ScoreNet: Consistency-driven Framework with Multi-side Information Fusion for Session-based Recommendation

Piao Tong[1], **Qiao Liu**[1*], **Zhipeng Zhang**[2], **Yuke Wang**[1], **Tian Lan**[1]
[1]University of Electronic Science and Technology of China
[2] Byte Dance
{piaot, yuke}@std.uestc.edu.cn, zhangzhipeng.work@bytedance.com, {qliu, lantian1029}@uestc.edu.cn

Abstract

Fusing side information in session-based recommendation is crucial for improving the performance of next-item prediction by providing additional context. Recent methods optimize attention weights by combining item and side information embeddings. However, semantic heterogeneity between item IDs and side information introduces computational noise in attention calculation, leading to inconsistencies in user interest modeling and reducing the accuracy of candidate item scores. These methods also often fail to leverage session-based re-interaction patterns, limiting improvements in score prediction during the decoding phase. To address these challenges, we propose **ScoreNet**, a consistency-driven framework with multi-side information fusion for session-based recommendation. ScoreNet explicitly models users' persistent preferences, generating consistent decoding scores for candidate items within a unified framework. It incorporates a multi-path re-engagement network to capture re-interaction behavior patterns in a semantic-agnostic manner, enhancing side information fusion while avoiding semantic interference. Additionally, a position-enhanced consistent scoring network redistributes attention scores within sessions, improving prediction accuracy, especially for items with limited interactions. Extensive experiments on three real-world datasets demonstrate that ScoreNet outperforms state-of-the-art models.

Introduction

Session-based recommender systems (SBRs) aim to predict the next item in personalized recommendations by analyzing users' recent interactions within anonymous sessions (Wang et al. 2021). Deep learning methods have significantly advanced the modeling of item transition patterns and the learning of session embeddings that capture user preferences (Hidasi et al. 2015; Li et al. 2017; Ren et al. 2019; Wu et al. 2019; Qiu et al. 2019; Xu et al. 2019; Chen and Wong 2020; Kang and McAuley 2018; Sun et al. 2019a; Zhou et al. 2020; Wang et al. 2024). Recently, there has been increasing interest in utilizing various side information (*e.g.*, category, brand) to enhance recommendation relevance by providing richer contextual information (Xie, Zhou, and Kim 2022). As a result, side information fusion in SBRs has garnered significant attention in both academia and industry (Liu et al. 2021; Wang et al. 2022, 2023; Liu et al. 2024).

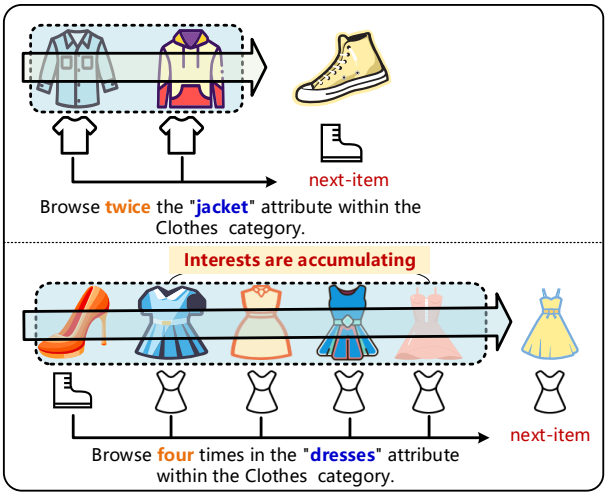


Figure 1: Motivating example. The figure illustrates users' persistent interest in items sharing similar side information, suggesting the potential enhancement for SBRs by fusing side information to model user-item re-interactions.

Current approaches integrating side information with self-attention mechanisms focus on improving the distribution of attention weights across session items to more accurately capture user preferences. These methods can be categorized into invasive and non-invasive techniques (Liu et al. 2021). Invasive methods like FDSA (Zhang et al. 2019) and ICAI-SR (Yuan et al. 2021b) combine session and side information embeddings through concatenation or weighted averaging. While promising, they risk causing a shift in embeddings away from pure item ID representations, potentially impairing consistency during decoding (Hou et al. 2022). To maintain decoding consistency and minimize disturbances from side information semantics, non-invasive methods like NOVA (Liu et al. 2021) and DIF-SR (Xie, Zhou, and Kim 2022) use fused representations only as keys and queries in the attention mechanism while keeping values confined to item features (Zheng et al. 2024, 2023).

Despite these advancements, several challenges remain. The semantic heterogeneity between item-specific (item

*Corresponding author

IDs) and attribute-specific (multi-side information) representations introduces interference during preference attention computation. This misalignment results in inconsistent attention weight distributions, disrupting the mechanism and causing inaccuracies in modeling user preferences and scoring candidate items. Furthermore, these methods often overlook distinct interaction patterns within multi-side information, limiting their ability to improve item score predictions. For example, users often re-interact with items sharing similar side information (*e.g.*, "clothing" category) during sessions, reflecting accumulated interests that clarify their subsequent interaction goals (see Figure 1). However, due to semantic heterogeneity between item IDs and side information, existing methods struggle to effectively capture such patterns, resulting in inconsistent modeling of user preferences and inaccuracies in decoding scores. Addressing persistent user preference patterns while ensuring a consistent semantic space between session representations and item IDs during prediction remains a significant challenge.

