The Framework of the Multi-Parameter Evaluation Index System for College Spoken English Based on Deep Learning Theory

Xiaofeng Wang

Abstract

When students in China learn English, speaking ability is regarded as the application and practical part of the comprehensive English competency and the most important improvement stage of college English learning. For this purpose, spoken English should be the absolute priority of the English departments in colleges and universities. However, based on the investigation of the comprehensive spoken English ability of college students in China, it is indicated that, restricted by the mother tongue environment and teaching conditions, the current college students’ spoken English proficiency is unprofessional and non-proficient over a long period of time. At present, China is in an important stage of the international development. As the backbone of the future of national society, college students are bound to bear the deeds of learning theory to construct the framework of the evaluation index system. The reference value of language ultimately simulates forward a deep learning theory that focuses on neuronal control parameter evaluation index system. This paper takes deep learning as the theoretical content, proceeds from pronunciation which is the weakest link in their spoken English, and makes a detailed analysis on its multi-parameter evaluation indexes. Furthermore, taking into account the differences between English phonetic pronunciation and Chinese phonetic transcription, this study focuses on constructing the neural network model, explaining the concept of the restricted Boltzmann machine and integrating the BP algorithm, and generates the differentiation of the multi-parameter evaluation index of college students’ spoken English. In the end, the article mainly builds on the needs of college students’ spoken English competence and establishes the reference framework of the evaluation index system.

Keywords: Deep Learning Theory, College Spoken English, Multi-Parameter Evaluation.

1. RESEARCH BACKGROUND

1.1 Literature Review

Currently, college English education obviously holds weaknesses. On the one hand, college English education in China always is always emphasized but in a moderate way. On the other hand, few talents in the field of English engage themselves in English education after the completion of their academic training. The above two aspects lead to a relatively backward situation of college English education resources in the long period of time in China. Furthermore, Chinese language is based on vowel pronunciation, and the pronunciation of many Chinese characters is rather difficult worldwide. For this reason, under the native language environment, it is extremely hard for Chinese undergraduates to practice spoken English while completely avoiding the influence of the native language environment. Therefore, the apparent phenomenon of Chinese students’ poor spoken English continues until now (Han, 2016). Fortunately, senior scholars in the international education community have put forward a deep learning theory that focuses on neuronal control, a method of process that builds on machine learning, adjusts the vector of the fixed value by collecting and analyzing the established conditions, and ultimately simulates the learning action of humans’ thinking under the same situation through the machine, so as to achieve the optimal adjustment of learning. When this method is applied to English learning, the spoken language quality of the test audio source can be evaluated and defined by means of the computer evaluation method. The reference value of spoken language quality involves a number of aspects, such as pitch, intonation, speech rate, etc., which are valuable contents that play a decisive role in the comprehensive score of spoken English (Xue, 2016). To this end, the machine-based neural analysis theory can maximize the practical value of the evaluation index system for college spoken English, and it is of practical value to introduce the deep learning theory to construct the framework of the multi-parameter evaluation index system.
1.2 Research Objective

This article takes Chinese college students as the main research object and studies their weakness in speaking ability. The actual embodiment of the study on the contemporary college students’ speaking ability indicates the obvious existing phenomenon of non-standard English pronunciation, intonation and other aspects of college students. For this reason, this paper applies the deep learning technology, takes BP algorithm and other computer technology as the main means, and constructs the framework of the multi-parameter evaluation index system for spoken language. The main objective is to adopt computer tools for detailed analysis and consequently achieve the main goal of improving college students’ spoken English competency. Specific means focus on the study of the multi-parameter quality evaluation on college students’ spoken language. However, considering that the spoken language evaluation indexes used by the education community are overly superficial and the contents of oral English assessment are insufficient, this paper combines the principle of objectivity, measures the weights of several arguments concerning the formation of spoken language that have not been thoroughly studied, and summarizes the framework and the model of the evaluation system based on the analysis result of the measurement process and feature parameters.

2. NEURAL NETWORK MODEL

2.1 Overview of Artificial Neural Network (ANN)

ANN is a type of artificial neural simulation system based on the artificial form. Its internal is connected by countless neurons that can simulate the reaction content of human nerves and participate in the artificial reproduction of thinking information, thinking process, forming process and other thinking logics in the form of computer (Li, 2016). ANN is a relatively mature simulation technology, which is applied in other analysis activities apart from teaching, such as business, construction and research, by virtue of its distribution structure, two-dimensional modelling and time sorting and other advantageous functions. At present, the scope of the neural network has spread to voice recognition and analysis module. Its functions for voice analysis include voice data compression, voice information synthesis, voice status, content identification, voice evaluation, etc. (Chen, 2014). Neurons (nodes/network) are the basic information processing units in a neural network structure, and its model structure can be divided into three basic elements, as illustrated in Figure 1.

