Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrating Adaptive Graph Propagation and Pretrained Text Semantic Alignment for Book Recommendation

Changhua Chen

cchlib@nuaa.edu.cn

Nanjing University of Aeronautics and Astronautics

Li Shi

Nanjing University of Aeronautics and Astronautics

Xiaotao Li

Nanjing University of Aeronautics and Astronautics

Article

Keywords: Book Recommendation from Interaction Sessions, Graph Neural Network, Adaptive Propagation Guided by Session Length, Text Semantic Alignment

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11562650/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Integrating Adaptive Graph Propagation and Pretrained Text Semantic Alignment for Book Recommendation

Changhua Chen[1*], Li Shi[1] and Xiaotao Li[1]

[1*]Library of Nanjing University of Aeronautics and Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China.

*Corresponding author(s). E-mail(s): cchlib@nuaa.edu.cn;

Abstract

Book recommendation from interaction sessions predicts a user's next book from an ordered sequence of interactions. Accurate prediction requires a session representation that captures both the transitions among interacted books and their textual semantics across different session lengths. Graph neural networks (GNNs) capture interaction transitions, while a fixed rule for updating node representations may not reflect differences in the amount of observed history. To address this, we develop a unified framework for session representation learning based on graphs. Embeddings of book text obtained from a pretrained model are projected into the item representation space and combined with trainable book ID embeddings, allowing textual information to enter graph propagation. A gate uses the effective session length to regulate updates to node representations. Features from successive graph layers are aggregated with learnable weights and complementary pooling operations to construct a session representation. This representation is used both to rank candidate books and to align with the aggregated textual semantics of the observed session through contrastive learning, connecting the components within a shared training objective. Experiments on Amazon Books, GoodReads, and Book Crossing show that the model achieves the highest NDCG@50 among six compared baselines, improving by 12.2%, 10.3%, and 2.8%, respectively, over the strongest baseline on each dataset.

Keywords: Book Recommendation from Interaction Sessions, Graph Neural Network, Adaptive Propagation Guided by Session Length, Text Semantic Alignment

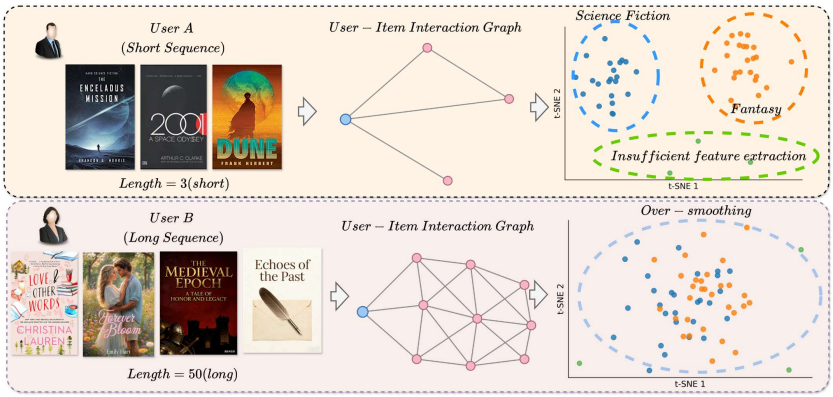

Fig. 1 Schematic illustration of how session length affects recommendation using graphs. Short sessions provide fewer interaction transitions, while repeated propagation in longer session graphs can reduce the distinctiveness of item representations.

1 Introduction

Online reading platforms contain large collections of books, making personalized recommendation important for helping users discover relevant titles(Edmunds and Morris 2000; Yang et al. 2025; Wayesa et al. 2023). In this work, we study the task of predicting the next book from interaction sessions: given an ordered sequence of books a user has interacted with, the model ranks candidate books for the next interaction. We refer to this ordered sequence of observed book interactions as a *session*. The order of interactions is informative because successive books can reflect the development of a user's immediate reading interests(Pan et al. 2026).

Graph neural networks (GNNs) provide a way to model this structure by representing books as nodes and successive interactions as directed edges(Li et al. 2025; Huang et al. 2025; Lou et al. 2025). However, a session representation cannot be determined by transition structure alone. The amount of behavioral evidence varies with session length: a short session contains only a few observed transitions, whereas a longer session requires information from multiple transitions to be combined without unnecessarily reducing differences between item representations. Book titles and other available textual attributes provide additional clues about the interests reflected in those interactions. A meaningful representation for predicting the next book should therefore preserve both the observed transition patterns and their semantic context. Figure 1 illustrates the different representation challenges associated with shorter and longer session graphs.

These considerations lead to a single modeling objective: learning a session representation that responds to the amount of interaction history, incorporates book semantics, and supports ranking of candidate books. We address this objective through a continuous computation and learning process. Text features obtained from a pretrained model first complement the trainable book ID embeddings, so that the initial item representations contain both textual information and patterns learned from interactions. The resulting features are propagated over the directed session graph, with a gate conditioned on session length controlling updates to node states. Outputs from

different propagation layers and sequence positions are then aggregated into one session representation. Finally, that representation serves both candidate book scoring and a contrastive objective that aligns it with the textual summary of the observed session. In this way, textual information participates in both graph feature learning and the training of the representation used for recommendation.

Based on this design, we develop a book recommendation model based on GNNs with an adaptive graph neural network module guided by session length (AGNM), a module for text semantic alignment and contrastive learning (MSACL), and a multiscale feature aggregation and hybrid pooling module (MFHP). These components respectively regulate how interaction information is updated, introduce semantic information into that process and its training objective, and organize the resulting features for scoring of candidate books. We evaluate the model on Amazon Books(Ni et al. 2019), GoodReads(Wan and McAuley 2018), and Book Crossing(Ziegler et al. 2005). The model achieves the highest NDCG@50 among the six compared baselines on all three datasets; the study of module combinations on Book Crossing further examines their contributions to recommendation quality.

