

Building a data-driven teaching platform for ESL vocabulary corpus in universities based on VR technology

Lijun Cheng*

School of Foreign Languages, Anyang Institute of Technology, Anyang 455000, China

Received: 20 September 2023 / Accepted: 27 February 2024

Abstract. With the rapid development of information technology, the application of VR to English language teaching has gradually become an emerging trend. As the driving force of English learning, the establishment of ESL vocabulary corpus is of great significance for making great progress in teaching in colleges and universities. By analyzing the characteristics of VR technology, the study establishes a data-driven teaching platform for ESL vocabulary corpus in colleges and universities based on this technology. Then, based on the improved graph neural network, the text classification of the ESL vocabulary corpus was carried out to achieve the integration of ESL vocabulary corpus resources. The results show that the improved graph neural network vocabulary text classification model converges at 145 iterations, and the convergence speed is improved by up to 100 generations; Among different datasets, the recognition accuracy of the SADE-GNN model on the MR, R8, SST1, and SUBJ datasets is 90%, 99%, 63%, and 96%, respectively. In the recognition and classification of ESL vocabulary corpora in universities, the accuracy of this model is stable at around 95%, with a maximum value of 97%. In practical teaching applications, students' English listening, reading, writing, and speaking scores have all increased to varying degrees, with the top 50% of students achieving 95 or above in both reading and speaking. The above results indicate that the teaching platform under the research model has high accuracy in text classification, can significantly improve the effectiveness of English teaching, providing new reference ideas for the construction of ESL vocabulary corpus and teaching method reform in universities.

Keywords: VR / ESL vocabulary / corpus / teaching platform / graph neural network

1 Introduction

The ESL (English as a Second Language) vocabulary bank in colleges and universities is an important element and foundation of English language teaching [1]. With the continuous development of corpus linguistics, data-driven learning has emerged as a new approach to foreign language learning, and it plays an important role in teaching English in higher education. Based on certain linguistic principles, a corpus is a large electronic text library with a certain capacity that collects continuous and naturally occurring fragments of language use discourse or text through random sampling [2]. Today's corpus, with its verifiability, adequacy and ease of use, is leading to a qualitative and quantitative leap in the use of linguistics. With the advent of web-based technology, the traditional English-Chinese approach and ESL vocabulary teaching in isolation from the discourse are no longer sufficient to meet the needs of English teaching. The data-driven learning mode based on corpus has gradually become an important development

direction for English vocabulary teaching and has received widespread attention in the field. The establishment of a data-driven teaching platform for ESL vocabulary corpus is an inevitable trend [3]. At the same time, with the progress of computer information technology, the conditions for the application of VR (virtual reality) technology in the teaching are becoming more and more adequate. Therefore, the study builds a data-driven teaching platform for ESL vocabulary corpus in universities through VR technology in order to further enhance the effectiveness of English teaching.

2 Related works

In recent years, the establishment of a data-driven teaching platform for the ESL vocabulary corpus through VR technology has received the attention of many professionals. Among them, many outcomes have been achieved on English vocabulary corpora. Yang and Coxhead [4] proposed to optimize the number of high-frequency, medium-frequency and low-frequency vocabulary in the New Concept English textbook series in response to the low

* e-mail: lijun_cheng2023@outlook.com

level of English learners' vocabulary knowledge, and analyzed the impact of learning progress on vocabulary. The outcomes showed that learning through vocabulary has a certain promotion effect on English teaching. Lee's [5] research team to address the problem of inadequate English vocabulary among university students, a data-driven learning model was developed based on assessing students' vocabulary levels and working memory capacity; A lexical reasoning strategy was identified; The outcomes showed that the method was beneficial to university students' vocabulary learning and working memory levels. Gholaminejad and Sarab [6] addressed the problem of developing disciplinary vocabulary lists by designing an academic glossary with a specific genre and time span, with paired word lists representing 7.1% of the corpus coverage. The outcomes show that the method provides an effective vocabulary list reference for students, teachers, and material developers. Lowphansirikul et al. [7] scholars trained a machine translation model for building a large-scale English-Thai dataset, collecting data from news, Wikipedia articles and web. Based on this, translation errors and preprocessing noise were eliminated and the quality of the corpus was evaluated by training the machine translation model. The outcomes proved that the method has high application efficacy. Kirk's [8] research team, for the International English Corpus Project, asked fundamental questions about the nature of English variants in a multilingual environment. And the importance of corpora was demonstrated through questionnaires filled out by investigators, providing a new approach for the pre-corpus database applications provides a new solution. Scholars such as Crosthwaite et al. [9] built a data visualization and corpus query platform to track students' use of a data-driven learning corpus database and to eliminate discrepancies between specific corpus features and query syntax. This was demonstrated to actively generate unique queries and improve monitoring of students' corpus use.

