Application of Deep Learning Algorithms in Junior High School English Listening Training

Huimin Duan

Xi' an International Studies University, Xi'an, China

619948817@qq.com

Abstract. To enhance the effectiveness of English listening training for junior high school students, this article explores the application of deep learning algorithms in listening training. Centered on deep learning models, the study analyzes their advantages in processing speech data, generating personalized training programs, and optimizing training outcomes. The results demonstrate that deep learning algorithms significantly improve listening comprehension, personalized teaching, and large-scale data processing, providing critical support for the intelligent development of English education.

Keywords: Deep learning algorithms; Junior high school English; Listening training

1 Introduction

In junior high school English education, the importance of listening training is self-evident. However, traditional training methods show evident shortcomings in personalization, efficiency, and large-scale data processing. With the rapid development of deep learning algorithms, exploring their application in English listening training becomes an inevitable choice, given their superior performance in big data processing and speech recognition. By optimizing algorithm models, the goal is to achieve precise analysis of students' listening abilities and personalized guidance, thereby significantly enhancing learning outcomes and bringing new technological innovations and practical significance to the field of English education.

2 Advantages of Deep Learning Algorithms in English Listening Training

Deep learning algorithms can efficiently process and analyze massive amounts of listening data through neural network models, offering more automation and intelligence than traditional methods. They can quickly generate personalized training programs in a short time. Secondly, the high accuracy of the algorithm is evident in its ability to precisely identify and predict students' weaknesses in listening training, providing

C. Lin et al. (eds.), *Proceedings of the 2024 9th International Conference on Modern Management, Education and Social Sciences (MMET 2024)*, Advances in Social Science, Education and Humanities Research 880,
https://doi.org/10.2991/978-2-38476-309-2_41

targeted training suggestions that significantly improve students' listening comprehension. Additionally, deep learning algorithms can handle complex speech signals, gradually improving listening training effectiveness by continuously optimizing model parameters, and reducing interference factors during the training process[1]. These advantages make deep learning algorithms more efficient, accurate, and adaptive in listening training, marking a significant technological innovation in the field of English listening education.

3 Deep Learning-Based Junior High School English Listening Training Model

3.1 LSTM-Based English Listening Training Model

The LSTM-based English listening training model leverages the characteristics of Long Short-Term Memory (LSTM) networks to effectively process and retain long-term dependency information in English listening training, thereby enhancing the model's ability to continuously improve students' listening comprehension[2]. The core of the LSTM network lies in its gating mechanism, which allows the network to intelligently decide which information to retain and which to forget during the processing of sequential data. Specifically, the LSTM unit includes three main gating structures: the forget gate, input gate, and output gate.

1.The Formula for the Forget Gate is as Follows:

$$f_t = \sigma(W_f \cdot [h_{t-1,} x_t] + b_f) \tag{1}$$

In the formula, f_t represents the activation vector of the forget gate, σ is the sigmoid function, W_f and b_f are the weight matrix and bias vector of the forget gate, respectively, h_{t-1} is the hidden state from the previous time step, and x_t is the input at the current time step.

2. The Formula for the Input Gate is as Follows:

$$i_t = \sigma(W_i \cdot [h_{t-1,} x_t] + b_i) \overline{C}_t = \tanh(W_C \cdot [h_{t-1,x_t}] + b_C) \tag{2}$$

Here, i_t represents the activation vector of the input gate, $\overline{C}_t$ is the value of the candidate memory cell,and W_i, W_C and b_i, b_C are the weight matrix and bias vector associated with the input gate.

3. The Formula for the Output Gate is as Follows:

$$o_t = \sigma(W_o \cdot \lfloor h_{t-1,} x_t \rfloor + b_o) \tag{3}$$

Here, o_t represents the activation vector of the output gate, and W_o and b_o are the weight matrix and bias vector of the output gate.

