Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 8389469, 8 pages
https://doi.org/10.1155/2021/8389469

Research Article

College Oral English Teaching Reform Driven by Big Data and Deep Neural Network Technology

Hui Liu

Department of Foreign Language, Zhanjiang University of Science and Technology, Zhanjiang 524000, China

Correspondence should be addressed to Hui Liu; liuhui19811117@126.com

Received 5 August 2021; Revised 16 August 2021; Accepted 18 August 2021; Published 18 September 2021

Academic Editor: Yuanpeng Zhang

The ultimate goal of English teaching is to cultivate the students' ability to communicate information in English, master good language learning methods, and become independent language learners and users. Therefore, successful English language teaching needs to be achieved through language communication training between teachers and students and between students. This article investigates the importance of promoting the reform of oral English teaching in China's English teaching environment. We believe that to promote the reform of oral English teaching, an oral teaching environment must be available. However, the current common problem in oral English teaching in colleges and universities is that the spoken conversation objects are not standard enough, or there is no person who can talk to. Therefore, an intelligent spoken dialogue system based on big data and neural network technology is particularly important, and the quality of dialogue depends on accurate spoken speech evaluation. We first extracted six features of pronunciation quality, fluency, content richness, topic relevance, grammar, and vocabulary richness. Secondly, we propose an evaluation model that connects specific TDNN layers in a feedforward manner, using the feature representation of target words in different TDNN layers, which can obtain richer context information and greatly reduce the amount of model parameters. Finally, we conducted a simulation experiment. The experimental results show that the proposed model is accurate in evaluating spoken English and can effectively assist the reform of spoken English teaching in colleges and universities, and its performance is better than SVM by 9.2%.

1. Introduction

In recent years, the application of information technology [1–3] in the field of education is more and more extensive. In oral English teaching, due to the increasing popularity of English teaching in China, traditional language teaching methods [4] can no longer meet people's needs, which is more and more obvious in colleges and universities. In this context, the computer-aided language learning system [5–7] based on big data and neural network [8–10] has become the focus of research. It can take the place of teachers for students' examination answers, classroom homework automation correction, so that teachers from repeated and time-consuming correction work out of the liberation. At present, such an automatic correcting system [11] has been able to achieve almost complete accuracy in the objective task. This form of question is typically used for multiple choice, fill-in, and other types of questions, and for writing and oral questions, automatic correction is still a research topic that needs to be broken through. The oral questions are separated into two types: one is retelling, reading, and reciting known content, and the other is "open speaking," in which the exam taker provides free rein to specific questions or subjects, through the examinee pronunciation and standard pronunciation for speech level comparative analysis, with the development of speech recognition technology [12–14].

A full evaluation of candidates' answers from several dimensions, including oral fluency, rhythm, intonation, vocabulary richness, and semantics, is also necessary for open spoken English, in addition to "pronunciation correctness." For a long time, open spoken English scoring [15, 16] and evaluation technology research has not yielded significant results. With the advancement of machine learning

technology [17–19], some researchers began to investigate how to use it to automatic speech evaluation, resulting in the development of the well-known SpeechRater automatic scoring system. The system combines feature engineering and machine learning algorithms [20] to achieve automatic scoring of open spoken language, and it has also set off a wave of research in this field. Although there is still a certain gap between the scoring results of the system and the teachers' manual scoring, it provides a good research idea for later researchers. Today, new changes have taken place in the field of artificial intelligence; the most significant sign of which is the maturity of deep learning technology. Deep learning [21, 22] employs a multilayer network to do data characterization and uncover richer features. The correlation between machine and manual scores has substantially improved, and scoring errors have grown fewer and smaller after a large number of researchers applied this technology to the realization of the oral score scoring model. The open oral scoring method can now be used on a practical level thanks to deep learning technologies.

Based on the foregoing observations, we discovered that big data and deep neural network technology-driven college oral English teaching reforms [23, 24] have become a trend. Teachers will be able to dedicate more energy to actual teaching work as a result of this, and the quality of instruction will be improved. As a result, an intelligent oral dialogue system based on big data and neural network technology is critical, and the quality of discourse is dependent on precise oral evaluation. We first extracted six features of pronunciation quality, fluency, content richness, topic relevance, grammar, and vocabulary richness. Secondly, we propose an evaluation model that connects specific TDNN layers in a feedforward manner. Using the feature representations of target words in different TDNN layers, we can obtain richer context information and greatly reduce the amount of model parameters. Finally, we conducted a simulation experiment. The experimental results show that the proposed model is accurate in evaluating spoken English and can effectively assist the reform of spoken English teaching in colleges and universities.