The outcomes of the application of VR technology in English language teaching and learning provide a technological input to the data-driven teaching and learning approach of the ESL vocabulary corpus proposed by the study. Yeh et al. [10], in order to make English language content and language skills further accessible to students, enhanced students' cross-cultural learning by producing VR content, including panoramic, interactive, audio and structured. The outcomes showed that students, with the help of VR technology features developed better intercultural learning and enhanced English learning ability and effectiveness. Cui [11] proposed the extraction of information features through target visual detection and deep learning for online teaching of an English vocabulary corpus, and the identification of recognizable components in images. The outcomes showed that the corpus was used in university English vocabulary teaching to promote students' independent learning ability, with classification accuracy of up to. El Jamiy and Marsh [12] addressed the problem of reduced depth perception in VR technology by improving depth perception, optimizing it with the visual cues it contains, and proposing a rendering system that renders in real scenes. The outcomes demonstrate the high performance of the method. Liao's [13] research team

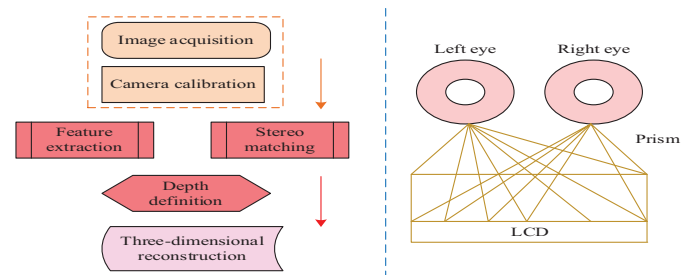


Fig. 1. Technical schematic diagram of VR.

designed a character-attentive fully convolutional network for recognizing image text in 2D space. This network is capable of recognizing text of arbitrary shapes, and implemented scene text recognition through semantic segmentation networks, demonstrating the accuracy of the method in the detection of text datasets. Xu et al. [14] proposed a new text detector for the problem of multi-directional text detection to detect irregular scene text. This detector separates adjacent text instances by encoding binary text masks and directional information, and shows a 28% performance improvement. Wang et al. [15] addresses the severe alignment problem arising from cutting-edge text recognition methods and proposes a decoupled attention network. This network contains a feature encoder, a convolutional alignment module and a text decoder, decoupled by alignment operations and using historical decoding outcomes, and experimental. The outcomes demonstrate its high accuracy in text recognition.

In summary, the construction of an English vocabulary corpus and the application of VR technology in English teaching have sufficient theoretical and implementation bases. While less research has been conducted on the combination of the two. Therefore, the study establishes a data-driven teaching platform for vocabulary corpus through VR technology in order to further promote the reform and progress of English teaching.

3 Construction of a data-driven teaching platform for ESL vocabulary corpus based on VR technology

3.1 Teaching ESL vocabulary corpus based on VR technology

Based on VR technology, establish an ESL vocabulary corpus teaching platform. Firstly, the characteristics and principles of virtual reality technology were analyzed. The basic characteristics of VR technology are interactivity, immersion, and imagination. Through computer simulation, users immersed in it can obtain dual cognition at both rational and perceptual levels. In a good virtual reality environment, users' creative catalysis and knowledge construction can be deepened [16]. In the virtual space composed of VR technology, users can broaden their cognitive range, absorb the necessary knowledge based on their own cognitive abilities and feelings, and diverge and broaden their thinking [17]. The application of VR

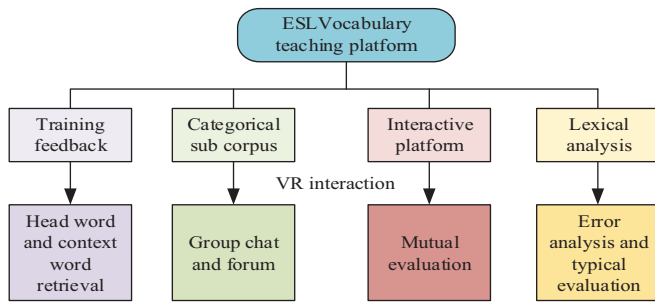


Fig. 2. ESL vocabulary corpus data driven teaching platform.

technology in education mainly aims to improve the quality of classroom teaching, strengthen students' learning motivation; and it can cultivate self-directed learning habits while changing traditional learning methods. It studied the construction of a vocabulary teaching platform based on the Roblox virtual platform. The technical principle of VR is shown in [Figure 1](#).

By analyzing the advantages and features of VR technology, a technical basis for data-driven teaching of ESL vocabulary corpus is provided. For ESL vocabulary teaching, the content mainly includes the form, meaning, sound and collocation of words, of which collocation is the most crucial part. In traditional vocabulary collocation teaching, teachers explain vocabulary through English-Chinese translation, usually from dictionaries and examples from teachers' own experience, so that students learn word collocation principles passively, mainly through rote memorization, rather than in frequent use of Chinese-English [18]. With this data-driven approach to learning, teachers can guide students to search the ESL vocabulary corpus for typical collocations of a particular word and summarize the collocation patterns to achieve a true mastery of contextual collocation usage. The data-driven approach of ESL vocabulary data-driven teaching allows learners to avoid overlooking errors due to linguistic intuition and to correct incorrect usage in time. Utilizing the advantages of data-driven vocabulary teaching in terms of word collocation, the study builds a teaching platform that connects students to each other and to each other through VR technology. And it operates as a corpus data-driven self-directed learning system, outcoming in a personalized and self-directed learning system characterized by "student-teacher-assisted-directed learning". This is a personalized and networked vocabulary teaching model featuring "student-teacher-assisted-autonomous learning", and students are able to access rich vocabulary information, completing the ESL vocabulary teaching platform architecture. A brief structure of the data-driven teaching platform for ESL vocabulary corpus in university with VR technology is demonstrated in [Figure 2](#).

