

Item Cluster-aware Prompt Learning for Session-based Recommendation

Wooseong Yang
University of Illinois Chicago
Chicago, USA
wyang73@uic.edu

Chen Wang
University of Illinois Chicago
Chicago, USA
cwang266@uic.edu

Zihe Song
University of Illinois Chicago
Chicago, USA
zsong29@uic.edu

Weizhi Zhang
University of Illinois Chicago
Chicago, USA
wzhan42@uic.edu

Philip S. Yu
University of Illinois Chicago
Chicago, USA
psyu@uic.edu

ABSTRACT

Session-based recommendation (SBR) aims to capture dynamic user preferences by analyzing item sequences within individual sessions. However, most existing approaches focus mainly on intra-session item relationships, neglecting the connections between items across different sessions (inter-session relationships), which limits their ability to fully capture complex item interactions. While some methods incorporate inter-session information, they often suffer from high computational costs, leading to longer training times and reduced efficiency. To address these challenges, we propose the CLIP-SBR (Cluster-aware Item Prompt learning for Session-Based Recommendation) framework. CLIP-SBR is composed of two modules: 1) an item relationship mining module that builds a global graph to effectively model both intra- and inter-session relationships, and 2) an item cluster-aware prompt learning module that uses soft prompts to integrate these relationships into SBR models efficiently. We evaluate CLIP-SBR across eight SBR models and three benchmark datasets, consistently demonstrating improved recommendation performance and establishing CLIP-SBR as a robust solution for session-based recommendation tasks.

CCS CONCEPTS

• Information systems → Recommender systems.

KEYWORDS

Recommendation System, Session-based Recommendation, Community Detection, Prompt Learning

1 INTRODUCTION

Session-based recommendation (SBR) has gained increasing attention due to their effectiveness in various online services [29], such as e-commerce, social media, and music platforms. Unlike traditional recommendation systems that model long-term static user

preferences, SBRs focus on short-term dynamic user preferences embedded within sessions, providing more timely and accurate recommendations. Deep learning-based approaches have recently driven significant advances in SBRs by utilizing their exceptional feature representation capabilities to improve the modeling of complex user preferences embedded in sessions. Specifically, recurrent neural networks (RNNs) have been used to capture sequential dependencies within sessions by processing item sequences in order [7, 8]. Several studies employ attention mechanisms to enhance SBRs by selectively focusing on the most relevant items within a session, thereby capturing key patterns in user behavior [9, 14]. Graph neural networks (GNNs) extend these capabilities by representing sessions as graphs, allowing the models to capture intricate relationships between items through graph structures [32, 35].

Despite the progress, most existing SBR approaches primarily focus on the sequential relations between items within a single session (*intra-session* item relationship), often neglecting the valuable information available across multiple sessions (*inter-session* item relationship). This significant **limitation in inter-session modeling** has been identified and proven in multiple studies [16, 22, 24, 30] as a critical issue in SBRs, hindering the accurate learning of user preferences and potentially affecting recommendation performance. To illustrate, consider three sessions for different users: session one {AirPods Pro, AirPods, AirPods Max}, session two {AirPods, Galaxy Buds, Beats Fit Pro}, and session three {Nike Dri-FIT Running Shirt, Adidas Ultraboost, Beats Fit Pro}. When focusing solely on intra-session information, we can observe that the sessions contain Apple earphones, earbuds, and running gears in each session, respectively. While the intra-session perspective provides immediate sequential insights, it struggles to capture more complex item relationships. For instance, the user of session three might also be interested in AirPods Pro instead of Beats Fit Pro, but it's challenging for the model to identify this solely from intra-session data, as there are no common items between session one and session three. In contrast, inter-session information can discover the deeper connections by linking Beats Fit Pro across session two and three, and AirPods across session one and two, bridging the gap between session one and three. This highlights the importance of inter-session information, as it enhances the understanding of user preferences beyond what intra-session information alone can achieve.

Although several SBRs attempt to incorporate inter-session item relationships and some have shown effectiveness, most still face

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/18/06
<https://doi.org/XXXXXXX.XXXXXXX>

limitation in computational efficiency. Specifically, GCSAN [34] strives to utilize inter-session information but fails to fully model complex relationships between items, which limits its overall recommendation performance. In contrast, models like GCEGNN [31] and LESSR [3] demonstrate effectiveness in capturing inter-session item relationships. However, they are constrained by high computational complexity, leading to prolonged training times. These limitations underscore the need for a robust framework that can fully exploit complex item relationships across sessions while balancing both effectiveness and efficiency in SBRs.

To address the aforementioned challenges, we propose **CLuster-aware Item Prompt learning framework for Session-Based Recommendation (CLIP-SBR)** that captures complex item relationships and efficiently embeds the information into the SBR models. To overcome the **limitation in inter-session modeling**, the first module, *item relationship mining*, constructs a global graph from session data. This graph models both intra- and inter-session item relationships in a single structure. Prior studies [3, 31, 32, 34, 35] have demonstrated the effectiveness of graph in capturing item relationships. Inspired by the approaches, the global graph represents items as nodes and item transitions as edges, allowing simultaneous modeling of relationships within and across sessions in one structure. Following graph construction, a community detection method [28] is used to identify item clusters that share similar user preferences. These clusters offer valuable insights, as items within the same cluster exhibit close relationships and similar user preferences.

To tackle the **limitation in computational efficiency**, which is the inefficiency of incorporating complex item relationships, we introduce the second module, *item cluster-aware prompt learning*. they suffer from high computational complexity. This module enhances the learning capabilities of SBR models efficiently by integrating learnable soft prompts for each identified item cluster. Prompt learning [17], originally introduced in natural language processing, began as a way to provide large language models with task-related information through natural language instructions, known as prompts. The concept evolved into learnable prompts, with discrete natural language prompts termed *hard prompt* and continuous, trainable prompts termed *soft prompts*. Soft prompts have demonstrated significant success due to their ability to provide task-related cues to pre-trained models during the tuning phase [13, 15, 25]. This approach has expanded to fields such as computer vision [5, 10, 36] and graph representation learning [4, 19, 26, 27], where soft prompts are used to deliver task-specific and data-related information during training. Notably, GPF [4] and SUPT [12] have proven the ability of soft prompt to deliver data-related information during tuning process. Inspired by the previous studies, we adopt soft prompts to integrate mined item relationship information into SBR models during training. By tailoring learnable soft prompts to specific item cluster, SBR models can effectively and efficiently learn user preferences embedded in session data, capturing both intra- and inter-session relationships. This comprehensive approach not only enhances recommendation performance in an efficient way but also ensures adaptability, making the CLIP-SBR framework universal across various SBR models. We examine CLIP-SBR on eight SBR baselines and three benchmark.

