www.cadanda.com
Computer-Aided Design AND APPLICATIONS

Improvement and Optimization Method of College English Teaching Level Based on Convolutional Neural Network Model in an Embedded Systems Context

Xi Li[1*] and Xiaoxin Huang[2]

[1]School of Foreign Languages, Shaoyang University, Shaoyang 422000, China
lixi198307@163.com
[2]Teaching Affairs Section, Shaoyang Open University, Shaoyang 422000, China
Xiaoxin.Huang2021@outlook.com

Corresponding author: Xi Li, lixi198307@163.com

Abstract. AI will be the science with the aim to explore the rules of human beings for activities. The AI will be employed in the teaching of schools and this will rise to the trouble that teaching ones are so small and the educating mode is very single. In this way, the teaching rate will be enhanced and the burden of teachers will be decreased. Students will arise the self-learning capabilities. In line with the pushing process of China's opening up and the globalization, with the aim to get through the obstacles of languages, it is very vital to speak English rather than others. During the procedure of current English teaching, the level of speaking will limit the improvement of the educating English. With the aim to deal with the trouble, the study will focus on the network of ways of counting and come up with the framework to realize the goal. The teaching in line with the AI ways of counting and neutral network , the one will study the features of these ones as well as SVM classification ability of deep convolutional. In this way, automatic detection of English spoken English pronunciation is realized, and the pronunciation caused by the position, action and pronunciation length of pronunciation-related organs are not standard, and the wrong ones for such pronunciation of learners are sorted out and located, and the pronunciation problems of learners will be identified, and different kinds of wrong ones are provided and corrected.

Key words: neural network; colleges and universities; English teaching; Embedded Systems Context
DOI: https://doi.org/10.14733/cadaps.2024.S8.212-227

1 INTRODUCTION

With the advancement of the reform for China as well as the fast growth of economic globalization, especially when the Chinese government proposed the "Belt and Road" and "Made in China 2025" and other major development strategies that benefit the world and more and more Chinese The company has gone abroad to become a multinational company. In order to break through the barriers of language communication and achieve better development prospects, it is more important to have high-level English compound talents [20]. Different from English learning in the past, in addition to focusing on written test scores, oral pronunciation is becoming more and more important[1]. In the 2018 National College English Ranking Test, the official addition of the oral test is a good reflection. However, spoken English pronunciation is largely influenced by the habit of native speakers. Although the official language of China is Putonghua, and China is a country with abundant language and cultural resources, the existing languages (dialects) can be divided into 8 major factions, and there are more than 100 dialects [2]. The diversity of language and culture has led to many pronunciation problems in foreign language learning, not only in spoken English, but even Mandarin pronunciation is also affected by dialect habits. Although there have been oral English courses since elementary school, most college students have not outstanding oral English ability, and even cannot cope with some simple daily communication [3].

The emergence of artificial intelligence will have such unprecedented differences of living space as well as human interaction big data as well as cloud computing, skills related technologies will be rapidly employed to various disciplines, making people's lives more and more convenient [4]. The penetration of one technology for the part of college English teaching continues to increase, subverting the long-standing traditional foreign language teaching model, changing the original ecological environment of college English teaching, innovating foreign language teaching application tools, and providing foreign language teaching and learning. With new resources, intelligent foreign language teaching has become the norm [5]. In the field of college English teaching, embedded systems are revolutionizing traditional foreign language teaching models. Intelligent foreign language teaching tools, often embedded systems themselves, are being employed to provide new resources and enhance the learning experience. These tools may include language learning apps, interactive platforms, or AI-powered tutors that adapt to individual learning styles and provide personalized feedback.

Pronunciation is an important part of oral English teaching. Pronunciation error detection and diagnostic feedback can analyze learners' pronunciation problems and give pronunciation correction suggestions, thereby improving language learners' pronunciation level and learning efficiency. Compared with classroom language teaching limited by time and space, computer automatic pronunciation error detection has many advantages such as real-time, convenient and efficient [6]. In view of the fact that most of the current academic researches on pronunciation error detection only focus on the detection of pronunciation errors, while ignoring the actual situation of the importance of feedback correction. The phoneme-level pronunciation error type caused by the non-standard of the students has carried out error detection research, combined with the machine learning algorithm to realize the automatic detection of spoken English pronunciation. The consequences show that the one can effectively correct students' spoken English pronunciation, effectively improve the level of English teaching in colleges and universities, and has a good application prospect.

2 RELATED CONCEPTS AND THEORETICAL BASIS

Artificial intelligence will be employed in the transformation of teaching environment, using artificial intelligence technology to build a smart learning experience, teaching method and teaching management environment for learners, integrating big data and deep learning algorithms into

different aspects of English teaching, in a smart classroom environment , the human-computer interaction enables learners to participate in high enthusiasm, and at the same time can develop a personalized learning plan [7]. Teaching intelligence is a fundamental change in the realization of teaching tools, teaching methods and even teaching content in an intelligent teaching environment. Using artificial intelligence technology to transform traditional teaching concepts has become the flow in the foreign language as well as foreign one for such artificial intelligence environment has given full play to its advantages [8].

