# System of Smart Home Based on Speech Recognition with Machine Learning

Baihua Li [1, *]

[1] Academy of Artificial Intelligence, Hefei Economic College, Hefei , Anhui, China.

* 276368350@qq.com

**Abstract.** This article designs and implements a system of smart home based on speech recognition with machine learning and Android. The STM32 microcontroller is combined with actuating components to control home devices by obtaining server control commands through the network. This article designs a smart home APP based on Android and Google Voice Search, optimizes the speech recognition algorithm by using the training results from Word2Vector model to improve the recall rate of the speech recognition of command samples. This APP displays environmental data according to the scenario for users, such as the temperature, humidity, and alert of smoke gas. The test results show that the recognition recall rate of the test data of control commands from users of different ages is increased to 94.4% with high stability.

**Keywords:** Speech recognition;word2 vector;Embedded technology; Android Studio.

## 1. Introduction

Users have more and more requirements for control methods of home equipment, and intelligence and networking are the direction of development of home control technology. The home control system designed in this paper uses speech recognition [1] technology to control home devices ,uses machine learning to optimize the recognition algorithm. In view of the inconvenience of manual operation of hardware equipment, Liu Mi proposed an intelligent voice control system based on ATM32[2], which controls home device through the combination of first-level trigger instruction and second-level instruction. However, there is no WIFI network to solve the user's remote control needs. In view of the networking of smart home devices with Zigbee protocol, Xie Yonghong et al. realized the graphical interface control of embedded microcontroller through the design of a voice-controlled home appliance system [3] based on ARM technology. However, the hardware structure is complex. Meanwhile traditional string matching is used in the speech recognition algorithm, this algorithm recognizes the voice commands of the elderly and young children with low recall and the error increases. This article designs and implements smart home hardware, uses android operating system with integrated Google Voice Search[4], and designs a combination of command library and machine model training to realize remote control of home devices, and at the same time, improves the recall rate of recognition of voice command.

## 2. System Hardware Components

### 2.1 System Architecture

The hardware devices in this system include an embedded microcontroller, various types of sensors, a WIFI communication module, motor drive module [5], these devices collect data, through the WIFI network, using the MQTT protocol to upload the data to the cloud platform, and accepts the control commands from the APP and the cloud platform. The ESP8266 WIFI module communicate with the microcontroller through a serial port, and communicate with the cloud server with MQTT protocol. The client APP connects to the WIFI network using the UDP protocol to control the hardware devices in the indoor environment.The system framework is shown in the figure below.

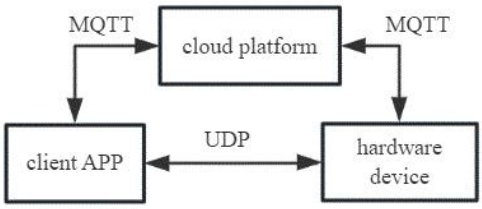


Fig. 1 System framework

The software of APP designs a similarity algorithm[6] to match the control command. The commands are vectorized using word2 vector to represent the commands. The algorithm is designed to solve the similarity between the input speech commands and the model training results, and the similarity threshold is set according to the model training results to improve the system's recall for positive sample recognition of user commands.

## 2.2 Central Processor

The microcontroller in this system is the minimum system of embedded 32-machine, The core of the module is STM32F103RC8T6. It is a 32-bit ARM microcontroller based on the Cortex-M3 core.

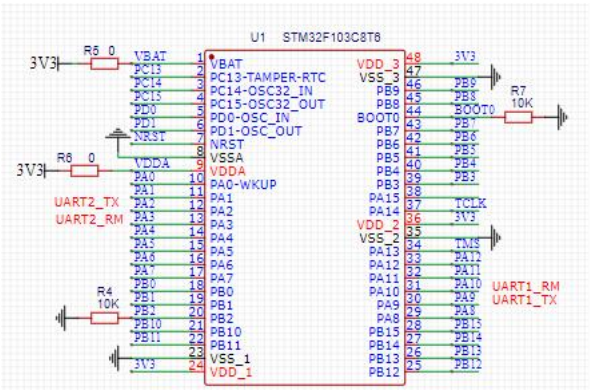


Fig. 2 Microcontroller

## 2.3 WIFI Module

The serial WIFI module with ESP8266 as the chip integrates multiple protocols for wireless communication, which can work in WLAN networks and WiFi networks [7], and can realize the connection and data transmission to a variety of different devices. The data transfer with the microcontroller is done through the serial port. The circuit diagram is shown in Figure 3.

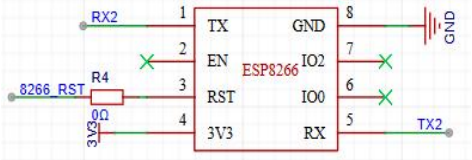


Fig. 3 WIFI module

## 2.4 Sensors & Actuators

Temperature and humidity sensor: DHT11 is connected to the GPIO port, and the calibrated data from the sensor is sent to the microcontroller.

