Research Article

Multidimensional Psychological Model Construction of Public English Teaching Based on Deep Learning from Multimodal Perspective

Chunping Nan

Northeast Petroleum University, Daqing 163318, China

Correspondence should be addressed to Chunping Nan; ncp@nepu.edu.cn

Received 4 June 2022; Revised 27 June 2022; Accepted 30 June 2022; Published 4 August 2022

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

Public English instruction focuses on language skill development while also enhancing students’ business professional knowledge and all-around ability under the direction of public English teaching objectives. From a knowledge standpoint, public English instruction contains a lot of tacit knowledge, making it appropriate for case teaching. This paper analyses the knowledge-locking factors in traditional case teaching based on deep learning from the multimodality perspective and builds a mental model to compensate for the shortcomings of the traditional case teaching. In this paper, a multidimensional psychological model is built that overcomes the limitations of conventional approaches. It is based on the multimodal feature extraction and modal selection algorithm of deep learning and multidimensional psychology in public English teaching. The method suggested in this paper outperforms the benchmark model created for the three evaluation indices and also demonstrates the benefits of multimodality. Create a fully integrated multidimensional mental model of the teaching and activity situations so that students can learn with a passion, become more engaged in practising their English, and ultimately develop the language skills required for employment.

1. Introduction

As the social economy continues to grow, various jobs have introduced a number of new and more demanding requirements for the professional abilities of talents. Institutions of higher learning have also taken the initiative to review the talent training programmes for various majors in accordance with the demands of social development and have demanded ongoing reforms to college English teaching in order to provide more ideal services for the training of professional talents, enhance student employment, and foster a significant number of demand-oriented talents for the society [1, 2]. With the vigorous promotion of teaching reform in higher education institutions in recent years, some progress has been made in the process of practise, but there are still some issues with the teaching content, methodologies, and talent team, which significantly impede the healthy and long-term development of higher English teaching. Public English instruction should improve students’ business professional knowledge and all-around ability while also focusing on the development of language skills in accordance with the objectives for public English teaching [3]. Public English instruction is suitable for case teaching because, from the standpoint of knowledge, it contains a lot of tacit knowledge. Explicit knowledge can be acquired through a clear reasoning process as a result of the systematization and standardization of that knowledge. The majority of knowledge acquired through traditional teacher instruction or student’s independent study falls under the category of explicit knowledge. A type of educational activity that is based on teacher input and motivated by student output is called public English teaching. The output effect of students is used to assess teaching effectiveness.

This paper examines the factors that cause knowledge locking in traditional case teaching based on in-depth learning, constructs the external connection of case teaching, suggests a multidimensional and multidimensional teaching model, and develops a psychological model to address the
shortcomings of traditional case teaching from the perspective of multimodality [4]. The issue of historically separating feature learning [5, 6] from feature selection is now resolved by the use of deep learning to extract features. We can directly learn high-level abstract features with semantic properties from the original data using the learning model with deep structure. Additionally, we can directly learn abstract semantic features with smaller feature dimensions by using the deep model to directly achieve the effect of dimension reduction while learning features. This model is essential for avoiding knowledge lock-in and dispelling common misconceptions about case teaching. Multimodality describes different pairings of two or more modes. Modality and multimodality are further defined as a specific type of information or a representation of that information, respectively. A study or dataset has multimodal attributes when it contains multiple modes. An inevitable by-product of the advancement of multimodal machine learning to this point is multimodal deep learning. Multimodal deep learning advances multimodal machine learning [7–9] with the deep learning method and has made significant strides. Multimodal deep learning inherits the learning tasks and learning objectives of prior multimodal machine learning.

