

Application and development prospect of AI speech recognition technology

Tianyu Jiang

Guangxi University of Financial and Economics

Tianyu.jty@gmail.com

Abstract. In our daily life, artificial intelligence technology has penetrated all aspects, for us, language communication is a kind of media, but also an important way for us to convey ideas and give orders, so now, language communication has become an important way for us to interact with artificial intelligence. This paper through social research and the use of voice command to the existing speech recognition technology products test method puts forward the current development problems and solutions, the current situation of artificial intelligence speech recognition technology and the future in the field of communication systems and the intelligent home application made a simple introduction.

Keywords: Artificial Intelligence, Speech Technology, Development Status.

1. Introduction

In computer systems, speech recognition, also known as automatic speech recognition (ASR), is a process of translating human speech into a written format. The application of speech recognition technology is not only to translate into words but more importantly to understand the meaning of the words and to be able to respond accordingly. It is also a technology for converting speech signals into computer-readable text. It is a field of acoustics, phonetics, linguistics, digital signal processing theory, information theory, computer science and many other interdisciplinary disciplines [1]. At present, intelligent speech recognition systems can only carry out language recognition under limited conditions, and as a cutting-edge application of artificial intelligence, speech recognition technology will be widely used in daily life, and even in the scientific research of artificial intelligence neighborhoods in the future.

2. The Background of artificial intelligence

At present, artificial intelligence is changing our way of life. A Voice recognition function is added to the phone, and we can control the phone through simple statements, and our lives have become more intelligent. Artificial intelligence technology has opened up new ways and opportunities for the progress of human society. Artificial intelligence plays an important role in all fields, such as industry, energy, environment, economy, education, and health care, and has now become a major force driving it forward in all fields of society. Especially in recent years, in some specific fields (such as image recognition, speech recognition, and other fields), the problem-solving ability of intelligent systems even exceeds the ability of human beings [2]. As a common AI application, automatic speech

recognition (ASR) can convert spoken words into text and process them to get their meaning. Since humans often speak in spoken, abbreviated, and acronyms, a lot of computer analysis of natural language is required to produce accurate transcription. Speech recognition technology faces many challenges, but the scope is shrinking.

3. Development history and current situation of intelligent speech recognition technology

In the 1950s, Davis et al. at Bell LABS studied the world's first experimental system that could recognize ten English digit sounds. In 1960, Chinese culture shines a light on the success of the first computer speech recognition system.

From 1970 to 1980, the focus of speech recognition technology gradually shifted to large vocabulary, conjunctive speech recognition, and the development of various speech algorithms.

After entering the 1990s, a variety of speech recognition products have come out in succession, and the intelligent speech recognition industry has developed rapidly.[3]

Nowadays, speech recognition technology is further mature, and speech recognition system is also applied in all aspects of life.

IBM officially launched the Chinese dictation machine system VIa VOice in 1997, the system for news speech recognition has high accuracy, and is a more representative Chinese continuous speech recognition system.[4]

In 2011, Microsoft Research proposed an acoustic model based on a context-related deep neural network and hidden horse quantum end model, which achieved significant performance improvement in large vocabulary continuous speech recognition tasks, and since then, a large number of researchers began to turn to deep learning in the field of intelligent speech research.

In 2016, the accuracy of machine recognition successfully reached the human level for the first time, and the artificial intelligence speech recognition technology will be successfully applied to various machines.

4. Basic principles of intelligent speech recognition

Speech recognition technology has three basic principles: first, the language information in the speech signal is encoded according to the time change pattern of the short-time amplitude spectrum; second, speech is readable, that is, its acoustic signal can be expressed by dozens of differentiated and discrete symbols without considering the information content that the speaker tries to convey; third, speech interaction is a machine recognition process, so it cannot be separated from the grammar, the semantic and pragmatic structure of the language. The patterned framework adopted by intelligent speech recognition is divided into the following four aspects: data collection, feature extraction, model training, and judgment rules.