The contributions of this work are summarized as follows:

1. We formulate a session graph encoder guided by session length that regulates updates to node states according to the effective number of interactions in each session while retaining residual information from previous layers.
2. We connect graph representations learned from user behavior with pretrained book text information through integration of features at the item level and a contrastive objective between session and text representations. This provides a semantic learning signal to the session representation used for prediction of the next book.
3. We construct a unified session representation from multiple graph propagation layers and complementary pooling operations. The resulting representation supports both recommendation scoring and semantic alignment, and the combined framework is evaluated on three book datasets.

2 Related Work

2.1 Graph Neural Networks for Book Recommendation

As a core technology to alleviate information overload in digital reading and improve user reading experience, book recommendation systems deliver personalized content by modeling users' historical interactions to uncover their latent reading preferences(Xu 2026). Early methods, mostly based on traditional paradigms like collaborative filtering and matrix factorization, struggled to effectively capture temporal dependencies and complex interest relationships in user interaction sequences(Kannikaklang et al. 2025; Duan et al. 2024). In recent years, recommendation based on interaction sequences has become a research hotspot, with researchers(Liang et al. 2025; Zhang et al. 2024b) greatly improving recommendation accuracy by modeling the temporal evolution of user behavior. Among these, Graph Neural Networks (GNNs) are widely used in book recommendation for their strong ability to model topological relationships

and propagate information(Pham et al. 2025). For recommendation from interaction sessions, a directed graph can be constructed from successive book interactions, allowing information to propagate among items observed in the same session.

However, existing book recommendation methods using GNNs still face significant performance bottlenecks(Peng et al. 2025a; Spillo et al. 2026). Models with fixed graph update rules may respond similarly to sessions with different numbers of interactions(Zheng et al. 2026). Short sessions provide limited transition evidence, while longer sessions require features from many items to be combined without losing distinctions between individual items. In addition, interaction records do not directly convey the semantic relations expressed in book text, such as titles and available descriptions(Peng et al. 2025b; Sun 2026). Session representations based on a single pooling rule may also emphasize one aspect of interaction history while discarding complementary information carried by other graph layers or sequence positions(Li et al. 2026; Pu et al. 2025). These observations motivate representing transition structure, textual meaning, and information from different propagation depths within the same objective for learning session representations.

2.2 Adaptive Graph Propagation Mechanisms in Sequence Recommendation

Graph propagation determines how information from neighboring items is incorporated into their representations. In session recommendation, the relevant neighborhood is shaped by the number and ordering of observed interactions: short sessions contain fewer transitions, while longer sessions may contain multiple overlapping paths. Gating mechanisms regulate the contribution of new information to a node state, and residual connections retain information from preceding layers(Chen et al. 2025; Elahi et al. 2025; Saghafi and Mirzaei 2026). Other adaptive propagation methods adjust layer weights or propagation depth according to graph characteristics(Lei et al. 2025; Liu et al. 2026).

Our setting uses a common propagation depth but conditions the update to node states on the effective session length. The coefficient determined by session length modulates the reset and update gates in each layer, allowing update strength to vary with the amount of observed interaction history. This mechanism is paired with aggregation across layers so that the final session representation can combine features from several propagation depths(Zhao et al. 2025; Wu et al. 2019).

2.3 Text Semantic Alignment and Multiscale Feature Aggregation in Recommendation Systems

Book text provides information that is complementary to interaction records. Titles and available descriptions can help associate books whose observed interaction histories are sparse or dissimilar(Wang et al. 2025b; Tang and Mi 2025; Wang et al. 2025a). Pretrained language models offer a way to encode this information, while contrastive learning can connect representations derived from interaction behavior and book text(Bellini et al. 2024; Song et al. 2025; Qin et al. 2026). For predicting the next

book from an interaction session, this connection can be established between a representation of the observed session and the aggregated textual features of the books within that session, rather than treating book text as an independent recommendation signal.

The representation of a session also depends on how node features are aggregated after graph propagation. Outputs from shallower and deeper layers reflect different extents of interaction neighborhoods, while attention pooling, mean pooling, and the state of the final item summarize different aspects of the ordered session(Feng and Jin 2026; Jiang et al. 2021; Zhang et al. 2024a). Combining these features is useful when a recommendation must reflect both repeated interaction patterns and the user's most recent action. In our framework, multiscale aggregation produces the session representation used for next item scoring and for text semantic alignment. This makes semantic supervision part of the training of the same representation that serves the recommendation task.

3 Methods

The proposed framework learns a session representation for prediction of the next book by connecting textual item information with ordered interaction patterns. Book text and ID embeddings jointly initialize the session graph, a graph encoder guided by session length updates its node features, and multiscale aggregation produces the final session representation. This representation supports both ranking of candidate books and a semantic alignment objective for each session. Figure 2 summarizes the computation from input construction to joint training.

3.1 Problem Definition

Let $S = [v_1, v_2, \ldots, v_L]$ denote a session of L valid book interactions, where $v_i \in \mathcal{V}$ is the book ID at position i and $\mathcal{V}$ is the catalog containing V books. The order of interactions is retained, and a book may appear more than once in the same session. Given S, the task is to rank candidate books according to their likelihood of being the next interaction v_{L+1}. Our model encodes the observed session into $s \in \mathbb{R}^d$ and computes a score for each candidate book. Padding items used during batch processing are excluded when determining L.

