ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

A Strategic Study on Overseas Chinese Language Teachers' Utilization of Artificial Intelligence Algorithms to Assist the Dissemination of Chinese Culture in Classroom Teaching

Jia Zuo[1,*] *and Dong Feng*[1]

[1] Office of International Exchange and Cooperation, Xianyang Polytechnic Institute, Xianyang, 712000, Shaanxi, China

SUMMARY: *This paper establishes a BERT-BiLSTM-ATT-CRF named entity recognition model containing four levels, mines the semantic relevance of Chinese cultural entity-relationships, and constructs a knowledge graph ontology covering a large number of Chinese cultural resources. A collaborative filtering recommendation algorithm integrating students' long- and short-term interest values and interest change degrees is proposed to improve the relevance and diversity of recommended resources based on the knowledge graph with students' interests, and accurately support Chinese culture dissemination in the classroom teaching of overseas Chinese language teachers. The study shows that the knowledge graph constructed based on an ontology with more than 60% coverage of Chinese culture contains 7 major categories of cultural knowledge resources. Students' short-term interest values in these 7 major categories of cultural resources range from 62.457 to 90.663, which are higher than the long-term values of 61.559 to 89.404, and the rate of interest change is faster. The recommendation algorithm that integrates students' long-term and short-term interests is able to provide students with a recommended list of Chinese culture with an accuracy of 94.165% and a diversity of 88.132% at the same time. Using this algorithm, overseas Chinese language teachers are able to integrate Chinese culture knowledge resources in a timely manner to complete classroom teaching on the basis of grasping students' current and long-term interests.*

KEYWORDS: *BERT-BiLSTM-ATT-CRF; knowledge graph; interest change degree; recommendation algorithm; Chinese culture*

1 Introduction

Chinese language international education aims to promote Chinese language knowledge and cultivate Chinese language learners worldwide [1]. International dissemination of culture, on the other hand, is to transmit the excellent traditional culture, values and ideological essence of the Chinese nation to people around the world [2]. Both are important in promoting communication and understanding between China and the world. However, the traditional Chinese language international education and Chinese culture dissemination have problems such as single strategy and lack of educational resources, which make the two mutually restrictive and are very unfavorable to the quality of Chinese language international education and the effect of Chinese culture dissemination.

In recent years, the Chinese government has introduced a series of policies emphasizing the combination of Chinese language international education and artificial intelligence (AI) to

*xyzy_zj515@163.com
https://doi.org/10.65102/is2026653 Published: 30 April, 2026

further enhance the teaching effect of Chinese language international education and expand the dissemination of Chinese culture. In this context, overseas Chinese language teachers have ushered in the opportunity to utilize AI in classroom teaching to assist Chinese language education and spread Chinese culture [3]. First of all, more accurate language analysis and feedback can be provided through natural language processing technology [4]. The student person can correct pronunciation errors in real time through the intelligent speech recognition system to improve speaking skills [5]. At the same time, the AI system can provide personalized practice and feedback based on students' speech, intonation, grammar and vocabulary usage to help students master Chinese faster [6, 7]. Regarding the application and impact of natural language processing (NLP) technology in the teaching of Chinese international education, the literature [8] examined the effect of the application of NLP tools in the teaching of Chinese as a foreign language, confirmed that it can significantly improve vocabulary, grammar and other language skills through mixed-methods analysis, and pointed out that although there are technological challenges, such as cultural adaptation, the learner's participation is high, and the effect is better than the traditional method, which emphasizes the technology's personalized teaching The positive impact of this technology on personalized instruction is emphasized. Literature [9] explored the design and evaluation of a teaching model for public university Chinese language classes based on natural language processing technology, constructed an integrated teaching method by integrating NLP tools, and analyzed its implementation effect in real classrooms, pointing out that the model can significantly improve students' language proficiency and learning satisfaction, and emphasizing the reference value of the technology for optimizing teaching practice. Literature [10] utilized the construction of a Chinese language learning system, analyzed its application effect in international Chinese language teaching by integrating the three modules of foundation, learning and tools, pointed out that the system could satisfy the needs of more than 90% of the learners, and emphasized its scientific value in enhancing the supply of teaching resources. Literature [11] designs and implements an intelligent educational assistance system for Chinese composition scoring, through integrating natural language processing technology for text analysis and semantic understanding, and studies its effectiveness in providing real-time feedback and personalized guidance, pointing out that the system can effectively improve scoring accuracy and learning efficiency, and emphasizing its potential for educational application. Literature [12] constructs a Chinese speaking teaching evaluation system based on natural language processing and knowledge mapping, and experiments show that its accuracy and personalized satisfaction rate are over 98%, and points out that the system can provide scientific feedback for teaching, and emphasizes its far-reaching impact on promoting the technological development of language education. Literature [13] examines the application of natural language-based intelligent human-computer interaction in teaching Chinese euphemisms to foreign students, and points out that the system can relieve international students' nervousness and improve their learning effect, while improving the recognition accuracy.

Secondly, the application of AI in Chinese international education is also reflected in the development of intelligent teaching systems [14]. These systems can automatically adjust the content and difficulty of teaching according to students' learning progress and ability, realizing truly personalized teaching [15]. Through big data analysis, AI teachers are able to identify students' learning habits and preferences, so as to design more appealing teaching programs for students [16]. In this regard, literature [17] conducts an in-depth study on the design of the intelligent teaching system for Chinese international education, realizes the personalized configuration and dynamic evaluation of learning resources through the introduction of elite optimization algorithms with feedback mechanism, and points out that the system can significantly improve teaching stability and learning efficiency, and better meet the needs of

Chinese language learners. Literature [18] analyzes the role of AI in revolutionizing the international Chinese language education ecosystem, and by exploring the challenges and opportunities it brings, studies the advantages and limitations of AI technology in digital teaching and puts forward practical strategies that teachers should flexibly combine the technology to enhance teaching effectiveness. Literature [19] discusses the application of AI technology in international Chinese language education, analyzes how it can overcome the resource and cultural barriers, and points out that intelligent curriculum design and learning management can improve the teaching effect, while emphasizing the need to develop towards cultural localization and learning personalization in the future, so as to promote the continuous improvement of education quality. Literature [20] uses deep learning to optimize the AI intelligent teaching system from the perspective of tailored teaching, analyzes its potential to improve teaching accuracy and personalization by incorporating learning styles and habits into student models, and points out that the system has a broad development prospect. Literature [21] examined the application effect of AI education system in Chinese language teaching, and by investigating the feedback from teachers and students in less developed regions, it pointed out that the system can provide personalized resources and learning analysis, and emphasized that it has a positive effect on promoting educational equity despite the challenges of equipment and interaction design. Literature [22] studies the application of AI and other digital technologies in international Chinese language education, analyzes their potential for innovation in teaching modes and teacher training, and points out the need to pay attention to the limitations of practice and data ethics, which provides theoretical and practical references for educational transformation.