2.1 The Principle and Characteristics of Human Voice

The oral cavity is the main cavity for humans to adjust the pronunciation of different voices. Changes in the position of the tongue during articulation can change the air passage to the nasal cavity or oral cavity in different directions [9]. The shape and volume of the nasal cavity are unchanged, while the mouth can change shape and volume with the movement of the lips and tongue, which is the main reason why humans can make various sounds. According to the research results, the actions form the tongue will influence the one, especially in such pronunciation of vowels. The tongue movement in vowel pronunciation of different language families is here as shown in Table 1.

vowel language family	*mandarin*	*English*	*Spanish*	*Japanese*	*Korean*
a	*central*	*rear*	*backward*	*backward*	*anterior*
e	*rear*	*rear*	*Center*	*front*	*Center, front*
i	-	-	-	-	-
o	*medium high low*	*Middle and high*	*too high*	*medium high low*	*medium high low*
u	*high*	*high*	*high*	*high*	*high*
ü	-	-	-	-	-

Table 1: Tongue Movements in Vowel Pronunciation in Different Language Families.

Automatic pronunciation error detection and correction, as the name implies, is to eliminate human subjective interference, and let the machine automatically detect errors in human language pronunciation and provide feedback and correction suggestions. It is a subject involving multiple disciplines and technologies [10]. When performing pronunciation error detection, if the text corresponding to the error-detected speech segment is known to the system, it is called text-related pronunciation error detection, and when the system does not know the corresponding text, it is called non-text-related pronunciation error detection. From the point of view, the system can only deal with the problem of pronunciation error detection of known text. According to the types of pronunciation errors, phonemic pronunciation errors can be divided into prosody errors and phonemic errors [11]. Most of the errors in pronunciation error detection, such as phoneme misreading, missing reading and insertion, are all phonemic errors. The pronunciation error detection in this paper is to provide direct corrective feedback for learners, so this paper focuses more about pronunciation-related organs as well as some movements.

2.2 Acoustic Characteristics of Speech Signals

The varying stationary one in a short period, and the information it contains will be the two ones: the first one is semantic information, and the second one is acoustic information [12]. There are

many studies on semantics, but they have little to do with this study. This article mainly involves acoustic information, and the key information contained in acoustic features is the basis of this study. Acoustic features have a certain uniqueness, and the characteristics of the signal can be more accurately expressed by analyzing and extracting these feature parameters. Therefore, the extraction of acoustic feature parameters of speech plays an important role in the pronunciation error detection system. Usually, the acoustic features of speech can be divided into two categories: the first category is amplitude, energy, and zero-crossing rate, which are the characteristics of speech signals in the time domain. The second category is the transformed frequency features, such as linear prediction coefficients (LPC), formant features, Mel cepstral coefficients (MFCC), etc. [13]. Generally speaking, it is not easy to analyze the characteristics carried by the speech signal in the time domain, so it is usually observed by converting it from the time domain to the characteristic distribution in the frequency domain. Linear prediction coefficients, formant features and Mel cepstral coefficients are mainly used in pronunciation error detection research [14].

(1) Linear prediction coefficient. Since the transition between sampling points is relatively smooth, linear predictive analysis can use this feature of speech to predict the current or future sample value based on the p sampling point values in the past. Set at time n, the sample value of the speech signal is s(n), and its predicted value is:

$$\hat{s}(n) = \sum_{i=1}^{p} a_i s(n-i) \tag{1}$$

The above formula is called a linear predictor. If P is the order of the linear predictor, its system function is as follows:

$$P(z) = \sum_{i=1}^{p} a_i z^{-1} \tag{2}$$

Linear prediction error. can be expressed as:

$$\varepsilon(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^{p} a_i s(n-i) \tag{3}$$

(2) Formant estimation. When studying vowels, the formant peak is a very good acoustic feature. According to acoustic knowledge, the vibrating vocal fold signal entering the vocal tract will cause vocal tract resonance. The resonance frequency at this time is called the formant frequency [15]. The information of the formant is contained in the spectral envelope curve, and it is generally considered that the maximum position on the spectral envelope curve of vowels and loud consonants is the formant. Different pronunciations have different formant positions. According to the height of the peak, it can be divided into the first formant, the second formant, the third formant and the fourth formant (F1, F2, F3 and F4). Generally, the average of each formant is taken. Value, average rate of change, maximum value, minimum value, mean square error, formant frequency and one-third and one-quarter quantiles of formant change were counted [16].

(3) Mel cepstral coefficient (MFCC). MFCC mimics the human speech production and auditory system. According to the theoretical research of human earphones, the hearing sensitivity of the human ear to sound waves varies with frequency. frequency. MFCC takes advantage of this feature [19]. Since this feature is developed on the human auditory model, MFCC is more stable than linear prediction based cepstral coefficients, and can maintain good performance when the signal-to-noise ratio is reduced. MFCC utilizes the human auditory model and converts the linear spectrum to a Mel-scale spectrum based on the non-linear characteristics of the frequency of the human ear. The relationship between ordinary frequency and Mel-scale frequency is:

$$Mel(f) = 2595 \times \lg(1 + \frac{f}{700}) \tag{4}$$

f is the frequency in Hz. The relationship between linear frequency and Mel-scale frequency, the distribution in the low-frequency part is denser than that in the high-frequency part, the Mel frequency gradually slows down with the increase of f, and in the Mel frequency domain, the human ear is more sensitive to the pitch. perception becomes a linear relationship. That is to say, the pitch change that the human ear can perceive is the same as the Mel frequency change trend of the audio signal itself. Based on this, a Mel-scale filter bank was designed, which can further reduce the dimension of the spectrogram into a Mel-spectrum. A Mel-scale filter bank consists of multiple equal-height triangular filters, which can be defined as:

$$H_f(k) = \begin{cases} 0 & k < f_{b_i-1} \\ \dfrac{k - f_{b_i-1}}{f_{b_i} - f_{b_i-1}} & f_{b_i-1} \le k \le f_{b_i} \\ \dfrac{f_{b_i-1} - k}{f_{b_{i+1}} - f_{b_i}} & f_{b_i} \le k \le f_{b_{i+1}} \\ 0 & k > f_{b_i+1} \end{cases} \tag{5}$$

