International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 1368-1382

https://doi.org/10.70517/ijhsa463106

Optimization and Interaction Enhancement of English Vocabulary Learning Paths Based on K-Nearest Neighbor Algorithm in Multi-task Learning Framework

Qingqing Zhang[1,*]
[1] Pingdingshan Polytechnic College, Ministry of Culture and Education, Pingdingshan, Henan, 467001, China
Corresponding authors: (e-mail: 15837565597@163.com).

Abstract With the continuous development and power of big data and artificial intelligence technology, the application of artificial intelligence technology in English learning is becoming more and more popular. In this paper, based on the knowledge of multi-task learning theory, the optimal learning path recommendation method based on K nearest neighbor algorithm is designed in order to enrich the current English teaching path. It is found that the English vocabulary learning path still has the problem of low degree of interaction, accordingly, a sentiment prediction model based on principal component analysis and K nearest neighbor algorithm is constructed, and the model is verified and analyzed. The prediction accuracy of the traditional K-nearest neighbor algorithm is 77.73%, while the prediction accuracy value of the model in this paper is 96.03%, which indicates that the introduction of principal component analysis algorithm on the basis of the traditional K-nearest neighbor algorithm can improve the prediction accuracy of students' emotions, and enhance the interactive effect of teaching by accurately capturing students' emotional state.

Index Terms multi-task learning, K-nearest neighbor algorithm, principal component analysis, English vocabulary learning

I. Introduction

With the accelerated development of globalization, English has become one of the global languages and a bridge for transnational communication and cultural dissemination in various countries [1], [2]. Whether in study, work, life or entertainment, the influence and penetration of English is indispensable [3]. Therefore, learning English has become a necessary skill in today's society, and an important part of it is English vocabulary learning [4].

The importance of vocabulary learning cannot be overstated for any non-native English learner, and it runs through the whole process of language learning [5]. A certain amount of vocabulary is the basis for the smooth performance of language activities such as listening, speaking, reading, writing and translating [6]. The vast majority of English learners recognize the importance of vocabulary learning, but traditional vocabulary teaching methods for learning vocabulary reveal many misunderstandings, such as many learners think that learning words means remembering their Chinese meanings, and that learning words can only be done by rote memorization, etc. [7]-[10]. In fact, vocabulary learning requires learners to be able to learn according to their own learning needs, to plan and selectively use learning methods suitable for themselves, and to achieve the best learning results by enhancing the practical interactions between classroom and after-class with teachers and classmates, and self-monitoring of the learning process [11]-[14]. With the development of artificial intelligence, the application of intelligent algorithms such as K nearest neighbor algorithm in English vocabulary has gradually been paid attention to, and the algorithms have an important reference value for the optimization of English vocabulary learning paths and the enhancement of interaction [15]-[17].

In this paper, through the theoretical study of multi-task learning framework, we have learned the advantages of multi-task learning in the process of English vocabulary learning. Firstly, through the form of questionnaires, students' English learning feature data are obtained, and their data are normalized due to the problem of different scales. By calculating the similarity and K value of the feature data, the optimal learning path recommendation is realized. Aiming at the problem of low interaction effect in the learning path, an emotion prediction model based on principal component analysis algorithm and nearest neighbor algorithm is designed to formulate a reasonable teaching plan by accurately capturing students' emotional state in the classroom, aiming to improve the interaction effect in the English teaching classroom.

II. An Exploration of English Vocabulary Learning under the Multi-task Learning Framework

II. A. Multi-task learning framework

Multi-task learning refers to the situation where the model learns while dealing with multiple tasks at the same time, which can significantly improve the learning efficiency and generalization ability of the model [18]. Compared with single-task learning, multi-task learning is more in line with the way human beings learn and more applicable to real-world scenarios. For example, in the field of natural language processing, a model can simultaneously perform multiple tasks such as sentiment categorization, text generation and machine translation. In the field of computer vision, a model can simultaneously perform multiple tasks such as target detection, image classification and image segmentation.

Traditional machine learning approaches optimize for specific metrics of a single task, usually by training models through hyperparameter tuning. However, this approach ignores information in the training data related to the task in question that can help improve model performance. To utilize this information, a new approach called multi-task learning is proposed. Multi-task learning aims to jointly learn multiple different but related tasks to maximize the efficiency of the model. This is done by sharing information between different tasks so that each task benefits from the others. In classification problems, multi-class classification refers to classification with more than two classes, while multi-label classification involves assigning a target set of labels to each instance. Although both concepts have the prefix “multiple”, in multitask learning we refer to multilabel classification.

II. A. 1) Structure of multi-task learning

The two learning modes commonly used for multi-task learning (MTL) in deep learning are hard parameter sharing and soft parameter sharing [19]. In hard parameter sharing, multiple tasks share parameters through multiple shared hidden layers, while their respective tasks are executed independently through their respective output layers. In soft parameter sharing, different tasks use different network layers to train parameter models belonging to their respective tasks and constrain each other by constraining hidden layers so that the corrected parameter models are similar. Hard parameter sharing is the most commonly used model for multi-task deep learning, which allows the shared parameters to be forced to be generalized to all tasks, reducing the risk of overfitting for each specific task. The more parameters shared, the lower the probability of overfitting and vice versa. In soft parameter sharing, each task has its own set of weights and biases, and the distance between these parameters needs to be adjusted in different models so that the parameters are similar between different models. The key to multi-task learning is how to design an appropriate loss function so that the model can perform well on multiple tasks. A common approach is to construct multiple loss functions, with one task corresponding to one loss function, and then weight and sum them as the loss function for multitask learning. Another method of loss function optimization in the algorithm, due to the different tasks with different learning rates, loss magnitude and complexity, it is difficult to achieve consistency in training, the loss function optimization algorithm can balance the training learning rate or loss magnitude, and this method can improve the model's performance and generalization ability, and is not prone to the phenomenon of overfitting.

II. A. 2) Loss Functions for Common Multitask Learning

Multi-task learning is a machine learning method designed to allow a model to solve multiple related tasks simultaneously. In multi-task learning, there are several common loss functions:

(1) Weighted loss function: for each task, different weights can be specified to weight the loss functions of different tasks according to their importance. For example, cross entropy can be used as the loss function and then a weight can be specified for each task.

(2) Balanced loss function: for each task, the loss function can be adjusted to achieve balance among different tasks.

(3) Hierarchical loss function: For each task, a different loss function can be used. For example, cross entropy, mean square error, or other loss functions can be used for each task, and these loss functions can be used as part of the hierarchical structure of the model.

II. A. 3) Advantages of multi-task learning

In the field of wireless perception, it would be cumbersome to use a single-task learning model if it is necessary to obtain the results of multiple tasks such as the identity, movement, and location of a person in the experimental environment at the same time. This requires splitting multiple tasks into multiple mutually independent individual tasks, constructing models for each task separately, and outputting them in parallel sequentially. Compared with single-task learning, multi-task learning has the following advantages:

(1) Improved model generalization ability. Multi-task learning can learn the common features between tasks through the hard-sharing mechanism, thus improving the generalization ability of the model, and at the same time, multi-tasking can avoid overfitting and reduce the complexity of the model.