To summarize, the main contributions lie in:

- We propose a graph-based method with community detection to design and extract complex item relationships from session data.
- We propose a novel approach to enhance SBRs by incorporating item cluster-aware prompt into the model training phase.
- To the best of our knowledge, CLIP-SBR is the first attempt to adopt a soft prompt to the SBR. Also, it is the pioneer work to introduce a soft prompt directly in the model training.
- We conduct extensive experiments on five representative SBRs and three benchmark datasets to demonstrate the superiority of the proposed CLIP-SBR framework.

2 RELATED WORKS

2.1 Session-based Recommendation

Session-based Recommendation (SBR) predicts the next user interaction based on recent behavior sequences, effectively capturing short-term user preferences. Recent advancements in deep learning have significantly enhanced SBRs by enabling more sophisticated modeling of item transitions within sessions. GRU4Rec [8] introduced Gated Recurrent Units (GRUs) for sequential dependency modeling, while models like NARM [14] and STAMP [18] leverage attention mechanisms to capture user intent and preference. CORE [9] addresses prediction inconsistencies by unifying session and item embeddings, reducing representation gaps. Graph neural networks (GNNs) have further improved SBRs by representing session sequences as graphs, as seen in SRGNN [34], which aggregates item information through gated GNNs, TAGNN [35] which employs target-aware attention to incorporate user interests. While these methods effectively model intra-session relationships, they often overlook the valuable inter-session information, limiting their ability to fully understand complex user behaviors. To address this, several inter-session SBRs have been proposed. GCSAN [34] uses self-attention on session graphs to capture intricate item relationships, while GCEGNN [31] combines item-level and session-level information using graph convolution and self-attention mechanisms. LESSR [3] introduces an edge-order preserving aggregation scheme and a shortcut graph attention layer to effectively capture global item relationships. Despite their effectiveness, they either inadequately capture inter-session information (GCSAN) or are computationally intensive (GCEGNN, LESSR). Our work proposes incorporating both intra- and inter-session item relationships using efficient cluster-aware prompts to enhance SBR performance.

2.2 Prompt Learning

Prompt learning is initially introduced in natural language processing (NLP) as a novel method to guide pre-trained language models in performing specific tasks by providing task-related information [17]. It originated with the use of discrete hard prompts: text-based instructions designed to provide task related information to pre-trained models [2]. However, due to the limitations of hard prompts that requires human intervention and limited adaptability, the concept has evolved to include continuous soft prompts. Soft prompts are learnable vectors that do not directly correspond to natural language but are optimized during training to enhance model performance on a given task [13, 15, 25]. In recent years, the effectiveness of soft prompts has been demonstrated beyond NLP, extending to domains such as computer vision (CV) and graph

representation learning. In the CV field, visual soft prompts have been well-studied and shown great potential in cross-task generalization, domain adaptation, and visual-language models [5, 10, 36]. Similarly, in the graph field, pioneering works [19, 26] align task-specific soft prompts and downstream tasks, such as link prediction. GPF [4] and SUPT [12] further enhance by introducing soft prompts to every node, thereby providing graph data-related information in tuning phase, enabling better model adaptation to complex graph structures and tasks. As a result, prompt learning has emerged as a versatile tool that can efficiently improve task performance by providing task- or data-related information across various fields.

3 PROPOSED METHOD

The CLIP-SBR aims to exploit both intra- and inter-session item relationships to effectively model the user preferences embedded in sessions and enhance existing SBRs with the information. Figure 1 presents the overall flow of CLIP-SBR, which comprises two main components: 1) *Item Relationship Mining*. This component constructs a graph structure from session data to model item relationships with sufficient depth. It then identifies item clusters based on sequential dependencies by employing existing community detection method. 2) *Item Cluster-aware Prompt Learning*. This component integrates learnable prompt vectors into item embeddings before they are fed into SBR models. The prompt vectors provide SBR models with information about item clusters during training. In this chapter, we first present the problem statement and then introduce the components of CLIP-SBR in detail.

3.1 Problem Statement

Session-based recommendation aims to predict the next item a user will interact with, based on the current session. Let $\mathcal{V} = \{v_1, v_2, \dots, v_{|\mathcal{V}|}\}$ represent the set of all unique items across sessions, where v_k represents the k -th item ($1 \leq k \leq |\mathcal{V}|$) in the sequence. We define a session set $\mathcal{S} = \{S_1, S_2, \dots, S_{|\mathcal{S}|}\}$ and we denote the t -th session ($1 \leq t \leq |\mathcal{S}|$) as $S_t = \{v_1^t, v_2^t, \dots, v_{|S_t|}^t\}$, where $v_l^t \in \mathcal{V}$ is the l -th interacted item ($1 \leq l \leq |S_t|$) in S_t . The objective of session-based recommendation is to predict the top- K items ($1 \leq K \leq |\mathcal{V}|$) that the user is most likely to interact with next. This translates into forecasting the $(|S_t| + 1)$ -th item based on the first $|S_t|$ items in the session.

3.2 Item Relationship Mining

As the first module of the CLIP-SBR framework, *Item Relationship Mining* captures complex relationships between items in sessions using a graph-based approach that models both intra- and inter-session item relationships, inspired by prior SBR studies [3, 31, 32, 34, 35]. We first construct session graphs to capture sequential dependencies within individual sessions. To incorporate inter-session relationships, we create a global graph by connecting item nodes if they are linked in any session graph, enabling a unified view of both intra- and inter-session dynamics. By applying a community detection algorithm to this global graph, we identify item clusters—groups of items with strong interconnections and shared user preferences. This clustering step is crucial for uncovering latent patterns and user behaviors, ultimately enhancing the accuracy and efficiency of the SBR training process.

3.2.1 Graph Construction. The graph construction process is a foundational step in our framework, designed to model the complex relationships between items in SBRs. We begin by constructing session graphs to capture intra-session item relationships. These session graphs serve as the basis for building a global graph, which captures inter-session item relationships, thereby providing a comprehensive representation of item dependencies across sessions.