The following are the main contributions points of this paper:

(1) This paper reforms the college oral English teaching based on big data and deep neural network technology and proposes a spoken language recognition model, which reduces the burden of teaching and improves the quality of teaching

(2) We propose an evaluation model that connects specific TDNN layers in a feedforward manner. Using the feature representations of target words in different TDNN layers, we can obtain richer context information and greatly reduce the amount of model parameters

(3) We carried out an experiment with simulation. The experimental results show that the proposed model is accurate in the assessment of spoken English and can help effectively reform language teaching at universities and colleges

The following is the general structure of the paper: The background is examined in Section 2. In Section 3, some details about the suggested algorithm's concepts and related submodules are presented. The experimental results are detailed in Section 4.

2. Background

As a medium for people to communicate with each other, the characteristics of spoken language are very convenient and concise. It is a main communication method for people to obtain information. Nowadays, people regard whether computers can understand the spoken language used in people's daily life as a research direction of artificial intelligence, and spoken language is the natural language used by people in daily life. The oral dialogue system is a tool for people to communicate with computers. The computer understands people's spoken language and makes corresponding answers. The spoken dialogue system has been widely used in the information query system. The main reason is that the price of manual customer service is more expensive than the dialogue system, and human resources are limited. The use of this system can greatly improve the efficiency of the system, so as to serve more people. The system can greatly facilitate people's daily life and, at the same time, improve work efficiency to a large extent.

Oral English comprehension research is critical for improving the effectiveness of the oral conversation system [25, 26]. Speech recognition, oral comprehension, dialogue management, text production, and speech synthesis are the most common modules. The spoken conversation system uses a speech recognition module to transform the user's voice into text, then converts each word in the text into a corresponding word vector, and finally classifies the entire sentence or each word so that the computer can extract the phrase's main semantics. The dialog management part analyzes the user's request to get the system's answer, and then, the text generation module generates texts based on the results of the dialog management. These texts are related to the time sequence. It can be seen that oral comprehension plays a key role in the performance of the oral dialogue system.

Whether a machine allows people to carry out related tasks through the use of spoken language is the criterion for judging whether the machine is truly "smart." The dialogue system conforms to the habit of humans using spoken language for interaction. Compared with traditional information acquisition methods, the dialogue system has great advantages. The user and the system can use multiple rounds of dialogue, such as inquiries, clarifications, and confirmations, to achieve the needs of information acquisition and emotional comfort in complex scenarios. With the continuous development of science and technology in the future, robots in related fields such as service, social networking, and industry will become new members of the future society, and human-machine dialogue technology is extremely critical for whether humans and machines can achieve "intelligence." With the development of artificial intelligence technology, the human-machine dialogue

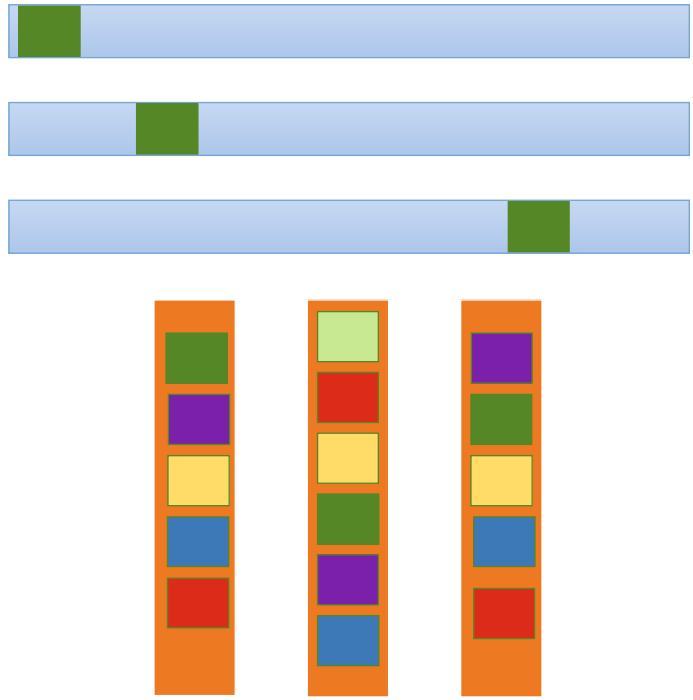

Figure 1: One-hot encoding and word embedding represent the word vector.

system will play an extremely important role in the future intelligent society.