To address these challenges, we propose **ScoreNet**, a novel and unified framework that integrates multi-side information to generate prediction scores across all candidate items during decoding. ScoreNet consists of two key components: (1) A Multi-Path Re-Engagement Network (**MPRE**) that calculates re-engagement scores for items likely to be re-interacted by users, explicitly modeling users' persistent interest preferences using multi-side information in a semantic-agnostic manner. It allows side information to influence candidate scores while remaining robust to semantic variations arising from different types of side information. (2) A Position-Enhanced Consistent Scoring Network (**PECS**) that addresses items with limited user interactions by ensuring a consistent semantic space between session and item ID representations through a position-enhanced encoder. Unlike existing methods that directly transform transition relationships into scores, ScoreNet models persistent behavior patterns in a semantics-agnostic manner and ensures decoding consistency by aligning session representations and item ID embeddings, thereby improving score accuracy and offering a novel perspective for SBRs. ScoreNet's superior performance compared to state-of-the-art models underscores the importance of maintaining decoding space consistency and directly integrating side information during the prediction phase.

Our contribution can be summarized as follows:

- We present ScoreNet, a novel framework that integrates side information directly into the calculation of candidate item scores, optimizing the use of side information while ensuring decoding consistency.
- We propose a multi-path re-engagement network that models persistent behavior patterns across different types of side information in a semantic-agnostic manner.
- We introduce a position-enhanced scoring network to address the challenge of items with limited interactions. Position-enhanced computation redistributes attention scores within sessions, ensuring session-item representation consistency for accurate predictions.
- Extensive experiments on three real-world datasets show ScoreNet's superior performance, demonstrating interpretability through attention distribution visualization.

Related Work

Session-based Recommendation

Early approaches like Markov chains and matrix factorization (Rendle, Freudenthaler, and Schmidt-Thieme 2010; Kabbur, Ning, and Karypis 2013; He and McAuley 2016) struggled with complex sequence patterns. The introduction of deep learning, particularly Recurrent Neural Networks (RNNs) (Quadrana et al. 2017; Ma, Kang, and Liu 2019; Yan et al. 2019; Hidasi et al. 2015; Li et al. 2017; Ren et al. 2019), significantly improved performance. GRU4Rec (Hidasi et al. 2015) leveraged Gated Recurrent Units (GRU) to capture user interests, outperforming traditional models. Graph Neural Networks (GNNs) further advanced SBRs by representing user interactions as graphs. SR-GNN (Wu et al. 2019) was pioneering in applying GNNs to SBRs, capturing complex item transitions with Gated Graph Neural Networks (GGNNs). Later models, including FGNN (Qiu et al. 2019), GC-SAN (Xu et al. 2019), GCE-GNN (Wang et al. 2020), and Atten-Mixer (Zhang et al. 2023) introduced attention mechanisms and enhanced information aggregation. The Attention Mechanism revitalized SBRs, with STAMP (Liu et al. 2018) emphasizing short-term preferences. SASRec (Kang and McAuley 2018) and Bert4Rec (Sun et al. 2019b) employed self-attention to capture contextual information. More recent models like DSAN (Yuan et al. 2021a), CORE (Hou et al. 2022) have further optimized these approaches, achieving state-of-the-art performance.

Side Information Fusion for Session-based Recommendation

Research on integrating side information into session embeddings has explored various methods. Liu et al. (Liu et al. 2021) classify these methods as invasive or non-invasive. Invasive techniques directly combine session embeddings with side information through vector concatenation or addition operations. For example, FDSA (Zhang et al. 2019) separately processes item and attribute sequences before merging them via concatenation, while ICAI-SR (Yuan et al. 2021b) employs an attention mechanism to integrate embeddings through a weighted average. However, this direct fusion can shift session embeddings, potentially affecting candidate item scoring based on vector similarity. Recent studies in favor of non-invasive methods like NOVA and DIF-SR (Liu et al. 2021; Xie, Zhou, and Kim 2022), which use integrated representations of items and side information as keys and queries in the self-attention mechanism while preserving the core item features. This approach maintains the consistency in session and candidate item embeddings, reducing bias in similarity computations. These advancements offer promising directions for improving session-based recommendation systems.

Methodology

Given a session $s = [v_1, v_2, ..., v_n]$, where v_t is the t-th interacted item and n is the session length. *Each in-*