4. The State Update Formula is as Follows:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t h_t = o_t * \tanh(C_t) \tag{4}$$

Here, C_t represents the memory cell state at the current time step, h_t is the hidden state at the current time step, and $*$ denotes the Hadamard product, which is the element-wise multiplication. These formulas collectively form the mathematical foundation of the LSTM model, enabling it to exhibit exceptional memory and forgetting capabilities in sequential data processing, thus providing strong technical support for junior high school English listening training[3].

3.2 Generation of Personalized Listening Training Plans

The generation of personalized listening training plans relies on a complex decision-making process that can be finely tuned through optimization algorithms. By employing a decision tree method combined with optimization algorithms, the listening training plan for each student can be effectively adjusted, allowing for more personalized training content[4]. In this context, the decision tree is used to evaluate and select the difficulty and type of listening training tasks, while the optimization algorithm determines the training sequence best suited to the student's current listening level. The core calculation formula of the decision tree combined with the optimization algorithm can be expressed as the minimization of a loss function, which includes adaptive adjustments based on student feedback. The specific formula is as follows, used for adjusting and optimizing the training plan:

$$L(\theta) = \sum_{i=1}^{N} (y_i - f(x_i; \theta))^2 + \lambda \|\theta\|^2 \tag{5}$$

In this formula, $L(\theta)$ represents the loss function, θ are the model parameters, x_i denotes the input features (such as the student's historical listening performance and personal preferences), y_i is the target output (the evaluation of training effectiveness), N is the total number of training samples, $f(x_i; \theta)$ is the prediction function based on parameters θ, and $\lambda \|\theta\|^2$ is the regularization term used to prevent model overfitting and enhance the model's generalization ability. By optimizing this loss

function, the training plan can be adjusted to better meet the student's needs, thereby improving the effectiveness and efficiency of the listening training[5].

3.3 Model Architecture and Design

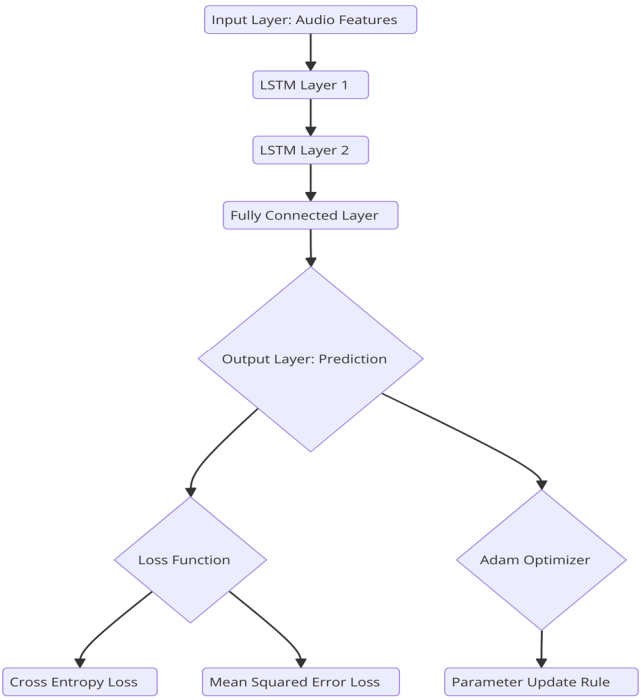

Fig. 1. Deep Learning Model Architecture

The model utilizes the cross-entropy loss function for classification tasks or the Mean Squared Error (MSE) loss function for regression tasks, depending on the specific training objectives[6]. For classification tasks, the cross-entropy loss function is an ideal choice because it measures the difference between the probability distribution of the model's output and the true distribution of the target. In terms of model optimization, the Adam optimizer is commonly used as it combines momentum and adaptive learning rate techniques, allowing for rapid convergence in the early stages of training while maintaining stable learning efficiency in later stages. The Adam optimizer updates parameters by adjusting the learning rate for each parameter through the calculation of first-moment estimates (i.e., the mean) and second-moment estimates (i.e., the uncentered variance)[7]. The specific parameter update rule for the Adam optimizer can be expressed as:

$$\theta_{t-1} = \theta_t - \frac{\eta}{\sqrt{\hat{\upsilon}_t} + \in} \hat{m}_t \tag{6}$$