Session Graph Construction. In this step, we transform each session sequence into a session graph, denoted as $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$. Here, $\mathcal{V}_t \subseteq \mathcal{V}$ represents the set of items that a user interacts with within session S_t . The edge set $\mathcal{E}_t = \{(v_i^t, v_j^t) \mid v_i^t, v_j^t \in \mathcal{V}_t\}$ captures the relationships between these items. Each edge e_{ij}^t is undirected and unweighted, capturing the sequential dependencies between adjacent items v_i^t and v_j^t within the session. This session graph effectively models intra-session item relationships by highlighting the item transitions within each session as edges.

Global Graph Construction. To capture inter-session item relationships, we construct a global graph by aggregating session graphs with inter-session connections. This global graph, denoted as $\mathcal{G}_g = (\mathcal{V}_g, \mathcal{E}_g)$, includes a node for each unique item in the dataset, where $\mathcal{V}_g = \mathcal{V}$. In this graph, edges are undirected and weighted, with the weight of an edge representing the frequency of item co-occurrences across different sessions. The graph aggregates connections from multiple session graphs, assigning the count of co-occurrences as the weight for each edge. Formally, the edge set is defined as $\mathcal{E}_g = \{(v_i^g, v_j^g) \mid v_i^g, v_j^g \in \mathcal{V}_g\}$, where each edge e_{ij}^g is associated with a weight w_{ij}^g , which quantifies the strength of the relationship between items v_i^g and v_j^g based on their co-occurrence frequency. The constructed global graph plays a crucial role in our framework by modeling both intra-session and inter-session item relationships in a single structure. We utilize this graph representation to detect clusters of items that are closely related.

3.2.2 Item Cluster Detection. The process of detecting item clusters within the global graph \mathcal{G}_g plays a crucial role in capturing the complex relationships between items. In this study, a *cluster* refers to a subset of nodes within the graph that are more densely connected to each other than to nodes outside the subset. This high internal connectivity helps uncover latent item dependencies and user behavior patterns, which are essential for modeling complex item relationships embedded in session data. The goal of this step is to identify item clusters that exhibit strong relationships, which are critical for capturing nuanced user preferences in session-based recommendation. To achieve this, we employ the Leiden algorithm [28], a community detection method known for its superior performance in handling large-scale and complex graphs. The algorithm enhances the Louvain algorithm [1], by improving detection quality, accelerating convergence speed, and increasing robustness against noise and resolution limits. Furthermore, it effectively handles unconnected and large-scale graphs, both of which are common in recommendation settings, making the Leiden algorithm particularly suitable for our task of item cluster detection.

We apply the Leiden algorithm to the global graph \mathcal{G}_g to partition items into clusters:

$$\text{Partition} = \text{Leiden}(\mathcal{G}_g) = \{(v_i^g : c_m) \mid v_i^g \in \mathcal{V}_g, c_m \in C\}. \quad (1)$$

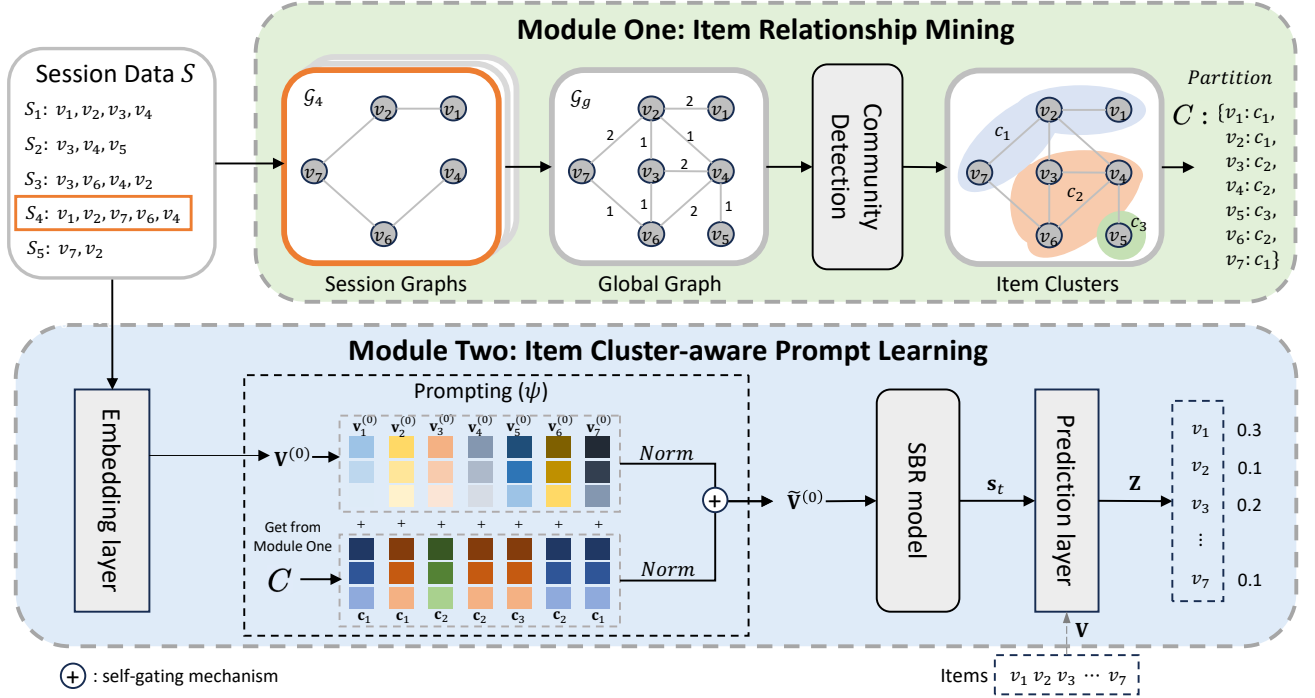


Figure 1: The proposed CLIP-SBR framework consists of two main modules. The *Item Relationship Mining* module constructs session graphs and combines them into a global graph to capture intra- and inter-session item relationships, followed by community detection to identify item clusters. The *Item Cluster-aware Prompt Learning* module enhances SBR models by integrating learnable soft prompts tailored to these clusters, embedding cluster-specific information to improve recommendation accuracy and efficiency.

