Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 2, No. 2, 2022

Research on Interactive System of Movie Subtitle Speech Based on Machine Learning Technology

Jiahang Li

Shandong Rizhao No. 1 Middle School, Rizhao, China

Abstract: The composition elements of subtitles, from the early single text, have developed into the present text, graphics, colors, animation, special effects and other combinations. With the development of speech technology and natural language understanding, speech interaction system has become a hot research field. Different from the traditional data interaction between keyboard, mouse and display, using hearing to transmit data makes the interactive system of movie subtitles more anthropomorphic and intelligent. It is the most natural and convenient means for human beings to exchange information with intelligent systems by incorporating machines and equipment with voice information processing capabilities into human voice interactive objects and endowing movie subtitle interactive systems with biological language recognition functions. A machine learning-based movie subtitle voice interactive system is constructed, which can well expand the application of voice interactive system and improve the user experience. In this paper, a movie subtitle voice interactive system based on machine learning technology is proposed, so as to better and effectively realize human-computer voice interaction.

Keywords: Machine learning; Movie subtitles; Voice interaction system.

1. Introduction

With the development of Internet technology, film and television works have gradually penetrated into people's lives, and people can watch film and television works anytime, anywhere through terminals such as computers, televisions and smart mobile devices [1]. With the development of speech technology and natural language understanding, speech interaction system has become a hot research field [2]. In the process of interaction, people usually transmit their commands to the computer through input devices such as keyboard and mouse. For inexperienced people, human-computer interaction becomes an obstacle, while the voice-based interaction system becomes a bridge between people and computers, and people can communicate with computers freely and conveniently [3]. Language is the most natural and portable way for human beings to communicate their thoughts, opinions and emotions. Cloud computing bears the long-standing dream of mankind to use computing power as infrastructure, and may change most information technology industries [4]. For example, we can treat software as a service to make IT more attractive, or change the way IT hardware is designed and purchased. It is the most natural and convenient means for human beings to exchange information with intelligent systems by incorporating machines and equipment with speech information processing capabilities into human speech interactive objects and endowing movie subtitle interactive systems with biological language recognition functions [5].

The composition elements of subtitles, from the early single text, have developed into the present text, graphics, colors, animation, special effects and other combinations. Auditory system is an important part of movie subtitle speech recognition system, and its purpose is to better complete the information interaction between people and movie subtitle interactive system [6]. Different from the traditional data interaction between keyboard, mouse and display, using hearing to transmit data makes the interactive system of movie subtitles more anthropomorphic and intelligent. The realization of traditional intelligent interactive technology is often realized on the platform of voice interactive system itself, such as simple speech recognition algorithm, video capture and basic processing, etc., and it is difficult to realize more complex algorithms [7]. Because of their high requirements for the operation speed of movie subtitle speech recognition system, and the problems of large-capacity data storage of pattern recognition system also limit the further development of offline speech interactive system [8]. As the leading direction of intelligent computer research and the key technology of man-machine language communication, speech processing technology has received extensive attention. In this paper, a movie subtitle voice interactive system based on machine learning technology is proposed, so as to better and effectively realize human-computer voice interaction.

2. Interactive system of movie subtitle speech based on machine learning

Voice interaction with the human movie subtitle speech recognition system, first of all, the movie subtitle speech recognition system needs to collect the interactive voice through the microphone, specifically through PyAudio component, which provides Python language version and is a cross-platform audio I/O library. Using PyAudio, you can play and record audio in Python programs. Dialogue management is at the core of the interactive system, which integrates speaker recognition and keyword recognition technologies to form a personified intelligent interactive system [9]. Compared with other single-chip microcomputers, it has a very complete operating system, and it carries its own interface, so it can use the corresponding programming to realize the effective application of various software. By networking, intelligent voice interaction with open cloud recognition technology and simple switch can be realized, and various software and hardware of voice interaction can be effectively controlled, and the combination of online and offline can also be effectively realized. The microphone

collects and generates the voice wave file, and transmits the wave voice file to the voice recognition server through the Internet. The recognized text will be sent to the movie subtitle speech recognition system through the Internet, and the movie subtitle speech interaction system will look up the best reply information in the knowledge base through the incoming text content and context, and send it back to the movie subtitle speech recognition system terminal through the Internet. The overall framework of the voice interaction system is shown in Figure 1.

