Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 18 No. 23 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijet.v18i23.45639

PAPER

A Comprehensive Model for Recommending Personalized Learning Resources for the Development of Linguistic Competence

Yingxin Ma(✉)

Department of Chinese, Cangzhou Normal University, Cangzhou, China

myx0810@caztc.edu.cn

ABSTRACT

With the continuous advancement of globalization and informatization, the linguistic competence of college students has become a key index for evaluating their comprehensive quality. Faced with diverse needs of students and educational environments, it is increasingly important and complex to accurately locate the linguistic competence goals of college students. Although existing research methods, such as standardized testing and teacher assessment, provide certain convenience, they rely on single data sources and have a certain degree of subjectivity, which limits their universality and accuracy. This study aimed to solve this problem by doing comprehensive research on two aspects: first, curriculum analysis based on relation extraction. A relation extraction model, such as Casrel, was used for advanced text analysis, which provided educators with more in-depth insights; second, personalized learning material recommendation based on text recommendation. Personalized learning paths were provided for students of different levels using the abstractness-based text recommendation algorithm. This study not only filled the gaps in existing research methods, but also provided a new, scientific and efficient solution, helping improve the quality of education and promote the formulation of scientific education policies.

KEYWORDS

linguistic competence, goal positioning, relation extraction, text recommendation, personalized learning, educational technology

1 INTRODUCTION

With the rapid development of globalization and informatization, linguistic competence has become one of the important indexes to measure the comprehensive quality of college students. However, faced with constantly changing educational environments and diverse needs of students, it is increasingly complex and urgent to accurately locate the linguistic competence goals of college students [1–6]. On the

Ma, Y. (2023). A Comprehensive Model for Recommending Personalized Learning Resources for the Development of Linguistic Competence. *International Journal of Emerging Technologies in Learning (iJET)*, 18(23), pp. 145–159. https://doi.org/10.3991/ijet.v18i23.45639

Article submitted 2023-07-13. Revision uploaded 2023-10-16. Final acceptance 2023-10-20.

one hand, traditional educational methods and curriculum structure often cannot adapt to this diverse trend. On the other hand, with the continuous maturity of big data and artificial intelligence technology, it is possible to solve this problem using more advanced analysis tools and methods.

Research on this problem helps not only educators set and evaluate teaching objectives more accurately, but also students engage in more targeted self-learning and career planning [7, 8]. By scientifically locating linguistic competence goals, educational institutions are able to allocate resources more effectively, implement teaching reforms, and improve the quality of education [9–12]. At the same time, this also provides strong data support for policy makers to promote more reasonable and efficient education policies.

Although several methods have been used to evaluate and locate the linguistic competence of college students, such as standardized testing, teacher assessment and self-assessment, these methods are often limited by single data sources and subjectivity [13–15]. In addition, most existing studies often focus on quantitative scores and levels, while neglecting multidimensional analysis of curriculum content and personalized needs [16–19]. Therefore, a more comprehensive and refined research method is needed to overcome these shortcomings and deficiencies.

The main research content of this study includes two aspects: first, curriculum analysis based on relation extraction, aiming to dig into curriculum structure and content in depth using an advanced text analysis method, such as the Casrel relation extraction model, thereby providing more insights and suggestions for educators; second, personalized learning material recommendation based on text recommendation, providing personalized learning paths and materials for students of different linguistic competence levels using the abstractness-based text recommendation algorithm. By doing comprehensive research on these two aspects, this study not only filled the shortcomings of existing research methods, but also provided a new, scientific and efficient solution for the linguistic competence goal positioning of college students, which helps promote the application of educational technology, improves the quality of education, and provides valuable research foundations for related fields.

2 CURRICULUM ANALYSIS BASED ON RELATION EXTRACTION

Traditional curriculum analysis methods are often superficial, making it difficult to grasp the internal connection between curriculum content and teaching objectives in a targeted manner. However, the relation extraction model accurately extracts the complicated relations between concepts from the text, thereby helping educators set and evaluate teaching objectives more accurately. For example, when analyzing educational materials, especially curriculum syllabuses, textbooks, and online tutorials, a concept map can be generated, which clarifies which topics and sub-topics are core and how they are interrelated. Relation extraction enables educators to clearly see which topics and sub-topics in the curriculum content are closely related and which need further emphasis or improvement, which is of great significance for optimizing the curriculum structure, especially in multidisciplinary or interdisciplinary curriculum design.

The relation extraction model in this study adopted a specific design structure, including a Chinese-xInet-base pre-training module, a head entity extraction module, a conditional layer normalization processing module, and a specific relation tail entity extraction module, which more comprehensively and accurately extracted the relations and information related to the linguistic competence goal positioning of college students from the text. The research goal of this study is college students, especially those who speak Chinese as their mother tongue. Therefore, the Chinese-xInet-base module specifically pre-trained for Chinese texts ensures that the model accurately understands and handles special details in Chinese language and culture. Accurate extraction of head entities is a crucial step in relation extraction. For example, teaching objectives, topics, or key concepts in curriculum analysis may serve as head entities. Accurate extraction of head entities helps further analyze the relations between these entities and other concepts or goals. However, educational text data may come from multiple different sources, formats, and structures. Therefore, a flexible and adaptable module is needed to handle these complexities. The conditional layer normalization processing module maintains the model's performance in different conditions, making the model more robust. With linguistic competence goal positioning as the research purpose of this study, both entities (e.g., teaching objectives or key concepts) themselves and the relations between these entities and other concepts or goals need to be understood. The specific relation tail entity extraction module enables the model to identify tail entities which have specific relations with the head entities more specifically.