3.2 Overall Architecture

As illustrated in Figure 2, the model follows a single path for learning session representations. First, pretrained book text features are projected into the ID embedding space to initialize item features. The observed transitions are then represented by a directed session graph. An adaptive graph neural network guided by session length (AGNM) propagates the combined features and returns node representations at several layers. A multiscale feature aggregation and hybrid pooling module (MFHP) converts these node representations into the session vector s. The module for text semantic alignment and contrastive learning (MSACL) also uses s, relating it to the textual content

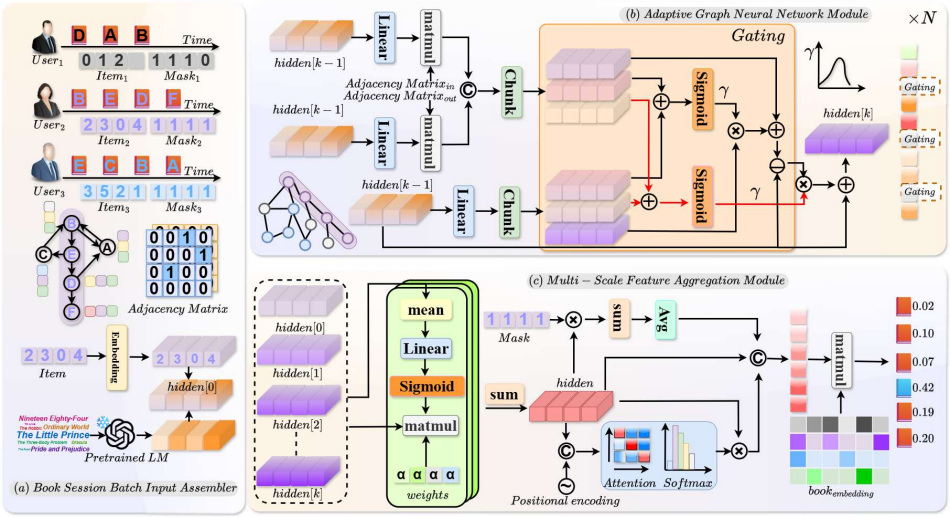

Fig. 2 Overview of the proposed session representation framework. Book ID and text features initialize the directed session graph; propagation guided by session length produces node features from different layers; aggregation forms a session vector used for both scoring of candidate books and alignment between session and text representations.

of the observed session during training. Thus, the prediction and alignment objectives act on the same session representation.

3.3 Item Representation with Textual Information (MSACL)

Book interaction records identify which items occur in a session, while book text provides information about the meaning of those items. For each book v, a pretrained text encoder supplies an embedding e^v_{text} based on its available textual information, such as its title. A trainable linear projection maps this feature to the space of dimension d used by the book ID embedding e^v_{id}. The initial item representation is

$$e^v = e^v_{id} + W_{proj} e^v_{text} + b_{proj}. \tag{1}$$

The ID embedding is learned from the recommendation data, whereas W_{proj} and b_{proj} connect the pretrained textual features to the recommendation model. The addition in Eq. (1) is elementwise. The resulting vector initializes the corresponding node in the session graph, allowing graph propagation to operate on features informed by both interaction data and book text.

We also retain the projected text embeddings as a semantic reference for each session. For a session S, the text summary t_S is the mean of the projected text embeddings over its L valid interaction positions. If a book occurs more than once, each occurrence contributes to this mean. During joint optimization, a contrastive objective aligns this text summary with the final session representation.

3.4 Graph Propagation Guided by Session Length (AGNM)

We convert each ordered session into a directed graph $\mathcal{G} = (\mathcal{V}_S, \mathcal{E})$. Each distinct book ID forms one node, and each consecutive pair $v_i \rightarrow v_{i+1}$ contributes a directed transition. Repeated transitions are recorded in the corresponding edge counts. Two adjacency matrices capture predecessor and successor relations separately: A_{in} describes incoming transitions, and A_{out} describes outgoing transitions. Each is obtained by normalizing the corresponding matrix of transition counts by its row sum, with a small constant $\epsilon = 10^{-9}$ to handle empty rows. Separate matrices preserve direction information during neighborhood aggregation.

Sessions contain different numbers of observed interactions. AGNM accounts for this difference by learning a session coefficient that regulates node updates:

$$\gamma = \sigma(g_\theta(\mathrm{CLAMP}(L, 1, 50))) \,. \tag{2}$$

Here, g_θ is a trainable scalar mapping, σ is the sigmoid function, and CLAMP restricts the input length to the range used in preprocessing. The resulting $\gamma \in (0, 1)$ is shared by the nodes of a session. It changes the strength of node updates without changing the number of graph layers.

At each propagation layer, the incoming and outgoing adjacency matrices aggregate features from the two transition directions. Each channel applies a trainable linear transformation to the previous node states, incorporates an embedding of item popularity, and adds a bias after neighbor aggregation. Their outputs, $h_{in}^{(l)}$ and $h_{out}^{(l)}$, are concatenated and passed to a node update based on a gated recurrent unit (GRU). The reset gate $r^{(l)}$ and update gate $i^{(l)}$ are calculated from these aggregated features and the previous state. For each gate, a sigmoid is applied to the sum of separate learned affine transformations of the concatenated neighborhood features and the previous node state. The resulting gate values are then multiplied by γ. A candidate state $\tilde{h}^{(l)}$ is obtained with a tanh transformation that uses the reset gate to control the contribution of the previous node state.

The final node update uses a residual connection and layer normalization:

$$h^{(l)} = \mathrm{LayerNorm}\Big(h^{(l-1)} + \mathrm{Dropout}\Big(i^{(l)} \odot \big(\tilde{h}^{(l)} - h^{(l-1)}\big)\Big)\Big) \,. \tag{3}$$

Here, $\odot$ denotes elementwise multiplication. A larger update gate allows a larger change toward the candidate state, while the residual term retains information from the previous layer. We use a fixed number D of propagation layers and pass all their outputs $\{h^{(1)}, \ldots, h^{(D)}\}$ to the subsequent aggregation module. The reported experiments use $D = 3$.