Mel cepstral coefficient (MFCC) is a very widely used audio signal representation method. The general process of MFCC extraction is shown in Figure 1. Take the logarithm of the Mel spectrum obtained above, and then do a discrete cosine transform (DCT) to get the Mel-frequency cepstral coefficients. The main reason for taking the logarithmic operation is that the sound signal emitted by humans is formed by convolution of the pitch information and the vocal tract information. After the Fourier transform, the relationship between the two becomes a multiplicative relationship. This can be further transformed into a linear plus relationship to separate the two. The conversion from Mel spectrum to MFCC coefficients can be expressed as:

$$C_j = \sum_{i=1}^{M} X_i \cos\left[j(i - \frac{1}{2}) \frac{\pi}{M} \right], i = 1, 2, \cdots, j \tag{6}$$

in,

$$X_i = \log_{10}(\sum_{k=1}^{N} |X(k)| H_i(k)), i = 1, 2, \cdots, M \tag{7}$$

In the formula, Xi is the logarithmic energy output by the i-th filter in the Mel-scale filter bank, and M and J are the number of filters and the number of cepstral coefficients, respectively. It can be seen that MFCC uses a set of feature parameters to represent the characteristics of the audio signal, which compresses the original two-dimensional signal into one-dimensional, and also causes some information loss in the process of extracting this set of coefficients. In addition, considering the strong learning ability of convolutional neural networks for image features, this paper only extracts the Mel spectrum of audio signals as shown in Figure 1.

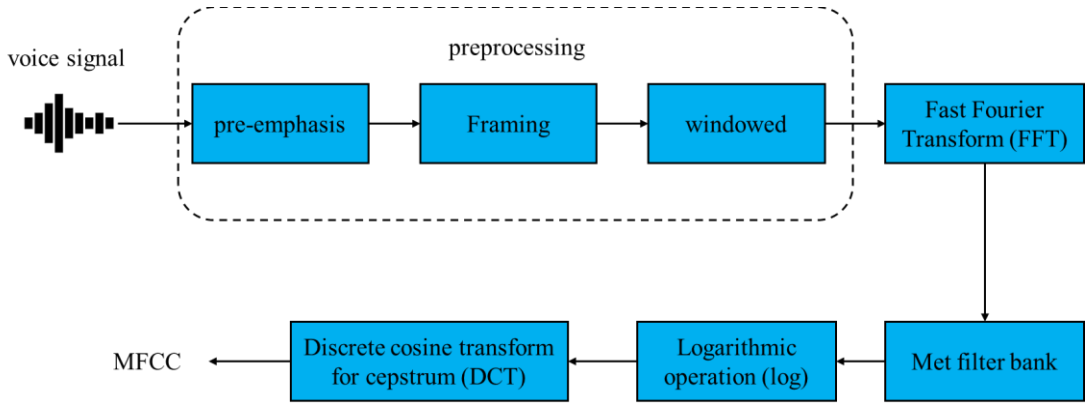

Figure 1: Flow Chart of Mel Cepstral Coefficient Extraction.

2.3 College Spoken English Teaching Based on Speech Recognition Technology

Speech recognition is a comprehensive technology covering many aspects, including signal processing technology, computer science technology, etc. The technical principle of speech recognition is image recognition, and its general process can be summarized as: preprocessing - feature extraction - pattern matching based on the speech model library - language processing under the language model library - complete recognition (see Figure 2) [18].

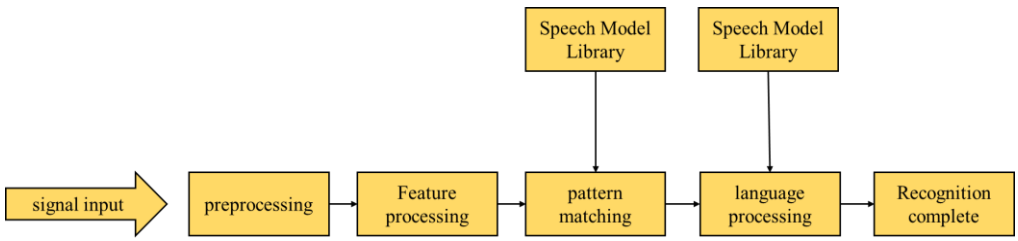

Figure 2: Demonstration Diagram of Speech Recognition Process.

The principle of speech recognition mainly covers the following contents: firstly, the information coding in the speech signal is converted according to the time of the amplitude spectrum; secondly, the speech can be represented by an acoustic signal with multiple independent symbols; finally, the speech interaction cannot be combined with grammar, Semantics are separated.

