ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Intelligent Analysis of College English Reading Comprehension Strategies Aided by Knowledge Graph

Zhen Wang[1,*]

[1] School of Foreign Language Studies, Nanyang Institute of Technology, Nanyang, Henan,473000, China

SUMMARY: *With the continuous advancement of digital teaching, English reading teaching in colleges and universities has put forward higher requirements for reading process identification, strategy analysis and accurate feedback. Aiming at the intelligent analysis task of college English reading comprehension strategies, this paper constructs a knowledge graph assisted analysis model, establishes a heterogeneous semantic network around reading text, topic type, behavior log and strategy label, and fuses text semantic features and learning behavior features. Gated fusion, graph reasoning and deep learning coding are used to realize strategy recognition such as skiming, scanning, inference and generalization. At the same time, the feedback adaptation, portrait update and joint optimization mechanisms are designed to form a closed loop of "recognition-feedback-correction-re-identification". The experiment is carried out based on 168 reading texts, 840 questions, 4120 strategy samples and 28640 behavior fragments. The results show that the accuracy of the model is 91.9%, the F1 value is 91.0%, the F1 value of the logical inference task is 90.2%, and the average response delay is 0.67 s. The results show that the knowledge graph driven multi-source fusion and graph reasoning method can effectively improve the accuracy, adaptability and teaching support value of college English reading comprehension strategy analysis.*

KEYWORDS: *Knowledge graph; College English reading; Reading comprehension strategy; Intelligent analysis*

1 Introduction

English reading teaching in colleges and universities has long undertaken multiple tasks such as vocabulary accumulation, discourse comprehension, logical inference and cross-cultural cognitive training, which is an important link in the formation of students 'comprehensive English ability. With the continuous improvement of digital teaching environment, reading teaching has gradually shifted from simply focusing on reading results to continuous analysis of reading process, comprehension path and strategy usage [1]. In practical teaching, students often show the characteristics of obvious differences in strategy use, unclear problem positioning, and fragmentation of reading behavior in reading comprehension. Some students rely on words to translate sentence by sentence, and it is difficult to grasp the structure of the text. Some of them can quickly obtain surface information, but their performance is insufficient in infering the author's intention, identifying implicit logic and integrating key information [2]. Although teachers can give guidance by virtue of experience, in the face of large-scale, multi-level and dynamically changing reading data, it is difficult to form a stable,

*hsgzwangzhen@163.com
https://doi.org/10.65102/is2026276
Published: 30 April, 2026

fine and traceable strategy analysis mechanism only relying on manual observation [3].

The development of artificial intelligence, educational data mining and natural language processing technology provides new methodological support for college English reading comprehension research [4]. In particular, due to the structural expression ability of entities, relations and semantic networks, knowledge graphs can organize the associations between vocabulary, grammar, topic concepts, discourse logic and reading strategies, providing a stronger semantic interpretation basis for modeling the reading comprehension process [5]. Compared with traditional analysis methods based on score statistics or questionnaire induction, knowledge graph can not only present the correlation structure between reading knowledge points, but also combine learning behavior data, text semantic features and strategy performance results to mine students' cognitive paths and weak links in reading, so as to improve the accuracy of reading strategy recognition and the pertinence of feedback [6, 7].

At present, related research has made certain progress in intelligent reading recommendation, learning portrait construction, text semantic analysis and other aspects, but there is still a lack of a more complete technical path for "how to form a collaborative support between knowledge graph and reading behavior analysis, strategy recognition model and teaching feedback mechanism" [8]. Some studies focus on text difficulty analysis and lack of joint modeling of learners 'behavior and strategy states. Although some studies introduce deep learning methods, they have shortcomings in the interpretability of results and the transferability of teaching [9, 10]. Based on this, this paper focuses on the intelligent analysis task of college English reading comprehension strategies, constructs a knowledge graph assisted analysis model, integrates the semantic features of reading text and learning behavior features, introduces graph reasoning and deep learning methods to realize reading strategy recognition, and further designs an intelligent analysis system for teaching applications and a feedback optimization mechanism. The purpose of this paper is to provide a research path with technical support, process explanation and practical application value for English reading teaching in colleges and universities.

2 Related work

Focusing on the formation mechanism and strategic influencing factors of English reading comprehension ability in colleges and universities, the existing research mainly focuses on three aspects: language foundation, cognitive processing and teaching intervention. Vettori et al. conducted a systematic review of English as a foreign language learners' exposative reading comprehension and pointed out that vocabulary, language ability, cognitive processing and higher-order thinking skills jointly affect reading comprehension performance, providing a comprehensive analysis framework for reading strategy research [11]. Nash et al. introduced dynamic vocabulary learning assessment into reading research, emphasizing that vocabulary development process has predictive value for reading comprehension level and weak reader identification [12]. Gunnerud et al. investigated the skill level and prediction model of bilingual children's reading comprehension, indicating that the language acquisition background would affect the structure of reading comprehension ability [13]. Babayiğit et al. further pointed out that although bilingual learners have advantages in some cognitive dimensions, vocabulary limitation will still weaken their reading comprehension effect [14]. Calafato et al. discussed the differences between EFL learners' language comprehension and text perception from the perspective of literary ability portrait, expanding the ability dimension of reading comprehension research [15].

In terms of teaching support and digital intervention, Shao et al. embedded metacognitive reading strategies into gamified interactive e-books and verified the promotion effect of the combination of digital resources and strategy training on reading comprehension and learning motivation [16]. Seifert pointed out through textbook analysis that reading comprehension ability was not fully presented in teaching resources, suggesting that there was still a structural gap in classroom reading training [17]. Kremmel et al. conducted replication research on the relationship between unknown word density and reading comprehension, further highlighting the fundamental role of vocabulary coverage on comprehension quality [18]. Orhan compared the effects of online and classroom problem-based learning in college EFL classrooms, and showed that teaching organization would affect learning outcomes and critical thinking development [19]. Michos et al. proposed that teachers' data literacy is an important predictor of the use of learning analytics data, which provided practical enlightenment for how the results of intelligent analysis of reading comprehension really enter into teaching decision-making [20].

In terms of intelligent analysis and computational model, Li et al. proposed a question-aware memory network for multi-hop question answering, indicating that complex semantic tasks can achieve more effective information integration through memory modeling [21]. Liao et al. discussed the problem of multi-hop reasoning on knowledge graphs, which provided a method basis for cross-sentence information association and implicit relationship mining in reading comprehension [22]. Feng et al. used contrastive learning to improve the robustness of machine reading comprehension, indicating that the deep model can still improve the stability by optimizing the training mechanism under noise interference and semantic perturbation [23]. Huang et al. aligned explicit knowledge and tacit knowledge into the process of human-like reading comprehension, and strengthened the reasoning depth and knowledge utilization effect in multi-hop question answering [24]. Hu et al. modeled session flow information based on graph neural networks and demonstrated the advantages of graph structure methods in capturing context dependencies in machine understanding tasks [25].

In general, the existing research fully reveals the influence of vocabulary, cognition, teaching methods and digital resources on reading comprehension, and also provides technical reference for multi-hop reasoning, graph modeling and robust optimization. However, the existing results mainly focus on reading performance interpretation, single strategy training or general machine reading comprehension tasks, and there are still few studies on the integrated modeling of "knowledge graph-reading behavior-strategy recognition-teaching feedback" in college English reading scenarios. Based on this, this paper attempts to combine knowledge graph, graph reasoning and deep learning methods to build an intelligent analysis model for college English reading comprehension strategies.

3 Construction of an intelligent analysis model for College English reading comprehension strategies assisted by knowledge graph

3.1 College English reading knowledge Graph Construction and semantic association representation

College English reading is not a linear process of vocabulary, syntax, topic and strategy separated from each other, but a dynamic cognitive activity of language knowledge, discourse structure, task requirements and learner behavior. If the traditional tabular statistics or single label labeling method is still used, it can only record the results of students' "correct answers or wrong answers", and it is difficult to reveal their real strategy paths in topic grasp, detail location, logical inference and discourse integration. In order to improve the structural degree and interpretability of reading comprehension strategy analysis, this paper integrates text knowledge, topic requirements, behavior trajectories and strategy labels in college English reading into the framework of knowledge graph, and constructs a reading semantic network for teaching applications. Figure 1 shows the construction process of college English reading knowledge graph.

