NP-MiSR: Neural Process-based Multi-Interest Learning for Session-Based Recommendation

Jun Bao[1,2*], **Junbo Wang**[1,2*], **Yiheng Jiang**[1], **Xiangfeng Liu**[1,2], **Mingyang Lv**[1,3], **Yuanbo Xu**[1†]

[1]MIC Lab, College of Computer Science and Technology, Jilin University, China
[2]College of Software, Jilin University
[3]College of Computer Science and Technology, Jilin University
yuanbox@jlu.edu.cn
{baojun9922,wangjb5522,Jiangyh22,liuxf5521,lvmy5521}@mails.jlu.edu.cn

Abstract

Session-based recommendation (SBR) aims to provide users with satisfactory suggestions via modeling preferences based on short-term, anonymous user-item interaction sequences. Traditional single interest learning methods struggle to align with the diverse nature of preferences. Recent advances resolved this bottleneck by learning multiple interest embeddings for each session. However, due to the pre-defining scheme of interest quantity (e.g. the number of interests), these approaches are deficient in adaptive ability towards distinctive preference patterns across different users. Moreover, these methods rely solely on the current session and ignore useful information from related ones. The short-term property of sessions would magnify the insufficient representation issue. To address these limitations, we propose a **N**eural **P**rocess-based **M**ulti-**i**nterest learning framework for **S**ession-based **R**ecommendation, namely **NP-MiSR**. To be specific, our method enables adaptive multi-interest representation learning through two complementary mechanisms: 1) **Neural Process-based Intra-session interest modeling**: We employ Neural Processes to model the distribution of interests within a session, where the fixed interest configurations are no longer needed. 2) **Cross-session context fusion**: We extract interest distributions of similar sessions as contextual priors to refine the current session's interest representation. Extensive experiments on three datasets demonstrate that our method consistently outperforms state-of-the-art SBR approaches with an average improvement of 38.8%. Moreover, the few-shot learning task reveals that NP-MiSR achieves a surprisingly favorable efficiency v.s. performance trade-off where utilizing only 10% of the training data attains 95% of the recommendation performance.

Code: https://github.com/xtdysss/NP-MISR

Introduction

Session-based recommendation (SBR) aims to infer user preferences based on their interactions within a single session and provide tailored item suggestions accordingly (Ludewig and Jannach 2018). A session typically refers to a short-term and time-ordered sequence of user-item interactions, such as consecutive product clicks during one browsing or shopping episode. This paradigm is especially valuable in scenarios involving anonymous or new users common in online platforms, where long-term user histories are unavailable (Choi et al. 2024).

Existing SBR methods can be divided into two categories. The first type is the single interest learning model, which generates a single session-specific embedding to represent the entire session's interest over the next item (Wu et al. 2019). However, such a scheme fails to account for the multi-interest nature on real-world scenarios (Shen et al. 2023). As shown in Figure 1, session S includes two interest scopes: clothing and electronics. Obviously, these methods overlook the diverse interests, compress them into a single interest vector, cause the loss of personality and degrade recommendation performance.

To address this bottleneck, the second type proposes to learn multiple interest representations for a single session. However, these methods have two major limitations: 1) requiring predefined interest configurations, e.g., the number of interests, which reduces model flexibility and adaptability (Lv, Liu, and Xu 2025), and 2) relying exclusively on the current session while neglecting semantically related sessions, resulting in suboptimal item representation. Moreover, existing multi-interest learning models approximate interest representations through deterministic parameterized functions, neglecting the inherent uncertainty in user interests (Jiang et al. 2025). As shown in Figure 1, these approaches can only provide a deterministic prediction while neglecting underlying user intent shifting. These shortcomings result in suboptimal recommendation performance.

In this paper, we focus on resolving the aforementioned limitations through a multi-interest learning framework for SBR that dynamically adapts to distinctive interest patterns, enhances item representations by extracting information from similar sessions and enables uncertainty-calibrated interest refinement.

Firstly, we leverage Neural Processes (NPs)(Garnelo et al. 2018) to model the intra-session interest distribution of user interests. This probabilistic framework eliminates the need for pre-defined interest quantities and dynamically adjusts to session-specific patterns. Moreover, we generate the current

*These authors contributed equally.
†Corresponding author.

