Building an Intelligent Annotation and Personalized Learning Platform for Ancient Chinese Poetry by Integrating ERNIE—Stu and Transformer—XL

Liping Lei[1], Xinyao Li[1], Wen Yan[1], Rui Zhang[1], Junyi Xu[2]
[1]College of Teacher Education, Weinan Vocational & Technical College, Weinan 714026, Shaanxi, China
[2]School of Humanities, Weinan Normal University (WNU), Weinan 714099, Shaanxi, China

Corresponding author: Junyi Xu (e-mail: llpchch@163.com)

ABSTRACT Ancient Chinese texts contain important records of cultural heritage, yet existing learning systems still exhibit insufficient semantic coherence, limited word-sense disambiguation, and weak personalized recommendation capability. This paper proposes an intelligent annotation and personalized learning platform for ancient poetry by integrating ERNIE-Stu and Transformer-XL. ERNIE-Stu is used for word embedding, entity recognition, and knowledgegraph-enhanced word-sense disambiguation, while Transformer-XL captures long-range semantic dependencies through segment-level recurrence. Based on syntactic dependency analysis, the platform generates hierarchical context -aware annotations and constructs learner representations from interaction features. Semantic similarity and a timedecay mechanism are further combined to support adaptive recommendation. The semantic fusion framework also provides methodological reference for heterogeneous information understanding and intelligent interpretation in electromagnetic information processing and knowledge-driven communication systems. Experimental results show F1- scores of 0.91 for content-word annotation and 0.87 for function-word annotation, a BERTScore-F1 of 0.798 for longtext discourse similarity, NDCG@5 above 0.70, and CTR above 0.45, confirming improvements in annotation completeness, learning adaptability, and digital cultural-knowledge dissemination.

INDEX TERMS ancient chinese texts, intelligent annotation, semantic modeling, knowledge graphs, electromagnetic information processing

I. INTRODUCTION

THE study and transmission of ancient texts, which are invaluable repositories of textile cultural heritage, face the dual pressures of linguistic complexity and cultural depth. Within these texts, descriptions of fabrics, weaving techniques, dyeing processes, and clothing symbolism are deeply intertwined with cultural implica- tions and linguistic expressions. This complexity directly impacts the learning outcomes for researchers and students in textile history and design, as well as broader cultural dissemination. Modern education demands greater adaptability and personalization of teaching content, while traditional teaching resources often remain at the level of static annotation and standardized content distribution, making it difficult to adapt to the diverse understanding paths and cognitive differences of different learners [1,2]. Intelligent platforms equipped with semantic understanding capabilities and content regulation mechanisms have become a crucial technological foundation for improving the quality and efficiency of ancient poetry and prose teaching [3,4].

Existing annotation systems for ancient poetry are limited by their inability to effectively address polysemy disambiguation (e.g., distinguishing specific textile terms like '绮' and '锦') and fully contextualize allusions related to craftsmanship, leading to fragmented annotations [5,6]. A lack of long-term context capture further hinders accurate identification of semantic extensions across sentences and poetic structures [7,8]. Furthermore, despite the educational value of user data, its low integration into content generation and recommendation results in slow feedback and prevents continuous optimization of learning outcomes [9,10]. Most existing approaches rely on pre-trained language models for basic semantic representation, supplemented by knowledge graphs for structural support, and expanded contextual information modeling capabilities through deep Transformer architectures. Some studies have applied recommendation modules based on behavioral preferences, attempting to factor in individual differences in annotation generation or content ranking [11,12]. However, these approaches have

yet to form a closed-loop structure in terms of model fusion, knowledge enhancement, and user feedback integration, making it difficult to meet the comprehensive requirements of ancient poetry and prose teaching, which require both complete annotation systems and personalized services [13,14].

II. RELATED WORK

The current intelligent understanding of cultural heritage texts, such as ancient Chinese poetry, faces three challenges. This challenge is not unique to literature, but is also seen in the digital analysis of historical archives in fields like textile conservation and history. First, in semantic modeling of ancient texts, while pretrained language models are expressive, they are limited in their integration of structural knowledge [15,16]. In the study of semantic representation, the integration of pre-trained models and structural knowledge is a key direction. Yu et al. [17] fine-tuned bert-base-chinese and SikuBERT in the task of proofreading Li Bai's poems and found that SikuBERT with domain priors has more advantages in maintaining the semantic consistency of ancient poetry texts. This method shows the transfer potential of pre-trained models in ancient text processing tasks. This method emphasizes the importance of lexical semantics to the accuracy of ancient text restoration. On the other hand, the multi-task graph convolutional network proposed by Sun et al. [18] realizes the joint extraction of entities and relations by simultaneously modeling the dependency information of nodes and edges. Current methods still lack higher-order structural coordination and fine-grained semantic alignment mechanisms in multi-layer semantic modeling and knowledge graph injection strategies [19,20]. In long texts, noise filtering and structural understanding have a key impact on the quality of annotation generation. Gan et al. [21] constructed a multi-layer denoising mechanism that integrated PageRank and Transformer, and combined it with the PolyLoss optimization strategy to significantly improve the accuracy of semantic recognition in Chinese long text matching. On this basis, Zhang et al. [22] proposed the cSEN model, which applied unsupervised syntax trees and modern Chinese analyzers for structural enhancement, thereby improving the model's sensitivity to ancient Chinese grammatical features. Existing research still has shortcomings in the dynamic adaptability of syntactic-semantic joint modeling and consistent annotation of crossstyle texts [23,24].

In terms of personalized recommendation, Chen [25] used collaborative filtering algorithms to model readers' preferences, improve the matching degree of ancient poetry recommendations, and verify the impact of user interest mining on enhancing the effect of digital communication. This method emphasizes behavioral data-driven preference modeling. At the same time, Ma et al. [26] designed a teaching platform based on the theory of multiple intelligences and verified its adaptability and promotion potential in literary teaching scenarios through cross-regional experiments. Current research still lacks in-depth integration in the collaborative modeling between individual cognitive characteristics and recommendation mechanisms [27,28]. To address these issues, this paper proposes a multi-scale semantic interaction mechanism driven by dependency syntactic graphs, building a lightweight graph-based neural semantic modeling framework. This approach combines local path attention with global sentence topology fusion to achieve deep semantic restoration and cross-genre structural consistency in classical Chinese. Furthermore, a translation candidate heatmap is applied to construct a semantic consistency visualization mechanism, improving the interpretability and accuracy of the annotation generation system [29,30].

III. CORE MODULE DESIGN AND FUNCTION IMPLEMENTATION

To clearly demonstrate the collaborative structure and functional flow between the system's modules, Figure 1 illustrates the overall architecture of the intelligent annotation and personalized learning platform for ancient poetry. The platform consists of three core layers: semantic modeling and structural enhancement, annotation generation and context modeling, and user modeling and recommendation adjustment. Each layer achieves collaborative collaboration through data transmission and information feedback. The system establishes a behavioral closed-loop within recommendation feedback, supports interest transfer modeling and weight update mechanisms, and improves the pertinence and adaptability of annotation output.

A. ENHANCED SEMANTIC REPRESENTATION OF ANCIENT POETRY

In the ancient poetry semantic representation enhancement module, the system directly pre-trains semantic modeling on the original text based on the ERNIE-Stu model, combining its knowledge enhancement capabilities to achieve refined extraction of entity-level semantics. After the text input is converted into a token sequence by the word segmenter, it is fed into the ERNIE-Stu encoder, where word vectors are generated by applying a cross-attention mechanism that incorporates structured knowledge nodes. The model also applies a triple embedding into the input sequence: position embedding, syntactic dependency embedding, and entity type embedding. This strengthens the ability to identify semantic boundaries between words while maintaining contextual consistency [31,32].