The main way of information interaction between humans and machines is intelligent speech recognition. This general-purpose technology has attracted widespread attention, and more and more experts and scholars have carried out active research on it. With the continuous progress and development of the times, teachers are increasingly applying artificial intelligence technology in the teaching process, so as to improve students' The effect of oral training. In natural speech recognition, it is mainly to establish related systems and models for natural language, etc., so that it collects and organizes the speech materials of the recognized objects, summarizes and forms samples, and adjusts the various items in the recognition system through a lot of training. content, so that the recognition system can carry out the recognition more accurately, and perform natural speech simulation recognition on the speech of each object. Trainers can train in any system and get corresponding evaluation results, so as to continuously develop and improve themselves. Artificial

intelligence provides suggestions for improvement through efficient and targeted evaluation results of oral language practitioners, as well as intelligent analysis of listening and speaking training process, listening comprehension ability, language expression ability, etc. After the oral English practitioner completes a single training session, the system will immediately give the training evaluation results and give the oral English practitioner real-time feedback, which can greatly enhance the oral English practitioner's enthusiasm and interest in continuing to participate in the training, and quickly improve the college English speaking practitioners. listening level[17].

3 METHODOLOGY

3.1 Convolutional Neural Networks

The classifier adopts Softmax mutual exclusion classifier. After learning and comparing a variety of activation functions, this paper will focus on researching and comparing two activation functions, ELU and ReLU [21]. The aim of the will repeat the process and rise the features of the image and at last through the layer the whole connection layer will be the output , the model we use is like this in the picture:

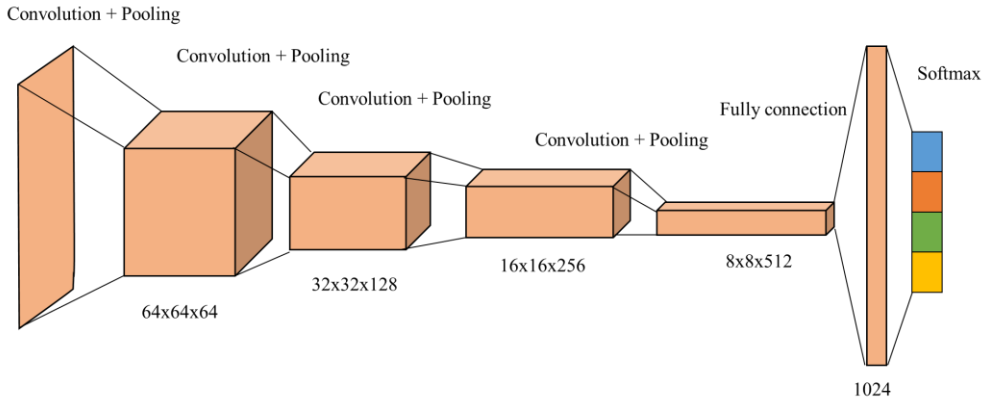

Figure 3: CNN Training Model.

3.2 Design of Neural Network Training Model

The samples of the training one will be finished and well designed in the process. We need to make sure the condition part. The condition has considered other good and deep ones in line with the neural networks in the test in a long time. they are mainly studied and compared, namely linear unit (ELU) and corrected linear unit (ReLU). The optimal controller compares Adam and RMSProp are two gradient descent methods, and the classifier adopts Softmax mutual exclusion classifier. After learning and comparing a variety of activation functions, this paper will focus on researching and comparing two activation functions, ELU and ReLU. Among them, ReLU is a very efficient and widely used activation function, the equation is as follows:

$$ReLU(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \tag{8}$$

Since the input x itself is taken in the positive range, the ELU alleviates the gradient dispersion problem to a certain extent (the derivative is equal to 1 in the range of x>0), and the equation is as follows:

$$f(x)=\begin{cases} x & if \quad x>0 \\ \alpha(\exp(x)-1) & if \quad x\le 0 \end{cases} \tag{9}$$

When stochastic gradient descent (SGD) is iteratively trained, only The samples of the training one will be finished and well designed in the process. We need to make sure the condition part. The condition has considered other good and deep ones in line with the neural networks in the test in a long time , that is, m=1 in the batch gradient descent method. The iterative process is as follows:

$$J(\theta_0,\theta_1)=\frac{1}{2}(h_0(x^{(i)})-y^{(1)})^2 \tag{10}$$

$$\theta_j:=\theta_j-\alpha\frac{\partial}{\partial\theta_j}J(\theta_0,\theta_1) \tag{11}$$

Therefore, the stochastic gradient descent method has a faster training speed, and the gradient descent method is also faster than the batch gradient descent method, and has a good learning convergence effect for a large number of data samples. Inevitably, the gradient of stochastic gradient descent cannot descend steadily, because the gradient changes only come from random partial samples, not from the overall macroscopic consideration. This means that there is an error in the gradient, and it affects the next gradient descent, which shows that the gradient changes irregularly in the iterative process.

The RMSProp algorithm introduces the concept of momentum on the basis of SGD, that is, the gradient of the first few times will also participate in the operation. Here, r is introduced to represent the cumulative gradient, and the r and parameter update process is as follows:

$$r=\rho r+(1-\rho)g\cdot g \tag{12}$$

$$\Delta\theta=-\frac{\alpha}{\delta+\sqrt{r}}g \tag{13}$$

$$\theta_j:=\theta_j+\Delta\theta \tag{14}$$

Among them, ρ represents the decay rate, g represents the gradient value, and δ represents the numerical stability, which is a small constant, about 10e-7, to prevent division by 0. The RMSProp algorithm enables automatic changes to the learning rate. If the gradient is relatively large this time, speed up the decay of the learning rate; otherwise, slow down the decay of the learning rate. RMSProp is suitable for dealing with non-stationary targets and is very suitable for CNN models.

