Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Artificial Intelligence-Driven Personalized College English Vocabulary Learning Path Generation and Optimization

Li Li[1], ✉

[1] Nanyang Medical College, Nanyang, Henan, 473000, China

ABSTRACT

In recent years, the construction of education informatization has been comprehensively promoted, and the personalized learning recommendation model has brought a new direction for the development of intelligent learning platform for college English vocabulary. This study constructs the KCPE-SR model based on collaborative filtering algorithm and knowledge graph, generates and optimizes the suitable personalized learning paths for learners through the interaction between learners of college English vocabulary and resources, and develops a personalized college English vocabulary learning system based on this model. The analysis of the application effect of the system reveals that the experimental class students' English vocabulary learning performance has been significantly improved with the help of the personalized learning system, and the students' English vocabulary knowledge mastery (20.00 points) and vocabulary comprehensive application ability (20.49 points) have also increased. The personalized college English vocabulary learning path generation and optimization system proposed in this paper is able to achieve accurate personalized recommendation of learning resources and can meet the needs of college English vocabulary learning.

Keywords: collaborative filtering algorithm, knowledge graph, KCPE-SR model, learning resource recommendation, personalized learning

1. Introduction

In recent years, the rapid development of artificial intelligence technology has triggered innovative changes in several fields. Generative AI is an advanced technology that simulates human learning and creation processes through deep learning and neural networks [1-2]. In terms of working principle, generative AI first learns patterns and structures of specific domains (e.g., language, image, etc.) by training with a large amount of data, and autonomously generates new instances based on these learned patterns [3-4]. As a cutting-edge AI technology, generative AI opens up a brand new way for language learning with its uniqueness of mimicking human language use and comprehension [5-6].

✉ Corresponding author.
E-mail address: lily5828@163.com (L. Li).
Received 10 May 2024; Revised 22 August 2024; Accepted 05 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-017

The learning of English vocabulary for college students plays a fundamental role in the improvement of their overall English proficiency. In the specific practice of English vocabulary learning, generative artificial intelligence not only has the ability to generate a large number of contextually relevant example sentences, but also provides learning materials and feedback according to the actual level and progress of the learners, thus greatly enhancing the richness and adaptability of the learning process [7-9]. Meanwhile, in the age of intelligence, college students improve the learning efficiency of English vocabulary with the help of mobile terminal software or intelligent platforms, so as to enhance their comprehensive English ability, which will play a positive role in their current professional expansion and future career development [10-12].

At the same time, the use of big data, artificial intelligence and other technical means, according to each student's situation to develop a personalized learning path, intelligent recommendation of learning content and testing, according to the test results of the continuous feedback iterative learning process, in order to achieve the highest efficiency of each student to achieve the learning goals [13-14]. From the current educational situation, Al technology plays an important role in the teaching process, plays an important role, such as the more familiar batch correction network, homework help and so on. The future trend of learning becomes AI technology to become an auxiliary tool for teachers' teaching, and teachers can better complete the teaching of English vocabulary and accomplish the teaching goals with the help of AI technology [15-16].

Literature [17] combined with literature survey, interviews and other methods to clarify that Bai Ci Zhan APP effectively stimulates students' motivation to learn English vocabulary and expands their English vocabulary to a certain extent. Literature [18] envisioned an English vocabulary test with artificial intelligence technology as the core logic, revealing that English vocabulary APP and website English vocabulary learning are effective and feasible. Literature [19], based on empirical analysis and research, learned that the introduction of AI technology in English vocabulary teaching effectively enhanced English vocabulary learning effectiveness and learning efficiency. Literature [20] argues that output reading has positive significance for students' English learning from the perspective of English vocabulary acquisition, and the study provides a certain reference for English vocabulary and English reading teaching. Literature [21] analyzed the research on English vocabulary teaching and pointed out that the English corpus tool helps students to improve their English vocabulary and English communication skills. Literature [22] conducted a comparative evaluation experiment and demonstrated that the English vocabulary learning model based on adaptive algorithms is better than the traditional vocabulary learning model. Literature [23] designed an intelligent English collaborative education game based on the learning analytics (LA) method and verified that the English teaching model stimulated students' motivation and interest in English vocabulary learning to a certain extent based on practical feedback. Literature [24] used questionnaire survey method and semi-structured interview strategy to examine and analyze the attitudes and perceptions of English vocabulary learners' learning strategies based on smartphones, and pointed out that the scientific and reasonable combination of the essence of traditional vocabulary learning strategies with network information technology and mobile terminals is more helpful to improve students' English vocabulary learning effects.

This paper collects all kinds of learning resources of college English vocabulary from the Internet and textbook resources, and constructs the construction of college English vocabulary learning resource base by categorizing the vocabulary learning knowledge points and the degree of difficulty. Combining the collaborative filtering recommendation algorithm and the knowledge concept path enhancement method to establish a personalized English vocabulary learning path generation and optimization model. Obtain the embedded representation of university English vocabulary learning resources through knowledge mapping, and then extract the knowledge concept path between vocabulary learning. The interaction function between vocabulary knowledge learning resources is calculated to predict learners' mastery of vocabulary knowledge points, personalized learning resources are recommended based on the prediction results, and the personalized recommendation results are optimized and trained using the alignment loss function. Subsequently, a personalized college English vocabulary learning path generation and optimization system is designed by combining the learning resource library and the personalized recommendation model. In this paper, we conduct quasi-experiments to study whether the learning system has a facilitating effect on college students' English vocabulary learning performance, knowledge mastery and comprehensive application ability.

2. Method

2.1. Constructing a Resource Base for Learning College English Vocabulary

2.1.1. Learning resource collection and screening. The university English vocabulary learning resources studied in this paper are mainly collected from the Internet, various kinds of teaching aids and test sets as well as supporting resources, and the collected resources include teaching videos, teaching courseware, teaching design, exercises and pictures related to knowledge points.

1) Internet. The Internet is the main way to collect college English vocabulary learning resources. There are a large number of college English vocabulary learning resources and some English vocabulary learning related websites on the Internet, download the related resources on these websites, and reorganize and classify them according to the actual needs of this paper's construction.

