An Efficient Speech Recognition Algorithm for Small Intelligent Electronic Devices

Zhichao Zheng[1], Xiaotao Lin[1], Weiwei Zhang[1], Jianqing Zhu[1*] and Huanqiang Zeng[2]
[1]College of Engineering, Huaqiao University, 269 Chenghua North Road, Quanzhou, Fujian, China
Email:{1695111058, jqzhu}@hqu.edu.cn
[2]College of Information Science and Engineering, Huaqiao University, 668 Jiemei Avenue, Xiamen, Fujian, China
Email: zeng0043@hqu.edu.cn

***Abstract*—The speech recognition technology makes it possible for people to communicate with intelligent electronic devices. However,existing speech recognition algorithms are overly complex for small intelligent electronic devices (e.g., mini speakers, intelligent toys, intelligent remote controls, etc.). For this, an efficient speech recognition algorithm is proposed. Firstly, the *Mel-scale Frequency Cepstral Coefficients* (MFCC) is applied to extract features of voices. Secondly, the *Support Vector Machines* (SVM) is used to train speech classification models. Finally, a speech database is collected to validate the proposed algorithm. The speech database contains 500 audio files of 10 speech commands for an electric motor car driving assistant and 550 audio files of 11 speech commands for a intelligent remote control. The proposed method is evaluated via a 5-fold cross-validation, and experiments show that the propose method acquires 94.20% and 88.73% average accuracy rates for the electric motor car driving assistant and the intelligent remote control, respectively.**

***Index Terms*—MFCC, SVM, Speech Recognition**

I. Introduction

The speech recognition technology allows people to communicate friendly with intelligent electronic devices. But, general speech recognition algorithms have a large computation load that usually completed on a remote server, which are not suitable for small intelligent electronic devices (e.g., mini speakers, intelligent toys and intelligent remote controls, etc.). Because these small intelligent electronic devices are cost controlled so that they can not accept a large computation load and additional communication chips. Therefore, it is urgent to design an efficient speech recognition algorithm for small intelligent electronic devices.

Erlin et al. [1] analyzed discriminative features of voices in detail, including loudness and tonal harmony, and designed the audio classifier according to the *nearest neighbor* (NN) criterion. Guo and Li [2] designed a SVM based multi-stage audio classifier. Lu et al. [3] proposed a hidden Markov model based audio classification method. Based on the work of Guo and Li [2], Lin used the wavelet transform to extract sub-band energy and pitch frequency [4] as features of voices.

In this paper, an efficient speech recognition algorithm is proposed. It uses *Mel-scale Frequency Cepstral Coefficients* (MFCC) for representing voices and applies *Support Vector Machines* (SVM) to classify speeches. Moreover, a speech database containing 500 audio files of 10 speech commands for an electric motor car driving assistant and 550 audio files of 11 speech commands for a intelligent remote control is collected to validate the proposed algorithm via a 5-fold cross-validation. The experiments show that the proposed method acquires 94.20% and 88.73% average accuracy rates for the electric motor car driving assistant and the intelligent remote control, respectively.

* Jianqing Zhu is the corresponding author.

Fig. 1. The flow chart of the proposed method.

II. Method

As shown in Fig. 1, the proposed approach consists of two steps(i.e., MFCC and SVM), which are introduced as follows.

A. MFCC

Due to MFCC [5] having an anti-noise ability, it is applied to extract features of speeches in this paper. Due to the paper length limitation, we can not detailly introduce MFCC. The details of MFCC can be found in [5]. However, one important step of MFCC should be noted. Because of the differences in speed and content of speech commands, the data lengths of speech commands are different, which is harmful to the following SVM. For this, during the MFCC step, frame energy is descending sorted, and only top-128 frames are selected to construct features of speeches, making each speech holds the same dimensional feature vector.

B. SVM

In practice, the interaction between human and small intelligent electronic devices does not require complex speech commands. The kinds of speech commands are limited to a specific electronic product. Therefore, we transform the speech recognition task as a classification task that aims to predict the class label of an input speech file. Considering that the SVM [6] has high efficiency and accuracy, it is applied to finish the speech classification task for small intelligent electronic devices in this paper.

978-1-7281-3038-5/19/$31.00 ©2019 IEEE

TABLE I
THE ENGLISH TRANSLATIONS PF SPEECH COMMANDS IN THE SPEECH DATABASE.

electric motor car driving assistant (Part I)		
1.the intelligent assistant	2.intelligent photography	3.start recoding video
4.stop recoding video	5.start the recoding	6.stop the recoding
7.electronic rearview mirror	8.lock video	9.return to the video
10.yadi yadi		
intelligent remote control (Part II)		
1.open the lock	2.closed lock	3.open the headlight
4.close the headlight	5.hello, xiao bao	6.please bright red
7.please bright yellow	8.please bright green	9.please bright blue
10.please bright purple	11.please run lantern mode	

TABLE II
ACCURACY RATE (%) OF THE PROPOSED METHOD.

dataset	fold-1	fold-2	fold-3	fold-4	fold-5	average
Part I	97.00	97.00	98.00	89.00	90.00	94.20
Part II	90.90	90.00	94.55	85.45	82.73	88.73

III. EXPERIMENT

A. Database

In this paper, a new speech database is collected to validate the proposed method. This speech database includes two part, the Part I is for an electric motor car driving assistant and Part II is for an intelligent remote control. As shown in Table I, both Part I and Part II consists of simple speech commands, and the language is the Chinese mandarin. 50 volunteers (i.e., 25 men and 25 women) from different provinces of China contribute their speeches to construct this database. As a result, the Part I contains 500 audio files of 10 speech commands, and the Prat II includes 550 audio files of 11 speech commands.