Here, $\text{Leiden}(\mathcal{G}_g)$ represents the execution of the Leiden algorithm on the global graph \mathcal{G}_g , yielding a dictionary-like structure *Partition*, in which each item v_i^g is assigned to a specific cluster c_m . The set of detected clusters is denoted as $C = \{c_1, c_2, \dots, c_{|C|}\}$, where each c_m represents a distinct cluster, and $|C|$ is the total number of clusters. Identifying item clusters enables us to capture intricate item relationships that are overlooked in many previous SBR studies [8, 9, 14, 32, 35]. By leveraging these clusters, we can better tailor the SBR process to reflect underlying user preferences and interactions across sessions. The identified item clusters are subsequently used in the second module of our framework, where they play a crucial role in integrating item cluster-aware prompt learning into session-based recommendation models.

3.3 Item Cluster-aware Prompt Learning

In this section, we introduce a method to integrate the mined information about item relationships into SBR models during training. By leveraging learnable prompts, which have proven effective in providing data-related information to models, we enhance the capability of SBRs to capture both intra- and inter-session item relationships.

First, session data is fed into the embedding layer of the existing SBR to obtain the initial item embeddings. Formally, we represent

the initial item embeddings as:

$$\mathbf{V}^{(0)} = \text{EmbeddingLayer}(\mathcal{V}), \quad (2)$$

where $\mathbf{V}^{(0)} = \{v_1^{(0)}, v_2^{(0)}, \dots, v_{|\mathcal{V}|}^{(0)}\}$ represents the set of initial item embeddings for the items in \mathcal{V} . Here, we assign a unique and randomly initialized learnable vector (soft prompt), called *cluster prompt*, to each detected item cluster. The cluster prompt for a cluster c_m is denoted as c_m . For each item v_k in \mathcal{V} , we retrieve its corresponding cluster prompt using the partition from the Leiden algorithm:

$$c_m = \text{Partition}(v_k), \quad (3)$$

where c_m represents the cluster to which item v_k belongs. The set of all cluster prompts is defined as $C = \{c_1, \dots, c_{|C|}\}$.

Next, we first normalize the initial item embedding $v_k^{(0)}$ and the cluster prompt c_m to ensure balanced contributions. This step prevents the cluster prompt from overwhelming the item embedding and enables the model to learn balanced information. The normalized item embedding $\hat{v}_k^{(0)}$ and normalized cluster prompt \hat{c}_m are then combined using a self-gating mechanism:

$$\psi(\hat{v}_k^{(0)}) = g_k \cdot \hat{v}_k^{(0)} + (1 - g_k) \cdot \hat{c}_m, \quad (4)$$

where $\hat{v}_k^{(0)} = \frac{v_k^{(0)}}{\|v_k^{(0)}\|}$ and $\hat{c}_m = \frac{c_m}{\|c_m\|}$ represent the normalized versions of the initial item embedding and cluster prompt, respectively, and ψ denotes the prompting function. The scalar g_k is a gating variable that controls the contribution of the item embedding and the cluster prompt. This self-gating mechanism dynamically balances the influence of both components based on their relevance, ensuring that the cluster prompt does not dominate and allowing the model to learn effectively.

The set of prompted item embeddings for $\tilde{\mathbf{V}}^{(0)}$ is defined as:

$$\tilde{\mathbf{V}}^{(0)} = \{\psi(\hat{v}_1^{(0)}), \psi(\hat{v}_2^{(0)}), \dots, \psi(\hat{v}_{|\mathcal{V}|}^{(0)})\}. \quad (5)$$

During the test phase, sessions in test dataset may contain unseen items. If a session includes both unseen items and items that are present in the global graph, new edges are added to the global graph to integrate the unseen items. If all items in a session are unseen, we assign the most frequent cluster to these items.

Prompted item embeddings are passed through the existing SBR model during training, providing information about item relationships. Formally, given sessions \mathcal{S} and prompted item embeddings $\tilde{\mathbf{V}}^{(0)}$, we apply the SBR model f to obtain session embeddings:

$$\mathbf{S}, \mathbf{V} = f(\mathcal{S}, \tilde{\mathbf{V}}^{(0)}), \quad (6)$$

where $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{|\mathcal{S}|}\}$ is a set of session embeddings with t -th element \mathbf{s}_t , and $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{|\mathcal{V}|}\}$ is a set of updated item embeddings with k -th element \mathbf{v}_k .

Next, we move to the prediction layer. In this layer, all items in \mathcal{V} are considered candidate items for recommendation. Let \mathbf{Z} represent the set of recommendation scores, where each element \mathbf{z}_k corresponds to the score for the target item $v_k \in \mathcal{V}$. Given session \mathcal{S}_t , the score \mathbf{z}_k is computed by taking the inner product between the session embedding and the item embedding:

$$\mathbf{z}_k = \mathbf{s}_t^\top \mathbf{v}_k, \quad (7)$$

where \mathbf{v}_k is the updated item embedding for item v_k . The softmax function is then applied to the unnormalized scores \mathbf{z}_k to obtain the final output probabilities:

$$\hat{y}_k = \text{Softmax}(\mathbf{z}_k), \quad (8)$$

where \hat{y}_k denotes the probability of item v_k being the next click in the current session. The loss function used for training is defined as the cross-entropy loss between the predicted probabilities \hat{y} and the one-hot encoded ground truth labels y . The cross-entropy loss measures the difference between the true distribution (given by the one-hot encoding) and the predicted distribution (given by the model's output). Formally, the loss is defined as:

$$\mathcal{L}(y, \hat{y}) = - \sum_{k=1}^{|\mathcal{V}|} y_k \log(\hat{y}_k), \quad (9)$$

where y_k is the one-hot encoded ground truth for item v_k (with $y_k = 1$ for the correct item and 0 for all others), and \hat{y}_k is the predicted probability that item v_k is the correct next item.

By incorporating learnable cluster prompts into item embeddings, our approach enhances the SBR model's ability to capture complex item dependencies, resulting in more accurate and effective recommendations.

Table 1: Statistics of datasets.

Statistic	Last.fm	Xing	Reddit
No. of users	966	11,399	13,850
No. of items	38,784	59,039	21,790
No. of sessions	294,402	90,286	458,292
Avg. of session length	12.86	5.82	4.49
Session per user	304.76	7.92	33.09

4 EXPERIMENTS

We have conducted extensive experiments to evaluate the proposed CLIP-SBR by answering the following four key research questions:

- **RQ1:** Does CLIP-SBR improve the recommendation performance of SBR baseline models?
- **RQ2:** Does the *Item Relationship Mining* module effectively capture and utilize complex item relationships?
- **RQ3:** Does the *Item Cluster-aware Prompt Learning* module effectively and efficiently integrate item relationships into the recommendation process effectively and efficiently?
- **RQ4:** How do different hyper-parameter settings impact the effectiveness of the CLIP-SBR framework?

