Research Article

Research on Scoring of Business English Oral Training Based on Deep Neural Network

Wei Duan

School of International Studies, Chengdu Jincheng College, Chengdu 611731, Sichuan, China

Correspondence should be addressed to Wei Duan; duanwei@cdjcc.edu.cn

Received 16 May 2022; Revised 28 June 2022; Accepted 9 July 2022; Published 22 August 2022

Academic Editor: Lianhui Li

Copyright © 2022 Wei Duan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A scoring approach of business English oral training based on a deep neural network is put forward. Based on speech recognition technology, automatic correction technology, deep learning technology, machine learning technology, and the scoring data of business English oral training generated by the general computerized examination platform system, this paper uses the audio data of students’ oral expression questions collected by the existing general computerized examination platform, and studies artificial intelligence technology, deep learning technology, and machine learning technology to train the intelligent scoring system, it realizes the intelligent scoring of oral expression questions, which plays a positive role in language training guidance and correction. Finally, the accuracy of the intelligent scoring system is verified by the correlation between the score of the intelligent scoring system and the manual score of teachers. The higher the correlation with manual scoring, the higher the scoring accuracy of the intelligent scoring system.

1. Introduction

With the rise of AI [1–3] to the national strategic level, AI technology has become more and more popular, and AI + education is also the key research direction of educational technology [3–8]. How to combine artificial intelligence technology [9, 10], deep learning technology [11, 12], machine learning technology [13] and education is a problem worthy of research in the field of educational technology [12–15]. English is a very important subject in the field of education. If artificial intelligence and business English skill training can be combined, it will be of great help to improve students’ interest in learning English, correct students’ English pronunciation, evaluate students’ academic performance, and guide students’ learning. Artificial intelligence technology developed to a research climax as early as the 1990s. Recently, with the release of Google’s deep learning engine TensorFlow and artificial intelligence technology, it became more popular after alpha dog defeated Li Shishi, which is also expected [16]. TensorFlow brings convenience and ease of use to deep learning and training, and this game also shows that artificial intelligence can still perform well in this scenario [17–19]. Business English speaking training score plays an important role in Business English teaching. Under the background of artificial intelligence, it is very urgent and necessary to research the scoring of business English oral training based on a deep neural network.


With the rapid development of the Internet, “Internet + education” has become a trend [20–22]. The business English oral skills training system trains and evaluates English through online learning, so that students have made great progress in all aspects of English listening, speaking, reading, and writing. After years of accumulation, a large number of valuable teaching data are stored in the system. These source data provide a basis for our work. The current situation of oral business English teaching and training is deeply analyzed from four aspects: teaching objectives, content, implementation, and assessment.
(1) Teaching objectives
At present, the formulation of oral English curriculum standards for business English Majors in most colleges and universities is dominated by teaching materials and focused on knowledge. The listed objectives include pronunciation rules, intonation, word quantity, and expressions commonly used in daily life and social life.

(2) Teaching content
Due to the limitations of employing foreign teachers or professional teachers without a business background in China, the teaching content is mainly based on the English teaching content itself, and there is no subject matter of oral English practice for business application background.

(3) Teaching implementation
Due to the reasons of teachers and teaching resources, business English mostly adopts large class teaching, and the number of people in each class is between 35 and 45. In terms of curriculum, oral English is a professional basic course, which is generally set up in the first and second semesters, with 4 class hours per week. Interviews and questionnaires with teachers and professional students in many colleges and universities show that the current teaching methods are mainly teachers’ explanation and demonstration (accounting for about 40% of the classroom time) and student group practice (accounting for 50%), with additional personal practice (accounting for 10%). At the same time, the per capita time of students’ special oral practice after the class is less than 20 minutes per week.

(4) Teaching evaluation
In Teacher-centered courses, students perform less alone and lack quantitative process records. At the end of the semester, the final assessment will be conducted only in the form of one-to-one oral daily dialogue, question and answer, and monologue.

According to the current situation of oral English teaching and training, combined with the training needs of business English talents, this paper analyzes the main problems existing in the current oral English teaching and training, including (1) the application of teaching objectives is not high; (2) the teaching content is not contemporary enough; (3) lack of pertinence in teaching implementation; (4) the process of teaching evaluation is not strong.