The Logistic function can only be applied to binary classification problems, but its polynomial regression, the Softmax classifier, can solve classification problems of multiple categories. Assuming that the Softmax function is, the input z is a vector of C dimension, then the output of the Softmax function is also a vector y of C dimension, with a value of 0 or 1.

$$y_c = \sigma(z)_c = \frac{e^{z_c}}{\sum_{d=1}^{C} e^{z_d}} \tag{15}$$

$$\sum_{c=1}^{C} y_c = 1 \tag{16}$$

The denominator in equation (16) acts as a regular term, which can make the sum of the probabilities of yc equal to 1. As the output layer of the neural network, C neurons in the Softmax function represent C categories. It can be seen from this that the Softmax function is a normalized exponential function. Since the sum of each classification probability is 1, the categories classified by the Softmax function are mutually exclusive. The logistic classifier of C combinations can also solve the classification problem of C categories, but these categories are not mutually exclusive.

3.3 Pronunciation Classification Error Detection Model Framework Based on Dbn-Svm

When the machine learning algorithm processes the sample data, it often faces the problem of unbalanced sample data. For the model training process, when the sample data volume differs greatly from class to class, the model with a large sample size is more likely to be discriminated by the model. It is easy to obtain a lower loss, which leads to the loss of motivation to continue optimizing the parameters, which will cause errors in the final training model results. The methods generally used to deal with the unbalanced distribution of sample data include up-sampling, down-sampling, data synthesis, weighting, One-Class, etc. Among them, OneClassSVM is a common method to solve data imbalance. Different from two-class classification, one-class SVM is a special classification method. Most samples have only one class of labels. Unlike ordinary SVMs, OneClassSVM only needs to determine the boundaries of such labeled samples, and the data outside the boundaries are classified into other classes. When judging multiple classes, it can be considered to model each class as a class, so that the influence of imbalanced sample data can be ignored. So it is mainly suitable for solving the problems of anomaly detection and unbalanced sample classification.

The construction process of the pronunciation error detection classification model based on DBN-SVM is shown in Figure 4. The construction process of the pronunciation classification error detection model is as follows:

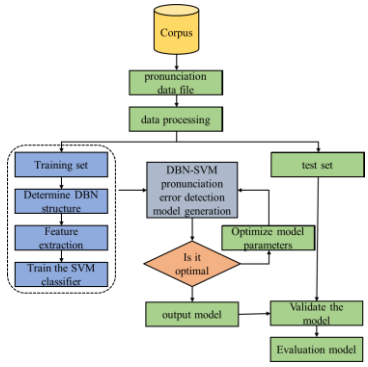

Figure 4: Flow Chart of Construction of Pronunciation Classification error Detection Model Based on DBN-SVM.

Step1: Data acquisition, first collect pronunciation data files from the corpus, use the Hidden Markov Model Toolbox (HTK) to force the alignment of the audio files with the reference text, and align the speech to the sentence, word and phoneme levels. By forced alignment, the alignment time

information at the phoneme level is obtained, and it is cut and separated according to the phoneme alignment time information. The phoneme data is obtained and processed as a dataset of pronunciation error detection classification model.

Step2: Determine the optimal DBN structure, use the training samples in the pronunciation classification error detection data set as the input of the visible layer of the DBN model, determine the number of hidden layers, batch parameters, iterations and learning rate of the DBN network to pre-train the DBN network. Use DBN to perform feature learning on pronunciation data and extract deep hidden features of pronunciation data.

Step3: Perform parameter training on RBM, which is the basic component of DBN. In this paper, the Contrastive Divergence (CD-K) algorithm is used to train the RBM in the pronunciation classification error detection model. The rapid learning of RBM can be roughly divided into four steps:

1 Determine the number n of hidden layer neurons in the RBM, the learning rate α of the RBM and the number of iterations according to the number and dimensions of the collected feature datasets for pronunciation classification error detection. And set the initial network parameter set of RBM.

2 Through Gibbs sampling, initialize the visible layer of the RBM with the input training samples, and calculate the state of each neuron in the hidden layer through the formula. After determining the hidden layer state, the visible layer neuron state is calculated. Then use the gradient descent algorithm to update the network to get the updated state of the hidden layer neurons of the network.

3 Parameter adjustment, from the state of each neuron in the visible layer and the state of each neuron in the hidden layer obtained in the second step, the updated parameter set of the network is obtained.

4 Under the number of iterations initially set in the first step, the optimal parameter set of the RBM network is obtained by repeating the above two steps of training.

Step4: Continue to perform the steps in Step4, train layer by layer, and complete the training of all the stacked RBMs in the DBN. Output the features extracted by the deep network.

Step5: Reduce the dimension of the output feature of each error type as the input vector of the One Class SVM classifier, and establish a one-class support vector machine model for each of the six types of pronunciation errors. Perform back-propagation to optimize the DBN network trained in the previous step, and fine-tune the DBN-SVM model parameters for the determined pronunciation classification error detection.

Step6: In the testing phase, input the test set of the pronunciation classification error detection feature data in Step 2 into the DBN-SVM pronunciation classification error detection model that has been constructed, and compare the actual pronunciation error category of the test set with the model classification error detection output. Calculate the classification error detection accuracy and error of each error type.

Step7: Model evaluation, calculate various evaluation indicators, as well as the Model evaluation, calculate various evaluation indicators, model parameters for the determined pronunciation classification error detection, and analyze and evaluate the pronunciation classification through the evaluation indicators [22].