Deep learning from a multimodal perspective investigates public English teaching, creates a multidimensional mental model, incorporates traditional case teaching into the following external knowledge networks, and unlocks locked knowledge channels by fusing together the complementary resources of this major, other majors at our school, off-campus resources, and network technologies, thereby fostering the transmission and transfer of heterogeneous tacit knowledge and improving. The goal of college English instruction is to improve students’ English abilities. Teachers can create flexible, multidimensional psychological models for the teaching process, implement the fundamental principle of teaching students in accordance with their aptitudes, stimulate the initiative of both teaching and learning effectively, and encourage students to engage in active learning and personalized learning according to the differences between the various teaching contents under deep learning [10]. In addition to realizing explicit knowledge sharing, the multimodal perspective of deep learning teacher cooperative teaching also recognizes the flow of tacit knowledge through interactions and communication between colleagues. Interdisciplinary educators have a variety of subject specialties and modes of thought. Some educators frequently dive deep into projects to conduct research and analysis and have specific practical experience. They play a significant role as dispersed knowledge carriers [11]. Strengthen situational instruction for students, make an effort to create a good professional environment for their learning based on deep learning, and fully integrate multidimensional mental models of teaching and activity situations so that students can be fully interested in learning and improve their interest by actively participating in practise, thus developing English skills required for employment.

I primarily present the following innovations in this paper.

(1) This paper constructs a multidimensional mental model. The multidimensional mental model is often used in picture text annotation. Firstly, an encoder is used to convert the input image into multiple vectors, each vector corresponds to an area of the image, and one vector is selected to align with the target sentence words, and the probability of other vectors aligning with the target sentence words is set to zero.

(2) The correlation between the multidimensional mental model theory and public English teaching is discussed. All public English teaching modes take the teaching collective as the main body, and through formal or informal channels, they exchange knowledge with the inside and outside of multidimensional mental models. By integrating the complementary resources of this major, our school, off-campus, and online channels, we can break through the locked knowledge channels and promote the transmission and transfer of tacit knowledge, thus improving the effect of case teaching.

2. Related Work

This chapter introduces the related work of multidimensional psychology in public English teaching, analyze the advantages and disadvantages of the methods involved, and then lead to the work content of this paper. The work done in this paper is based on the theoretical and practical basis provided by these existing studies.

Ramachandran et al. put forward that in the process of carrying out multidimensional psychological reform of English Teaching in institution of higher learning Based on employment orientation, active cooperation is required from all aspects, such as the importance attached to it by the leaders of the University, the strong support provided by the academic affairs office, and the cooperation of each college with English teaching. Therefore, it is very necessary to carry out multidimensional psychological management of English teaching in institution of higher learning through diversified teaching management models [12]. Huang et al. put forward the multidimensional psychology of web-based case public English teaching, which not only broke the bottleneck of traditional fixed class teaching in space and time, but also broke through the time and space limitations of knowledge spillover [13]. Bayoudh et al. put forward that in the multidimensional psychological process of public English teaching, teachers always put students in the central position. Specifically, they can choose the integrated teaching method of "teaching, learning, and doing" to effectively mobilize students’ initiative and enthusiasm in teaching. In terms of learning arrangement, they should give consideration to the extracurricular autonomous learning and classroom learning, and effectively strengthen the guidance to students so that they can successfully carry out extracurricular English learning, so as to cultivate their ability of autonomous learning [14]. Adwaith et al. put forward that compared with traditional case teaching, multidimensional case teaching in public English teaching is conducive to exploratory learning. Compared with traditional case
teaching mode, multidimensional case teaching is the integration of multiple knowledge resources in multiple situations. It requires learners to constantly learn to think, analyze, seek, discover, and use new knowledge in a larger range, so as to stimulate and promote the cultivation of learners’ innovative ability [15]. Aceto et al. proposed to support the multidimensional psychology of public English teaching in local schools through obligation, which not only improved the allocation of resources, but also interacted and communicated with teaching groups through multidimensional psychological teaching in cooperation with schools, which accelerated the flow and transformation of heterogeneous knowledge, thus realizing the spiral rise of knowledge [16]. Born et al. proposed that in the multidimensional psychological process of professional English module teaching, in view of the practical characteristics of students’ willingness to learn in a situational teaching atmosphere, English practice teaching should be emphasized in the teaching process to reflect the cultivation of students’ professional English application ability [17]. Scherer et al. put forward that most of the university teachers have been in the front line of teaching and scientific research for a long time, lacking first-hand contact with society and enterprises, and relatively lacking in “practical” knowledge. By actively absorbing these institutions and actively participating in case teaching, it is not only conducive to improving the multidimensional psychological ability of professional teachers in practical teaching, but also enables case teaching to be embedded in the external knowledge network [18]. Mitra et al. proposed to strengthen the guidance for students to learn strategic and empirical knowledge, on the one hand, to teach students how to apply multidimensional English psychology; on the other hand, make them master learning methods, so that students can learn how to better use professional English [19]. Seidler et al. proposed that in the past, the acquisition channel of students’ public English teaching knowledge was relatively single, and most of it came from teachers’ teaching, that is, teachers first built relevant multidimensional psychological models for students, on this basis, revised the understanding of new knowledge, and continuously supplemented and integrated [20]. Lui et al. proposed that the multidimensional psychological content of public English teaching should be reformed under the guidance of employment, which is the key to the reform English teaching. When setting the multidimensional psychological content of teaching, we should emphasize the adaptability to the job. The higher the adaptability and the more scientific the content selection, the more consistent it will be with the direction of curriculum reform. At the same time, it will enable students to smoothly adapt to the job requirements in the future [21].