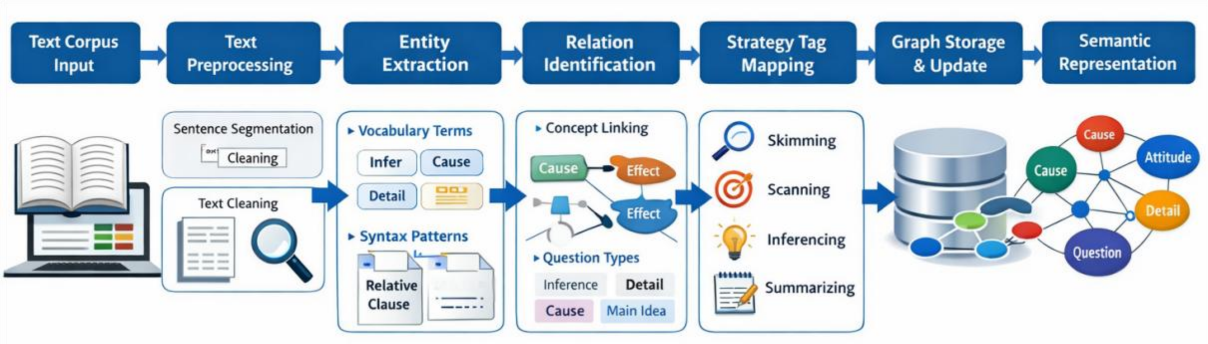

Figure 1: Flowchart of constructing the College English reading knowledge graph

In the construction of the graph, the data sources mainly include reading texts, topic stem options, teacher labeling results, and learning platform logs. The reading text provides lexical knowledge, syntactic structure, discourse theme and logical connection information, the question stem options reflect the test direction of the topic, teacher labeling can provide supervision signals for strategy identification, and the platform log records behavioral characteristics such as stay time, look back frequency, skip reading sequence and answer delay. Through the unified coding of multi-source data, the originally scattered information can be organized into an interrelated graph structure of "knowledge point-topic type-strategy behavior". Table 1 shows the core entities and relationship types of College English reading knowledge graph.

Table 1: Table of core entities and relation types of College English reading knowledge graph

Entity Type	Example Nodes	Relation Type	Relation Description
Lexical Knowledge	inference, contrast, evidence	belongs_to	Lexical items belong to specific semantic function categories.
Syntactic Structure	attributive clause, inversion, concession	affects	Syntactic structures affect comprehension difficulty and processing paths.
Discourse Unit	topic sentence, supporting detail, conclusion	located_in	Information points are located in specific discourse positions.
Logical Relation	cause-effect, comparison, transition	connects	Logical relations connect contextual semantic units.
Question Type	main idea, detail, inference, attitude	corresponds_to	Question types correspond to different strategy requirements.
Reading Strategy	skimming, scanning, inferencing, summarizing	supports	Strategies support question solving and information integration.
Behavioral Feature	dwell time, regression count, answer latency	reflects	Behavioral features reflect strategy execution states.

In order to formally represent the structure of knowledge graph, this paper defines the knowledge graph of college English reading as follows:

$$\mathrm{G} = (\mathrm{V}, \mathrm{E}, \mathrm{T}, \mathrm{X}) \tag{1}$$

where $\mathrm{V} = \{\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_\mathrm{N}\}$ denotes the set of entity nodes, E denotes the set of edges, T denotes the set of relation types, and $\mathrm{X} \in \mathbb{R}^{\mathrm{N}\times\mathrm{d}}$ denotes the node feature matrix. The expression maps vocabulary, syntax, discourse, question type, strategy and action nodes into a unified semantic space, which lays a data foundation for subsequent graph reasoning and strategy recognition.

Only node connection is not enough to reflect the real strength relationship between reading knowledge, and it also needs to calculate the semantic association weight between nodes. In this paper, the joint modeling method of semantic similarity and context co-occurrence strength is used to express the association weight of nodes v_i and v_j as follows:

$$w_{ij} = \lambda \cdot \frac{h_i^\top h_j}{|h_i||h_j|} + (1-\lambda) \cdot c_{ij} \tag{2}$$

where hi and hj are node embedding vectors, cij is the co-occurrence strength of nodes in the same sentence, the same paragraph or the same topic environment, and $\lambda \in [0,1]$ is the balance coefficient. This formulation enables the graph to both preserve deep semantic proximity and reflect contextual dependencies in real reading tasks. For example, "turning connective" and "author attitude question" may not be close to each other in the semantic space, but they have stable co-occurrence relationship in the reading task, so they can still form a high weight connection.

After node modeling and edge weight calculation, the system further uses relation constraints to achieve semantic propagation. Let the state of node v_i in the l-th layer graph representation be z_i^l, then its update form can be written as follows:

$$\mathrm{z}_\mathrm{i}^{(\mathrm{l}+1)} = \sigma\left(\sum_{\mathrm{j}\in\mathrm{N}(\mathrm{i})} \alpha_{\mathrm{ij}}^{(\mathrm{l})} \mathrm{W}_{\mathrm{t}_{\mathrm{ij}}}^{(\mathrm{l})} \mathrm{z}_\mathrm{j}^{(\mathrm{l})} + \mathrm{b}^{(\mathrm{l})}\right) \tag{3}$$

Here, N(i) is the neighborhood set of node v_i, $\mathrm{W}_{\mathrm{t}_{\mathrm{ij}}}^{(\mathrm{l})}$ is the transformation matrix corresponding to relation type t_{ij}, $\alpha_{\mathrm{ij}}^{(\mathrm{l})}$ represents the contribution weight of neighbor nodes to the current node, and $\sigma(\cdot)$ is the nonlinear activation function. This propagation mechanism can compress the multi-hop association between discourse structure-question type requirements and strategy selection into the node representation, so that the graph has the reasoning ability of serving reading comprehension strategy recognition.

Based on the above process, the college English reading knowledge graph constructed in this paper achieves three functions. First, the reading knowledge points are transformed from scattered items into a computable and traceable relationship network. The second is to establish a stable mapping between text semantic information and strategy labels to enhance the interpretation ability of the subsequent model for reading paths. The third is to provide a structural interface for the access of learning behavior data, so that the stay time, review frequency and question type bias can be connected back to the specific knowledge node and strategy node. Therefore, the knowledge graph not only completes the structured organization of reading resources, but also provides a clear semantic support for the subsequent fusion modeling of text features and learning behavior features.

3.2 The fusion modeling of reading text semantic features and learning behavior features

In the analysis of college English reading comprehension strategies, relying solely on the characteristics of the text content can only explain "what difficulties the material itself has", but it is difficult to answer how students read, what aspects of comprehension deviation occur, and what kinds of strategies are insufficiently used. If we only analyze the learning behavior log, it is easy to deviate from the text semantic environment, and it is impossible to explain the cognitive reasons behind the dwell time, the number of returns and the answer delay. Based on this, this paper puts the semantic features of reading text and learning behavior features into a unified modeling framework, and forms a joint representation for reading comprehension strategy recognition through multi-source feature alignment, temporal coding and gated fusion. Figure 2 shows the fusion framework of reading text semantic features and learning behavior features.

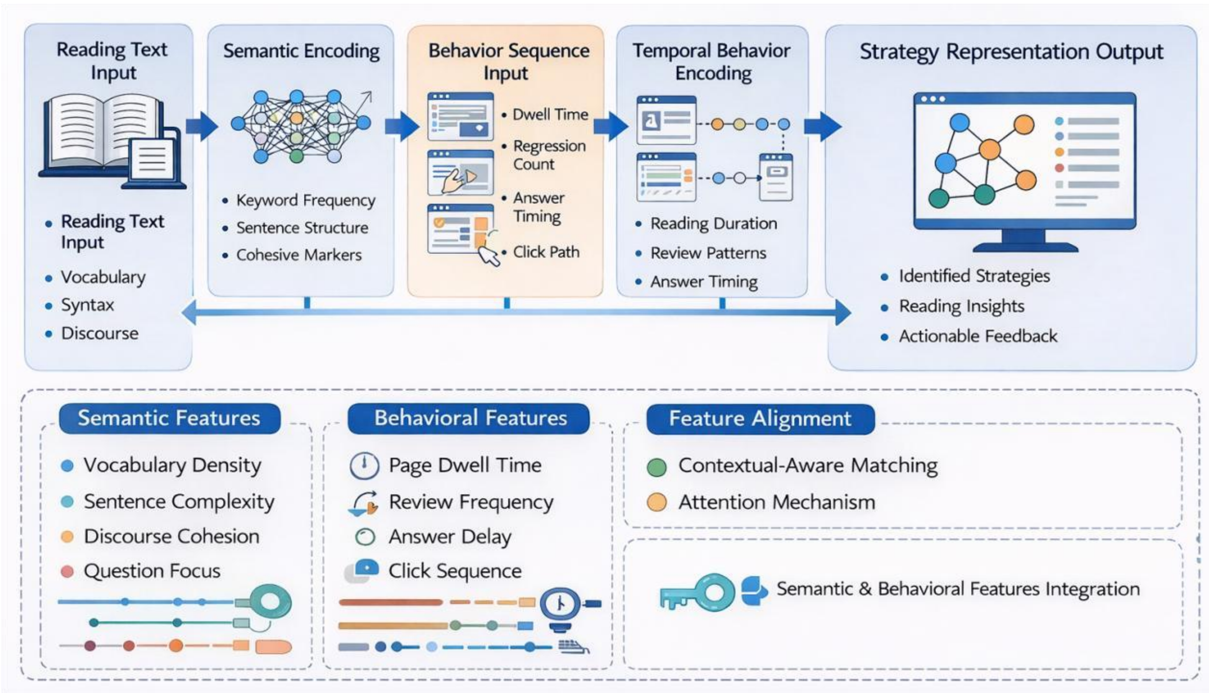

Figure 2: Framework diagram of the fusion of reading text semantic features and learning behavior features