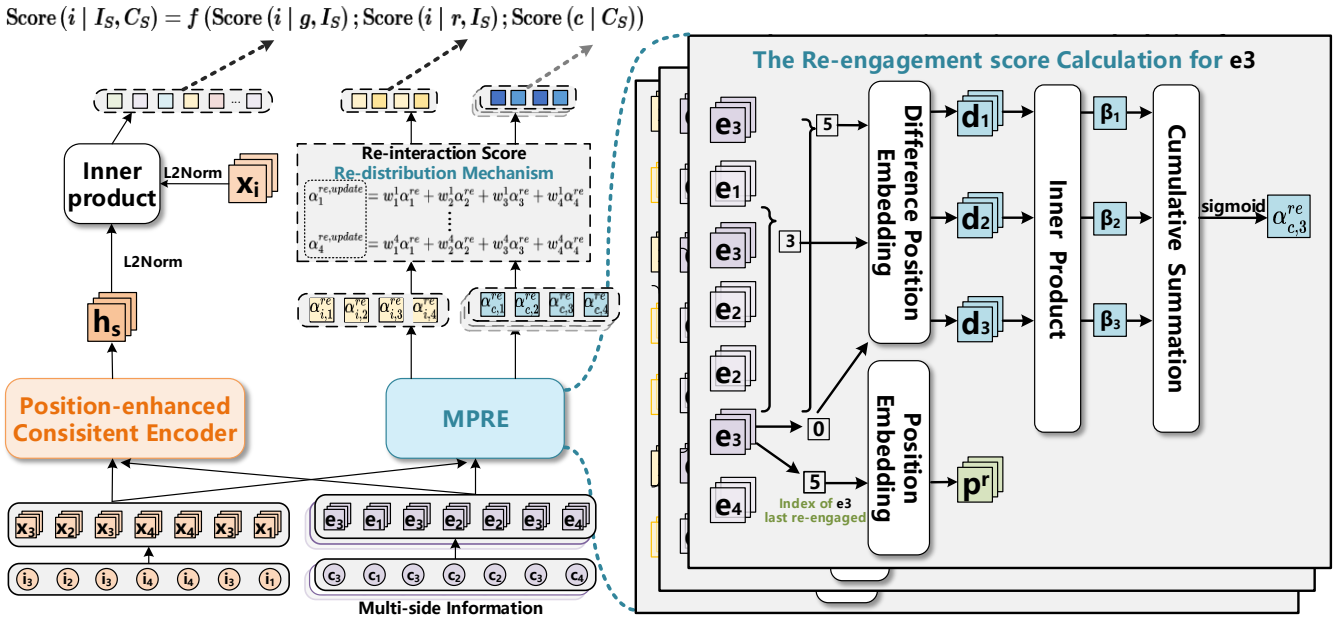


Figure 2: Overview of the proposed ScoreNet.

teraction item v_t includes M types of side information: $v_t = (i_t, c_t^{(1)}, ..., c_t^{(M)})$, where $c_t^{(j)}$ denotes the j-th type of side information for the t-th interaction, and i_t represents the item ID. All session item IDs are drawn from the $I = [i_1, i_2, ..., i_{|I|}]$. We represent item IDs and side information with low-dimensional, dense embedding vectors. Candidate item IDs are represented by $X = [x_1, x_2, ..., x_{|I|}]$, where $x_i \in \mathbb{R}^d$ is the embedding vector for the i-th item in the dictionary, and $x_t \in \mathbb{R}^d$ represents the item in the t-th session interaction. Side information is represented by $E^{(m)} = [e_1^{(m)}, e_2^{(m)}, \ldots, e_{|C|}^{(m)}]$, where $e_t^{(m)} \in \mathbb{R}^d$ represents the m-th type of side information, and C denotes the number of unique side information values for a given attribute type (*e.g.*, the number of unique side information values under the 'brand' attribute). The session-based recommendation system computes the score and predicts probabilities for all candidate items: $\hat{y} = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_{|I|}]$, where $\hat{y}_i$ is the probability for the i-th item in I. The system recommends the top items with the highest probability[1].

Framework

We propose ScoreNet (seen as Figure 2) to generate decoding scores for all candidate items within a unified decoding space, explicitly incorporating side information to model user preference patterns, as described in Eq. 1:

$$\begin{aligned}\textbf{Score}\,(i \mid I_S, C_S) = \mathbf{f}(&\textbf{Score}\,(i \mid g, I_S)\,; \\ &\textbf{Score}\,(i \mid r, I_S)\,; \textbf{Score}\,(c \mid C_S))\end{aligned} \tag{1}$$

$\textbf{Score}\,(i \mid r, I_S)$ represents the re-engagement mode, which models users' persistent interests in items with re-engaged attributes by computing the decoding score for these items. $\textbf{Score}\,(c \mid C_S)$ indicates the weight score of specific side information that frequently engaged in the session. $\textbf{Score}\,(i \mid g, I_S)$ captures the generalization mode, calculating the weight score for items with limited user interaction sequences. It models semantic dependencies between item IDs and related side information, ensuring consistency between sessions and candidate item ID representations. It also incorporates position-enhanced computation to redistribute attention scores within sessions, further refining the recommendation process. $\mathbf{f}$ is a unified scoring function that aggregates the different scoring components to balance the influence of re-engagement and generalization, to ensure that both persistent and transient user preferences are captured effectively.

Multi-Path Re-Engagement Network

We introduce the Multi-Path Re-Engagement Network (**MPRE**), a model designed to independently process multi-side information sequences, effectively capturing user re-interaction patterns without relying on specific semantic meanings. MPRE generates item scores by leveraging the user's overall re-engagement behavior.

The independent processing of various side information sequences within sessions. As expressed in the following equations:

$$\begin{aligned}\text{Score}(i|r, I_S) &= \text{MPRE}([I_S; X], \Phi_I) \\ \text{Score}(c|C_S^{(1)}) &= \text{MPRE}([C_S^{(1)}; E^{(1)}], \Phi_{C1}) \\ \text{Score}(c|C_S^{(2)}) &= \text{MPRE}([C_S^{(2)}; E^{(2)}], \Phi_{C2}) \\ &\vdots \\ \text{Score}(c|C_S^{(k)}) &= \text{MPRE}([C_S^{(k)}; E^{(k)}], \Phi_{Ck})\end{aligned} \tag{2}$$

Here, I_S represents the items the user has interacted with,

[1] https://github.com/pppiao/ScoreNet.

and $C_S^{(1)}, C_S^{(2)}, ..., C_S^{(k)}$ denote sets of different side information $c \in C_S$ within the session. The Score functions compute re-engagement probabilities for a specific item or side information. The parameters $\Phi_I, \Phi_{C1}, \Phi_{C2}, ..., \Phi_{Ck}$ are non-shared network parameters modeling re-interaction behavior with various side information types.

We further elaborate on the calculation of the re-engagement score using the specific side information $c \in C_S^{(j)}$. The absolute position embedding p^r for the index i_w represents the user's most recent interaction. If a user's last repeat interaction with certain side information is far from the most recent one, it may signal a shift in user interest. These embeddings model the user's re-interaction characteristics via an attention mechanism.