3.5 Multiscale Session Representation (MFHP)

Graph features obtained at different propagation layers summarize different ranges of transition information. Instead of selecting the final layer alone, MFHP combines the layers using trainable scalar scores $\alpha_1, \ldots, \alpha_D$. Their normalized weights α_l' are obtained using softmax and determine the relative contributions of different depths.

A feature gate g is computed by applying a learned transformation and a sigmoid to the mean of the layer outputs. The aggregated node features are

$$h_{multi} = g \odot \sum_{l=1}^{D} \alpha'_l h^{(l)}, \qquad \alpha'_l = \frac{\exp(\alpha_l)}{\sum_{k=1}^{D} \exp(\alpha_k)}. \tag{4}$$

The scores α_l and the parameters of the feature gate are learned through the joint training loss. Unlike the coefficient γ determined by session length, which regulates graph updates, g weights dimensions of the features extracted by those updates.

Next, the aggregated features of distinct book nodes are mapped back to their positions in the original interaction sequence. Consequently, repeated books occupy their original positions even though they share a graph node. For each session, only its valid positions are considered in the following pooling operations. Attention pooling combines each node feature with a learned positional embedding, calculates an attention score using a sigmoid function, and normalizes the scores over valid positions using softmax. Mean pooling averages these valid states to summarize the overall interaction history. Pooling of the last item retains the hidden state at the final valid position, preserving information about the most recent interaction. The three summaries are concatenated and projected to the common session vector:

$$s = W_{fusion}[h_{attn}; h_{avg}; h_{last}] + b_{fusion}. \tag{5}$$

Here, $[\,;\,]$ denotes concatenation. The vector s is used without an additional session encoder in both the prediction objective and the text alignment objective. It is therefore the common representation through which the graph propagation, semantic features, and multiscale aggregation are trained together.

3.6 Next Book Prediction and Joint Optimization

For candidate book v, the model computes a ranking score from the session representation and the trainable book ID embedding:

$$\hat{y}_v = s^\top e^v_{id} + b_v, \tag{6}$$

where b_v is a trainable bias specific to candidate v. The item features enriched by text are used to encode the observed session, while candidate scoring uses the ID embedding. Scores over the catalog are converted into a probability distribution by softmax for training. The main objective $\mathcal{L}_{ce}$ is cross entropy with label smoothing. For a session with target book v^+, its smoothed target distribution assigns probability $1-\beta$ to v^+ and $\beta/(V-1)$ to each other book, with $\beta = 0.1$ in the reported setting.

The semantic objective compares the final session vector with the mean projected text embedding t_S constructed from the books in each session. For a batch of N sessions, the representation of each session and its own text summary form a positive pair. Text summaries of other sessions in that batch are comparison pairs. We use the

contrastive loss

$$\mathcal{L}_{contrast} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(z_{ii})}{\sum_{j=1}^{N}\exp(z_{ij})}, \quad z_{ij} = \frac{\mathrm{sim}(s_i, t_{S_j})}{\tau}, \tag{7}$$

where sim denotes cosine similarity and $\tau = 0.1$ is the temperature. This objective relates the session representation used for recommendation to the textual semantics of the books in the observed session. The full training objective is

$$\mathcal{L}_{total} = \mathcal{L}_{ce} + \lambda_{ssl}\mathcal{L}_{contrast}. \tag{8}$$

The coefficient λ_{ssl} balances the two losses. Their gradients jointly update the trainable book ID embeddings, text projection, graph encoder, and multiscale aggregation parameters. After training, recommendations for the next book are produced by ranking the candidate scores in Eq. (6).

4 Experiments

4.1 Datasets

We evaluate the proposed model on three book recommendation datasets: Amazon Books(Ni et al. 2019), Book Crossing(Ziegler et al. 2005), and GoodReads(Wan and McAuley 2018). Among these, Amazon Books is a large dataset of implicit interactions from online retail that covers many users and book categories; Book Crossing is a publicly available book rating dataset with sparse interactions and a broad user population, posing a greater challenge to the model's generalization ability; GoodReads is a subset of a reading community interaction dataset, featuring complete user behavior sequences and rich book attribute information, making it appropriate for modeling tasks involving recommendation from interaction sessions. For the GoodReads experiments, a random subset containing 10% of the original data is used for model training under the available memory budget.

Following standard experimental procedures in the field of session recommendation, this paper applies uniform preprocessing to the raw dataset: users and books with fewer than five recorded interactions are filtered; each user's interactions are ordered by timestamp and grouped into sequences containing at most 50 interactions; shorter sequences are padded with the reserved zero ID for batch processing. The resulting samples are divided into training, validation, and test subsets in an 8:1:1 ratio. Session lengths used by the gate and pooling operations are calculated before padding.

4.2 Evaluation Criteria

Consistent with existing research in recommendation from interaction sessions, this paper employs Recall@K and Normalized Discounted Cumulative Gain (NDCG@K) to evaluate recommendation performance. Here, K is set to 10, 20, and 50 to assess retrieval and ranking quality at different recommendation list lengths; higher values indicate better performance.

Table 1 Performance comparison across three datasets. Best results are shown in bold and second highest results are underlined; gray shading indicates our method.