4.1 Experimental Setup

4.1.1 Datasets. We employ three benchmark datasets that are widely used in the session-based recommendation.

- *Last.fm*¹ contains the complete listening behavior of approximately 1,000 users collected from Last.fm. In this paper, we focus on the music artist recommendation. We consider the top 40,000 most popular artists and group interaction records within an 8-hour window from the same user as a session, following former studies [3, 6].
- *Xing*² gathers job postings from a social networking platform, including interactions with the postings by 770,000 users. We split user's records into sessions following [23].
- *Reddit*³ is a dataset collected from social media that includes tuples consisting of a user name, a subreddit where the user commented on a thread, and a timestamp of the interaction. The interaction data was segmented into sessions using a 60-minute time threshold, as outlined in [20].

Using the preprocessed data provided in [21] as a basis, we further process by following the previous studies [3, 23, 31], removing sessions with fewer than three interactions to exclude less informative data. Additionally, we retain only users with five or more sessions to ensure sufficient historical data. We allocated 10% of the sessions as the test set and the penultimate 10% as the validation set, with the remaining sessions used for training. Furthermore, for a session $\mathcal{S}_t = \{v_1^t, v_2^t, \dots, v_{|\mathcal{S}_t|}^t\}$, we generate sequences and corresponding labels by a sequence splitting preprocessing, i.e., $(\{v_1^t\}, v_2^t), (\{v_1^t, v_2^t\}, v_3^t), \dots, (\{v_1^t, v_2^t, \dots, v_{|\mathcal{S}_t|-1}^t\}, v_{|\mathcal{S}_t|}^t)$ for training, valid and test across all the three datasets. The statistics of datasets, after preprocessing, are summarized in Table 1.

¹<http://ocelma.net/MusicRecommendationDataset/lastfm-1K.html>

²<http://2016.recsyschallenge.com/>

³<https://www.kaggle.com/colemaclean/subreddit-interactions>

Table 2: Comparison of MRR@5, Recall@5 and average improvement (%) on the three datasets. Our proposed models are highlighted in gray. The best results are highlighted in boldface, and the best results among *intra-session* models are underlined.

Category	Model	Last.fm			Xing			Reddit		
		MRR@5	Recall@5	Imp.(%)	MRR@5	Recall@5	Imp.(%)	MRR@5	Recall@5	Imp.(%)
intra-session	GRU4Rec	7.70	13.39	-	10.33	15.71	-	35.45	46.40	-
	CLIP-GRU4Rec	10.07	17.24	+29.77	11.12	16.90	+7.61	40.21	50.47	+11.10
	NARM	5.50	9.74	-	9.66	15.41	-	33.08	44.48	-
	CLIP-NARM	8.10	13.96	+45.30	12.04	17.81	+20.11	38.50	48.79	+13.04
	CORE	10.51	17.70	-	10.27	17.79	-	29.33	46.54	-
	CLIP-CORE	10.67	18.11	+1.92	12.65	19.59	+16.65	31.33	47.56	+4.51
	SRGNN	9.02	14.66	-	14.03	19.37	-	38.04	47.75	-
	CLIP-SRGNN	12.12	18.87	+31.54	14.18	19.72	+1.44	39.70	49.73	+4.26
	TAGNN	9.25	15.30	-	12.68	18.37	-	40.07	51.10	-
	CLIP-TAGNN	12.82	19.77	+33.91	<u>14.70</u>	<u>21.07</u>	+15.31	41.22	52.01	+2.33
inter-session	GCSAN	4.47	8.69	-	13.02	16.82	-	36.03	44.13	-
	CLIP-GCSAN	7.05	11.59	+45.54	15.65	20.34	+20.56	39.07	48.21	+8.84
	GCEGNN	10.81	17.32	-	12.83	19.41	-	38.17	48.38	-
	CLIP-GCEGNN	11.05	17.60	+1.92	14.36	21.58	+11.55	37.93	48.19	-0.51
	LESSR	7.06	11.05	-	14.67	18.96	-	39.05	47.90	-
	CLIP-LESSR	6.85	10.96	-1.89	14.94	19.07	+1.21	39.07	48.21	+0.35

Many prior studies use sampled metrics by ranking relevant items alongside a smaller subset of randomly selected items to accelerate calculations. However, this sampling technique can lead to discrepancies compared to the full, unsampled counterparts [11]. Therefore, in line with previous works [33], we assess each method on the entire item set without sampling, ranking all items that the user has not interacted with based on their similarity scores.

4.1.2 Baselines. To evaluate the effectiveness of our approach, we compare its performance against a diverse set of well-established methods. The first five baselines focus solely on *intra-session* information, while the last three incorporate both *intra-session* and *inter-session* information. We assess our method across various model architectures, including those based on RNNs, GNNs, and attention mechanisms, to demonstrate the applicability and effectiveness of our method across different model architectures.

- **GRU4Rec** [8] is a RNN-based model that employs the Gated Recurrent Unit (GRU) to model interaction sequences.
- **NARM** [14] is a RNN-based model that improves over GRU4Rec by applying an attention mechanism alongside GRU.
- **CORE** [9] unifies the representation space for both encoding and decoding processes by using a representation-consistent encoder ensuring that session embeddings and item embeddings are in the same space.
- **SRGNN** [32] transforms session sequences into directed unweighted graphs and applies a gated graph convolutional layer to identify item transition patterns.
- **TAGNN** [35] is a variant of SRGNN that utilizes target-aware attention network to generate session embeddings by incorporating the features of the candidate item during prediction.
- **GCSAN** [34] extracts local context information using a GGNN and then employs a self-attention mechanism to capture explicit dependencies.

- **GCEGNN** [31] integrates both global context and item sequences from the current session to produce session embeddings through multi-level graph neural networks.
- **LESSR** [3] introduces shortcut connections between items within a session and incorporates sequence information into graph convolution with the help of GRU.

4.1.3 Evaluation Metrics. To evaluate the performance of the recommendation, we use two metrics widely adopted in SBR evaluation: Mean Reciprocal Rank (MRR@ k) and Recall (Recall@ k) following [14, 23, 24], where $k = 5$.