Smoke Sensor: The MQ-2 smoke sensor measures the concentration of combustible gas in the air. The buzzer alarms when there is a change in the sensor output signal.

Actuators: The fan is controlled by a stepper motor. curtains are controlled by delay, and two relays are used to simulate the state of the curtains.

# 3. System Software Design

The software of the system includes three parts: the hardware programming control, communication between the APP client and hardware, and the app client.

## 3.1 Hardware Programming Control

The hardware programming control mainly realizes sensor data acquisition, control of home equipment and communication between the sensors and the cloud platform. The control system collects sensor data, sets the smoke concentration alarm threshold, and alarms when it exceeds the threshold; The WIFI module accesses the network by command, and the hardware accesses the cloud platform through the WIFI network.

## 3.2 Communication Between the Hardware and the APP Client

The hardware programming control system connects to the home internet environment via the ESP8266. The communication between the hardware and the APP client is carried out according to the UDP protocol.

In remote control, the system uses the MQTT protocol to communicate with the cloud server, and uses the protocol to upload the underlying hardware device data. SQL database technology is used in the cloud server to store and manage equipment data. The MQTT client is integrated into the APP control system, which initiates a request to the server to obtain data and issues the user's control command. The MCU parses the control command according to the protocol and executes the corresponding command. The communication flow is shown in Fig.4.

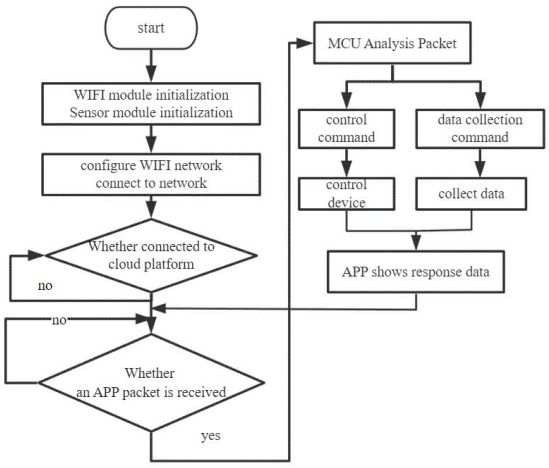


Fig. 4 Communications flow chart

## 3.3 The APP Client

The APP control system was designed and implemented in Android Studio, completing the client APP interface design, network communication, and home device control [8].

The system setting completes the system alarm and system connection functions. Speech recognition uses Google Voice Search to compare the user's voice control commands with the command library designed by the system, and optimizes the algorithm by using the model training results to obtain more accurate recognition results. Room management can control the fans, lights, and other devices in different rooms. Environmental monitoring is to summarize the temperature, humidity, smoke concentration, and other data information in the home environment. The database is an important foundation of APP control system to realize the storage and management of user data and equipment data.

This paper uses the Ali cloud server as a data relay station. Through the construction of the cloud server, a reliable long connection between the server and the client is realized to ensure that the client can obtain timely data messages. The abnormal data of the sensor is sent to the APP using the

message push service in the cloud service, and the user can query the abnormal information of the device in the system Settings. This way can improve the efficiency of message transmission and save the power consumption[9] of cell phones. The APP interface is shown in Fig.5.

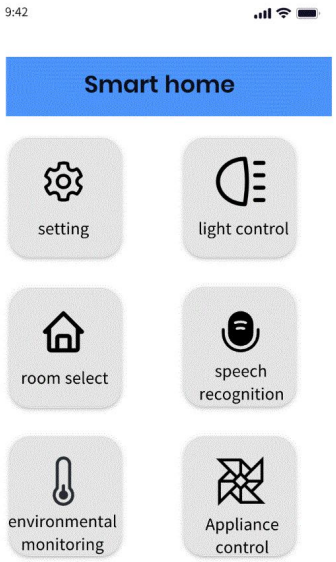


Fig. 5 APP UI

### 3.4 Algorithm Optimization For APP Speech Recognition

The design of APP speech recognition is realized with the Google Voice Search. The recognition process is: firstly, obtain the user's real intention through software recognition; secondly, control the device according to the recognition results. Because there is some error in Google software identification, the specific command library designed by the system matches the software identification results. The syntax rules of a command library are:

Operation + Modifier property + object + property 1.

Design the command library according to syntax rules. The operation library describes the "open" and "close" actions of the voice signal; The modification property corresponds to the room, and the object is the home device; Attribute 1 is the mode parameter of the home device.

When matching control commands of users with command libraries, the most critical algorithm is the string similarity algorithm [10]. The traditional algorithm divides the similarity calculation process into three operation types: modify, add, and delete characters. Using dynamic programming method to solve the dynamic equation to calculate the distance, and then calculate the similarity of the two strings. The detection results will match the command only when they are all similar, and the accuracy rate of command recognition and recall rate are low.