Here, θ_t represents the parameter at time t, η is the learning rate, $\hat{m}_t$ and $\hat{\upsilon}_t$ are the bias-corrected first and second moment estimates, respectively, and $\in$ is a small constant added for numerical stability. The model consists of several key components: the input layer, multiple LSTM layers, one or more fully connected layers, and the output layer[8]. The input layer receives the raw English listening audio features, the LSTM layers process the sequential data and capture temporal dependencies, the fully connected layers map the outputs of the LSTM layers to a higher feature space, and the output layer produces the final prediction results based on the task requirements, such as classification or regression, as shown in Figure 1.

4 Experimental Results and Analysis

4.1 Experimental Environment and Dataset

In this experiment, the hardware environment included a high-performance computer equipped with an NVIDIA GeForce RTX 3090 GPU, an Intel Core i9-12900K CPU, 64GB of RAM, and 2TB of NVMe SSD storage[9]. The software environment consisted of the Ubuntu 20.04 operating system, TensorFlow 2.10 as the deep learning framework, Python version 3.8.13, along with CUDA 11.6 and cuDNN 8.4 to fully utilize the GPU's computational power. The experimental dataset was derived from the publicly available TED-LIUM 3 speech recognition dataset, which includes audio and transcribed texts from TED talks. To meet the needs of junior high school English listening training, this study extracted content suitable for junior high school students from the original dataset. A total of 300 audio samples were collected, with each sample ranging from 3 to 10 seconds in length. The dataset was divided into a training set (80%), a validation set (10%), and a test set (10%), consisting of 240, 30, and 30 samples, respectively.

During the data preprocessing phase, the audio samples underwent noise reduction, normalization, and feature extraction using Mel-Frequency Cepstral Coefficients (MFCC). All audio samples were resampled to 16kHz mono to ensure data consistency. The text portion was tokenized, stop words were removed, and the text was converted into corresponding word vectors, as shown in Table 1.

Table 1. Dataset Size and Feature Statistics

Dataset Type	Number of Samples	Total Audio Duration (minutes)	Average Sample Duration (seconds)
Training Set	240	36	9
Validation Set	30	4.5	9
Test Set	30	4.5	9

4.2 Model Performance Evaluation

To comprehensively evaluate the model's performance on junior high school English listening training tasks, accuracy, recall, precision, and F1 score were calculated on

both the training and test sets[10]. The following sections will present detailed performance results across different datasets, with multiple tables providing a thorough analysis of the experimental outcomes.Firstly, on the training set, the model exhibited high accuracy and recall, indicating strong generalization capabilities in recognizing and understanding the training data. Specific data is illustrated in Figure 2.

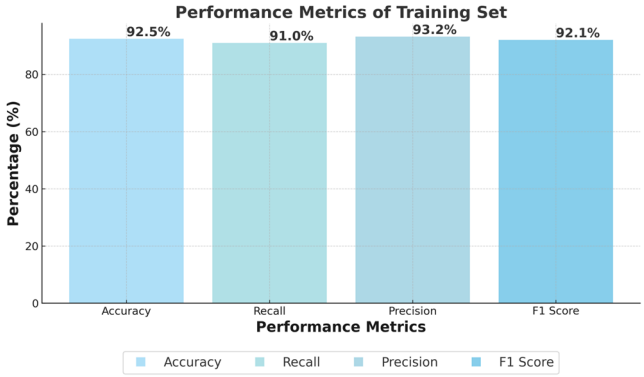

Fig. 2. Performance Metrics on the Training Set

As shown in Figure 2, the model achieved an accuracy of 92.5% on the training set, indicating that the model can accurately predict students' listening comprehension abilities. The recall rate was 91.0%, suggesting that the model correctly identified most of the correct samples in the training set. The precision rate was 93.2%, indicating that the majority of the predicted samples were correct. The F1 score, which combines precision and recall, was 92.1%, further validating the model's stability and effectiveness.

