International Journal of Science and Engineering Applications
Volume 12-Issue 04, 107 – 109, 2023, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1204.1036

Phonetic Feature Extraction and Recognition Model in Korean Pronunciation Practice Based on AdaBoost

Chunying Wang
College of Humanities and Foreign Languages
Zhejiang Shuren University
Hangzhou, Zhejiang 310015, China

Abstract:This paper proposes a phonetic feature extraction and recognition model in Korean pronunciation exercises based on AdaBoot moments. Feature extraction algorithms have a great influence on speech emotion recognition algorithms, among which Mel-frequency Cepstral Coefficients (MFCC) is the most commonly used feature in speech emotion recognition. The feature extraction method and the influence of SVM kernel function and parameter selection on the recognition result are studied, and the existing speech feature extraction algorithms and their respective advantages and disadvantages are analyzed, as well as the influence of different kernel functions, kernel parameters and penalty parameters on the recognition performance. An improved grid optimization method is used to further improve the recognition time of voice information.

Keywords: Phonetic Feature Extraction, Recognition Model, Korean Pronunciation, AdaBoost

1. INTRODUCTION

Korean language teaching in higher vocational colleges is an important part of higher vocational education and teaching in our country. The Korean language major in higher vocational education in my country was generally established in 2005. After 10 years of construction and development, a Korean language teaching team of a certain scale has been formed, and a large number of high-quality skilled talents who understand Korean and know skills have been trained and delivered for social development. The training goal of Korean language professionals in higher vocational colleges focuses on the cultivation of students' practical skills in listening, speaking, reading, writing, and translation of the Korean language and the cultivation of their communicative skills. Vocational Korean majors are mainly engaged in Korean language communication or elementary Korean training after graduation, such as Korean tour guides, foreign secretaries, international business assistants, and Korean language training teachers, etc., where oral communication and language coordination are required. These positions require students to meet the phonetic standards and fluency in Korean pronunciation. The Korean pronunciation level directly affects the employment situation of Korean majors in higher vocational colleges [1-6].

Voice is one of the most important natural attributes of human beings. Due to the physiological differences of the vocal organs and acquired behavior differences between different people, the voice of each speaker has strong personal characteristics. This allows people to analyze the voice signal to identify the speaker. In the beginning, people used their ears to identify the identity of the speaker to whom the voice belongs, the so-called "knowing people by hearing". With the rapid development of computer technology, people use computers to automatically recognize the identity of the speaker to which the voice belongs. This is speaker recognition. Recently, with the extensive development of electronic technology, various electronic terminals have penetrated into people's daily lives, playing an increasingly important role, making humans and electronic terminals need more humane interaction methods. At present, the human-computer interaction mode has undergone major changes, and users are no longer satisfied with the cold interaction mode that only uses keyboards, mice, switches and other machinery. Although the recent rapid development of touch screen technology has made human-computer interaction more convenient and more diverse, people generally hope that the human-computer interaction process will be more humane and intelligent, and the human-computer interaction interface will be more friendly and vivid. It is inevitable that computers have the ability to perceive thinking similar to that of humans. Obviously, enabling computers to understand human emotions is a very important step in accomplishing the above tasks. And the online movies we usually watch, as well as some computer application tutorials [7-16].

In addition, with the continuous update and development of educational diversification, distance education has formed a new situation of online learning between students and teachers. Then, with the addition of voice emotion recognition, students can better conduct contextual learning, and computers can also provide real-time feedback based on "calculated" emotions, thereby improving learning efficiency. In the past 40 years, domestic and foreign researchers have conducted a lot of exploration and research in the field of speech emotion recognition, and they have also been used in education, medicine, traffic safety, criminal investigation, service industry, entertainment industry and other fields. However, speech emotion recognition is still in the initial stage of development, so there are still problems that need to be solved. Although Korean teachers in higher vocational colleges have realized the importance of developing good phonetic skills and have carried out a series of reforms in phonetic classroom teaching methods, judging from the overall current situation of the implementation of teaching method reforms, there are still many problem. The main manifestations are: there are more studies on teaching methods, but few real implementations; experienced old teachers have formed a fixed teaching model in the teaching process for many years, and they have less focus on improving teaching methods. Teachers lack education and teaching theory and practical experience, and it is difficult for teachers to enter the role W in the reform of teaching methods in a short period of time. How to further promote the reform

of Korean phonetics teaching methods in higher vocational colleges, choose suitable teaching methods under the guidance of scientific theory, and standardize classroom teaching method management, is an important topic that Korean teachers in higher vocational schools must study and think about. The complexity of the human voice system. Different people have different speaking habits [17-24].