4 EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Data Acquisition and Evaluation Indicators

After separating the phonemes from the collected speech data, extracting the acoustic features and processing them into a feature data set, the 6 error types that were manually annotated by the corpus experts were raising, lowering, and lowering the tongue. A total of 1290 pronunciation samples were collected in the experiment for fronting, backing, lengthening and shorting. Table 2 is a sample information table.

data set / error type	Training set	test set
	1032	258
	256	64
Raising	256	64
lowing	256	64
shorting	80	20
backing	64	16
lengthening	120	30

Table 2: Sample Information Sheet.

Fine-tune the DBN-SVM model parameters for the determined pronunciation classification error detection: 1) Correct Acceptance (CA), fine-tune the DBN-SVM model parameters for the determined pronunciation classification error detection ; 2) False Rejection (FR), that is, the number of correct pronunciations judged as the current mispronunciation type; 3) Correct other (CO), that is, the number of correct pronunciations judged as other mispronunciation types; 4) Correct Rejection (CR), that is, the number of the current mispronunciation type is judged as the current mispronunciation class; 5) Falsely accepted (FA), that is, the number of the current mispronunciation type is judged to be the correct pronunciation type; 6) False other (FO), That is, the current pronunciation error type is judged as the number of other pronunciation error types. There are three main evaluation indicators:

That is, the number of the current mispronunciation type is judged as the current mispronunciation class:

$$Accuracy = \frac{CR + CA}{CR + FA + FO + CA + FR + CO} \tag{17}$$

$$Recall = \frac{CR}{CR + FA + FO} \tag{18}$$

that is, the number of the current mispronunciation type is judged as the current mispronunciation class

$$Farate = \frac{FA}{CR + FA + FO} \tag{19}$$

4.2 Comparison and Analysis of Experimental Results

We divided a total of 1290 error samples of six types and 320 pronunciation data samples of correct pronunciation data into the one, and used the training set to divid a total of 1290 error samples of six types and 320 pronunciation data. After training, use Test the model on the labeled test set. In the experiment, 150, 130 and 100 neuron nodes are used to model different error types under a single hidden layer. By comparing the output results, it is found that better results can be obtained when the number of nodes in the first layer is 130. When the number of nodes in the first layer is determined, the number of hidden layers is increased, and the number of neurons in the second and higher layers is determined by the experimental method. Finally, the optimal model is obtained by adjusting other parameters. The number of hidden layers is 2. The number of layer neuron nodes is 130-100. The learning rate is 10-5 and the number of RBM iterations is 25. With the increase of Epochs of different sample batching times, the loss trend of Raising-type training set and validation set in the model classification error detection results decreases rapidly, and finally reaches a plateau as shown in Figure 5. Similarly, with the increase of epoch, the correct rate of pronunciation classification error detection in the validation set also increased to more than 0.7, as shown in Figure 6, and the correct rate in the test set was 82.5%.

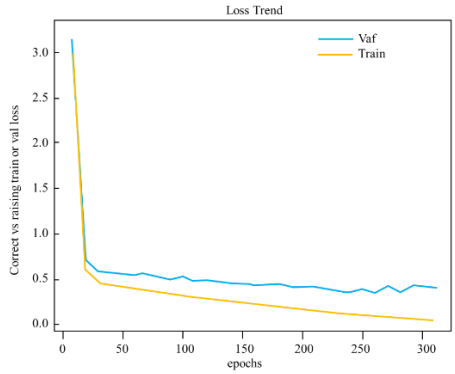

Figure 5: Loss Trend of Raising Error Test Results.

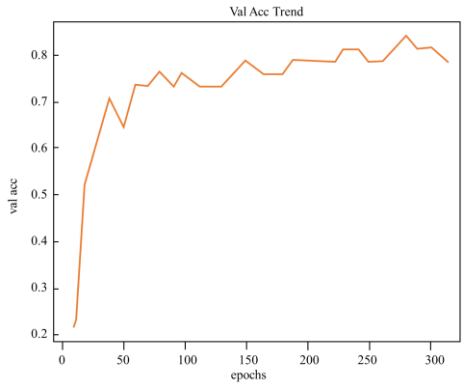

Figure 6: Error Detection Accuracy of Raising Error Test Results.

The results of pronunciation classification error detection of other error types such as Lowing and Shorting are shown in Figure 7 to 10. The results of the test set show that with the increase of the number of Epochs of sample batch processing, the loss trend of the training set and the validation set decreases rapidly, while the test set The error detection accuracy rate can basically rise to between 75% and 85%, indicating that the model constructed by this method has a good recognition effect on pronunciation errors.

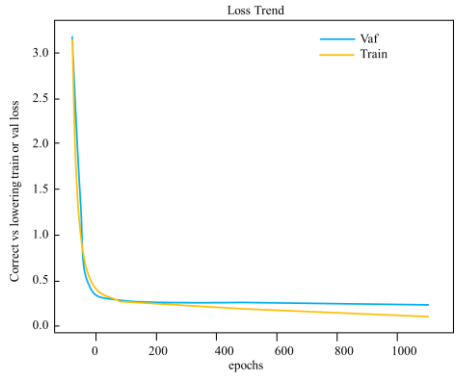

Figure 7: Lowing Error Test Results Loss Trend.