To enhance the structural perception of semantic representation, a structure enhancement module is applied. This module reconstructs the inter-word connection graph G = (V, E) by transforming the topological structure obtained from the dependency syntactic tree. The dependency syntactic tree is generated through the LTP parser, and the parser output is directly converted into dependency edges by mapping head–dependent links into the adjacency format required by the structural module. V represents the set of word nodes

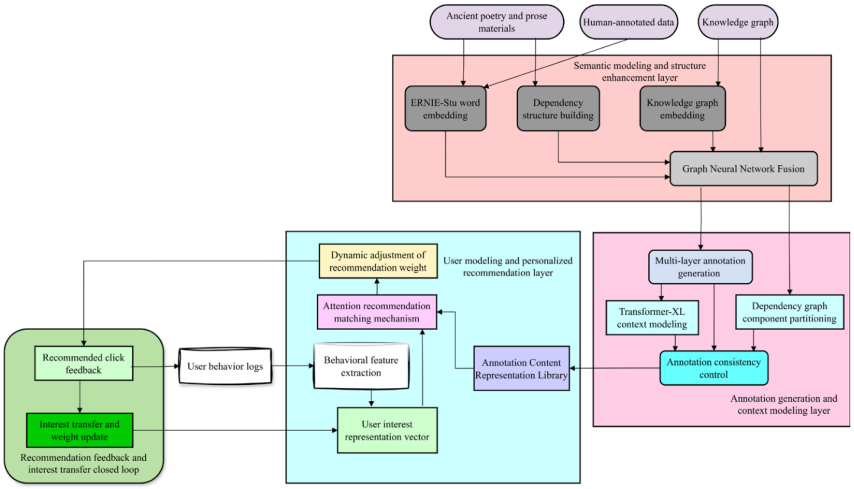

FIGURE 1. System architecture of the ancient poetry and prose intelligent annotation and personalized learning platform

after word segmentation, and E represents the syntactic dependency edges. A graph convolutional network (GCN) is applied to propagate and fuse features based on the graph structure. A two-layer GCN is used in the structural module, and each layer updates node representations through normalized adjacency propagation with a shared activation function. The first layer enlarges the receptive field for syntactic dependency aggregation, and the second layer outputs the fused structural representation that enters the residual fusion stage The propagation process for updating the node representation H' is as follows:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\right) \tag{1}$$

In this formulation, the input matrix H has a dimension of N by d, where N is the number of nodes and d is the feature width, and the updated matrix H' has a dimension of N by d' after the linear mapping of the GCN layer. $\tilde{A} = A + I$ is the adjacency matrix with self-loops added; $\tilde{D}$ is the degree matrix; $W^{(l)}$ represents the weight matrix of the l-th layer GCN; σ is the nonlinear activation function. After fusing semantic embedding and structural graph embedding, the system uses residual connection to perform feature fusion and finally forms a multi-dimensional word representation vector E_i, which is defined as:

$$E_i = \mathrm{LayerNorm}(T_i + S_i + G_i) \tag{2}$$

T_i is the context vector output by Transformer; S_i is the syntactic enhancement vector; G_i is the knowledge graph entity vector.

To further verify the impact of the semantic enhancement mechanism on word meaning differentiation, Figure 2 shows the distribution changes of word vectors for ancient poetry before and after semantic modeling using t-SNE dimensionality reduction. Unenhanced word vectors exhibit fuzzy clustering boundaries within semantic categories and a relatively chaotic distribution. The horizontal and vertical axes in the figure represent the principal component coordinates after dimensionality reduction, which illustrate the relative position of word vectors in two-dimensional space and do not correspond to specific semantic dimensions. The t-SNE projection for Figure 2 is generated with a perplexity of 35, and a fixed random seed of 42 is used to ensure consistent spatial distribution across runs. The visualization is based on a sample of thirty thousand word vectors drawn from the corpus after semantic enhancement and structural integration.

To further quantify the fundamental complexity of ancient poetry in structural modeling, it analyzes the distribution of key structural indicators across different genres. Table 1 presents basic statistical results for Tang poetry, Song lyrics, and Yuan opera on dimensions such as syntactic edge density, average entity connectivity, polysemy frequency, and dependency tree depth. The statistical data in Table 1 is derived from a corpus consisting of 18,420 Tang poems, 9,365 Song lyrics, and 1,127 Yuan opera texts, and each subset is processed by removing entries with missing sentence boundaries or ambiguous structural markers to maintain consistency in dependency parsing and entity extraction. This serves to support the rationality of input for the structural enhancement strategy and the design of compositional parameters.

As shown in Table 1, Yuan opera has higher syntactic edge density and dependency tree depth than Tang poetry and Song ci, indicating a more complex structure. The average entity connectivity and polysemy frequency also show an upward trend, reflecting the increased semantic ambiguity brought about by increased colloquialism and narrative. These differences reflect the uneven distribution of structural complexity and semantic ambiguity across different literary genres, providing a foundation for parameter setting for structure enhancement mechanisms and designing word sense disambiguation strategies.

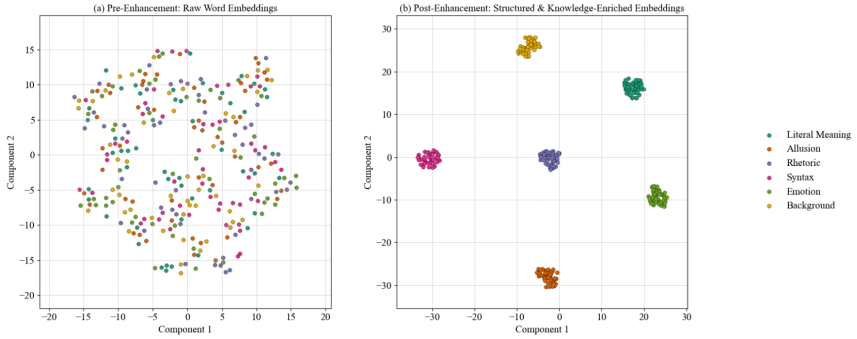

FIGURE 2. t-SNE dimensionality reduction distribution of ancient poetry word vectors before and after semantic embedding: (a) Unenhanced: word vector distribution generated using only ERNIE-Stu; (b) Enhanced: word vector distribution after integrating structural information and knowledge graph

TABLE 1. Statistical index table of ancient poetry and prose structure information

Genre	Syntactic Edge Density (Edges/Words)	Average Entity Connectivity (Edges/Entity)	Polysemy Frequency (%)	Average Dependency Tree Depth
Tang Poetry	0.78	3.2	15.6	4.1
Song Lyrics	0.81	3.7	18.4	4.6
Yuan Opera	0.85	4.1	21.3	5.2
Overall Average	0.81	3.67	18.43	4.63

B. ANNOTATION CONTENT GENERATION MECHANISM

In the annotation content generation mechanism, the system input is a sequence of word vectors processed by the semantic enhancement module $\{E_i\}_{i=1}^{n}$, where each vector has integrated entity semantics and structural information. To achieve high-precision annotation parsing based on the language characteristics of ancient poetry, the system adopts a joint modeling strategy of word sense disambiguation and syntactic dependency, combining contextual information to achieve structured generation of multi-layer annotations. The word sense disambiguation process constructs a semantic candidate set based on the knowledge graph $S_i = \{s_{i1}, s_{i2}, ..., s_{im}\}$, and uses the cosine similarity between the vector weighted center within the context window C_i and the candidate semantic center vector v_{ij} to select the optimal semantic [33,34]. The calculation formula is as follows:

$$\hat{s}_i = \underset{s_{ij} \in S_i}{\operatorname{argmax}} \frac{C_i \cdot v_{ij}}{\|C_i\| \|v_{ij}\|} \tag{3}$$

In formula (3), $\hat{s}_i$ is the final semantic label of the i-th word. To improve semantic consistency, the calculation of the window center vector C_i applies a contextual attention mechanism, which assigns a dynamic weight w_{ij} to each context word vector E_j, defined as follows:

$$C_i = \sum_{j \in N(i)} w_{ij} \cdot E_j, \quad w_{ij} = \frac{\exp(E_i \cdot E_j)}{\sum_{k \in N(i)} \exp(E_i \cdot E_k)} \tag{4}$$

$N(i)$ represents the context word set centered on i.

The structural relationship identification stage employs a deep bidirectional dependency parser to generate a structural dependency graph G=(V,E). Node representations are initialized by the upstream module's word vectors, and edge weights are assigned based on the dependency probability matrix. The system then uses relationship clustering to decompose the grammatical structure into three functional units: the interpretation, background, and rhetoric layers. Traversing the directed subgraph formed by nodes in each unit yields the candidate sequence of annotation components for that layer.