In addition, AI can greatly enrich the resources for teaching Chinese as a foreign language [23]. Through machine learning algorithms, AI is able to filter teaching materials suitable for learners of different levels from a huge amount of online resources, including video, audio, text and other forms [24]. These resources can not only provide the basic elements of language learning, but also enable students to understand Chinese culture and society, and enhance their interest and motivation in learning [25]. For the research on AI-enriched teaching resources, literature [26] describes the current status of AI applications in educational resource generation and teaching aids, discusses its potential to promote personalized learning and enhance teaching equity, and points out the challenges it faces, such as ethics and data privacy, and emphasizes the critical impact of technological convergence on future development. Literature [27] discusses the application of AI in customizing teaching resources from the perspective of K-12 teachers, analyzes its specific methods in developing literacy and enriching extracurricular activities, and points out six challenges that must be addressed, such as ethics and equity, and emphasizes the need for teachers to incorporate innovative strategies to ensure the inclusive development of educational resources. Literature [28] proposes a solution for automatically generating and adapting accessible educational resources using AI, which analyzes the actual needs of students with disabilities, emphasizes the accurate annotation of metadata such as image descriptions and video subtitles, and points out that the method can improve the accessibility and search efficiency of the resources, and provides technical support for the promotion of digital inclusion. Literature [29] generated an improved vocabulary resource DICO-2 by enriching the semantic information of the French educational vocabulary resource DICO and recalculating the word polarity using multiple machine learning models, and pointed out that it can significantly improve the classification effect of opinion mining in the field of education, and provide technical support of natural language processing for the enrichment of teaching resources by AI.

In this paper, we design and implement the BERT-BiLSTM-ATT-CRF named entity recognition model, which is applied to entity naming for ontology construction in Chinese culture domain. We utilize the bi-directional transformer structure of BERT to extract word vectors containing rich contextual information from the training samples, and input them into the BiLSTM layer to capture long-distance bi-directional semantic features through forward and backward LSTM, and then the Attention mechanism mines and assigns different weights to the bi-directional semantic features, and then combines them with the CRF's transfer matrix to sort the word labeling information of the text, and then outputs the final predicted Naming results. Calculate the interest value and interest change degree of students' short-term and long-term interests, and classify the interest tendency preference of different students. Incorporate students' short-term and long-term interests into the resource recommendation algorithm, calculate the Jaccard similarity between the resource recommendations and students' interests, and output a diverse list of recommendations in the Chinese culture knowledge map that best fit students' interests for overseas Chinese teachers' classroom teaching.

2 Chinese Culture Knowledge Mapping

2.1 Design and construction of Chinese cultural ontology

In this paper, we propose the process of constructing an ontology for the Chinese culture domain by combining the ontology construction method and the actual situation of the Chinese culture domain, to build a knowledge map containing a large amount of Chinese culture content, and to increase the Chinese culture resources available for classroom teaching by overseas Chinese language teachers. Figure 1 shows the process of constructing the ontology of Chinese culture domain. First, the domain scope is clarified according to the characteristics of the ontology domain and the teaching needs, and then information is collected to determine whether there is a readily available ontology model for the domain scope; if so, the ontology attributes and ontology relations can be defined directly. If there is no directly available domain ontology model, it is necessary to list the important terms of the ontology domain and define the hierarchical structure of the domain ontology classes, and then finally define the relationships and attributes of the ontology classes.

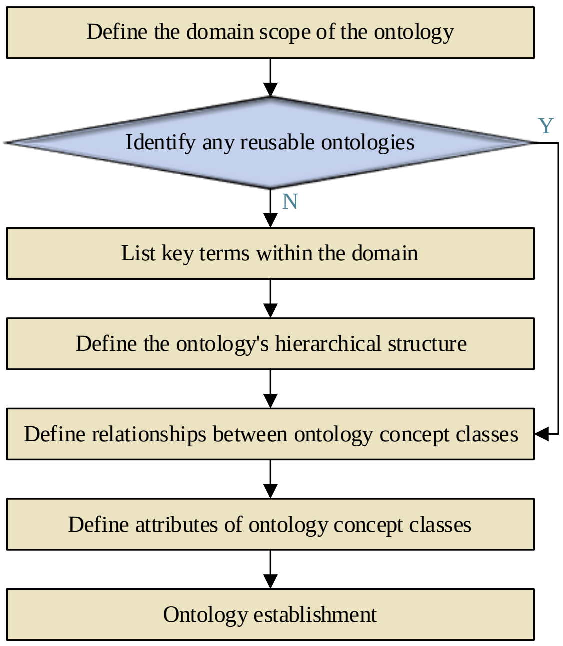

Figure 1: Process of constructing the ontology in the field of Chinese culture

2.2 Chinese Named Entity Recognition Based on BERT-BiLSTM-ATT-CRF

The named entity recognition model LSTM-CRF based on deep learning can only take into account the information above at the current moment when processing the information, and cannot capture the information below, which will greatly reduce the semantic understanding of words. In order to further improve the performance of the named entity recognition model, we seek to improve it from the perspective of word vectors, and propose the BERT-BiLSTM-CRF named entity recognition model that generates word vectors by using the BERT pre-training model, which is mainly characterized by migrating a large amount of semantic information through embedding and then fine-tuning the downstream task to generate word vectors that can characterize the meaning of the word in multiple senses. The main feature is to migrate a large amount of semantic information through embedding and then fine-tune the downstream task to generate word vectors that can characterize multiple meanings of a word.

In this paper, the attention mechanism is introduced on the basis of the BERT-BiLSTM-CRF model to do the Chinese named entity recognition task for Chinese culture ontology construction and knowledge graph building, to highlight the important features of the context, to help the model to better perform the entity recognition task, and to improve the accuracy of the content recommendation of knowledge graphs and the classroom teaching of overseas teachers. Figure 2 shows the structure of the BERT-BiLSTM-Att-CRF model.The overall design idea of the BERT-BiLSTM-Att-CRF model is as follows: use the BERT model as a pre-training model, generate word vectors based on the contextual semantic information with the help of the powerful feature extraction ability of the Transformer structure in BERT; utilize the BiLSTM's The forward LSTM and backward LSTM in BiLSTM are utilized to fully capture the semantic information of the context of the feature vectors; The results are input to the Attention layer, which gives more weight to the part of the semantic sequence that is more helpful in recognizing the named entity through the attention mechanism, and further learns the potential semantic features of this part of the information to improve the recognition accuracy; and combines with the state transfer matrix in the CRF layer to sequence the output results and obtain the globally optimal sequence.