The relation extraction model was constructed mainly in two key steps, aiming to accurately locate the linguistic competence goals of college students. Head entities were extracted in the first step. A pre-training XLNet model was first used to analyze the text related to education or curriculum. The model identified all possible head entities in the text, which may be teaching objectives, curriculum topics, or other key concepts related to linguistic competence. Once all possible head entities were identified, the model randomly selected one of these entities for further analysis. The random extraction increased the coverage of the model for various different relations, thereby making the analysis results more comprehensive. Specific relations and tail entities were identified in the second step. For each selected head entity, the model launched a specific relation marker. The task of this marker was to identify all possible relations between head and tail entities, which may include but are not limited to curriculum content relevant to the teaching subjective, and prior knowledge required for the curriculum topic. The model identified not only various relations related to the head entities, but also tail entities under these specific relations. For example, if the head entity is "reading comprehension", the tail entity may include "vocabulary", "grammar knowledge", etc. These two steps together form a powerful and flexible relation extraction framework, which meets the complex and multi-level research needs of linguistic competence goal positioning of college students.

The working principles of each module in the model were introduced in detail below.

The Chinese-xInet-base module especially pre-trained for Chinese environments was selected, because this study focused on college students with Chinese as their native language. After receiving a Chinese text related to education or linguistic competence as the input, the Chinese-xInet-base module encoded the input text using the pre-trained weight, which converted the natural language text into a series of high-dimensional vectors. After completing the encoding, the module began to identify various entities in the text, which may be teaching objectives, key concepts, or any other factors related to linguistic competence. In addition, the module attempted to capture the internal relations between these entities, though the subsequent specific relation marker would go through this step in more detail. Finally, the module output a series of possible head entities and related information, providing a basis for subsequent relation extraction steps.

This study aimed to locate the linguistic competence goals of college students. Therefore, head entities, which may be teaching objectives, core skills, topics, or keywords, were usually considered as the elements most directly related to these goals. The head entity extraction module received the encoded and preliminary entity identification data from the Chinese-xInet-base pre-training module. After obtaining a list of all possible entities, the module used a scoring mechanism to rank these entities. The scoring may be based on the frequency of entities that occurred in the text, contextual relevance, or other customized indexes. According to research needs, the head entity extraction module selected one or more head entities in a random or targeted way for subsequent analysis. Random selection helped increase the diversity and coverage of the model, while targeted selection was used to focus on a specific topic or goal. Finally, the module output the selected head entities and their related metadata (e.g., their positions in the text and their association with other entities, etc.) for subsequent relation extraction and analysis. Figure 1 shows the architecture of the head entity extraction module.

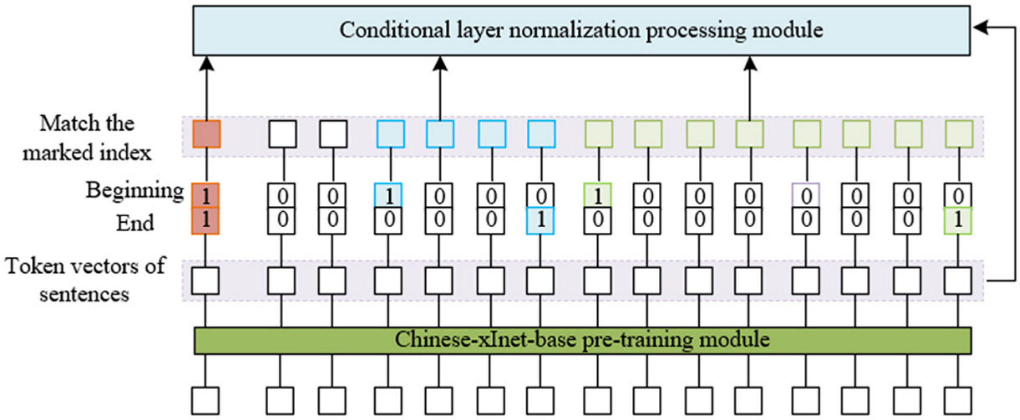

Fig. 1. Architecture of the head entity extraction module

In a multi-step model process, the scale and distribution of data may change, leading to a decrease in the model's performance. The conditional layer normalization processing module helps stabilize the data distribution, making it easier for processing and analysis. In this study, the module received the output from the head entity extraction module or the Chinese-xInet-base pre-training module, then made statistical analysis of each feature dimension of the input data, and calculated the mean value and variance. The module dynamically adjusted the normalization parameters based on conditions related to head entities or specific relations, which allowed the model to focus more on features related to the current head entities or specific relations in case of relation extraction. The input data was normalized using the calculated mean value and variance as well as conditional parameters. This step typically involved subtraction of the mean value, division by variance, and possibly scale and displacement transformations. The normalized data was passed to the next module, which was usually the specific relation tail entity extraction module, for further analysis and processing. Let $\hat{z}$ be the output vector of the normalized conditional layer, $\omega(z)$ be the mean value of the input vector, $\delta(z)$ be the standard deviation of the input vector, Q_ε be the weight of the input vector in the ε dimension, and Q_α be the weight of the input vector in the α dimension, then the execution process of this module was given by the following equations:

$$\hat{z} = (\varepsilon + CON_\varepsilon)\frac{z - \omega(z)}{\delta(z) + \gamma} + (\alpha + CON_\alpha) \tag{1}$$

$$CON_\varepsilon = Q_\varepsilon \cdot CON, CON_\alpha = Q_\alpha \cdot CON \tag{2}$$