This dependency structure is vital, serving both to identify word disambiguation points and as the foundation for hierarchical annotation. The dependency graph integrates the path relationships between all three component types, reflecting their multi-layered, cross-nested annotation logic, as illustrated in Figure 3. Specifically, the interpretation layer focuses on the semantic backbone; the background layer expands context; the rhetorical layer connects components to boost expressiveness, ensuring the annotation content is both structurally organized and logically supported.

Annotation content generation is accomplished by a multi-channel fusion network, which takes as input a sequence of structural units and their contextual feature tensors and outputs three types of annotation vectors: word meaning annotation, rhetorical markers, and contextual background. The system is designed to jointly minimize the semantic redundancy between the three types of annotations and achieves consistency constraints for multi-annotation fusion using the following loss function:

$$L_{\text{fusion}} = \sum_{i=1}^{n} \sum_{(a,b)} \lambda_{ab} \cdot \left(1 - \frac{A_i^a \cdot A_i^b}{\|A_i^a\| \|A_i^b\|}\right) \tag{5}$$

A_i^a and A_i^b represent the representation vector of the i-th word in annotation dimensions a and b, respectively, and λ_{ab} is the weight coefficient of the consistency regularization

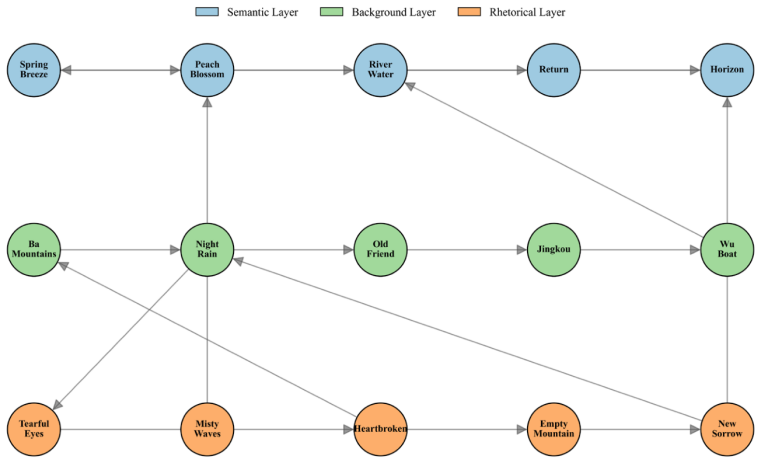

FIGURE 3. Multi-layer annotation component dependency graph based on dependency structure

term between dimensions. In the model implementation, λ_{ab} is assigned through a fixed-value strategy determined in advance, and its value is selected by evaluating validation performance under multiple preset candidates during initial calibration, after which the chosen value remains unchanged throughout training. To quantify the dataset's coverage across six annotation categories and three stylistic dimensions, it constructs a label distribution statistics table to characterize the annotation's extensibility at different structural levels. Table 2 shows the average annotation length, dependency depth, semantic similarity, and node overlap rate for the six core annotation layers in the training corpus. The statistical results show that the allusion and context layers have the highest average annotation length and dependency depth, indicating a broader semantic scope and more complex structural development. The rhetorical and syntactic layers have relatively low semantic similarity, resulting in more discrete annotation content, while the sentiment and word meaning layers exhibit higher semantic concentration. The node overlap rate is higher in the word meaning and syntactic structure layers, reflecting the more frequent reuse of lexical nodes in cross-layer semantics and stronger boundary coupling in these two annotation types. Table 2 reveals the differentiated characteristics of multi-layer annotations in terms of structural complexity and semantic coupling, providing structural support for the representation layering mechanism and consistency control strategy in the annotation fusion module.

C. LONG TEXT CONTEXT MODELING MODULE

In the long-text context modeling module, the system takes a sequence of semantically annotated vectors as input and uses the Transformer-XL architecture to implement semantic dependency modeling across sentence and paragraph levels. The original annotation representation is $\{E_i\}_{i=1}^{n}$ organized into input blocks in paragraph order, and the Transformer-XL inter-segment state memory mechanism is used to persistently encode the historical context of each segment [35,36]. The system applies a position encoding recalculation strategy, dynamically offsetting the original position index to support cross-segment semantic propagation, while retaining the historical state vector M_{t-1} as the global prior for the current paragraph. Its cross-segment information transfer mechanism is expressed as follows:

$$h_t^{(l)} = \mathrm{MultiHead}\left(q_t^{(l)}, [h_{t-1}^{(l)}, M_{t-1}^{(l)}], V\right) \tag{6}$$

$q_t^{(l)}$ represents the query vector of the l-th layer of the current segment; $h_{t-1}^{(l)}$ is the output of the previous segment; $M_{t-1}^{(l)}$ is the historical memory state; V is the set of attention value vectors. To ensure stable convergence of long-distance dependencies during context modeling, a segment-level residual memory fusion mechanism is applied to gate the output of each layer to form an updated state vector Z_t. The gating mechanism is defined by the following formula:

$$Z_t = \sigma(W_z H_t + b_z) \odot M_{t-1} + (1 - \sigma(W_z H_t + b_z)) \odot H_t \tag{7}$$

H_t is the output concatenation matrix of all positions in the current layer; $\odot$ represents element-wise multiplication; σ is the sigmoid function; W_z is the gating parameter.

The system uses a context-focused mechanism to project long text output into local context vectors and global annotation representations. When generating annotation content, a semantic alignment loss function is applied to constrain the consistency of each word annotation within its context. The loss function is defined as follows:

$$L_{\mathrm{align}} = \sum_{i=1}^{n} \left(1 - \frac{A_i \cdot C_i}{\|A_i\| \|C_i\|}\right) \tag{8}$$

A_i represents the target vector output by the annotation generation module, and C_i is the center of the context vector output by the context modeling. The cosine similarity between the two is used as a measure of semantic consistency.

TABLE 2. Multi-layer annotation structure distribution and semantic consistency statistics table

Annotation Level	Average Annotation Length (Words)	Average Dependency Depth (Layers)	Internal Semantic Consistency (COS Similarity)	Internal Node Overlap Rate (%)
Meaning Level	4.9	2.3	0.812	11.6
Syntactic Structure Level	3.7	3.1	0.789	14.2
Emotion Level	5.6	3.8	0.841	9.5
Context Level	7.2	4	0.806	7.3
Rhetoric Level	6.3	3.5	0.772	6.8
Allusion Level	8.5	4.2	0.794	5.4

D. MATCHING USER MODELING WITH LEARNING CONTENT

In the user modeling and learning content matching module, the system constructs a multi-dimensional user feature vector based on interaction behavior logs to capture learning preferences and knowledge mastery, supporting dynamic matching of personalized annotation content. The raw input consists of the user's click sequence, page dwell time, annotation expansion depth, and historical task completion. The behavior data used in the recommendation module is obtained from a preliminary deployment phase during which the system was released to a controlled group of invited users, and the interaction traces accumulated in this phase are stored as the behavior log for model training. A simulation-based augmentation procedure is applied to enlarge the initial set by generating interaction samples that follow the same distributional properties observed in the collected logs, ensuring that the recommendation module receives stable supervision during the optimization stage. This is encoded into a behavior representation matrix via an embedding layer $B \in \mathbb{R}^{m \times d}$, where m is the number of behavior types, and d is the feature dimension. This matrix is used to extract sequential patterns through a bidirectional gated residual network, outputting a unified user interest vector $u \in \mathbb{R}^d$, which serves as the key discriminant vector for subsequent annotation screening.

The system applies a content representation library $C = \{c_1, c_2, ..., c_n\}$, each of which c_i represents the semantic and structural features of an annotation content, which is jointly constructed by the upstream semantic enhancement and context modeling modules. The matching process calculates the relevance score between the user vector and each annotation item through the attention mechanism:

$$s_i = \frac{u \cdot c_i}{\|u\| \cdot \|c_i\|} \tag{9}$$

s_i represents the cosine similarity between the user and the i-th annotation. The system normalizes and sorts the similarity scores and selects the top k high-scoring items to form a candidate set $\mathcal{S}_u$, which is the optimal annotation matching result set for the user under the current learning task.

To model the long-term migration trends of users' interests, the system applies a memory update mechanism after each learning cycle. Let u_t be the updated user vector in the current cycle and u_{t-1} be the persistent memory of the previous cycle. The system uses an exponential smoothing strategy to model preference migration. The update formula is as follows:

$$u_t = \lambda \cdot u_{t-1} + (1 - \lambda) \cdot u'_t \tag{10}$$

$\lambda \in [0, 1]$ represents the historical interest retention coefficient, and u'_t is the interest vector extracted from the current period behavior.