(2) Improve training efficiency. Compared with single-task learning in which feature extraction and model training are performed individually for different tasks, multi-task learning can learn multiple tasks at the same time, which reduces repetitive operations, training time, and the waste of computational resources.

(3) Improve data utilization. Multi-task learning can make full use of the data of each task, improve data utilization, and further improve the performance and generalization ability of the model.

(4) Strong migratability. Multi-task learning can learn common features between tasks through the hard parameter sharing mechanism, and these features have good transferability and can be applied to other related tasks.

II. B.English Learning Path Design Based on K Nearest Neighbor Algorithm

Under the multi-task learning framework, combined with the K nearest neighbor algorithm, it can accurately analyze the learning interest and knowledge level of each college student, so as to create a personalized learning path for them. In this paper, through the characteristics of the students' English vocabulary learning features to predict the learning path, we determine that the KNN algorithm, which does not depend on the classification boundaries, is more suitable for the study of the classification method of the characteristics of the English vocabulary learning, and finally, based on the weighted KNN algorithm Finally, the optimal path recommendation method for English vocabulary learning is designed. At the same time, it can monitor the learning situation of college students in real time and continuously optimize the English vocabulary learning path according to their reading time, reading volume and correct rate, helping them to build up self-confidence and stimulate learning interest in the learning process.

II. B. 1) Data collection

The main methods of data collection are experimental method, observation method, network information collection method and survey method. Experimental method refers to obtaining information by designing and implementing experiments. Observation method refers to on-site research and recording of data through field sampling, in-depth site, meeting and other ways. Network information collection method refers to the process of obtaining data by searching, integrating and screening network information resources. Survey method refers to the method of investigating through telephone, meeting, mail, questionnaire and so on.

Due to the privacy of English learning information in colleges and universities, this paper chooses the method of questionnaire survey for the data collection of students' English vocabulary learning characteristics. Classification questions need to correspond to a label for each sample, so when developing the questionnaire, it is necessary to design questions related to the learning path way that students may choose. Therefore, the questionnaire survey includes the collection of students' English vocabulary learning characteristics and the collection of labels.

In order to facilitate the implementation and statistics of the questionnaire, the English vocabulary learning features are categorized and the discrete feature variables are classified according to the attributes of the variables. For example, the attribute features of vocabulary are categorized into verbs and nouns, etc. Such categorization has no difference in degree and order among the variables and is incompatible with each other, which is called unordered categorical variables denoted as X^{uno}. For continuous type feature variables, they are discretized by the equal width method. Such differences in order and degree among categorical variables are called ordered categorical variables and are denoted as X^{con}.

II. B. 2) Data processing

The data of students' English vocabulary learning features have different magnitudes, such as reading time and reading volume features, which have different magnitude units and large value gaps, and since KNN is a classification decision through the distance between the feature vectors, too large a gap between the feature values will affect the data classification results. Therefore, it is necessary to standardize the data so that the indicators of each traveler feature are in the same order of magnitude, i.e., data normalization. Since there are ordered categorical variables and unordered categorical variables, the normalization of different categories of data needs to be classified and discussed.

(1) Normalization of ordered categorical variables

Since the ordered categorical variables of EFL vocabulary features are discrete to continuous variables, the normalization can be done by using the mean-variance normalization or the most-valued normalization. The mean-variance normalization method is to normalize the data into a standard normal distribution, which is suitable for data with no obvious boundary distribution. Maximum value normalization is to map the data between 0 and 1 by the maximum and minimum values of the data and is suitable for data with an obvious maximum value. Ordered categorical variables of features have maximum and minimum values. Therefore the ordered categorical variable feature data is processed by the most value normalization, which is calculated as shown in equation (1):

$$X^{con^*} = \frac{X^{con} - X_{\min}^{con}}{X_{\max}^{con} - X_{\min}^{con}} \tag{1}$$

where $X_{\min}^{con}$ denotes the minimum value of the ordered categorical variable of the class and $X_{\max}^{con}$ is the maximum value of the ordered categorical variable of the class.

(2) Logical Determination of Unordered Categorical Variables

For the unordered categorical variables of students' English vocabulary learning characteristics, it is difficult to normalize the variables because there is no connection between the variables, and it is necessary to establish a determination function, as shown in equation (2):

$$S(X_i^{uno,train}, X_i^{uno,test}) = \begin{cases} 0 & X_i^{uno,train} = X_i^{uno,test} \\ 1 & X_i^{uno,train} \neq X_i^{uno,test} \end{cases} i = 0,1,...,N \tag{2}$$

where $X_i^{uno,train}$ is the traveler feature corresponding to the i th unordered categorical variable in the training set, $X_i^{umo,test}$ the student's English vocabulary learning feature corresponding to the i th unordered categorical variable in the test set, and N is the dimensionality of the student's English vocabulary learning feature.

II. B. 3) Calculation of similarity

The commonly used distance metrics for KNN are Manhattan distance, Euclidean distance, etc [20]. Since the data design in the questionnaire survey uses a categorization process so that the difference in data values is not too large, i.e., the dimensional indicators of the experimental data are at the same scale level under the data categorization, the Euclidean distance is applied as a measure of the distance between the samples.

The Euclidean distance is the distance of the feature vectors between sample points. For the convenience of statistics, the ordered categorical variables are made to be sorted in the first place, and the unordered categorical variables are made to be sorted in the second place during the experiment. If there are n_1 ordered feature variables and n_2 unordered feature variables in the N-dimensional students' English vocabulary learning features, the numbering order of the students' English vocabulary learning features will be: $X_1^{con}, X_2^{con}, ..., X_{n_1}^{con}, X_{n_1+1}^{umo}, X_{n_1+2}^{uno}, ..., X_N^{uno}$. Then the formula for the Euclidean distance between the test set samples and the training set samples is shown in equation (3):

$$DIS = \sqrt{\sum_{i=1}^{n_1} (X_i^{con,t\min} - X_i^{con,test})^2 + \sum_{i=n_1+1}^{N} [S(X_i^{uno,train}, X_i^{uno,test})]^2} \tag{3}$$

II. B. 4) Feature weighting

After calculating the distance between the data in the test set and all the data in the training set, the first K training set data with the smallest distance from the test set data can be found, and the target value Y to be predicted in the test set will be obtained by finding the mean value of the target value Y of these K rows of data in the training set. If only the method of finding the arithmetic mean is taken, there may be a situation where the training set data that is farther away from the test set data is dominant because of its larger number. Therefore, the accuracy of the recommended English vocabulary learning path can be theoretically improved by using the weighted KNN algorithm [21]. The inverse function of the distance between samples is commonly used as the weights, so that the closer to the test set data the greater the weight of the first K training set data. The weights are calculated as shown in equation (4):

$$weight = \frac{1}{DIS + D\varepsilon} \tag{4}$$

where weight is the weight of a data feature in the first K students' English vocabulary learning data in the training set. $D\varepsilon$ is a constant to prevent the algorithm from being sensitive to noise.