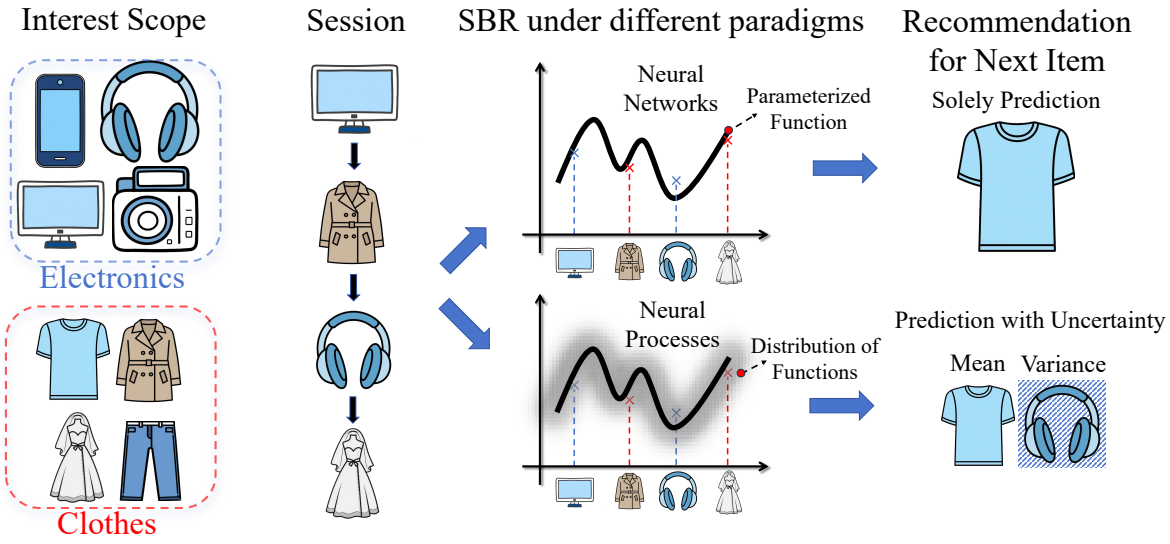

Figure 1: An illustrative example explains the differences between various paradigms in interest modeling. Neural Networks learn a single parameterized function and thus make Solely Predictions, whereas Neural Processes learn the distribution of functions and accordingly produce Predictions with Uncertainty that include both mean and variance components.

interest representation from the distribution via a resampling mechanism where the uncertainty estimation is naturally included for accommodating evolving intents.

Secondly, we incorporate cross-session dependencies to enrich item representations. Specifically, we aggregate relevance-filtered session embeddings as contextual priors for NP through similarity-based session retrieval. Such priors contain patterns of identical items across distinct sessions, enabling a multi-perspective item characterization that comprehensively captures dynamic roles and contextual dependencies.

To this end, we propose a **N**eural **P**rocess-based **M**ulti-**i**nterest learning framework for **S**ession-based **R**ecommendation, namely **NP-MiSR**. The key contributions can be summarized as follows:

- To the best of our knowledge, we are the first to establish the contact between Neural Processes and session-based recommendation, which enables the functional distribution modeling of diverse session interests.
- The proposed NP-MiSR achieves adaptive multiple interests learning for different sessions with uncertainty included, and augments item representations via comprehensively considering contextual references.
- Empirical validation on three benchmark datasets demonstrates the consistent superiority of NP-MiSR, where the average improvement achieves up to 38.8%. The ablation study proves the effectiveness of various components in NP-MiSR. Moreover, the few-shot learning task reveals that our proposed algorithms is capable with a surprisingly good efficiency v.s. performance trade-off that utilizing only 10% of the training data attains 95% of the recommendation performance.

Related Work

Session-based Recommendation

With the in-depth research in the field of session-based recommendation, various session recommendation methods have been proposed. (Hidasi et al. 2015) et al. first introduced the Gated Recurrent Unit for Session-based Recommendation (GRU4REC), which applies Recurrent Neural Networks (RNNs) to session-based recommendation. Inspired by Transformer (Vaswani et al. 2017), the Self-Attention based Sequential Recommendation model (SASRec) (Kang and McAuley 2018) stacks multiple layers to capture item correlations.

However, sequence-based methods infer user preferences through temporal ordering of given sequences, thus failing to model complex item transition patterns (e.g., non-adjacent item transitions). To address this limitation, (Wu et al. 2019) proposed a Gated Graph Neural Network model (SR-GNN) that learns item embeddings on session graphs, then obtains representative session embeddings by combining learned item embeddings with attention mechanisms. (Wang et al. 2020) et al. developed the Global-Context Enhanced Graph Neural Network (GCE-GNN), which captures both global and local session information by constructing global and session graphs.