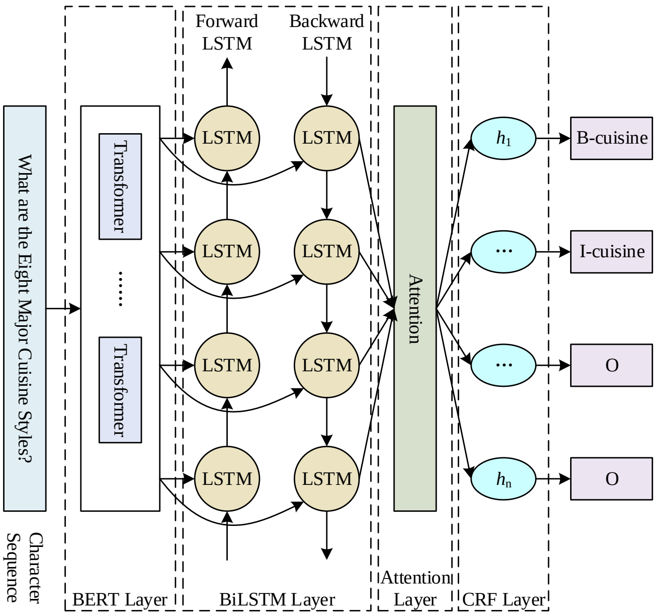

Figure 2: The model structure of BERT-BiLSTM-Att-CRF

2.2.1 BERT layer

BERT utilizes the attention mechanism contained in the bi-directional Transformer structure for pre-training, and learns deep bi-directional representations by combining captured word-to-word and sentence-to-sentence contextual features to dynamically capture contextual semantic feature information and obtain semantically rich word vectors.

The input to BERT can be a single sentence or sentence pair.The word embedding of BERT is obtained by synthesizing three parts: token embedding, fragment embedding and positional embedding. Before being fed into the token embedding layer, the input sequence first needs to complete the tokenization process, which refers to inserting [CLS] tags at the beginning of the input sequence and [SEP] tags at the end of it. The token embedding layer is mainly used to represent each character as a vector of a certain dimension, i.e., word vector representation. The fragment embedding layer is mainly used to distinguish the vector representation of two different sentences in a sentence pair. The order information of the input sequence is learned by introducing the positional embedding layer.

The BERT model pre-training consists of two tasks, the first task is to randomly mask a small number of words in a given utterance and replace them with [MASK] markers at the original position to allow the model to predict the word based on the context. The second task is to determine whether there is a relationship between the contextual sentences in the training sample. This task is based on the [MASK] masking task: two sentences are randomly obtained from the pre-training corpus; the model is allowed to predict whether there is a relationship between Sentence A and Sentence B, and the predictions are labeled using different tags.

2.2.2 BiLSTM layer

The BiLSTM structure consists of a forward LSTM and a backward LSTM.With the joint action of the three gate mechanisms of the LSTM, the LSTM cell state is continuously updated, the LSTM structure is operationalized, and the model is able to capture long-distance dependencies in utterances.

By combining two unidirectional LSTMs with opposite directions (forward LSTM and backward LSTM) into a bidirectional LSTM, it is possible to capture both forward and backward semantic feature information, and the more adequately the semantic feature information is grasped, the more it will help in the backward process tasks. The processing flow of this layer is roughly as follows: after the sentence vectors obtained from the BERT model, the results are fed into the BiLSTM model, and after the LSTM and backward LSTM structures, the forward and backward features are obtained respectively, and at last, the results of the two features are spliced together and combined for output, i.e., bidirectional semantic features.

2.2.3 Attention layer

The advantage of Attention mechanism is that it can devote more attention to the processing of key information and reduce the attention to useless information, thus improving the efficiency of resource utilization. Although the BiLSTM in the upper layer of the model can get the context information, it pays the same attention to all the information in the context, and cannot catch the key, and then there is a situation of "ignoring" some key information in the context. In view of this, an Attention layer is introduced behind the BiLSTM layer to further explore the potential semantic feature information between texts. The process is as follows: after obtaining the sentence vectors through the BERT layer, extract the contextual semantic features through the BiLSTM layer; through the Attention mechanism and combined with the importance of the feature vectors, assign the corresponding weights to better highlight the important potential semantic features in the context; and output semantically rich and focused contextual feature vectors.

2.2.4 CRF layer

For the last layer of the model, its main purpose is to predict the annotation information of the text based on the previous feature vectors of the text sequence.There is a transfer matrix in the CRF, which takes into account the transfer probability of the word labels so as to ensure the order and interdependence of the word labels, and constraints are added to ensure the legitimacy of the word labels. Therefore it was decided to use the CRF layer as the output layer of the whole model.

In the CRF layer, for each input $X=\left(x_1,x_2,\cdots,x_n\right)$, a sequence of predicted labels $y=\left(y_1,y_2,\cdots,y_n\right)$, the score of this prediction is shown in equation (1).

$$s(X,y)=\sum_{i=0}^{n}A_{y_i,y_{i+1}}+\sum_{i=1}^{n}P_{i,y_i} \tag{1}$$

where P_{i,y_i} denotes the state probability that the output of the i th position is y_i, and $A_{y_i,y_{i+1}}$ denotes the transfer probability from y_i to y_{i+1}.

Then, all scores are normalized as shown in equation (2).

$$P\left(y\middle|X\right)=\frac{e^{s(X,y)}}{\sum_{\tilde{y}\in Y_x}e^{s(X,\tilde{y})}} \tag{2}$$

In order to maximize the probability of correctly predicting the sequence, a log-likelihood function is used on both sides of Eq. (2), and the result of calculating the loss function is shown in Eq. (3) below.

$$\log\left(P\left(y\middle|X\right)\right)=s(X,y)-\log\left(\sum_{\tilde{y}\in Y_x}e^{s(X,\tilde{y})}\right) \tag{3}$$

Next, the Viterbi algorithm is utilized to solve the s scores corresponding to all possible y sequences, and finally the optimal solution is taken as the prediction output according to equation (4).

$$y^{*}=\arg\max_{\tilde{y}\in Y_x}s\left(X,\tilde{y}\right) \tag{4}$$

3 A Study of the Diversity of Recommendation Algorithms Incorporating Students' Long- and Short-Term Interests

3.1 Recommendation Algorithm Diversity Optimization Process

In order to better meet the diversified needs of recommendation lists, increase the possibility of overseas Chinese language teachers to disseminate Chinese culture in classroom teaching, and improve the traditional user-based recommendation method, this chapter proposes a collaborative filtering recommendation algorithm that integrates knowledge graph after classifying the short-term and long-term interests of student users. Considering the classification of different students' long-term and short-term interest tendencies, we calculate the degree of interest and the change of interest, and utilize the back-and-forth connection between knowledge points in the knowledge graph to extend the possibility of recommending Chinese culture course resources, which helps the recommender system to avoid content narrowing and improve the diversity of the recommender list while losing as little as possible of the recommendation accuracy. Figure 3 shows the process of optimizing the diversity of the recommendation algorithm.