Next, we will analyze the model's performance on the test set to evaluate its practical application on unseen data. The performance metrics for the test set are shown in Figure 3.

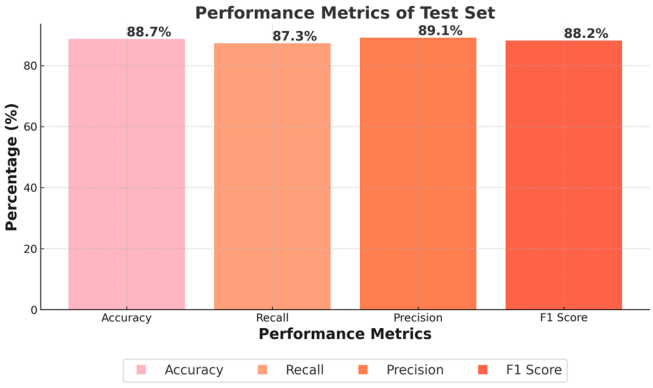

Fig. 3. Performance Metrics on the Test Set

As shown by the data in Figure 3, the model achieved an accuracy of 88.7% on the test set, slightly lower than on the training set but still maintaining a high level of predictive capability. The recall rate was 87.3%, indicating that the model could effectively identify correct listening comprehension outcomes on unseen data. The precision rate was 89.1%, similar to the performance on the training set, suggesting that most of the predicted results on the test set were accurate. The F1 score, a composite metric, was 88.2%, demonstrating that the model maintained good balance on the test set.

4.3 Comparative Analysis with Traditional Methods

Traditional methods often rely on fixed template exercises and manual grading, which, although somewhat effective, show significant limitations in personalization, efficiency, and large-scale data processing capabilities. The following data analysis demonstrates the advantages of deep learning methods in practical applications. We compared the accuracy and recall rates of the two methods on the training set. The deep learning method showed a clear advantage in the training set, as shown in Table 2.

Table 2. Performance Comparison Between Deep Learning and Traditional Methods on the Training Set

Performance Metric	Deep Learning Model (%)	Traditional Method (%)
Accuracy	92.5	78.4
Recall	91	76.2
Precision	93.2	79.1
F1 Score	92.1	77.6

As shown in Table 2, the deep learning model's accuracy is 92.5%, significantly higher than the 78.4% of the traditional method, indicating that the deep learning model can more accurately predict listening comprehension. In terms of recall, the deep learning model reached 91.0%, also significantly outperforming the traditional method's 76.2%, meaning that the deep learning model could more comprehensively identify correct listening comprehension outcomes. The comparisons of precision and F1 score further reinforce this conclusion, showing that the deep learning method outperforms traditional methods across all metrics.Next, we compared the performance on the test set, as shown in Table 3.

Table 3. Performance Comparison Between Deep Learning and Traditional Methods on the Test Set

Performance Metric	Deep Learning Model (%)	Traditional Method (%)
Accuracy	88.7	72.5
Recall	87.3	70.4
Precision	89.1	74
F1 Score	88.2	72.1

The data in Table 3 indicates that even on the unseen data in the test set, the deep learning model maintained a significant advantage. The deep learning model's accuracy was 88.7%, showing a significant improvement compared to the 72.5% of the

traditional method, highlighting its strong generalization ability when handling new data. In terms of recall, the deep learning model achieved 87.3%, far surpassing the 70.4% of the traditional method, demonstrating its superior ability to identify correct samples. The precision and F1 score also stood out, further proving that the deep learning method outperforms traditional listening training methods across all performance metrics.

4.4 Analysis of Students' Listening Ability Improvement

By comparing students' performance before and after the experiment, we can visually observe the degree of improvement in listening ability and validate the practical effectiveness of the deep learning model. First, the average scores and standard deviations of students before and after the listening training were assessed, as shown in Table 4.