II. B. 5) Target value projections

When testing the unknown sample P_u, after selecting the first K training set data with the closest distance to P_u, the number of different labels of the transportation modes corresponding to these K training sets are summed using weight weight, and the maximum value in the result corresponds to the transportation mode, which is the recommended English vocabulary learning path as shown in equation (5):

$$y = \arg\max \sum_{i=1}^{K} weight_i \delta(y, f(P_k))|_{y \in f(P)} \tag{5}$$

Formally, the K closest samples of P_u to the training set are denoted as $P_1, P_2, ..., P_k, f(P)$ is. category mapping function, $\delta(y^{test}, y^{train})$ is the impulse function, $\delta(y^{test}, y^{touin}) = 1$ if and only if $y^{test} = y^{tmin}$, otherwise, $\delta(y^{test}, y^{train}) = 0$.

II. B. 6) Determination of K-value

The value of K in the KNN algorithm has a direct impact on the prediction results. a small value of K increases the learning estimation error, makes the prediction results sensitive to the near neighboring instance points, and is prone to overfitting. If the K value is too large, the more distant instance points will also affect the prediction results, leading to an increase in the error rate of prediction. So the KNN algorithm needs to choose the appropriate K value. Commonly used K value selection methods include leave-one-out method and K-fold cross-validation method.

(1) Leave-one-out method

The leave-one-out method means leaving one sample for testing each time, and the steps are as follows:

Step 1: In the selected training set of n samples, number each sample. Keep sample number 1 as the test set and the other n-1 samples as the training set. After the training is completed the test is performed using sample number 1 and the prediction results are recorded.

Step 2: Select 1 sample as the test set in the order of the sample serial number, repeat step 1 and record the prediction results respectively.

Step 3: The average of the classification rate obtained from n tests of the training set is used as the recommended accuracy. The complexity of the leave-one-out method is $o(n)$, which is more cumbersome, but it is tested for each sample, highly utilized, and suitable for datasets with a small number of samples.

(2) K-fold cross validation method

Step 1: Divide all the training set data into k disjoint subsets, assuming that the training set has: n samples,then each subset will have n/k samples.

Step 2: Each time 1 copy from the divided subset is selected as the test set and the remaining k-1 copies are used as the training set.

Step 3: Calculate the distance between the training set and each test sample to get the classification rate.

Step 4: Repeat the above steps to get the classification rate for k tests and calculate the average as the recommended accuracy of this classifier.

The complexity of the K-fold crossover method is $o(k)$, which reduces the complexity compared to the leave-one-out method, and at the same time improves the utilization of the samples, and is suitable for data sets with a large number of samples due to the need to partition the data set. For the method of this paper, the larger the data volume of traveler characteristics, the higher the credibility of the results of the prediction of transportation modes, so the K-fold-crossing method is applied to determine the K value.

II. B. 7) Recommended accuracy

After determining the number distance measure, it is possible to classify the English vocabulary learning samples of the test set students. If the classification result P_m for the unknown sample y_c agrees with the label y_m for P_m, the classification is proved to be correct, and vice versa, it is judged to be incorrect. The proportion of the total number of correctly classified sample test sets is called the recommended accuracy. It is shown in Equation (6):

$$ACC = \frac{n_{correct}}{n_{test}} \tag{6}$$

where $n_{correct}$ is the number of correctly recommended samples and n_{test} is the total number of test sets.

II. C.KNN-based Emotional Interaction for English Vocabulary Learning

The current English vocabulary learning process lacks teacher-student interaction and emotional communication, and the emotional information brought by facial expressions, sounds and gestures is missing in the learning transmission process. All these lack of emotional information will affect the emotional interaction between teachers and students. On the one hand, it is difficult for students to feel the teacher's attention to them, and they are prone to confusion and laziness in learning. On the other hand, it is difficult for teachers to understand students' feelings and control their learning process effectively. Aiming at the problems described above, a weighted KNN-PCA-based emotional interaction design for English vocabulary learning is proposed, and it judges and understands the emotional state by capturing and predicting the facial expressions of online learners, so as to complete the learning interaction process and enhance the learners' online learning effect.

II. C. 1) PCA-based feature frame extraction

Principal Component Analysis (PCA) algorithm is used for the generation of feature frames in online multimedia English teaching.PCA is a widely used dimensionality reduction technique in computer vision, which maximizes the separation of all projection samples by selecting the dimensionality reduction through linear projections.

Let $\{x_1, x_2, \cdots, x_N\}$ be the set of N sample images in n-dimensional space. Assume that a linear transformation maps the n-dimensional space to an m-dimensional feature space, where $m < n$. Further, let the feature vector $y_j \in \mathbb{R}^m$ denote the linear transformation, then we have:

$$y_j = Q^T x_j, j = 1, 2, \cdots, N \tag{7}$$

where $Q \in R^{n\times m}$ is the orthogonal matrix. In addition, the scattering matrix S_T is calculated as follows:

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \tag{8}$$

where $\mu \in R^n$ is the average facial image.

Further, a linear transformation Q^T is applied to obtain the scattering matrix of the new feature vectors $\{y_1, y_2, \cdots, y_N\}$, which is denoted as $Q^T S_T Q$. To maximize the determinant of the scattering matrix of the projection sample, the projection Q_{opt} is chosen as follows:

$$Q_{opt} = \arg\max_Q | Q^T S_T Q | = [W_1, W_2, \cdots, W_m] \tag{9}$$

where w_i is the i th n-dimensional feature vector of S_T. The first orthogonal dimension of this feature space captures the largest variance in the database, while its last dimension captures the smallest variance in the database. These feature vectors having the same dimension as the original image will be considered as feature frames.

Next PCA is applied to the frame sequence of each video to select a representative set of feature frames. To select the optimal subset, the following steps are performed:

(1) Select the corresponding eigenvectors with non-zero eigenvalues to create the optimal feature space.

(2) Discard the bottom 40% of the feature vectors.

(3) It is assumed that the first 3 feature vectors are affected by lighting conditions, which reduces the classification performance. Therefore, all feature vectors except the first 3 feature vectors are used.

(4) Calculate the minimum number of feature vectors that guarantee that the energy e is greater than a typical threshold. Let e_i be the cumulative energy of the i th feature vector, then we have:

$$e_i = \sum_{j=1}^{i} \lambda_j / \sum_{j=1}^{k} \lambda_j \tag{10}$$

where: k is the number of non-zero eigenvalues. λ is the eigenvalue.