2) All kinds of teaching aids and test sets. The learning resources in the teaching aids and test sets have very good pertinence, applicability and differentiation of difficulty level, and the resources in the teaching aids and test sets are manually made into electronic resources for collection and storage.

3) Supporting resources. Most textbooks and teaching aids have supporting learning resources, such as courseware, lesson plans, micro-videos and so on. These supporting resources are highly systematic, applicable and targeted, and these learning resources are collected and organized according to the actual construction needs.

To screen the collected learning resources, firstly, for the media-type learning resources such as teaching videos, teaching courseware and teaching design, it is necessary to ensure that the content of the resources is correct, adequate, vivid and easy to understand. Secondly, for the exercise resources, it is required that the question types are rich, the amount of questions is moderate, the contents are typical, the knowledge points examined are clear and the difficulty should be graded.

2.1.2. Classification of learning resources:

1) Classification of Knowledge Points. Reasonable classification of university English vocabulary learning resources is the key to building a university English vocabulary learning resource base. Classifying learning resources correctly and reasonably according to a certain structure can facilitate the system to provide learners with learning resources around a certain knowledge point as well as facilitate the system to find learning resources related to the demanded knowledge point. There are two common ways to categorize learning resources, one is to categorize learning resources according to the chapter structure of the textbook, and the other is to categorize learning resources according to the knowledge structure. The learning resources of college English vocabulary corresponding to the knowledge points are categorized according to the structural relationship among the first-level knowledge points, second-level knowledge points, third-level knowledge points and fourth-level knowledge points.

2) Difficulty Level Classification. Different knowledge points corresponding to the learning resources (mainly exercise resources) there is a difference in the degree of difficulty, in order to quantitatively analyze the degree of examination of exercises and tests on the knowledge points, according to the direction of the content of the resources examined to divide the degree of difficulty of the learning

resources. The difficulty level of learning resources is divided into three categories: easy, medium and difficult.

2.2. Personalized Recommendation Optimization Strategy for English Vocabulary Learning Paths

Under the collaborative filtering recommendation framework [25], this paper proposes the Knowledge Concept Path Enhancement [26] personalized recommendation and optimization model for English vocabulary learning path (KCPE-SR), which introduces the knowledge concept path to enhance the personalized recommendation of English vocabulary learning path based on the application of recurrent neural network for sequence modeling of learner representations to achieve the simultaneous enhancement of the recommendation accuracy and explanatory nature. First, the embedded representation of English vocabulary learning is obtained through the knowledge graph [27] representation learning technique. For the specified learner p_i and English vocabulary study q_j, the N English vocabulary studies that learner p_i has interacted with recently at a known t moment are extracted based on the knowledge graph to extract the knowledge concept paths between these English vocabulary studies and English vocabulary study q_j, which is then combined with the knowledge graph embedding representation to obtain the embedded representation of the knowledge concept paths. Then, the embedding representation $p_{i,t}$ of learner p_i at moment t is obtained by sequence modeling, and the embedding representation $q_{ij,t}$ of English vocabulary learning q_j when facing learner p_i at moment t is obtained by fusing the knowledge concept paths.Finally, the interaction probability of the two at moment t is predicted by establishing the interaction function between $p_{i,t}$ and $q_{ij,t}$.

1) Knowledge Graph Embedding Representation. In order to learn the embedding representation of entities and knowledge concepts in the knowledge conceptual graph, this study applies the TransE technique, i.e., the ternary $(h,r;t)$ in the knowledge graph satisfies $e_h+e_r\approx e_t$, where $e_h,e_t\in\mathbb{R}^{d}_{ò}$ is the embedding representation of the entities, $e_r\in\mathbb{R}^{d}_{r}$ is the embedding representation of the semantic relations, and $d_e=d_r$. The TransE model defines a distance metric function $dm(\cdot)$ on the ternary (h,r,t), see Eq:

$$dm(h,r,t)=\|e_h+e_r-e_t\|_2 \tag{1}$$

The real ternary can satisfy a low value of $dm(h,r;t)$. In general, the TransE model learns by optimizing the ranking loss function:

$$\mathrm{L}_{KGE}=\sum_{(h,r,t,t')}^{\mathrm{G}'}-\ln(\sigma(dm(h,r,t')-dm(h,r,t))) \tag{2}$$

where $G'=\{(h,r,t,t')\,|\,(h,r;t)\in\mathrm{G},(h,r,t')\notin\mathrm{G}\},(h,r,t')$ denotes the random replacement of the tail entity t in (h,r,t) with $t'\neq t$. Through the TransE Knowledge Graph representation, an initial representation, denoted as $\tilde{e}_h,\tilde{e}_r,\tilde{e}_t\in\mathbb{R}^{d_ò}$, will be obtained for each ternary (h,r,t). In particular, $\tilde{q}_j\in\mathbb{R}^{d_ò}$ is used to denote the embedded representation of English vocabulary learning in subject j.

2) Knowledge concept path mining and representation. Given that stemmed knowledge concept mapping introduces noisy data when applied to an individual's English vocabulary learning recommendation task, this study extracts knowledge concept paths by introducing meta-paths [28]. Specifically, the model proposed in this section uses two types of meta-paths to describe the learner's course selection process for:

$$\begin{cases} learner \xrightarrow{iweract} course_1 \xrightarrow{comains} concept_1 \\ \xrightarrow{prerequisite} concept_2 \xleftarrow{comtains} course_2 \\ learner \xrightarrow{iweract} course_1 \xrightarrow{comtains} concept_1 \\ \xleftarrow{upper} concept_2 \xrightarrow{upper} concept_3 \xleftarrow{comtains} course_2 \end{cases} \tag{3}$$