Multi-Interest Learning

Pioneered by MaxMF (Weston, Weiss, and Yee 2013), Multi-Interest Representation (MIR) gained momentum with capsule-network-based methods: MIND (Li et al. 2019) utilized dynamic routing to cluster interests, while ComiRec (Cen et al. 2020) introduced self-attention to balance diversity and relevance. Subsequent works enriched MIR through temporal modeling (PIMIRec (Chen et al. 2021)), regu-

larization (Re4 (Zhang et al. 2022)), and advanced training strategies (REMI (Xie et al. 2023)). (Jiang et al. 2025) proposed NP-Rec, designed for long-term sequential multi-interest recommendation. Our NP-MiSR, although also utilizing neural processes, is specifically designed for the anonymous, short-session setting.

For session-aware scenarios, graph-based MIR methods emerged: MI-GNN (Wang et al. 2023) jointly modeled historical and current behaviors, whereas TMI-GNN (Shen et al. 2023) increased item correlation density via interest nodes. To address the over-generation of redundant interests, DMI-GNN (Lv, Liu, and Xu 2025) further introduced distance regularization between interest vectors, enforcing sparsity in multi-interest discovery. Despite significant progress, determining the optimal number of interests and modeling the latent interests of individual users with uncertainty remain critical challenges.

Methodology

In this section, we start from introducing neural processes(NPs), and followed by a detailed description of NP-MiSR.

Basic Definition

Let $\mathcal{V} = \{v_1, v_2, \cdots, v_m\}$ be all of items. Each anonymous session, which is denoted by $S = \{v_1^s, v_2^s, \cdots, v_l^s\}$, consists of a sequence of interactions (i.e., items clicked by a user in chronological order), where v_i^s denotes item v_i clicked within session S, and the length of S is l. A session set $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$ is defined as a set containing n such anonymous sessions.

Neural Processes (NPs)

NPs aim at mapping an input $\boldsymbol{x}_i \in \mathbb{R}^{d_x}$ to the corresponding output $\boldsymbol{y}_i \in \mathbb{R}^{d_y}$ based on an (infinite) family of conditional distributions. In particular, one may condition on an arbitrary number of observed *Contexts* $(\boldsymbol{X}_C, \boldsymbol{Y}_C)$ to model an arbitrary number of *Targets* $(\boldsymbol{X}_T, \boldsymbol{Y}_T)$. The arbitrary property requires the mapping procedure to be non-sensitive towards the order of contexts or targets. The conditional distribution is

$$p(\boldsymbol{Y}_T|\boldsymbol{X}_T, \boldsymbol{X}_C, \boldsymbol{Y}_C) = \int p(\boldsymbol{Y}_T|\boldsymbol{X}_T, \boldsymbol{r}_C, z)p(z|\boldsymbol{s}_C)dz \tag{1}$$

where $\boldsymbol{r}_C = r(\boldsymbol{X}_C, \boldsymbol{Y}_C) \in \mathbb{R}^d$ and $\boldsymbol{s}_C = \psi(\boldsymbol{X}_C, \boldsymbol{Y}_C) \in \mathbb{R}^d$ are the finite dimensional representations. $r(\cdot)$ is an order invariant *deterministic* function which aggregates contexts, and $\psi(\cdot)$ is the *latent* one of the same properties. Given the observation $(\boldsymbol{x}_C, \boldsymbol{y}_C)$, the global latent $z \in \mathbb{R}^d$ accounts for incorporating uncertainties in the predictions $\boldsymbol{Y}_T$ which is modeled by a factorized Gaussian parameterized $\boldsymbol{s}_C$. Given a random subset of contexts C and targets T, NPs learn the parameters in the encoder-decoder architecture by maximizing the following ELBO with reparameterization trick (Kingma, Welling et al. 2013),

$$\log p(\boldsymbol{Y}_T|\boldsymbol{X}_T, \boldsymbol{X}_C, \boldsymbol{Y}_C) \geq \mathbb{E}_{q(z|\boldsymbol{s}_T)}[\log p(\boldsymbol{Y}_T|\boldsymbol{X}_T, \boldsymbol{r}_C, z)] - D_{KL}(q(z|\boldsymbol{s}_T)||q(z|\boldsymbol{s}_C)) \tag{2}$$

where q, r and s form the encoder part, and the likelihood p is referred as the decoder.

NP-MiSR

As shown in Fig.2, NP-MiSR primarily consists of a session representation learning layer and the NP model. The NP model further contains a deterministic encoder and a latent encoder. Below, we will provide detailed descriptions of each component of the model, as well as its training and inference procedures.