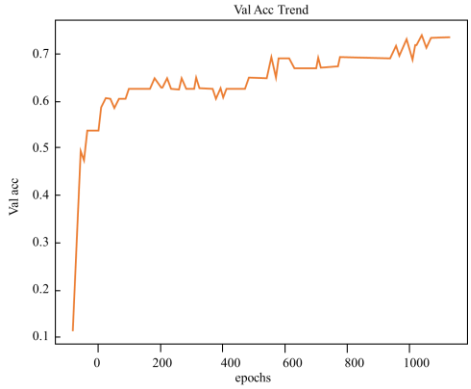

Figure 8: Error Detection Accuracy of Lowing Error Test Results.

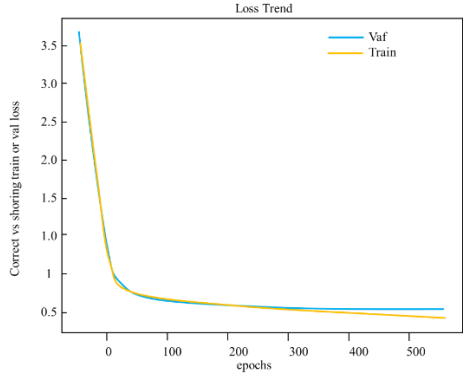

Figure 9: Shorting Error Test Results Loss Trend.

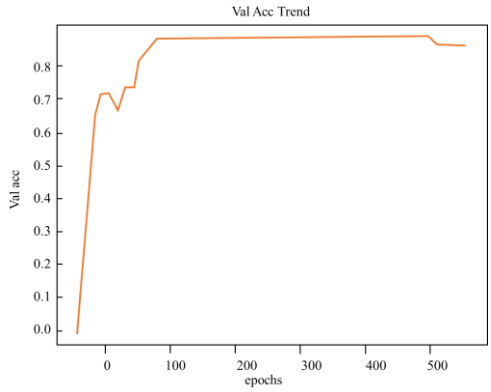

Figure 10: Shorting Error Test Results Error Detection Accuracy.

The error detection of the above three types of errors shows the effect of the model on the recognition of pronunciation errors. With the aim to further highlight such stress of the ways of counting, we also tested the same data samples with different feature extraction algorithms combined with support vector machine classifiers in the experiment, including these three ones, SVM parameter settings are the same as in this chapter, it will be employed to get out of the principal component features of pronunciation data and the features extracted by LDA are used as the input of SVM respectively. We take the training time, error detection accuracy and standard error in the test results of the three error types of Lengthening, Fronting and Backing to compare who they behave in the algorithm. Such results between these two algorithms and the model in this chapter will be here in the following ones:

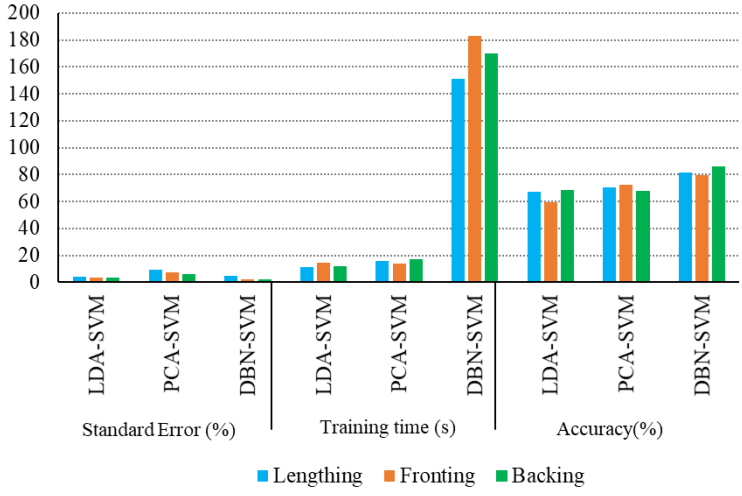

Figure 11: Performance Comparison Results of Three Algorithms (Changed into Grouped Histogram).

From Figure 11, such will be clearly noted that one pronunciation one error detection, such error detection accuracy of this model is basically unmoved at about 80%, ahead of the other two algorithms. The error detection accuracy of LDA-SVM The lowest, basically rarely reaches more than

70%, and the model error detection accuracy rate based on principal component analysis is moderately stable to between 65% and 75%. However, in terms of training time, compared with other methods, because this model adopts a deep neural network structure, the feature dimension is high, and the iteration of the model during training requires more resources, which is not dominant in training time. In terms of standard error, the same model differs between different types of pronunciation errors. The standard error of this model is larger than that of the other two error types in terms of phoneme lengthening, indicating that individual fluctuations in sample error detection results too large. LDA-based methods are relatively balanced in terms of standard errors.

5 CONCLUSION

People employ the class learning as the old and traditional way to study. There are at least 10 or more students, so it is not practical to teach the tones of English from the first one the last one. Generally, classroom teaching can only be done in terms of different kinds of things for English. We are short of teaching guidance language learning which is the shortcomings for the real class. The failure to obtain correct and incorrect feedback for pronunciation exercises after class has caused the trouble and it is in our country in the past few years. It is a common problem for English learners that they dare not speak and are afraid of making mistakes. Teachers' one-to-many teaching cannot accurately correct students' behavior English pronunciation has led to a low level of English teaching in schools.

With such development, there are much ways to learn English. The core of computer-assisted pronunciation training is pronunciation error detection and feedback correction. Because the previous pronunciation error detection focused on phoneme insertion, misreading and missing reading, etc. Typical errors, and there are very few errors in the learner's pronunciation action. With the aim to give advice for improving such learner's pronunciation more intuitively, one combines the one to construct a pronunciation error detection model to fill the gap in pronunciation action error detection. by classifying pronunciation errors and clarifying learners' pronunciation problems, it becomes a reality for ones for different types of errors. Through such effective classification and correction guidance for pronunciation errors, it can effectively correct students' oral English pronunciation, can arise the level of English for schools, and has a good application prospect.