On the text side, the system extracts word density, syntactic complexity, discourse cohesion, topic sentence distribution, logical connectives strength and topic pointing information from the reading materials, and combines the context encoding model to form the semantic representation. Suppose that the reading text is segmented and encoded to obtain the word vector sequence $\{e_1, e_2, \ldots, e_n\}$,whose semantic aggregation representation is defined as follows:

$$r_s = \sum_{i=1}^{n} \alpha_i e_i, \quad \alpha_i = \frac{\exp(u_i)}{\sum_{j=1}^{n} \exp(u_j)} \tag{4}$$

Here, r_s represents the text semantic representation, α_i is the attention weight of the ith word in the current reading task, and u_i reflects the importance of the contribution of the word to the overall understanding. The expression can highlight the role of topic sentences, logical connectives, evaluative words and key evidence information in reading processing, so that the model no longer treats all words equally, but is closer to the focus mechanism in real reading.

On the behavior side, the system constructs the behavior sequence based on the learning platform log, focusing on preserving the information such as page stay time, review times, click order, topic switching path and answer delay. Considering that different behaviors do not contribute consistently to policy judgment, this paper does not adopt a simple concatenation method, but dynamically fuses it with semantic constraints through time-aware encoding. Let rb be the behavior representation vector, then the fusion result of text semantic representation and behavior representation is denoted as follows.

$$h_f = g \odot r_s + (1 - g) \odot r_b, \quad g = \sigma(W_s r_s + W_b r_b + b) \tag{5}$$

Here, h_f is the final fusion representation, g is the gating coefficient, $\odot$ represents element-wise multiplication, W_s, W_b, and b are trainable parameters, and $\sigma(\cdot)$ is the Sigmoid function. This mechanism can adaptively adjust the proportion of text information and behavior information according to the characteristics of the current sample. When students were faced with logical inference problems and looked back frequently, the model would enhance the explanatory role of behavioral characteristics. When the text syntax is complex, the discourse transition is dense and the behavior fluctuation is small, the system will increase the influence weight of semantic features. The multi-source feature composition and variable description are shown in Table 2.

Table 2:Multi-source feature composition and variable description table

Feature Category	Variable Name	Description	Data Source
Lexical Feature	Vocabulary Density	Proportion of keywords and academic words in a unit of text	Reading Text
Syntactic Feature	Sentence Complexity	Complexity of long sentences, subordinate clauses, and nested structures	Reading Text
Discourse Feature	Discourse Cohesion	Strength of inter-sentence cohesion and inter-paragraph coherence	Reading Text
Logical Feature	Logical Connector Strength	Distribution strength of connectors such as cause-effect, contrast, and progression	Reading Text
Question Feature	Question Orientation	The direction of assessment regarding main idea, details, or inference	Test Items
Behavioral Feature	Dwell Time	Time spent on a page or an item	Platform Logs
Behavioral Feature	Regression Count	Number of look-backs and rereading actions	Platform Logs
Behavioral Feature	Click Sequence Depth	Depth of the click path during reading and answering	Platform Logs
Behavioral Feature	Answer Latency	Response time from reading to answering	Test Records
Fused Feature	Fused Representation	Joint representation result of semantic and behavioral features	Model Output

Based on the multi-source feature design in Table 2, the text feature describes "what to read", and the behavior feature describes "how to read", and the two jointly determine the discriminant boundary of reading comprehension strategy identification. Compared with the method only using semantic encoding, fusion modeling can more accurately distinguish the two cases of "ununderstood text" and "improper strategy use", which have similar surface results but different internal mechanisms. For example, if a student loses points on the topic question, if he/she also shows insufficient retention of the topic sentence and frequent skipping between paragraphs, he/she is more likely to have weak skiming strategy. If the length of stay is sufficient but the reader returns to the transition relationship many times, it may be related to the insufficient ability of logical integration. Therefore, the fusion model not only improves the accuracy of policy recognition, but also enhances the teaching interpretation of the analysis results, and provides a more discriminative input representation for subsequent graph reasoning and deep learning recognition modules.

3.3 Reading Comprehension Strategy Identification Method Based on Graph Reasoning and Deep Learning

After integrating the semantic features of reading texts and learning behavior features, how to further transform the multi-source information into discriminative strategy categories is the core link in the intelligent analysis of college English reading comprehension. Traditional classification models usually use the processing method of vector concatenation and then directly enter the full connection layer, which can complete the basic recognition task, but it is insufficient to excavate the association path between "lexical clues, discourse relations, question type requirements and behavior performance", especially in the information processing that needs to be integrated across sentences such as inference questions, attitude questions and topic questions. It is prone to the problem of fuzzy policy boundary and unstable discrimination. In order to improve the accuracy and interpretability of reading comprehension strategy recognition, this paper introduces the graph reasoning mechanism and deep learning coding framework on the basis of fusion features, and constructs a recognition model of "graph structure modeling-relationship propagation-deep representation-strategy classification". Figure 3 shows the structure of the policy recognition model based on graph reasoning and deep learning.

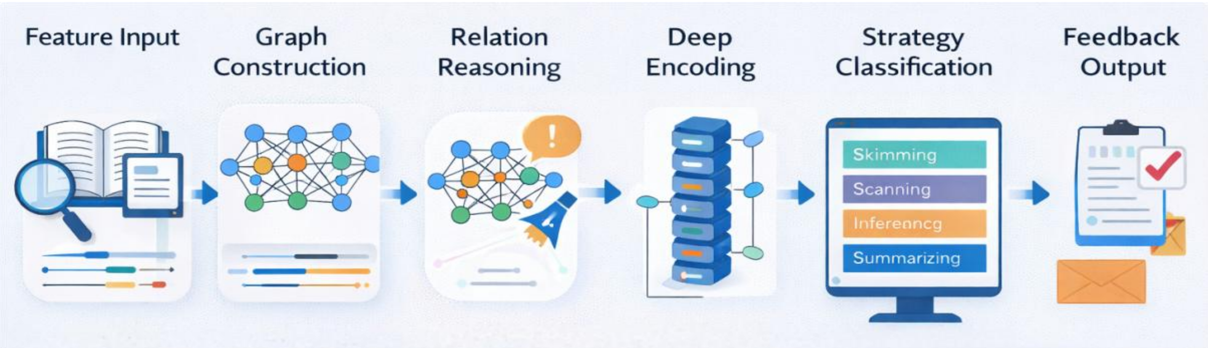

Figure 3: Structural diagram of policy recognition model based on graph reasoning and deep learning

In the graph structure modeling phase, the system maps lexical nodes, syntactic nodes, discourse nodes, question type nodes, action nodes and strategy candidate nodes into a unified graph space, and builds a heterogeneous graph based on semantic association, logical connection and action reflection relationship. Let the representation of node v_i in the initial layer be $h_i(0)$, then its neighborhood relationship-aware representation can be written as follows:

$$m_i^{(l)} = \sum_{j \in \mathcal{N}(i)} \beta_{ij}^{(l)} W_{r_{ij}}^{(l)} h_j^{(l)} \tag{6}$$

Here, N(i) is the neighborhood set of node v_i, $W_{r_{ij}}^{(l)}$ is the mapping matrix corresponding to the relationship type r_{ij}, and $\beta_{ij}^{(l)}$ is the neighbor node contribution coefficient. This formula emphasizes the differences in the role of different relationship types in policy recognition. For example, "located_in" reflects more discourse position constraints, and "reflects more mapping ability of behavior features to policy states.