The platform contains four main functional sections, divided into ESL vocabulary acquisition module, teacher-student interaction module, ESL vocabulary training feedback module and categorized sub-corpus module. Contextual co-occurrence search and central word search are the main functions of the categorical corpus module, which can improve learners' efficiency and relevance in learning ESL vocabulary. The teacher-student interaction

module provides learners with the necessary help through chat and forum for asynchronous tutoring and learning of ESL vocabulary, asynchronous communication and online synchronization with the help of VR technology. The vocabulary training feedback module provides a platform for students and teachers to evaluate each other and for students to evaluate each other, inspiring learners to participate and cooperate. The vocabulary assessment module enables learners to communicate with each other through VR technology, including the analysis of errors and special meanings of ESL vocabulary. And it is to deepen students' understanding of ESL vocabulary in an interactive context using VR technology. Thus, the data-driven teaching platform for the ESL vocabulary corpus built with VR technology achieves better learning outcomes for students.

3.2 Lexical text classification based on graph neural networks

After establishing an ESL vocabulary corpus teaching platform through VR technology, the graphics processing technology in VR technology is optimized to further improve the classification accuracy of ESL vocabulary texts and thereby improve teaching effectiveness. The research of graph neural networks in deep learning has gained widespread attention due to their superior classification accuracy, for feature recognition and classification of ESL vocabulary texts. When using graph neural networks for text classification, the first step is to remove markers and stop words, which is a preprocessing process [19]. Then it initializes the word features to represent vertex embedding, and generate corresponding graphs for each document. Finally, the word feature information is propagated and combined in the context, and the classification prediction module outputs the classification results. However, graph neural network realizes classification based on graph text conversion features, and there is a over-fitting problem caused by less training data sets, so it needs to be improved [20]. Easy Data Augmentation (EDA) can enhance the text data by simply adding noise, reduce the over-fitting phenomenon of the system, and has good universality. Therefore, the study introduces a combination of EDA and self-attention mechanism, and proposes a Graph Neural Network based on Self Attention and Data Enhancement (SADE-GNN), which is applied to ESL vocabulary corpus classification. The basic structure of the SADE-GNN model proposed in the study is shown in [Figure 3](#).

The proposed SADE-GNN model retains the GNN graphical coding properties, maintains the model's stable performance with relatively little training data through EDA data augmentation techniques, and introduces self-attention to strengthen the interconnections between word levels. Through the operation of this model, both learning and language can be carried out smoothly, and the models are very similar. The model consists of four main parts, the first being the data sampling part, which is a combination of self-attention and EDA; The second is the part that performs the graph construction operation through a sliding window; The third part uses gated neural networks for word feature interaction, and finally the text prediction classification through two multi-layer perceptrons. In the data sampling

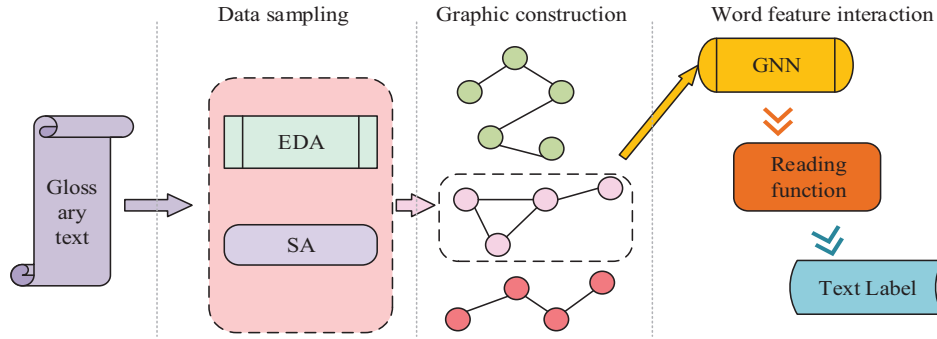


Fig. 3. Basic structure of SADE-GNN model proposed.

part, the EDA technique performs data augmentation through synonym substitution, random insertion, random swapping, and random deletion, yielding an effective amount of data several times larger than the original data [21]. The amount of word variation in the EDA operation is calculated as demonstrated in equation (1).

$$n = l\alpha. \quad (1)$$

In equation (1), l represents the sentence length; α represents the percentage parameter and n is the word variation. The EDA operation is followed by a self-attentive mechanism, which calculates the similarity between the key vector matrix and the query vector matrix according to the scoring function. And it performs a numerical conversion operation on the outgoing similarity score by the softmax function, and finally uses the weighting coefficients to weight the summation vector matrix Value. The attention score is calculated as demonstrated in equation (2).

$$s_i = F(Q, k_i). \quad (2)$$

In equation (2), s_i represents the attention score; s_i is the query vector and F is the scoring function. k_i represents the keyword for the query operation. Generally speaking, the query vector matrix and the key vector are vectors of different lengths, and additive attention can be used as the scoring function. Then given the value range of the query vector and key vector, $F = w_e^T \tanh(W_Q Q + W_K K)$ can be obtained. Where W_Q, W_K represents the learnable parameters. The input information probabilities are demonstrated in equation (3).

$$a_i = \text{soft max}(s_i) = \frac{\exp(s_i)}{\sum_{j=1}^N \exp(s_j)}. \quad (3)$$