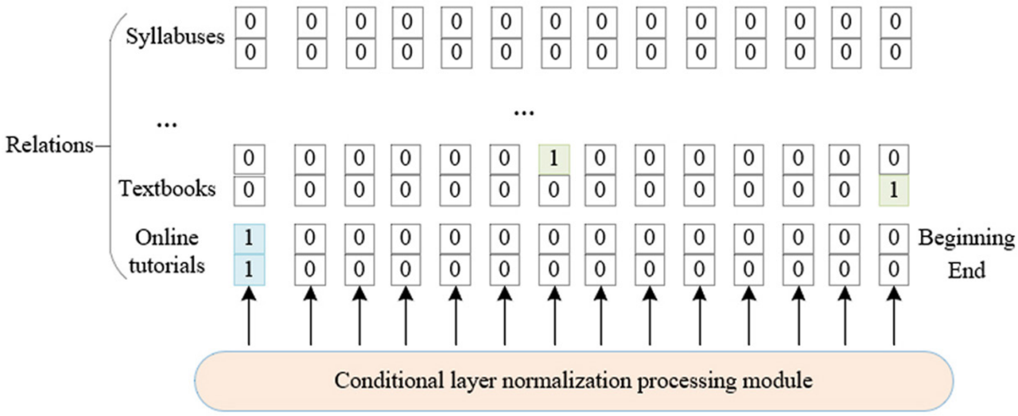

Fig. 2. Architecture of the specific relation tail entity extraction module

The specific relation tail entity extraction module was set up, aiming to specifically extract tail entities from identified head entities and preset specific relations, providing more accurate relation modeling capability, which is a key step in achieving research objectives. After receiving the normalized data from the conditional layer normalization processing module and obtaining information on head entities, the module defined in advance or dynamically identified specific relations that may be related to the head entities, such as "possessing skills", "the need to master the topic", etc. The module performed feature mapping on specific relations and normalized data, and assigned weights using the neural network or other machine learning algorithms. Based on feature mapping and weight allocation, the module began to identify and rank possible tail entities. In this step, one or more algorithms were usually used to optimize the identification accuracy of tail entities. Finally, the module output a list of tail entities related to the head entities and specific relations for subsequent data analysis and interpretation, thereby achieving the research objectives. Figure 2 shows the architecture of the specific relation tail entity extraction module. The specific process of this module was divided into two parts: beginning and end of marking. Let $O_u^{OB_{BE}}$ be the probability value of the starting position of the tail entity, i.e., the u-th token in sentence sequence $\hat{a}$ in the e-th relation, Q_{BE}^e be the transformation matrix for predicting the starting position of the tail entity in the e-th relation, and n_{BE}^e be the bias term, then the equation was given as follows:

$$O_u^{OB_{BE}} = sigmoid(Q_{BE}^e \hat{a} + n_{BE}^e) \tag{3}$$

Let M_{SU} be the loss extracted by head entities, M_{OB} be the loss extracted by tail entities in specific relations, M be the overall loss of the model, i.e., the sum of M_{SU} and M_{OB}, and b be the length of the input sentence, then the specific equations were given as follows:

$$M_{SU} = -\frac{1}{b} \sum_{j \in \{SU_a BE, SU_r bf\}} \sum_{u=1}^{b} t_u \log(o_u^j) + (1 - t_u) \log(1 - o_u^j) \tag{4}$$

$$M_{OB} = -\frac{1}{b} \sum_{j \in \{OB_a BE, OB_r bf\}} \sum_{u=1}^{b} t_u \log(o_u^j) + (1 - t_u) \log(1 - o_u^j) \tag{5}$$

$$M_{TO} = M_{SU} + M_{OB} \tag{6}$$

The constructed relation extraction model was used for curriculum analysis in the following operation process:

Step 1: Data preprocessing and standardization. Before any analysis was made, text data related to curricula (e.g., curriculum descriptions, objectives, syllabuses, etc.) was preprocessed and standardized for subsequent analysis.

Step 2: Model processing. Chinese-xInet-base was used to extract features from the preprocessed text data. After receiving the output from the Chinese-xnet-base pre-training module, the head entity extraction module identified possible head entities, such as specific learning topics or skills. Then the data obtained from the previous module was normalized at the conditional layer. Furthermore, for each identified head entity and predetermined specific relations (e.g., "knowledge points to be mastered", "competence involved", etc.), the tail entities were identified and ranked.

Step 3: Result integration and interpretation. Based on the identification results of head and tail entities, a graph of relations between curricula and various teaching objectives and learning outcomes was constructed.

Step 4: Visualization and report generation. The relation graph and other analysis results were visualized, and corresponding reports were generated, thereby enabling educators and researchers to more intuitively understand the relations between curricula and linguistic competence of college students.

3 PERSONALIZED LEARNING MATERIAL RECOMMENDATION BASED ON TEXT RECOMMENDATION

The personalized learning resource recommendation model recommends corresponding learning materials based on the learning needs and ability level of each student, thereby effectively improving teaching quality and the learning efficiency of students. Moreover, by recommending learning resources that are more in line with students' interests and needs, their learning participation and satisfaction can be enhanced, thereby achieving the goal positioning of their linguistic competence more rapidly.

To accurately analyze and recommend personalized learning materials, a text concept set concerning learning resources should be first constructed, including but not limited to keywords, phrases, topics, etc. That is, the set should be constructed from a large number of teaching materials, papers, online curricula and so on using natural language processing techniques, such as topic modeling and keyword extraction.