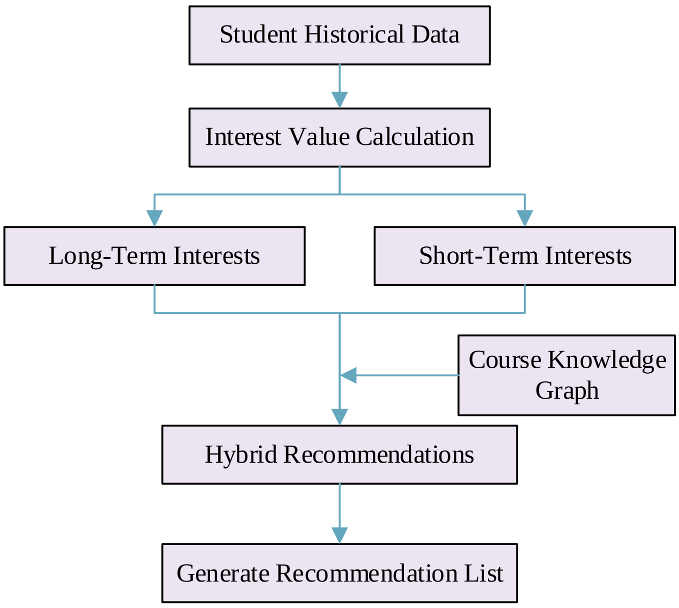

Figure 3: Process of optimizing the diversity of recommendation algorithms

3.2 Analysis of students' interests

3.2.1 Analysis of long and short-term interests

In a broad sense, a user's short-term interest refers to a change in the user's interest orientation that occurred within a short period of time in the past, and is not inherent in the user's long-term interests. Short-term interests tend to be more easily influenced by external environmental factors. A high proportion of short-term interests is characterized by greater variability and a wider range of interests. The user's long-term interest refers to the user in the past period of time with a more constant interest orientation, with long-term interest in the user's interest is not very variable and the scope of interest is relatively narrow. When making recommendations, we can generate a diversified recommendation list that integrates long and short-term interests according to a certain weighting.

3.2.2 Calculation of interest values

The definitions of the symbols used in the recommendation algorithms in this chapter are first explained as follows:

$U=\{u_1,u_2,\cdots,u_n\}$ is the set of all student users, $I=\{i_1,i_2,\cdots,i_n\}$ is the set of all course resources, the set of categories to which course resource i belongs is $L_i=\{l_1,l_2\cdots l_n\}$, the rating of course resource i by user U is $r_{u,i}$, the set of all resources in which student users

U have generated rating behaviors is defined as the set of resources of interest to student users C_u, the probability of interest in a resource category based on the history of student users' rating data is defined as P_u, and the probability of interest in a resource category for a certain resource category in the data set l, the student user U 's interest in it is $P_{u,l}$, and the calculation formula is as follows:

$$P_{u,l} = \frac{\sum_{i \in C_u, l \in L_i} \frac{1}{|I_i|}}{|C_u|} \tag{5}$$

where $P_{u,l} \in P_u$. If a resource i belongs to resource category l only, it is considered that i has a weight value of 1 for resource category l, and if i belongs to n resource categories at the same time, it is considered that i has a weight value of $\frac{1}{n}$ for each resource category to which it belongs. Where the larger the value of $P_{u,l}$, the higher the degree of interest of the student user in the resource category, and vice versa. For all the interest categories of the student user, according to formula (5), the value is calculated one by one and added up to get the specific interest degree of the student in a resource category and then classify the student's interest degree. If a student has ratings in most of the resource categories, it means that the student has a wide range of interests; if he/she has ratings in only a few resource categories, it means that the student has a narrow range of interests. The overall student user interest level is calculated as shown in Equation (6).

$$D_u = \sum_{l=1} P_{u,l} \log_{|l|} P_u \tag{6}$$

After calculating the overall interest value, D_u is normalized. A higher value of overall interest indicates that the student user is interested in more categories of course resources, while a lower value indicates that the student user is only interested in a few categories of course resources.

In addition, the change of interest is a major factor in determining the classification of short- and long-term interests. If a student user has a wide distribution of interests but has not added or deleted a resource category over a long period of time, it means that the student has a high degree of interest but not a high degree of interest change, and the student can be regarded as a user who tends to have long-term interests. The cosine similarity of the course resource categories was used to measure the similarity between category l of the student user ratings and set L of the resource categories, and the similarity was calculated to determine the change in the student user's interests.

$$Sim(l,L) = COS(l,L) = \frac{\sum_{i=1}^{n}(l * L)}{\sqrt{\sum_{i=1}^{n} l_i^2} * \sqrt{\sum_{i=1}^{n} L^2}} \tag{7}$$

Set the similarity threshold, and the resource categories l exceeding the threshold are i ones. Define the degree of change in student user interest denoted by C.

$$C = \frac{i}{n-i} \tag{8}$$

Since students' long-term and short-term interests are mainly determined by both the degree of interest and the degree of change in interest over a certain period of time, this paper puts students' short-term and long-term interests on an equal footing. By setting a constant coefficient $\alpha(0.0 < \alpha < 1.0)$ to balance the influence of primary and secondary weights, this paper puts students' short-term and long-term interests on an equal footing. T is the overall interest value of the student user, which is used to indicate the ability of the student user to accept new resources. The calculation formula is shown in equation (9):

$$T = \alpha D_u + (1-\alpha)C \tag{9}$$

Set the value domain division of T to categorize student users' interest orientation preferences, for example, select student users with low interest values to be considered as a group of students with a preference for long-term interests, which have a concentrated and low variability in the rating resource categories, and those with a high level of interest to be considered as a group of students with a preference for short-term interests, which have a discrete and high variability in the rating resource categories.

3.3 Integration of students' long and short-term interests

Fusion recommendation is to fuse the long and short-term interests of student users, and when recommending new resources, it is necessary to consider whether the new resource category is associated with the historical resource category rated by student users, and whether the new resource category is directly accepted by student users. At the same time, it is necessary to consider the categorization of student users with different interests.

Since each resource may not belong to one resource category alone, Equation (11) gives the ratio of the public category to the total category for any two resources. For the calculation of the similarity between the new resource i and the list of student users' historical interest resources, Jaccard similarity is used in this paper. As shown in equation (10).

$$sim(i, j) = \frac{|I_i \cap I_j|}{|I_i \cup I_j|} \tag{10}$$

$$sim(i, R) = \frac{\sum_{j \in R} sim(i, j)}{|R|} \tag{11}$$

Based on the data of all ratings of student users for historical resources, the predicted ratings of students for a resource l are first predicted, and the category to which the resource l belongs is a non-empty set of intersections with the set of resource categories of historical interests of the student users, which are the associated resource categories. The formula for calculating the predicted ratings is as follows:

$$U_{g,l} = \frac{\sum_{i \in C_u, I_i \cap I_l \neq \varnothing} \sum_{C \in \{I_i \cap I_l\}} \frac{r_{u,i}}{|I_i|}}{\sum_{i \in C_u, I_i \cap I_l \neq \varnothing} 1} \tag{12}$$

where $\frac{r_{u,i}}{|I_i|}$ is the ratio of student user's ratings for resource i to the total number of categories, which represents the ratio of ratings.