In order to avoid the average processing of different neighbor nodes, the model further introduces the relation attention mechanism to give higher weights to key information nodes. It is calculated as follows:

$$\beta_{ij}^{(l)} = \frac{\exp\left(\text{LeakyReLU}\left(a^\top [W h_i^{(l)} \| W h_j^{(l)} \| e_{r_{ij}}]\right)\right)}{\sum_{k \in N(i)} \exp\left(\text{LeakyReLU}\left(a^\top [W h_i^{(l)} \| W h_k^{(l)} \| e_{r_{ik}}]\right)\right)} \tag{7}$$

Here, a is the trainable attention vector, // denotes the concatenation of vectors, and $e_{r_{ij}}$ is the relation embedding. This mechanism can automatically amplify the joint effect between logical connectives, supporting sentences and looking back behavior in inference question recognition, and improve the contribution degree of topic sentence-discourse structure-skiming behavior in topic question recognition.

After the multi-layer graph propagation is completed, the model inputs the graph inference results and the above-mentioned fusion features into the deep encoding layer to obtain a more stable policy discrimination representation. Let the graph inference output be g_i and the multi-source fusion be hf, then the final strategy representation is defined as follows:

$$z_i = \text{ReLU}(W_c [g_i \| h_f] + b_c) \tag{8}$$

Here, W_c and b_c are linear transformation parameters. The formula further compresses the structural reasoning results and the joint features of semantic behavior into a unified discriminant space, so that the model not only retains the reasoning depth brought by the graph relationship, but also maintains the sample-level individual behavior differences.

In the strategy classification stage, the reading comprehension strategies are divided into skimping, scanning, inferencing, and summarizing, and the probabilities of each strategy are output by Softmax. For sample i, the prediction result can be expressed as follows:

$$\hat{y}_i = \text{softmax}(W_o z_i + b_o) \tag{9}$$

Here, $\hat{y}_i$ is the policy probability distribution, W_o and b_o are the output layer parameters. Through this layer, the system can not only give the final strategy label, but also output the confidence level of each candidate strategy, which provides a continuous basis for subsequent teaching feedback, rather than a simple single label judgment.

Considering the unbalanced distribution of different strategy categories in the samples, if the ordinary cross-entropy loss is directly used, it is easy to cause the high-frequency categories to dominate the model training. To this end, this paper introduces a joint optimization objective of class weight constraint and regularization term, and its loss function is written as follows:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{c=1}^{C} \omega_c \, y_i c \log \hat{y}_i \, c + \mu \|\Theta\|_2^2 \tag{10}$$

Here, ω_c is the weight coefficient of the CTH strategy, $y_i c$ is the true label, $\hat{y}_i$ c is the predicted probability, Θ is the set of model parameters, μ is the regularization coefficient. The proposed objective function can suppress parameter overfitting while ensuring classification accuracy, so that the model can maintain stable recognition ability under the condition of uneven sample size and behavior noise.

Through the above modeling process, the strategy recognition method constructed in this paper has three advantages. First, the graph reasoning mechanism enhances the association mining ability of cross-sentence, cross-segment and cross-action features, and can identify complex strategies that are difficult to judge by only relying on shallow text features. Secondly, the deep coding layer realizes the collaborative compression of structural information and individual behavior information, which improves the robustness of policy classification. Thirdly, the design of probability output and weighted loss enables the model to adapt to the problems of unbalanced strategy samples and large differences in behavior patterns in real teaching environments. Therefore, this method provides a stable recognition core for the integration and feedback optimization of the subsequent reading comprehension strategy intelligent analysis system.

3.4 Integration and feedback optimization mechanism of reading comprehension strategy intelligent analysis system

What is really needed in the college English reading scenario is not only to judge what strategies students use, but also to convert the recognition results into executable, traceable and iterative teaching feedback, and to enable the system to continuously correct the analysis bias in continuous interaction. Therefore, this paper constructs an intelligent reading comprehension strategy analysis system based on the above model, which organizes data access, graph invocation, feature fusion, policy recognition, feedback generation and result regurgitation into a unified closed loop, so that the system has both online analysis ability and feedback optimization ability.

From the perspective of system structure, the overall framework is composed of data access layer, semantic support layer, policy analysis layer, feedback generation layer and result reflux layer. The data access layer is responsible for collecting reading texts, test data, teacher labeling information and platform logs, and completing the format unification and timestamp alignment. The semantic support layer calls the reading knowledge graph to associate and retrieve text entities, logical relationships, question type nodes and behavior nodes. The policy analysis layer receives the semantic features and behavior features from the front-end, and combines the graph reasoning model to complete the policy recognition. The feedback generation layer outputs personalized suggestions according to the current strategy status of students. Results The reflux layer records the revision behavior of students in the subsequent reading, and re-sends it to the system to form a dynamic update chain.

In order to measure the matching degree between the current recognition results and the feedback content, this paper defines the feedback fitness function as follows:

$$\phi_i = \eta s_i + (1 - \eta)\rho_i \tag{11}$$

Here, ϕ_i represents the feedback fitness of the ith learning sample, si represents the strategy identification confidence, ρ_i represents the degree of correlation between the feedback content and the learner's problem state, and $\eta \in [0,1]$ is the regulation coefficient.

On this basis, in order to further characterize the learner's weakness in a certain strategy dimension, this paper constructs a strategy deviation intensity index:

$$\delta_{ic} = \alpha(1 - \hat{y}_{ic}) + (1 - \alpha)b_{ic} \tag{12}$$

Here, δ_{ic} represents the deviation strength of learner i on the type c reading strategy, $\hat{y}_{ic}$ is the predicted probability of the model for this strategy, b_{ic} is the abnormal degree of the strategy calculated from the behavior log, and $\alpha \in [0,1]$ is the balance parameter.

Considering that a learner may have multiple strategy shortboards at the same time, the system also needs to prioritize different feedback tasks. To this end, this paper defines the allocation weight of each type of feedback as follows:

$$\pi_{ic} = \frac{\exp(\delta_{ic})}{\sum_{k=1}^{C} \exp(\delta_{ik})} \tag{13}$$

Here, π_{ic} represents the priority of the system to assign feedback content of type c to learner i, and C is the total number of strategy categories. Through this normalization mechanism, the system can automatically determine whether skimming training, scanning training or inferring training should be strengthened.

In the process of feedback generation at the student side, the system forms hierarchical feedback according to different strategy weak items. For students with insufficient skim reading, the feedback focused on fast positioning of discourse structure and identification of topic sentences. For students who lack of searching, detail information search and keyword tracking are highlighted. For students with weak inference strategies, context clue integration, attitude recognition and implicit relationship judgment are strengthened. At the same time, the teacher terminal interface will synchronously display the strategy distribution, error clustering and problem type correlation graph at the class level, so that teachers can understand the learning problem from two scales of individual and group.

Considering that students' reading status will constantly change with training, the system needs to update the learner strategy profile online. Let the learner's strategy state vector after the TTH round of interaction be qi(t), then its update form is denoted as follows:

$$q_i^{(t+1)} = (1 - \gamma)q_i^{(t)} + \gamma f_i^{(t)} \tag{14}$$

Here, $f_i^{(t)}$ represents the actual behavior improvement result observed after the current round of feedback, and γ is the update rate. In order to more specifically measure the actual gain after feedback intervention, this paper further defines the learning improvement gain as follows:

$$G_i^{(t)} = \omega_1 \Delta A_i^{(t)} + \omega_2 \Delta R_i^{(t)} - \omega_3 \Delta L_i^{(t)} \tag{15}$$

where $G_i^{(t)}$ represents the comprehensive improvement gain of learner i after the TTH feedback, $\Delta A_i^{(t)}$ represents the improvement of reading accuracy, and $R_i^{(t)}$ represents the decrease amplitude of invalid look back behavior, $\Delta L_i^{(t)}$ represents the average answer delay change, and ω_1,ω_2,ω_3 are the corresponding weights.

In addition to the feedback content itself, the efficiency of system operation also determines the actual teaching usability. English reading training in colleges and universities usually has the requirements of concurrent access, batch analysis and instant feedback in class. If the system response delay is too high, even if the recognition is accurate, it is difficult to embed into the real teaching scene. To this end, this paper first defines the system average response delay as follows:

$$D_{avg} = \frac{1}{N}\sum_{i=1}^{N}(t_i^{out} - t_i^{in}) \tag{16}$$

Here, t_i^{in} and t_i^{out} represent the input and output time of the ith request, respectively, and N is the total number of requests. This metric is used to measure the response efficiency of the system in a real deployment and to provide performance constraints for immediate feedback in a classroom setting.