2. THE PROPOSED METHODOLOGY

2.1 The AdaBoost

The AdaBoost algorithm is an ensemble learning algorithm. The idea is to first train multiple different classifiers against the same training set. These classifiers are called weak classifiers; then, the trained weak classifiers are combined and integrated. As a result of classification, a classifier with better performance is established. This classifier is called a strong classifier; finally, the classification result is given by the strong classifier. In the training process, each training data has an initial weight. After each weak classifier is trained, the weight of all training data is adjusted according to the classification result of this weak classifier, and each weak classification is given by calculation. The weight of the device. The new training set whose weight has been adjusted in this training is handed over to the next weak classifier for training. Finally, the weighted sum of all weak classifiers is the decision result of the strong classifier.

This paper proposes the AdaBoost-CRC algorithm, which sends the voice features into the base classifier CRC to train the model and test. The test results of each round will affect the weight distribution of the next round of data. Based on the correct classification of data, the weight is low, and the misclassified data is weighted high. In order to focus on the misclassified data in the next study. The experiment enhances the heterogeneity of the integrated detection model from the feature layer and the base classifier layer respectively. In the feature layer, the voiced frame threshold is randomly selected in the specified range to enhance the feature perturbation effect of the speech layer; the base classifier layer uses the number of generated base classifiers (weak classifiers), and the above multiple different base classifiers are integrated together to form the final output result. This method effectively improves the traditional CRC problem in the case of small samples; at the same time, for the problem of class imbalance, this paper uses the random generation of the number of meta-samples of dictionary atoms to keep the two types of samples selected in a consistent manner and the number of selections is random, which enhances the diversity of base classifiers.

2.2 The Korean Pronunciation Practice

Korean phonetics teaching is the foundation of Korean teaching and the key to the entrance of Korean spoken language. Since all Korean majors in higher vocational education in our country are taught from scratch, students have little understanding of Korean language knowledge before they start their professional studies. This course in higher vocational colleges is called "Korean Intensive Reading"). In the first two to H weeks of the course, special oral speech teaching will be conducted. The content of Korean phonetics teaching in higher vocational colleges includes: Korean pronunciation (consonants, vowels and radio), Korean phonetic rules, word spelling and sentence expression.

In the learning process of higher vocational students, most of them have insufficient learning motivation, poor self-control ability, and have not formed good learning habits and learning methods. In particular, students majoring in Korean in higher vocational colleges learn Korean from scratch. The language learning starts late, with tight schedules, and heavy tasks. It is very easy to be afraid of difficulties and tired of learning. Phonetics is the foundation of foreign language learning. Korean teachers in higher vocational schools must grasp the principles of the following aspects in the phonetic teaching stage to design teaching methods to help students complete the phonetic learning tasks as smoothly as possible. In this paper, the number of atomic element samples of the base classifier CRC dictionary is generated in a random range to meet the integration diversity requirements.

2.3 The Korean Speech Feature Extraction and Recognition Modeling

The Mel frequency cepstrum coefficient is a characteristic parameter that combines the human ear's auditory perception characteristics and speech production. In fact, the level of the sound heard by the human ear is not linearly proportional to the frequency of the sound. Experiments show that Mel has a linear relationship with frequency when the frequency is below 1kHz, and Mel has a logarithmic relationship with frequency when it is above 1kHz. Therefore, the frequency scale division that conforms to the human auditory system should have a higher frequency resolution in the low frequency part and a lower frequency resolution in the high frequency part.

This chapter will use two experimental strategies to test HuWSF: speaker independent (SI) and speaker dependent (SD). In a speaker-independent experiment, all sentences of one speaker are used as test data for each crossover, all sentences of other speakers are used as training data, and sentences of each speaker are used as test sentences once. In the speaker dependence experiment, all sentences in the database are equally divided into 5 parts, one part is used as test data, the other part is used as training data, and each part is used as test data once. The process lasts 10 times, and the average of the 10 results is used as the final result. MFCC, or Mel frequency cepstral coefficient, is a feature widely used in automatic speech and speaker recognition. Different from ordinary cepstrum analysis, the analysis of Mel frequency cepstrum coefficients (or perceived frequency domain cepstrum coefficients, MFCC) focuses on the hearing mechanism of the human ear, and analyzes the frequency spectrum of speech based on the results of auditory experiments. High recognition rate and good noise robustness.

We humans cannot recognize frequencies above 1Khz, so the proposal of MFCC is based on the frequency difference that human ears can distinguish. So far, Mel frequency cepstral coefficients have shown better results in speech recognition and the subjective tone and frequency modeling of audio signals, and they have become the most basic and indispensable feature extraction in the field of speech emotion recognition. Salient features. In the experiment, we used Matlab20I6.r.a to extract MFCC. Cultivating students' good listening, distinguishing and pronunciation skills is the fundamental purpose of Korean phonetics teaching in higher vocational schools. In order to achieve the purpose of sending one, a lot of "imitation and practice" must be adopted. Listening is the fundamental method of Korean phonetic learning.