(5) Calculate the stretch value s_i of the i th eigenvector, defined as the ratio of the i th eigenvalue λ_i to the largest eigenvalue λ_I, expressed as follows:

$$s_i = \lambda_i / \lambda_i \tag{11}$$

Note that the typical threshold for the stretch value s_i is 0.01. In addition, most keyframe selection methods rely on motion differences. Motion differences can be calculated as the amount of motion between different parts of the face (e.g., lips and eyebrows). However, this method requires exhaustive numerical calculations. The author uses the amount of variance explained by each principal component obtained from a single video, and the variation in the emotional facial expression video corresponds to the localized variation of the facial expression in the time domain. This means that the variance explains the difference in motion associated with the dynamics of facial expressions.

II. C. 2) Prediction of Emotional Distribution

Assuming that there are M sentiment categories $C_1, \cdots, C_M$ and N training images, $x_1 \cdots, x_N$ (which also denote the corresponding features of the images), the distribution of the sentiment of P_n is denoted by $P_n = \{P_{n1}, \cdots P_{nm} \cdots, P_{nM}\}^T$ denotes the sentiment distribution of X_n, where P_{nm} denotes the probability that the sentiment expressed by x_n is c_m, and for each image there is $\sum_{m=1}^{M} P_{nm} = 1$. Assuming that y is a test image, the goal of this paper is to find the sentiment distribution $p = \{P_1, \cdots, P_M\}^T$ of y, i.e.:

$$f(\{x_n, p_n\}_{n=1}^{N}, y) \to p \tag{12}$$

Very distant training sets have little effect on y, considering all training sets slows down the run, and irrelevant training samples can mislead the algorithm's classification. The $k-$ nearest neighbor algorithm weighted by distance is a very effective inductive inference method, with closer samples weighted more heavily. It is robust to noise in the training data and it is also very effective when given a sufficiently large training set. By taking a weighted average of k nearest neighbors, the effect of isolated noisy samples can be eliminated.

The weighted k-nearest neighbors select only the basis functions corresponding to the k training images that are most similar to the test image for weighting, and the contributions of the k nearest neighbors are weighted to assign larger weights to the closer ones. $P_k(k = 1, \cdots K)$ denotes the sentiment distribution of the K training images closest to the test image, which is regarded as the basis function, and the sentiment distribution of the test image y is computed by summing up the distance-weighted sums of the basis functions p, i.e.:

$$p = \frac{\sum_{k=1}^{K} s_k p_k}{\sum_{k=1}^{K} s_k} \tag{13}$$

where s is the similarity between the test sample and the training sample:

$$s = e(-\frac{d(x_k, y)}{\beta}) \tag{14}$$

d is the Euclidean distance and β is the average distance between y and k training images. Algorithm: weighted k-nearest neighbor sentiment distribution prediction algorithm.

Input: training set (x_n, p_n), test set y.

Output: sentiment distribution of the test set p.

(1) Calculate the distance d between the test set images y and each image in the training set.

(2) Select the top k images that are closest to y in increasing order of distance $x_1 \cdots x_k$.

(3) $\beta = \frac{1}{k}\sqrt{(x_1 - y)^2 + \cdots + (x_k - y)^2}$ is brought into Eqn. (11) to compute the similarity s.

(4) Calculate the emotional distribution of test image y at the end of $p = \frac{\sum_{k=1}^{K} s_k p_k}{\sum_{k=1}^{K} s_k}$.

III. Exploratory Analysis of English Vocabulary Learning Examples

III. A. Exploratory Analysis of English Vocabulary Learning Paths

III. A. 1) Experimental environment

Due to the complexity of processing the data set selected for this experiment, we choose our own computer for the experiment, and the corresponding experimental hardware and software platforms are shown below:

(1) The experimental hardware platform is as follows:

Processor Intel (R), Core i7-8550U, CPU@2.00GHz, memory 16GB.

(2) The experimental software platform is as follows:

Ubuntu18.04, Python3.6, Tensorflow, Selenium, numpy, Pandas.

III. A. 2) Comparison algorithms

(1) Probabilistic Decomposition Model PMF: Currently, the most used in recommender systems is based on the matrix decomposition model, and the PMF model is a probabilistic model introduced on the basis of this model, which explains the hidden relationship between user and item from the perspective of the probabilistic generating process, and thus serves as a benchmark model for most new model comparisons.

(2) Item CF model: the recommendation algorithm based on content filtering refers to recommending similar content to a user based on the content that the user has consumed and then recommending similar content to that user.

(3) User CF model: The purpose of user-based similar recommendation is to find items that are of interest to a collection of users but not behaved by the target user.

III. A. 3) Experimental evaluation criteria

This paper synthesizes the influencing factors of various aspects and presents three representative recommended indicators. As shown below:

The accuracy rate of recommendation is an important index for evaluating the recommendation list; if the recommendation accuracy rate cannot be guaranteed, all other indexes are meaningless even if they are good. For the recommendation task of Top-K, the corresponding accuracy rate formula is shown in (15):

$$Percision@K=\frac{\sum_{u\in U}\left|R(u)\cap T(u)\right|}{|U|\times K} \tag{15}$$

where $R(u)$ is the set of optimal paths for English vocabulary learning recommended to user u for prediction, and $T(u)$ is the set of optimal paths for English vocabulary learning by user u in the test set, $|U|$ denotes all users in the test set, and K denotes the number of optimal paths recommended by users u.

Diversity This metric measures the recommendation model's ability to capture similar interests based on the user's historical preferences; the user has his or her historical interests, but such interests are not necessarily static. The higher the value of this indicator, the fewer the number of optimal paths of the same type in the push list, which is favorable for expanding the list of users' interests. The overall diversity of the corresponding computational recommendation model Top-K is as follows:

$$Diversity@K=\frac{1}{|U|}\sum_{u\in U}\left(1-\frac{\sum_{i=1}^{K}\sum_{j=1}^{K}sim(i,j)}{\frac{1}{2}K(K-1)}\right) \tag{16}$$

where $|U|$ denotes the total number of users of the test data, K denotes the number of optimal paths recommended by the recommendation model to the user for English learning, and $sim(i,j)$ denotes the similarity between the optimal path i and the optimal path j, which is generally chosen to be measured by the cosine similarity. The specific calculation is shown in Equation (17):

$$sim(i,j)=\frac{v_i,v_j}{\|v_i\|\|v_j\|} \tag{17}$$

where v_i and v_j denote the optimal path type vectors corresponding to the optimal paths i and optimal paths j respectively.