To model the user's re-engagement behavior, we define a difference vector D, representing the behavioral pattern for side information c: $D = [i_w - i_1, i_w - i_2, i_w - i_3, ..., i_w - i_w]$, where $[i_1, i_2, i_3, ..., i_w]$ are recent re-engagement indices. For example, if the indices are $[1, 2, 4, ..., 8]$, the difference vector D is computed as $[7, 6, 4, ..., 0]$. These values are then mapped to position embeddings $[d_1, d_2, d_3, ..., d_w]$ to model re-engagement behaviors focusing on interaction frequency rather than item sequence. The re-engagement score α_c^{re} for side information c in the set $C_S^{(j)}$ reflecting the accumulative interest in the side information, is calculated as follows:

$$\alpha_c^{re} = \sum_{m=1}^{w} p^r \cdot d_m^{\mathsf{T}} \tag{3}$$

MPRE also introduces a score redistribution mechanism to adjust re-engagement scores $\alpha^{re} \in \mathbb{R}^n$ based on the user's overall re-engagement behavior across all side information in $C_S^{(j)}$. The updated score $\alpha^{re,update}$ is calculated as:

$$\alpha^{re,update} = \begin{bmatrix} w_1^1 \alpha_1^{re} + w_2^1 \alpha_2^{re} + \cdots + w_n^1 \alpha_n^{re} \\ \vdots \\ w_1^n \alpha_1^{re} + w_2^n \alpha_2^{re} + \cdots + w_n^n \alpha_n^{re} \end{bmatrix} \tag{4}$$

where w_i^j represents the re-assigned score.

Finally, the re-engagement score for side information $c \in C_S^{(j)}$ is given by:

$$\text{Score}(c|C_S^{(j)}) = \begin{cases} sigmoid(\alpha^{re,update}) & \text{if } c \in C_S^{(j)} \\ 0 & \text{if } c \notin C_S^{(j)} \end{cases} \tag{5}$$

This score reflects the user's expected re-engagement with side information c. If a candidate lacks historical interaction data, the score is set to 0, indicating no expected re-engagement.

Position-Enhanced Consistent Scoring Network

We design a Position-Enhanced Consistent Scoring Network (PECS) within ScoreNet to improve prediction accuracy for items with limited user interactions. In each session, every item is associated with multiple side information attributes. The interaction embedding E_t at time step t is computed using a fusion function $\mathcal{F}$:

$$E_t = \mathcal{F}(x_t, p_t, e_t^{(1)}, e_t^{(2)}, \ldots, e_t^{(k)}) \tag{6}$$

where p_t is the position embedding for time step t, and x_t and $e_t^{(i)}$ represent the item and side information embeddings, respectively. We employ addition for fusion:

$$\mathcal{F}_{\text{add}}(x_t, p_t, e_t^{(1)}, e_t^{(2)}, \ldots, e_t^{(k)}) = x_t + p_t + \sum i = 1^k e_t^{(i)} \tag{7}$$

A $\mathcal{L}$-layer Transformer is then used to update the embeddings across all time steps:

$$F = \text{Transformers}([E_1; E_2; \ldots; E_n]) \tag{8}$$

where $\mathcal{F} \in \mathbb{R}^{n \times d'}$ represents the updated embeddings, with n as session length and d' as the output dimension. The updated embeddings capture user behavior and are transformed into semantic importance scores using learnable parameters:

$$\alpha^e = \mathbf{w} \cdot F^\top \tag{9}$$

where $\alpha^e \in \mathbb{R}^n$ is a score vector, and $\mathbf{w} \in \mathbb{R}^{d'}$ is a learnable parameter. To preserve positional information, a position-aware attention mechanism is applied:

$$\alpha^p = p_n \cdot p_t^\top \tag{10}$$

where $p_n \in \mathbb{R}^d$ is the position embedding for the last interaction position n, and $\alpha^p \in \mathbb{R}^n$ represents the position scores for all time steps t. The final weights for items in the session are computed by combining semantic and position scores:

$$\begin{aligned} \alpha_t &= \frac{\exp(\alpha_t^e + \alpha_t^p)}{\sum_{j=1}^n \exp(\alpha_j^e + \alpha_j^p)} \\ h_s &= \sum_{t=1}^{n} \alpha_t \cdot x_t \end{aligned} \tag{11}$$

where $h_s \in \mathbb{R}^d$ is the final session representation. This method ensures consistency between session and candidate item embeddings. The score for each candidate item as:

$$\begin{aligned} \hat{h_s} &= \text{L2Norm}(h_s), \quad \hat{x_i} = \text{L2Norm}(x_i) \\ \text{Score}(i|g, I_S) &= \begin{cases} 0 & \text{if } i \in I_S \\ \hat{h_s} \cdot \hat{x_i}^\top & \text{if } i \in (I - I_S) \end{cases} \end{aligned} \tag{12}$$

where x_i represents the embedding of the i-th candidate item, L2Norm denotes the L2 Normalization function. The score $\text{Score}(i|g, I_S)$ is zero for previously clicked items and is determined by the cosine similarity between the session embedding and candidate item embedding for new items.

Unified Scoring Mechanism for Prediction

We propose a unified scoring mechanism that integrates both re-engagement and semantic scores to generate consistent predictions for all candidate items:

$$\begin{aligned} \text{Score}(i|I_S) &= \text{Score}(i|g, I_S) + \text{Score}(i|r, I_S) \\ S_{cs}(j) &= \text{Score}(c|C_S^{(j)}) \cdot \mathbb{1}_{\{c_i = c \text{ and } c \in C_S^{(j)}\}} \\ \text{Score}(i|I_S, C_S) &= \alpha_1 \cdot Score(i|I_S) + \frac{\alpha_2}{k} \sum_{j=1}^{k} S_{cs}(j) \end{aligned} \tag{13}$$

where α_1 and α_2 are hyperparameters that balance the contributions of re-engagement behavior and semantic preference modeling, with $\alpha_1 + \alpha_2 = 1$; The indicator function $\mathbb{1}_{\{c_i^{(j)}=c^{(j)}\}}$ equals 1 if the candidate item i has the side information c that re-engaged in historical interactions, and 0 otherwise.

