 There is noise interference

At present, there are still some problems in solving noise interference in speech recognition system. If there is noise around the speaker, or if the speaker's tone, mood and intonation make the pronunciation inaccurate or unclear, the speech system cannot effectively recognize the speech information.

3.2 The endpoint detection level is not high

In the process of signal recognition, endpoint detection technology is very important. Apart from noise interference, even in a quiet state, speech signal recognition errors may occur, which are mainly caused by endpoint detectors. In a certain sense, in order to improve the speech recognition technology, the most important thing is to optimize the endpoint detection technology, and in order to meet this requirement, it is necessary to explore more stable speech parameters [3].

4. Analysis of Speech Recognition Methods for Artificial Intelligence Deep Learning

4.1 Speech pickup and feature extraction in speech recognition methods

For speech pickup, there are two main links, one is sampling and the other is endpoint detection. The former refers to collecting samples of sound data, and then converting the analog signals collected by the converter into digital audio, which is the first step of voice pickup. In this step, the sound card uses a frequency twice as high as the frequency when collecting voice signals, so as to avoid distortion caused by low frequency.

Endpoint detection, also known as speech detection and speech boundary detection, refers to recognizing noise and speech during collecting speech samples in noisy environment, reducing speech coding speed, reducing equipment energy consumption and communication broadband, thus improving recognition rate. At present, there are two problems to be solved in the end-point detection. One is the noise problem in the environment. How to distinguish the sound signal from the noise accurately and efficiently means that it is necessary to explore the speech parameters with stable characteristics and determine the features of speech extraction. Another problem is the cutting of leading and trailing edges, that is, there is a time delay from when people make sounds to before speech detection. Therefore, the beginning and end of speech waveform will be regarded as mute reduction, resulting in the difference between the restored speech and the original speech [4].

4.2 Simulation training and recognition judgment in speech recognition methods

It is mainly related to two aspects: first, training. It refers to taking the fixed recognition method as a standard, obtaining the corresponding speech parameters by training or accumulating the same kind of methods, and then saving the obtained speech parameters as reference templates, and all the reference templates together form a reference template library, which serves as an important reference for speech recognition. Secondly, compare the recognition sample with the reference template, and take the one

with the highest similarity as the recognition word. There are three main methods of comparison: (1) The results extracted from speech feature training are directly saved and used as a reference standard. In the recognition process, the vector sequence of the word to be recognized is obtained by extracting the input speech signal and training features, and then the vector sequence is compared with each template, and the smallest distance is judged as the required character. (2) Judging Chinese characters to be recognized mainly by dynamic graph. To solve the problem of dynamic time matching of speech, we choose a scientific method to divide the sequence of language features into n segments, calculate the average position of each segment of feature sequence, get n feature vectors, and use them as templates [5].

5. Optimize artificial intelligence deep learning speech recognition methods

5.1 Strengthen the targeted feature recognition of speech system

For speech recognition system, extracting speech signal features is a very important step, which can quantify huge speech signals, and then explore some features that can represent speech signals, and then analyze and process acoustic models. In the deep learning research of artificial intelligence, only the image recognition effect is ideal, which can guarantee the effect of speech recognition. Compared with other methods, it has obvious advantages. The training method of deep learning is special, which can provide many good initial weights and stresses to the neural network. On the one hand, it can ensure that the neural network model will not have local optimal solution during the training period, and can converge to reasonable extremum. On the other hand, the deep neural network can also describe the essential characteristics of the original speech information, improve the distinguishability of the data, and improve the recognition ability of the speech recognition system. In addition, deep learning can obtain depth features, and even if the dimension is reduced, the original information will not be damaged, which can ensure good factor resolution. Using depth neural network to express data layer mapping can obtain the depth essence of the original data, and then strengthen the recognition ability of speech recognition system.

5.2 Repeatedly carry out simulation training of speech recognition

During the pre-processing of speech using deep neural network, it is necessary to carry out the model training of relevant data. Because the number of layers of the network model is very deep and the structure of the network is extremely complex, many parameters need to be adjusted during the training, which means that the encoder model is needed to avoid over-fitting or local optimization. The ultimate goal of speech recognition simulation training is to grasp the eigenvalues of speech, so that after inputting the corresponding data, cycle training can be carried out to obtain better recognition effect. In addition to adding simulation training templates to the speech recognition system, it is also necessary to identify the words in the template library and put in a large number of articles with very high similarity. Through this training, the effectiveness of speech recognition can be effectively improved and primary mistakes can be avoided. Constantly enriching the methods and modes of speech recognition can improve the matching effect of speech recognition and greatly improve the performance of speech recognition. It can be seen that strengthening the simulation training of speech recognition is an effective method to improve the effect of speech recognition [6].