E. PERSONALIZED RECOMMENDATION AND DYNAMIC ADJUSTMENT STRATEGY

The personalized recommendation and dynamic adjustment strategy module builds a dynamic recommendation distribution function based on user behavior vectors and semantic representations of annotation content. This function controls the presentation order and weight update cadence of learning content, achieving a bidirectional match between content and cognitive needs. The recommendation candidate set consists of the matching set $\mathcal{S}_u$ output in Section Personalized Recommendation and Dynamic Adjustment Strategy. Based on the user's current task status and behavioral feedback signals, the system assigns a weight vector $\omega \in \mathbb{R}^k$ to the content in this set. A time decay function is used to account for the dynamic changes in behavioral value, constructing a recommendation strength function as follows:

$$\omega_i^{(t)} = \frac{e^{r_i^{(t)} - \tau \cdot \Delta t_i}}{\sum_{j=1}^{k} e^{r_j^{(t)} - \tau \cdot \Delta t_j}} \tag{11}$$

$r_i^{(t)}$ is the user's interaction score for the i-th comment at time step t; Δt_i is the time interval between the comment and the current one; τ is the time sensitivity coefficient, which is used to adjust the dominance of recent feedback in the recommendation.

The recommendation ranking process is driven by the fusion scoring function $g(i, t)$, which integrates two factors: interest relevance and content difficulty gradient, and is defined as follows:

$$g(i, t) = \alpha \cdot s_i + (1 - \alpha) \cdot (1 - \delta_i) \tag{12}$$

s_i is the cosine similarity between the user preference vector and the annotation semantic vector; The user preference vector is generated through an embedding aggregation process in which the long-term interaction traces in the behavior log are mapped into the semantic space

through the text encoder, and the resulting representations are fused by weighted pooling to form a stable preference representation for the subsequent similarity computation. δ_i is the estimated difficulty value of annotation item i; α is a weighting factor that controls the balance between interest satisfaction and cognitive challenge in the recommended content. Difficulty assessment constructs a scoring function based on text length, syntactic depth, and dependency density, and is normalized. The difficulty value δ_i is obtained through an automatic estimation process that maps each annotation item into a unified scoring scale. The scoring function is defined as $F_i = \alpha_1 l_i + \beta_1 d_i + \gamma_1 c_i$, where l_i denotes the text length of the annotation, d_i denotes the syntactic depth derived from the parsed dependency tree, and c_i denotes the dependency relation density extracted from the same structure. The parameters α_1, β_1, and γ_1 are fixed coefficients determined during the preliminary calibration stage. The normalized difficulty value is given by $\delta_i = (F_i - F_{\min})/(F_{\max} - F_{\min})$, where $F_{\min}$ and $F_{\max}$ denote the minimum and maximum F_i across all annotation items. The recommended list is finally sorted in descending order by $g(i,t)$, and the top L items are selected for display.

The dynamic adjustment mechanism in the recommendation process applies the memory decay of the weight update trajectory and the nonlinear response mapping of the user feedback signal to construct the cumulative weight update amount Ω_t, which is defined by the following formula:

$$\Omega_t = \gamma \cdot \Omega_{t-1} + \sum_{i=1}^{k} \omega_i^{(t)} \cdot f_i^{(t)} \tag{13}$$

$f_i^{(t)}$ is the feedback activation value of annotation content i at time t, and γ is the historical memory retention factor. The system Ω_t adjusts the frequency of subsequent recommendation updates based on the changing trend, achieving adaptive control of the content switching rhythm to the learning state.

IV. EXPERIMENTAL SETUP AND EVALUATION CONFIGURATION

A. ENVIRONMENT AND PLATFORM CONFIGURATION

The experiment is deployed on a local server with graphics computing capabilities, using a dual-core Intel Xeon Silver processor and four NVIDIA RTX A6000 graphics cards, each with 48GB of video memory. This supports large-scale parallel training of Transformer structures, and the system environment is Ubuntu 20.04 LTS. The platform's operating environment is unified with Python 3.9, and the deep learning framework selected is PaddlePaddle 2.5, due to its native compatibility with the ERNIE-Stu model and its excellent dynamic graph execution efficiency. The model relies on the pre-trained language model ERNIE-Stu-Base, which integrates knowledge graph entity annotation and structural attention mechanisms, and has the dual capabilities of entity recognition and semantic embedding. The long text modeling component applies the Transformer-XL structure, and the core dependency module is re-implemented based on a customized Segment-Level Memory mechanism to enhance cross-segment state memory capabilities.

To support the system's online learning behavior modeling and real-time recommendation functions, a Redis cache system is deployed on the server side to achieve rapid updates of user state vectors and intermediate storage of content matching results, enabling dynamic adjustment of recommendation paths in personalized learning tasks. All explicit parameters and implicit state variables during training and inference are centrally managed in PaddleNLP's parameter service module, supporting breakpoint recovery and multi-card weight synchronization to ensure experimental repeatability and result consistency under large-scale multi-round training. The overall platform architecture is implemented in the form of module decoupling.

B. DATASET CONSTRUCTION AND PREPROCESSING PROCESS

Experimental data is sourced from publicly available sources such as the Complete Tang Poems, Three Hundred Song Poems, and Selected Yuan Operas. Texts are extracted from both the National Language Resource Monitoring Corpus and the CNKI Ancient Chinese Text Database, ensuring comprehensive coverage of diverse literary genres, structural features, and linguistic styles. The curated dataset consists of 18,420 Tang poems, 9,365 Song lyrics, and 1,127 Yuan opera texts. A total of 162,480 sentences are segmented from these materials, producing 214,306 annotation items across all categories. The dataset is divided into a training set, a validation set, and a test set following a ratio of 8 to 1 to 1. The distribution of annotation categories remains consistent across the three subsets through stratified sampling, and the proportions of allusion, rhetorical, structural, and imagery-related annotations match their global frequencies in the full dataset. To evaluate the dataset's coverage across six annotation categories and three genres, a statistical analysis characterized the extensibility of annotations using metrics like sentence coverage, overlap, and cross-tag density. The structural distribution statistics across genres and annotation categories are summarized in Table 3. The results show that word meaning tags are core to the annotation structure, with over 75% coverage across all genres. Syntactic tags are also highly universal, maintaining a stable coverage above 64%. Conversely, background tags exhibit comparable sentence coverage in Tang poetry and Song Ci, while their coverage decreases in Yuan opera. Allusion tags show a progressive increase in sentence coverage from Tang poetry to Song Ci and further to Yuan opera, reflecting genre-dependent differences in narrative expansion and intertextual reference density. Specifically, background annotation coverage reaches 48.5% in Tang poetry and 51.2% in Song Ci before declining to 45.6% in Yuan opera, whereas allusion annotation coverage increases from 25.1% in Tang poetry to 28.5% in Song Ci and 30.7% in Yuan opera, forming a clear

cross-genre gradient. Rhetoric tags are sparsely distributed, consistently below 20% coverage and registering the lowest overlap and cross-density. Crucially, the cross-label ratio is highest for the Word Meaning and Syntactic categories (e.g., Yuan Qu word meaning reaching 1.9), indicating that the co-occurrence of multiple labels is concentrated in these fundamental dimensions. This distribution pattern provides a corpus-level basis for modeling genre-sensitive annotation density and for designing feature coupling and representation fusion strategies aligned with structural and intertextual variation.

The text segmentation process is uniformly completed using a combination of the Jieba word segmenter and a custom classical Chinese dictionary. Proprietary entries are added to the word segmentation vocabulary to improve the accuracy of structural recognition. The dependency relations in the preprocessing stage are produced with LTP and converted into a unified adjacency representation shared with the structural module, which serves as the input for the structure enhancement module. Entity recognition uses a dual-path superposition of rules and models. Initial entities are first extracted based on vocabulary matching; then, the ERNIE-Stu named entity recognition function is called to complete unmatched segments, and entity IDs and labels are uniformly output. The contextual segmentation strategy is based on dynamic division based on semantic turning points and punctuation positions, adapting to the Transformer-XL inter-segment memory mechanism's sensitivity to the flow of the text. The maximum sentence length is set to 240 characters for different corpora, and the annotation anchor points are retained for target alignment training.