By using T to indicate the high or low total interest value of the student user. If the student is a user who favors long-term interests, the T weight is assigned lower, and if the student favors short-term interests and the recommended resource is a new resource category, a larger weight influence is assigned to the T value, which guarantees personalization. The overall predicted score of the student user is given by Equation (13).

$$R_g = \lambda \frac{U_g}{MG} + (1-\lambda) T Div(i,R) \tag{13}$$

where the maximum rating (MG) is the highest rating given to a resource by all student users, where U_g is used to divide with it in order to normalize the relevance resource scores. The $\lambda(0.0 \leq \lambda \leq 1.0)$ parameter is used to balance the percentage of relevance versus diversity of the resource category. A λ of 0.0 indicates the lowest relevance and highest diversity; a λ of 1.0 indicates the highest relevance and lowest diversity. $sim(i,R)$ indicates the similarity between the resource i and the list of student user resources, the higher the similarity, the lower the diversity of the resource i to the list of student user resources. The diversity is calculated as shown in equation (14).

$$Dif(i,R) = 1.0 - sim(i,R) \tag{14}$$

The student long and short-term interest fusion recommendation module optimizes the diversity of recommendation lists through the mutual adjustment of U_g and T, while guaranteeing personalization and maintaining a certain level of accuracy.

4 Resource Recommendation Practices Based on Chinese Cultural Knowledge Mapping and Students' Interests

4.1 Ontology-supported Chinese Culture Knowledge Mapping Establishment

4.1.1 Ontology coverage assessment

In the ontology-based construction of Chinese culture knowledge graph, the coverage of the ontology determines the size of the knowledge graph and the subsequent diversity level of resource recommendation. In order to verify the coverage of Chinese culture ontology constructed based on BERT-BiLSTM-ATT-CRF for Chinese named entity recognition in this paper, the original corpus data and post-construction ontology data are evaluated and calculated. Figure 4 shows the calculation results of the coverage of various types of Chinese cultural ontologies. The ontologies include 13 categories: person, craft, organization, dynasty, decree, place, writings, event, title, spirit, carrier, doctrine, and way.The coverage of Chinese cultural knowledge corresponding to the 13 categories of ontologies ranges from 62.33% to 96.22%, which are all higher than 60%. Among them, 82.37% for dynasties, 96.22% for places, 94.51% for events, 88.17% for spirits, and 87.58% for doctrines, the coverage of these five categories of ontologies is above 80%, which is in line with the characteristics of the long history of China and the richness of its cultural achievements, and proves that the constructed ontologies have a high level of coverage while taking into account a high level of credibility.

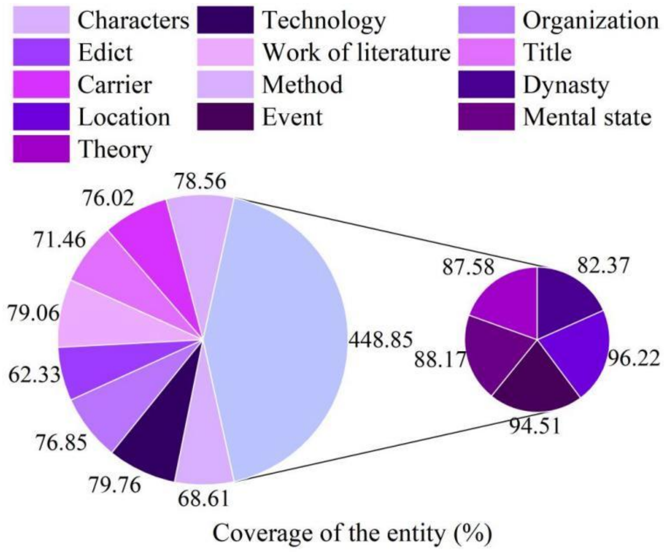

Figure 4: Coverage of various Chinese cultural elements

4.1.2 Thematic Clustering and Feature Word Analysis

The large amount of ontological feature word data in the training text is clustered thematically to construct and complete a knowledge map of Chinese culture, which increases the available resources for overseas Chinese language teachers to disseminate Chinese culture in classroom teaching. Figure 5 shows the theme clustering results. The size of the arrow represents the weight size of the theme in the overall corpus, which is the frequency of the theme in all texts. As can be seen from Figure 5, when clustering the topics into 7 categories, there is no overlap and the distance between the arrows of the topics and their rays is large except for the center point, which indicates that the similarity between the topics is low and the topic clustering is more effective. The theme weights, in order of magnitude, are: 4>1>7>6>2>3>5.

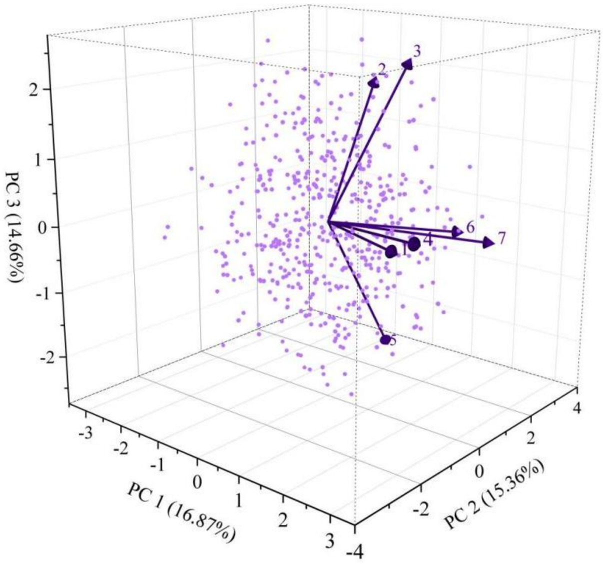

Figure 5: Cluster results of the topic classification

Table 1 presents the statistical results of the top 10 characteristic words for each topic. The top 10 frequently-occurring characteristic words for Topic 4 are Huaxia, dynasty, era name, historiography, political events, great unification, ethnicity, system, people-oriented thought, and civilization, which are in line with the characteristics of Chinese history. Therefore, Topic 4 is named "Chinese History". Based on the specific characteristic words, Topic 1 is named "Chinese Art"; Topic 7 is named "Chinese Festivals"; Topic 6 is named "Chinese Folk Customs"; Topic 2 is named "Chinese Artifacts"; Topic 3 is named "Chinese Economy"; and Topic 5 is named "Chinese Cuisine".