On this basis, this paper further constructs the joint optimization objective of analysis accuracy and response performance:

$$J = \lambda_1(1 - Acc) + \lambda_2 D_{avg} + \lambda_3(1 - Sta) \tag{17}$$

Here, Acc represents the policy identification accuracy, Sta represents the system stability index, and λ_1,λ_2, and λ_3 are weight parameters. The objective function integrates the recognition quality, timeliness and operation stability into the same evaluation framework, which is helpful to adjust the parameters of the system in different deployment environments. For example, in the after-class large-scale batch analysis scenario, the accuracy weight can be appropriately increased. In the classroom immediate feedback scenario, the time delay control can be moderately emphasized to make the system output more in line with the teaching rhythm.

4 Experimental design and result analysis

4.1 Data set construction and experimental environment configuration

In order to verify the effectiveness of the intelligent analysis model of knowledge graph-assisted reading comprehension strategy proposed in this paper, the experimental data are constructed from the real scenario of college English reading teaching, and the data sources include reading texts, supporting test questions, teacher strategy annotation results and learning platform behavior logs. Reading texts were selected from 168 exposative, argumentative and applied texts in college English courses, with a length of 450-900 words. There are 840 questions, covering such types as topic understanding, detail location, logical inference, author attitude and word sense guessing. Based on teacher review and classroom observation, students 'strategy performance in sample tasks such as skiming, scanning, inference and generalization were labeled, and 4120 strategy samples that could be used for supervised training were formed. At the same time, the platform log records the behavior information such as stay time, look back times, click order, and answer delay, and collects a total of 28640 valid behavior fragments.

After data preprocessing, the samples are divided into training set, validation set and test set according to the ratio of 8：1：1 to ensure that model training, parameter tuning and result evaluation are independent of each other. The experimental platform is developed with Python 3.10 and PyTorch 2.1, the graph structure processing module is implemented based on

PyTorch Geometric, and the knowledge graph storage is Neo4j 5.0. The hardware environment is Intel Xeon Silver processor, 64 GB memory and NVIDIA RTX 4090 graphics card with 24 GB video memory. The model training batch size is set to 32, the initial learning rate is set to 0.001, the optimizer is selected AdamW, the training rounds are set to 80, and the early stopping mechanism is triggered when the validation set has no improvement for 10 consecutive rounds. The configuration can better support the stable operation of multi-source feature fusion, graph inference propagation and policy classification tasks, and provide a unified environment for the analysis of subsequent experimental results.

4.2 Data preprocessing and knowledge graph labeling mechanism

In order to ensure that reading text, question information and behavior log can be co-modeled in a unified semantic space, the original data are standardized pre-processed before the experiment. On the text side, it completed clause segmentation, word segmentation, lemming and stop symbol cleaning, and marked clause boundaries, logical connectives and topic sentence positions in long sentences. The behavior side performs missing repair and abnormal truncation on stay time, look back times, click sequence and answer delay. Min-max normalization is applied to continuous features of the form:

$$x_i^* = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{18}$$

Here, x_i^* is the normalized result, and x_{min} and x_{max} represent the minimum and maximum value of the feature, respectively. This processing helps to reduce the interference of different dimensions on model training.

In the knowledge graph annotation stage, this paper maps lexical knowledge, syntactic structure, discourse unit, topic type, behavior feature and strategy category into heterogeneous nodes, and generates relationship edges jointly by teacher annotation and rule matching. In order to improve the reliability of annotation, the node-relation joint confidence is defined as follows:

$$\psi_{ij} = \theta a_{ij} + (1 - \theta) r_{ij} \tag{19}$$

Here, a_{ij} represents the manual annotation agreement, r_{ij} represents the rule matching score, and θ is the balance coefficient. Only the annotation results whose confidence is higher than the threshold are retained to be written into the graph, so as to ensure the stability of subsequent graph reasoning and strategy identification.

4.3 Evaluation index system and comparative experimental scheme

In order to comprehensively evaluate the performance of the proposed model in the intelligent analysis task of college English reading comprehension strategies, an evaluation index system is constructed from three dimensions of recognition accuracy, classification stability and system operation efficiency. In terms of recognition performance, accuracy rate, precision rate, recall rate and F1 value are selected to measure the discrimination ability of the model for strategy categories such as skimming, scanning, inference and generalization. The F1 value is defined as follows:

$$F1 = \frac{2PR}{P + R} \tag{20}$$

Here, P represents precision and R represents recall. This index can comprehensively reflect the recognition quality of the model under the condition of class imbalance. In order to further investigate the reliability of the system in continuous analysis tasks, this paper also records the average response time delay and policy output consistency, which are used to evaluate the stability of the model in teaching application scenarios.

In the contrast experiment design, this paper sets up four groups of methods for comparison. The first baseline model is a machine learning classification method using only text semantic features, which is used to test the effect of traditional shallow modeling. The second baseline model is a deep learning method that only combines text features and behavioral features, which is used to compare the performance changes caused by multi-source feature fusion. The third baseline model is a deep classification model without introducing the reasoning mechanism of knowledge graph, which is used to verify the contribution of graph structure modeling. The fourth group is the fusion model of knowledge graph assisted graph reasoning and deep learning proposed in this paper. All models were run under the same data partition, the same training rounds, and the same hardware environment to ensure comparable results. At the same time, the ablation analysis is set up to remove the knowledge graph module, behavior feature module and feedback constraint module respectively, so as to further observe the influence of each key component on the final policy recognition performance.

4.4 Analysis of reading comprehension strategy identification effect and ablation experiment

The experimental results of different models in the reading comprehension strategy recognition task are shown in Table 3. In order to present the overall recognition effect and the contribution of key modules at the same time, the baseline method, the knowledge graph removal module, the behavior feature removal module and the full model are included in the unified comparison.

Table 3: Comparison of reading comprehension strategy recognition effect and ablation experiment results

Model/Method	Accuracy/%	Precision/%	Recall/%	F1/%	Avg. Delay/s
Traditional Text Feature Classification Model	81.6	80.9	79.8	80.3	0.41
Deep Model with Text Semantics and Behavioral Features	86.8	86.1	85.4	85.7	0.58
Without Knowledge Graph Module	88.2	87.6	87.1	87.3	0.64
Without Behavioral Feature Module	87.4	86.9	86.2	86.5	0.61
Proposed Model	91.9	91.3	90.8	91.0	0.67

It can be seen from Table 3 that the model in this paper achieves the best results on all indicators, the strategy recognition accuracy reaches 91.9%, and the F1 value reaches 91.0%, which are 10.3 and 10.7 percentage points higher than the traditional text feature classification model respectively, indicating that the collaborative modeling of knowledge graph, semantic features and learning behavior can significantly enhance the ability of reading strategy discrimination. From the ablation results, the F1 value decreased to 87.3% after removing the knowledge graph module. After removing the action feature module, the F1 value drops to 86.5%, indicating that both graph reasoning and action information have stable gains for policy recognition. Although the average response delay of the full model rises to 0.67 s, it is still within the acceptable range.

5 Discussion

5.1 Performance comparison with existing English reading comprehension analysis methods

In order to further test the comprehensive performance of the proposed method in the task of college English reading comprehension strategy recognition, this paper compares and analyzes the proposed method with the traditional text feature classification model, BiLSTM reading strategy recognition model, Transformer encoding model and semantic behavior fusion deep learning model in a unified data division and experimental environment. This paper focuses on the differences of different methods in terms of Accuracy, Precision, Recall and F1, so as to evaluate the practical role of knowledge graph, graph reasoning mechanism and multi-source feature fusion in improving the overall recognition performance. The comprehensive performance comparison of different methods is shown in Figure 4.