3. Methodology

This section will elaborate on the college oral English teaching reform algorithm driven by big data and deep neural network technology. The core link is the evaluation of spoken English. This section will conduct a detailed analysis of natural language processing and neural networks.

3.1. Natural Language Processing. In the spoken language scoring system studied in this article, a neural network model needs to be used to score the candidates' spoken language content. The neural network cannot directly process text data, so it is necessary to convert the text into numerical data that the neural network can recognize. A commonly used method is one-hot encoding, which associates each word with a numeric vector of length N. The numeric vector corresponding to each word has only one element of 1, and the rest of the elements are 0. For example, the one-hot codes of the three words "me," "he," and "she" are [1,0,0], [0,1,0], and [0,0,1], respectively. Although this method is simple, it has the following two main disadvantages: (1) Since there is only one bit in each numeric vector to identify a word; if there are N words in the text, an N-dimensional vector needs to be used to encode it. Therefore, when the number of nonrepeated words in the text is large, the dimensionality of the vector will be large. At the same time, the number of nodes in the neural network will increase, and the calculation will become more complicated. (2) One-hot encoding, a simple encoding method, cannot describe the semantic relationship between words and thus cannot provide more information for subsequent neural network calculations. What a useful information. In order to solve the above problem, word embedding appears, which can use lower-dimensional vectors to represent words. At the same time, for words with similar meanings, their vector representations are also similar. As shown in Figure 1, the word vectors are represented by one-hot encoding and word embedding, respectively. It can be found that the latter can embed richer information into lower-dimensional vectors.

The network topology presented in Figure 2 is commonly employed when utilizing neural networks to handle natural language challenges. The word embedding layer is the network's first layer, and it turns the words in the input text into a word vector representation, such as a word with 20 letters. If the length of the word embedding vector is 50, the text of a word will become a 2-dimensional matrix of 2050 after the word embedding layer.

3.2. Recurrent Neural Network. The recurrent neural network (RNN) is a special artificial neural network, which is different from the feedforward neural network of the general structure. It is a neural network with internal loops. This structure enables information to circulate in the network, so unlike CNN and other networks, their output only considers the impact of the previous input and does not consider the impact of the input at other times. In RNN, the output is at every time linked not just to the input, but to the input in the preceding moment, just like the "memory" function is available in the network. Consequently, RNN is very suited for serial data processing, particularly text data. If the state is regarded at every moment as a layer of the feedforward neural network, then the cyclical network can be seen as a feedforward neural network that shares weight

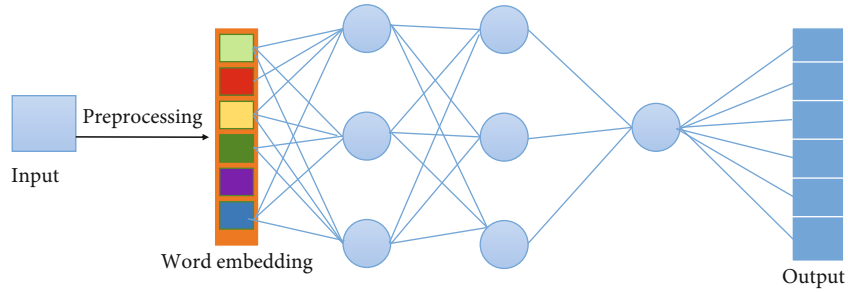

Figure 2: Schematic diagram of the neural network model architecture with word embedding layer.