In equation (3), a_i represents the input information probability. The weighted summation is demonstrated in equation (4).

$$\text{Attention}[(K, V), Q] = \sum_{i=1}^N v_i a_i. \quad (4)$$

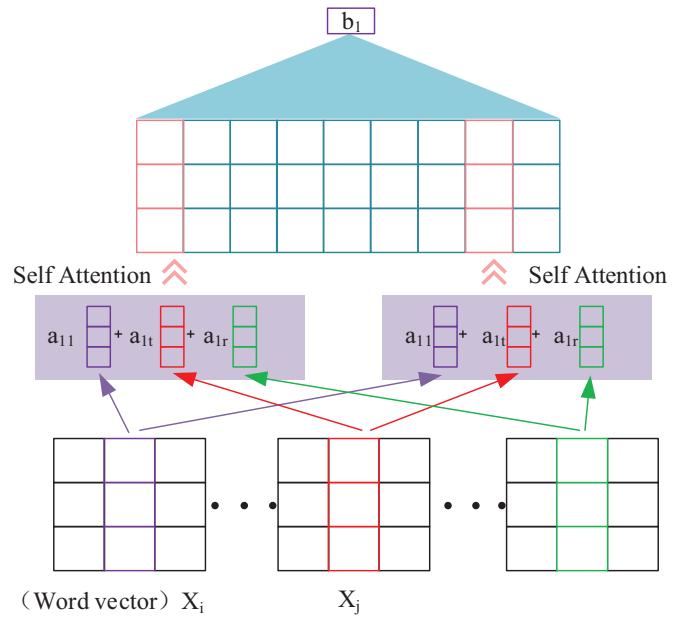


Fig. 4. Flow diagram of self-attention mechanism.

In equation (4), K represents the key vector matrix and V is the value vector matrix. The self-attentive key parameters are demonstrated in equation (5).

$$\text{Att}(Q, K, V) = \omega V(QK^T). \quad (5)$$

In equation (5), ω represents the weighting factor. For each of the three matrices in the self-attentive mechanism, each row corresponds to a corresponding vector representation, which is obtained by multiplying the input serial numbers by $W^{(q)}, W^{(k)}$ and $W^{(v)}$, as demonstrated in Figure 4.

As can be seen in Figure 4, in the self-attention mechanism, the model calculates the attention weight of each element in the sequence (for example, each word in the sentence) to other elements. These weights reflect how dependent the current element is on other elements in generating output. Therefore, considering the uniqueness of each element, experiments were conducted to determine

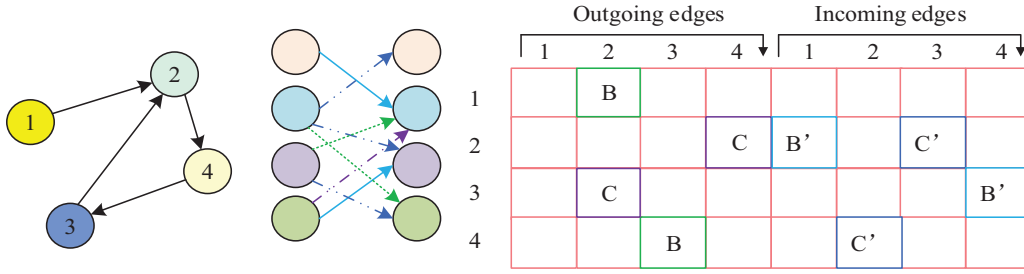


Fig. 5. Construction Diagram of Adjacency Matrix in Gated Neural Network.

the attention weight of each element through the relative importance between different elements. In addition, in the graph construction section, the words selected in the sentence are represented as vertices and the co-occurrence between words is used for graph construction. The co-occurrence is the correlation between words in a sliding window, the size of the sliding window is set to 3 and the edges are all undirected. The dense connection of graphs may outcome in ambiguous information about word features, so the gated neural network first initializes the word features as vertex embedding [22]. A separate subgraph representation is then implemented for all documents, thus allowing for complete propagation of word feature information in the context. Once the image construction part is completed, the embedding representation of the word nodes is clearly presented through gated neural network learning. All nodes in the GGNN receive information about neighbouring nodes through edges and connect this information to their own information to update the information [23]. The adjacency matrix is the main means by which the gated neural network updates information, and the adjacency matrix is constructed as demonstrated in Figure 5.

The first step in the recursive process of the propagation model is demonstrated in equation (6).

$$h_v^1 = [x_v^T, 0]^T. \quad (6)$$

In equation (6), v represents the node; x_v is the input feature of the node, and h_v^1 represents the initial state of the node, which is a D-dimensional vector. The outcome of the update of the current node and the neighbouring nodes is demonstrated in equation (7).

$$a_v^t = b + A_{v:}^T [h_1^{T(t-1)} \dots h_{|V|}^{(t-1)T}]^T. \quad (7)$$

In equation (7), a_v^t represents the update outcome; A is the matrix corresponding to the node; $A_{v:}^T$ is the first two columns of the matrix A ; $[h_1^{T(t-1)} \dots h_{|V|}^{(t-1)T}]$ is the matrix obtained by performing augmentation on all node features at time $t-1$. The control forgetting information is calculated as demonstrated in equation (8).