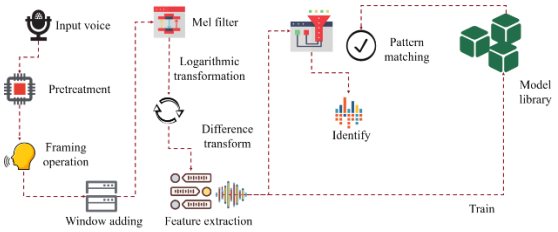

Figure 1. Framework and process of speech recognition system

Errors in keyword recognition and speaker recognition can be corrected by certain strategies in dialogue management to improve the success rate of system interaction, so how to build a dialogue management module is the key to successfully realize the system. At present, most voice interaction systems only use front-end voice keyword or whole sentence recognition and back-end dialogue management, and dialogue management only receives single input information, which leads to single feedback of voice system and is not humanized enough. This system integrates speaker recognition and voice location into the system, which can receive more detailed user information. Using the effective connection of multiple interfaces such as Internet speech recognition, offline speech recognition and cloud open speech recognition, we can effectively collect external speech information and some simple action information, and transmit them to the main board for certain processing [10]. After the processed information is output, a simple switch and speaker are used to give some feedback and interaction to the information, and then the switch equipment and networking connection are used again to realize the effective control of the wireless switch. Through speaker recognition and judgment, the personality information database is introduced, and different users are fed back according to their different preferences, so that the speech system has a certain memory function. The subtitle speech recognition system plays the reply speech file through the audio output interface to complete the final speech data output.

3. Speech recognition process and parameter estimation

With the continuous improvement of speech processing technology, the processing algorithm becomes more and more complex, and the real-time system requires higher computing speed and storage capacity. In the specific operation process, relevant software systems can be used to effectively identify keyword information. In the working process, the module usually sends out the corresponding bytes in time as long as it captures some voice information. The processing process of speech recognition system is that the user inputs a speech signal through a speech input device at first. The design of the system architecture needs to first define the requirements, select appropriate technologies for planning, and use complete tools to realize a complete system. The voice data is transmitted to the cloud server through the network, and the recognition results are obtained and returned by using cloud computing technology. Because of the abundant resources in the cloud, the knowledge base resources can be better utilized to complete the interaction between people and the voice system, and online video resource retrieval. The overall process of the speech recognition model is shown in Figure 2.

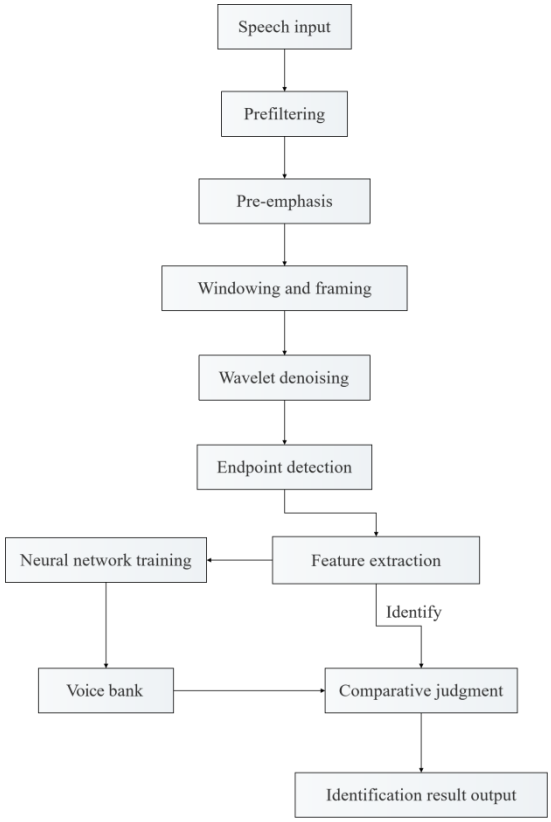

Figure 2. The overall framework and process of voice interaction system

Voice interaction and touch interaction are similar in system architecture. The terminal of voice interaction is a smart speaker, and the terminal of touch interaction is a mobile phone. Apart from the interaction, smart speakers and mobile phones can be regarded as the terminal devices of smart video users. After the text is effectively identified, it can be sent to the movie subtitle speech recognition system by using the Internet, and the movie subtitle speech recognition system can fully identify the content of the text by combining its context, and then find the most matching data in the whole database. Intelligent video system is an intelligent system that integrates many technologies, including not only communication transmission protocol, terminal wireless networking technology and equipment automation control technology, but also terminal control software for managing equipment. After the speech signal is input, it is necessary to process the speech signal, and the processing process is speech enhancement, which generally exists in the speech processing system as a preprocessing or front-end processing module. The parameter estimation process of speech recognition model is shown in Figure 3.