The machine learning and WMDare used to optimize the speech recognition algorithm. The word2 vector model dicovered by Mikolov et al. [11] trains the user’s voice command, and the string similarity algorithm Word Mover's Distance (WMD)[12] calculates the distance between two documents (or a group of words). Word2Vec is a word embedding model based on a neural network, and its word vector function can capture the semantic knowledge of language well. WMD can calculate the minimum "effort" required to move words from the first statement to the second statement. The evaluation indicators are accuracy, recall, and F1. Accuracy reflects the percentage of positive sample predictions that are correct; The recall rate reflects how many samples are correctly found; F1 measures both accuracy and recall, and usually the bigger the F1, the better.

Combining the word2vec and WMD into algorithm design requires three steps: Firstly, preprocess the text and train a Word2Vec model that maps words to vectors. Secondly, the voice signals are converted into words that are sent into the word2vec model to obtain word vectors, and then the WMD algorithm is used to calculate the distance between these word vectors and the restored target word vectors to carry out the semantic search. The system sets the threshold

according to the F1 index of model training results. When the distance is less than the threshold, the user's voice command can be considered to match the target command. The algorithm flow is shown in the figure below.

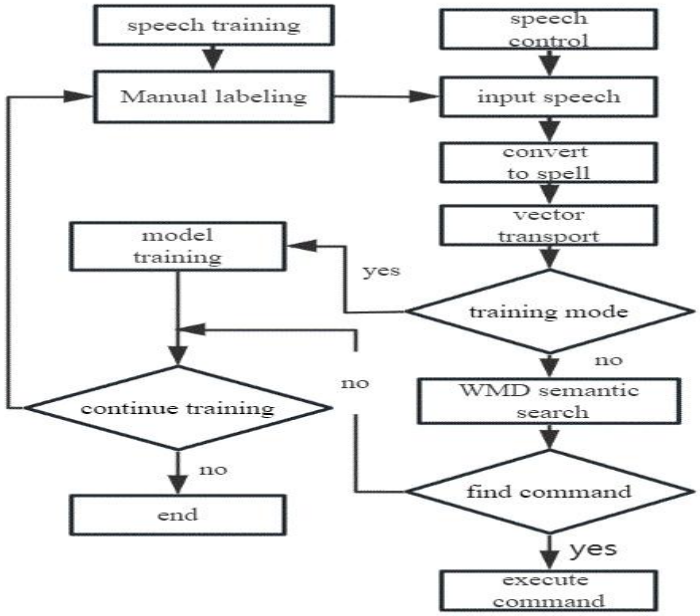


Fig. 6 Algorithm flow

For commands that are higher than the threshold value, you can add such commands to the model training set. When the user inputs this command again in the future, the WMD algorithm can match the relevant commands from the training model.

## 4. System Test

The following figure shows the test results before the introduction of the machine learning method. There are 100 test samples, including 50 positive samples (voice input has matching commands) and 50 negative samples (voice input has no matching commands). Arrange 5 users(different genders, different ages) to pronounce according to the test sample and record the test results.

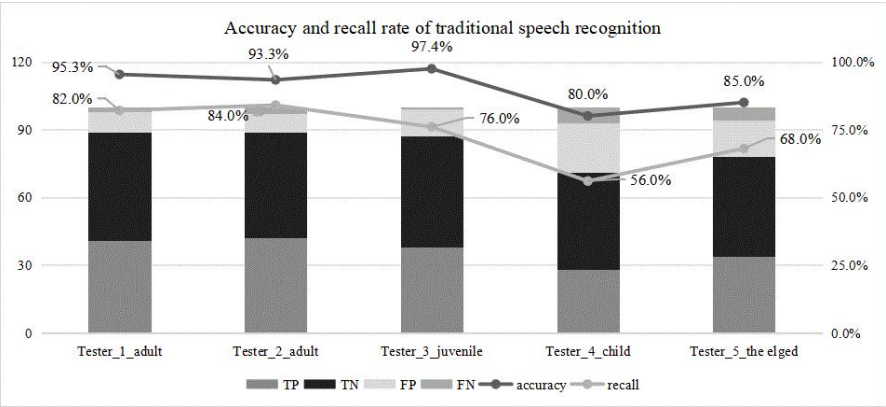


Fig. 7 Recognition of traditional string similarity algorithm

TP (True Positive)-predicts the positive class to be positive. TN-The negative class is predicted to be negative.

It can be seen from the test results that the traditional voice recognition method has only a 73.2% recall rate and 90.6% accuracy. This means that nearly 27% of correct voice commands are not correctly recognized. It can also be intuitively seen from the figure that the recall rate of children

and the elderly will fall to around 60% due to non-standard pronunciation, and the use experience of different users will have huge differences.