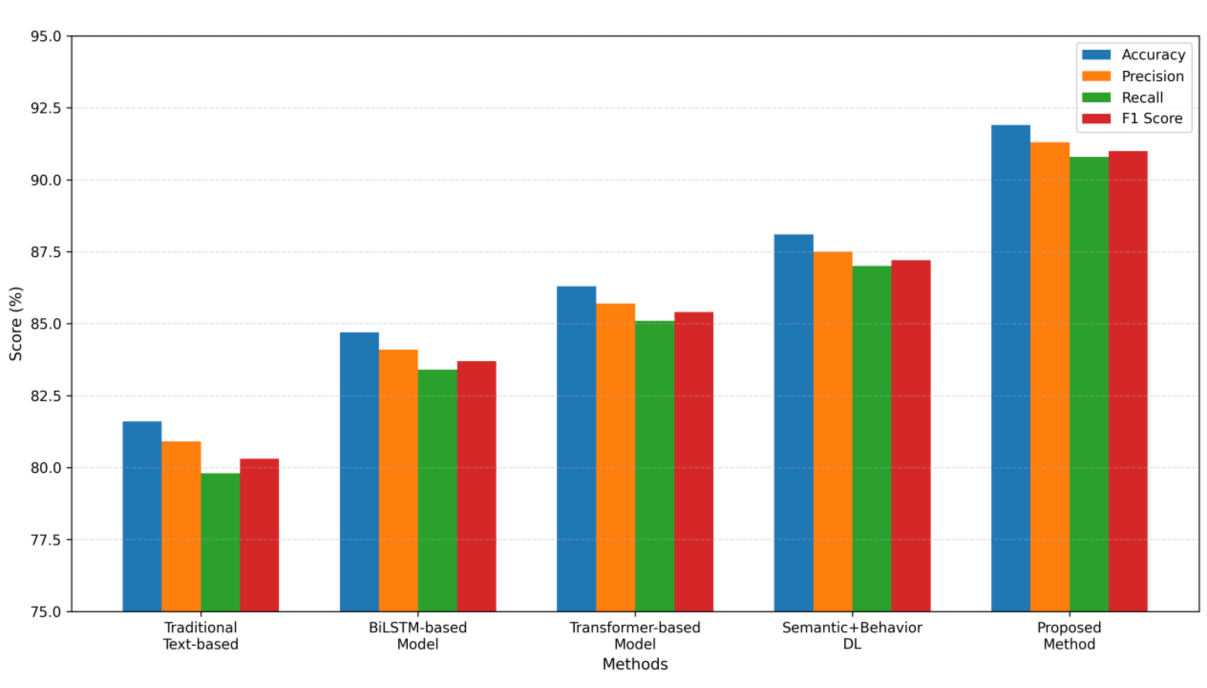

Figure 4: Bar chart of comparison of comprehensive performance of different methods

It can be seen from Figure 4 that the proposed method maintains the optimal level in the four core indicators, with Accuracy, Precision, Recall and F1 reaching 91.9%, 91.3%, 90.8% and 91.0%, respectively. Compared with the traditional text feature classification model, the accuracy is increased by 10.3 percentage points, and the F1 is increased by 10.7 percentage points. Compared with the BiLSTM reading strategy recognition model, the accuracy is increased by 7.2 percentage points and F1 is increased by 7.3 percentage points. Compared with the Transformer encoding model, the accuracy is increased by 5.6 percentage points, and the F1 is increased by 5.6 percentage points. Compared with the semantic behavior fusion deep learning model, the Accuracy and F1 of the proposed method are still improved by 3.8 and 3.8 percentage points, respectively. This indicates that only relying on sequence modeling or shallow feature representation is difficult to fully describe the semantic association, question type constraints, and behavior differences in reading text, and the graph reasoning mechanism driven by knowledge graph can more effectively integrate multi-source information, so as to perform better in comprehensive discrimination ability.

5.2 Adaptive analysis of knowledge graph driven reading strategy analysis

In addition to overall performance, RCI models also need to maintain stable performance across different task types, especially in topic understanding, detail localization, logical inference, author attitude judgment, and word sense guessing tasks, which face different semantic depth and reasoning requirements. In order to investigate the adaptability of our method in various reading task scenarios, we further compare the F1 performance of traditional text feature classification model, BiLSTM reading strategy recognition model, Transformer encoding model, semantic behavior fusion deep learning model and our method in different tasks. The comparison of model adaptability under different reading task scenarios is shown in Figure 5.

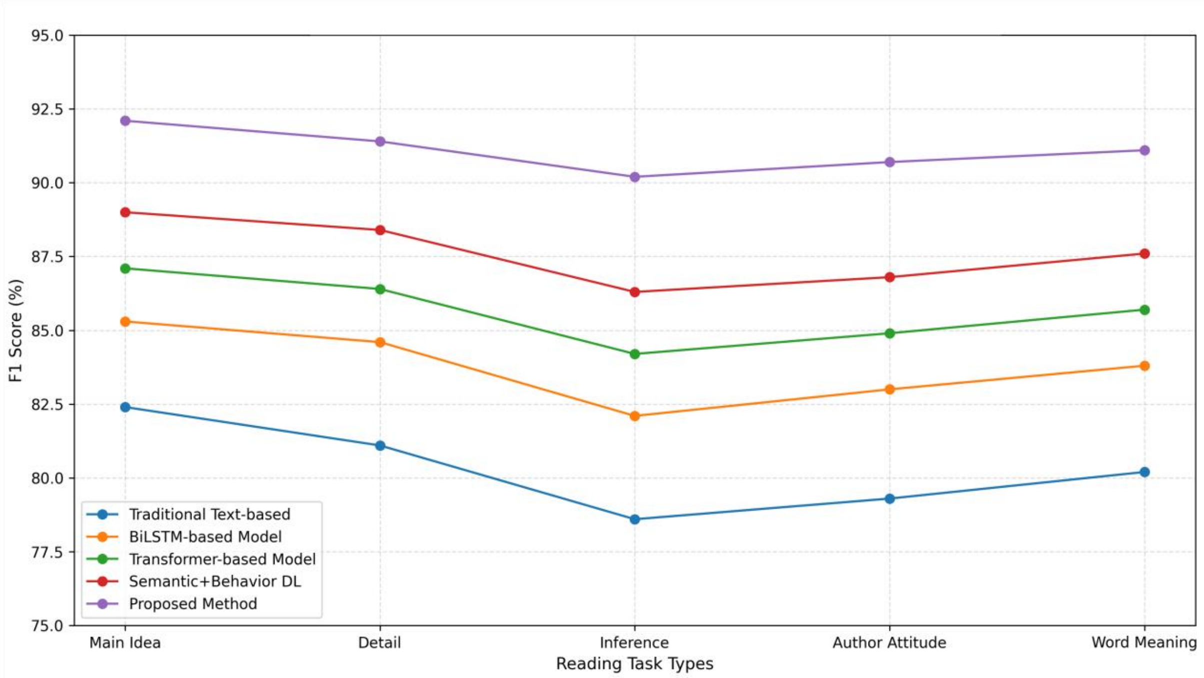

Figure 5: Line chart comparing model adaptability for different reading task scenarios

It can be seen from Figure 5 that the proposed method maintains the highest recognition level in all five types of reading tasks, and the F1 score of topic understanding, detail location, logical inference, author attitude judgment and word sense guessing tasks reaches 92.1%, 91.4%, 90.2%, 90.7% and 91.1%, respectively. Among them, the logical inference task was the most difficult overall, and the scores of each model decreased, but the F1 of the proposed method still maintained 90.2%, which was 11.6 percentage points higher than that of the traditional text feature model, 8.1 percentage points higher than that of the BiLSTM model, and 6.0 percentage points higher than that of the Transformer model. It shows that knowledge graph relation modeling has obvious advantages for cross-sentence reasoning and implicit semantic recognition. In the task of author attitude judgment, the proposed method is 3.9 percentage points higher than the semantic behavior fusion deep learning model. In the task of word sense guessing, it is 5.4 percentage points higher than that of the Transformer model, indicating that the graph driven semantic association representation can enhance the adaptation ability of the model to complex texts, deep logic and fine-grained semantic cues. On the whole, the proposed method not only has higher overall performance, but also has more balanced performance and stronger adaptability in different reading task scenarios.

5.3 Evaluation of system resource overhead and computational efficiency

While the performance of the model is improved, the system resource overhead and operating efficiency also directly affect its actual deployment value in college English reading teaching scenarios. This paper further compares the differences in response efficiency between the traditional text feature classification model, the BiLSTM model, the Transformer model, the semantic behavior fusion deep learning model and the proposed method, and comprehensively evaluate the system availability by combining the video memory occupation and training time. The comparison of average response delay of different methods is shown in Figure 6.