The first type of meta-paths indicates that learners keep learning more difficult knowledge concepts, and therefore search for English vocabulary learning that contains subsequent knowledge points based on the knowledge points contained in the already learned English vocabulary learning through the precedence relation. The second type of meta-path represents the process of learners learning different lower level knowledge concepts under the same higher level knowledge concept. Without considering the first learner English vocabulary learning interaction edge, the remaining part of the meta-path can represent the learning order relationship between English vocabulary learning, and the path information is all static information in the knowledge concept map, so the set of paths between any two English vocabulary learning that match the above meta-paths are denoted as $KCP^{Ql}(q_j, q_{j'})$ and $KCP^{Q2}(q_j, q_{j'})$, respectively. given that learner p_i has recently interacted with the N English vocabulary learning $[q_{i,t-N},...,q_{i,t-3},q_{i,t-3},q_{i,t-1}]$, and that the given English vocabulary learning q_j corresponds to the set of knowledge concept paths as $KCP(p_i,q_j,t)=\bigcup_{k=0}^{N-1}\{p_i \xrightarrow{imeraca} path \mid path \in KCP^{Q1}(q_{i,I-N+k},q_j)\cup KCP^{Q2}(q_{i,I-N+k},q_j)\}$. In order to embed the representation of $KCP(p_i,q_j,t)$, the set of paths is divided into two groups according to the categories of the meta-paths, the first group is denoted as $KCP^1(p_i,q_j,t)=\bigcup_{k=0}^{N-1}\{p_i \xrightarrow{interact} path \mid path \in KCP^{1}(q_{i,l-N+k},q_j)\}$, and the second group as $KCP^2(p_i,q_j,t)=\bigcup_{k=0}^{N-1}\{p_i \xrightarrow{interact} path \mid path \in KCP^{2}(q_{i,I-N+k},q_j)\}$, which are represented by the recurrent neural network and the aggregation layer as $kcp^1_{ij,t} \in \mathbb{R}^{d_f}$ and $kcp^2_{ij,t} \in \mathbb{R}^{d_f}$, respectively, as shown in the following equation:

$$
\begin{cases}
kcp^1_{ij,t} = \sigma\left(\frac{1}{|KCP^1(p_i,q_j)|}\sum_{path_m \in KCP^1(p_i,q_j)} GRU^1(emb^1(path_m))\right) \\
\qquad emb^1(path_m) = \left[\tilde{q}_{m,1}, \tilde{e}_{m,1}, \tilde{e}_{m,2}, \tilde{q}_{m,2}\right] \\
kcp^2_{ij,t} = \sigma\left(\frac{1}{|KCP^2(p_i,q_j)|}\sum_{path_m \in KCP^2(p_i,q_j)} GRU^2(emb^2(path_m))\right) \\
\qquad emb^2(path_m) = \left[\tilde{q}_{m,1}, \tilde{e}_{m,1}, \tilde{e}_{m,2}, \tilde{e}_{m,3}, \tilde{q}_{m,2}\right]
\end{cases}
\tag{4}
$$

where $emb(\cdot)$ function maps the English vocabulary learning and knowledge concepts in the path into corresponding embedding representations based on the knowledge graph embedding representations, $GRU^1(\cdot)$ and $GRU^2(\cdot)$ are two different trainable GRU recurrent neural networks, and $\sigma(\cdot)$ is a Sigmoid function.

3) Path-enhanced English vocabulary learning representation of knowledge concepts. The learner embedding representation is acquired by the GRU recurrent neural network [29] acting on the English vocabulary learning embedding representation, denoted as $p_{i,t} \in \mathbb{R}^{d_e}$. Based on the learner embedding representation and the knowledge concept path representation, the English vocabulary learning representation is augmented as shown in the following equation:

$$
\begin{cases}
q_{ij,t} = \alpha_1 \tilde{q}_j + \alpha_2 kcp^1_{ij,t} + \alpha_3 kcp^2_{ij,t} \\
\alpha = soft\max(p_{i,t}\cdot \tilde{q}_j, p_{i,t}\cdot kcp^1_{ij,t}, p_{i,t}\cdot kcp^2_{ij,t})
\end{cases}
\tag{5}
$$

where weight α models the distribution of learner p_i's attention to the learning strategy at moment t.

4) Learner English vocabulary learning interaction prediction. Based on the above module, the model obtains the English vocabulary learning representation $q_{ij,l}$ and the learner representation $p_{i,l}$. Based on the collaborative filtering framework, the interaction probability $\hat{y}_{ij,I}$ between learner p_i and English vocabulary learning q_j at moment t is obtained by vector dot product with the formula:

$$\hat{y}_{ij,t} = p_{i,t} \cdot q_{ij,t} \tag{6}$$

In this paper, we also use the ranking loss function [30] to optimize the recommendation results, and the core idea is that the observed learner English vocabulary learning interactions have higher prediction probabilities than the unobserved interactions, see Equation (7):

$$\mathrm{L}_{CF} = \sum_{(p_i, q_j, q_{j'})} -\ln(\sigma(\hat{y}_{ij} - \hat{y}_{ij'})) \tag{7}$$

where $\mathrm{Y}' = \{((p_i, q_j, q_j')) \mid (p_i, q_j) \in \mathrm{Y}^+, (p_i, q_{j'}) \in \mathrm{Y}^-\}, \mathrm{Y}^+$ denotes the observed learner English vocabulary learning interactions ($y_{ij} = 1$) and Y^- denotes a random subset of unobserved learner English vocabulary learning interactions sampled. For the training of the entire KCPE-SR model, end-to-end training is performed by using the Adam optimizer with the optimization objective of minimizing both L_{CF} and L_{KGE}, see Equation (8):

$$\mathrm{L} = \mathrm{L}_{CF} + \lambda_1 \mathrm{L}_{KGE} + \lambda_{thera} \|\Theta\|_2 \tag{8}$$

where Θ contains the parameters in the embedded representation layer of entity relationships in the knowledge graph, the recurrent neural network layer of the embedded representation of knowledge concept paths, and the recurrent neural network layer of the learner sequence modeling, and λ_1 and λ_{rhera} are manually set hyperparameters.