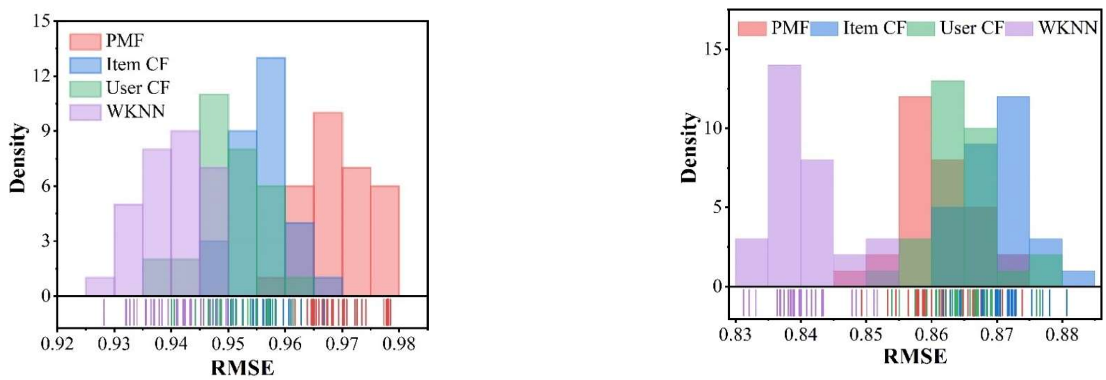
The coverage rate is of little practical significance in terms of looking at the recommendation list alone, but this metric plays an important role in the learning and improvement of the recommendation model itself. The corresponding formula is shown in (18):

$$Coverage@K=\frac{\left|U_{u\in U}R(u)\right|}{I} \tag{18}$$

where U is the set of all test data users, $R(u)$ is the list of Top-K recommendations for users u by the recommendation model, and I is the set of all optimal paths for English vocabulary learning.

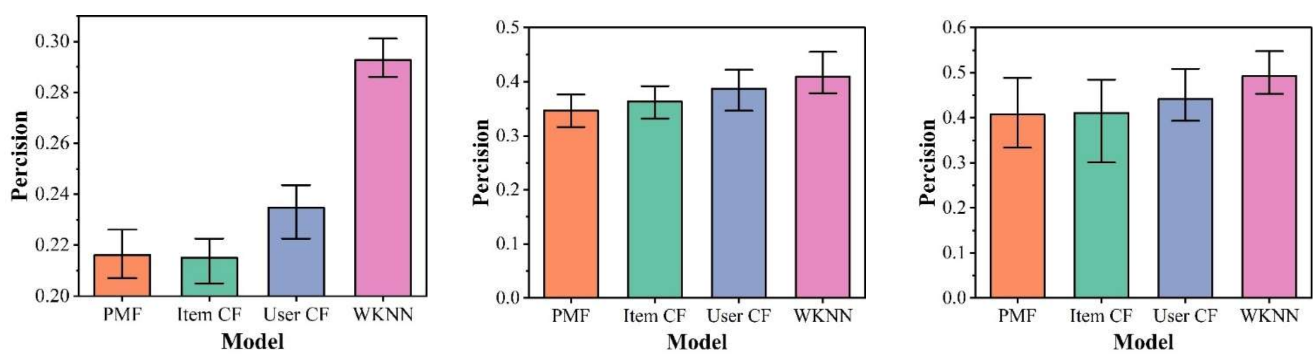
III. A. 4) Analysis of results

In this paper, the recommendation accuracy, diversity and coverage of PMF, Item CF, User CF and Weighted KNN (WKNN) models are tested on datasets A and B respectively for five Top-K recommendations without using the value of K. At the same time, when each model converges, the RMSE of the training error is shown in Fig. 1, in which (a) ~ (b) are datasets A and B, respectively. Item CF model Compared with the commonly used PMF algorithm, the training error RMSE has been reduced more significantly, by 2.08% on dataset A and 3.43% on dataset B. The training error RMSE is reduced by 2.08% on dataset A and 3.43% on dataset B, respectively. Meanwhile, when the User CF model was used, the RMSE was again reduced a lot, by 0.185% and 0.256% on the two datasets, respectively, relative to the PMF model, which illustrates that the improvement of the User CF model is still more obvious. Finally, using the weighted KNN model for recommendation, the RMSE error of training waits to be further reduced, and the RMSE error of recommendation on datasets A and B is 0.9325 and 0.8423.