3. Design of Intelligent Scoring System
Oral Business English training includes two parts, English learning and English examination. English learning refers to that students can use the system to conduct self-test and practice relevant English topics when learning English at ordinary times. English test is organized by teachers. Students are evaluated before, during, and at the end of each semester. After the examination, the students will participate in the questionnaire, which includes learning interests, whether they pass CET-4 or CET-6, the difficulty of the topic, their satisfaction with the examination system, etc. The oral business English training system will collect students’ test data, in which the objective question system can directly give scores through the standard answers set in advance, and the subjective questions need to be corrected manually by teachers. This puts forward higher requirements for teachers’ energy. The purpose of this study is to save teachers’ energy and make subjective questions score automatically.

The topic types of business English oral training include single choice, multiple choice, cloze, oral expression, and English writing. Among them, subjective questions include oral expression questions and English writing questions. English writing can be corrected according to the automatic correction system. The intelligent correction of oral expression questions realizes the automatic scoring of the business English oral training system.

Through the interview with English teachers, the needs for an intelligent scoring system are determined. There are two main aspects of demand as follows:

(1) Accuracy
Since it is automatic scoring, if the accuracy of scoring is not high enough to meet the expectations of teachers, the scoring work will lose its significance. The total score of oral expression questions is 10 points. Compared with the manual score of teachers, the error of no more than 1 point is acceptable and the total score of the whole English test is 100, so the error for the total English score is less than 1%, which can meet the expectations of teachers.

(2) Time consuming
If the time of the intelligent scoring system is too long, which is longer than the time of teachers’ manual correction system, it will consume teachers’ time longer and lose the significance of saving teachers’ time and energy. It takes no more than one minute to batch correct the oral expression questions of about 30 students in a class, which is the ideal automatic correction time for teachers.

3.1. Data Preparation Design. Based on the above demand analysis, the overall design of the intelligent scoring system is shown in Figure 1.

First, the audio data is processed. The corresponding features are extracted from it, then input into the neural network or linear regression model, and finally, the score is output. The language used for all processes is mainly python.

The python language has now become the mainstream language for machine learning programming. This language is more superficial than Java, C++, and C#. It also acts as a “glue” language to “glue” together functional modules written in other languages. Python has a rich software library, which is very helpful for machine learning. NumPy is an essential library in python, which provides some
commonly used functions and standard libraries for data calculation. For example, operations related to matrices and arrays can be done using NumPy, and the code is simple and easy to maintain.

3.1.1. Data Processing. The data processing is divided into two parts. One part is processing the oral expression questions answers (audio data), and the other part is processing the data in the system (including student learning data and test data).

(1) Audio data processing
First, the audio data is recognized by speech recognition, and then the text is corrected according to the requirements of composition correction according to the automatic correction scheme. Natural language processing includes text segmentation, sentence segmentation, construction of dictionaries, and word frequency statistics. The processed data is then subjected to data cleaning in preparation for the subsequent feature extraction. The flowchart of audio data processing is shown in Figure 2.

The flow of natural language processing is shown in Figure 3. The text recognized by the audio is firstly segmented and sentenced through speech recognition, the function words that have no real meaning, such as a and an, are removed, and only actual words such as nouns and adjectives are left. Then, a dictionary is constructed to count the frequency of each word.

(2) System data processing
The system data processing of the college English skill training system is divided into two parts as follows: learning data processing and examination data processing. After the data is extracted, the data needs to be cleaned because the extracted data will have a lot of dirty data and empty data, and these data need to be cleaned so as not to affect the results of subsequent model training. The system data processing is shown in Figure 4.

3.1.2. Feature Extraction. After the feature is extracted, too many features will lead to a complex model. Moreover, some features have little correlation with the target variable, so it is necessary to use the methods of feature transformation and feature selection to extract features.

Feature transformation methods are divided into normalization, discretization, and dimensionality reduction. Feature selection includes filtering, encapsulation, and integration. The definition of feature selection is to select a suitable feature set from many features, making the model make the evaluation index higher in the evaluation stage; that is, the model is more accurate, and the quality is better.

(a) The first is the filtering method. The filtering method can assign a weight to the feature through some common statistical methods, such as the chi-square test, $T$-test, information entropy and information gain, correlation coefficient, and covariance. This weight means that the more significant the correlation between the feature and the output result, the greater the impact on the result; the smaller the correlation, the less the effect on the result.

(b) The encapsulation method refers to selecting different feature subsets among many features. Some features can be excluded according to human experience and combined to achieve different effects. Then, according to different features, the effect to be predicted, the features of each combination of each group are evaluated, and the feature with the best effect is selected. In this way, a subset of features with
better effects can be selected; that is, some irrelevant features are excluded, or features that have less influence on the target result.