B. Implementation Detail

The hardware is a notebook with an i5-6300HQ CPU and 8.00 GB memory. The software is MATLAB 2016b. Regarding to the MFCC step, the cepstrum coefficient is 12; alpha value is 0.97; the frequency range is 100-800 Hz. Regarding to the SVM step, the libsvm [6] toolbox is applied. The linear kernel is used and the penalty factor C is set as 1.5. The 5-fold cross-validation is implemented to evaluate the proposed method.

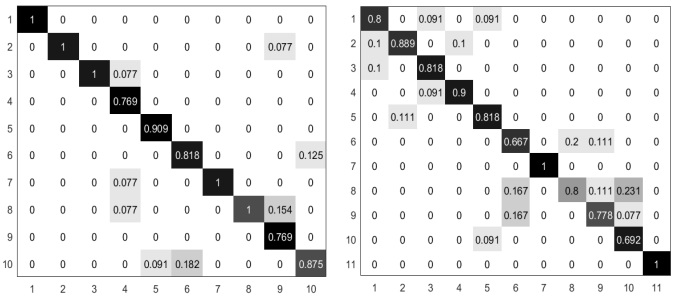
C. Experimental Result

As shown in Table II, on the Part I (i.e., electric motor car driving assistant), the average accuracy rate of the proposed method is 94.20%, and on the Part II (i.e., intelligent remote control), the average accuracy rate of the proposed method is 88.73%. Moreover, Fig. 2 shows the confusion matrixs of Part I and Part II. It can find that only a small number of samples are not on the diagonal, which shows that most of samples are classified correctly.

In addition to the accuracy rate, for processing one sample, the average time cost of the MFCC step is 12.80 ms and the SVM step is 2.87 ms. Consequently, the total time for processing one sample is only about 16 ms, which shows that the proposed method is able to be expected to run on small intelligent electronic devices.

IV. CONCLUSION

In this paper, an efficient speech recognition algorithm using *Mel-scale Frequency Cepstral Coefficients* (MFCC) and *Support Vector Machines* (SVM) is proposed. Experiments



(a) (b)

Fig. 2. Confusion matrixs. (a) is for the Part I (electric motor car driving assistant) and (b) is for the Part II (intelligent remote control).

are evaluated on a newly collected speech database by a 5-fold cross-validation. The speech database contains 500 audio files of 10 speech commands for an electric motor car driving assistant and 550 audio files of 11 speech commands for a intelligent remote control. The proposed method acquires 94.20% and 88.73% average accuracy rates for the electric motor car driving assistant and the intelligent remote control, respectively.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under the Grants 61976098, 61602191, 61871434, 61802136 and 61876178, in part by the Natural Science Foundation of Fujian Province under the Grant 2018J01090, in part by the Natural Science Foundation for Outstanding Young Scholars of Fujian Province under the Grant 2019J06017, in part by the Open Foundation of Key Laboratory of Security Prevention Technology and Risk Assessment, People's Public Security University of China under the Grant 18AFKF11, in part by the Science and Technology Bureau of Quanzhou under the Grants 2018C115R, 2017G027 and 2017G036, in part by the Promotion Program for Young and Middle-aged Teacher in Science and Technology Research of Huaqiao University under the Grants ZQN-PY418 and ZQN-YX403, and in part by the Scientific Research Funds of Huaqiao University under the Grants 16BS108.

REFERENCES

[1] E. Wold, T. Blum, D. Keislar, and J. Wheaten, "Content-based classification, search, and retrieval of audio," *IEEE MultiMedia*, vol. 3, no. 3, pp. 27–36, Fall 1996.

[2] G. Guo and S. Z. Li, "Content-based audio classification and retrieval using svm learning," in *IEEE Pacific-Rim Conference on Multimedia(PCM)*, vol. 8, no. 5, 2000, pp. 619–625.

[3] L. Jian, C. Yi-song, S. Zheng-xing, and Z. Fuyan, "Automatic audio classification by using hidden markov model," *Journal of Software*, vol. 13, no. 8, pp. 1593–1597, 2002.

[4] Chien-Chang Lin, Shi-Huang Chen, Trieu-Kien Truong, and Yukon Chang, "Audio classification and categorization based on wavelets and support vector machine," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 644–651, Sep. 2005.

[5] F. Zheng, G. Zhang, and Z. Song, "Comparison of different implementations of mfcc," *Journal of Computer Science and Technology*, vol. 16, no. 6, pp. 582–589, 2001.

[6] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

978-1-7281-3038-5/19/$31.00 ©2019 IEEE