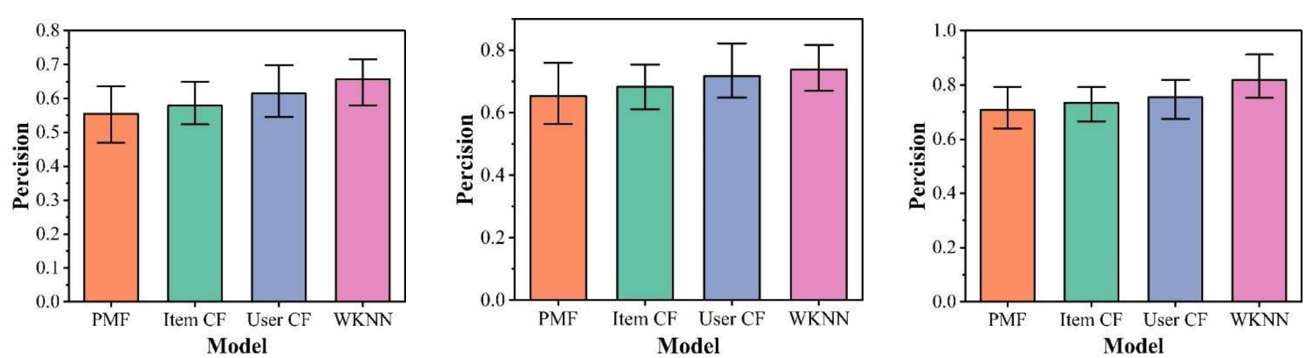


(a)Data set A (b)Data set B

Figure 1: The training error RMSE of each model



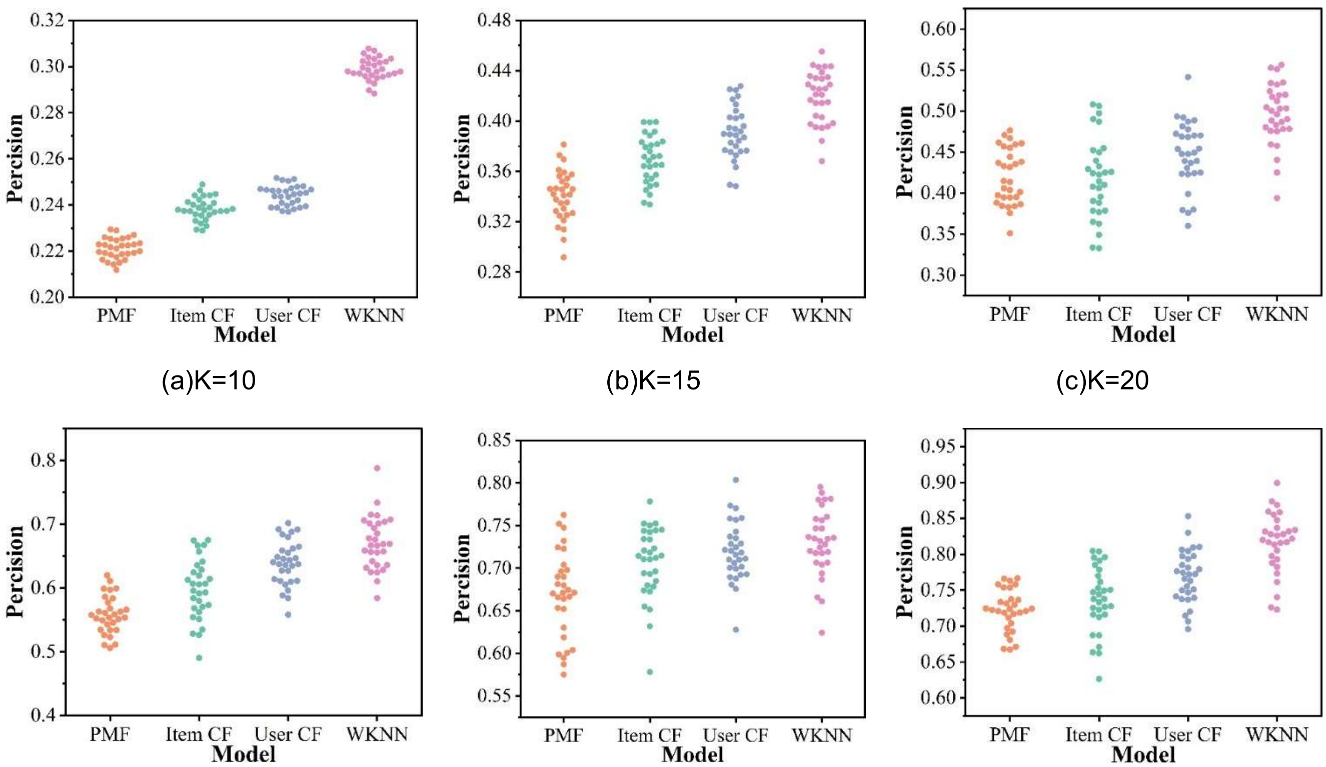
(a)K=10 (b)K=15 (c)K=20



(d)K=25 (e)K=30 (f)K=35

Figure 2: Recommended accuracy comparison experiment on dataset A

For the final Top-K recommendation of the model, the performance on the three indicators of recommendation accuracy, diversity and coverage. The recommendation accuracy comparison experiment on dataset A is shown in Fig. 2, where (a)~(f) are 10, 15, 20, 25, 30, and 35 respectively, and the results of the recommendation accuracy comparison experiment on dataset B are shown in Fig. 3. The weighted KNN recommendation model has a relatively obvious improvement in the recommendation accuracy rate compared to other models, and when the recommended K value is increasing, the trend of this enhancement will be significantly weakened instead, without a great deal of change. It also shows that when a small number of recommendations are made, the effect of the KNN algorithm is significant, but when the number of recommendations becomes more, the effect of the KNN algorithm can basically meet the user needs.



(a)K=10 (b)K=15 (c)K=20

(d)K=25 (e)K=30 (f)K=35

Figure 3: Recommended accuracy comparison experiment on dataset B

For diversity, the comparison experiment of the overall diversity index of recommendations on dataset A is shown in Fig. 4, and the comparison experiment of the overall diversity index of recommendations on dataset B is shown in Fig. 5. The Item-CF model has higher overall diversity than the PMF model, and after the improvement of the Item-CF model, the diversity of the User-CF model on dataset A improves significantly, and the diversity of the User-CF model on dataset B improves not significantly, probably due to the increase of the dataset, which makes the recommendation more and more dense. The diversity improvement is not very obvious, probably due to the increase of the dataset, the recommendations are getting denser and denser so that the diversity improvement of the recommendations is not significant. Similarly, after using the weighted KNN algorithm, the diversity enhancement on dataset A is significant and the diversity enhancement on dataset B is average. Thus, the weighted KNN algorithm shows little improvement in accuracy with increasing K values, but a more significant improvement in diversity of recommendations. From the perspective of diversity, the recommendation model based on the weighted KNN algorithm can ensure that the accuracy rate is not lost in the recommendation process, which makes the diversity of recommendations better and provides students with a wider range of English vocabulary learning materials.

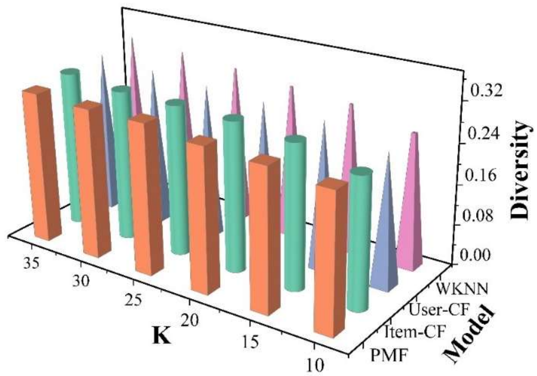


Figure 4: Dataset A recommends comparison of the overall diversity index

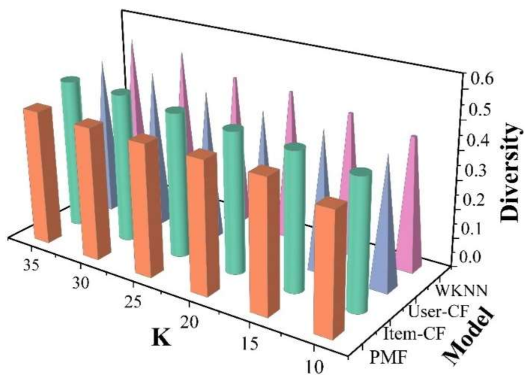


Figure 5: Dataset A recommends comparison of the overall diversity index

For the coverage rate, the higher the coverage rate, the wider the range of English learning paths recommended to the user, the higher the probability of covering different types of optimal paths, which can effectively solve the long-tail effect existing in the recommendation model. The comparison experiment of the overall coverage rate of recommendation in dataset A is shown in Fig. 6, and the comparison experiment of the overall coverage rate of recommendation in dataset B is shown in Fig. 7. The Item-CF model, relative to the PMF model, does not have much enhancement when the value of the recommended K value is relatively small, but with the increasing value of the recommended K value, this enhancement has a more obvious effect. Similarly, User-CF model for Item-CF model, as the K value continues to increase, this growth trend is also increasing. When comparing this paper's model with User-CF, this paper's model is better on dataset B. The difference in coverage between the two on dataset A is not significant.