With the introduction of machine learning, the accuracy and recall rates are improved to the following levels.

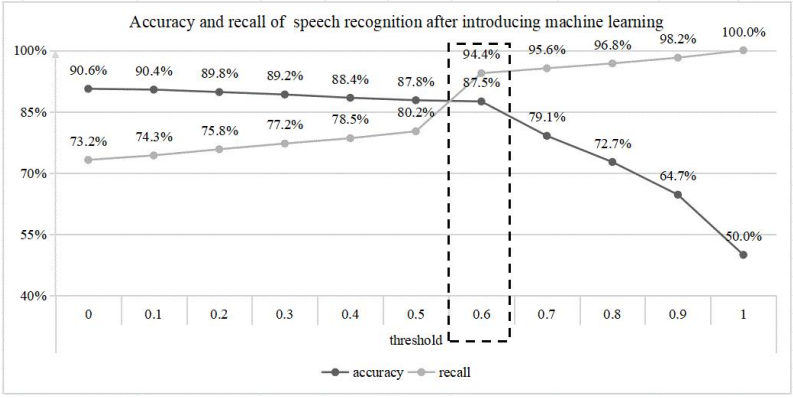


Fig. 8 Accuracy and recall with machine learning

After introducing the machine learning, based on the same test sample, the optimal index F1 is obtained when the threshold is set at 0.6; The recall rate increased by 21.2% with a 3.1% loss in accuracy. The results are compared in the following Table.

Table 1. Two Algorithms Comparing

| Algorithm | Accuracy | Recall |
|---|---|---|
| traditional algorithms | 90.6% | 73.2% |
| Word2Vec + WMD | 87.5% | 94.4% |
| result comparison | -3.1% | 21.2% |

## 5. Conclusion

This article designed and implemented a smart home system which can control home equipment indoors and remotely. It designed and implemented the APP control software based on the cloud platform and Android Studio, optimized the speech recognition algorithm by using a machine learning model, and improved the recall rate of the voice recognition at the APP control terminal. The test results showed that the recall rate increased by 21.2% when the accuracy of recognition of voice control commands were lost by 3.1%. The probability that positive samples of voice control commands of users of different ages that were recognized increased to 94.4%, and the user experience was significantly improved. The designed system used the functions of data storage management and device control management of the cloud platform to build a server, enhanced the stability, and realized the two-way remote communication with the hardware devices and the app client.

## Acknowledgments

Provincial Key Project of Scientific Research in Colleges and Universities of Anhui Province (N atural Science), project number :2022AH052622.

## References

[1] Sheng Zhentao, Yan Kengyang. Design and production of Smart Home Assistant based on Speech recognition [J]. Digital World,2019(8):134-135.

[2] LIU Mi. Design of Intelligent Voice Control System Based on STM32 [J]. Industrial Instrumentation and Automation Equipment.2022(04):14-18.

[3] Xie Yonghong, Li Wei. Design of voice-controlled Home Appliance System Based on ARM [J]. Electronic Design Process, 2019,30(13):175-179.

[4] Zhang Y , Park D S , Han W ,et al.BigSSL: Exploring the Frontier of Large-Scale Semi-Supervised Learning for Automatic Speech Recognition[J]. 2021.DOI:10.48550/arXiv.2109.13226.

[5] Zhou Saiqing, Song Wenlu, Yang Cong. Hardware Design of Intelligent Home Audio Control System based on STM32 MCU [J]. Journal of Xi 'an Siyuan University,2016(6): 239-242.

[6] Kong Sheng, Wang Yu. Research on topic sentence extraction algorithm based on sentence similarity [J], Journal of Information Technology, 2019, 30(6): 605-609..

[7] Li Yunhong, Zhang Heng, Zhang Yalin, et al. WiFi terminal system design based on ARM [J]. Computer and Digital Engineering,2017,45(3):498-543.

[8] Deng Xin. Design of Control Terminal Based on Speech Recognition [D]. Xi 'an: Xi 'an Polytechnic University,2019.

[9] Chen Xixiang, Huang Wu, Li Deying. Design of Smart Home Control System based on Speech Recognition [J]. Automation & Instrumentation, 2021,36(7):91-95.

[10] Wang Weiming, Liang Dongying. Chinese sentence similarity calculation based on Semantic dependency matching [J]. Journal of Shenzhen Polytechnic of Information Technology, 2014 (1): 56-61.

[11] Mikolovt, Sutskever I, Chen K,etal. Distributed Representations of Words and Phrases andtheir Compositionality [J]. Andances in Neural Information Processing System, 2013,26:3111-3119.

[12] Kusner M, Sun Y, Kolkin N, Weinberger K. From word embeddings to document distances. In International conference on machine learning 2015 Jun 1 (pp. 957-966).