Furthermore, the information quantity corresponding to the text of the target learning resources should be calculated. In this study, the information quantity of target text was calculated mainly in two key links: text scanning mechanism and calculation of conceptual information quantity. Text scanning first divided the text into small information units or concepts using semantic segmentation, then constructed the semantic link network by simulating the recall and associative processes of the human brain, and dynamically assigned weights during this process to simulate the forgetting mechanism of the human brain. In the process of calculating the conceptual information quantity, frequency analysis was first made to calculate the frequency that each concept occurred in the text. Then the relative importance of these concepts in specific texts and overall corpora was evaluated through local-global alignment, such as the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm. Finally, the information entropy formula was used for quantification. The combination of these two links not only helped achieve the linguistic competence goal positioning of college students more accurately, but also improved the accuracy and effectiveness of personalized learning resource recommendation.

As a key link in calculating the information quantity in the target text, the text scanning mechanism pays special attention to simulating the recall, association, forgetting and other steps of the human brain during the reading process. These simulation processes help more accurately locate the linguistic competence goals of college students and recommend personalized learning resources. When scanning the learning resource text, the text scanning mechanism made calculations in the following steps:

Step 1: Recall process, which simulated the recall mechanism of the human brain after exposure to information. The pre-constructed knowledge graph or semantic network was used to mark the keywords or concepts that occurred in the text and link them to relevant concepts.

Step 2: Associative process, which simulated how the human brain automatically associated with other relevant information when it came into contact with a certain information element. Specifically, natural language processing techniques were used to determine the similarity or correlation between different concepts in the text, thereby obtaining a conceptual impression. Let k be the adjacent concept of u, $OQ(u \rightarrow k)$ be the impression value propagated from concept u to concept k, and μ be the retention rate, then the conceptual impression was quantified using the following equation:

$$OQ(u \rightarrow k) = \begin{cases} \left(1 - \dfrac{1}{\mu}\right) * Q[u] * \dfrac{ME_{j,F}(u,k)}{\sum_k ME_{j,F}(u,k)} & IF\ (Q[u] > MIN) \\ 0 \quad IF(Q[u] \leq MIN) \end{cases} \tag{7}$$

Step 3: Forgetting process, which simulated the human brain in forgetting non-core or unrelated information. Specifically, inactive or low-weight nodes (concepts) in the semantic link network were removed or their weights were reduced, which manifested as the conceptual impression value changing with the forgetting time as well as the number of repetitions and associated concepts. Let DY be the forgetting time, V_{FO} be the coefficient reflecting the forgetting process speed, η be the lowest value of V_{FO}, $BM_j(u)$ be the number of adjacent concepts of u after scanning the j-th sentence, and YZ_β be the minimum value of β, then there were:

$$HQU_j(u) = V_{FO} * HQU_{j-1}(u) = \left(1 - \frac{1}{\eta * (1 + (DY - \beta)^2)}\right) * hqu_{j-1}(u) \qquad (DY \geq 1) \tag{8}$$

$$\beta = MAX\left(MIN_\beta, \log_2 \frac{\prod_k (1 + HE_{j-1}(u,k))}{MK_j(u) + \sum_j HE_{j-1}(u,k)}\right) \tag{9}$$

Step 4: Joint step, which combined the results of recall, association and forgetting to make preparations for calculating the conceptual information quantity and recommending personalized resources in the next step. Specifically, the results of recall, associative and forgetting processes were used to encode the text again or extract features, thereby generating a comprehensive and dynamic representation of text information. Let $EO_{j-1}[]$ be the global conceptual impression value of recall during the recall process, $MQU_j[]$ be the change in conceptual impression value during the associative process, and $QD_j[]$ be the change in conceptual impression value during the forgetting process, then the corresponding global conceptual impression value was calculated using the following equation:

$$HQU_j[\ ] = \begin{cases} EO_{j-1}[\ ] + MQU_j[\ ] + QD_j[\ ] & (j > 1) \\ MQU_j[\ ] & (j = 1) \end{cases} \tag{10}$$

The equation for HQU in the learning resource text was given as follows:

$$UL_j(u) = \sum_{k=1}^{j} HQU_k(u) \tag{11}$$

Let W be the related concept set which is no more than three layers from concept v in distance, $O(v,o)$ be the value of information quantity brought by concept o to concept v, and $O(v,o)$ be the correlation calculation between the conceptual impression value on the shortest path between concepts v and o and the concept, then the equation for the conceptual information quantity was given as follows:

$$GNXX(v) = \sum\nolimits_{o \in W} O(v,o) \tag{12}$$

Let $LF(s)$, $LF(n)$ and $LF(v)$ be the impression values of concepts s, n, and v; $SF(v, n)$ and $SF(n, s)$ be the correlation between concepts v and n, and that between n and s. The equation for $O(v, s)$ was given as follows:

$$O(v,s) = LF(v) * SF(v,n) * LF(n) * SF(n,s) * LF(s) \tag{13}$$

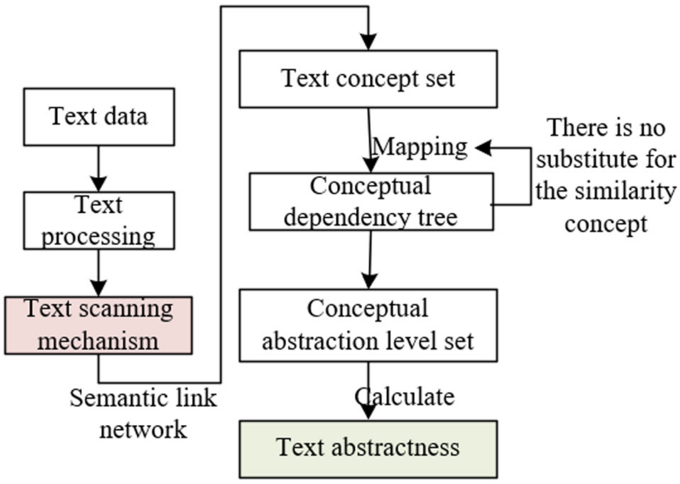

Fig. 3. Algorithm flowchart

This study calculated the abstractness of learning resource texts based on the high information quantity concept set of the text. Figure 3 shows the algorithm flowchart. The specific steps were described as follows:

Step 1: Text data preprocessing, including text cleaning, word segmentation, part of speech tagging, etc.