This paper investigates multimodal deep-learning-based public English teaching and builds a multidimensional mental model. The mental model has the following significant cognitive characteristics: the basic organisational form of knowledge in the mind, or mental model, is the knowledge structure of cognitive subjects. This abstract knowledge structure is made up of large and small classes of knowledge, or “knowledge sets,” of abstract knowledge. To prevent knowledge lock-in and dispel the widespread misconception of case teaching, multidimensional case teaching combined with multimodal deep learning is essential in public English education. Deep learning has a significant positive impact on the development and enhancement of students’ comprehensive ability as well as the enlargement of their business knowledge. Multimodal teaching significantly compensates for the shortcomings of traditional case teaching. According to the multidimensional psychological model used in the deep learning case study for teaching public English, cooperative teachers share knowledge with one another, communicate with one another, and discuss topics related to the course as they are teaching it. This allows them to make up for the gaps in the professional knowledge of business English teachers and fill in any knowledge gaps. Since the development of the multimodal mental model theory, linguists have started to use the multidimensional mental model theory to investigate pragmatic reasoning in various communicative contexts with the goal of conducting in-depth research on human communicative activities and improving human communication. In addition to the selection of the negotiation team members, the responsibilities of each team member and the standards of excellence for team members should all be included in the psychological model that teachers share with students in practical teaching activities. In the course of continuous learning and communication, teachers from various fields will establish a regular system of in-depth communication that will encourage their heterogeneous knowledge to be shared covertly, enhancing and promoting the development of cooperative teachers’ abilities and enabling the internalisation of knowledge.

3. Multimodal Deep Learning

Numerous research institutions are continually exploring and innovating their current models, improving datasets, speeding up the operation of multimodal deep learning models, and enhancing the precision of output prediction because this is a research area of deep learning with significant development potential. The term “single-mode deep-learning representation” refers to the linear or nonlinear mapping of single-mode data to produce high-order semantic feature data [22]. The primary processing objects in the single-mode representation are sentences, images, videos, sounds, and other modes. Different modes have various neural network models that are applicable to them. Multimodality has the ability to judge information obtained accurately based on its memory, but it is challenging to accurately describe multimodal perception in language. Multimodal representation is the processing of modal information using modal interaction semantic representation or modal constraint semantic representation, resulting in representational outcomes that are the same or similar for modal information with the same or similar semantics [23]. College English teaching that integrates listening, speaking, reading, writing, and translation through multimodal deep learning is advantageous to the natural connection between output and input. To put it in another way, multimodal deep
learning teaching encourages the limitless and zero distance interaction between new knowledge acquired through multimodal interaction and language materials based on long-term memory or existing input knowledge, so that students can combine or design practical situations to do things in English based on reading texts and listening materials, or to achieve multimodal information processing and decoding with the help of multimodal deep learning teaching.