in time. Figure 3 shows the expanded structure of a single-layer RNN network, where t is time, x_t represents the input at time t, h_t is the state of the hidden layer at time t, and o_t represents the output of the RNN network at time t; matrix W, V, U represents the weight matrix, and the calculation equations of h_t and o_t are as follows:

$$h_t = fuc(Uh_{t-1} + Wx_t + b), \tag{1}$$

$$o_t = Vh_t. \tag{2}$$

Although the traditional RNN network can process sequence data, it has a more serious problem: when the input sequence is long, there will be a problem of gradient disappearance during the error back propagation process, so that the network will eventually become unable to train. Therefore, the traditional RNN model is only suitable for processing short-sequence data. In order to solve the problem of insufficient "long-term memory" ability of the traditional RNN network, many researchers began to explore how to improve the model. Hochreiter and Schmidhuber proposed LSTM in 1997. This model solves the above problems well. The LSTM network introduces a new state c_t (also called a memory unit) internally for the circular transmission of information.

At each time t, c_t records the historical information up to the current time. The state h_t of the hidden layer and the state c_t of the memory unit. The calculation equation is as follows:

$$c_t^* = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right), \tag{3}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ c_t^*, \tag{4}$$

$$h_t = o_t \circ \tanh\left(c_t\right), \tag{5}$$

where f_t, i_t, and o_t are three gate controllers, and the gate control mechanism is a method to allow information to pass through selectively. The value range of the door controller is between 0 and 1, which means that a certain proportion of information is allowed to pass. $\circ$ represents the matrix dot product operation, c_{t-1} is the state of the memory unit at the previous moment, and c_t^* represents the candidate state of c_t.

3.3. TDNN Model. The TDNN algorithm is similar to the standard back spread algorithm in terms of training, and it is a quick algorithm. The TDNN is a multilayered network with abstract capacity at each layer and the ability to achieve input sequence in time. TDNN is time invariant, and the network learning process does not necessitate precise input data placement. The benefit of TDNN is that each layer's TDNN shares weights, making the model easier to train. The time range of the context of the sequence collected by TDNN becomes wider and wider as information flows to higher layers of TDNN. As a result, TDNN has a distinct edge in certain jobs where context information is critical.

In the phoneme recognition paper, a time delay neural network is first proposed. It is a neural feedback network of multilayered systems. The neural network time delay is a hierarchical neural feedforward network. The network can take the proximity data of the current frame into account and complete the function extraction of the current frame. The time delay neural network can be multilayered; each layer has a strong ability to abstract features; it can represent the relationship between features in time; with time invariance, it is more convenient to learn by sharing weights. In the process of learning, it is not required to carry out precise time positioning of the learned marks. The training method of the time delay neural network is the traditional back propagation algorithm.

The time and space complexity of TDNN is the same as that of the convolutional neural network. The time complexity of a single convolutional layer is

$$\text{Time} \sim O\left(M^2 \times K^2 \times C_{\text{in}} \times C_{\text{out}}\right), \tag{6}$$

where M is the side length of each convolution kernel's output feature map and K is the side length of each convolution kernel, i.e., the number of output channels of the network's upper layer. The output feature map area M^2 determines the time complexity of each convolutional layer, as shown in the formula, input C_{in}, output C_{out}, and the convolution kernel K^2. Three parameters determine the size of the output feature map: the size of the input matrix X, the size of the convolution kernel K, padding, and stride. The following is the calculation formula:

$$M = \frac{(X - K + 2 * \text{Padding})}{\text{Stride} + 1}. \tag{7}$$

The convolutional network's overall time complexity, D, is the number of convolutional layers in the network model; ℓ is the network's ℓth convolutional layer; and C is the

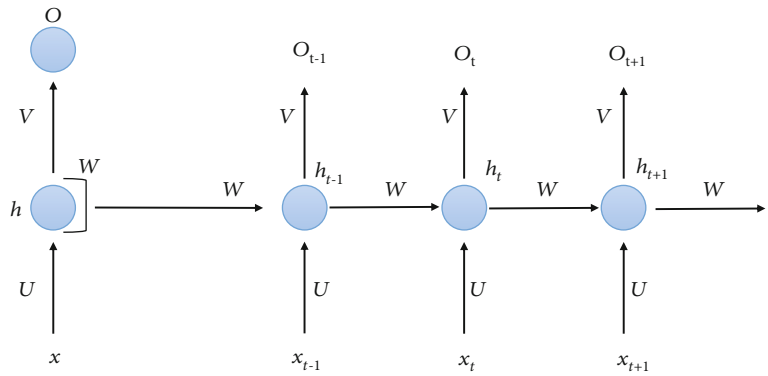

Figure 3: Schematic diagram of a single-layer RNN structure.