![Neuron model diagram](image)

Figure 1. Neuron model diagram

(1) In the first set of connection, all the values that represent the quantity play an intense display role. Meanwhile, the value has two poles: positive and negative. The value of the positive pole represents activation and the value of the negative pole stands for inhibition. (2) The second element is the summator. Its actual function is to derive the sum of the corresponding synaptic weights (input signal to neuron pattern). (3) The third element is the incentive function, whose aims to limit the output amplitude of neurons (Jia, 2014). The specific restriction means are as follows. Function is first used to limit the range of of activity of the output signal. After the output signal is limited to a fixed value range, a relatively easily determined closed interval can be formed. At the same time, neurons model can also add the threshold value of another angle that is represented by δ. The actual role of the threshold is to achieve the increase and decrease of the excitation function through the basic setting of the positive and negative poles (Li, 2014). Specific neurons (in the artificial mode) can be expressed by the following formula:

\[ a = \sum_{i=1}^{r} b_i o_i \]

\[ q = m(a + \delta) \]
In the above formula, \( o_1, o_2, \ldots, o_r \) stands for \( r \) inputs of the individual neuron; \( o_i \) represents the connection strength (connection weight) of the \( i \)-th input; \( \delta \) is the bias of neurons (threshold); \( q \) is the output of neurons. On this basis, artificial neuron has a multi-input single-output nonlinear structure.

2.2 BP NEURAL Network

2.2.1. Overview of BP

BP neural network is a kind of algorithm based on the theory of neural network, also known as the error backward propagation algorithm. However, the algorithm was not widely used until 1986, so its actual effect is relatively more volatile (Wu, 2013). Figure 2 demonstrates the BP neural network model.

![BP neural network model](image)

**Figure 2. BP neural network model**

Based on the BP neural structure chart in Figure 2, BP neural is a multi-layer neural network model, and its layers can be divided into three categories: the input layer, the middle layer and the output layer. Each connection shows a regular state of connection, but there is no interconnection between neurons within the single and unified organizational structure (Yang, 2013). Meanwhile, the artificial learning process of BP neural network is explained from two aspects, namely the two directions of the propagation mode (forward and backward) and the error reduction process.

(1) Error Formation Process of the Forward and Backward Propagation Mode

As the existence of human-like neural movement model, neurons must be responsible for receiving external information and communicating in various neurons (Xu, 2015). The observation of the middle layer in a multi-layered structure indicates that the middle layer plays the role of information transfer, and the actual process of redirection can redefine the number of layers (hidden layers) according to the information requirements. Besides, the final communication of information needs to be completed by the hidden layer. After the degree elevation processing of the neural information in various layers, the status and the course of action can be classified as the forward mode of transmission. When the result value of information transmission remains the same in the process of signal transmission, correlation intervention would occur to the state of neurons in each layer, and the interference of the neurons in the lower level is the most obvious. In the intervention state, once the output layer fails to obtain the expected output value in the fixed phase, the information structure propagated by the neurons would inevitably enter the backward propagation mode of the error.

(2) Error Reduction Process

Error refers to the difference between the expected output value and the actual output value, and applies the
passing process of the output layer to continuously modify the value range among various levels in a descending manner. Meanwhile, the motion trail of the error value is reversed. Therefore, when the information is communicated in the forward direction and the error returns in the opposite direction, all layers can complete the learning process based on error correction by judging the difference (Zheng, 2011). The periodicity of the process can be selectively activated. In other words, the termination time of the learning process is the time when the error value is reduced to the optimal level, which is subject to the computer’s preset number of times in most cases.

2.2.2. BP Learning Algorithm

In order to achieve a more detailed analysis on the results, this section first discloses the structure of the learning process based on BP algorithm.

![Figure 3. BP algorithm learning flow chart](image)

2.3 Deep Learning Neural Network

2.3.1 The Basic Ideas of Deep Learning

The ideas of deep learning cover three points of view. First, all layers of the network should apply the unsupervised learning mode. Secondly, when the unsupervised learning deepens the training level, the corresponding training results should be recorded at the top of the layer. Thirdly, supervised learning can be used to adjust all layers, regardless of content differences (Wang, 2016). When the activity model of supervised learning is top-down, the neural network takes the initial parameters of the initial layer as the main reference point and gradually implements supervised learning into multiple layers of the model.

2.3.2 Overview of Restricted Boltzmann Machine (RBM)

RBM is a conceptual model image that can be formed by a stochastic neural network. The actual definition of
random refers to random neurons, rather than other random contents (Hao, 2016). Meanwhile, the RBM stands for only two forms of neural output models—active and inactive. Besides, 0 and 1 are basically used to represent its basic form, and the specific state value of the probability is determined based on the calculation of the internationally unified probabilistic method. Figure 4 indicates the conventional model diagram of the RBM.

![Figure 4. BM diagram](image)

3. MULTI-PARAMETER PRONUNCIATION QUALITY EVALUATION

3.1 Process of Pronunciation Quality Evaluation

Pronunciation quality evaluation on speech sound involves subjective evaluation and objective evaluation (Figure 5).