3.1. Unimodal Representation. Three categories can be used to categorize the representation of sentence modes: the bag representation of word sequence modes, the low-dimensional representation of word sequence modes, and the unique representation of word modes. The representation of sound modes is divided into the extraction of sound feature vectors and the high-order representation of extracted feature vectors. The representation of visual modality under deep learning is divided into the representation of image modality and representation of video modality. Many datasets gathered in the real world exhibit high-dimensional unstructured characteristics in the conventional single-mode environment. These original data from in-depth learning cannot be used directly for model calculation; instead, they must first go through the data cleaning process, in which, the unstructured original data is structurally processed, the invalid data is removed, and the original data is represented in a form that can be used for calculation. A further requirement of this modal representation is that the data language be able to reflect the semantic information of words or words [24, 25]. The unique representation of word mode only performs a simple vectorization replacement of the words or words contained in the sentence mode. Word sequence modes are represented in low-dimensional space. The semantic representation of word sequence modality, or mapping the word sequence into the semantic space, is referred to as the low-dimensional space representation of word sequence modality. The conventional approach to dealing with issues arising from high-dimensional data in in-depth learning is to first use manually created feature extraction rules for single feature extraction, deal with issues arising from too high dimension based on these extracted primary features and deal with issues arising from too high dimension by using a method based on feature space mapping changes for dimension reduction. Convolutional neural networks, a type of deep learning technique created specifically for images on the foundation of multilayer neural networks, have produced outstanding results in image processing. The structure of neural network for obtaining video modal representation is shown in Figure 1.

Convolutional neural network processes a continuous image frame as video. It uses a convolutional neural network to fuse the time and space information in this continuous image and generates the vector representation of this continuous image frame before the full connection layer of the convolutional neural network. The depth map in the image is transformed into 3D point cloud data by coordinate transformation, and the 3D points generated by the depth map are set together, and the covariance matrix Cov of the following points $\rho$ is decomposed by eigenvalue.

$$\text{Cov} = \frac{1}{k} \sum_{i=1}^{k} \begin{bmatrix} \rho_i - g \\ \rho_i - g \\ \rho_i - g \end{bmatrix}^T \begin{bmatrix} \rho_i - g \\ \rho_i - g \\ \rho_i - g \end{bmatrix},$$  \hspace{1cm} (1)

where $g$ is the three-dimensional centroid of $k$ nearest points of $\rho$, $g = 1/k \sum_{i=1}^{k} \rho_i$.

For the simplicity of calculation, the consistency of normal direction is adjusted.

$$n^*_p = n_p \cdot \text{sign}(n_p, \rho),$$  \hspace{1cm} (2)

where $(n_p, \rho)$ represents the inner product between $n_p$ and $\rho$, which reflects the angle between them, and the symbolic function sign is used to adjust $n_p$ to be consistent with $\rho$ direction.

When the original data dimension of deep learning is too high and the number of training datasets is huge, the amount of computation is unacceptable, and these methods are difficult to be applied in a big data environment. Moreover, this kind of method is still based on the primary feature extraction of the original unstructured data, and its algorithm itself cannot be directly applied to the original data.