number of output channels of the ℓth convolution kernel, which is also the current convolutional layer. The number of output channels of the convolutional layer of the $\ell - -1$ layer is the input channel C for the ℓth convolutional layer. It can be seen that the convolutional network's time complexity is the sum of the time complexity of all convolutional layers.

$$\text{Time} \sim O\left(\sum_{\ell=1}^{D} M_\ell^2 \times K_\ell^2 \times C_{\ell\text{-}1} \times C_\ell\right). \tag{8}$$

The space complexity of the convolutional network is described as the parameter quantity of the model, which is expressed as the size of the model.

$$\text{Space} \sim O\left(\sum_{\ell}^{D} K_\ell^2 \times C_{\ell-1} \times C_\ell\right). \tag{9}$$

The size K of the convolution kernel, the number of channels C, and the number of network layers D of the model are the only variables that influence the model's space complexity. The size of the input data has no bearing on the complexity of space.

3.4. Our Model. On the basis of multilayer TDNN, this paper proposes a parallel structure of the TDNN network (as shown in Figure 4). Multiple layers of time-delayed neural networks can be stacked to obtain more contextual information, but this causes gradient explosion and gradient dispersion issues. In response to this issue, the residual convolutional neural network has performed well on image classification tasks, demonstrating that the residual structure can reduce gradient dispersion or explosion by using jump connections. In this paper, by quoting the residual structure, the number of network layers of the model can be deepened, and its performance on the task of image classification has been significantly improved. We compared the experimental results of multilayer time-delayed neural networks and discovered that increasing the number of layers does not improve the model's performance, but rather decreases it. As a result, we add a residual structure to the multilayer time delay neural network model in order to improve the model. The context information of the current word is well captured by TDNN, and the longer contextual information is captured by stacked multilayered delayed neural networks. The residual structure can combine the low-level network's features with the high-level network's features to improve feature representation.

In the word embedding layer, in many NLP tasks, a common practice is to construct a dictionary of words in the dataset. Each id of the dictionary corresponds to a specific word, and the id of each input word is converted to a D -dimensional real value vector, which is called a word vector. We splice consecutive W word vectors as the representation of the word vector of the current word, and W is the size of the spliced word window. w is the offset of the spliced context, $W = 2w + 1$. During splicing, we fill in the embedding representation of the filling symbol if there are not enough words before or after the target word. As a result, the input at t in the sequence is

$$E_t = [e_{t-w}, \cdots, e_{t-1}, e_t, e_{t+1}, \cdots, e_{t+w}]. \tag{10}$$

The vector of the entire sentence can be represented as an input matrix $s \in R^{N \times W \times D}$ for a sentence containing N single words.

$$E_t = [e_{t-w}, \cdots, e_{t-1}, e_t, e_{t+1}, \cdots, e_{t+w}]. \tag{11}$$

In addition, the normalized probability distribution is obtained using the SoftMax activation function at the network's last layer, and the cross-entropy-based objective function used in this article is as follows:

$$L = -\frac{1}{N}\sum_{t=1}^{N}\sum_{c=1}^{C} y_{t,c} \log y'_{t,c}, \tag{12}$$

where $y'_{t,c}$ is the true probability distribution of the cth label of the tth word in the sample and $y_{t,c}$ is the probability of the cth label of the tth word. N denotes the number of words in the sample, while C denotes the number of semantic categories.

When it comes to predicting semantic labels for oral English comprehension, the target word's background