4.1.4 Implementation Details. We implement CLIP-SBR using a popular open-source recommender systems library *RecBole*⁴. All methods are optimized with the Adam optimizer, and we employ early stopping with a patience of 50 epochs to prevent overfitting, using MRR@5 as the indicator. The batch size is tuned from {64, 128, 256, 512, 1024} and learning rate from {0.001, 0.005, 0.01, 0.05}. For the Leiden algorithm used in item cluster detection, the resolution value is set to 1, with its impact discussed in detail in Section 4.2.4.

4.2 Results

4.2.1 Overall Performance (RQ1). Table 2 presents the experimental results of the eight baselines and their corresponding implementations using the CLIP-SBR framework on three real-world datasets. The best result in each column is highlighted in bold, and the top result within the *intra-session* category is underlined. We applied the CLIP-SBR framework to both *intra-session* and *inter-session* baselines to assess its impact. The results show that the CLIP-SBR leads to performance improvements across most of the baselines.

In the *intra-session* baselines, we observed substantial improvements across the board, with the most significant performance gains in the NARM model. For the *inter-session* baselines, CLIP-SBR demonstrated notable improvements in GCSAN, while having a

⁴<https://recbole.io>

smaller impact on GCEGNN and no improvement on LESSR. Further analysis of the impact of applying CLIP-SBR to inter-session baselines is provided in Section 4.2.3.

Moreover, the *inter-session* baselines (GCSAN, GCEGNN, LESSR) demonstrate comparable or even better performance than the *intra-session* baselines. Notably, GCEGNN outperforms all *intra-session* models on the Last.fm dataset, and LESSR achieves the highest metrics on the Xing dataset. These results highlight the potential of incorporating inter-session information to improve recommendation accuracy, emphasizing the importance of considering both intra-session and inter-session item relationships when developing robust session-based recommendation models.

The application of the CLIP-SBR framework to the *intra-session* baselines resulted in performance improvements in all but one case. Specifically, NARM saw a substantial improvement of 45.30% on Last.fm, 20.11% on Xing, and 13.04% on Reddit, marking it as one of the most improved models. TAGNN also demonstrated significant gains, particularly on Last.fm with a 33.91% increase, 15.31% on Xing, and 2.33% on Reddit. Similarly, GRU4Rec achieved notable enhancements of 29.77% on Last.fm, 7.61% on Xing, and 11.10% on Reddit. CORE exhibited improvements of 1.92%, 16.65%, and 4.51% across the respective datasets, while SRGNN showed increases of 31.54%, 1.44%, and 4.26% on the three datasets. When comparing intra-session CLIP-SBR models with inter-session models, the CLIP-SBR models generally outperformed in many instances. Notably, CLIP-TAGNN consistently surpassed all inter-session models across. CLIP-SRGNN also outperformed inter-session models on Last.fm and Reddit dataset, and outperformed GCSAN in MRR@5 and Recall@5, and GCEGNN in MRR@5. This indicates that the CLIP-SBR framework is highly effective in leveraging and designing inter-session information, leading to better overall performance.

We also applied CLIP-SBR to the *inter-session* models to assess its effectiveness. The impact varied across the three models: it had a substantial effect on GCSAN, a slight effect on GCEGNN, and almost no effect on LESSR. This suggests that the CLIP-SBR is less beneficial for GCEGNN that already capture inter-session information effectively, while providing significant improvements for GCSAN that do not sufficiently address it. For LESSR, the lack of improvement indicates that the model may not effectively utilize the additional information provided by CLIP-SBR. The impact of CLIP-SBR on *inter-session* models is further discussed in Section 4.2.3.

4.2.2 Impact of Item Relationship Mining (RQ2). We evaluated the impact of the *Item Relationship Mining* module in CLIP-SBR by assessing performance across seven different scenarios.

- **C:** The original CLIP-SBR applies distinct learnable prompts for each item's cluster, embedding them into the item representations. This allows the model to capture group-level item relationships and patterns across sessions.
- **U:** User-specific prompts are added to item embeddings, maintaining a consistent prompt for a given user across all sessions. This helps the model capture user preferences and long-term behavioral patterns across sessions.
- **S:** Session-specific prompts are applied to all items within a session, helping the model learn session-level patterns that reflect user intent or context within a specific session.

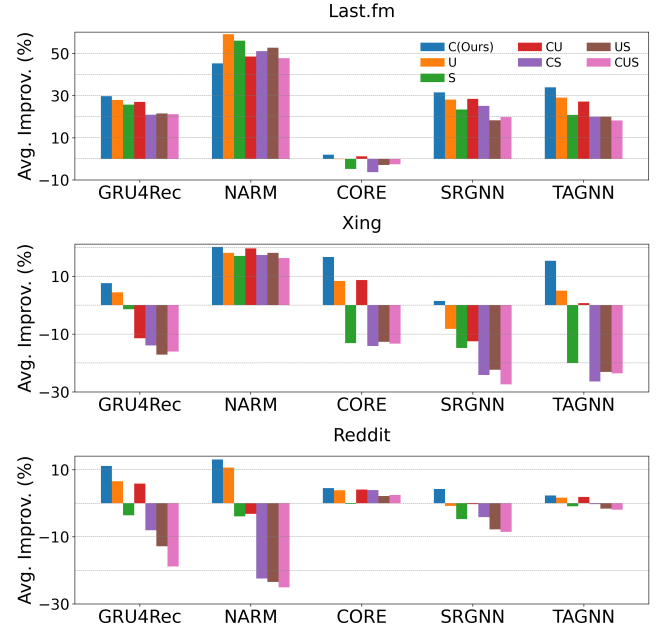


Figure 2: Improvement (%) comparison of CLIP-SBR variants on the three datasets. The blue bar represents the original CLIP-SBR, while the other colors indicate various CLIP-SBR variants. "Avg. Improv." refers to the average of improvement percentage calculated from MRR@5 and Recall@5.

- **CU:** Combines both cluster and user-specific prompts. Cluster prompts capture relational structure among items, while user prompts represent long-term user preferences, enabling a deeper understanding of user behavior and item relations.
- **CS:** Combines cluster and session-specific prompts, adding both to item embeddings. This captures the dynamic context of the session (via session prompts) while preserving item relationships within clusters (via cluster prompts).
- **US:** Combines user and session-specific prompts to balance short-term session intent with long-term user preferences. This enables the model to capture how user behavior shifts across sessions while keeping user-specific patterns maintained.
- **CUS:** Combines cluster, user, and session-specific prompts, creating a comprehensive embedding. This setup captures group-level item relationships, user preferences across sessions, and dynamic session behavior simultaneously.