$$z_v^t = \sigma(U^z h_v^{t-1} + W^z a_v^t). \quad (8)$$

In equation (8), z_v^t is the control forgetting information. The control of newly generated information is demonstrated in equation (9).

$$r_v^t = \sigma(U^r h_v^{t-1} + W^r a_v^t). \quad (9)$$

In equation (9), r_v^t represents the control of newly generated information. And the newly generated information is demonstrated in equation (10).

$$\tilde{h}_v^t = \tanh[(r_v^t \odot h_v^{t-1})U + W a_v^t]. \quad (10)$$

In equation (10) \tilde{h}_v^t is the newly generated information that determines which past information the new information is generated from. The final node state obtained by updating is demonstrated in equation (11).

$$h_v^t = z_v^t \odot h_{v\tilde{}}^t + (1 - z_v^t) \odot h_v^{t-1}. \quad (11)$$

In equation (11), h_v^t is the final node state; $(1 - z_v^t) \odot$ is the past information selected for forgetting, and $z_v^t \odot$ is the new information selected for remembering. The word nodes are sufficiently updated to become a graph representation of the document by encoding. The information weights of the word features in the document are obtained using the soft attention layer, and the graph vector is obtained by multiplying the information weight matrix with the word features by the non-linear feature transform, as demonstrated in equation (12).

$$h_v = \sigma[f_1(h_v^t)] \odot \tanh[f_2(h_v^t)]. \quad (12)$$

In equation (12), h_v represents the graphical vector representation; h_v^t is the word feature; f_1 is the soft attention mechanism; f_2 represents the non-linear feature transformation. All documents are then concatenated and passed through a maximum pooling layer where the relational words have the maximum weight, as demonstrated in equation (13).

$$h_g = \text{Maxpooling}(h_1 \dots h_v) + \frac{1}{|V|} \sum_{v \in V} h_v. \quad (13)$$

In equation (13), h_g represents the graph vector and $(h_1 \dots h_v)$ is all documents. Using the softmax layer to predict the graph vector and minimising the loss through

the cross-entropy function, the first i element of the One-shot label \hat{y}_g is calculated as demonstrated in equation (14).

$$\hat{y}_g = \text{soft max}(b + Wh_g). \quad (14)$$

In equation (14), b and W represent the weight and deviation respectively. The final prediction outcomes are demonstrated in equation (15).

$$\Gamma = -\sum_i y_{g_i} \log(\hat{y}_g). \quad (15)$$

Thus, after the read-out function is executed, the whole process of lexical text classification by the proposed SADE-GNN model is finally completed.

4 Analysis of the effectiveness of the data-driven teaching platform for the ESL vocabulary corpus

The study establishes a teaching platform for ESL lexical corpus through VR technology and analyses the effect of its application. The main purpose was to validate the effect of improved graph neural networks in VR technology for lexical text classification, i.e. to test the performance of the SADE-GNN model. Therefore, the study selected four thousand pieces of data about English vocabulary from the GitHub software platform and divided them into four data sets, named MR data set, R8 data set, SST1 data set and SUBJ data set respectively. MR data is in the field of movie reviews. As a two-category data set, it contains positive and negative reviews. The R8 data set is in the news domain and is collected from the Reuters newswire. SST1 is the social domain, obtained from the Stanford Sentiment Treebank. SUBJ is a subjective data set, which divides sentences into two categories based on subjective and objective criteria. To ensure the quality of the data set, it is necessary to ensure the quality of the data set throughout the entire process of the experiment to ensure the accuracy, consistency and completeness of the data. Regularly update and maintain the obtained data set to ensure good timeliness of the data in the data set. Three models were also selected for comparison with the SADE-GNN model, namely CNN (Rand), BiLSTM and GCN (Texting). The CNN (Rand) model is based on a convolutional neural network and uses pre-trained word embedding to extract key information from sentences. After the convolutional layer of the CNN (Rand) model, the activation function is used to introduce the Rand nonlinear function, so that the model can learn more complex data patterns. To prevent over-fitting, CNN uses Dropout regularization technology. BiLSTM is a bi-directional LSTM structure and extracts information through pre-trained word embedding, while GCN (Texting) uses a separate graph for each document and then uses GCN for text classification. The experimental environment is demonstrated in Table 1.

The text classification outcomes of the four selected models in the four datasets are demonstrated in Figure 6. In Figure 6, the horizontal coordinates are the text

Table 1. Experimental environment.

Experimental environment	Detailed information
Development language	Python 3.6
Development framework	TensorFlow-gpu 1.13.1
CPU	i7-8700
GPU	RTX-2070
Operating system	Windows 10

classification accuracies and the vertical coordinates indicate the four datasets MR, R8, SST1, and SUBJ. It can be seen that the accuracy of the SADE-GNN model in the MR dataset is up to 90%, while the CNN (Rand), BiLSTM and GCN (Texting) models are 70%, 75% and 78% respectively, with an improvement of 20%, 15% and 13% for the SADE-GNN model respectively. In the SST1 dataset, the accuracy of the CNN (Rand), BiLSTM and GCN (Texting) models showed relatively small differences in accuracy, all below 60%, while the SADE-GNN model exceeded 60%, at approximately 63%, with a maximum improvement of 9%. In dataset R8, the accuracy of CNN (Rand), BiLSTM and GCN (Texting) were all above 80%, with GCN (Texting) approaching 95%, while the SADE-GNN model reached the maximum at 99%. In the dataset SUBJ, BiLSTM and GCN (Texting) were both between 80% and 90% accurate, with GCN (Texting) at roughly 93%, but the SADE-GNN model was still ahead of the first three models at roughly 96%.