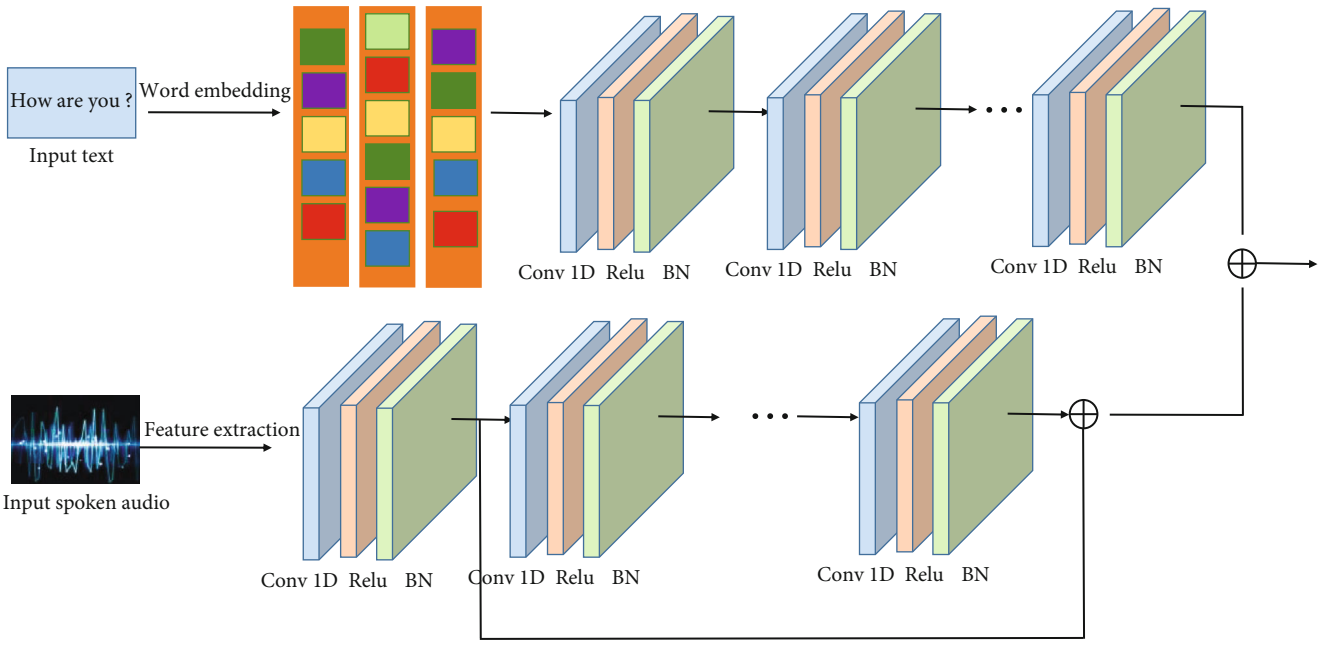

Figure 4: Schematic diagram of the proposed spoken English evaluation model.

information is crucial. It is vital to collect the target word's context information because the same term can have multiple labels in different circumstances. The delay offset of each layer of a delay neural network can be adjusted, changing the range of background information that the extracted features can acquire. The proposed model can successfully extract the target word's contextual information through delay migration.

4. Experiments and Results

4.1. Experimental Environment. The experiment uses the Linux operating system Centos 6.5 version and uses the deep learning tool Pytorch 0.4 version in the Python 3.6.5 environment under Anaconda. Pytorch is widely used in the field of deep learning; its code is simple and easy to write and can run on CPU and GPU. All the networks in this article are implemented using the Pytorch deep learning toolkit under the Python 3.6.5 environment. The learning rate is 0.01, and we batch processed 1,000 samples each time.

4.2. Dataset. The data used in this article is derived from the oral examination of a situational English course at a university. We extracted 650 test takers' answers to the same open-ended oral question from the examination data from 2012 to 2018 (each recording is about 60 seconds or so)) and the teacher's manual scoring data (10-point system, including decimals). The system designed in this paper needs to score the spoken language pronunciation and the spoken language content separately, so we also asked the teacher to score the examinees' spoken language separately from these two aspects. In addition, before training and testing the model, we also need to convert the recording format. The recording files collected from the oral test are all in mp3 format, and the audio attributes are 16 bit and 16 kHz sampling rate. FFmpeg is an open source tool that specializes in processing video and audio streams. We use this tool to convert recordings in mp3 format to pcm format. Finally, we divide the entire dataset into two groups for training and testing: the "training set" contains 500 pieces of data, and the "test set" contains 150 pieces of data.