Dataset	Model	Recall			NDCG		
		@10	@20	@50	@10	@20	@50
Amazon Books	LINE	0.0706	0.1048	<u>0.1660</u>	0.0425	0.0520	<u>0.0658</u>
	GCMC	0.0787	0.1099	0.1596	0.0422	0.0501	0.0599
	LightGCN	0.0384	0.0628	0.1147	0.0220	0.0287	0.0403
	DGCF	0.0573	0.0844	0.1299	0.0297	0.0365	0.0455
	SGL	**0.0822**	<u>0.1139</u>	0.1636	<u>0.0446</u>	<u>0.0526</u>	0.0625
	DiffRec	<u>0.0805</u>	0.1122	0.1617	0.0435	0.0515	0.0613
	Ours	0.0758	**0.1166**	**0.1944**	**0.0451**	**0.0564**	**0.0738**
Book Crossing	LINE	0.0880	0.1162	0.1813	0.0575	0.0655	0.0804
	GCMC	0.0356	0.0578	0.1190	0.0201	0.0267	0.0409
	LightGCN	0.0824	0.1173	0.1916	0.0489	0.0590	0.0761
	DGCF	0.0942	0.1321	<u>0.2022</u>	0.0589	0.0696	<u>0.0858</u>
	SGL	0.0915	0.1304	0.2009	0.0496	0.0594	0.0734
	DiffRec	<u>0.1046</u>	<u>0.1359</u>	0.1898	**0.0629**	<u>0.0707</u>	0.0814
	Ours	**0.1120**	**0.1581**	**0.2336**	<u>0.0616</u>	**0.0733**	**0.0882**
GoodReads	LINE	0.0191	0.0336	0.0673	0.0105	0.0146	0.0223
	GCMC	0.0310	0.0547	0.1080	0.0166	0.0232	0.0350
	LightGCN	0.0342	0.0600	0.1178	0.0179	0.0251	<u>0.0380</u>
	DGCF	0.0286	0.0489	0.0905	0.0154	0.0210	0.0303
	SGL	0.0318	0.0558	0.1091	0.0168	0.0236	0.0355
	DiffRec	<u>0.0384</u>	<u>0.0665</u>	<u>0.1249</u>	<u>0.0186</u>	<u>0.0257</u>	0.0372
	Ours	**0.0426**	**0.0744**	**0.1424**	**0.0206**	**0.0285**	**0.0419**

4.3 Experimental Parameters

All experiments were conducted on an NVIDIA A5000 24GB GPU, with model development and training performed using the PyTorch deep learning framework. We trained the model using the AdamW optimizer, with an initial learning rate of 1×10^{-3}, a weight decay coefficient of 1×10^{-5}, a batch size of 512, and a maximum of 100 training epochs. An early stopping strategy was employed: training was terminated when the validation set's NDCG@20 metric showed no improvement for five consecutive training epochs to prevent overfitting. The model was configured with 3 GNN layers, a hidden layer dimension of 64, a dropout rate of 0.2, and an initial adaptive gating coefficient of 0.5. The multiscale aggregation weights were automatically learned and optimized through joint training of the model.

4.4 Performance Comparison with Mainstream Methods

We compare the proposed model with six representative recommendation baselines, covering classical neural embedding methods, collaborative filtering models based on

graphs, recommendation methods using self supervision, and generative recommendation models. Specifically, these include: LINE(Tang et al. 2015): a classical neural embedding model that models the structural features of nodes through first order and second order proximity to provide a foundational embedding representation for recommendation using graphs; GCMC: a graph convolutional matrix factorization model that combines graph convolutional operations with a collaborative filtering framework to achieve joint modeling of user and item interactions; LightGCN(He et al. 2020): a graph convolutional recommendation model that removes nonlinear transformations from conventional graph convolutions, retaining only neighborhood aggregation operations, and serves as a core baseline in the field of recommendation using graphs; DGCF(Wang et al. 2020): a disentangled graph collaborative filtering model that captures different latent factors underlying user and item interactions; SGL(Wu et al. 2021): a graph recommendation model using self supervision that constructs a contrastive learning loss enhanced by graph structure, effectively mitigating the data sparsity issue in recommendation systems; DiffRec(Wang et al. 2023): a generative recommendation method based on diffusion models that models the distribution characteristics of user and item interactions through a diffusion process, improving the robustness of recommendations.

We compare the proposed model with these baselines on the three book datasets. The experimental results are shown in Table 1, with the best results for each metric highlighted in bold. Table 1 shows that the proposed model obtains the highest NDCG@50 on all three datasets and achieves the best value on most of the reported Recall@K and NDCG@K measures. Specifically, on the Amazon Books dataset, our method improves Recall@50 and NDCG@50 by 17.1% and 12.2%, respectively, over the strongest baseline; on the sparsely observed Book Crossing dataset, our method achieves an improvement of 15.5% in Recall@50 compared with DGCF and 2.8% in NDCG@50 compared with the strongest baseline for that metric, indicating improved ranking performance under the reported setting with sparse data. On the GoodReads dataset, our method achieves the best results on all Recall@K and NDCG@K metrics, with Recall@50 improving by 14.0% over the strongest baseline, DiffRec, consistent with the benefits of learning session representations.

The results show that regulating graph updates, incorporating information derived from book text, and aggregating multiple propagation levels can be used together to form a session representation for ranking candidate books. The relative contributions of these components are examined in the following ablation study.

4.5 Ablation Experiment

To examine how the components contribute to the shared session representation, we compare the three modules and their combinations on Book Crossing. Using Recall@20 and NDCG@20 as key metrics, we evaluate the performance of individual modules, various module combinations, and the full system. The results are shown in Table 2.

Among the configurations containing only one module, AGNM achieves the highest Recall@20 and NDCG@20 (0.1215 and 0.0612). Pairing AGNM with MSACL or MFHP further improves both metrics, and using all three modules yields the highest

Table 2 Ablation study of module combinations on the Book Crossing dataset. AGNM, MSACL, and MFHP denote the three proposed modules; ✓ indicates inclusion and ✗ indicates removal.