(c) The inheritance method refers to the premise that the model has been trained. Then, the data information is trained through this model to learn the characteristics that make the model accuracy the best. For example, the logistic regression model in the machine learning model can determine the weight of each input feature according to the model to determine the importance of the impact of each feature on the target variable; in this way, the importance of the features can be obtained, and the features can be filtered. The method of feature processing is shown in Figure 5.

The design of the correlation calculation model is shown in Figure 6.

3.2. Intelligent Scoring System Design. The dataset was randomly divided into training and test sets using Python’s random method. The training set accounted for 80%, with a total of 3,122 samples, and the test set accounted for 20%, with a total of 781 samples.

3.2.1. Linear Regression Model Design. According to the principle and usage scenarios of linear regression, experiments were carried out using the linear regression model [23–26] as shown in Figure 7.

3.2.2. Deep Neural Network Model Design. There are many mainstream deep learning open source tools on the market [27–29]. TensorFlow was developed by Google and has been widely used in recent years. On GitHub, you can see that TensorFlow has considerable attention and collection. Other deep learning tools such as Caffe and Torch do not have as much attention as TensorFlow, which shows that TensorFlow plays a pivotal role in the minds of deep learning developers and researchers. Many researchers study the training effect of these deep learning tools. Generally speaking, it is not objective. Each deep learning tool has its advantages and disadvantages. The specific situation needs to be analyzed in detail. This research mainly uses TensorFlow to train the neural network model.

The name of TensorFlow has already explained its two most essential components, Tensor and Flow. These two words are translated into Chinese as tensor and flow, respectively, which are explained from the perspective of the data model and calculation model. Tensor represents multidimensional data, that is, the data model. Flow refers to the
Flow computing performed by tensor Tensor, which means the computing model. Each calculation in TensorFlow is a node on the graph, and the edges between nodes describe the dependencies between nodes.

In Python, “import TensorFlow as tf” is generally used to load TensorFlow. If you can use TensorFlow more easily in programming. The TensorFlow program is divided into two parts, the first part needs to define all the calculations in the calculation graph, and the second part is to perform calculations.

There are many ways to install TensorFlow, which can be installed using Docker, installing using Pip, or compiled from source code. TensorFlow is divided into CPU version and GPU version. I installed the CPU version; the method used is Pip installation. The following is a detailed introduction to my installation method.

Pip is a small tool for Python that can easily and conveniently install and manage Python packages. Install TensorFlow using Pip - a three-step process. The first step is to install Pip. Select the appropriate version of Pip for your operating system to install. The second step is to find the proper installation package for TensorFlow. Because TensorFlow is divided into CPU and GPU versions, the GPU version requires the computer to support CUDA (Compute Unified Device Archive, a computing platform); I installed the CPU version. Currently, GPU installation still has specific restrictions on the environment. After the first two steps are completed, you can proceed to the third step, install TensorFlow through Pip. During installation, the commands installed will vary slightly depending on the version of Python. The steps of the deep neural network model are shown in Figure 8.

The steps of the deep neural network model (Figure 8) are as follows: input the training data into the neural network, and then use the neural network to train an intelligent scoring model continuously, and then input the test data into the intelligent scoring model to obtain the preliminary results of the intelligent scoring, compare this result with the teacher’s score, and then feedback and optimize the model, improve the model, and finally determine the intelligent scoring model; then input the test data into the intelligent scoring model and finally determine the score and then output it.

3.2.3. Fusion Design of Linear Regression Model and Deep Neural Network Model. The linear regression model and the deep neural network model are respectively assigned a weight and then summed to calculate the fused value. The final weighted prediction results in the experiment are \( p = Lr\_model \times \alpha + dnn\_model \times \beta, \alpha + \beta = 1 \), of which \( Lr\_model \) and \( dnn\_model \) are the results of the linear regression model and the deep neural network model, respectively. The values of \( \alpha \) and \( \beta \) change dynamically during each training and are related to relevant parameters and datasets.
3.3. Experimental Results of the Intelligent Scoring System

3.3.1. Model Evaluation. The weight value results of the linear regression model are shown in Table 1.

The result of the weight value of the deep network is shown in Table 2.

The weight value results after the linear regression model-neural network model fusion are shown in Table 3.

The dataset is randomly divided into a training set and test set, of which the training set accounts for 80%, with a total of 3,122 samples, and the test set accounts for 20%, with a total of 781 samples. The random sampling is divided into five categories, and the five categories are marked as A, B, C, D, and E, respectively.