4.3. Evaluation Index. The evaluation criteria used in this article are F1, precision, and recall, and the calculation equations are as follows:

$$\mathrm{F1} = \frac{2 \times P \times R}{P + R} \times 100\%, \tag{13}$$

$$P = \frac{TP}{P_{\mathrm{all}}} \times 100\%, \tag{14}$$

$$\mathrm{Recall} = \frac{TP}{T_{\mathrm{all}}}. \tag{15}$$

4.4. Experimental Results. In order to prove the effectiveness of the proposed algorithm, we compared SVM and BP neural network, and the comparative experimental results are shown in Table 1.

As can be seen from Table 1, the proposed algorithm achieves competitive results. Compared with SVM and BP algorithm, the accuracy of this paper is improved by 9.2% and 5.4%, respectively, and that of F1 is improved by 7.5% and 6.4%, respectively, indicating that the proposed algorithm is effective.

4.5. Ablation Experiments. The slot value filling results can be seen from Table 2 using the neural network of multilayer delays. The experimental delay D settings, the number of convolutionary kernels, and the size of the word concatenation window W are identical to those of the single-layer

Table 1: Compare experimental results.

Methods	*P*	*R*	F1
SVM	0.8102	0.8207	0.8514
BP	0.8435	0.8525	0.8624
Ours	0.8925	0.9017	0.9214

Table 2: Results of ablation experiments.

Layers	*P*	*R*	F1
1	0.8425	0.8936	0.9125
2	0.8169	0.8814	0.9111
3	0.8925	0.9017	0.9214
4	0.8755	0.8936	0.9125
5	0.8852	0.8745	0.9005

TDNN model to deliver best performance. To observe the effect of the TDNN network layers as a result of slot filling, just stack TDNN layers from 1 to 5. As shown in the table, the F1, with 3-layer TDNN, amounted to 92.14%. The F1 value obviously drops to 90.05% by increasing the number of network layers from 3 to 5. The experimental results show that the number of layers of TDNN could simply increase and that dependency relationship cannot be captured more effectively. It is not only difficult to train the profound network structure model but also can cause gradient problems.

5. Conclusion

The importance of promoting the reform of spoken English teaching in our country's English teaching environment is discussed in this article. An oral teaching environment is seen to be important to facilitate the reform of spoken English teaching. However, a widespread problem in oral English instruction in colleges and universities is that the spoken conversation objects are not standardized enough, or there are no individuals with whom to converse. As a result, an intelligent oral dialogue system based on big data and neural network technology is critical, and the quality of discourse is dependent on precise oral evaluation. We began by identifying six characteristics: quality of pronunciation, fluency, content richness, issue relevance, grammar, and vocabulary richness. Second, we present a feedforward evaluation approach that connects certain TDNN layers. We can obtain richer context information and greatly reduce the number of model parameters by using feature representations of target words in different TDNN layers. Finally, we carried out a simulation test. The findings of the experiments suggest that the proposed model is accurate in evaluating spoken English and can effectively assist the reform of spoken English teaching in colleges and universities, and its performance is better than SVM by 9.2%.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This research was supported by the Guangdong higher education reform project "Research on the Innovation Mode of College English Micro-class under the Era of Internet Plus" (Guangdong Higher Teaching Document (2018) No. 180); Research on the Cultivation of College Students' English Autonomous Learning Ability and Strategies Under the Background of Educational Informatization (DDXK202101ZXM); Advanced English: Viewing, Listening and Speaking I"—"Curriculum Ideological and Political" Demonstration Course (PPJH202118YLKC); Research on Ideological and Political Teaching Design and Practice of " Advanced English: Viewing, Listening and Speaking" Course from the Perspective of "Three- Complete Education" (ZLGC202057); and College English—Online and Offline Mixed First-class Course (PPJH202104YLKC).

References

[1] S. A. Asongu and N. M. Odhiambo, "Basic formal education quality, information technology, and inclusive human development in sub-Saharan Africa," *Sustainable Development*, vol. 27, no. 3, pp. 419–428, 2019.

[2] S. F. Alfalah, "Perceptions toward adopting virtual reality as a teaching aid in information technology," *Education and Information Technologies*, vol. 23, no. 6, pp. 2633–2653, 2018.

[3] P. Paul, A. Bhuimali, and P. S. Aithal, "Indian higher education: with slant to information technology—a fundamental overview," *International Journal on Recent Researches In Science, Engineering & Technology*, vol. 5, no. 11, pp. 31–50, 2017.