AGNM	MSACL	MFHP	Recall@20	NDCG@20
✓	✗	✗	0.1215	0.0612
✗	✓	✗	0.0987	0.0525
✗	✗	✓	0.0964	0.0508
✓	✓	✗	0.1386	0.0657
✓	✗	✓	0.1352	0.0643
✗	✓	✓	0.1023	0.0547
✓	✓	✓	**0.1581**	**0.0733**

scores (0.1581 Recall@20 and 0.0733 NDCG@20). These results show that the complete configuration performs better than the evaluated partial configurations. They are consistent with the intended division of labor: graph propagation forms features of interaction patterns, semantic supervision relates them to book text, and multiscale aggregation organizes them into the session representation used for ranking.

4.6 Visual Analysis

The model is examined on the Book Crossing test set from five perspectives: propagation depth, adaptation to session length, the distribution of learned item representations, textual similarity, and the contribution of individual graph layers.

4.6.1 The Effect of the Number of GNN Layers

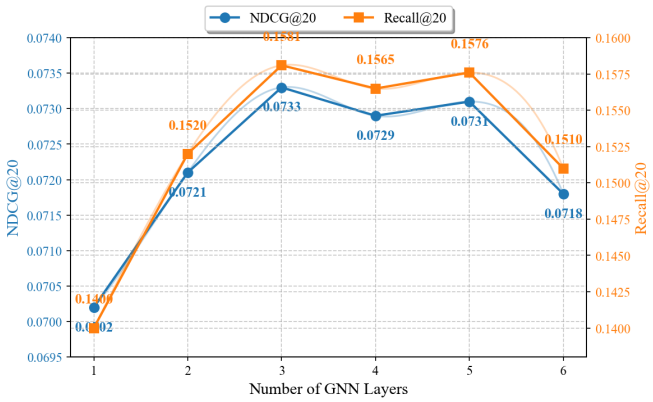

Fig. 3 Recall@20 and NDCG@20 for different numbers of graph propagation layers. Both metrics attain their highest values at three layers.

The number of graph propagation layers determines the range of interaction neighborhoods contributing to each item representation. We trained models containing one

to six graph layers while keeping the remaining experimental settings unchanged. Recall@20 and NDCG@20 were used to compare these configurations, as presented in Figure 3.

Both metrics increase as the depth changes from one to three layers and reach their highest values at three layers. Performance decreases with four layers and recovers slightly with five, although neither setting matches the three layer configuration. A further decline occurs at six layers. Among the evaluated configurations, three propagation layers provide the best recommendation performance and are therefore used in the main experiments.

4.6.2 Analysis of the Adaptive Gating Mechanism

The adaptive gate regulates the extent to which newly aggregated information changes node states according to the number of valid interactions in a session. We extracted the learned coefficient γ for different session lengths, excluding padding positions, and plotted its variation in Figure 4.

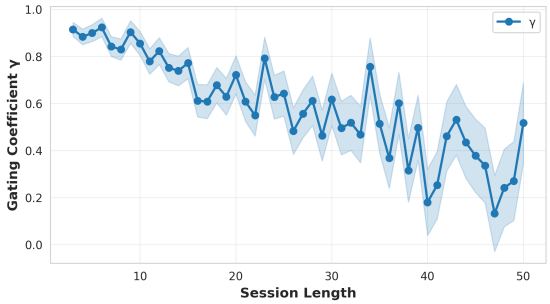

Fig. 4 Learned gating coefficient across effective session lengths. The shaded area serves as a visual guide rather than a statistical interval.

The coefficient is generally close to 0.9 for shorter sessions. Its overall level decreases as the number of interactions grows, with local fluctuations that become more noticeable at longer lengths. Because γ scales the reset and update gates, larger values allow stronger changes to node states, whereas smaller values limit the magnitude of these updates. The learned response varies with session length and follows an overall downward trend without being strictly monotonic.

4.6.3 t-SNE Visualization of Item Representations

To examine the distribution of learned item features, we selected representations associated with sampled sessions and projected them into two dimensions using t-SNE. Different colors identify the sessions from which the items were sampled, as illustrated in Figure 5.

Most samples form relatively compact groups, and the groups associated with different sessions occupy distinguishable regions of the projected space. This distribution suggests that the learned representations retain differences among the selected interaction contexts after graph propagation. The position and spread of each group provide a visual description of how these sampled representations are organized.

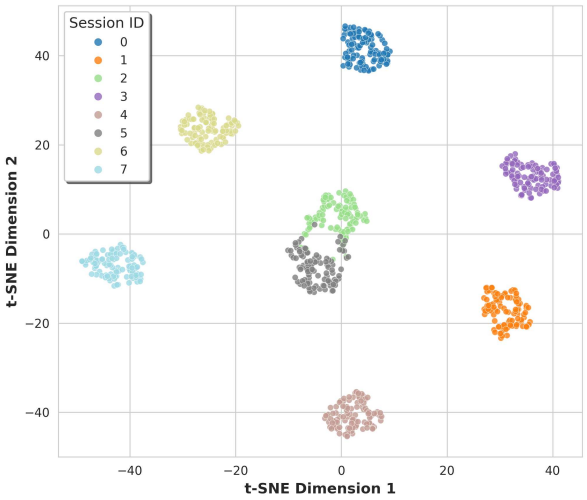

Fig. 5 Two dimensional t-SNE projection of learned item features, colored by sampled session.

4.6.4 Text Semantic Similarity Analysis

Text embeddings provide information about the semantic relationships between books and an observed session. We selected a session, averaged the projected text embeddings of its interacted books, and calculated cosine similarity between this session text summary and candidate book embeddings. Figure 6 presents the 20 books with the highest similarity scores.