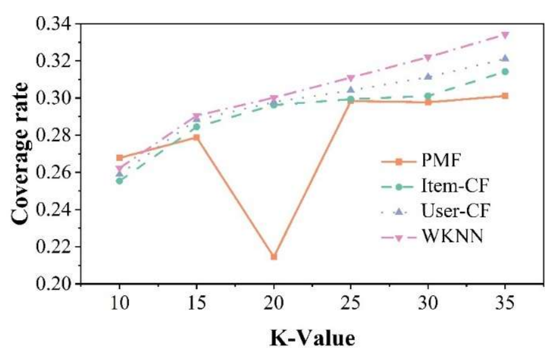


Figure 6: Data set A recommends overall coverage comparison

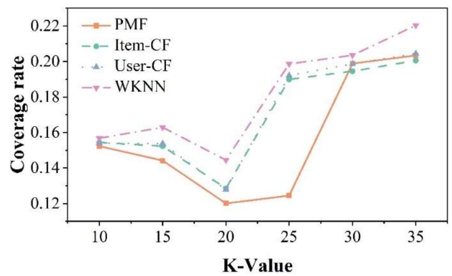


Figure 7: Data set B recommends overall coverage comparison

Comprehensively analyzing the above experiments, the model in this paper ensures the accuracy of the optimal path recommendation for English vocabulary learning, on the basis of which we can appropriately increase the K-value of the recommendation, so that the recommended materials are more diversified, so that the students can obtain more learning fun. At the same time, we can also improve the coverage of the optimal path recommendation for English vocabulary learning, solve the long-tail effect in the recommendation, and recommend some novel materials to students, such as some high-quality learning materials, which not only improves the students' learning performance, but also enriches the path of English vocabulary learning, and jointly promotes the quality of English teaching to improve and develop.

III. B. Exploratory Analysis of Emotional Interaction in English Vocabulary Learning

Accompanied by the development of intelligent technology, online English vocabulary teaching has gradually become an important application scene in the teaching field, and the research on emotional interaction in the online English vocabulary teaching environment is imminent. Researchers are eager to use the data in the online English vocabulary teaching scene to realize the automatic prediction of emotions through the principal component analysis algorithm and the K-nearest neighbor algorithm, in order to help teachers realize accurate teaching and improve the interactive effect of English vocabulary courses. The specific analysis process is as follows:

III. B. 1) Feature extraction

The database selected for the experiment is the standard German Berlin library, and the four extracted features are short-time average amplitude, short-time average energy, short-time over-zero rate and fundamental frequency. In order to reduce the influence of individual differences in emotional expression on the recognition results, the extracted emotional features are normalized using the principal component algorithm, and the normalized emotional features are used as training samples and test samples in the experiment. In the training phase, 40 images were selected as samples for each of the four emotions for feature parameter extraction, i.e., short-time average energy, short-time average amplitude, short-time over-zero rate and base frequency were extracted for happy, angry, neutral and sad. The emotional features are shown in Figure 8-figure supplement 11. Through the data in the figure, it is found that the fundamental frequency of Neutral and Sad are relatively close to each other, but there is a significant difference in the short-time average energy. And the short-time average energy of anger and happiness are closer, but the base frequency of anger is higher, which indicates that the K nearest neighbor algorithm using the contribution of feature parameters weighted with the Euclidean distance is reasonable for interaction emotion prediction.

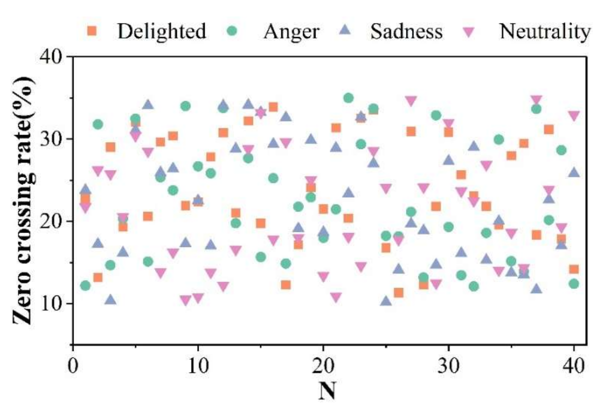


Figure 8: Zero crossing rate characteristics of three kinds of emotions

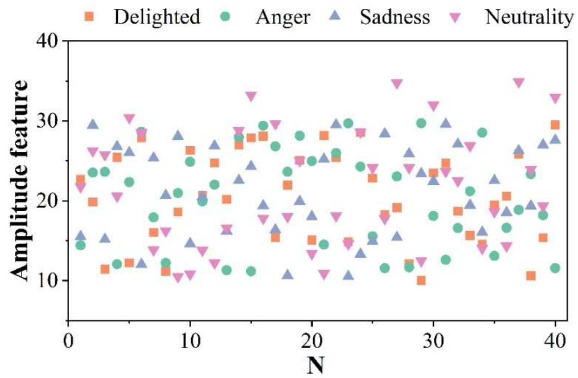


Figure 9: Average amplitude characteristics of three emotions

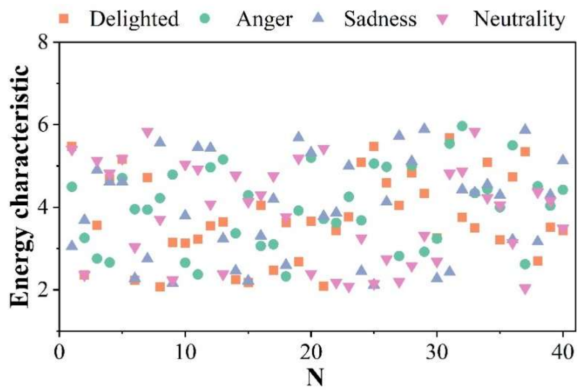


Figure 10: Average energy characteristics of three emotions

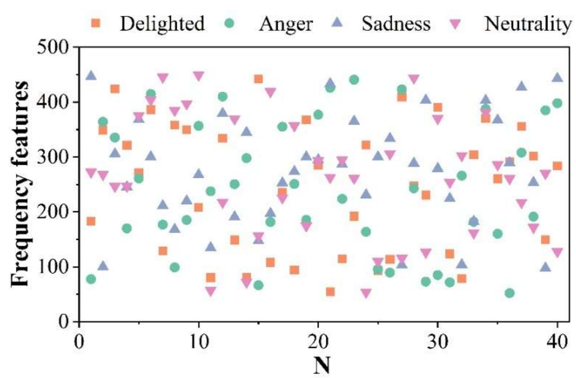


Figure 11: Fundamental frequency mean characteristics of three kinds of emotion

III. B. 2) Analysis of predicted effects

After the extraction of the feature parameters of the four emotions, the contribution of each parameter to the recognition of different emotional states is shown in Table 1. The traditional K-nearest neighbor algorithm and principal component weighted K-nearest neighbor algorithm were used for prediction, and the two methods were compared through experiments to analyze their respective characteristics, and the comparison results of emotion prediction are shown in Table 2. From the table, it can be seen that the prediction effect of the principal component weighted KNN algorithm on the three emotions of anger, happiness and neutrality have been improved to different degrees, this is because happiness and anger are easily confused with each other in the process of prediction, and the improved algorithm is able to correct the wrong discrimination well, and the prediction of neutrality is also improved at the same time. The prediction of sadness was not improved, because the emotion sadness is easy to predict under these 4 emotional features, more obvious than the other 3 emotions. After extracting and mixing the features, the resolution of the features can be improved effectively by giving corresponding weights to the characteristics of different features. The principal component weighted KNN method can better correct the misjudgment produced in the process of emotion interaction prediction, and overall the emotion prediction of English teaching interaction has a greater degree of improvement, and the technology can effectively improve the effect of student interaction in the English vocabulary course, which is conducive to the improvement of the quality of English teaching.