Xi Li, https://orcid.org/0009-0005-3001-5734
Xiaoxin Huang, https://orcid.org/0009-0005-6974-8417

REFERENCE

[1] Chen, X.; Zhang, F.: An Experimental Report on Computer Network Assisted College English Teaching, 2021. https://doi.org/10.1145/3482632.3483112
[2] Cui, M.; Yaguang, W. U.; Office, D.; et al.: Evaluation of Innovation and Entrepreneurship Teaching Ability of College Teachers Based on Improved BP Neural Network, Journal of Jianghan University (Natural Science Edition), 2018.
[3] Feng, L. U.; University, W. N.: Development and Application of English MOOc System Based on Neural Network, Automation & Instrumentation, 2019,
[4] Geng, L.: Evaluation Model of College English Multimedia Teaching Effect Based on Deep Convolutional Neural Networks, Mobile Information Systems, 2021. https://doi.org/10.1155/2021/1874584
[5] Guo, C.; Liu, Y.: BP Neural Network-Based Evaluation on University Teachers' Teaching Quality; proceedings of the EBIMCS 2020: 2020 3rd International Conference on E-Business, Information Management and Computer Science, F, 2020. https://doi.org/10.1145/3453187.3453348

[6] Han, B.; Yao, Q.; Yu, X.; et al.: Co-teaching: Robust Training of Deep Neural Networks with Extremely Noisy Labels, 2018.
[7] Kim, J. Z.; Lu, Z.; Nozari, E.; et al.: Teaching Recurrent Neural Networks to Modify Chaotic Memories by Example, 2020,
[8] Lang, A.: Evaluation Algorithm of English Audiovisual Teaching Effect Based on Deep Learning, Mathematical Problems in Engineering, 2022. https://doi.org/10.1155/2022/7687008
[9] Li, J.: Based on Multimedia and Network Social Environment--A study of Pragmatic Acquisition in College English Teaching, Contemporary Education Research (100 Photos), 5(4), 2021. https://doi.org/10.26689/jcer.v5i4.2046
[10] Liu, T.; Ning, L.: Deep Convolutional Neural Network and Weighted Bayesian Model for Evaluation of College Foreign Language Multimedia Teaching, Wireless Communications and Mobile Computing, 2021, 3, article e3080, 1-7. https://doi.org/10.1155/2021/1859065
[11] Liu, T.: Convolutional Neural Network-Assisted Strategies for Improving Teaching Quality of College English Flipped Class, Wireless Communications and Mobile Computing, 2021, https://doi.org/10.1155/2021/1929077
[12] Ma, X.: Study on College English Online Teaching Model in Mixed Context Based on Genetic Algorithm and Neural Network Algorithm, Discrete Dynamics in Nature and Society, 2021, 2021 https://doi.org/10.1155/2021/8901469
[13] Qiao, W.: Evaluation Model of Flipped Classrooms Teaching Based on AHP and BP Neural Network. Journal of Wenzhou Vocational & Technical College, 2018,
[14] Su, X.: Design and Implementation of English Aided Instruction Platform Based on Intelligent Expert System; proceedings of the 2019 International Conference on Intelligent Transportation, Big Data & Smart City, 2019. https://doi.org/10.1109/ICITBS.2019.00102
[15] Such, F. P.; Rawal, A.; Lehman, J.; et al.: Generative Teaching Networks: Accelerating Neural Architecture Search by Learning to Generate Synthetic Training Data, F, 2019.
[16] Wahyono, I. D.; Asfani, K.; Mohamad, M. M.; et al.: A Novel Intelligent Learning for Teaching using Artificial Neural Network, proceedings of the 2020 4th International Conference on Vocational Education and Training, 2020. https://doi.org/10.1109/ICOVET50258.2020.9230093
[17] Wang, J.; Department, B. C.: Application of Optimized BP Network in College English Teaching Evaluation, Microcomputer Applications, 2018.
[18] Wang, Y.; Zhang, Y.; Dong, Z.; et al.: Neural Network-Based Approach for Evaluating College English Teaching Methodology, Mathematical Problems in Engineering, 2022, 2022.
[19] Yang, Z.; Kang, L.; Guo, Y.; et al.: Compact Real-valued Teaching-Learning Based Optimization with the Applications to Neural Network Training, Knowledge-Based Systems, 2018, S0950705118302995. https://doi.org/10.1016/j.knosys.2018.06.004
[20] Yin, W.: Study on the Construction of English ICAI Course Based on BP Neural Network Algorithm, Journal of Physics: Conference Series, 1992(2), 2021, 022185. https://doi.org/10.1088/1742-6596/1992/2/022185
[21] Zhang, L. H.: Research on English Pronunciation Recognition Based on Neural Network; proceedings of the International Conference on Intelligent Transportation, F, 2018. https://doi.org/10.1109/ICITBS.2018.00182
[22] Zhu, Y.; Zhou, D.: Online and Offline Teaching Effect of C Language in Colleges and Universities Based on Principal Component Analysis and Neural Network; proceedings of the CIPAE 2021: 2021 2nd International Conference on Computers, Information Processing and Advanced Education, F, 2021. https://doi.org/10.1145/3456887.3457054