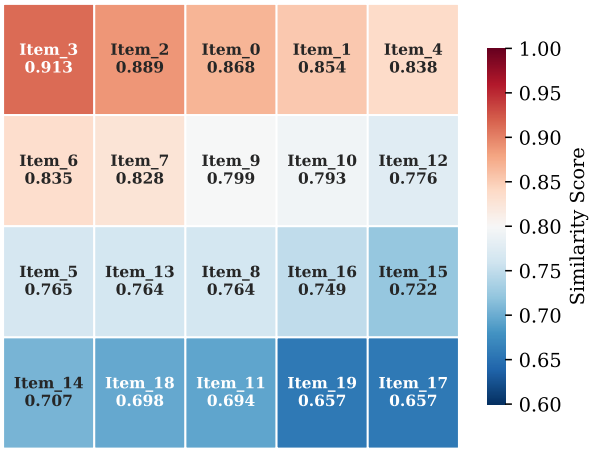

Fig. 6 Cosine similarity between a session text summary and its 20 most similar book embeddings.

The scores range from 0.657 to 0.913. The highest value corresponds to a book whose projected text representation is particularly close to the selected session summary, while the lower values indicate weaker relative similarity within this group.

Variation remains evident even among the 20 closest candidates, reflecting differences in the semantic relationships encoded by the projected text features.

4.6.5 Multiscale Aggregation Weight Analysis

Multiscale aggregation combines features from propagation layers using learned weights. We extracted the normalized layer weights from the trained model to examine their relative contributions to the final session representation. The values are 0.45 for the first layer, 0.35 for the second, and 0.20 for the third, as presented in Figure 7.

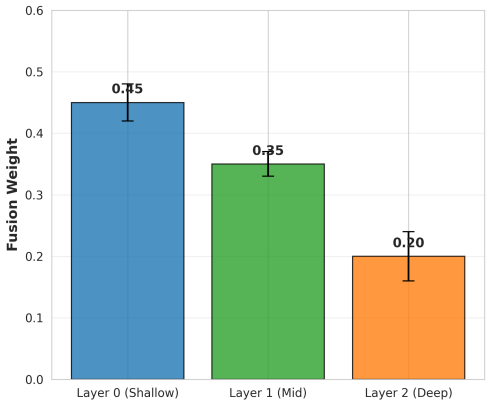

Fig. 7 Learned contributions of the three graph propagation layers to feature aggregation.

The first propagation layer receives the largest weight and therefore contributes the greatest proportion to the aggregated representation. The second and third layers also retain nonzero weights, allowing information from wider interaction neighborhoods to participate in the representation. The distribution indicates that the model combines the three propagation levels with different learned contributions.

5 Conclusions

This paper studies recommendation of the next book through a unified session representation that connects interaction transitions with book semantics. A graph encoder guided by session length controls updates to node states, projected text features enrich the item representations and provide an alignment objective for each session, and multiscale aggregation produces the shared representation used for ranking. On Amazon Books, Book Crossing, and GoodReads, the combined model obtains the highest NDCG@50 among the compared methods. Future work will investigate more lightweight session encoders and the extension of the approach to book recommendation across domains.

Funding. This work was supported by the Library Innovation Research Project of Nanjing University of Aeronautics and Astronautics (Grant No. TSG202509) and the Graduate Education and Teaching Reform Research Project of Nanjing University of Aeronautics and Astronautics (Grant No. 2025YJXGG49).

Author contributions. C.C. contributed to the conception and design of the study, secured funding, supervised the project, and reviewed and edited the manuscript. L.S. contributed to the methodology, model development, software implementation, and preparation of the original draft. X.L. conducted the experiments, performed data analysis and visualization, and participated in manuscript review and editing. All authors read and approved the final version of the manuscript.

Data availability. The datasets analysed during this study are publicly available from their respective sources: Amazon Books (https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/index.html), Goodreads (https://cseweb.ucsd.edu/~jmcauley/datasets/goodreads.html), and Book-Crossing (https://dbis.informatik.uni-freiburg.de/forschung/projekte/2nd-Gen-RS/index.html). Processed data and session splits are available from the corresponding author upon reasonable request, subject to the original data providers' terms of use.

References

Bellini V, Di Sciascio E, Donini FM, et al (2024) A qualitative analysis of knowledge graphs in recommendation scenarios through semantics-aware autoencoders. Journal of Intelligent Information Systems 62(3):787–807

Chen W, Ye G, Wang Y, et al (2025) Smoothness really matters: A simple yet effective approach for unsupervised graph domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 15875–15883

Duan Y, Tu Y, Lu Y, et al (2024) Improving graph collaborative filtering with view explorer for social recommendation. Journal of Intelligent Information Systems 62(6):1703–1724

Edmunds A, Morris A (2000) The problem of information overload in business organisations: a review of the literature. International Journal of Information Management 20(1):17–28

Elahi E, Anwar S, Al-Kfairy M, et al (2025) Graph attention-based neural collaborative filtering for item-specific recommendation system using knowledge graph. Expert Systems with Applications 266:126133

Feng J, Jin X (2026) Fusion of diffusion models and intent learning in sequential recommendation. Journal of Intelligent Information Systems 64(2):829–855

He X, Deng K, Wang X, et al (2020) Lightgcn: Simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, pp 639–648

Huang J, Qin J, Yu Y, et al (2025) Beyond graph convolution: Multimodal recommendation with topology-aware mlps. In: Proceedings of the AAAI conference on artificial intelligence, pp 11808–11816

Jiang N, Duan F, Chen H, et al (2021) Mafi: Gnn-based multiple aggregators and feature interactions network for fraud detection over heterogeneous graph. IEEE Transactions on Big Data 8(4):905–919

Kannikaklang N, Thamviset W, Wongthanavasu S (2025) Bigcan: A novel srs-based bidirectional graph convolution attention network for dynamic user preference and next-item recommendation. Expert Systems with Applications 265:126016