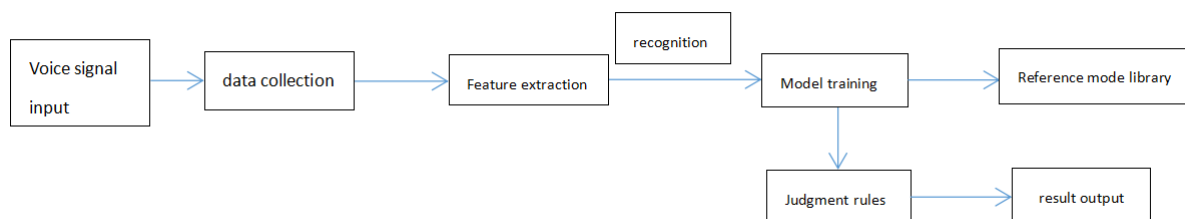


Figure 1. Basic principles of intelligent speech recognition.

In figure 1, the speech to be recognized is first input by the microphone, and then transformed into a specific speech signal. The data collection is carried out by the collection section. The collection process includes noise removal, environmental noise reduction, etc. The feature extraction part is used to identify and extract the main information of the speech signal. The common main information includes pitch frequency, short-time amplitude, etc. Model training is to allow the speaker to eliminate

redundant information by speaking several times to achieve the purpose of retaining the main information, and then form a reference pattern library according to certain rules for information comparison. The judgment rule is to compare the function calculation and other aspects of the model and to judge the semantic information of the input information according to certain rules.

5. The technical problems faced by intelligent speech technology

1) Environmental noise: at present, the phonetic system of recognition accuracy rate dependence of the high and low for the environment is strong, according to the test in a very quiet environment can be concluded that intelligent speech recognition accuracy is above 95%, and in a certain environment noise, the noise often affects the sound and tone of voice, etc. The author will make some discussion on the interference of environmental noise:

First, for a small amount of noise, we can use the line data or deliberately add noise training, which can be completely absorbed. Only under 20dB, the spectrum characteristics of the noisy sample and the pure sample are too different, and the convergence of the model learning is not good.

Second, we can consider reducing the overall volume of the audio file, so that we can better identify the main part of the audio file.

Third, as far as I know currently on the market the most intelligent speech recognition system is based on the traditional approach to identifying and noise reduction, I know we can go on the other hand, we can make the system take the initiative to learn audio and then by capturing the main body of audio noise, recognition and separation of audio noise to improve the accuracy of speech recognition.

2) Voice standard degree: at present, the intelligent speech recognition system in machine learning and deep learning for different accent and pronunciation recognition rate has improved, but now most people talk is not standard, there are a large number of local accents, and oral English, makes the voice system exist in the process of recognition errors. In terms of pronunciation, the pronunciation of accents is different. English is mainly on stress, while Chinese is on the flat tongue syllables; In terms of semantic recognition, the pause, rhythm and tone are different at different times; Finally, there is the generalization problem. After learning one or more dialects, the intelligent speech recognition program also needs to recognize other unseen dialects. There are two solutions to this problem:

First, learning representative dialects from different regions. After learning a representative dialect, the program can recognize and extract commands with similar pronunciation, intonation, and tone in the same region and surrounding areas. In addition, we can let the machine freely choose the library similar to the command speech, tone, and intonation for calling and recognition.

The second is to use the generalization model, only learn the Mandarin model, and then adjust the parameters based on the place where the speaker is. In addition, we can also specify the position of the speaker during training, add which dialect, analyze the differences between different dialects, and enrich the language library of the system.

3) Recognition speed, keyword detection, and screening: short sentences, modal words (ah, ah, etc.), and long responses to commands of medium and long sentences by ordinary people in the process of speaking. Therefore, we need to cover these factors more comprehensively through training data. The more comprehensive the coverage, the more ideal the speech recognition effect will be. When the training set data is limited, the data augmentation method can be used to expand the training set data to make the training set data diverse, so as to improve the accuracy of speech recognition. The acoustic waves collected by the system have several characteristics: pitch, loudness and quality. Only small changes should be made to the data enhancement. Too large changes will result in "extra data", so you can start with the following: increasing noise, increasing reverberation, changing pitch, and time stretching.

6. What people think of intelligent speech recognition technology

At present, with the rapid development of artificial intelligence and speech recognition, People's Daily life and work, and study have been greatly changed, through voice direct control of the machine, the

machine can understand our language and complete our instructions, so it can greatly improve our work efficiency.