To generate the final recommendation probabilities for each candidate item, we apply a softmax function:

$$\hat{y} = \text{Softmax}(\tau \cdot \text{Score}(i|I_S, C_S)) \tag{14}$$

where τ is a scaling factor to prevent over-smoothing. he model is trained using a cross-entropy loss function:

$$\mathcal{L}(y, \hat{y}) = -\sum_{i=1}^{|I|} y_i \log(\hat{y}_i) \tag{15}$$

where $|I|$ denotes the set of candidate items, and y is the one-hot encoded vector of the user's actual clicks. Model parameters are updated through backpropagation.

Experiments and Analysis

Experiments Setup

Datasets To evaluate our model, we utilize three real-world e-commerce datasets: **Diginetica**, **Tmall**, and **Retailrocket**. These datasets provide diverse scenarios for performance assessment. The Diginetica dataset from the 2016 CIKM Cup includes transaction data and item categories. The Tmall dataset from the IJCAI-15 competition contains anonymous shopping logs, where we use category and brand as side information attributes. The Retailrocket dataset, released by a Kaggle competition, spans six months of browsing activity, but most side information beyond item categories is missing. We follow the data preprocessing steps outlined in previous studies (Wu et al. 2019; Yuan et al. 2021a; Hou et al. 2022) for a fair comparison. Key statistics of the datasets are summarized in Table 1.

Dataset	Diginetica	Retailrocket	Tmall
# clicks	858,108	710,586	377,166
# train	526,135	433,648	351,268
# test	44,279	15,132	25,898
# items	40,840	36,968	40,728
avg.len.	5.97	5.43	6.69

Table 1: Statistics of the datasets used in experiments.

Baselines We compare ScoreNet with the following representative methods:

(1) ID-based:

- **GRU4Rec (Hidasi et al. 2015)** is a session-based recommendation model using GRU layers.
- **NARM (Li et al. 2017)** combines GRU with attention mechanisms to model item transitions.
- **STAMP (Liu et al. 2018)** models users' short-term interests only using attention mechanisms focused on the last-clicked item.
- **SR-GNN (Wu et al. 2019)** constructs item transitions as a directed graph and learns item representation.
- **RepeatNet (Ren et al. 2019)** incorporates a repeat-explore mechanism into RNNs to capture the repeat-explore recommendation intent in a session.
- **DSAN (Yuan et al. 2021a)** introduces an adaptive sparse function to enhance the attention mechanism.
- **CORE (Hou et al. 2022)** linearly generates session embeddings to avoid representation inconsistencies.
- **Atten-Mixer (Zhang et al. 2023)** constructs combinations of recent items of varying lengths.

(2) ID-based with side information:

- **ICAI-SR (Yuan et al. 2021b)** implements an invasive side information fusion framework with attention-based Item-Attribute Aggregation.
- **NOVA (Liu et al. 2021)** refines attention distribution using a non-invasive approach with side information.
- **DIF-SR (Xie, Zhou, and Kim 2022)** employs a non-invasive method, decoupling attention calculations for items and side information to enhance performance.

Evaluation metrics We evaluate our model using two common metrics following (Wu et al. 2019; Ren et al. 2019; Hou et al. 2022): P@K (Precision at K) and MRR@K (Mean Reciprocal Rank a K), with K set to 10 and 20. **P@K** measures the proportion of relevant items among the top K recommendations. **MRR@K** calculates the average reciprocal rank of the first relevant item in the top K recommendations.

Hyper-parameter Settings We optimize all hyper-parameters using grid search on the corresponding validation datasets, including the learning rate η, temperature coefficient τ, and embedding dimension d. Specifically, we explore the ranges: $\eta \in \{0.0001, 0.0005, 0.001, 0.01\}$, $\tau \in \{8, 10, 12, 14\}$, and $d \in \{64, 100, 128, 200, 256\}$. Our experiments identify the optimal values as $\eta = 1e-3$, $\tau = 12$, and $d = 28$. We employ Adam as the optimizer and implement the model in PyTorch with a batch size of 256. The parameter α_2, controlling the influence of the side information re-interaction from the MPRE module on the final recommendation, is set to 0.3, while α_1 is set to 0.7.

Overall performance

Table 2 shows that ScoreNet significantly outperforms baseline models across all datasets. Several key insights can be drawn from these results: (1) **Contextual feature modeling is essential for improving recommendation performance.** GNN-based models like SR-GNN surpass earlier models (*e.g.*, GRU4Rec, STAMP) by modeling complex dependency to learn item embeddings. DSAN exceeds GNN models by leveraging Transformers for better contextual feature extraction. (2) **Consistency between session and candidate item representations enhances accuracy.** The CORE model maintains spatial consistency while using Transformers' sequence extraction, improving accuracy over DSAN. (3) **Maintaining representation consistency while using side information offers significant benefits.** DIF-SR outperforms ICAI-SR and NOVA by decoupling item and side