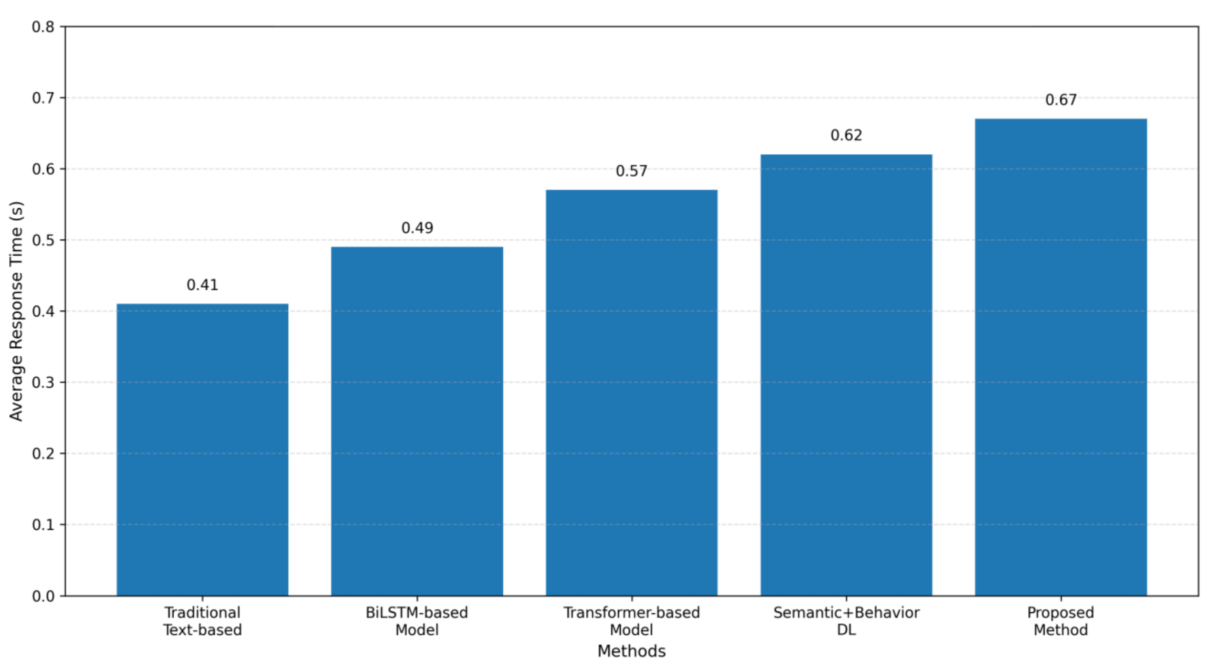

Figure 6: Bar plots comparing average response delays of different methods

It can be seen from Figure 6 that the average response time of each method gradually increases with the increase of model structure complexity. The average response delays of the traditional text feature classification model, the BiLSTM model, the Transformer model, the semantic behavior fusion deep learning model and the proposed method are 0.41 s, 0.49 s, 0.57 s, 0.62 s and 0.67 s, respectively. Although the time delay of the proposed method is the highest, it is still controlled within 0.7 s, which can meet the needs of immediate feedback in classroom diagnosis and after-class training. Further statistics show that the video memory occupation of the proposed method in the training phase is about 11.2GB, which is higher than 9.8GB of the Transformer model and 7.4GB of the BiLSTM model, and the training time of a single round is about 5.1 min, but it gains higher recognition accuracy and stronger task adaptability. On the whole, the proposed method achieves a reasonable balance between resource consumption and computational efficiency, and has good engineering deployment feasibility.

5.4 Analysis of scalability and stability in multi-scene teaching applications

In order to further investigate the application potential of the system in the real teaching environment, this paper starts from four typical scenarios of instant reading in class, independent training after class, unit test analysis and teacher diagnosis support, and comprehensively compares the six dimensions of strategy recognition accuracy, response speed, feedback timeliness, teaching adaptability, system stability and scalability. To analyze the operating performance of the system under multi-scenario conditions. The stability and scalability of the system under multi-scenario teaching application are shown in Figure 7.

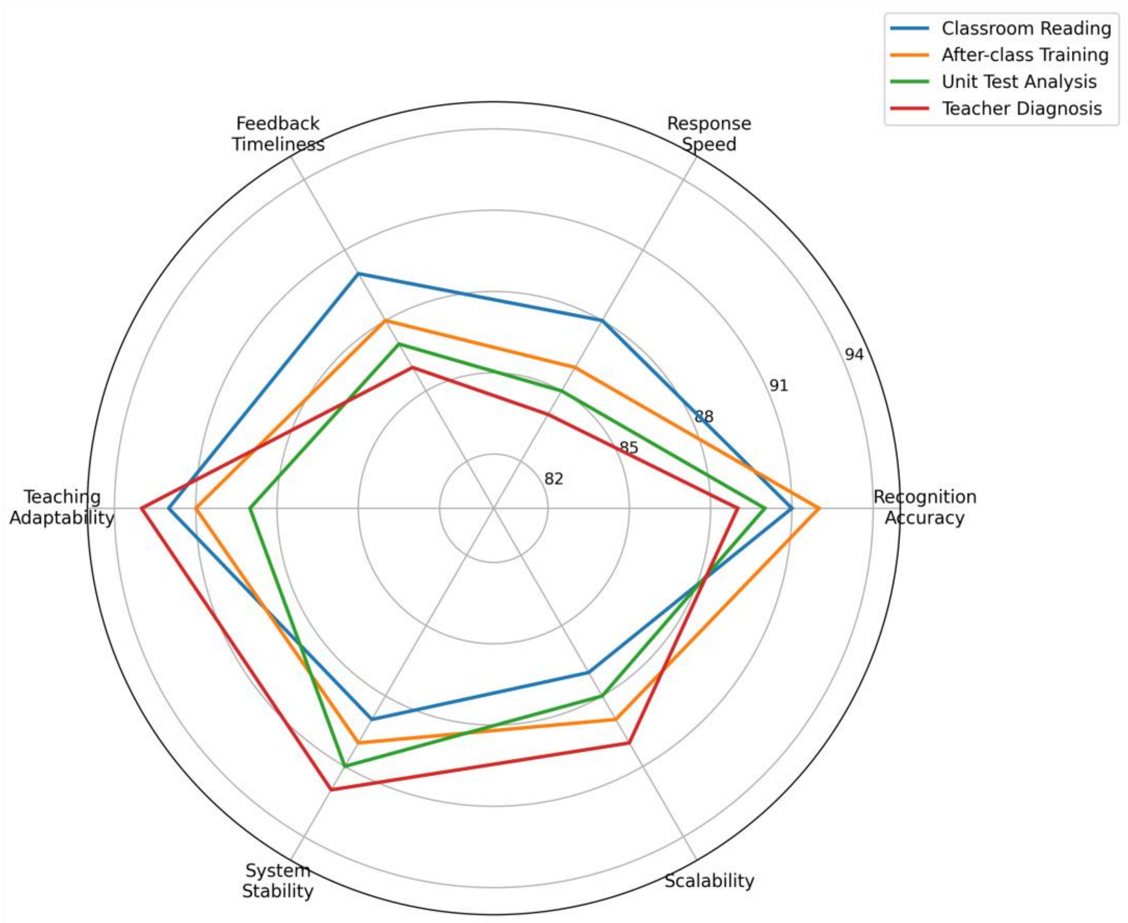

Figure 7: Radar chart of system stability and scalability under multi-scenario teaching application

It can be seen from Figure 7 that the proposed system maintains a high level in all four types of teaching scenarios, and the overall indicators are mainly distributed between 84 and 93. Among them, the classroom instant reading scene performs well in recognition accuracy, feedback timeliness and teaching adaptability, reaching 91, 90 and 91 respectively. The recognition accuracy of after-class autonomous training scenarios reaches 92, and the scalability is 88, indicating that the system is suitable for long-term autonomous learning support. The unit test analysis scenario reaches 91 in terms of system stability, showing good batch processing ability. The teacher diagnosis support scenarios reach 93, 92 and 89 in terms of teaching adaptability, system stability and scalability, respectively, indicating that the system has strong advantages in assisting teachers in strategy diagnosis and teaching decision-making. On the whole, the performance of the system in different teaching scenarios is balanced, which not only has good stable operation ability, but also has strong potential for scene migration and expansion.

5.5 Limitations of the method and future optimization direction

Although the intelligent analysis model of college English reading comprehension strategy assisted by knowledge graph constructed in this paper has shown good results in terms of strategy recognition accuracy, task adaptability and teaching feedback effectiveness, there are still some areas that need to be further improved from the perspective of research completeness and practical application requirements. First, the current experimental data mainly come from common reading materials in college English courses. Although the text genres cover explanatory texts, argumentative essays and applied texts, there are still insufficient samples in interdisciplinary professional English, academic paper excerpts and difficult long texts, which will limit the generalization ability of the model to complex reading scenarios to a certain extent. Second, although the entity relations in knowledge graphs can well support tasks such as topic understanding, detail location and logical inference, the expression of implicit cultural context, discourse rhetorical function and cross-segment deep semantic cohesion are still insufficient, resulting in the problem that the recognition boundaries are not clear enough in a few higher-order understanding tasks. Third, learning behavior data currently mainly rely on platform logs, which can reflect explicit processes such as stay time, review frequency and answer delay, but the description of learners' real cognitive state, attention changes and strategy switching timing is still relatively indirect, which also makes some feedback suggestions remain at the level of "result correlation" and have not yet fully reached the level of more detailed process interpretation.