The inter-network communication between subtitle network and broadcast network is realized by accessing IP in the two LANs with their respective switches as the main communication equipment. When the subtitle editing workstation or subtitle broadcasting workstation needs to obtain the information of the on-air program list of the

broadcasting network, it sends out an access application. After receiving it, the subtitle database confirms it, and then communicates with the broadcast network switch through the subtitle network switch. The speech recognition system is implemented from bottom to top. First, the functions of each module are realized separately, and then the functions of each module are integrated together for debugging. After receiving the problem of the movie subtitle speech recognition system, you can search the text data or speech data in the cloud speech system in time, and then send it to the movie subtitle speech recognition system in speech format, and the movie subtitle speech recognition system can play these subtitles.

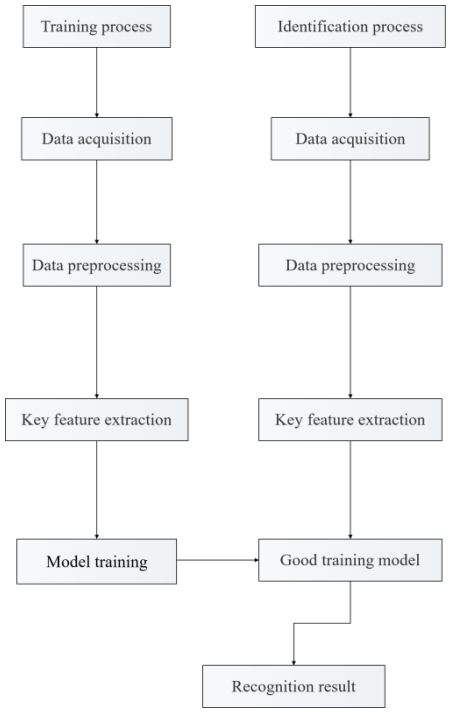

Figure 3. Parameter estimation process of speech recognition model

4. Conclusions

With the development of Internet technology, film and television works have gradually penetrated into people's lives, and people can watch film and television works anytime and anywhere through terminals such as computers, televisions and smart mobile devices. In fact, speech recognition is an advanced technology in which the speech system uses its own hardware or software system to effectively process the received information and then form the corresponding speech or file. The realization of traditional intelligent interactive technology is often realized on the platform of voice interactive system itself, such as simple speech recognition algorithm, video capture and basic processing, etc. It is difficult to realize more complex algorithms. In this paper, a movie subtitle voice interactive system based on machine learning technology is proposed, so as to better and effectively realize human-computer voice interaction. The centralized management of film and television materials broadcast by subtitles and the realization of multi-task and multi-window network operation have greatly improved the efficiency of broadcast workflow. After the text is effectively identified, it can be sent to the movie subtitle speech recognition system by using the Internet, and the movie subtitle speech recognition system can fully identify the content of the text by combining its context, and then find the most matching data in the whole database. This technology can effectively promote the natural and friendly relationship between human-computer interaction, and is the main development direction of human-computer interaction in the future.

References

[1] Pan Huilin, Han Zhiyan, Wang Shurui, et al. Emotion recognition method based on two-channel speech to image [J]. Electronic Design Engineering, 2021, 29(15):5.

[2] Qi Baoliang, Wang Xu, Lin Yupu, et al. Design of a parking guidance system with voice interaction function [J]. Computer Measurement and Control, 2019, 27(6):5.

[3] Hou Yong, Wang Zheng, Shu Qiaoye, et al. A voice assistant system to prevent voice errors of power grid dispatchers [J]. Microcomputer Application, 2019, 35(12):4.

[4] Yu Lei, Li Taotao. Design of intelligent speech control system based on ROS [J]. Electronic Measurement Technology, 2019, 42(23):5.

[5] Wang Tao, Wang Jiaxin, Zhang Xinke. Intelligent guide companion based on voice interaction and positioning system [J]. Application of Single Chip Microcomputer and Embedded System, 2019, 19(11):4.

[6] Li Jian, He Hanwu, Wu Yueming, et al. Voice Interactive AR Operation Guidance System for Electronic Taxation [J]. Modern Electronic Technology, 2021, 44(22):5.

[7] Zhu Lili, Wang Jianxiang, Miao Peixian, et al. Design and implementation of intelligent voice sand table control system based on STM32 [J]. Electronic Design Engineering, 2019, 27(14):5.

[8] Wenquan, Xia Xiaochun, Li Yao. System implementation method of airborne speech recognition technology [J]. Measurement and Control Technology, 2018, 37(09):3.

[9] Xu Xiufu, Lu Xiaonan. Design of smart home voice control system for Android phone [J]. Application of Single Chip Microcomputer and Embedded System, 2018, 18(1):5.

[10] Niu Heng, Chen Chile, Zhang Wenjing, et al. Multifunctional blind guide system based on image perception and location matching [J]. Automation and Instrument, 2021, 36(10):5.