There are three models as follows: the linear regression model (lr), the deep neural network model (DNN), and the model after the fusion of the two models (lr_DNN). We calculate the respective Recall and Precision to evaluate the model; the results are shown in Table 4. It can be seen from the table that the Recall and Precision after the fusion of the linear regression model and the deep neural network model are significantly better than the deep neural network model and the linear regression model.

3.3.2. Evaluation of Scoring Accuracy. The dataset was randomly divided into a training set and test set, of which the training set accounted for 80%, with a total of 3,122 samples, and the test set accounted for 20%, with a total of 781 samples. The method for judging whether the score is accurate or not is shown in Figure 9.

Among the 781 samples, 711 samples were predicted correctly, and 70 samples were predicted wrong. Therefore, the calculated accuracy rate is 91.04%, and the formula for calculating the accuracy rate is as follows:

\[
\text{Accuracy rate} = \frac{\text{accurate number of scores}}{\text{total number of scores}} \times 100\% = 91.04\%
\]

The main reason for the accuracy of the predicted samples is that the established intelligent scoring model can objectively describe the students' verbal expression ability, and teachers' scores are relatively concentrated, generally ranging from 7 to 9 (10-point scale). The main reason for the inaccuracy of the predicted samples is that the usual grades of some students cannot reflect their learning situation. For
example, student A is not interested in English but has a higher score on the oral English expression question. It is usual for the prediction results to have a certain error. Due to the diversity of samples, it is acceptable for the accuracy of a certain model to have the error within a reasonable range.

4. Conclusions

With the deepening of China’s opening to the outside world, business English majors put forward higher requirements for oral application ability. Making full use of various educational information technologies to promote the normalization, informatization, and networking of oral teaching and training is an important measure to improve the oral application ability of business applications. Based on the analysis of the current situation of applied oral English teaching and training, this paper combs the prominent problems in the current teaching, such as low application of objectives, insufficient timeliness of the content, insufficient pertinence of implementation, and weak evaluation process, and puts forward a scoring method of business English oral English training based on deep neural network.

This paper explores a new teaching mode. Focusing on business application, this model promotes the transformation of teaching objectives from knowledge to ability, introduces scene application and other rich teaching contents, promotes the implementation of “online and offline” hybrid teaching, and implements curriculum formative assessment and evaluation, so as to promote student-centered teaching reform and improve students‘ oral application ability. After one academic year’s pilot application in many business English application courses, the new model has better results in classroom effect, student investment, performance assessment, innovative practice, and practice.

Although it is based on the student’s oral test data in the business English oral skill training system for speech recognition, automatic correction, machine learning modeling, and neural network modeling, there will still be many problems in the actual operation process. The limitations of this study are as follows: inaccurate speech recognition results, inaccurate oral score, and limitations of the deep learning model.

1. Although speech recognition has been claimed to have an accuracy of more than 95% in the official documents, in the actual process, the output results of speech recognition will be affected by the examination environment, the influence of surrounding students, and the noise nearby. If the speech recognition result is not accurate, it will have a certain impact on the oral score.

2. At present, most oral evaluation still adopts the subjective evaluation of teachers, and the oral evaluation itself is a difficult problem to overcome. Different teachers have different scoring standards, and there are differences in the scores between teachers. Most of them take the form of average scores to calculate the student’s scores. Therefore, in the process of automatic recognition of oral scoring, there will be a difference between the system scoring standard and the teacher scoring standard. In short, it is difficult to quantify teachers’ scoring standards, and it is difficult to reach an agreement between the scoring standards of the automatic correction system and teachers’ scoring standards.

3. The following research can optimize the scoring model and select the most appropriate model with the in-depth study of neural networks. Moreover, with the gradual increase of the computing power of CPU and GPU, the model can be more complex. In this way, even if the complexity of the model is increased and the amount of calculation becomes larger, the results can be calculated quickly.

### Table 4: Evaluation of three models.

<table>
<thead>
<tr>
<th></th>
<th>lr_DNN (%)</th>
<th>DNN (%)</th>
<th>lr (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>A</td>
<td>91.23</td>
<td>93.52</td>
<td>89.12</td>
</tr>
<tr>
<td>B</td>
<td>92.43</td>
<td>92.41</td>
<td>87.35</td>
</tr>
<tr>
<td>C</td>
<td>90.69</td>
<td>91.90</td>
<td>89.31</td>
</tr>
<tr>
<td>D</td>
<td>92.34</td>
<td>92.28</td>
<td>88.81</td>
</tr>
<tr>
<td>E</td>
<td>91.83</td>
<td>90.31</td>
<td>91.04</td>
</tr>
</tbody>
</table>

![Figure 9: Evaluation of scoring accuracy.](image)
Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