The accuracy variation outcomes of the four models in the selected dataset for different number of iterations are demonstrated in Figure 7. In Figures 7a–7d shows the text classification accuracy outcomes of the SADE-GNN, CNN (Rand), BiLSTM and GCN (Texting) models respectively, with the horizontal axis showing the number of iterations and the vertical axis showing the text classification accuracy. Figure 7a showcase that in all four datasets, the accuracy of the SADE-GNN model decreases after the number of iterations increases to 145, i.e. the best outcomes are achieved around 145, with the highest accuracy above 90%. Figure 7–b–7–d illustrates that the CNN (Rand) converges at roughly 245 iterations, while the BiLSTM and GCN (Texting) reach their best outcomes at 200 and 190 iterations, respectively, i.e. convergence occurs. Combining the outcomes in Figure 7, it indicates that the SADE-GNN model converges significantly faster than the other three models, with faster convergence and better performance.

The four selected models were then applied to the university ESL lexical corpus for comparison of classification time and accuracy. To enhance the credibility of the outcomes, the study conducted three separate recognition classifications in this corpus, and the outcomes obtained are demonstrated in Figure 8. Figure 8a shows the classification accuracy outcomes of the four models, and Figure 8b shows the comparison of classification times. Figure 8a illustrates that the classification accuracy of CNN (Rand) was below 85% for all three times, BiLSTM was the highest at 85%, GCN

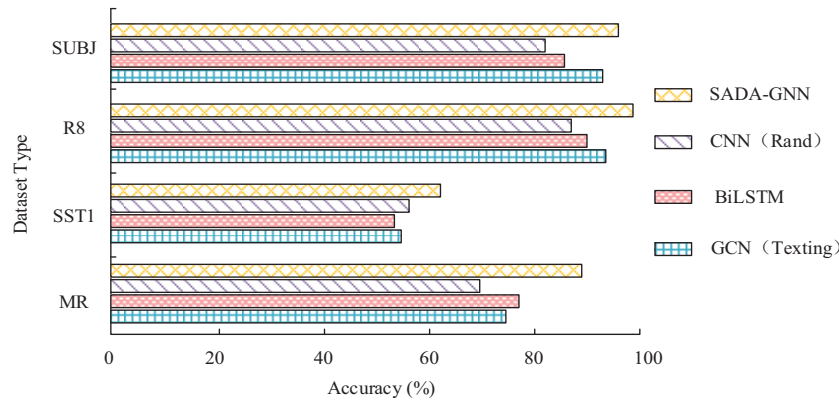


Fig. 6. Text classification outcomes of four selected models in four datasets.

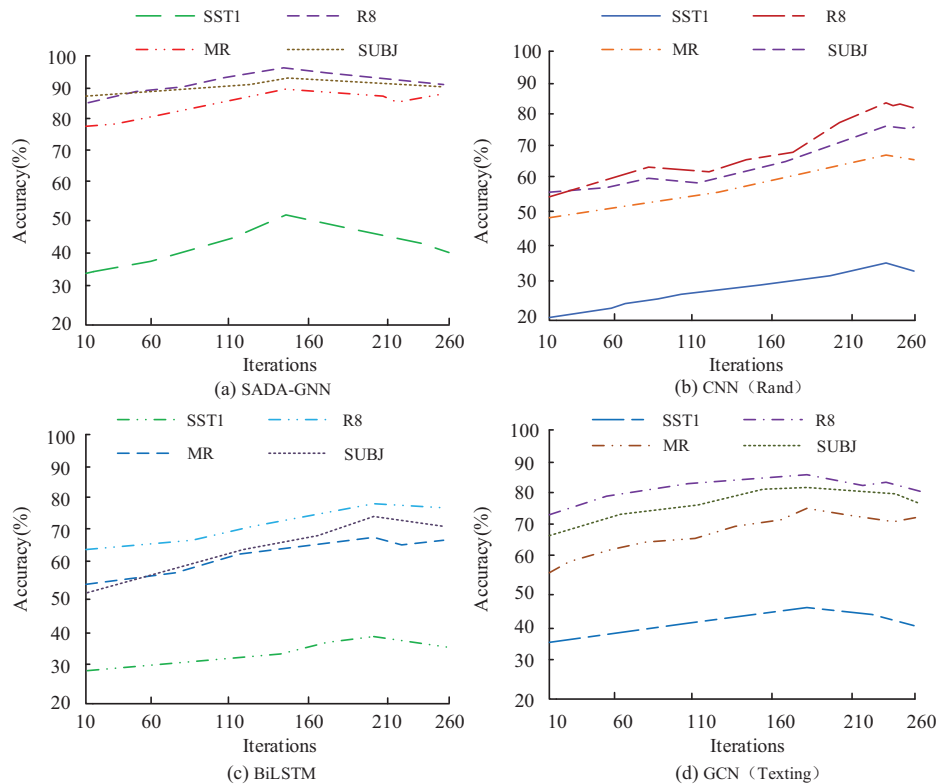


Fig. 7. Outcomes of accuracy change of four models in the selected dataset under different iterations.