Table 1: The contribution degree of emotion characteristic parameters to emotion state

Contribution degree	Short-time energy	Short-time mean amplitude	Short time zero crossing rate	.Fundamental note frequency
Anger	0.2615	0.3551	0.8124	0.9308
Sadness	0.2834	0.3262	0.8317	0.9207
Delighted	0.9346	0.9271	0.9434	0.9625
Neutrality	0.2617	0.3553	0.8126	0.9306

Table 2: Comparison results of affective prediction

Method	Anger	Sadness	Delighted	Neutrality	Mean Value
Traditional KNN	76.33%	65.25%	91.52%	77.81%	77.73%
Improved KNN	95.35%	91.24%	97.88%	99.66%	96.03%

IV. Conclusion

In this paper, the optimization of English vocabulary learning path can be achieved by constructing the optimal path recommendation model for English learning under the multi-task learning framework. Considering the existence of classification boundaries in the research data, for this reason, the optimal path recommendation model for English vocabulary learning based on weighted KNN algorithm is proposed. Although the students' English vocabulary learning path is optimized, the interaction between students and teachers on top of the course is low. To address this problem, a weighted KNN-PCA combination is used to predict the emotional state of online learners so as to complete the learning interaction process, which has achieved the purpose of improving the classroom interaction effect. Comprehensive research data and models are used to explore and analyze the optimization and interaction of English vocabulary learning path. When K=35, the optimal learning path recommendation accuracy of this paper's model on datasets A and B is 0.842 and 0.846, which is much better than the other three models, proving that this paper's model can provide students with optimal learning paths. In addition, in the process of sentiment prediction, the prediction rate of the traditional KNN algorithm is 77.73%, while the prediction rate of the combined PCA-KNN algorithm is 96.03%, which indicates that the method of this paper can effectively improve the effect of students' interactions on the English vocabulary course, and it can provide reference for the development and construction of English teaching in schools.

References

[1] Rao, P. S. (2019). The role of English as a global language. Research journal of English, 4(1), 65-79.

[2] Melitz, J. (2016). English as a global language. In The Palgrave handbook of economics and language (pp. 583-615). London: Palgrave Macmillan UK.

[3] Mohammed, M. A. A., & Abdalla, M. (2020). English language and globalization. Intl J of Novel Research in Education and Learning, 7(1), 5-11.

[4] Simonnet, E., Loiseau, M., & Lavoué, É. (2025). A Systematic Literature Review of Technology-Assisted Vocabulary Learning. Journal of Computer Assisted Learning, 41(1), e13096.

[5] Yuan, X., & Tang, X. (2025). Effects of the sequential use of L1 and bilingual subtitles on incidental English vocabulary learning: A cognitive load perspective. British Journal of Educational Psychology.

[6] Aisyah, N., Thohir, L., Zamzam, A., & Melani, B. Z. (2025). English Vocabulary Learning Strategies Used by The Eleventh-Grade Students of MAN 1 BIMA. Jurnal Ilmiah Profesi Pendidikan, 10(1), 40-47.

[7] Teng, M. F. (2025). Exploring self-regulated vocabulary learning strategies, proficiency, working memory and vocabulary learning through word-focused exercises. The Language Learning Journal, 53(1), 22-39.

[8] Biseko, J. M. (2025). Vocabulary learning in EFL context: do primary school English Subject textbooks provide structured support?. Cogent Education, 12(1), 2455047.

[9] Normurodovna, M. A. (2025, January). BUILDING VOCABULARY: EFFECTIVE STRATEGIES FOR LEARNERS. In International Conference on Medical Science, Medicine and Public Health (pp. 4-11).

[10] Xie, S. (2025). The influence of learning motivation on English vocabulary learning. In Addressing Global Challenges-Exploring Socio-Cultural Dynamics and Sustainable Solutions in a Changing World (pp. 227-232). Routledge.

[11] Kulsum, E. M., & Fitri, F. P. C. (2025). Exploring Vocabulary Learning Strategies in EFL Contexts. JEPAL (Journal of English Pedagogy and Applied Linguistics), 5(2), 187-198.

[12] Li, M., & Wang, T. (2025). Identifying Optimal Learning Strategies: Application of the Asymptotic Retention Rate Model in College Students' Vocabulary Learning. The Asia-Pacific Education Researcher, 1-12.

[13] Wang, X., & Feng, L. (2025). Examining the Influential Mechanism of English as a Foreign Language (EFL) Learners' Flow Experiences in Digital Game-Based Vocabulary Learning: Shedding New Light on a Priori Proposed Model. Education Sciences, 15(2), 125.
[14] Sasson, A., Mor, D., & Schiff, R. (2025). Words at first exposure: How EFL proficiency influences incidental and intentional vocabulary learning in an unfamiliar language. Reading and Writing, 1-21.
[15] Lou, Y. (2019). Storage and allocation of English teaching resources based on k-nearest neighbor algorithm. International Journal of Emerging Technologies in Learning (iJET), 14(17), 102-113.
[16] Zhao, J., Mao, H., Mao, P., & Hao, J. (2024). Learning path planning methods based on learning path variability and ant colony optimization. Systems and Soft Computing, 6, 200091.
[17] Zheng, F. (2025). Improving english vocabulary learning with a hybrid deep learning model optimized by enhanced search algorithm. Egyptian Informatics Journal, 29, 100619.
[18] Qiqi He,Qiuju Yang,Hang Su & Yixuan Wang. (2024). Multi-task learning for segmentation and classification of breast tumors from ultrasound images. Computers in Biology and Medicine,173,108319-.
[19] Xin Wen,Yang Li & Zemin Zheng. (2024). Scalable efficient reproducible multi-task learning via data splitting. Statistics and Probability Letters,208,110071-.
[20] S Vaisali,C Maheswari,S Shankar & R Naveenkumar. (2025). Predicting the performance of assistive device for elderly people using weighted KNN machine learning algorithm. Journal of back and musculoskeletal rehabilitation, 10538127251317602.
[21] Yonghua Wang,Jingyi Lu & Kaidi Zhao. (2020).A locally weighted KNN algorithm based on eigenvector of SVM. International Journal of Wireless and Mobile Computing,19(3),