Step 2: Text scanning processing. The semantic link network was generated dynamically by simulating the data recall, associative, and forgetting processes of the human brain.

Step 3: Constructing the text semantic link network. Natural language processing techniques and algorithms were used to conduct in-depth text analysis and construct the semantic link network, thereby capturing the relations between various concepts in the text.

Step 4: Calculating the information quantity of the concept set. Information entropy or other information theory methods were used to calculate the information quantity of each concept.

Step 5: Mapping of the concept set. Statistical methods or machine learning algorithms were used to standardize or normalize the concept set.

Step 6: Abstraction level analysis of the concept set. The text semantic link network was used to analyze the relation complexity between concepts, thereby determining the abstraction level.

Step 7: Abstractness calculation. Based on the information quantity and abstraction level of the concept set, the text abstractness $LE(a)$ was calculated using a mathematical model. Let $UD(j)$ be the abstraction level of j in the conceptual dependency tree, and B be the number of concepts with high information quantity extracted from the text, then the equation was given as follows:

$$LE(a) = \frac{\sum_{j \in N} UD(j)}{B} \tag{14}$$

Finally, this study recommended personalized learning resources using the abstractness-based text recommendation algorithm. The specific operation process was described as follows:

Step 1: Linguistic competence assessment of students. The linguistic competence of students was assessed through online testing, historical data analysis, or teacher assessment.

Step 2: Construction of the text resource database. Learning resources of different types and difficulty levels were collected and classified, and the abstractness of each type of resources was calculated.

Step 3: Personalized need analysis of students. The interests and learning goals of students were known through questionnaire surveys, online behavior analysis, or interaction with teachers.

Step 4: Operation of the abstractness-based recommendation algorithm. The text resource database was ranked using the previously calculated text abstractness. According to the linguistic competence and personalized needs of students, the most suitable learning resources were selected using the algorithm. The deep learning method was further used for recommendation optimization.

Step 5: Feedback and adjustment of recommendation results. The effect of recommendation results was evaluated through the online interactive behavior and feedback information of students. The algorithm was adjusted as needed.

4 EXPERIMENTAL RESULTS AND ANALYSIS

The data in Table 1 shows the performance of different pre-training models in relation extraction tasks of different types of learning resources. In terms of learning resources of basic teaching materials and teaching auxiliary materials, the Chinese-xInet-base pre-training module has the best performance in precision (64.18%), recall (62.83%), and F1 (63.92%). The ELECTRA (the abbreviation of Efficiently Learning an Encoder that Classifies Token Replacements Accurately) Chinese pre-training model has the second best performance, but there is a significant gap compared with Chinese-xlnet-base. Bidirectional Encoder Representations from Transformers (BERT) and Enhanced Representation through Knowledge Integration (ERNIE) have relatively low performance. In terms of learning resources of practical application texts, the Chinese-xInet-base pre-training module also has the best performance in precision (67.39%), recall (62.39%), and F1 (64.58%). The ERNIE Chinese pre-training model performs relatively well, but its performance is still lower than that of

Chinese-xInet-base. ELECTRA has slightly poor performance in recall, but performs better in precision.

Table 1. Performance results of different pre-training models

Models	Basic Teaching Materials and Teaching Auxiliary Materials			Practical Application Texts		
	Precision	Recall	F1	Precision	Recall	F1
BERT Chinese pre-training model	61.25%	58.16%	61.47%	61.28%	62.33%	62.55%
ERNIE Chinese pre-training model	61.86%	61.28%	61.33%	62.33%	62.18%	62.36%
ELECTRA Chinese pre-training model	62.39%	61.48%	62.83%	64.36%	61.26%	62.14%
Chinese-xInet-base pre-training module	64.18%	62.83%	63.92%	67.39%	62.39%	64.58%