2.3. Personalized Learning System for College English Vocabulary

Based on the above proposed college English vocabulary learning library construction method and personalized learning path recommendation and optimization model, this paper constructs a college English vocabulary personalized learning system as shown in Figure 1. The learning system constructed in this study mainly consists of five sub-modules, namely, determining personalized learning objectives, learning common core knowledge from textbooks, personalized learning tasks, blended classroom learning, and classified multiple teaching evaluation.

Accurately matching English vocabulary learning materials through the corpus, with high vocabulary recurrence rate of extracurricular reading materials and co-core video learning parts, creating more opportunities to encounter core professional vocabulary for the second or third time in a short period of time, and contributing to vocabulary acquisition. The 15-20 seed words function of the Sketch Engine corpus platform is utilized to quickly match learning materials with high reproduction rates, overcoming the obstacles and bottlenecks of English teachers who do not belong to the professional discourse circle, and the time-consuming and inefficient search for professional information, and solving the difficulties of English teachers through technological means. The purpose is to improve the efficiency of students' vocabulary acquisition. Through multimodal personalized assignments, we can break the time and space limitations of English vocabulary classroom, create a relaxing English vocabulary learning environment, improve English vocabulary application and output opportunities, enhance the comprehensive ability of listening, speaking, reading and writing, and overcome the situation of “mute English”. At the same time, in the multimedia era, students' information literacy is high, and reading pictures and watching images have natural affinity for them. The form of homework to present what they think and feel after reading in the way of group cooperation and video shooting is novel and unique for them, and at the same time, it is favorable to be used for personal oral presentation and ability enhancement, which breaks through the limitations of classroom teaching with a large number of students and fewer opportunities for students to practice their English vocabulary, and expands limited classroom learning time to outside the classroom. This is a breakthrough from the limitations of classroom teaching where there are many students and few opportunities for individual students to practice English vocabulary. The group preparation of PPT, report outline, group discussion plan, etc. is conducive to the exercise of English vocabulary input skills such as written expression ability. Meanwhile, in the process of college English vocabulary learning, extracurricular learning is organized through group cooperative learning and inquiry-based learning, which aims to cultivate students' sense of cooperation, critical thinking ability, independent learning, etc., and plays a role in the improvement of students' comprehensive qualities such as analytical and problem-solving ability and quality of thinking.

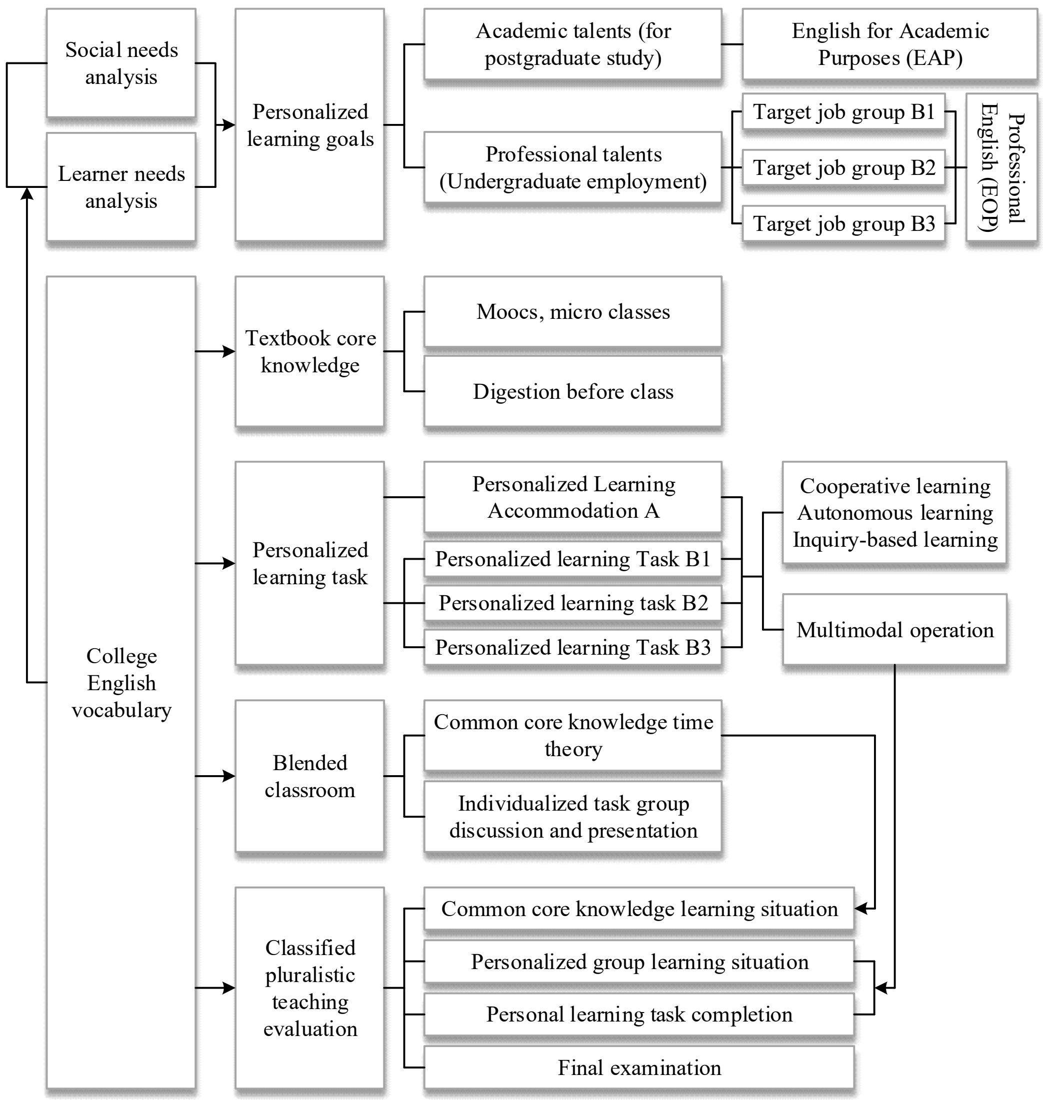


Fig. 1. English vocabulary personalized learning system

3. Results and discussion

3.1. Quasi-experimental research design

In order to study the guiding significance of the Personalized Intelligent Teaching Model of College English Vocabulary in English Vocabulary Learning at the college level, a teaching practice was carried out for first-year students based on this model. This teaching experiment was carried out in a college in H city in September 2023 for a period of four months, and two classes of freshman year were selected to carry out teaching activities.