[4] V. Bošković Marković, "Traditional language teaching versus ICT oriented classroom," in *In Sinteza 2019-international scientific conference on information technology and data related research*, pp. 627–632, Singidunum university, 2019.

[5] L. Qiu, "Computer-aided English teaching platform based on secure shell framework," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 16, 2019.

[6] U. Afini, C. Supriyanto, and R. A. Nugroho, "The development of Indonesian POS tagging system for computer-aided independent language learning," *International Journal of Emerging Technologies in Learning*, vol. 12, no. 11, 2017.

[7] M. K. Ahmed, "Multimedia aided language teaching: an ideal pedagogy in the English language teaching of Bangladesh," *American International Journal of Social Science Research*, vol. 3, no. 1, pp. 39–47, 2018.

[8] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.

[9] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, p. 1129, 2021.

[10] L. Geng, "Evaluation model of college English multimedia teaching effect based on deep convolutional neural networks," *Mobile Information Systems*, vol. 2021, Article ID 1874584, 8 pages, 2021.

[11] M. Willsey, A. P. Stephenson, C. Takahashi et al., "Puddle: a dynamic, error-correcting, full-stack microfluidics platform," in *In Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 183–197, Providence, 2019, April.

[12] M. W. Ok, K. Rao, J. Pennington, and P. R. Ulloa, "Speech recognition technology for writing: usage patterns and perceptions of students with high incidence disabilities," *Journal of Special Education Technology*, 2020, in press.

[13] H. Isyanto, A. S. Arifin, and M. Suryanegara, "Performance of smart personal assistant applications based on speech recognition technology using IoT-based voice commands," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 640–645, Korea, 2020, October.

[14] E. Armas, R. Álvarez, and G. Romero, "Aids based on speech recognition technology for people with motor disabilities and reduced mobility," *Revista Politécnica*, vol. 43, no. 1, pp. 15–22, 2019.

[15] H. Chung, Y. K. Lee, S. J. Lee, and J. G. Park, "Spoken English fluency scoring using convolutional neural networks," in *In 2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA)*, pp. 1–6, korea, 2017, November.

[16] K. Evanini, M. Mulholland, E. Tsuprun, and Y. Qian, *Using an automated content scoring system for spoken CALL responses: the ETS submission for the Spoken CALL Challenge. In Proceedings of the Seventh SLaTE Workshop*, Stockholm, Sweden, 2017.

[17] L. Hussain, I. A. Awan, W. Aziz et al., "Detecting congestive heart failure by extracting multimodal features and employing machine learning techniques," *BioMed research international*, vol. 2020, 19 pages, 2020.

[18] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, pp. 1–1, 2021, in press.

[19] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia.*, p. 1, 2021, in press.

[20] M. Li, G. Zhou, W. Cai et al., "MRDA-MGFSNet: network based on a multi-rate dilated attention mechanism and multi-granularity feature sharer for image-based butterflies fine-grained classification," *Symmetry*, vol. 13, no. 8, p. 1351, 2021.

[21] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Multiscale graph sample and aggregate network with context-aware learning for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4561–4572, 2021.

[22] Z. Huang, P. Zhang, R. Liu, and D. Li, "Immature apple detection method based on improved Yolov3," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 9–13, 2021.

[23] R. A. Rashid, S. B. Abdul Rahman, and K. Yunus, "Reforms in the policy of English language teaching in Malaysia," *Policy Futures in Education*, vol. 15, no. 1, pp. 100–112, 2017.

[24] M. Y. Amin, "English language teaching methods and reforms in English curriculum in Iraq; an overview," *Journal of University of Human Development*, vol. 3, no. 3, pp. 578–583, 2017.

[25] E. J. Hwang, B. A. Macdonald, and H. S. Ahn, "End-to-end dialogue system with multi languages for hospital receptionist robot," in *In 2019 16th International Conference on Ubiquitous Robots (UR)*, pp. 278–283, Korea, 2019, June.

[26] O. W. Kwon, Y. K. Kim, and Y. Lee, "Task graph based task-oriented dialogue system using dialogue map for second language learning," in *Future-Proof CALL: Language Learning as Exploration and Encounters – Short Papers from EUROCALL 2018*, pp. 153–159, Research-publishing.net, 2018.