(Texting) maintained around 85% with the highest close to 88%; SADE-GNN was stable at around 95% with the highest at 97% for the third time. Figure 8b showcases that both CNN (Rand) takes more than 25s, BiLSTM and GCN (Texting) both run in the range of 20–25s, with the latter being relatively more stable and the former running in a decreasing trend. The SADE-GNN model, on the other hand, maintained a running time of 10–15s, with a maximum of 15s and a minimum of 13s, making it a more efficient operation.

Finally, the proposed SADE-GNN model was applied to the teaching of English in a university in China, and the teaching effects before and after using it were evaluated from four aspects: listening, reading, writing and speaking,

and the statistical outcomes are demonstrated in Figure 9. Figures 9a and 9b show the comparison of the learning effects of the top 50% and bottom 50% of students before and after the application of the teaching platform, respectively. It reveals that after the application of the teaching platform, the top 50% of students showed different degrees of improvement in English listening, reading, writing and speaking, all of which scored above 90, and both reading and speaking scored 95 and above. At the same time, the bottom 50% of students in this teaching platform all improved their scores to the range of 70 to 80, and were able to improve their listening scores to 82. Figure 9 shows that the SADE-GNN model can effectively improve students' learning outcomes.

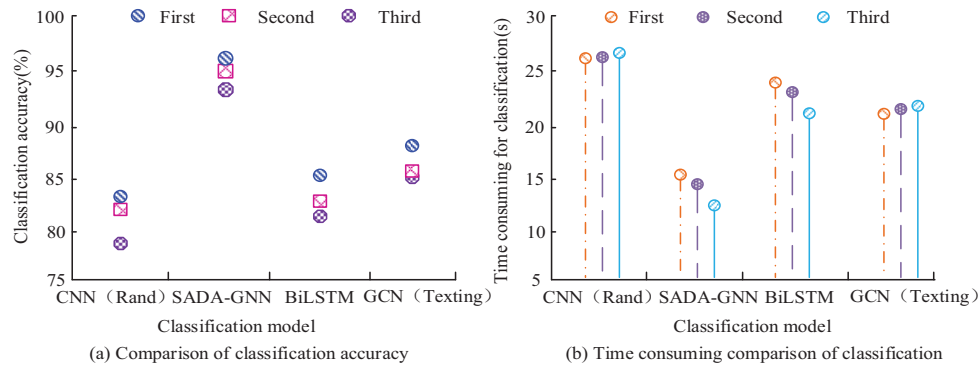


Fig. 8. Comparison of recognition time and classification accuracy of four models in ESL lexical corpus.

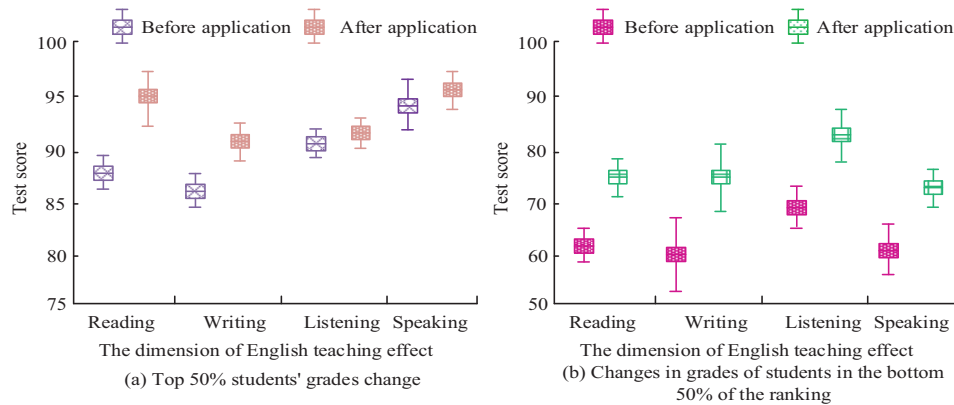


Fig. 9. Changes in students' English scores before and after the application of ESL vocabulary corpus data driven teaching platform.

5 Conclusion

English teaching in the new era cannot be supported by new technologies, and enhancing teaching effectiveness through VR technology has become one of the development directions of education reform. The study builds a data-driven teaching platform based on the principles of VR technology and the characteristics of the art, using the ESL vocabulary corpus of universities as content, and classifies the corpus with lexical texts through improved graph neural networks. The outcomes show that the proposed SADE-GNN model achieves the highest accuracy of 90%, 99%, 64% and 93% in the four datasets of MR, R8, SST1, and SUBJ respectively, outperforming the other three models; In the accuracy variation of different iterations, convergence occurs at roughly 245 iterations for CNN (Rand), and at 245 iterations for BiLSTM and GCN (Texting) reached the best outcomes at 200 iterations and 190 iterations respectively; While the accuracy of the proposed SADE-GNN model started to decrease at the 145th iteration, with a faster convergence rate; In the classification of ESL vocabulary corpus in universities, the model consumed time within 10–15 s and the accuracy rate was stable at around 95%; In the application of English teaching in universities, the students' English scores all improved, with the top 50%

and bottom 50% of students' scores remaining stable within the range of 90–97 and 70–80 respectively, obtaining better teaching outcomes. However, the study did not incorporate different word embedding approaches to learn better text representation, so further exploration in this area is needed.