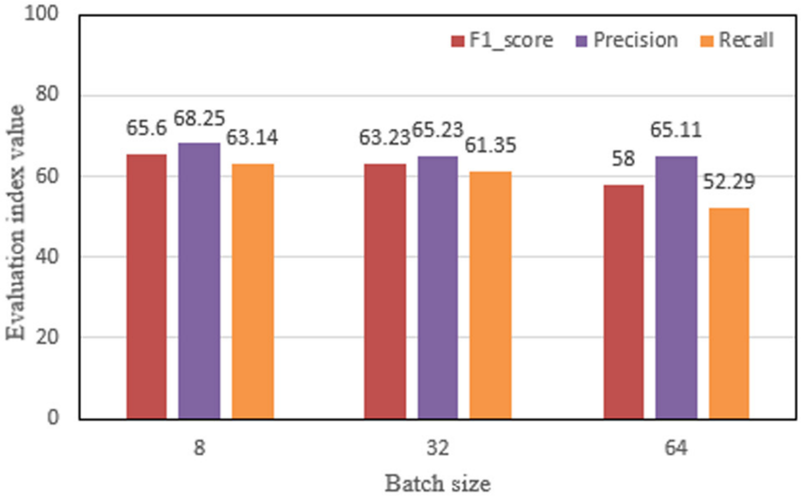

Fig. 4. Performance comparison diagram with different batch sizes

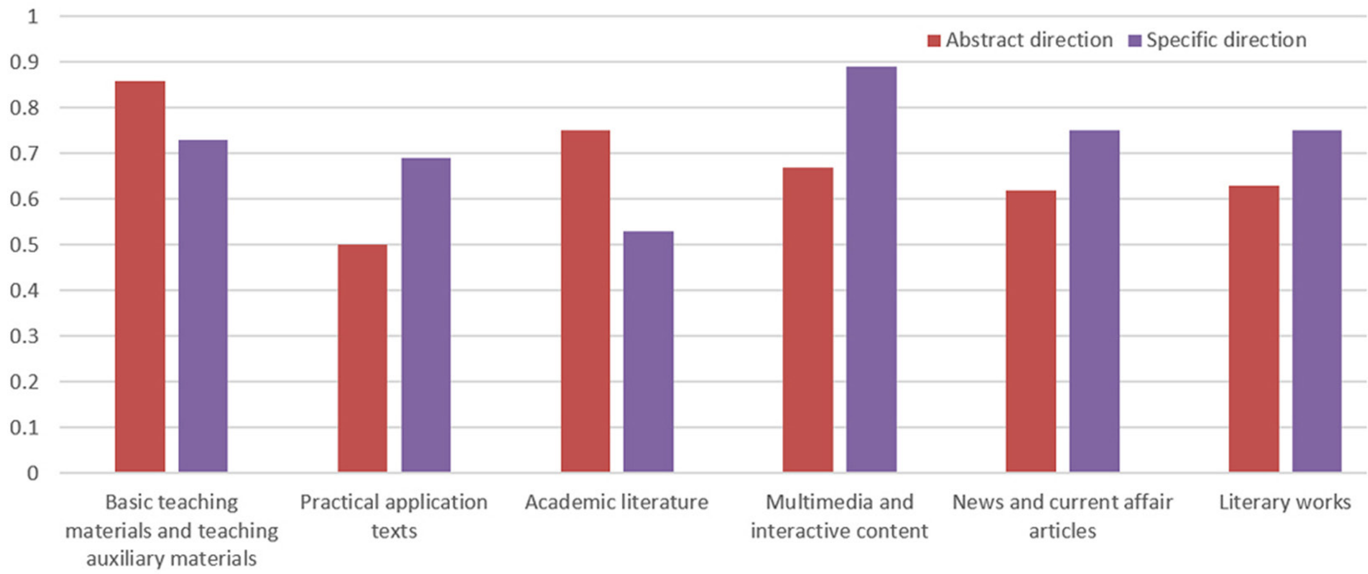

Fig. 5. Accuracy experiment histogram of the recommendation algorithm

Table 2. Accuracy of different learning resource recommendation methods

Resource Types	Methods	Text	Abstract
Basic teaching materials and teaching auxiliary materials	Collaborative filtering recommendation algorithm	0.72	0.85
Basic teaching materials and teaching auxiliary materials	Content-based recommendation algorithm	0.51	0.56
Basic teaching materials and teaching auxiliary materials	Matrix decomposition recommendation algorithm	0.53	0.64
Basic teaching materials and teaching auxiliary materials	The algorithm proposed in this study	0.8	0.87
Practical application texts	Collaborative filtering recommendation algorithm	0.68	0.51
Practical application texts	Content-based recommendation algorithm	0.66	0.5
Practical application texts	Matrix decomposition recommendation algorithm	0.5	0.65
Practical application texts	The algorithm proposed in this study	0.77	0.84

It can be seen from the data in Figure 4 that the performance of the relation extraction model varies along with different batch sizes. When the batch size is smaller, such as 8, the model has relatively good performance in the three indexes of F1-score, precision, and recall. As the batch size increases, the overall performance of the model decreases. Especially in terms of recall, the model's performance decreases significantly when the batch size is 64. The selection of batch size has a significant impact on the performance of the relation extraction model. In this research scenario, the smaller batch size (e.g., 8) is more suitable for relation extraction tasks.

According to the data in Figure 5, the accuracy of the recommendation algorithm varies for various types of learning resources in both abstract and specific directions. It can be seen from the figure that basic teaching materials and teaching auxiliary materials have the highest recommendation accuracy in the abstract direction, which is related to the universality and foundational nature of their content. Multimedia and interactive content have the highest recommendation accuracy in the specific direction, because these types of resources typically contain more specific information with strong operability. There are significant differences in the accuracy of practical application texts and academic literature in both abstract and specific directions, which means that these resource types are more suitable for recommendations in a specific direction. The accuracy of news and current affair articles as well as literary works is relatively average in both directions, because these types of resources typically cover multiple topics and styles, which leads to good application potential in both abstract and specific directions.