Session Representation Learning Layer For each item v_i^s in session $S \in \mathcal{S}$, we obtain its corresponding item representation $\mathbf{h}_{v_i^s}$ through an embedding layer. We then employ a session representation model to aggregate item representations within the session, ultimately deriving the comprehensive session representation $\mathbf{h}^s$. To ensure experimental consistency, in subsequent experiments, we use **GCE-GNN** as the "SBR Model" in Fig.2 to perform session representation extraction.

NP-MiSR Pipeline Next, we will introduce our model from two perspectives: training mode and inference mode.

Training Mode. Considering the typically short length of session-based recommendations poses challenges for effective sequence modeling, we opt to perform modeling at the session set level rather than individual session level. Specifically, for an input session set $\mathcal{S}$, each session $S \in \mathcal{S}$ is processed through the session representation learning layer described in Section to obtain session representations $\mathbf{h}^S$ as x_S, while using the embedding of the next item $h_{v_{l+1}}$ as y_s.

We concatenate each pair $(\boldsymbol{x_s}, \boldsymbol{y_s})$ along the last dimension to form the session-label pair collection $[\boldsymbol{X}, \boldsymbol{Y}] \in \mathbb{R}^{|\mathcal{S}| \times 2d}$.

At each iteration, we select one session-label pair from the set as the target pair $[\boldsymbol{X}_T, \boldsymbol{Y}_T] \in \mathbb{R}^{1 \times 2d}$. We then construct context pairs $[\boldsymbol{X}_C, \boldsymbol{Y}_C] \in \mathbb{R}^{C \times 2d}$ by selecting the *Top-C* session-label pairs with the highest similarity between $\boldsymbol{X}$ and $\boldsymbol{X}_T$ from the remaining pairs. In practice, it is feasible to sparsely sample a small subset from the dataset as the scope for the model to retrieve contexts, thereby accelerating both learning and inference. For any target pair $[\boldsymbol{X}_T, \boldsymbol{Y}_T]$, its corresponding context pairs $[\boldsymbol{X}_C, \boldsymbol{Y}_C]$ can be represented by the following formula:

$$\begin{aligned} \mathcal{C}_T &= \left\{ (\boldsymbol{x}_c, \boldsymbol{y}_c) \mid c \in \underset{s \neq T}{\arg\max^{(C)}} \operatorname{sim}(\boldsymbol{x}_s, \boldsymbol{X}_T) \right\}, \\ [\boldsymbol{X}_C, \boldsymbol{Y}_C] &= \operatorname{Stack}\left(\boldsymbol{x}_c \oplus \boldsymbol{y}_c \right)_{(\boldsymbol{x}_c, \boldsymbol{y}_c) \in \mathcal{C}_T}, \end{aligned} \tag{3}$$

where $\arg\max^{(C)}$ denotes the indices of the *Top-C* largest values, $\oplus$ represents the vector concatenation operation, and $\operatorname{sim}(\cdot, \cdot)$ refers to the similarity computation function.

For the *Deterministic Encoder* that takes only the obtained context pair $[\boldsymbol{X}_C, \boldsymbol{Y}_C]$ as input, it first processes the context using a MLPs $\boldsymbol{r}(\cdot)$, then performs order invariant aggregation over outputs of each session to obtain the deterministic representation $\boldsymbol{r}_C \in \mathbb{R}^d$.

For the *Latent Encoder*, which takes both the context and target as inputs (using the context as an example here), it first