Table 4. Comparison of Students' Average Scores and Standard Deviations Before and After Listening Training

Testing Phase	Average Score (points)	Standard Deviation (points)
Before Training	65.3	8.7
After Training	78.9	5.6

As shown in Table 4, the average score of students before the listening training was 65.3 points, with a standard deviation of 8.7 points, indicating a wide variance in listening ability among students. After undergoing training with the deep learning model, the average score increased to 78.9 points, and the standard deviation decreased to 5.6 points, indicating a significant overall improvement in listening ability and a reduction in the variance among students, showing more balanced progress. Next, we analyzed the progress of students in different score ranges, as shown in Table 5.

Table 5. Comparison of Listening Scores Before and After Training for Students in Different Score Ranges

Score Range	Number of Students	Average Score Before Training (points)	Average Score After Training (points)	Score Improvement (points)
Below 60	50	54.2	69.8	15.6
60-80	150	71.5	82.3	10.8
Above 80	100	85.6	91.2	5.6

As shown in Table 5, students in the below-60 score range saw the greatest improvement, with a 15.6-point increase, indicating that the deep learning model significantly benefits students with weaker foundations. Students in the 60-80 score range also saw substantial improvement, with an average increase of 10.8 points, while students in the above-80 score range showed a more moderate improvement, with an average increase of 5.6 points. This suggests that the deep learning model is beneficial for students of all proficiency levels, with particularly notable effects for those with weaker foundations.

To further validate the training effect, we conducted a statistical analysis of students' accuracy and error rates before and after the training. Before the training, the students' average accuracy was 62.4%, and the error rate was 37.6%. After the training, the accuracy significantly increased to 81.7%, and the error rate decreased to 18.3%. This data further proves the effectiveness of the deep learning model in improving students' listening abilities, especially in reducing listening errors.

5 Conclusion

Deep learning algorithms have demonstrated outstanding advantages in personalization and intelligence in junior high school English listening training, effectively improving students' listening comprehension abilities and significantly optimizing training outcomes when dealing with complex speech signals. In the future, further exploration of more refined algorithm models and broader dataset applications is expected to enhance listening training outcomes comprehensively, providing solid technical support for the digital development of English education.

References

1. Xiangying K , Shuang B ,Yalin G .THE APPLICATION OF FLIPPED CLASSROOM IN ADVANCED ENGLISH LISTENING TRAINING[J]. Psychiatria Danubina, 2021, 33(S8): 409-411.
2. Blagoja D ,Brett M .Three Bottom-up Listening Training Ideas for the English as a Lingua Franca Classroom[J].The Center for ELF Journal,2019,536-53.
3. Suzuki C .A Study of the Listening Comprehension Ability of Native Speakers of English : Inquiry into a Training Method for Japanese Students[J].Language Laboratory, 2017, 18(0):1-10.
4. Grohe A ,Weber A .Learning to Comprehend Foreign-Accented Speech by Means of Production and Listening Training[J].Language Learning,2016,66(S2):187-209.
5. Ma ,Jianhe. Pure Listening Training—A Tentative Innovation of English Learning[J]. Theory and Practice in Language Studies,2011,1(6):729-731.
6. Ma T . Communicative Listening Training in English— Features, Strategies and Methods [J]. Journal of Language Teaching and Research,2010,1(4):464.
7. Choi F F .Personal Multimedia Player for English Listening Training[J].Ubiquitous Learning: An International Journal,2009,1(1):67-76.
8. Suri S S .Teachers' Strategies to Enhance Deeper Learning Skills in English Language Classes[J].International Journal of Linguistics, Literature and Translation, 2024, 7(3):118-125.
9. Yongqing H ,Anand P ,K.S. S C , et al. Evaluation of English online teaching based on remote supervision algorithms and deep learning[J].Journal of Intelligent & Fuzzy Systems, 2021,40(4):7097-7108.
10. Lei W ,Wang L ,Cao X , et al. English Letter Recognition Based on TensorFlow Deep Learning[J].Journal of Physics: Conference Series,2020,1627(1):012012-.