3.1.1. Quasi-experimental subjects. The experimental subjects were the first-year students of a university in H. The study took the first-year English major class 1 as the experimental class, which consisted of 69 students, including 42 female and 27 male students. The control class was 68 students in the class 5 of freshman English majors, including 38 girls and 30 boys. Using a one-way comparative isogroup experimental method, in which the personalized learning system is the independent variable in the research process, the dependent variable is the students' college English vocabulary learning effectiveness, knowledge mastery, and academic performance, and the control variable is that both classes are taught English vocabulary teaching English vocabulary learning by the same teacher.

3.1.2. Quasi-experimental methods:

1) Quasi-experimental hypothesis. This study constructs an English vocabulary personalized learning system assisted by artificial intelligence recommendation algorithm, and carries out the design of university English vocabulary personalized teaching under the application of this system. In order to test the practical application effect of university English vocabulary personalized wisdom teaching mode in English vocabulary classroom teaching, this paper intends to carry out the following hypotheses.

(1) English classroom teaching under the personalized teaching mode of college English vocabulary improves students' English vocabulary academic performance.

(2) The application of college English vocabulary personalized teaching mode in English classroom teaching can improve students' English vocabulary knowledge mastery rate.

(3) Personalized teaching of college English vocabulary helps to improve students' comprehensive application of English vocabulary.

2) Quasi-experimental design. This study aims to measure and count the differences in learners' academic performance, knowledge mastery, and comprehensive application ability after going through college English vocabulary smart teaching and traditional classroom, thus launching a quasi-experimental study on the personalized teaching model of college English vocabulary. The effectiveness of the personalized learning path generation and optimization system was verified through pre- and post-test papers and questionnaires. To carry out the quasi-experimental study, a control class and an experimental class were selected for teaching. The experimental class was taught English vocabulary under the personalized teaching model, while the control class received English vocabulary teaching under the traditional teaching model. Before implementing the quasi-experiment, pre-test papers were used to obtain the pre-test scores of the two classes, and post-test papers were administered to the two classes at the end of the semester to obtain the changes in students' scores during the period. A post-test questionnaire was also administered to the two classes at the end of the semester to clarify the differences in knowledge acquisition and vocabulary application skills between the two classes.

3.2. English Vocabulary Learning Achievement Test Results:

1) Analysis of changes in the number of exams taken. Between September 1 and December 30, 2023, a total of 20 vocabulary exams were completed by freshman English majors 1 and 5 at School H. The number of students completing the exams varied from one to two. It is important to clarify that this training remains voluntary, so the number of people completing the online English vocabulary exams varies each time. This paper plots the trend of the change in the number of people participating in the 22 vocabulary exams in the two classes as shown in Figure 2.The total number of students completing the online English vocabulary exams in Class 8 shows a significant decrease from 12. Due to the

voluntary nature of the exam, a significant number of students gave up taking the exam because of the longer time interval. The number of students taking the English vocabulary exam in class 8 declined more from the 12th (61) to the 20th (51). This shows that the students in this class do not have a coherent learning of English vocabulary and have not formed a good habit of independent learning. Once interrupted for a long period of time, it is difficult to regain the passion for learning. On the other hand, the number of students taking the 20th vocabulary test in Class 1, where the personalized English vocabulary learning system is applied, remains at 65.

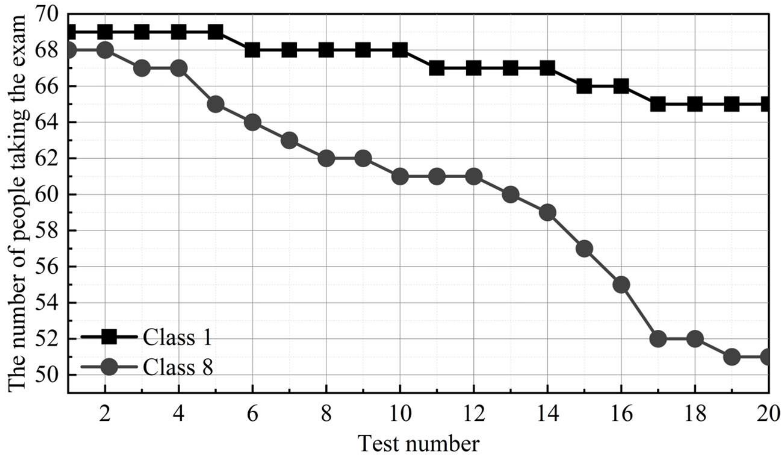


Fig. 2. The total number of English vocabulary tests has changed

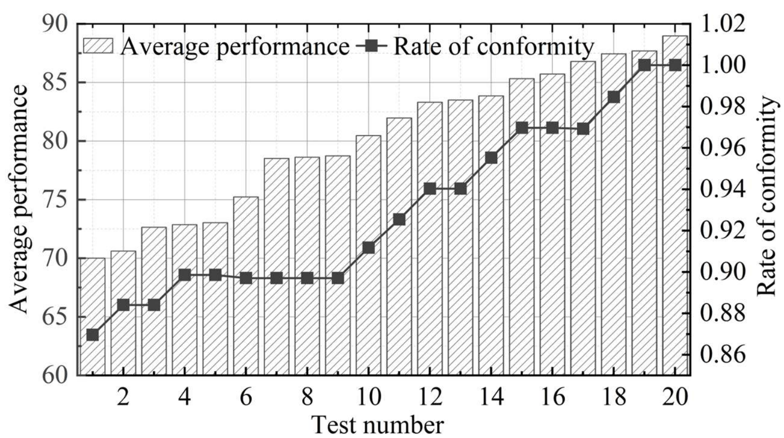
2) Changes in the distribution of English vocabulary test scores. In addition, the results of segmented statistics on students' scores are shown in Table 1. Specifically analyzing the distribution of the number of people in each achievement band, it can be found that from the 3rd examination, the number of students in class 1 who scored above 90 in the examination showed a rapidly increasing trend, increasing from 9 to 19 at the first time. The medium and high scores achievement band always had a sizable number of students, while the number of students in this score band below 60 participating in the vocabulary exam decreased significantly. With this finding, it can be further concluded that there is a strong correlation between students' self-discipline and their English vocabulary test scores. Conversely, poorer self-discipline in English vocabulary learning will also be reflected in each test score, and the negative feedback between the two will easily lead to students' gradual self-abandonment. Therefore, this paper argues that the application of personalized college English vocabulary teaching system can stimulate students' interest and self-discipline in learning English vocabulary and achieve the effect of improving vocabulary scores.