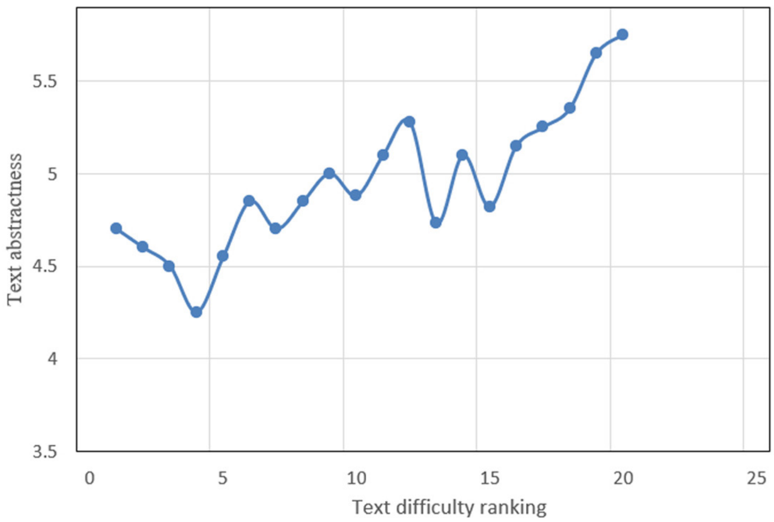

Fig. 6. *(Continued)*

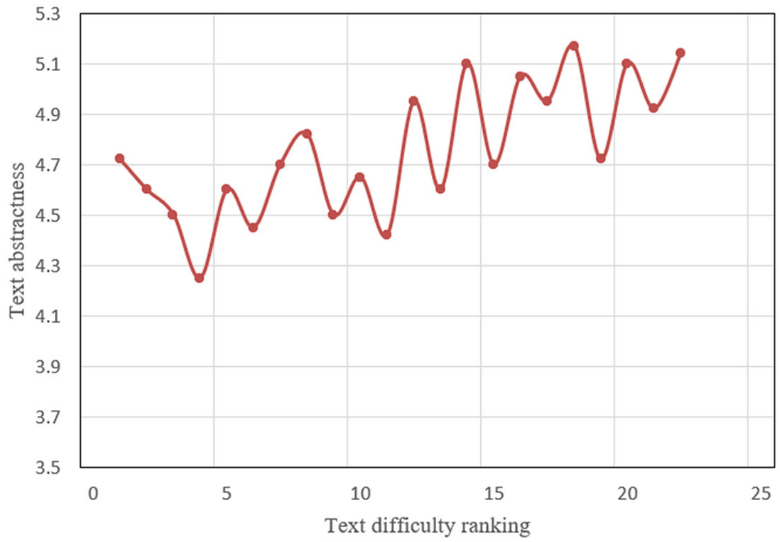

Fig. 6. Relation diagram between text abstractness and difficulty of different types of resources

According to the data in Table 2, the accuracy of different types of learning resources varies when using different recommendation methods, especially in terms of text and abstract. For basic teaching materials and teaching auxiliary materials, the algorithm proposed in this study has the best performance, which is followed by the collaborative filtering recommendation algorithm. The content-based and matrix decomposition methods have relatively weak performance. For practical application texts, the proposed algorithm also has the best performance, while the collaborative filtering and content-based recommendation algorithms have some shortcomings in certain aspects.

According to the text difficulty ranking data in Figure 6, basic teaching materials and teaching auxiliary materials and practical application texts have different scores at different difficulty levels. It can be seen from the figure that the difficulty scores of basic teaching materials and teaching auxiliary materials fluctuate less as the difficulty increases, but significantly increase when the difficulty level is high (e.g., 15–20). The difficulty scores of practical application texts increase significantly with the increase in difficulty level, especially at the highest difficulty level (e.g., 15–20). These data indicate that practical application texts are more complex and challenging at high difficulty levels, while basic teaching materials and teaching auxiliary materials are relatively more stable, which is valuable for recommending texts of different difficulty levels based on the competence and needs of learners.

5 CONCLUSION

This study focused on recommending personalized learning resources based on the text recommendation model, and using different pre-training models for relation extraction. Specifically, this study covered multiple aspects, including constructing the text concept set, calculating the information quantity of target texts, calculating the text abstractness based on the high information quantity concept set of texts, and recommending personalized learning resources based on abstractness. The experimental results showed that the Chinese-xInet-base pre-training module performed better in the relation extraction model for different types of learning resources, compared with other pre-training models, such as BERT, ERNIE and ELECTRA. It was found from the experiment that the model obtained the highest F1-score when the batch size was smaller (i.e., 8). However, the model's performance decreased as the batch size increased. For different types of learning resources, such as basic teaching materials, practical application texts, academic literature, etc., the algorithm proposed in this study generally achieved high recommendation accuracy. Practical application texts were more complex and challenging at high difficulty levels, while basic teaching materials and teaching auxiliary materials were relatively more stable.

This study successfully proposed and verified a comprehensive model for recommending personalized learning resources based on multiple dimensions, including the text concept set, information quantity, abstractness, etc. The experimental results showed that the model not only performed well in different pre-training models, but also made effective personalized recommendations based on different types and difficulty levels of learning resources. Therefore, this study has important application value in the fields of educational technology and personalized learning.

6 REFERENCES

[1] I. Stevanus and L. Parida, "The impact of gadget usage on the social and linguistic development of primary school students," *International Journal of Online and Biomedical Engineering*, vol. 19, no. 11, pp. 159–172, 2023. https://doi.org/10.3991/ijoe.v19i11.40903

[2] S. Qin, L. Orchakova, Z.-Y. Liu, Y. Smirnova, and E. Tokareva, "Using the learning management system 'modular object-oriented dynamic learning environment' in multilingual education," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 3, pp. 173–191, 2022. https://doi.org/10.3991/ijet.v17i03.25851

[3] I. Belousova and S. Bairamova, "Linguistic and cognitive analysis of inter-cultural business discourse from the perspective of 'dialogue of cultures' in the formation of inter-cultural communicative competence," *International Journal of Emerging Technologies in Learning*, vol. 15, no. 9, pp. 173–189, 2020. https://doi.org/10.3991/ijet.v15i09.12539