User interaction data used in the recommendation and behavior modeling experiments were collected during the system's internal pilot stage. The platform ran a controlled pre-release test for a continuous three-month period, during which 312 participants were recruited, of whom 300 generated complete and valid click sequences, browsing traces, annotation expansion records, and task completion histories. The interaction logs produced by these 300 users formed the source of the historical behavioral data used in the CTR analysis, NDCG evaluation, and adaptation experiments for different user groups.

C. PARAMETER SETTING AND TRAINING PROCESS

The training process utilizes a multi-stage distributed scheduling mechanism, performing hierarchical optimization for the semantic enhancement, structural modeling, and contextual understanding submodules. The pre-trained language model uses a frozen entity encoder strategy, updating weights only in the structural fusion and contextual propagation layers to avoid disrupting the stability of the pre-trained parameter distribution. The entity encoder is based on the RoBERTa-large Chinese model released through public repositories, and its parameters are fixed during training to maintain stable semantic representations. The AdamW optimizer is selected, with an initial learning rate of 3e-5 and a weight decay coefficient of 0.01. A linear warmup and gradual annealing strategy is employed, with the warmup step number set to 10% of the total training steps. Each round of training combines the semantic alignment loss, annotation consistency loss, and recommendation reranking loss to minimize the three objectives, with a weight coefficient ratio of 4:3:3 to ensure gradient coordination and stable model convergence during the multi-task optimization process. The training batch size is 32; the maximum sequence length is 256; the number of gradient accumulation steps is set to 2 to support effective modeling of long texts under memory constraints. The Transformer-XL inter-segment state cache length is set to 128, and the maximum backtracking depth of the historical state window is 4, matching the average sentence length distribution characteristics of the corpus to support con- tinuous semantic propagation across chapters. A full sample iteration is performed in each cycle, with an upper limit of 50 rounds. The early stopping condition is that there is no significant improvement in the performance indicators of the validation set for 5 consecutive rounds. The evaluation indicators cover F1-score, BERTScore-F1, NDCG@5, and content matching accuracy, and independent monitoring is performed in the word meaning annotation, semantic consistency, recommendation ranking, and user behavior mapping modules, respectively.

V. EXPERIMENTAL RESULTS ANALYSIS

A. COMPARATIVE ANALYSIS OF WORD SEMANTIC ANNOTATION PERFORMANCE

To enhance contextual understanding and structural adaptation in the task of annotating the meaning of ancient Chinese poetry, this experiment compared the performance of five models across different word meaning categories. Model 1, BERT, uses a bidirectional Transformer for basic semantic encoding but has limitations in handling long-range dependencies. Model 2, ERNIE, incorporates a knowledge augmentation strategy to optimize word meaning recognition accuracy. Model 3, BiLSTM-CRF, relies on sequence labeling and condi- tional random fields to model local dependencies. Model 4, ERNIE-Stu, combines entity recognition with semantic modeling to enhance word sense disambiguation. Model 5, the proposed ERNIE-Stu and Transformer-XL fusion model, combines structural augmentation with global context awareness. In task design, the semantic boundaries between content word meaning (S1) and function word meaning (S2) are clear, focusing on lexical recognition accuracy. Allusion meaning (S3) and background meaning (S5) rely on deep semantic links and knowledge expansion. Rhetorical meaning (S4) involves implicit semantic expression and precise analysis of syntactic structure. Syntactic function word meaning (S6) places higher demands on modeling contextual dependencies. Figure 4 illustrates the adaptability of each model for these different types of word meaning annotation tasks.

In an accuracy analysis, the proposed ERNIE-Stu and Transformer-XL fusion model achieved scores of 0.88 and

TABLE 3. Structural distribution statistics of six types of annotation tags under three types of writing styles

Literary genres	Tag Type	Sentence coverage (%)	Average annotation length (words)	Annotation overlap (%)	Cross-label ratio (per sentence)
Tang poetry	Word Meaning	83.2	4.6	12.7	1.4
Tang poetry	Background	48.5	7.3	6.2	0.8
Tang poetry	Rhetoric	13.6	6.1	3.4	0.3
Tang poetry	Allusion	25.1	8.2	9.1	0.6
Tang poetry	Sentence	32.7	5.8	4.9	0.5
Tang poetry	Syntax	64.4	3.7	10.6	1
Song ci	Word Meaning	79.5	5.1	15.3	1.6
Song ci	Background	51.2	8.1	7.4	1
Song ci	Rhetoric	16.9	6.7	3.7	0.4
Song ci	Allusion	28.5	9	10.5	0.7
Song ci	Sentence	41.3	6.5	5.8	0.6
Song ci	Syntax	69.8	4	11.2	1.2
Yuan opera	Word Meaning	76.1	5.4	18.1	1.9
Yuan opera	Background	45.6	7.8	7.2	0.9
Yuan opera	Rhetoric	18.3	7.1	4.5	0.4
Yuan opera	Allusion	30.7	9.3	11.3	0.8
Yuan opera	Sentence	39.2	6.8	6.7	0.7
Yuan opera	Syntax	72.5	4.3	12.1	1.4

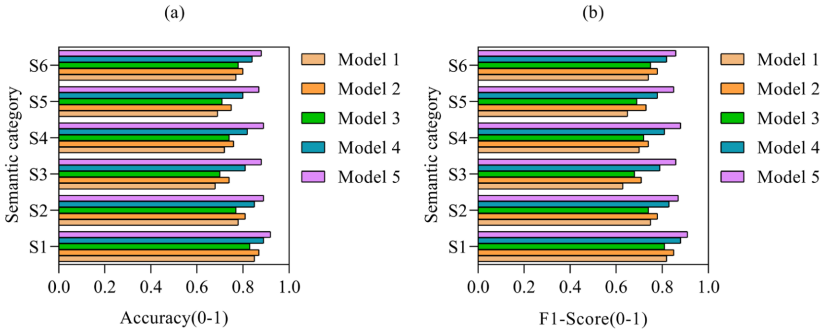

FIGURE 4. Performance comparison of different models in each semantic category annotation task: (a) Accuracy; (b) F1-score

0.89 for allusive meaning and rhetorical meaning, respectively. This performance difference was significantly superior to the ERNIE model's scores of 0.74 and 0.76 in the corresponding categories. This performance difference stems from the effective capture of complex structural information by long-range semantic dependencies and contextual enhancement strategies. Model 5 reaches accuracy values of 0.87 for background meaning and 0.88 for syntactic functional meaning, significantly higher than the BiLSTM-CRF model's scores of 0.71 and 0.78 in similar tasks, demonstrating the importance of structural enhancement and dependency modeling in resolving word semantic ambiguity. Model 5 achieves F1-scores of 0.91 for content word meaning and 0.87 for function word meaning, demonstrating that the model maintained a stable balance between precision and recall in basic semantic recognition and local dependency processing. Model 5 yields F1-scores of 0.86 for allusive meaning and 0.85 for background meaning, demonstrating the combined benefits of the knowledge graph and context fusion mechanism in disambiguation and context adaptation. Comprehensive analysis shows that this method demonstrates excellent adaptability and generalization across various word meaning annotation tasks, validating the effectiveness of the semantic enhancement and context dynamic modeling strategies.

B. DISCOURSE-LEVEL SEMANTIC COHERENCE ASSESSMENT

To evaluate semantic matching and user consistency in ancient poetry recommendation, two comparative paths were established. The first path included general language models: Model 1 (RoBERTa-Base), Model 2 (BERT-Base), Model 3 (GPT-2), and Model 4 (XLNet-Base). The second path featured fusion optimization models: Model 5 (ERNIE 3.0 + Transformer-XL), focusing on knowledge and long text; Model 6 (ERNIE-Stu + Transformer), using student model compression; Model 7 (the proposed ERNIE-Stu + Transformer-XL), which integrates deep semantic understanding, positional awareness, and contextual expansion, with specific structural and parameter improvements for ancient Chinese poetry. Figure 5 presents the BERTScore-F1 trend and manual consistency scores across different text lengths for these models. A panel of three annotators with academic backgrounds in classical Chinese conducted the manual scoring. Each annotator independently evaluated the semantic coherence of the generated content. Inter-annotator agreement was assessed using Cohen's Kappa,

which reached 0.82, indicating a high level of scoring consistency. A double cross-annotation procedure was applied to ensure reliability, with disagreements resolved through a subsequent consensus check.