Model	Diginetica				Tmall				Retailrocket			
	P@10	P@20	MRR@10	MRR@20	P@10	P@20	MRR@10	MRR@20	P@10	P@20	MRR@10	MRR@20
GRU4Rec	26.17	39.27	9.69	10.59	9.74	10.93	5.78	5.89	38.35	44.01	23.27	23.67
STAMP	33.49	46.47	13.99	14.89	22.63	26.47	13.12	13.36	42.95	50.96	24.61	25.17
SR-GNN	37.72	50.50	16.75	17.63	23.41	27.57	13.45	13.72	43.21	50.32	26.07	26.57
RepeatNet	33.30	43.17	16.65	17.33	42.74	47.94	18.74	19.14	45.84	51.17	29.08	29.45
DSAN	40.29	53.76	17.05	18.69	26.66	32.32	17.90	18.29	49.05	56.54	30.21	30.74
CORE	41.03	54.36	17.95	18.87	32.97	39.31	18.72	19.16	49.27	56.76	30.03	30.56
Atten-Mixer	40.16	53.86	17.28	18.27	30.11	37.24	18.01	18.62	49.02	56.01	28.05	28.57
ICAI-SR	39.39	52.72	16.82	17.73	25.69	31.27	14.07	14.45	48.21	55.93	28.50	29.03
NOVA	39.69	53.86	17.09	18.03	28.20	33.66	15.51	15.89	48.63	56.38	29.29	28.57
DIF-SR	40.30	53.48	17.25	18.17	29.14	34.92	16.26	16.67	48.96	56.80	29.80	29.25
ScoreNet	**42.76***	**56.27***	**19.38***	**20.32***	**44.81***	**52.24***	**22.73***	**23.25***	**52.09***	**59.37***	**32.31***	**32.81***
Improv(%)	+4.22	+3.44	+7.97	+7.68	+2.34	+6.95	+4.55	+21.35	+5.72	+4.52	+3.46	+7.36

Table 2: Performance comparison of ScoreNet and baseline models across three datasets. Bold values indicate the best overall results; underlined values show the best baseline performance. All values are reported as percentages, with the "%" symbol omitted for brevity. "*" denotes that ScoreNet significantly outperforms the best baseline in a paired t-test ($p < 0.05$).

Model	Diginetica		Tmall		Retailrocket	
	P@20	MRR@20	P@20	MRR@20	P@20	MRR@20
ScoreNet	**56.27**	**20.32**	**52.24**	**23.25**	**59.37**	**32.81**
.-no-MPRE	54.66	19.25	39.84	19.55	57.25	31.06
.-RepeatNet	55.46	18.95	52.03	21.35	58.79	32.02
.-no-MPRErd	56.08	20.18	51.94	22.85	58.91	32.75

Table 3: Performance comparison of ScoreNet variants with and without MPRE module.

information embeddings, avoiding representation shifts, and enhancing performance. (4) **ScoreNet's superior performance stems from three strategies:** First, It maintains **decoding consistency** with a unified scoring mechanism that integrates side information in a semantic-agnostic manner, ensuring consistent prediction scores. Second, compared to RepeatNet, the MPRE module introduces **persistent preference patterns** and models re-interactions with multi-side information independently, revealing user interest patterns without relying on semantics. Third, The PECS module **optimizes attention weights** using side information, generating consistent session embeddings and avoiding nonlinear encoder inconsistencies.

Abltion Studies

Effect of MPRE module We conduct experiments to evaluate the effectiveness of the MPRE module.

- **.-no-MPRE**: ScoreNet without the MPRE module.
- **.-RepeatNet**: ScoreNet with MPRE replaced by RepeatNet for GRU-based re-interaction modeling.
- **.-no-MPRErd**: ScoreNet without MPRE's re-interaction score redistribution mechanism.

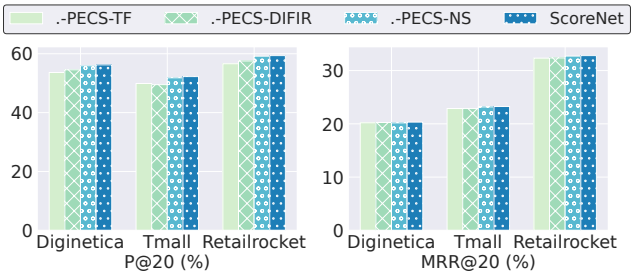
Table 3 presents the results for P@20 and MRR@20 across the three datasets for these ScoreNet variants. The analysis reveals several key findings: First, removing MPRE significantly decreases performance across all datasets, with an average drop of 9.61% in P@20 and 8.52% in MRR@20, underscoring the importance of modeling re-engagement at both item and side information levels. Second, replacing MPRE with RepeatNet results in a performance decline, showing that GRU-based feature extraction is less effective in capturing re-engagement characteristics like frequency and recency. Third, removing the re-interaction score redistribution mechanism reduces performance across all datasets, emphasizing the importance of accounting for global user re-engagement tendencies.

Additionally, varying the hyperparameter α_2 in ScoreNet reveals an inverted U-shaped curve (as shown in Figure 4), with peak performance at $\alpha_2 = 0.3$. This finding highlights that incorporating side information benefits captures persistent user interests. However, the increase of α_2 leads to performance degradation, indicating that over-reliance on side information can dilute core item features. Therefore, balancing core item features with side information is crucial for optimal model performance.

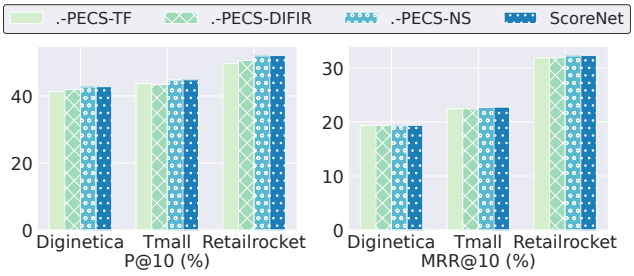
Effect of PECS module To assess the effectiveness of the PECS module, we perform the following experiments:

- **.-PECS-TF**: Replace PECS with a general Transformer in ScoreNet, using its compound embedding from the last time step as the session embedding.
- **.-PECS-DIFIR**: Replace PECS with DIF-SR.
- **.-PECS-NS**: Remove all side information from PECS.