Lei S, Chang X, Yu Z, et al (2025) Feature-structure adaptive completion graph neural network for cold-start recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 12022–12030

Li M, Yan Z, Chen Y, et al (2026) Multi-granular graph learning with fine-grained behavioral pattern awareness for session-based recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 23030–23038

Li W, Song X, Tu Y (2025) Graphdrl: Gnn-based deep reinforcement learning for interactive recommendation with sparse data. Expert Systems with Applications 273:126832

Liang S, Wen W, Feng Y, et al (2025) Modeling multi-seasonal multi-behavior dependency for temporal recommendation. ACM Transactions on Intelligent Systems and Technology 16(4):1–23

Liu X, Zou L, Zhang X, et al (2026) Adadcl: An adaptive disentangled contrastive recommendation method. IEEE Transactions on Computational Social Systems 13(1):569–583

Lou J, Rong C, Chen H, et al (2025) Efficient federated graph aggregation for privacy-preserving GNN-based session recommendation. Scientific Reports 15:23394

Ni J, Li J, McAuley J (2019) Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pp 188–197

Pan LW, Pan WK, Wei MY, et al (2026) A survey on sequential recommendation. Frontiers of Computer Science 20(3):2003606

Peng T, Yuan H, Zhang Y, et al (2025a) Tagrec: Temporal-aware graph contrastive learning with theoretical augmentation for sequential recommendation. IEEE Transactions on Knowledge and Data Engineering 37(5):3015–3029

Peng Y, Gao C, Zhang Y, et al (2025b) Denoising alignment with large language model for recommendation. ACM Transactions on Information Systems 43(2):1–35

Pham P, Bui QT, Nguyen NT, et al (2025) Topological data analysis in graph neural networks: Surveys and perspectives. IEEE Transactions on Neural Networks and Learning Systems 36(6):9758–9776

Pu D, Zhang Y, Qian Z, et al (2025) A graph neural network recommendation algorithm based on multi-scale attention and contrastive learning. Scientific Reports 15:32207

Qin X, Cui Y, Sun S, et al (2026) Disentangle to fuse: Toward content preservation and cross-modality consistency for multi-modality image fusion. IEEE Transactions on Image Processing 35:1756–1770

Saghafi F, Mirzaei S (2026) HGCN-GRU: temporal memory learning for hypergraph sequential recommendation. Scientific Reports Published online 22 October 2026

Song Q, Hu J, Xiao L, et al (2025) Diffcl: A diffusion-based contrastive learning framework with semantic alignment for multimodal recommendations. IEEE Transactions on Neural Networks and Learning Systems 36(10):18587–18597

Spillo G, Musacchio E, Musto C, et al (2026) Gotta embed them all! - knowledge-aware recommendations fusing heterogeneous multimodal item embeddings. Journal of Intelligent Information Systems 64(2):909–935

Sun J (2026) Disentangled graph–text collaborative filtering with prototypical cold-start for scientific article recommendation. Scientific Reports 16:30565

Tang J, Qu M, Wang M, et al (2015) Line: Large-scale information network embedding. In: Proceedings of the 24th international conference on world wide web, pp 1067–1077

Tang MT, Mi CG (2025) Improving ancient chinese word segmentation with knowledge-enhanced prompting for large language models. International Journal of Intelligent Systems 2025(1):9612240

Wan M, McAuley J (2018) Item recommendation on monotonic behavior chains. In: Proceedings of the 12th ACM conference on recommender systems, pp 86–94

Wang J, Xie H, Zhang S, et al (2025a) Multimodal fusion framework based on knowledge graph for personalized recommendation. Expert Systems with Applications 268:126308

Wang W, Xu Y, Feng F, et al (2023) Diffusion recommender model. In: Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval, pp 832–841

Wang X, Jin H, Zhang A, et al (2020) Disentangled graph collaborative filtering. In: Proceedings of the 43rd international ACM SIGIR conference on research and

development in information retrieval, pp 1001–1010

Wang Y, Shi X, Zhao X (2025b) MLLM4Rec: multimodal information enhancing LLM for sequential recommendation. Journal of Intelligent Information Systems 63(3):745–761

Wayesa F, Leranso M, Asefa G, et al (2023) Pattern-based hybrid book recommendation system using semantic relationships. Scientific Reports 13:3693

Wu J, Wang X, Feng F, et al (2021) Self-supervised graph learning for recommendation. In: Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval, pp 726–735

Wu S, Tang Y, Zhu Y, et al (2019) Session-based recommendation with graph neural networks. In: Proceedings of the AAAI conference on artificial intelligence, pp 346–353

Xu F (2026) Smart library personalized resource proactive recommendation system integrating user profiling and knowledge graphs. Scientific Reports 16:20128

Yang X, Wang Y, Chen J, et al (2025) Dual test-time training for out-of-distribution recommender system. IEEE Transactions on Knowledge and Data Engineering 37(6):3312–3326

Zhang B, Fan C, Liu S, et al (2024a) The expressive power of graph neural networks: A survey. IEEE Transactions on Knowledge and Data Engineering 37(3):1455–1474

Zhang Z, Zhang H, Zhang Z, et al (2024b) Context-embedded hypergraph attention network and self-attention for session recommendation. Scientific Reports 14:19413

Zhao R, Zhang Y, Ju S, et al (2025) Adaptive user multi-level and multi-interest preferences for sequential recommendation. World Wide Web 28(1):20

Zheng Z, Ai T, Lu Y (2026) Graph-augmented and over-smoothing-resistant contrastive clustering for short text. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp 35050–35058

Ziegler CN, McNee SM, Konstan JA, et al (2005) Improving recommendation lists through topic diversification. In: Proceedings of the 14th International Conference on World Wide Web, pp 22–32