As text length increased from 0–30 words to over 210 words, Model 7 maintained a consistently high score across the entire range, reaching 0.798 in the 210+ word range, a significant improvement over Model 2 (0.720) and Model 4 (0.720). This demonstrated that the integrated long-text modeling mechanism effectively mitigated the semantic degradation caused by context dilution. Models 5 and 6 showed a slight decrease in scores as the text length entered the medium-long range and did not maintain a stable advantage. This is speculated to be related to the loss of semantic granularity caused by their structural depth or compression strategies. In contrast, the curves of Models 1 and 3 fluctuated significantly, suggesting a lack of effective structural perception of semantic boundaries in long texts. In manual consistency scoring, Model 7 showed a concentrated distribution in the manual consistency scores, with the score range remaining stable around 4.8, significantly higher than the central tendency range of Models 1 and 3, demonstrating low dispersion and high confidence. This demonstrates that the proposed approach exhibits stronger consistency in semantic expression and user cognitive fit, validating its comprehensive adaptability in personalized learning recommendation scenarios.

C. USER CONTENT MATCHING EVALUATION

To verify the recommendation strategy's adaptability, a joint evaluation using CTR and NDCG@5 was conducted, comparing five semantic modeling and ranking strategies: Strategy 1 (TF-IDF, sparse retrieval); Strategy 2 (BERT-base, contextual); Strategy 3 (RoBERTa, improved generalization); Strategy 4 (ERNIE-Stu, knowledge embedding); Strategy 5 (the proposed ERNIE-Stu + Transformer-XL, combining knowledge and long-term dependency modeling). Annotation content was categorized into five types: Word Meaning (A1), Historical Context (A2), Rhetorical Techniques (A3), Allusion Source (A4), and Emotional Intent (A5). To measure adaptation across different user levels and multi-dimensional annotations, a CTR-NDCG two-dimensional scater plot was constructed, with user groups clustered and colored to highlight local distribution and overall performance differences. The graph presents behavioral data from 300 users. User groups are categorized based on three metrics: cumulative learning time on the platform, annotation depth, and task completion rate. After normalization, a unified proficiency index is generated. Indices below 0.4 are classified as the basic group, between 0.4 and 0.7 as the intermediate group, and above 0.7 as the advanced group. The three user groups are represented in the scater plot with different dot sizes, demonstrating their compatibility with CTR and NDCG levels on the same coordinate system. This is shown in Figure 6:

In this figure, the proposed strategy, ERNIE-Stu + Transformer-XL, exhibited high density clustering in the upper right region across all annotation dimensions and user groups. NDCG@5 generally exceeded 0.70, and CTR remained consistently above 0.45, outperforming all compared strategies. RoBERTa and ERNIE-Stu formed the second tier, showing some adaptability for intermediate and advanced users in the two annotation dimensions of rhetoric and allusion source. However, their overall performance weakens in the dimension of sentiment analysis, with a significant decrease in CTR, indicating model limitations in their ability to capture sentiment. The distribution presents a near-linear relationship between CTR and NDCG. This pattern arises from the controlled evaluation setting, where annotation-driven relevance signals influence both ranking order and click responses. The uniform task structure reduces stochastic behavior in user clicks and produces a coupled trend between ranking quality and interaction outcomes, forming a monotonic distribution rather than a dispersed pattern common in open-domain recommendation environments.

D. ANALYSIS OF THE HIERARCHICAL INTEGRITY OF ANNOTATION INFORMATION STRUCTURE

To test the models' ability to capture the structural hierarchy of ancient poetry annotations across different genres, five models were evaluated using three distinct classical corpora: Tang poetry (metrical, image-focused), Song ci (emotional, diverse grammar), and Yuan dynasty opera (structured prose, colloquial). The five models tested represent progressive levels of complexity: Model 1 (ERNIE-Stu, shallow semantic), Model 2 (ERNIE-Stu + Knowledge Graph), Model 3 (Transformer-XL, long-context only), Model 4 (ERNIE-Stu + Transformer-XL, joint semantic-context), and Model 5 (complete system with user modeling). Figure 7's radar chart visualizes each model's coverage across six annotation dimensions for these three corpora, allowing a structural integrity comparison.

Model 5 demonstrated superior structural integrity across all three corpora. Its coverage of word meaning annotations in the Tang poetry corpus reached 0.88, and its coverage of background knowledge reached 0.72, significantly exceeding the other models. This is due to its ability to contextually couple historical entities with situational semantics. Model 4 demonstrated a balanced performance across rhetoric and grammatical structures, with a rhetorical coverage of 0.61 and a sentiment coverage of 0.67 in Song ci, demonstrating its well-suited structure for genres rich in emotional expression. Model 3 maintained grammatical structural stability across all three corpus subsets, reaching 0.64 in the Yuan opera corpus, demonstrating its generalizability to non-standard spoken language patterns. Model 1 exhibited deficiencies in the background knowledge dimension across all subsets, with coverage never exceeding 0.40, reflecting its limited ability to model external knowledge. Model 2 showed improvement in the knowledge-related dimension, but its syntactic and rhetorical dimensions remained middling. Overall, the fusion model with knowledge en-

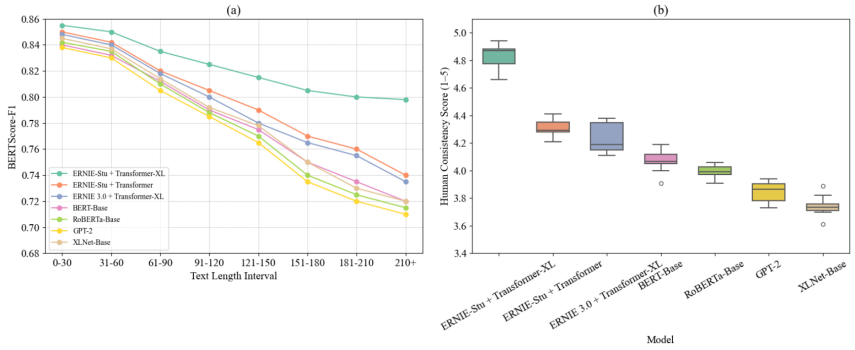

FIGURE 5. Performance comparison of the fusion model and the general model under different evaluation indicators: (a) BERTScore-F1 trend under different text lengths; (b) Manual consistency evaluation

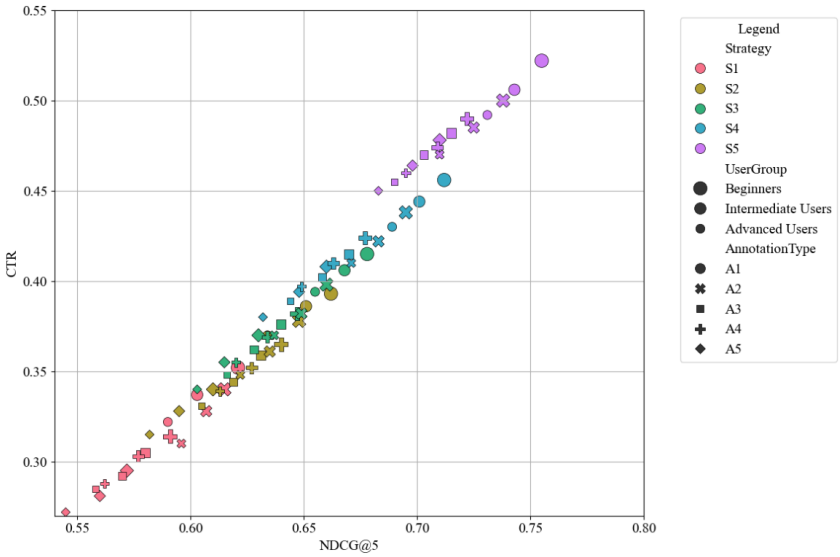

FIGURE 6. CTR-NDCG distribution of different recommendation strategies under five types of annotation content and three types of user groups

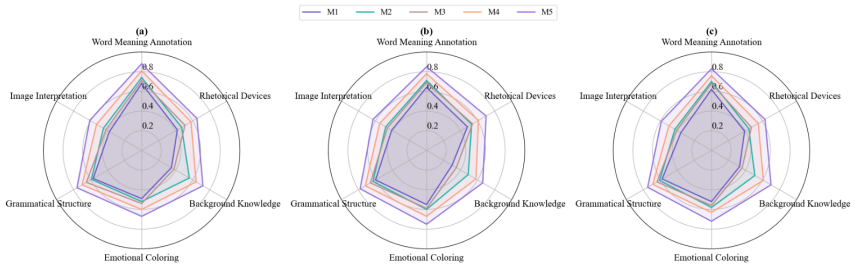

FIGURE 7. Radar chart comparing the coverage of six-dimensional annotation structures of different models in Tang poetry, Song lyrics and Yuan opera corpora: (a) Anthology of Tang poetry; (b) Anthology of Song poetry, (c) Anthology of Yuan opera

hancement and context modeling capabilities showed better dimension coverage and corpus adaptability in the task of structured annotation of ancient poetry and prose.