Further research can be carried out in three directions. On the one hand, to further expand data sources, increase cross-course, cross-difficulty and cross-college samples, and build a reading strategy analysis corpus with wider coverage, so as to improve the transfer ability and robustness of the model. On the other hand, we continue to improve the structure of the knowledge graph by introducing discourse relations, rhetorical functions, cultural background knowledge and cross-paragraph reasoning chains, which makes the graph more complete in higher-order semantic expression, thereby enhancing the reasoning depth under complex tasks. At the same time, a more fine-grained reading process perception model can be constructed by combining eye movement data, process evaluation and generative feedback mechanism, so that the system can be further developed from "recognition strategy" to "explanation strategy formation" and "support strategy improvement". With the continuous expansion of multi-source data fusion capabilities and intelligent teaching scenarios, the knowledge graph-driven reading comprehension strategy analysis method is expected to play a greater role in precision teaching, personalized diagnosis and continuous learning support of college English.

6 Conclusion

Focusing on the intelligent analysis of college English reading comprehension strategies, this paper constructs a technical framework consisting of knowledge graph modeling, multi-source feature fusion, graph reasoning and deep learning recognition, system integration and feedback optimization. In this study, lexical knowledge, syntactic structure, discourse unit, question type information, behavior characteristics and strategy labels are integrated into the unified graph space. Through the joint modeling of semantic representation and behavior log, the structural expression ability and teaching interpretation ability of reading strategy recognition are improved. The experimental results show that the accuracy and F1 value of the proposed model are 91.9% and 91.0%, respectively 10.3 and 10.7 percentage points higher than those of the traditional text feature classification model. It maintains a high level in five tasks: topic understanding, detail location, logical inference, author attitude and word meaning guessing. The average response time of the system is controlled within 0.67 s, and it shows 84-93 stability and scalability in multi-scene teaching applications. In the future, the cross-course samples can be expanded, the expression of higher-order semantic relations can be improved, and the model generalization and continuous feedback ability can be improved by combining more fine-grained process data.

References

[1] Zheng, L., Niu, J., Long, M., & Fan, Y. An automatic knowledge graph construction approach to promoting collaborative knowledge building, group performance, social interaction and socially shared regulation in CSCL. British Journal of Educational Technology, 2023, 54(3): 686–711. DOI: 10.1111/bjet.13283

[2] Abu-Salih, B., & Alotaibi, S. A systematic literature review of knowledge graph construction and application in education. Heliyon, 2024, 10(3): e25383. DOI: 10.1016/j.heliyon.2024.e25383

[3] Zhang, P. Effects of highlights and annotations on EFL learners' reading comprehension: an application of computer-assisted interactive reading model. Computer Assisted Language Learning, 2024: 1–33. DOI: 10.1080/09588221.2024.2410166

[4] Shang, J., Huang, Y., Xu, M., Huang, Y., Shen, X., Wang, G., Wang, Y., & Zhang, L. Competition between human learners and ChatGPT: enhancing university EFL students' reading comprehension and critical thinking through competitive questioning. Computer Assisted Language Learning, 2025: 1–35. DOI: 10.1080/09588221.2025.2577356

[5] Muche, T., Simegn, B., & Shiferie, K. Self-efficacy and metacognitive strategy use in reading comprehension: EFL learners' perspectives. The Asia-Pacific Education Researcher, 2024, 33(1): 219–227. DOI: 10.1007/s40299-023-00721-5

[6] Samiei, F., & Ebadi, S. Exploring EFL learners' inferential reading comprehension skills through a flipped classroom. Research and Practice in Technology Enhanced Learning, 2021, 16: 12. DOI: 10.1186/s41039-021-00157-9

[7] Geng, X., & Yamada, M. Using learning analytics to investigate learning processes and behavioural patterns in an augmented reality language learning environment. Journal of Computer Assisted Learning, 2023, 39(2): 532–546. DOI: 10.1111/jcal.12760

[8] Liu, S., Sui, Y., You, Z., Shi, J., Wang, Z., & Zhong, C. Reading better with AR or print picture books? A quasi-experiment on primary school students' reading comprehension, story retelling and reading motivation. Education and Information Technologies, 2024,

29: 11625–11644. DOI: 10.1007/s10639-023-12231-4

[9] Kaygısız, Ç. Comparision of digital and printed text reading process. Education and Information Technologies, 2025, 30: 22709–22733. DOI: 10.1007/s10639-025-13668-5

[10] Shi, Y., & Lee, B. Chinese EFL middle school learners' reading skills: a latent profile analysis. The Asia-Pacific Education Researcher, 2025, 34: 1677–1688. DOI: 10.1007/s40299-025-00980-4

[11] Vettori, G., et al. Key language, cognitive and higher-order skills for L2 reading comprehension of expository texts in English as foreign language students: a systematic review. Reading and Writing, 2024, 37: 2481–2519. DOI: 10.1007/s11145-023-10479-3

[12] Nash, H., Dixon, C., Clarke, P., et al. Dynamic assessment of word learning as a predictor of vocabulary, reading comprehension and risk status for the poor comprehender reading profile. Reading and Writing, 2025, 38: 2633–2659. DOI: 10.1007/s11145-024-10603-x

[13] Gunnerud, H. L., Foldnes, N., & Melby-Lervåg, M. Levels of skills and predictive patterns of reading comprehension in bilingual children with an early age of acquisition. Reading and Writing, 2022, 35: 2365–2387. DOI: 10.1007/s11145-022-10286-2

[14] Babayiğit, S., Hitch, G. J., Kandru-Pothineni, S., et al. Vocabulary limitations undermine bilingual children's reading comprehension despite bilingual cognitive strengths. Reading and Writing, 2022, 35: 1651–1673. DOI: 10.1007/s11145-021-10240-8

[15] Calafato, R., & Hunstadbråten, S. Literature in language education: exploring EFL learners' literary competence profiles. English Teaching & Learning, 2025, 49: 537–559. DOI: 10.1007/s42321-024-00193-w

[16] Shao, J., Abdul Rabu, S., & Chen, C. The impact of gamified interactive e-books incorporating metacognitive reading strategies on Chinese elementary students' mathematical reading comprehension, word problem-solving performance, and general reading motivation. Education and Information Technologies, 2025, 30: 22893–22929. DOI: 10.1007/s10639-025-13660-z

[17] Seifert, S. Is reading comprehension taken for granted? An analysis of Austrian textbooks in fourth and sixth grade. Technology, Knowledge and Learning, 2021, 26: 383–405. DOI: 10.1007/s10758-021-09490-w

[18] Kremmel, B., Indrarathne, B., Kormos, J., & Suzuki, S. Unknown vocabulary density and reading comprehension: replicating Hu and Nation (2000). Language Learning, 2023, 73(4): 1127–1163. DOI: 10.1111/lang.12622

[19] Orhan, A. Online or in-class problem based learning: which one is more effective in enhancing learning outcomes and critical thinking in higher education EFL classroom? Journal of Computer Assisted Learning, 2024, 40(5): 2351–2368. DOI: 10.1111/jcal.13033

[20] Michos, K., Schmitz, M. L., & Petko, D. Teachers' data literacy for learning analytics: a

central predictor for digital data use in upper secondary schools. Education and Information Technologies, 2023, 28: 14453–14471. DOI: 10.1007/s10639-023-11772-y

[21] Li, X., Alazab, M., Li, Q., et al. Question-aware memory network for multi-hop question answering in human–robot interaction. Complex & Intelligent Systems, 2022, 8: 851–861. DOI: 10.1007/s40747-021-00448-0

[22] Liao, J., Zhao, X., Tang, J., et al. To hop or not, that is the question: towards effective multi-hop reasoning over knowledge graphs. World Wide Web, 2021, 24: 1837–1856. DOI: 10.1007/s11280-021-00911-5

[23] Feng, J., Sun, J., Shao, D., et al. Improving the robustness of machine reading comprehension via contrastive learning. Applied Intelligence, 2023, 53: 9103–9114. DOI: 10.1007/s10489-022-03947-w

[24] Huang, G., Long, Y., & Luo, C. Improving multi-hop question answering with prompting explicit and implicit knowledge aligned human reading comprehension. International Journal of Machine Learning and Cybernetics, 2025, 16: 8103–8118. DOI: 10.1007/s13042-025-02712-y

[25] Hu, J., Wu, L., Chen, Y., et al. GraphFlow+: exploiting conversation flow in conversational machine comprehension with graph neural networks. Machine Intelligence Research, 2024, 21: 272–282. DOI: 10.1007/s11633-023-1421-0