We trained these seven types of models for each baselines on each dataset and evaluated their performance improvement using MRR@5 and Recall@5 metrics. The results, as shown in Figure 2, reveal that in almost all cases, the original CLIP-SBR models with cluster prompts (C) exhibited the most significant improvement compared to the baseline models. Notably, models using user-specific prompts (U) showed the second highest improvement, indicating that incorporating user-specific prompts effectively provided personalizing information during the training phase, thereby enhancing the performance of the SBR models. In contrast, other scenarios, especially those involving combinations of prompts, did not yield

Table 3: Comparison of MRR@5 and average training time (in seconds) for the proposed methods and inter-session baselines on the three datasets. Our proposed models are highlighted in gray.

Model	Last.fm		Xing		Reddit	
	MRR@5	Time	MRR@5	Time	MRR@5	Time
CLIP-GRU4Rec	10.07	745.11	11.12	6.51	40.21	30.66
CLIP-NARM	8.10	791.55	12.04	6.97	38.50	34.26
CLIP-CORE	10.67	660.77	12.65	14.72	31.33	82.55
CLIP-SRGNN	12.12	619.79	14.18	56.33	39.70	196.71
CLIP-TAGNN	12.82	1416.41	14.70	197.62	41.22	290.83
GCSAN	4.47	1032.88	13.02	92.70	36.03	435.70
GCEGNN	10.81	625.71	12.83	60.60	38.17	260.64
LESSR	7.06	758.59	14.67	57.21	39.05	382.98

similar improvements and often resulted in performance decreases. This decline can be attributed to the added prompts acting as noise, disrupting the training of the SBR models. For instance, models with combined prompts showed inconsistent or negative performance, as observed in the CORE on the Last.fm dataset and most models on the Xing dataset. This effect is even more pronounced on the Reddit dataset, where most models with combined prompts experienced a decline in performance. In summary, among all prompt-based approaches, cluster prompts (C) proved to be the most effective. This outcome suggests that the *Item Relationship Mining* module successfully identifies item communities that share similar user preferences by effectively considering both intra- and inter-session item relationships. These findings highlight the superiority of the module in mining item communities within sessions and underscore its critical role in the CLIP-SBR framework, confirming its effectiveness in enhancing session-based recommendation.

4.2.3 Impact of Item Cluster-aware Prompt Learning (RQ3). To evaluate the impact of the *Item Cluster-aware Prompt Learning* module, we assessed its effectiveness on *intra-session* and *inter-session* SBR models. As shown in Table 2 and Figure 2, our approach demonstrates substantial performance improvements, highlighting both the effectiveness and efficiency of the module. In terms of effectiveness, applying the *Item Cluster-aware Prompt Learning* module led to significant performance improvements in *inter-session* baselines. Specifically, CLIP-GCSAN showed notable gains, effectively compensating for GCSAN’s limitations in capturing inter-session information, particularly on the Last.fm dataset. For GCEGNN, the module provided moderate improvements, suggesting that while GCEGNN already captures inter-session information effectively, the prompts still offer additional value. Conversely, CLIP-LESSR showed no improvement over the original LESSR, indicating that LESSR’s inherent handling of inter-session information leaves little room for further enhancement from the prompts. Regarding efficiency, the *Item Cluster-aware Prompt Learning* module demonstrated significant advantages in terms of training time. As shown in Table 3, several CLIP models achieved better or comparable performance (MRR@5) with significantly less training time compared

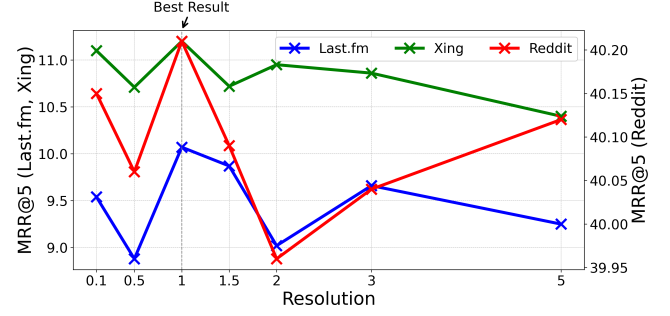


Figure 3: Performance comparison of CLIP-GRU4Rec with different resolution parameters on the three datasets. The MRR@5 scale for Last.fm and Xing is shown on the left y-axis, while the scale for Reddit is on the right y-axis.

to the inter-session baselines (GCSAN, GCEGNN, LESSR). Among the CLIP models, CLIP-SRGNN shows a good balance between effectiveness and efficiency. It outperformed the inter-session baselines on Last.fm and Reddit, achieving higher MRR@5 with reduced training times—619.79 seconds on Last.fm and 196.71 seconds on Reddit, compared to GCSAN’s 1032.88 seconds, GCEGNN’s 625.71 seconds, and LESSR’s 758.59 seconds on Last.fm, as well as the significantly longer times of the inter-session baselines on Reddit. For the Xing dataset, CLIP-SRGNN achieved comparable performance to the inter-session models while still requiring considerably less time—56.33 seconds versus 92.70 seconds for GCSAN and 60.60 seconds for GCEGNN. The CLIP-TAGNN achieves the best performance on MRR@5 at the expense of longer training time. This highlights the efficiency of the *Item Cluster-aware Prompt Learning* module, which achieves comparable or superior performance with reduced training time.

4.2.4 Impact of hyperparameter. The *Item Relationship Mining* module in CLIP-SBR uses the Leiden algorithm for item cluster detection, with the resolution parameter controlling the granularity of community detection. We evaluated CLIP-GRU4Rec using various resolution values 0.1, 0.5, 1, 1.5, 2, 3, 5, with the MRR@5 results shown in Figure 3. Although no clear pattern emerged regarding how resolution changes affected performance, the model consistently performed best with a resolution of 1 across all datasets. Based on these findings, we use a resolution value of 1 in our implementation, as it effectively balances fine-grained and broader item relationships.

5 CONCLUSION

In this paper, we introduced the CLuster-aware Item Prompt learning framework for Session-Based Recommendation (CLIP-SBR), aimed at overcoming the limitations of existing SBR methods by effectively capturing both intra- and inter-session item relationships. The framework comprises two key modules: the first module identifies item clusters from session data, while the second module incorporates cluster-specific prompts to enhance the learning capabilities of SBR models. Extensive experiments demonstrated that CLIP-SBR consistently outperforms baseline models, highlighting its effectiveness.