3. CONCLUSIONS

It is verified that the method of using MFCC + ΔMFCC + E feature combination in speech recognition can improve the comprehensive recognition performance of support vector

machine; the influence of the selection of kernel function and parameters in SVM on the recognition rate has been deeply studied. The simulation results show that the selection of RBF kernel function and Appropriate relevant parameters can improve the accuracy of speech recognition; in addition, the improved grid optimization algorithm can effectively shorten the recognition time and improve the real-time performance.

4. REFERENCES

[1]Zhang Yan. Feature parameter extraction and recognition algorithm research in speaker recognition [D]. Nanjing University of Science and Technology, 2018.

[2] Chen Hongheng. Research on Speech Recognition Method Based on Deep Learning [D]. Harbin University of Science and Technology, 2019.

[3] Xu Ji, Cheng Gaofeng, Pan Jielin, et al. A speech recognition system and method based on a hybrid acoustic model:, CN109754790A[P]. 2019.

[4] Wang Jinhua, Ying Na, Zhu Chendu, et al. Speech emotion recognition algorithm based on spectrogram extraction of deep space attention features[J]. 2021(2019-7):100-108.

[5] Wang Jinhua, Ying Na, Zhu Chendu, et al. Speech emotion recognition algorithm based on spectrogram extraction of deep spatial attention features[J]. Telecommunications Science, 2019, 35(7):9.

[6] Chen Shu, Yu Haibo. Application of an improved feature extraction method in speech recognition [J]. Sensors and Microsystems, 2018, 037(005):154-157.

[7] Zhang Xingming, Yang Kai. Research on speech feature parameter extraction in speaker recognition in deep learning [J]. Modern Computer, 2021(8):6.

[8] Liu Zhengchen. Research on Speech Generation Method Combining Pronunciation Features and Deep Learning [D]. University of Science and Technology of China, 2018.

[9] Wu Junqing, Ni Jiancheng, Wei Yuanyuan, et al. Rxk ensemble model for statistical features in speech emotion recognition[J]. Journal of Qufu Normal University: Natural Science Edition, 2020, 46(2):6.

[10] You Yongbin. An acoustic recognition model, method and system for Chinese and English mixed speech:, CN110930980A[P]. 2020.

[11] Zou Yuexian, Luo Danqing. A speech emotion recognition model and recognition method based on joint feature representation: CN108899051A[P]. 2018.

[12] An Tongxun, Park Chiyan, Li Minming, etc. Methods and devices for updating language models and performing speech recognition:, CN106409284B[P]. 2019.

[13] Li Wei, Qian Binghua, Jin Xingming, et al. A voiceprint recognition method and device:, CN106098068B[P]. 2019.

[14] Pan Wenjiao. Comparison of Korean Chinese Phonetics and Research on Korean Phonetics Teaching Methods[J]. Shanxi Youth, 2018, 000(021): 26-27.

[15] Li Yuqing, Xu Cheng. Application of intelligent speech analysis technology based on auditory attention model and BiLSTM and CNN [C]// China Computer Users Association Network Application Branch 2019 23rd Annual Conference on New Network Technologies and Applications Proceedings. 2019.

[16] Ji Xuan, Yu Meng, Zhang Chunlei, etc. Target speech extraction method, device, equipment, medium and joint training method: CN111179911A[P]. 2020.

[17] Zhang Hua, Dai Meixiang, Dai Guojun, et al. A speaker recognition method based on speech feature fusion and GMM: CN110415707A[P]. 2019.

[18] Li Na, Wang Jun. A voice identity feature extractor, classifier training method and related equipment: CN109584884A[P]. 2019.

[19] Wen Bing, Chen Zisang. The English-Chinese Phonetic Comparison Teaching Model and Its Feasibility Research——Take Vowel Comparison as an Example[J]. Journal of Shantou University (Humanities and Social Sciences Edition), 2019, 035(001): 46-55.

[20] Li Na, Ge Wancheng. Model training and performance evaluation of speech keyword recognition system[J]. Information and Communication, 2020(3):3.

[21] Feng Meng, Jia Yanming, Zhang Weiyu, et al. A method for evaluating advanced English pronunciation skills based on speech recognition: CN110858482A[P]. 2020.

[22] Guo Weitong, Yang Hongwu, Gan Zhenye. A voice and facial feature extraction method and system for depression detection: CN109171769A[P]. 2019.

[23] Xiong Huiyuan, Chen Caiting, Liu Sheng, et al. Argument recognition method based on audio analysis and deep learning:, CN110956953A[P]. 2020.

[24] Xi Zexi, Zhao Lijun. A voice service quality detection method, model training method and device: CN111538809A[P]. 2020.