Funding

The research is supported by The General Projects of Humanities and Social Sciences Research in Universities of Henan Province (Project Name: Research on the Application of GPT-4 in EFL Vocabulary Learning; Grant Number: 2024-ZDJH-833).

Conflict of interest

The author reports there are no competing interests to declare.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Author contribution statement

All work for this article was completed by Lijun Cheng.

References

1. E. Bonner, H. Reinders, Augmented and virtual reality in the language classroom: practical ideas, *Teach. Engl. Technol.* **18**, 33–53 (2018)
2. T. Polić, E. Krelja Kurelović, Corpus-based vocabulary learning in technical English, *Int. J. Comput. Linguist.* **12**, 35–55 (2021)
3. L. Hongyan, A study on corpus-based EFL vocabulary teaching, *ISLLAC: J. Intensive Stud. Lang. Liter. Art Culture* **2**, 21–25 (2018)
4. L. Yang, A. Coxhead, A corpus-based study of vocabulary in the new concept English textbook series, *RELC J.* **53**, 597–611 (2022)
5. H. Lee, M. Warschauer, J.H. Lee, Toward the establishment of a data-driven learning model: role of learner factors in corpus-based second language vocabulary learning, *Mod. Lang. J.* **104**, 345–362 (2020)
6. R. Gholaminejad, M.R.A. Sarab, Academic vocabulary and collocations used in language teaching and applied linguistics textbooks: a corpus-based approach, *Terminology* **26**, 82–107 (2020)
7. L. Lowphansirikul, C. Polpanumas, A.T. Rutherford, S. Nutanong, A large English-Thai parallel corpus from the web and machine-generated text, *Lang. Resour. Eval.* **56**, 477–499 (2022)
8. J. Kirk, G. Nelson, The International Corpus of English project: a progress report, *World Englishes* **37**, 697–716 (2018)
9. P. Crosthwaite, L.L.C. Wong, J. Cheung, Characterising postgraduate students' corpus query and usage patterns for disciplinary data-driven learning, *ReCALL* **31**, 255–275 (2019)
10. H.C. Yeh, S.S. Tseng, L. Heng, Enhancing EFL students' intra-cultural learning through virtual reality, *Interact. Learn. Environ.* **30**, 1609–1618 (2022)
11. J. Cui, Application of deep learning and target visual detection in English vocabulary online teaching, *J. Intell. Fuzzy Syst.* **39**, 5535–5545 (2020)
12. F. El Jamiy, R. Marsh, Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality, *IET Image Process.* **13**, 707–712 (2019)
13. M. Liao, J. Zhang, Z. Wan, F. Xie, J. Liang, P. Lyu, X. Bai, Scene text recognition from two-dimensional perspective, *Proc. AAAI Conf. Artif. Intell.* **33**, 8714–8721 (2019)
14. Y. Xu, Y. Wang, W. Zhou, Y. Wang, Z. Yang, X Bai, Textfield: learning a deep direction field for irregular scene text detection, *IEEE Trans. Image Process.* **28**, 5566–5579 (2019)
15. T. Wang, Y. Zhu, L. Jin, C. Luo, X. Chen, Y. Wu, M. Cai, Decoupled attention network for text recognition, *Proc. AAAI Conf. Artif. Intell.* **34**, 12216–12224 (2020)
16. B. Yildirim, E.S. Topalcengiz, G. Arikan, S. Timur, Using virtual reality in the classroom: reflections of STEM teachers on the use of teaching and learning tools, *J. Educ. Sci. Environ. Health* **6**, 231–245 (2020)
17. C. Norberg, M. Nordlund, A corpus-based study of lexis in L2 English textbooks, *J. Lang. Teach. Res.* **9**, 463–473 (2018)
18. M. Siddiq, L.M.Q. Arif, S.C. Shafi, A survey research analysis of effectiveness of vocabulary learning through English vocabulary corpus, *Int. J. Educ. Pedagogy* **3**, 1–13 (2021)
19. R. Dhaya, Improved image processing techniques for user immersion problem alleviation in virtual reality environments, *J. Innov. Image Process.* **2**, 77–84 (2020)
20. S. Long, X. He, C. Yao, Scene text detection and recognition: the deep learning era, *Int. J. Comput. Vis.* **129**, 161–184 (2021)
21. Y. Meng, J. Shen, C. Zhang, Weakly-supervised hierarchical text classification, *Proc. AAAI Conf. Artif. Intell.* **33**, 6826–6833 (2019)
22. D.S. Sachan, M. Zaheer, R. Salakhutdinov, Revisiting lstm networks for semi-supervised text classification via mixed objective function, *Proc. AAAI Conf. Artif. Intell.* **33**, 6940–6948 (2019)
23. S. Minaee, N. Kalchbrenner, E. Cambria et al., Deep learning-based text classification: a comprehensive review, *ACM Comput. Surv.* **54**, 1–40 (2021)

Cite this article as: Lijun Cheng, Building a data-driven teaching platform for ESL vocabulary corpus in universities based on VR technology, *Int. J. Simul. Multidisci. Des. Optim.* **15**, 4 (2024)