3.2. Multimodal Representation. A single-modal representation serves as the foundation for multimodal representation, and connecting all modal representations in series is the quickest and the most straightforward way to do so. As multimodal research has grown in popularity recently, the technique for obtaining multimodal representation has advanced quickly. Modal interaction semantic representation and modal constraint semantic representation make up the multimodal representation of deep learning. Constrained semantic representation differs from coordinated semantic representation in that the representations of other modalities are constrained by the single-modal representation of one modality in order for those representations to contain the semantic data of that modality. The deep learning attention model’s models are continuously updated throughout the development process. On the basis of the global attention model, the local attention model, the static attention model, and the dynamic attention model were developed. The heterogeneity of various data modalities in the deep multimodal feature learning task is one of the most significant issues. The primary goal of the majority of multimodal learning methodologies is to project different modalities into the same subspace, despite the fact that the original data of various modalities are in various feature spaces. In revealing the implicit hierarchical feature expression and transforming the original data, the deep learning model demonstrates outstanding advantages. This paper builds a multidimensional mental model of public English teaching based on deep learning from a multimodal perspective, and it carefully examines the connections between each element in the input mode and a specific element.
in the target mode. In Figure 2, the multidimensional mental model is displayed. The multidimensional mental model is often used in picture text annotation. First, the encoder is used to convert the input image into multiple vectors, each vector corresponds to an area of the image, a vector is selected to align with the target sentence word, and the probability of aligning other vectors with the target sentence word is set to 0.

It is assumed that there is a classless input vector $x \in \mathbb{R}$, which is mapped by the nonlinear activation function $f_\theta$ to obtain the hidden layer representation $y \in \mathbb{R}$.

$$y = f_\theta(Wx + b),$$

(3)

where network parameters $\theta = \{W, b\}$ and $f_\theta(x) = 1$ are Sigmoid functions. $W \in \mathbb{R}$ is the weight matrix and $b \in \mathbb{R}$ is the offset vector. Map again to reconstruct the vector $z \in \mathbb{R}$

$$z = f_\theta(W_y + b).$$

(4)

In order to reduce the number of parameters to be trained, the network parameter $\theta = \{W, b\}$ can be limited to $W = W^T$, where $W^T$ is the transpose of the weight matrix $W$, and the network parameters can be adjusted through training.

For the average activation of all training samples, there are

$$\rho_j = \frac{1}{m} \sum_{i=1}^{m} \rho_j.$$

(5)

For the sparsity of data representation, the sparsity limiting parameter $\rho$ is added, and $KL$ relative entropy is used as the penalty factor, which can be expressed as

$$KL(\rho | \rho_j).$$

(6)

Therefore, after implementing the sparsity restriction and adding the penalty factor, the overall cost function $J$ can be expressed as

$$J_{sparse}(W, b),$$

(7)

where, $W$ is the weight matrix and $b$ is the offset vector. The input training dataset is trained, and the weight matrix $W$ and the offset vector $b$ are adjusted to make the formula reach a minimum.

In order to get a coded representation that can reveal the internal structure of the data as much as possible, the optimization goal is the reconstruction error between the original input and the reconstructed input.

$$\text{Loss}(x, x') = \|x - x'\|.$$

(8)

When the whole network is fine-tuned under supervision, the back propagation algorithm is used to optimize the model. Define the loss function as

$$\infty = -\sum_{j} \sum_{i} \log$$

(9)

where $m$ represents the number of modes, and $N$ represents the number of training samples. For $k$ classification problems, the probability that the input vector $h_m$ belongs to the category $i$ is

$$p(y = y^{(i)} | h_m, T, b_{root}).$$

(10)

Only the parameters of one subnetwork are adjusted at a time, and the parameters of the other subnetworks are fixed. After the weight update is completed, the network to which the next mode belongs is adjusted until all modes are adjusted. The auxiliary layer is only used for joint fine-tuning of

Figure 1: Get the neural network structure diagram of video modality representation.
the root network, and it will be cancelled after the network training is completed.

3.3. The Correlation between Multidimensional Psychological Model Theory and Public English Teaching. The cognitive context of communication between teachers and students includes public English instruction. In communication, teachers and students are on an equal footing as speakers and listeners, respectively. Teaching activities have the communicative effect of allowing teachers and students to share a multidimensional psychological model. The primary body in the public English teaching mode is the teaching collective, which engages in formal and informal knowledge exchanges with both the inside and outside of the multidimensional psychological model. To increase the impact of case teaching, complementary resources from the major, the school, off-campus locations, and network channels are integrated. This opens up previously locked knowledge channels and encourages the transmission and transfer of tacit knowledge. 