Table 1: Statistics of the top 10 characteristic words for each theme

Topic	Characteristic Words (TOP 10)	Naming
Topic4	China, Dynasty, Regnal Name, History, Political events, Monarchical unification, Ethnicity, System, People-oriented thinking, Civilization	Chinese history
Topic1	Calligraphy, Chess, Guqin, Painting, Carving on a board, Tea Art, Kung Fu, Opera, Dance, Handicrafts	Chinese art
Topic7	Spring Festival, Lantern Festival, Qingming Festival, Dragon Boat Festival, Qixi Festival, Ghost Festival, Mid-Autumn Festival, Double Ninth Festival, Spirit Festival, Chinese New Year's Eve	Chinese festivals
Topic6	Pasting Spring Festival couplets, Giving red envelopes, Staying up late, Eating tamales, Competing in dragon boat races, Placing mugwort, Making zongzi, Organizing lantern festivals, Cremating ancestors, Offering sacrifices	Chinese folk customs
Topic2	Brush pen, Jade pendant, inkstone, Ink, Lanterns, Xuan paper, Amber bag, Pipa, Erhu, Dish	Chinese artifacts
Topic3	Economic center Reform, Opening-up, South, Gross Domestic Product, Economic Model, Digital transformation, Supply side reform, Income gap, Five-year plan, Market economy	Chinese economy
Topic5	Cuisine style, Abundant resources, Flavor styles, Guangdong Cuisine, Chinese cuisine, Cooking techniques, Having a long history, Color, aroma, taste, Seasonal cuisine, Cooking temperature contro	Chinese cuisine

4.1.3 Results of Knowledge Graph Construction

From the aforementioned clustered themes and their feature words, it can be seen that the knowledge of various aspects of Chinese culture is basically covered in the constructed ontology with a high degree of richness. Figure 6 is a visualization of the knowledge map of Chinese culture based on the results of ontology construction and theme clustering. k1~29 are the abbreviations of the ontologies contained in the 7 categories of themes. The constructed knowledge map with appropriate distance between the 7 categories of themes better connects the ontologies of different categories as well as the subordinate feature words (mauve circles), which facilitates overseas Chinese language teachers to use recommendation algorithms to quickly search for relevant Chinese culture knowledge that students are interested in, generate recommendation lists, and disseminate Chinese culture in classroom teaching.

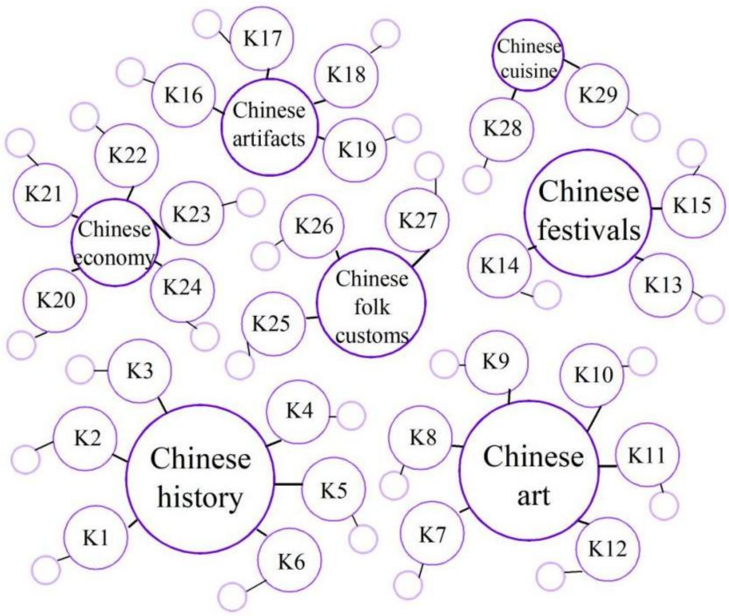

Figure 6: Chinese culture knowledge graph constructed based on ontology(Parts)

4.2 Analysis of students' long and short-term interests

4.2.1 Data pre-processing

Before completing the diversified resource recommendation based on knowledge graph and students' long and short-term interests using recommendation algorithms, the students' course resource rating data are preprocessed and analyzed, such as outliers removal, to calculate the students' long and short-term interest values and the degree of change of interest.

Table 2 shows the pre-processed student interest resource ratings. The 22 categories of Chinese cultural resources of interest to students shown in the table include important regions, ethnic characteristics, time of political change, artistic achievements, year numbers of each dynasty, guqin playing style, origin of the Spring Festival, Mid-Autumn Festival folklore, process of making dumplings, location of economic centers, influence of the Five-Year Plan, status of Cantonese cuisine, cooking style, achievements of reform and opening up, digital conversion style, pipa playing characteristics, ink raw materials, kungfu overseas dissemination effects, tea tasting process, types of paintings, Chongyang Festival customs, dumpling ingredients, food flavors, and landscape painting characteristics. Students' ratings of these Chinese culture learning resources ranged from 4.794 to 9.819 out of 10.000. The difference in resource ratings amounted to 5.025, representing that different students' interests in different Chinese culture learning resources also varied widely.

Table 2: Post-preprocessing student interest resource scores(Parts)

Scoring resources	Scoring	Scoring resources	Scoring
Important area	8.127	Cooking method	9.093
National characteristics	9.819	The achievements of reform and opening up	9.504
Time for political change	4.794	Digital conversion method	9.279
Artistic achievements	6.742	Characteristics of the Pipa Performance	9.196
Dynastic reigns of various dynasties	7.795	Ink raw materials	8.491
The playing method of the guqin	8.139	The overseas dissemination effect of Kung Fu	7.574
The Origin of the Spring Festival	7.485	Tea tasting process	8.936
The customs of the Mid-Autumn Festival	5.879	Types of Painting	6.942
The process of making dumplings	6.315	Chongyang Festival Customs	6.295
Economic center location	6.702	Ingredients for Zongzi	7.752
The impact of the five-year plan	8.827	Food flavor	9.312
The status of Cantonese cuisine	8.291	Characteristics of landscape paintings	8.908

4.2.2 Analysis of Students' Long and Short-term Interest Levels and Changes in Interest Levels

The preprocessed scores of resources of interest were normalized to analyze the interest values and degree of change of interest of students for short-term resources of interest and long-term resources of interest types. Figure 7 shows the interest values and the degree of change in interest for students' short-term resources of interest. Figure 8 shows the interest values and the degree of change in interest for students' long-term interest resources. The interest values of students' short-term interest resources range from 62.457 to 90.663, which is slightly higher than that of long-term interest resources from 61.559 to 89.404.At the same time, the degree of change of students' interest in short-term interest resources ranges from 56.389% to 82.586%, which is also higher than that of long-term interest resources from 51.149% to 57.534%. Students will be more interested in the resources in the short term, but the change of interest in these resources is faster. Overseas Chinese language teachers should grasp the timing of short-term changes in students' interests as well as the stability of long-term interests when disseminating corresponding knowledge of Chinese culture in classroom teaching.