Table 1. The distribution of vocabulary test scores

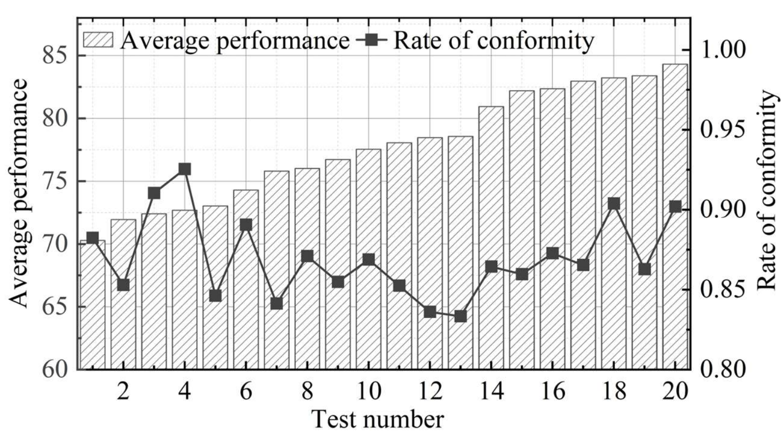
Test number	<60		60-75		75-90		>90	
	Class 1	Class 8	Class 1	Class 8	Class 1	Class 8	Class 1	Class 8
1	9	8	29	15	22	24	9	21
2	8	10	29	26	19	14	13	18
3	8	6	26	21	16	17	19	23
4	7	5	26	20	17	15	19	27
5	7	10	26	27	16	17	20	11
6	7	7	25	16	16	19	20	22
7	7	10	24	10	16	17	21	26
8	7	8	23	26	14	24	24	4
9	7	9	23	11	12	22	26	20

10	6	8	23	13	13	15	26	25
11	5	9	22	16	13	21	27	15
12	4	10	21	23	14	19	28	9
13	4	10	18	25	17	19	28	6
14	3	8	17	29	18	19	29	3
15	2	8	16	11	17	24	31	14
16	2	7	13	28	17	11	34	9
17	2	7	10	24	13	17	40	4
18	1	5	7	17	16	13	41	17
19	0	7	6	16	15	21	44	7
20	0	5	6	28	7	15	52	3

In order to analyze the overall performance of the English vocabulary exams of the two classes, this paper plots the results of the trend changes in the average scores and pass rates of the 20 English vocabulary exams of the two classes as shown in Fig. 3, with (a) and (b) showing the results of the analysis of the performance of the students of Class 1 and Class 8, respectively. The results of the 20th exam results of the students who took the exam in class 1 (88.97) are overall higher than the average score of the students in class 8 (84.31). Relating the average scores to the overall number of students who took the exam as mentioned above can be used to analyze the quality of teaching and learning in both classes. In this paper, we believe that the higher average grade and pass rate of English vocabulary exam of students in class 1 can reflect that the personalized college English vocabulary learning path generation and optimization system assisted by artificial intelligence technology can help to improve the English vocabulary scores of students in colleges and universities, and that hypothesis A is valid.



(a) Class 1



(b) Class 8

Fig. 3. Change in English vocabulary test scores

3.3. Analysis of English Vocabulary Learning Ability

According to the established research program, in order to better understand whether there is a difference in the level of English vocabulary knowledge mastery between the two classes after the experiment, this experiment conducted a post-test and questionnaire survey of English vocabulary knowledge for the experimental subjects at the end of the experiment. The questionnaire was designed to categorize the assessment of college English vocabulary learning ability into three dimensions of evaluation indexes, namely, the dimension of vocabulary knowledge (W1) and the dimension of comprehensive application ability of English vocabulary (W2). The vocabulary knowledge dimension consists of five aspects: fluent spelling of English vocabulary (W11), using English vocabulary in conversation (W12), mastering a certain amount of vocabulary (W13), familiarizing oneself with multi-vocabulary utterances (W14), and standardizing the pronunciation of vocabulary (W15), respectively. The W2 dimension consists of conversations in English in different contexts (W21), understanding of the foreign cultural backgrounds (W22), using English vocabulary in daily life anxiety (W23), mastering the meaning of vocabulary in different contexts (W24), and clarifying the differences between Chinese and English cultural backgrounds (W25). The assessment questions were designed using a Likert scale, with a score of 1-5 representing five levels of agreement. The scoring of each dimension was tallied using the sum of the test question scores, and the level of proficiency in that dimension was measured by the level of statistical scores.