On the one hand, English teachers should be able to instruct students in basic English, and on the other, they should be able to instruct students in a variety of majors. We should actively establish a specialized team of teachers in order to increase quality and work efficiency and lessen the pressure on teachers' professional teaching. In the past, there were only a few ways for students to learn, and the majority of those ways were through their teachers. Specifically, teachers would build relevant psychological models for their students, revise those models in light of new information, and constantly supplement and integrate their learning. The public teaching department or the foreign language institute is primarily in charge of the traditional multidimensional psychological model, and the public foreign language teaching and research office is primarily in charge of carrying out the related teaching work. The multidimensional psychological model unlocks the knowledge network that has been impenetrable. Students can master the most recent management, finance, economy, and legal knowledge and comprehend the most recent trends in international business activities by being exposed to new heterogeneous knowledge from the outside environment. In addition to helping to avoid knowledge locking and provide the necessary outside knowledge for the creation of collective knowledge, it can also aid in removing the case teaching bottleneck.

In the course of implementing the reform, the reform group frequently discusses how to best advance the reform of English teaching, identifies problems as they arise and proposes solutions, chooses the most practical solutions through mutual communication, and successfully fortifies their division of labour and cooperation in order to successfully advance the English teaching reform over time. In order to raise the professionalism of English teachers, specific training and education should be implemented concurrently. On the other hand, it is urged that teachers continue their education and learning, particularly through distance learning or by taking graduate courses while still working full-time. Furthermore, we can strengthen the talent pool of English teachers in schools and take the lead within the entire teaching team by bringing in high-calibre English teaching talent. Teachers can use the multidimensional psychological model to divide their class into groups and group members when teaching public English. The students who are the best at organizing their groups are chosen as group leaders by each group. Each member of the group is given a specific task by the group leader, who also monitors how well they are completed. To ensure group
completion and organise class presentations, teachers only need to regularly locate the group leader. Once students’ assignments have been completed, teachers should honestly and objectively evaluate their performance and assign grades as evidenced. Students today have stronger personalities because they are in college. The growth of the Internet has allowed students in the modern age to quickly find out a lot of information, though it might be superficial. To guide, modify, deepen, and connect them into a shared model between teachers and students, teachers in public English instruction should first understand the basic and dispersed knowledge structure of students’ prior relevant knowledge.

4. Analysis and Discussion of Results

In order to verify the multimodal fusion feature extraction ability of the model, the method in this paper and NUS-object are used to test the discrimination ability of each single mode, and the results are shown in Table 1. To evaluate the discrimination ability of multimodal fusion features extracted by the proposed model, this paper first gives the classification accuracy obtained by applying different algorithms to experimental datasets, as shown in Tables 1 and 2.

The experimental results from Tables 1 and 2 show that compared with other algorithms, this method has the best classification accuracy in the image classification dataset. For NUS-WIDE-object dataset, it is not difficult to find that the classification accuracy of features using text modality is higher than that using image modality. The features of the six modes of the dataset all come from images. The spliced feature vectors can be classified by SVM classifier and PCA+SVM, and the classification accuracy is higher than that obtained by a single feature. The classification accuracy obtained by multicores learning method is low. However, the classification accuracy of the model proposed in this paper is still the highest among several methods.

A good feature should not only have good discrimination ability, but also keep as much information as possible in the original data in the lowest dimension, so this paper also tests the feature dimension reduction ability of the model for the original multimodal data. The classification accuracy results on the obtained experimental data are shown in Figure 3.

As can be seen from Figure 3, the model proposed in this paper can extract low-dimensional features from the original high-dimensional features, effectively integrate a variety of original different features, and have low dimensions while maintaining strong discrimination ability. On the NUS dataset, the proposed model can reduce the original 745 dimension features to 300 dimensions, and still has the highest classification accuracy. This experiment uses literature [8], literature [10], and this method to study the recall rate of multidimensional psychology in public English teaching. The experimental results are shown in Figure 4.