6.1. Conduct a social survey of the population

The author conducted a social survey, the survey of the population more than 75 percent of the people have used intelligent voice assistant, but only about 20 percent of the people use it more than five times a week, it can be seen that intelligent speech recognition technology still has a lot of room for development. In the voice assistant in people often use more than fifty percent of users react less intelligent speech recognition available instruction, and a lot of machines can't be connected, different vendors of products are also hard to connect, led to the choice and buying different manufacturers and different kinds of excellent products, through the voice of different machines at the same time control will be more complex. In this regard, I think that for different machines with voice recognition, we can adopt a unified algorithm or system, just like different brands of mobile phones using the Android system can understand each other; Or use a master controller, which acts like a middleman to recognize and communicate commands, so that you can just install speech recognition on the master controller and control all the machines that are connected to it.

6.2. Speech recognition applications were tested using speech commands

In order to better reflect on the problems associated with using the voice assistant, I have tested some voice recognition applications by using some voice commands.

Table 1. Test instructions and results.

Instruction	Voice Assistant feedback	Whether to complete the instruction
Order a takeaway	Open the takeaway APP	no
Call at 12345	Start calling	yes
Open the calendar	Open the calendar APP	yes
Open the software	More detailed commands are required	no
Navigation to the nearby park	Multiple nearby park addresses are shown	yes
Set an alarm clock for 9 am tomorrow morning	Set the alarm	yes
Broadcast the weather forecast	Intelligence voice broadcast	yes

As shown in table 1, when the author uses some simple commands and detailed commands, such as reporting today's weather and setting a 9 o'clock alarm clock every morning. However, when I use some more complex commands and not detailed commands, the voice recognition assistant cannot help me complete them.

Therefore, through the survey and test results show that the current audience of intelligent voice recognition assistants is very wide, but, the use frequency of people is not high. Moreover, only some simple and detailed commands can be completed. Therefore, nowadays, intelligent speech recognition technology has been fully integrated into people's Daily life, so, as interdisciplinary voice recognition technology is gradually becoming an important means of human interaction. The ability of speech recognition and the way to execute different adult commands still need to be matched and learned from large amounts of data.

7. The future of intelligent speech recognition technology

At present, influenced by smartphones and other value-added Internet services, AI voice recognition technology has been widely applied in many aspects such as medical treatment, military, office management, and daily life.

For example, in the telephone and communication system, the intelligent voice interface is changing the telephone from a simple service tool into a service "service provider" and life "partner". By using telephone and communication network, people can easily query and extract related information from the remote database system through voice command; smart home has driven artificial intelligence into every bit of life, people can control the machine's automatic cleaning through voice or can change the air conditioning temperature through voice, the smart home began to walk into people's lives; in autonomous driving, cars will determine and change the route and destination by identifying people's voice. It can be seen that intelligent speech recognition technology has affected people's clothing, food, housing, transportation, and other aspects.

8. Conclusion

This paper mainly introduces the application and future development of intelligent speech recognition assistants, and investigates and studies some existing problems of this technology, but this paper does not give a detailed overview of deep learning and database optimization.

Nowadays, the leading enterprises in various fields are more committed to the development of intelligent voice technology, such as Google, Microsoft, Apple, IBM, Facebook, Intel, China's BAT and other artificial intelligence voice recognition technology as the next technology tipping point, have invested huge amounts of investment in research and development and competition [5]. A variety of technology products such as Baidu assistant, Google Assistant, Apple SIRY and others products have been published. In the future, both neural networks and deep learning have made great progress in artificial intelligence, especially the application of deep learning algorithms, which greatly improve the accuracy of speech recognition and provide other possibilities for language awareness [6].

References

- [1] Dai Wei; Liu Hong; Chinese Speech Recognition Technology Based on neural Network, Journal of Sichuan Normal University (Natural Science Edition), 2022, 45 (01), 131-135.
- [2] application in China in the Era of intelligent education: Interpretation and enlightenment of the American report of "Planning the Future, Welcome the Era of Artificial Intelligence", Audio-chemical Education Research, 2017, 38 (3), 123-128.
- [3] Chen X M. Huang nan shan; Yang Can; Research Progress of Speech Recognition Technology, Modern Computer (Professional Edition), 2008, (09), 43-45.
- [4] Gao Xintao; Lovely lili Chen; Development Status and Application Prospect of Speech Recognition Technology, Gansu Science & Technology Review, 2007, (04), 13-13+76+3.
- [5] Qian Weijue; Internet "Independent" Innovation is getting Better, Nanfang Daily, 2015, (05).
- [6] Xu Xin; The Future Way of Speech Recognition, The Communication World, 2016, (16), 59.