3.3.1. Analysis of vocabulary knowledge acquisition. The statistical results of the software statistical analysis in measuring the topic scores in the English Vocabulary Knowledge Dimension (W1) are shown in Table 2. In terms of the mean scores of English vocabulary knowledge dimension, the mean scores of English vocabulary knowledge dimension of students in class 1 and class 8 were 20.00 and 17.40, and class 1 was slightly higher than class 8. In order to test whether there is a difference between the experimental class and the control class in the English vocabulary knowledge dimension after the experiment, the original hypothesis was proposed, and the results of the analysis of the differences between the two classes in the English vocabulary knowledge dimension after the experiment are shown in Table 3. Using independent samples t-test to test the variance of English vocabulary knowledge dimensions of the two classes, under the assumption of variance chi-square condition, according to the results of the variance chi-square test of the post-test scores of the experiment, F=0, significance Sig=0.965>0.05, which indicates that the variance of the two classes is homogeneous, and the test of variance can be carried out. In the result of "T-test for equality of means", the T-value is -0.526, significance Sig=0.048<0.05, hypothesis B is valid, which means that there is a significant difference between the scores of English vocabulary knowledge dimensions of the two classes.

Table 2. Knowledge dimension measurement of English vocabulary

Class		N	Min	Max	Mean	SD
Class 1	W11	69	2.62	4.86	4.45	1.137
	W12	69	2.43	4.83	4.43	1.073
	W13	69	2.08	4.93	3.71	1.075
	W14	69	2.43	5	3.58	1.113
	W15	69	2.44	4.94	3.83	0.888
	W1	69	12	24.56	20	4.821
Class 8	W11	68	1.52	4.24	3.05	0.871

	W12	68	1.65	4.46	3.13	0.98
	W13	68	1.33	4.21	3.06	1.007
	W14	68	1.75	4.11	3.65	1.081
	W15	68	1.83	4.01	3.37	0.895
	W1	68	8.08	21.03	17.40	4.136

Table 3. Independent sample T test of English vocabulary knowledge

Dimension		Variance equivalence test		The average value of t is tested					
		F	Sig	t	Sig	Mean difference	Standard error difference	Lower limit	Upper limit
W1	Assumed variance homogeneity	0.000	0.965	-0.632	0.048	-0.526	0.953	-2.364	1.263
	Unassumed variance homogeneity	——	——	-0.632	0.048	-0.526	0.953	-2.364	1.263

3.3.2. Analysis of the Results of the Measurement of Comprehensive English Vocabulary Application Skills. The statistical results of the scores of the topics measured in the dimension of comprehensive application ability of English vocabulary (W2) through the software statistical analysis are shown in Table 4. On the average score of the dimension of comprehensive application ability of college English vocabulary, the average total score of students in class 1 is 20.49, which is 4.49 points higher than the comprehensive application ability score of students in class 8. In the comparison of the scores of the sub-dimensions of the comprehensive application ability of English vocabulary, the scores of the students of class 1 and class 8 in the dimension of mastering the meaning of words in different contexts are 4.15 and 2.92, which indicates that the two classes have opened up a big gap in the scores of the sub-dimensions of the comprehensive application ability of English vocabulary.

Table 4. The analysis of the dimensional analysis of lexical comprehensive application

Class		N	Min	Max	Mean	SD
Class 1	W21	69	1.67	4.75	3.75	0.978
	W22	69	2.31	4.59	4.15	1.073
	W23	69	3	4.96	4.15	1.075
	W24	69	1.5	4.63	4.15	1.113
	W25	69	1.99	4.69	4.29	0.888
	W2	69	10.47	23.62	20.49	5.698
Class 8	W21	68	1.4	4.18	3.34	0.963
	W22	68	1.42	3.8	3.36	0.98
	W23	68	1.23	3.81	3.34	1.007
	W24	68	1.35	3.92	2.92	1.081
	W25	68	1.29	3.83	3.12	0.895
	W2	68	6.69	19.54	16.08	4.265

In order to test whether there is a significant difference in the dimension of comprehensive English vocabulary use ability between Class 1 and Class 8 in the post-test, the original hypothesis was formulated that there is a significant difference between the two classes in the dimension of comprehensive English vocabulary use ability after the experiment. The results of the test of variance of the two classes' comprehensive English vocabulary use ability using the independent samples t-test are shown in Table 5. Under the assumption of chi-square condition, F=0.026, significance Sig=0.958>0.05, indicating that the variance is homogeneous and can be analyzed by ANOVA. In the "T-test for equality of means", the significance Sig=0.032<0.05, the original hypothesis is valid, which means that under the application of personalized English learning path recommendation and optimization system, the students of class 1 are significantly higher than those of class 8 in terms of the comprehensive application of English vocabulary at university.

Table 5. Independent sample test results

Dimension		Variance equivalence test		The average value of t is tested					
		F	Sig	t	Sig	Mean difference	Standard error difference	Lower limit	Upper limit
W2	Assumed variance homogeneity	0.026	0.958	-0.528	0.032	-1.236	0.854	-2.958	1.326
	Unassumed variance homogeneity	——	——	-0.528	0.032	-1.236	0.854	-2.958	1.326

4. Conclusion

This paper uses collaborative filtering algorithm and knowledge graph to realize the recommendation and optimization of personalized path in college English vocabulary learning, and designs an intelligent learning system for college English vocabulary based on this method. Experiments are carried out to explore the effectiveness of the learning system, and it is found that under the assisted learning of the system, the English vocabulary learning performance of the students in the experimental class gradually improves, and the number of students scoring more than 90 points in the last exam reaches 52. At the same time, the test of English vocabulary learning ability found that in the average score of English vocabulary knowledge dimension, the average score of English vocabulary knowledge dimension of the experimental class and the control class was 20.00 and 17.40, and the score of vocabulary comprehensive application ability of the experimental class was higher than that of the control group students. This shows that the personalized English learning path recommendation and optimization system is of great help to students' English vocabulary learning.

Future research can improve the construction of the learner model in the university English vocabulary personalized learning system, and achieve more accurate and personalized vocabulary learning resources recommendation through the introduction of the clustering model of learners' learning styles and interests.

References

[1] Han, C. (2019). Design and Application of the Intelligent Learning Platform for College English Based on the New Engineering Concept. In Application of Intelligent Systems in Multi-modal Information Analytics (pp. 60-66). Springer International Publishing.

[2] Wei, C. W., Kao, H. Y., Lu, H. H., & Liu, Y. C. (2018). The effects of competitive gaming scenarios and personalized assistance strategies on English vocabulary learning. Journal of Educational Technology & Society, 21(3), 146-158.