As can be seen from Figure 4, among the three methods, the recall rate of the method proposed in this paper is the lowest, followed by literature [10], and finally literature [8]. As can be seen from Figure 4, when the number of experiments reaches 120, the recall rate of literature [8] is 87.48%, that of literature [10] is 83.56%, and that of this method is 63.25%. It shows that in small-scale datasets, pretraining the model on large-scale similar datasets can effectively help the model learn the underlying features of the picture, while fine-tuning can make the model better adapt to specific task datasets, thus improving the model performance and generalization ability. In this experiment, literature [8], literature [10], and this method are used to study the change of F value of multidimensional psychology in public English teaching. The experimental results are shown in Figure 5.

As can be seen from Figure 5, when the number of experiments reaches 150, the recall rate of literature [8] is 98.25%, that of literature [10] is 93.25%, and that of this method is 70.15%. On the whole, the method proposed in this paper is superior to the designed benchmark model among the three evaluation indexes, and it also shows the advantages of multimodality. Firstly, the experiment verifies the feature extraction ability of the proposed multimodal neural network algorithm framework for heterogeneous features, as shown in Table 3 and Table 4.

From the experimental results in Tables 3 and 4, it can be clearly seen that the algorithm proposed in this paper is obviously superior to the other methods. At the same time, the classification effect obtained by using multiple data modalities is obviously better than that obtained by using only a single-data modality, which also confirms the effective utilization of complementary information in multimodal data by the proposed model. However, the classification accuracy obtained by training SVM classifier with all the refined feature dimensions alone is higher than other methods, but slightly lower than that obtained by feature screening, which shows that the proposed method can effectively remove redundant information without losing useful information.

In order to verify the effectiveness of the proposed model in modal selection, this paper observes the modal weight information obtained by different methods. The weight of each mode, but other methods will get a weight for each feature dimension. The weights of each mode are obtained on three experimental datasets by using the methods of literature [8], literature [10], and this paper. The experimental results are shown in Figure 6.

The method suggested in this paper can successfully filter out noise modes and useless modes, as shown by Figure 6. The proposed method applies to the three experimental datasets and assigns the appropriate weights to each mode while assigning zero weights to some noise modes and higher weights to some modes associated with the current classification task. Because different data modes are heterogeneous, the other methods, in contrast, assign incorrect weights to various data modes, we can use the conventional methods to produce the right results only by successfully removing this heterogeneity. It is decided if the modal weight is not zero after obtaining the weight for each independent mode. A series of experiments are conducted to see if this weight can accurately depict the mode’s significance to the current learning task. Figure 7 illustrates how the classification accuracy varies with the addition of various modes.
As can be seen from Figure 7, for the features extracted from the network, even only a few feature groups are needed to achieve high classification accuracy, which also confirms the existence of information redundancy in multimodal data. However, with the addition of 0-weighted feature groups, the classification accuracy basically remains unchanged or fluctuates slightly, and the features of these groups have no effect on the current classification task. Moreover, after the multimodal neural network proposed in this paper is extracted, the modes originally in the noise group have no effect on the classification task, but have no negative impact, which proves the effectiveness of the proposed model again.
Table 3: Classification accuracy of datasets classified by using unimodal features.