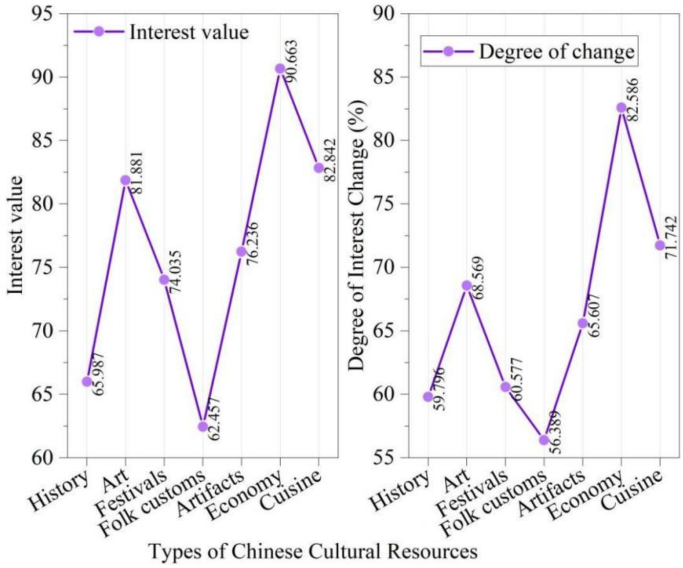

Figure 7: Short-term interest value of resources and the degree of interest change

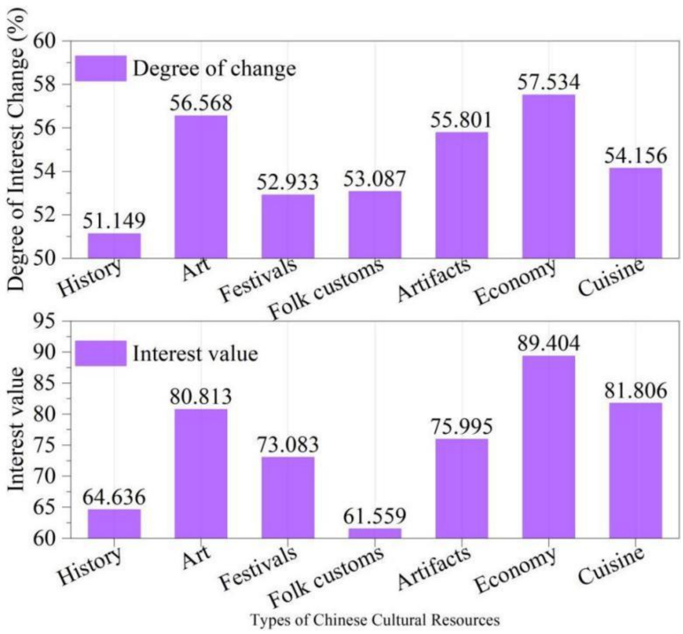

Figure 8: Long-term interest value of resources and the degree of interest change

4.3 Comparison of resource recommendation effect of different algorithms

4.3.1 Resource Recommendation Accuracy Comparison

The proposed recommendation algorithm incorporates students' short-term and long-term interest values and interest changes into the proposed recommendation algorithm to recommend diversified teaching resources for 100 students in the classes led by Chinese language teachers at Confucius Institute A overseas. At the same time, in order to verify the application effect of the proposed algorithm, the same type of recommendation algorithms: content-based recommendation algorithm (CBR), association rule-based recommendation algorithm (ARBR), utility-based recommendation algorithm (UBR), and knowledge-based recommendation algorithm (KBR) are chosen as comparison algorithms to provide resource recommendations for the 100 students' learning in Chinese culture courses as well. Figure 9 compares the resource recommendation accuracy of the five recommendation algorithms.

In order of resource recommendation accuracy size, the recommendation algorithm of this paper (94.165%) > CBR (88.374%) > UBR (75.608%) > ARBR (65.691%) > KBR (62.584%). The recommendation algorithm that integrates students' long and short-term interests and Chinese culture knowledge mapping is better able to accomplish learning resource recommendation based on students' interests and needs compared to other algorithms.

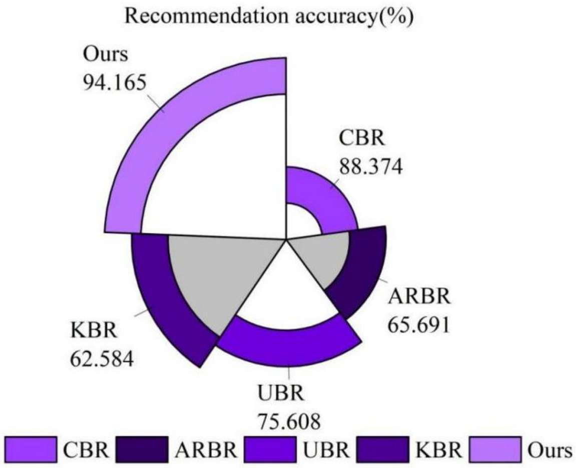

Figure 9: Resource recommendation accuracy of 5 recommended algorithms

4.3.2 Comparison of diversity of resource recommendations

Figure 10 shows the diversity of resources recommended by the five recommendation algorithms. The diversity of resources recommended by the recommendation algorithm that integrates students' long-term and short-term interests reaches 88.132%, which is 33.557% higher than that of CBR (54.575%), 7.225% higher than that of ARBR (80.907%), 31.924% higher than that of UBR (56.208%), and 22.950% higher than that of KBR (65.182%). The resources recommended by the recommendation algorithm that integrates students' long and short-term interests cover a wider variety of topics to meet students' changing interests, which can help overseas Chinese language teachers better utilize the algorithm to accomplish the dissemination of Chinese culture.

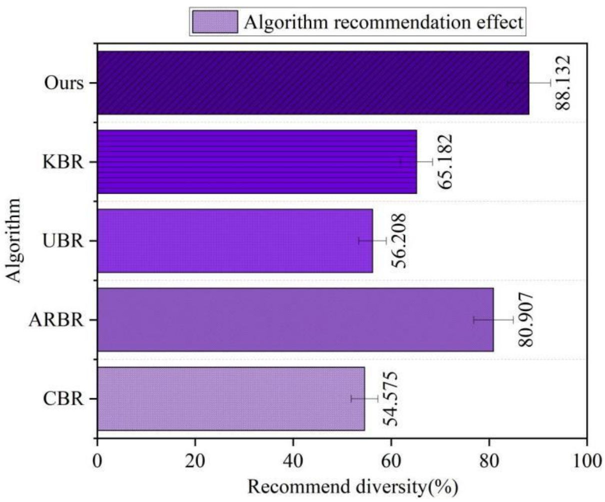

Figure 10: Diversity of resource recommendations provided by 5 algorithms

5 Conclusion

This paper integrates students' short-term and long-term interests in Chinese cultural resources into the recommendation algorithm, generates a diversified list of recommended Chinese cultural resources that meets the needs of overseas Chinese language teachers for cultural dissemination and students' interests, and accurately enhances overseas students' enthusiasm for traditional and modern Chinese culture. With the support of 94.165% recommendation accuracy and 88.132% recommendation diversity, overseas Chinese language teachers can timely display seven categories of Chinese culture-related content in the classroom, such as “Chinese history, Chinese art, Chinese festivals, Chinese folklore, Chinese artifacts, Chinese economy, and Chinese cuisine”, and build a communication window for overseas students to learn about China and love China, This will help overseas students to understand and love China.