Figure 3 compares the performance of ScoreNet with PECS against its variants, demonstrating the effectiveness of the PECS module and underscoring the importance of maintaining consistency between session and candidate item representations. **.-PECS-TF** shows a significant performance decline compared to **.-PECS-DIFIR** and ScoreNet, indicating that directly fusing side information with item embeddings can disrupt item ID decoding consistency (Hou et al. 2022). **.-PECS-DIFIR** generally outperforms **.-PECS-TF**, showing the effectiveness of DIF-SR in avoiding concatenation of side information and item embeddings, the underperformance compared to ScoreNet, suggesting that nonlinear operations in the Transformer might introduce bias when computing candidate item similarity. **.-PECS-NS** outperforms both **.-PECS-TF** and **.-PECS-DIFIR** across all



(a) Results on P@20 and MRR@20



(b) Results on P@10 and MRR@10

Figure 3: Performance analysis of ScoreNet variants with or without the PECS Module and its component.

datasets, likely due to its linear weighting of original item embedding via attention scores, which preserves core item features. The performance gap with ScoreNet can be attributed to refining attention scores with side information.

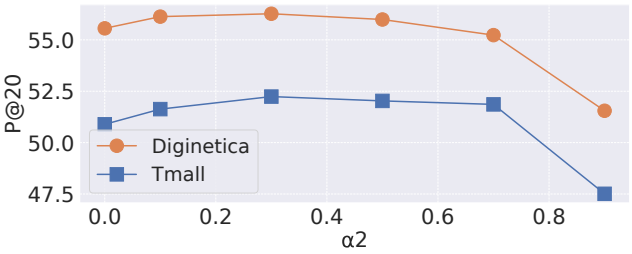


Figure 4: Performance variation with hyper-Parameter α_2.

Contributions of Different Side Information

To evaluate the impact of side information on ScoreNet, we conducted experiments on the Tmall dataset, with results summarized in Table 4. Removing all side information significantly reduces performance, highlighting the critical role of item attributes in enhancing recommendation accuracy. Among individual attributes, brand information improves performance more than category information (0.10% on P@20), indicating that finer-grained attributes are more effective for optimization. Integrating all attributes yields the best results, demonstrating the synergistic effect of combining multiple attributes and highlighting ScoreNet's robustness across diverse e-commerce scenarios.

Visualization of Attention Distribution

We visualize and analyze ScoreNet's re-engagement attention scores for side information in Figure 5 to evaluate its

Side Info	P@10	P@20	MRR@10	MRR@20
None	44.40	50.87	22.70	23.17
Categorie	44.41	50.97	22.78	23.20
Brand	44.64	51.97	22.62	23.15
All	**44.80**	**52.24**	**22.81**	**23.25**

Table 4: Performance of ScoreNet with different side information on Tmall dataset.

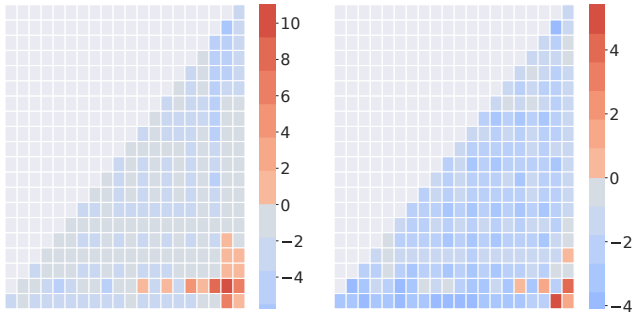


Figure 5: Visualization of attention distribution. The horizontal axis represents the differential value between the index of the last re-interaction and each session index, while the vertical axis denotes the index of the last re-interaction for specific side information on an item. The horizontal axis is reversed, with zero on the right and the maximum on the left. The intensity of the red color indicates the magnitude of the weight value. (Left: Diginetica; Right: Retailrocket.)

interpretability. The analysis reveals a strong correlation between a user's most recent re-interaction with an item attribute and the attribute's attention weight, with attention shifting from right to left. In the Diginetica dataset, the highest attention score appears at the last repeated visit (index 19) with a differential index of 1, emphasizing the significant influence of recent repeated interactions on predictions. These results demonstrate ScoreNet's ability to capture temporal dynamics and prioritize persistent user preferences, enhancing both performance and interpretability in session-based recommendations.

Conclusion and Future Work

In this work, we propose ScoreNet, which consists of MPRE and PECS, designed to address diverse interaction patterns in sessions while maintaining consistent decoding scores. MPRE models users' persistent preference patterns in a semantics-agnostic manner, calculating re-engagement scores for re-interacted items while avoiding interference caused by semantic heterogeneity. PECS enhances predictions for items with limited interactions by ensuring a consistent semantic space between session representations and candidate item ID embeddings using a position-enhanced encoder. These components together enable ScoreNet to effectively model diverse interaction scenarios and improve recommendation performance. We validated ScoreNet across three diverse datasets, demonstrating that it achieves state-of-the-art performance.

Acknowledgements

We would like to thank the anonymous reviewers for their valuable discussion and constructive feedback. This work was supported by the National Natural Science Foundation of China (U22B2061, U2336204), the National Science and Technology Major Project of the Ministry of Science and Technology of China (2022YFB4300603) and Sichuan Science and Technology Program (2023YFG0151).

References

Chen, T.; and Wong, R. C.-W. 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.

He, R.; and McAuley, J. 2016. Fusing similarity models with markov chains for sparse sequential recommendation. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 191–200. IEEE.

Hidasi, B.; Karatzoglou, A.; Baltrunas, L.; and Tikk, D. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939*.

Hou, Y.; Hu, B.; Zhang, Z.; and Zhao, W. X. 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.

Kabbur, S.; Ning, X.; and Karypis, G. 2013. Fism: factored item similarity models for top-n recommender systems. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 659–667.