E. VERIFICATION OF DYNAMIC ADAPTABILITY OF RECOMMENDATION SYSTEM

To test the dynamic adaptability of learning paths, an experiment compared user access behavior under fixed and dynamic recommendation strategies across eight hierarchical annotation tasks. The tasks range from word meaning interpretation (T1), contextual disambiguation (T2), syntactic analysis (T3), historical reference interpretation (T4), rhetorical style recognition (T5), imagery and emotional analysis (T6), thematic extraction (T7), to extended appreciation (T8). Their difficulty is defined by linguistic depth, semantic abstraction, and required cultural knowledge, forming a stable progression from T1 to T8. Six user groups were defined based on learning goals (G1: beginners, G2: background interest, G3: exam-oriented, G4: art appreciation, G5: knowledge expansion, G6: highly engaged), showing varied task demands. The grouping was derived from the

behavioral statistics collected during the pre-experimental stage, using users' task completion tendencies, annotation depth preferences, and browsing stability as the core indicators. The six groups contained 128 users in total, with 26 in G1, 21 in G2, 18 in G3, 22 in G4, 17 in G5, and 24 in G6. The division followed a consistent threshold for each indicator to ensure that users with similar learning intentions and interaction patterns were assigned to the same group. The static strategy used a linear task order, while the dynamic strategy adjusted the sequence in real time according to user profiles and behavioral feedback to improve path fit. Figure 8 presents the heatmap of access frequency distribution for these groups under both strategies.

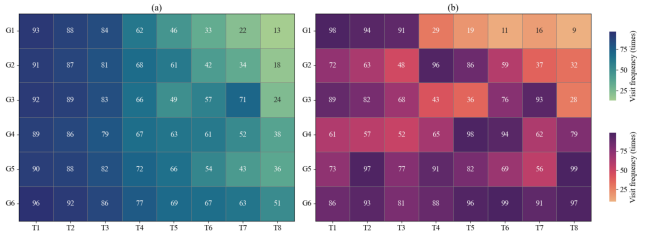

FIGURE 8. Adaptability distribution of task paths in the personalized learning platform for ancient poetry and prose: (a) Static recommendation strategy; (b) Dynamic adjustment strategy

Under the static recommendation path, user access heavily concentrated in the first three tasks (e.g., beginners at 93 and 88 visits), with engagement decreasing monotonically, failing to cover specialized interests (background/art tasks remained just above 60 visits). Conversely, the dynamic recommendation strategy showed significantly personalized path distribution. Beginners maintained high focus (up to 98 visits in initial tasks), but specialized users showed high engagement in relevant advanced tasks: background users increased visits to allusion/rhetoric tasks (up to 96 and 86 times), and art users focused on rhetoric/artistic conception (up to 98 and 94 times). Highly active users showed high engagement across all paths (up to 97 times). This demonstrates that the dynamic strategy effectively breaks linear constraints, significantly enhancing path adaptability and aligning content matching with personalized learning needs.

VI. CONCLUSION

This paper develops an intelligent annotation and personalized learning platform, integrating ERNIE-Stu and Transformer-XL, specifically to extract textile heritage content from ancient Chinese poetry. The system achieves closed-loop optimization of semantic modeling, annotation, and recommendation by utilizing ERNIE-Stu for knowledge graph-enhanced disambiguation and Transformer-XL for long-text context modeling, with content dynamically adjusted via user interest vectors. The model demonstrates high performance: annotation accuracy reaches 0.88 and 0.89 (allusive/rhetorical meaning), surpassing the ERNIE baseline. Discourse coherence achieves a BERTScore-F1 of 0.798 for long texts. Recommendation metrics are robust (NDCG@5 > 0.70, CTR > 0.45), demonstrating its advantage in complex context analysis and personalized content delivery. These results demonstrate the advantages of this approach in disambiguating word meanings in complex contexts, maintaining long-term semantic coherence, and matching multi-dimensional annotations with user preferences. In the context of digital humanities and textile cultural studies, the platform can effectively improve the completeness and accuracy of annotations for complex historical texts. It enhances the alignment of recommended content with the cognitive needs of diverse users, from textile history researchers to students of classical literature. This work provides a viable technical path for the digital preservation and intelligent dissemination of textile-related knowledge embedded in cultural heritage, demonstrating significant application value for interdisciplinary fields that bridge AI, cultural studies, and textile history.

AUTHOR CONTRIBUTIONS

Conceptualization – Liping Lei, Xinyao Li, Wen Yan, Rui Zhang and Junyi Xu; methodology – Liping Lei, Xinyao Li, Wen Yan, Rui Zhang and Junyi Xu; investigation – Liping Lei, Xinyao Li, Wen Yan, Rui Zhang and Junyi Xu; writing-original draft preparation – Liping Lei, Xinyao Li, Wen Yan, Rui Zhang and Junyi Xu. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by,

1. 2023 Research Project on Teaching Reform of Undergraduate and Higher Continuing Education in Shaanxi Province by Shaanxi Provincial Department of Education–Research and Practice on the Characteristics of "New Teacher Training" Education (No. 23BZ063)

2. 2023 Vocational Education Research Orientation Project of Shaanxi Province Chinese Vocational Education Society–Research and Practice of Part-time Teachers' Team Construction and Management Mechanism in Vocational Schools (No. ZJS202306)

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] E. Du Plooy, D. Casteleijn, and D. Franzsen, "Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement," Heliyon, vol. 10, no. 21, pp. 1-24, 2024, doi: 10.1016/j.heliyon.2024.e39630.

[2] A. Sikarskie, "Digital research methods in fashion and textile studies," London, UK: Bloomsbury Publishing; 2020.

[3] Y. Wei, L. Hu, Y. Zhu, J. Zhao, and B. Wu, "Knowledge-guided transformer for joint theme and emotion classification of Chinese classical poetry," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 1, no. 1, pp. 1-12, 2024, doi: 10.1109/TASLP.2024.3487409.

[4] A. P. Wibawa and F. Kurniawan, "Advancements in natural language processing: Implications, challenges, and future directions," Telematics and Informatics Reports, vol. 16, no. 1, pp. 100173-100190, 2024, doi: 10.1016/j.teler.2024.100173.

[5] C. Pramartha, G. D. Saputra, I. W. Supriana, I. G. Putra, I. W. Arka, and M. A. Raharja, "Enabling Cultural Heritage Preservation: A Semantic Web Prototype for Digitizing and Documenting Balinese Woven Fabric Artifacts," Proceedings of the First International Conference on Applied Mathematics, Statistics, and Computing (ICAMSAC 2023); 21-22 November 2023; Denpasar, Indonesia (Hybrid), 2024, doi: 10.2991/978-94-6463-413-6_12.

[6] E. Alba, M. Gaitán, A. León, D. Mladenić, and J. Brank, "Weaving words for textile museums: the development of the linked SILKNOW thesaurus," Heritage Science, vol. 10, no. 1, pp. 59, 2022, doi: 10.21203/rs.3.rs-1043107/v1.

[7] M. C. Hyer, "Textiles and Textile Imagery in Early Medieval English Literature: Traditions and Contexts," Suffolk, UK: Boydell & Brewer; 2025, doi: 10.1515/9781805436126.