<table>
<thead>
<tr>
<th>Mode/method</th>
<th>LSS</th>
<th>RGSIFT</th>
<th>SIFT</th>
<th>SURF</th>
<th>Da CAF</th>
<th>CQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.4942</td>
<td>0.5021</td>
<td>0.4087</td>
<td>0.5272</td>
<td>0.8033</td>
<td>0.3921</td>
</tr>
<tr>
<td>MMNN + SVM</td>
<td>0.5506</td>
<td>0.5313</td>
<td>0.4380</td>
<td>0.5643</td>
<td>0.8402</td>
<td>0.4772</td>
</tr>
<tr>
<td>Mode/method</td>
<td>Gaussian</td>
<td>Uniform</td>
<td>Chi2</td>
<td>F-dist</td>
<td>Beta</td>
<td>LSS + N</td>
</tr>
<tr>
<td>SVM</td>
<td>0.1063</td>
<td>0.1121</td>
<td>0.1057</td>
<td>0.1063</td>
<td>0.1121</td>
<td>0.3162</td>
</tr>
<tr>
<td>MMNN + SVM</td>
<td>0.1071</td>
<td>0.1102</td>
<td>0.1163</td>
<td>0.1088</td>
<td>0.1071</td>
<td>0.4025</td>
</tr>
</tbody>
</table>

Table 4: Classification accuracy of datasets classified by using unimodal features.

<table>
<thead>
<tr>
<th>Mode/method</th>
<th>CH</th>
<th>CORR</th>
<th>EDH</th>
<th>WT</th>
<th>CM</th>
<th>LDA31</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.2425</td>
<td>0.3065</td>
<td>0.2893</td>
<td>0.3062</td>
<td>0.2856</td>
<td>0.5061</td>
</tr>
<tr>
<td>MMNN + SVM</td>
<td>0.3031</td>
<td>0.3737</td>
<td>0.3073</td>
<td>0.3732</td>
<td>0.3393</td>
<td>0.5432</td>
</tr>
<tr>
<td>Mode/method</td>
<td>Gaussian</td>
<td>Uniform</td>
<td>Chi2</td>
<td>F-dist</td>
<td>Beta</td>
<td>CH + N</td>
</tr>
<tr>
<td>SVM</td>
<td>0.1042</td>
<td>0.1043</td>
<td>0.1022</td>
<td>0.1066</td>
<td>0.1053</td>
<td>0.2222</td>
</tr>
<tr>
<td>MMNN + SVM</td>
<td>0.1051</td>
<td>0.1051</td>
<td>0.1051</td>
<td>0.1051</td>
<td>0.1051</td>
<td>0.2237</td>
</tr>
</tbody>
</table>

Figure 6: Weights of different modes obtained by different methods on different datasets.

Figure 7: Variation of classification accuracy with the addition of different modes. (a) Awa dataset, (b) Article method dataset, and (c) NUS wide object dataset.
5. Conclusions

The current teaching reform in higher education, in a nutshell, includes the curriculum teaching reform. As a defining feature of its instruction, the higher education curriculum always adheres to an employment orientation. In this paper, a multimodal feature extraction and modal selection algorithm based on the multidimensional psychological depth learning and public English teaching from the multimodal perspective is proposed to develop a multidimensional psychological model, which overcomes the limitations of conventional approaches. When 150 experiments have been conducted, this method has a recall rate of 70.15 percent, compared to 98.25 percent for literature 8 and 93.25 percent for literature 11, respectively. Overall, the method suggested in this paper outperforms the designed benchmark model in each of the three evaluation indices, and it also demonstrates the benefit of multimodality. With the help of the deep learning method, the multimodal data is imaginatively transformed into a modal independent abstract representation, and the structural sparse method is then used to further select the feature dimensions in the abstract representation. The future employment of college students should be the driving force behind English instruction in institutions of higher learning. Only through methodical, all-encompassing reform and innovation can English teaching in various majors be managed better over time, and the calibre of instruction raised to foster a large number of talents for the society. In addition to traditional case teaching, multidimensional case teaching based on in-depth learning is also used. By putting more emphasis on multidimensional case teaching, the traditional case teaching is not being undermined. In order to ensure that case teaching is effective, teachers should fully utilize their positive roles in accordance with the business English teaching content, training objectives, and available resources.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author does not have any possible conflicts of interest.

Acknowledgments

This study was supported by Research on Chinese College Students’ Vocabulary Application Ability in English Writing Based on English Corpus, No. 2020YDQ-15.

References