![Figure 5. Comparison of subjective evaluation and objective evaluation of speech sound quality](image)

3.2 Evaluation Indexes

3.2.1 Intonation

Compared with the practicability of spoken English, intonation evaluation attaches more attention to the integrity of sentences, the fluency of the speech and the basic faulty wording or formulation (Wu, 2013). This evaluation process employs the MFCC coefficient as the evaluation comparison coefficient and takes the speech recognition model based on deep belief theory as the recognition module of examination and evaluation. The computer technology inside the model is evaluated by comparing with the assessed voice through the preset correct pronunciation, so as to determine whether the input voice information conforms to the intonation standard of the spoken language assessment. The index system of tone evaluation is shown in Figure 6.
3.2.2 Speech Rate

Speech rate represents the speed of spoken English. The greatest role of spoken language is to communicate. When the participants involved in the test itself possess a certain foundation of grammar and vocabulary knowledge, a faster speed can become the bonus point of their spoken English competence (Hu, 2016). However, most students experience slow or tactile language proficiency, and different types of college students are completely different in speech rate under different test states, knowledge retention and emotional states. For this reason, the paper applies the formula of speech rate evaluation index (time-length ratio) to play some professional and protective role on the standardization of the evaluation indexes of speech rate measurement of college students. The formula is as follows:

$$\Delta = \frac{L_{std}}{L_{test}}$$

In this formula, $\Delta$ indicates the speed difference (in terms of duration) of the sentence to be measured and the reference sentence; $L_{std}$ represents the length of the reference sentence; $L_{test}$ stands for the duration of the sentence to be measured.

Figure 7 indicates the further comparison between $\Delta$ and the set speech rate threshold. It is noteworthy that, the duration of both short-term energy and short-term average zero-crossing rate applies the double threshold detection method for pretreatment, which can effectively exclude silent noise interference.

3.2.3 Rhythm

The rhythm of spoken English has obvious pronunciation features. First, in most cases, the more stressed
syllables exist in sentences, the slower the rhythm of English sentences is, the clearer the rhythm of pronunciation is (Zhao, 2010). Secondly, the pronunciation of the stressed syllables is inevitably vague and ephemeral. Furthermore, the length of the sentence depends on the specific number of stressed syllables. In case of excessive number of stressed syllables, the overall semantic meaning of the article is bound to be deep with an emphasis on the main functions. To this end, the evaluation contents include basic and loud spoken English rhythm patterns and clearly and understandable pronunciation. The specific evaluation mechanism is illustrated in Figure 8.

Figure 8. Rhythm evaluation

In order to achieve a softer evaluation, the short-term energy value of the voice signal is specifically extracted, so that assessors can adopt more targeted means to objectively evaluate the sense of rhythm of college students’ oral conversation, as indicated in the following formula:

$$ E_n = \sum_{m=-\infty}^{\infty} [s(n)\varphi(n-m)]^2 $$

(3)

3.2.4 Tone

The key point of tone is the high or low pitch of the human voice. When people express different contents in spoken English, the actual expression of the pitch is different. For this purpose, the actual meaning of the tone is to highlight the representation of human language emotion. English voice contains rising tone, rising-falling tone, falling tone, falling-rising tone and level tone. The analogy of the distinction between tones can recognize and score the accuracy of oral emotion expression of college students under the precondition of the preset reference content of the speech system. The specific identification contents can be carried out based on the following paragraph.

The first step is speech framing. The number of frames is an important unit of the subsequent tone analysis, where the time domain function method is adopted to extract the pitch of the sentence data. Secondly, the extremely unstable and abnormal frames in the pitch are excluded, and then the DTW algorithm is used to compare the fitting degree between the to-be-measured sentences and the standard sentences. In the end, scores are given to the to-be-measured contents of spoken English in view of the number of the frames and the fitting degree. Figure 9 demonstrates the specific evaluation index system.

Figure 9. Tone evaluation
4. CONCLUSION

This essay focuses on thoroughly understanding the content of the evaluation system for contemporary college students’ spoken English, which plays an absolutely impetus role in the structural realization of the evaluation system after an in-depth discussion of neural networks and its related algorithms. This paper analyzes the weak points of college students’ spoken English and limits the research points to pitch, speech rate, rhythm, intonation and other aspects, obtains the index evaluation and evaluation formation diagram based on the theory of deep learning after analyzing the specific content and formation of each weakness feature or the intervening factors, and draws the conclusion after repeated finalization of the content. Under the simulated condition of neural network, when the factor of undergraduates’ mastery of grammar and vocabulary is excluded, based on the in-depth learning theory, the advanced study on pitch, speech rate, rhythm, etc. can be a valuable promotion for college students’ improvement of oral English competence.

REFERENCES

Chen B. (2014). The dynamic evaluation of the intervention effect of cases of oral English ability of poor students of, in Huazhong Normal University, 9(13), 85-86.
Han Y.M. (2016). Output driven - A Study of the teaching mode of College English spoken under the input driven theory, Harbin Normal University, 12(03), 12-14.
Jia Q.S. (2014). the application of context in college oral English teaching, Huazhong Normal University, 14(22), 95-96.
Li G. (2014). A research on the application of formative assessment in the evaluation of college oral English learning effect, Sichuan foreign language university, 8(14), 75-76.
Li J. (2016). The application of output oriented method in college oral English teaching, North Central University, 25(08), 61-64.