Funding

This work was supported by 2025 Shaanxi Higher Vocational Education Teaching Reform Research Program:Research and Practice on Demand-Oriented Talent Cultivation of Medical and Nursing Majors for Central Asian Countries under the Background of "Vocational Education Going Global" (NO. 25GZ021)

About the Author

Jia Zuo was born in Xianyang, Shaanxi, P.R. China, in 1984. She obtained a master’s degree from Xi'an International Studies University, China, in 2010. She is currently working in Xianyang Polytechnic Institute as an associate professor. Her main research direction is Second Language Aquisiton and Testing, Internationalization of Higher Vocational Education. She has already published 17 essays and 3 textbooks related to English teaching, Englsih translation and internationalization of higher vocational education.
xyzy_zj515@163.com

Dong Feng was born in Xianyang, Shaanxi Province, P. R. China, in 1988. He obtained a master’s degree in Teaching and Curriculum from Michigan State University in the United States. He is currently working in Xianyang Polytechnic Institute as a lecturer. His main research direction is Teaching Chinese to Speakers of Other Languages and internationalization of vocational education.
dfeng0125@163.com

References

[1] Yalun, A. (2019). International Promotion of Chinese Language in the New Era. International Education Studies, 12(7), 67-79.

[2] Ying, W. (2016). Review of the Confucius Institutes’ strategy for the dissemination of Chinese culture. Chinese Education & Society, 49(6), 391-401.

[3] Lai, Y. (2024). ChatGPT in International Chinese Language Education: Opportunities, Potential Problems, and Solution Strategies. World Journal of Educational Studies, 48.

[4] Li, X. (2025). Development and implementation of an intelligent assisted teaching system

for Chinese English based on natural language processing and reinforcement learning. Computers and Education: Artificial Intelligence, 100466.

[5] Gao, X. (2023). Technology-Enhanced International Chinese Education: An Overview. International Journal of Chinese Language Teaching, 4(3), 1-11.

[6] Jing, Y. (2023). Design and Application of Teaching System for Chinese Language and literature major drove by interpreted AI technology. Applied Artificial Intelligence, 37(1), 2203573.

[7] Xue, M. (2022). Research on the application and influence of digital technology in international Chinese language education. Academic Journal of Humanities & Social Sciences, 5(3), 105-108.

[8] Ou, E. Y., Jiang, S., & Wang, H. (2024). A Digital Teaching Paradigm: Natural Language Processing Integrates to Teaching Chinese as The Second Language. Rajapark International Journal, 1(3), 1-14.

[9] He, H. (2024, August). Research on Design and Evaluation of Teaching Model of College Chinese Language Public Course based on Natural Language Processing Technology. In 2024 3rd International Conference on Information Economy, Data Modelling and Cloud Computing (ICIDC 2024) (pp. 64-72). Atlantis Press.

[10] Liu, Y. (2022). Open university Chinese language and literature teaching model based on NLP technology and mobile edge computing. Mobile Information Systems, 2022(1), 4149492.

[11] Li, X. (2024, August). Design and Implementation of a Chinese Language Educational Support System Based on Natural Language Processing. In 2024 7th International Conference on Education, Network and Information Technology (ICENIT) (pp. 6-10). IEEE.

[12] Ma, X., Hu, Y., & Li, M. (2024). Design of assessment algorithm and model for Chinese spoken language teaching based on natural language processing and knowledge graph. Theoretical and Natural Science, 36, 20-26.

[13] Wang, Y. (2024). Analysis of Euphemism Teaching in Teaching Chinese as a Foreign Language Under the Environment of Modern Educational Technology. International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 18(1), 1-16.

[14] Chen, J., Huang, Y., Xu, J., & He, D. (2025). Constructing a New" Teacher-AI" Collaborative Teaching Paradigm in International Chinese Language Education Enabled by Generative AI. Journal of Computing and Electronic Information Management, 18(1), 71-78.

[15] Shan, H. (2023). AI-assisted Chinese language teaching system in distributed sensor networks. Scalable Computing: Practice and Experience, 24(4), 1107-1116.

[16] Sumo, D. S., & Bah, M. L. (2021). Chinese language education in the era of artificial intelligence; innovation development, pedagogy & the smart classroom. Education Quarterly Reviews, 4(4).

[17] Jia, Z. (2022). Designing an intelligent teaching system of Chinese as a foreign language under the internet background. Scientific programming, 2022(1), 3610081.

[18] Xie, J. (2023). Opportunities and challenges brought by artificial intelligence to second language teaching: A case study of international Chinese language education. Advances in Engineering Innovation, 3, 41-46.

[19] Li, Q. (2025, March). Analysis of the Application Status and Trends of Artificial Intelligence Technology in International Diversified Higher Education—Taking Teaching Chinese as a Foreign Language as an Example. In 2024 International Conference on Social Sciences and Educational Development (ICOSSED 2024) (pp. 445-453). Atlantis Press.

[20] Kang, B., & Kang, S. (2022). Construction of Chinese language teaching system model based on deep learning under the background of artificial intelligence. Scientific programming, 2022(1), 3960023.

[21] Niu, S. J., Luo, J., Niemi, H., Li, X., & Lu, Y. (2022). Teachers' and students' views of using an AI-aided educational platform for supporting teaching and learning at Chinese schools. Education Sciences, 12(12), 858.

[22] Qi, H. (2024). New Pathways of International Chinese Education in the Digital Intelligence Era. Journal of Modern Social Sciences, 1(2), 231-241.

[23] Xia, J., Ge, Y., Shen, Z., & Rahman Najar, M. (2024). The auxiliary role of artificial intelligence applications in mitigating the linguistic, psychological, and educational challenges of teaching and learning chinese language by non-chinese students. International Review of Research in Open and Distributed Learning, 25(3), 116-133.

[24] Arif, M., Ismail, A., Aqib, M., Shaoan, M. R., Ali, W., & Okafor, M. U. (2025). The impact of AI-driven tools on foreign students' Chinese language acquisition: A case study at Southwest University. Language, Technology, and Social Media, 3(1), 150-168.

[25] Zhang, L. (2024). BENEFITS AND CHALLENGES IN IMPLEMENTING ARTIFICIAL INTELLIGENCE FROM CHINESE EFL PRIMARY SCHOOL TEACHERS'ATTITUDE AND PERCEPTION. From the editors, 48.

[26] Li, X., & Jiang, Y. (2024). Artificial intelligence in education: Opportunities, current status, and prospects. Geographical Research Bulletin, 3, 146-174.

[27] Almuhanna, M. A. (2025). Teachers' perspectives of integrating AI-powered technologies in K-12 education for creating customized learning materials and resources. Education and Information Technologies, 30(8), 10343-10371.

[28] Ingavélez-Guerra, P., Robles-Bykbaev, V. E., Perez-Muñoz, A., Hilera-González, J., & Oton-Tortosa, S. (2022). Automatic adaptation of open educational resources: an approach from a multilevel methodology based on students' preferences, educational special needs, artificial intelligence and accessibility metadata. IEEE Access, 10, 9703-9716.

[29] Oussaid, M., Lazib, S., Lazib, L., Bouarab-Dahmani, F., & Cullot, N. (2021, December).

Educational Lexical Resource Enrichment Using Machine Learning Classifiers. In 2021 22nd International Arab Conference on Information Technology (ACIT) (pp. 1-8). IEEE.