5.3 Combining acoustic features with sports features to identify

In recent years, artificial intelligence technology has made rapid progress and development, and human-computer interaction has become the focus of attention, eager to integrate some feelings when communicating. Therefore, emotion recognition of speech needs to be added to the speech recognition system. If you want to analyze emotional information in speech, it is difficult to extract information, which is a particularly complicated process. If you only use speech to identify people's emotions during speech, it has great limitations, so you also need to identify human facial expressions, data of pronunciation organs, and contact sports and voice learning to identify voice emotions. In this case,

there will be a lot of data to be collected, which will help to improve the accuracy of speech recognition by finding the features in these data and fusing them into the speech recognition system [7].

6. Development prospect of speech recognition method for deep learning of artificial intelligence

6.1 Widely used in people's work and life

At present, electronic products play an important role in people's work and life, which provides great convenience. Speech recognition has made many electronic products improve users' application satisfaction rate, while some users feel that there are problems that need to be improved. If the related research on speech recognition methods of artificial intelligence deep learning can be more widely applied to electronic products, it will significantly improve the breadth of deep network processing information and data in the future man-machine interface, and at the same time, it can significantly reduce the interference caused by noise to speech recognition.

6.2 Closer to 'human intelligence'

In the future, the speech recognition method of artificial intelligence deep learning will refer to the algorithm of deep neural network, and be closer to the process and mode of obtaining information, analyzing and processing information of human brain, and then create a more powerful engine, which will obviously improve the perception ability and cognitive ability. From the perspective of perception, it is possible to collect the perceptual information of vision, hearing and reading in the future, and expand the scope of perception. From the cognitive point of view, the judgment of language input will be more and more accurate, and scientific decisions can be made with the help of powerful logic network and reasoning network, and at the same time, effective output can be developed to create an advanced mode of human-computer interaction [8].

6.3 Promote the progress and development of the artificial intelligence industry chain

The speech recognition method of artificial intelligence deep learning will recognize speech signals more accurately in the future, and can analyze speech signals accurately to form decision-making output, thus optimizing the user experience to the greatest extent. Not only that, but also voice software will advance to a new level, which will help improve the accuracy of products and expand the scope of its application. Finally, the voice service forms of the voice recognition system will become more and more abundant and gradually develop into a mature industrial chain.

7. Conclusions

To sum up, at present, speech recognition technology based on artificial intelligence deep learning still faces some problems, such as noise interference and low endpoint detection level, which affects the quality of speech recognition. Therefore, we can make full reference to the above-mentioned speech recognition technology strategy of optimizing artificial intelligence deep learning, optimize the process of obtaining information, analyzing and processing information in speech recognition system, improve the matching effect of speech recognition, greatly improve the performance of speech recognition, and promote the sound development of artificial intelligence speech recognition.

References

[1] Zhang Guofeng. Analysis of Speech Recognition Methods for Deep Learning under Artificial Intelligence[J]. Electronics and Software Engineering,2020(11): 176-177.
[2] Zhang Liang. Analysis of Speech Recognition Methods for Deep Learning under Artificial Intelligence[J]. Computer Products and Circulation,2020(06):121.
[3] Liu Zubin. Discussion on Oral English Evaluation Path of Artificial Intelligence Speech Recognition[J]. Information Recording Materials,2019,20(11):92-95.

[4] Na Tian. Application and Challenge of Artificial Intelligence Technology in Language Learning[J]. Modern Information Technology,2019,3(19):109-110.
[5] Cui Juan, Wu Lei. Analysis of Speech Recognition Method Based on Artificial Intelligence Deep Learning[J]. Information Recording Materials,2019,20(09):168-169.
[6] Yuan Yifeng, Xue Xue, Kuang Fuhua. Application of Intelligent Building Management System Based on Artificial Intelligence[J]. Mechanical & Electrical Engineering Technology, 2019,48(05): 33-36.
[7] Bi Xinwen. Speech Recognition Method Based on Deep Learning[J]. Electronics and Software Engineering,2019(08):245.
[8] Huang Tianyun. Speech Recognition Method Based on Artificial Intelligence Deep Learning[J]. Information Recording Materials,2017,18(09):20-21.