Kang, W.-C.; and McAuley, J. 2018. Self-attentive sequential recommendation. In *2018 IEEE International Conference on Data Mining (ICDM)*, 197–206. IEEE.

Li, J.; Ren, P.; Chen, Z.; Ren, Z.; Lian, T.; and Ma, J. 2017. Neural attentive session-based recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1419–1428.

Liu, C.; Li, X.; Cai, G.; Dong, Z.; Zhu, H.; and Shang, L. 2021. Non-invasive Self-attention for Side Information Fusion in Sequential Recommendation. *arXiv preprint arXiv:2103.03578*.

Liu, Q.; Zeng, Y.; Mokhosi, R.; and Zhang, H. 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.

Liu, Q.; Zhao, X.; Wang, Y.; Wang, Y.; Zhang, Z.; Sun, Y.; Li, X.; Wang, M.; Jia, P.; Chen, C.; Huang, W.; and Tian, F. 2024. Large Language Model Enhanced Recommender Systems: Taxonomy, Trend, Application and Future. arXiv:2412.13432.

Ma, C.; Kang, P.; and Liu, X. 2019. Hierarchical gating networks for sequential recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 825–833.

Qiu, R.; Li, J.; Huang, Z.; and Yin, H. 2019. Rethinking the item order in session-based recommendation with graph neural networks. In *Proceedings of the 28th ACM international conference on information and knowledge management*, 579–588.

Quadrana, M.; Karatzoglou, A.; Hidasi, B.; and Cremonesi, P. 2017. Personalizing session-based recommendations with hierarchical recurrent neural networks. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 130–137.

Ren, P.; Chen, Z.; Li, J.; Ren, Z.; Ma, J.; and De Rijke, M. 2019. Repeatnet: A repeat aware neural recommendation machine for session-based recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 4806–4813.

Rendle, S.; Freudenthaler, C.; and Schmidt-Thieme, L. 2010. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th international conference on World wide web*, 811–820.

Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; and Jiang, P. 2019a. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1441–1450.

Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; and Jiang, P. 2019b. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM international conference on information and knowledge management*, 1441–1450.

Wang, S.; Cao, L.; Wang, Y.; Sheng, Q. Z.; Orgun, M. A.; and Lian, D. 2021. A survey on session-based recommender systems. *ACM Computing Surveys (CSUR)*, 54(7): 1–38.

Wang, Y.; Du, Z.; Zhao, X.; Chen, B.; Guo, H.; Tang, R.; and Dong, Z. 2023. Single-shot feature selection for multi-task recommendations. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 341–351.

Wang, Y.; Xu, D.; Zhao, X.; Mao, Z.; Xiang, P.; Yan, L.; Hu, Y.; Zhang, Z.; Wei, X.; and Liu, Q. 2024. GPRec: Bi-level User Modeling for Deep Recommenders. *arXiv preprint arXiv:2410.20730*.

Wang, Y.; Zhao, X.; Xu, T.; and Wu, X. 2022. Autofield: Automating feature selection in deep recommender systems. In *Proceedings of the ACM Web Conference 2022*, 1977–1986.

Wang, Z.; Wei, W.; Cong, G.; Li, X.-L.; Mao, X.-L.; and Qiu, M. 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.

Wu, S.; Tang, Y.; Zhu, Y.; Wang, L.; Xie, X.; and Tan, T. 2019. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 346–353.

Xie, Y.; Zhou, P.; and Kim, S. 2022. Decoupled Side Information Fusion for Sequential Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference*

on Research and Development in Information Retrieval, 1611–1621.

Xu, C.; Zhao, P.; Liu, Y.; Sheng, V. S.; Xu, J.; Zhuang, F.; Fang, J.; and Zhou, X. 2019. Graph contextualized self-attention network for session-based recommendation. In *IJCAI*, volume 19, 3940–3946.

Yan, A.; Cheng, S.; Kang, W.-C.; Wan, M.; and McAuley, J. 2019. CosRec: 2D convolutional neural networks for sequential recommendation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2173–2176.

Yuan, J.; Song, Z.; Sun, M.; Wang, X.; and Zhao, W. X. 2021a. Dual sparse attention network for session-based recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 4635–4643.

Yuan, X.; Duan, D.; Tong, L.; Shi, L.; and Zhang, C. 2021b. ICAI-SR: Item Categorical Attribute Integrated Sequential Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1687–1691.

Zhang, P.; Guo, J.; Li, C.; Xie, Y.; Kim, J. B.; Zhang, Y.; Xie, X.; Wang, H.; and Kim, S. 2023. Efficiently leveraging multi-level user intent for session-based recommendation via atten-mixer network. In *Proceedings of the sixteenth ACM international conference on web search and data mining*, 168–176.

Zhang, T.; Zhao, P.; Liu, Y.; Sheng, V. S.; Xu, J.; Wang, D.; Liu, G.; and Zhou, X. 2019. Feature-level Deeper Self-Attention Network for Sequential Recommendation. In *IJCAI*, 4320–4326.

Zheng, Z.; Chao, W.; Qiu, Z.; Zhu, H.; and Xiong, H. 2024. Harnessing large language models for text-rich sequential recommendation. In *Proceedings of the ACM on Web Conference 2024*, 3207–3216.

Zheng, Z.; Sun, Y.; Song, X.; Zhu, H.; and Xiong, H. 2023. Generative Learning Plan Recommendation for Employees: A Performance-aware Reinforcement Learning Approach. In *Proceedings of the 17th ACM Conference on Recommender Systems*, 443–454.

Zhou, K.; Wang, H.; Zhao, W. X.; Zhu, Y.; Wang, S.; Zhang, F.; Wang, Z.; and Wen, J.-R. 2020. S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 1893–1902.