[3] Zhang, X., & Chen, L. (2021). College English smart classroom teaching model based on artificial intelligence technology in mobile information systems. Mobile Information Systems, 2021(1), 5644604.

[4] Shi, X. (2017). Application of Multimedia Technology in Vocabulary Learning for Engineering Students. International Journal of emerging technologies in learning, 12(1).

[5] Govindasamy, P., Yunus, M. M., & Hashim, H. (2019). Mobile assisted vocabulary learning: Examining the effects on students' vocabulary enhancement. Universal Journal of Educational Research, 7(12), 85-92.

[6] Liu, J., & Liu, M. (2024). Application of CLT in College English Vocabulary Teaching. International Journal of Mathematics and Systems Science, 7(3).

[7] Dindar, M., Ren, L., & Järvenoja, H. (2021). An experimental study on the effects of gamified cooperation and competition on English vocabulary learning. British Journal of Educational Technology, 52(1), 142-159.

[8] Li, J., Cummins, J., & Deng, Q. (2017). The effectiveness of texting to enhance academic vocabulary learning: English language learners' perspective. Computer Assisted Language Learning, 30(8), 816-843.

[9] Zhou, J. (2019, June). Application investigation and technological development of intelligent product in college English teaching. In Journal of Physics: Conference Series (Vol. 1237, No. 2, p. 022069). IOP Publishing.

[10] Li, R. (2021). Does game-based vocabulary learning APP influence Chinese EFL learners' vocabulary achievement, motivation, and self-confidence?. Sage Open, 11(1), 21582440211003092.

[11] Baker, D. L., Ma, H., Polanco, P., Conry, J. M., Kamata, A., Al Otaiba, S., ... & Cole, R. (2021). Development and promise of a vocabulary intelligent tutoring system for Second-Grade Latinx English learners. Journal of Research on Technology in Education, 53(2), 223-247.

[12] Zhang, Z. (2018). A cognitive study of college students' English vocabulary based on electroencephalogram. NeuroQuantology, 16(5).

[13] Wang, B. T. (2017). Designing mobile apps for English vocabulary learning. International Journal of Information and Education Technology, 7(4), 279.

[14] Duong, T. M., Tran, T. Q., & Nguyen, T. T. P. (2021). Non-English majored students' use of English vocabulary learning strategies with technology-enhanced language learning tools. Asian Journal of University Education, 17(4), 455-463.

[15] Gao, X. (2021). Role of 5G network technology and artificial intelligence for research and reform of English situational teaching in higher vocational colleges. Journal of Intelligent & Fuzzy Systems, 40(2), 3643-3654.

[16] Yu, L. (2021). Intelligent Recommendation System for English Vocabulary Learning–Based on Crowdsensing. Applied Mathematics and Nonlinear Sciences.

[17] Wang, L. (2020). Investigation of the present situation of intelligent app in college students' vocabulary learning. Journal of Language Teaching and Research, 11(5), 764-768.

[18] Liao, L. (2022). Artificial intelligence-based English vocabulary test research on cognitive web services platforms: User retrieval behavior of English mobile learning. International Journal of E-Collaboration (IJeC), 19(2), 1-19.

[19] Wang, Y., Wu, J., Chen, F., Wang, Z., Li, J., & Wang, L. (2024). Empirical assessment of AI-powered tools for vocabulary acquisition in EFL instruction. IEEE Access.

[20] Zhang, W., & Wang, X. (2023). Intelligent Algorithm Evaluation of Incidental English Vocabulary Acquisition in Complex Reading Tasks. Journal of Electronics and Information Science, 8(3), 17-32.

[21] Cui, J. (2020). Application of deep learning and target visual detection in English vocabulary online teaching. Journal of Intelligent & Fuzzy Systems, 39(4), 5535-5545.

[22] Chang, L. (2024, June). Improvement of Vocational English Vocabulary Learning Based on Adaptive Algorithms. In Proceedings of the 2024 International Conference on Intelligent Education and Computer Technology (p. 1).

[23] Tlili, A., Hattab, S., Essalmi, F., Chen, N. S., Huang, R., Chang, M., & Solans, D. B. (2021). A smart collaborative educational game with learning analytics to support English vocabulary teaching. IJIMAI, 6(6), 215-224.

[24] Yang, J. (2021). On the Strategies of English Vocabulary Learning Based on Smart-Phone. In Application of Intelligent Systems in Multi-modal Information Analytics: 2021 International Conference on Multi-modal Information Analytics (MMIA 2021), Volume 1 (pp. 275-283). Springer International Publishing.

[25] Tao Wu,Yu Guo,Shaohua Huang,Lijun Ma,Xifeng Guo & Jiahui Zheng. (2025). A two-channel collaborative filtering process template recommendation algorithm: RCAN - GGCNII - 2C. Advanced Engineering Informatics103033-103033.

[26] Flanagan Jane. (2021). From concepts and systematic reviews: a path to enhancing nursing knowledge development. International Journal of Nursing Knowledge(4),217-217.

[27] John S. Erickson,Henrique Santos,Vládia Pinheiro,Jamie P. McCusker & Deborah L. McGuinness. (2025). LLM experimentation through knowledge graphs: Towards improved management, repeatability, and verification. Journal of Web Semantics100853-100853.

[28] Wenyi Xiao,Huan Zhao,Vincent W. Zheng & Yangqiu Song. (2024). Meta-path based proximity learning in heterogeneous information networks. Data Mining and Knowledge Discovery(1),8-8.

[29] N. Cimmino,D. Amato,R. Opromolla & G. Fasano. (2025). A recurrent neural network-based approach for ballistic coefficient estimation of resident space objects in low earth orbit. Advances in Space Research(2),2088-2107.

[30] Yang Jiahao,Feng Shuo,Zhang Wenkai,Zhang Ming,Zhou Jun & Zhang Pengyuan. (2024). An efficient loss function and deep learning approach for ranking stock returns in the absence of prior knowledge. Information Processing and Management(1).