[4] O. M. A. Aldalalah, "The effectiveness of the different patterns of the digital story in the development linguistic intelligence skills and mindfulness among jordanian primary students during the corona pandemic," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 19, pp. 4–20, 2021. https://doi.org/10.3991/ijim.v15i19.22575

[5] O. Novikova, Y. Kalugina, A. Mustafina, and I. Edeleva, "Linguistic analysis of Insta, Twit Posts and LJ Blogs in the context of their functions (Based on the Russian Language)," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 5, pp. 66–86, 2021. https://doi.org/10.3991/ijim.v15i05.20013

[6] E. Vdovina and L. Khalyapina, "Issues of training multilingual professionals in higher education," in *International Conference on Professional Culture of the Specialist of the Future*, Petersburg, Russia, 2022, pp. 346–358. https://doi.org/10.1007/978-3-031-26783-3_29

[7] I. R. Banday, M. Zaman, S. M. K. Quadri, S. A. Fayaz, and M. A. Butt, "Big data in academia: A proposed framework for improving students' performance," *Revue d'Intelligence Artificielle*, vol. 36, no. 4, pp. 589–595, 2022. https://doi.org/10.18280/ria.360411

[8] O. Zhukova, V. Mandragelia, T. Alieksieienko, A. Semenenko, and E. Skibina, "Digital technologies for introducing gamification into the education system in the context of the development of industry 4.0," *Ingénierie des Systèmes d'Information*, vol. 28, no. 1, pp. 141–147, 2023. https://doi.org/10.18280/isi.280114

[9] N. S. B. Johan, K. Si Na, Z. Tasir, T. W. Koh, T. W. Liew, and S. S. Lee, "The implementation of digital gamification in English classroom: teachers' and students' perspectives," in *Proceedings of the 2022 5th International Conference on Education Technology Management*, Lincoln, United Kingdom, 2022, pp. 169–175. https://doi.org/10.1145/3582580.3582607
[10] M. C. Campoy-Cubillo and M. Esbrí-Blasco, "Pedagogical potential of online dictionaries in metaphor and idiom language instruction," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 21, pp. 214–229, 2022. https://doi.org/10.3991/ijet.v17i21.32445
[11] M. H. Chen, W. F. Chen, and L. W. Ku, "Application of sentiment analysis to language learning," *IEEE Access*, vol. 6, pp. 24433–24442, 2018. https://doi.org/10.1109/ACCESS.2018.2832137
[12] N. Kanwal, S. Amin Qadir, and K. Shaukat, "Writing instructions at a university and identity issues: A systemic functional linguistics perspective," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 6, pp. 275–285, 2021. https://doi.org/10.3991/ijet.v16i06.19485
[13] X. Guo, "Research on the system design of cooperative foreign language teaching mode under data analysis," In *Application of Big Data, Blockchain, and Internet of Things for Education Informatization: First EAI International Conference*, BigIoT-EDU 2021, Virtual Event, Proceedings, Part I 1. Springer Cham, 2021, pp. 191–201. https://doi.org/10.1007/978-3-030-87900-6
[14] H. Du, "Research on the application of blackboard writing images in college English writing teaching aided by 'POA' theory," *Mathematical Problems in Engineering*, vol. 2022, 2022. https://doi.org/10.1155/2022/6829622
[15] W. Ling, "Improving EFL college students' metacognitive writing ability through multimedia software," in *2021 International Conference on Internet, Education and Information Technology (IEIT)*, Suzhou, China, pp. 417–422, 2021. https://doi.org/10.1109/IEIT53597.2021.00099
[16] H. Peng, "On the cultivation of pragmatic competence for college students in the information age," in *Journal of Physics: Conference Series*, vol. 1176, no. 4, p. 042020, 2019. https://doi.org/10.1088/1742-6596/1176/4/042020
[17] L. Jiang and F. S. Alotaibi, "Investigation of English grammar application ability of college students using dynamic nonlinear system theory," *Fractals*, vol. 30, no. 2, p. 2240082, 2022. https://doi.org/10.1142/S0218348X22400825
[18] D. Zhang, "Investigation and analysis of translation ability research on the psychological obstacles of college students in English learning under the background of wireless network communication," *Wireless Communications and Mobile Computing*, vol. 2021, 2021. https://doi.org/10.1155/2021/1836014
[19] Q. He, "Computer-assisted language learning and the improvement on reading ability," in *2015 International Conference on Logistics, Informatics and Service Sciences (LISS)*, Barcelona, Spain, pp. 1–4, 2015. https://doi.org/10.1109/LISS.2015.7369816

7 AUTHOR

Yingxin Ma is a Hui nationality female, was born in Cangzhou City, Hebei Province in August 1983. She is an Associate Professor in the Department of Chinese at Cangzhou Normal University and holds a doctorate from Hebei Normal University. Her research focuses on Chinese and applied linguistics. She has led one ministerial, seven provincial, and four municipal projects, published four academic books, and over twenty papers. Her work has been recognized with awards like the "Excellent

Doctoral Dissertation of Hebei Province", "Social Science Excellent Achievement Award of Hebei Province", and "Teaching Achievement Award of Hebei Province". Additionally, she has been named "Top Young Talent in Hebei Province", a third-level candidate of the "333 Talent Project" in Hebei Province, and a "Top Professional and Technical Talent in Cangzhou City" (E-mail: myx0810@caztc.edu.cn; ORCID: https://orcid.org/0009-0008-7868-7398).