REFERENCES

- [1] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.
- [2] Tom B Brown. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165* (2020).
- [3] Tianwen Chen and Raymond Chi-Wing Wong. 2020. Handling information loss of graph neural networks for session-based recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 1172–1180.
- [4] Taoran Fang, Yunchao Zhang, Yang Yang, Chunping Wang, and Lei Chen. 2024. Universal prompt tuning for graph neural networks. *Advances in Neural Information Processing Systems* 36 (2024).
- [5] Chunjiang Ge, Rui Huang, Mixue Xie, Zihang Lai, Shiji Song, Shuang Li, and Gao Huang. 2023. Domain adaptation via prompt learning. *IEEE Transactions on Neural Networks and Learning Systems* (2023).
- [6] Lei Guo, Hongzhi Yin, Qinyong Wang, Tong Chen, Alexander Zhou, and Nguyen Quoc Viet Hung. 2019. Streaming session-based recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 1569–1577.
- [7] Balázs Hidasi and Alexandros Karatzoglou. 2018. Recurrent Neural Networks with Top-k Gains for Session-based Recommendations. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (Torino, Italy) (CIKM '18)*. Association for Computing Machinery, New York, NY, USA, 843–852. <https://doi.org/10.1145/3269206.3271761>
- [8] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2–4, 2016, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1511.06939>
- [9] Yupeng Hou, Binbin Hu, Zhiqiang Zhang, and Wayne Xin Zhao. 2022. Core: simple and effective session-based recommendation within consistent representation space. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*. 1796–1801.
- [10] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. 2022. Visual prompt tuning. In *European Conference on Computer Vision*. Springer, 709–727.
- [11] Walid Krichene and Steffen Rendle. 2020. On sampled metrics for item recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 1748–1757.
- [12] Junhyun Lee, Woosong Yang, and Jaewoo Kang. 2024. Subgraph-level Universal Prompt Tuning. *arXiv preprint arXiv:2402.10380* (2024).
- [13] Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The Power of Scale for Parameter-Efficient Prompt Tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 3045–3059.
- [14] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural attentive session-based recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1419–1428.
- [15] Xiang Lisa Li and Percy Liang. 2021. Prefix-Tuning: Optimizing Continuous Prompts for Generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 4582–4597.
- [16] Zihao Li, Xianzhi Wang, Chao Yang, Lina Yao, Julian McAuley, and Guandong Xu. 2023. Exploiting explicit and implicit item relationships for session-based recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 553–561.
- [17] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.
- [18] Qiao Liu, Yifu Zeng, Refuoe Mokhosi, and Haibin Zhang. 2018. STAMP: short-term attention/memory priority model for session-based recommendation. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1831–1839.
- [19] Zemin Liu, Xingtong Yu, Yuan Fang, and Xinming Zhang. 2023. Graphprompt: Unifying pre-training and downstream tasks for graph neural networks. In *Proceedings of the ACM Web Conference 2023*. 417–428.
- [20] Malte Ludewig, Noemi Mauro, Sara Latifi, and Dietmar Jannach. 2019. Performance comparison of neural and non-neural approaches to session-based recommendation. In *Proceedings of the 13th ACM conference on recommender systems*. 462–466.
- [21] Yitong Pang, Lingfei Wu, Qi Shen, Yiming Zhang, Zhihua Wei, Fangli Xu, Ethan Chang, Bo Long, and Jian Pei. 2022. Heterogeneous global graph neural networks for personalized session-based recommendation. In *Proceedings of the fifteenth ACM international conference on web search and data mining*. 775–783.
- [22] Ruihong Qiu, Zi Huang, Jingjing Li, and Hongzhi Yin. 2020. Exploiting cross-session information for session-based recommendation with graph neural networks. *ACM Transactions on Information Systems (TOIS)* 38, 3 (2020), 1–23.
- [23] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing session-based recommendations with hierarchical recurrent neural networks. In *proceedings of the Eleventh ACM Conference on Recommender Systems*. 130–137.
- [24] Massimiliano Ruocco, Ole Steinar Lillestøl Skrede, and Helge Langseth. 2017. Inter-session modeling for session-based recommendation. In *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems*. 24–31.
- [25] Taylor Shin, Yasaman Razeghi, Robert L Logan IV, Eric Wallace, and Sameer Singh. 2020. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 4222–4235.
- [26] Mingchen Sun, Kaixiong Zhou, Xin He, Ying Wang, and Xin Wang. 2022. Gppt: Graph pre-training and prompt tuning to generalize graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1717–1727.
- [27] Xiangguo Sun, Hong Cheng, Jia Li, Bo Liu, and Jihong Guan. 2023. All in one: Multi-task prompting for graph neural networks. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2120–2131.
- [28] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports* 9, 1 (2019), 5233.
- [29] Shoujin Wang, Longbing Cao, Yan Wang, Quan Z Sheng, Mehmet A Orgun, and Defu Lian. 2021. A survey on session-based recommender systems. *ACM Computing Surveys (CSUR)* 54, 7 (2021), 1–38.
- [30] Ting-Yun Wang, Chiao-Ting Chen, Ju-Chun Huang, and Szu-Hao Huang. 2023. Modeling cross-session information with multi-interest graph neural networks for the next-item recommendation. *ACM Transactions on Knowledge Discovery from Data* 17, 1 (2023), 1–28.
- [31] Ziyang Wang, Wei Wei, Gao Cong, Xiao-Li Li, Xian-Ling Mao, and Minghui Qiu. 2020. Global context enhanced graph neural networks for session-based recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 169–178.
- [32] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 346–353.
- [33] Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. 2022. Contrastive learning for sequential recommendation. In *2022 IEEE 38th international conference on data engineering (ICDE)*. IEEE, 1259–1273.
- [34] Chengfeng Xu, Pengpeng Zhao, Yanchi Liu, Victor S Sheng, Jiajie Xu, Fuzhen Zhuang, Junhua Fang, and Xiaofang Zhou. 2019. Graph contextualized self-attention network for session-based recommendation.. In *IJCAI*, Vol. 19. 3940–3946.
- [35] Feng Yu, Yanqiao Zhu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2020. TAGNN: Target attentive graph neural networks for session-based recommendation. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 1921–1924.
- [36] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. 2022. Learning to prompt for vision-language models. *International Journal of Computer Vision* 130, 9 (2022), 2337–2348.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009