[8] C. Xiao, P. Zhang, X. Han, G. Xiao, Y. Lin, Z. Zhang, et al., "Infllm: Training-free long-context extrapolation for llms with an efficient context memory," Advances in Neural Information Processing Systems, vol. 37, no. 1, pp. 119638-119661, 2024, doi: 10.52202/079017-3801.

[9] C. C. Y. Yang and H. Ogata, "Personalized learning analytics intervention approach for enhancing student learning achievement and behavioral engagement in blended learning," Education and Information Technologies, vol. 28, no. 3, pp. 2509-2528, 2023, doi: 10.1007/s10639-022-11291-2.

[10] H. Yaseen, A. S. Mohammad, N. Ashal, H. Abusaimeh, A. Ali, and A. A. A. Sharabati, "The impact of adaptive learning technologies, personalized feedback, and interactive AI tools on student engagement: The moderating role of digital literacy," Sustainability, vol. 17, no. 3, pp. 1133-1160, 2025, doi: 10.3390/su17031133.

[11] L. C. Wang, X. Y. Zeng, L. Koehl, and Y. Chen, "Intelligent fashion recommender system: Fuzzy logic in personalized garment design," IEEE Transactions on Human-Machine Systems, vol. 45, no. 1, pp. 95-109, 2014, doi: 10.1109/THMS.2014.2364398.

[12] S. Chakraborty, M. S. Hoque, and S. M. Surid, "A COMPREHENSIVE REVIEW ON IMAGE BASED STYLE PREDICTION AND ONLINE FASHION RECOMMENDATION," Journal of Modern Technology & Engineering, vol. 5, no. 3, pp. 212-233, 2020, [Online]. Available: https://jomardpublishing.com/UploadFiles/Files/journals/JTME/V5N3/ChakrabortyS.pdf.

[13] M. Mejeh and M. Rehm, "Taking adaptive learning in educational settings to the next level: Leveraging natural language processing for improved personalization," Educational Technology Research and Development, vol. 72, no. 3, pp. 1597-1621, 2024, doi: 10.1007/s11423-024-10345-1.

[14] Y. Yao and H. González-Vélez, "AI-Powered System to Facilitate Personalized Adaptive Learning in Digital Transformation," Applied Sciences, vol. 15, no. 9, pp. 4989-5016, 2025, doi: 10.3390/app15094989.

[15] J. Hou and S. Zhang, "Exploring Thematic Diversity in Classical Chinese Poetry: A Novel Dataset and a BERTenhanced Ensemble Learning Approach," ACM Journal on Computing and Cultural Heritage, vol. 17, no. 4, pp. 1-19, 2025, doi: 10.1145/3685679.

[16] Z. Li, Y. Jian, Z. Xue, Y. Zheng, M. Zhang, and Y. Zhang, "Text-enhanced knowledge graph representation learning with local structure," Information Processing & Management, vol. 61, no. 5, pp. 103797-103813, 2024, doi: 10.1016/j.ipm.2024.103797.

[17] H. Yu, C. Gao, X. Li, and L. Zhang, "Ancient Chinese Poetry Collation Based on BERT," Procedia Computer Science, vol. 242, no. 1, pp. 1171-1178, 2024, doi: 10.1016/j.procs.2024.08.179.

[18] Q. Sun, K. Zhang, L. Lv, X. Li, K. Huang, and T. Zhang, "Joint extraction of entities and overlapping relations by improved graph convolutional networks," Applied Intelligence, vol. 52, no. 5, pp. 5212-5224, 2022, doi: 10.1007/s10489-021-02667-x.

[19] D. Du, J. Ding, and Y. Liu, "Knowledge graph construction of Chinese embroidery evolution based on associating cultural space and critical incidents under intangible cultural heritage," The Electronic Library, vol. 43, no. 3, pp. 283-302, 2025, doi: 10.1108/EL-02-2024-0036.

[20] A. Jana and R. Rout, "Sambalpuri Sarees in the Digital Loom: Metadata Schema and Natural Language Query Interface Design," Journal of Library Metadata, vol. 1, no. 1, pp. 1-20, 2026, doi: 10.1080/19386389.2026.2636260.

[21] L. Gan, L. Hu, X. Tan, and X. Du, "TBNF: A transformer-based noise filtering method for Chinese long-form text matching," Applied Intelligence, vol. 53, no. 19, pp. 22313-22327, 2023, doi: 10.1007/s10489-023-04607-3.

[22] S. Zhang, P. Wang, Z. Li, J. Hou, and Q. Hu, "Confidence-based Syntax encoding network for better ancient Chinese understanding," Information Processing & Management, vol. 61, no. 3, pp. 103616-103632, 2024, doi: 10.1016/j.ipm.2023.103616.

[23] C. Yu, H. Xue, Y. Jiang, L. An, and G. Li, "A simple and efficient text matching model based on deep interaction," Information Processing & Management, vol. 58, no. 6, pp. 102738-102754, 2021, doi: 10.1016/j.ipm.2021.102738.

[24] Y. Wei, M. Li, Y. Zhu, Y. Xu, Y. Li, and B. Wu, "A diachronic language model for long-time span classical Chinese," Information Processing & Management, vol. 62, no. 1, pp. 103925-103941, 2025, doi: 10.1016/j.ipm.2024.103925.

[25] C. Chen, "A Personalized Recommendation Method for Ancient Chinese Literary Works Based on a Collaborative Filtering Algorithm," Mobile Information Systems, vol. 2022, no. 1, pp. 4460628-4460635, 2022, doi: 10.1155/2022/4460628.

[26] Y. Ma and G. Yu, "Construction of a Teaching Platform of Modern and Contemporary Chinese Literature From the Perspective of Diversified Intelligence," International Journal of e-Collaboration (IJeC), vol. 20, no. 1, pp. 1-18, 2024, doi: 10.4018/ijec.357638.

[27] G. Lv, "Personalized Recommendation Algorithm of Literary Works Based on Annotated Corpus," Mobile Information Systems, vol. 2022, no. 1, Art. no. 5198612, 2022, doi: 10.1155/2022/5198612.

[28] S. Shirkhani, "Image-based fashion recommender systems," 2021, [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1603327&dswid=-636.

[29] G. Wu, Z. Lu, X. Zhuo, X. Bao, and X. Wu, "Semantic fusion enhanced event detection via multi-graph attention network with skip connection," IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 7, no. 3, pp. 931-941, 2023, doi: 10.1109/TETCI.2022.3229414.

[30] S. Y. Cui, B. W. Yu, X. Cong, T. W. Liu, Q. F. Tan, and J. Q. Shi, "Label-aware Chinese event detection with heterogeneous graph attention network," Journal of Computer Science and Technology, vol. 39, no. 1, pp. 227-242, 2024, doi: 10.1007/s11390-023-1541-6.

[31] M. Yasunaga, A. Bosselut, H. Ren, X. Zhang, C. D. Manning, P. S. Liang, et al., "Deep bidirectional languageknowledge graph pretraining," Advances in Neural Information Processing Systems, vol. 35, no. 1, pp. 37309-37323, 2022.

[32] J. Huang and S. Liu, "Lexicon-enhanced transformer with spatial-aware integration for Chinese named entity recognition," Complex & Intelligent Systems, vol. 11, no. 8, pp. 1-17, 2025, doi: 10.1007/s40747-025-01953-2.

[33] J. Cheng, W. Tong, and W. Yan, "Capsule network improved multi-head attention for word sense disambiguation," Applied Sciences, vol. 11, no. 6, pp. 2488-2502, 2021, doi: 10.3390/app11062488.

[34] S. Kwon, D. Oh, and Y. Ko, "Word sense disambiguation based on context selection using knowledge-based word similarity," Information Processing & Management, vol. 58, no. 4, pp. 102551-102567, 2021, doi: 10.1016/j.ipm.2021.102551.

[35] X. Chen, P. Cong, and S. Lv, "A long-text classification method of Chinese news based on BERT and CNN," IEEE Access, vol. 10, no. 1, pp. 34046-34057, 2022, doi: 10.1109/ACCESS.2022.3162614.

[36] T. Yang, L. Hu, C. Shi, H. Ji, X. Li, and L. Nie, "HGAT: Heterogeneous graph attention networks for semi-supervised short text classification," ACM Transactions on Information Systems (TOIS), vol. 39, no. 3, pp. 1-29, 2021, doi: 10.1145/3450352.