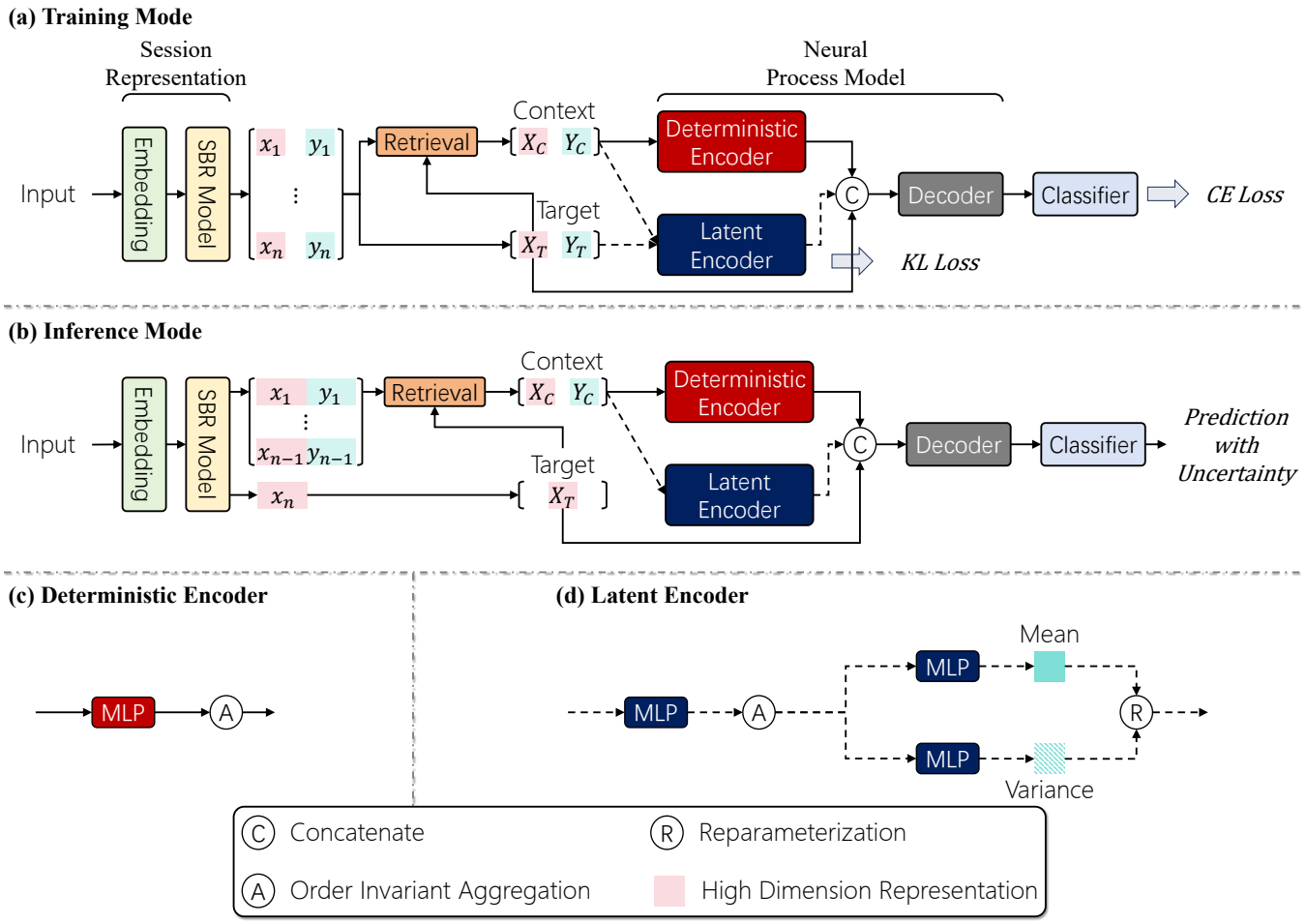

Figure 2: Overview of NP-MiSR: It primarily comprises a session representation learning model and an NP model. The session representation extraction layer can be replaced by any session model capable of extracting comprehensive session representations. In the figure, parts (a) and (b) are the flowcharts of the model's training mode and inference mode, respectively, while parts (c) and (d) demonstrate the detailed implementations of the deterministic encoder and latent encoder.

projects the input into a latent space via MLPs $\psi(\cdot)$. Subsequently, an aggregator is applied to aggregate the representations. The aggregated output is then passed through two separate MLPs to obtain the mean vector $\boldsymbol{\mu}$ and variance vector $\boldsymbol{\sigma}^2$, forming the prior distribution $q(z|\boldsymbol{s}_C)$. Then, it samples K latent representations via reparameterization trick $z_C \in \mathbb{R}^{K\times d}$. Towards the target input, the latent performs in the same manner to get the posterior distribution $q(z|s_T)$ and latent representations $z_T \in \mathbb{R}^{K\times d}$.

For the *Decoder*, which takes the target session representation $\boldsymbol{X}_T$, deterministic representation $\boldsymbol{r}_C$, and posterior latent representation z_T as inputs, we maximize the evidence lower bound (ELBO). To achieve this, $\boldsymbol{X}_T$ and $\boldsymbol{r}_C$ are replicated K times. These replicated representations are concatenated with z_T along the last dimension to form the Decoder's input. The Decoder then produces K prediction vectors $\{\hat{\boldsymbol{h}}_k\}_{k=1}^K$, which are aggregated via mean pooling to obtain the final session representation $\boldsymbol{h'}_S$. The uncertainty is computed as the entropy of mean(Wang et al. 2022). The processing process of the Decoder can be described by the following formula:

$$\boldsymbol{h'}_S = \frac{1}{K}\sum_{k=1}^{K} \text{Decoder}\left(\left[\boldsymbol{X}_T^{(k)}, \boldsymbol{r}_C^{(k)}, z_T\right]\right), \tag{4}$$

where $\boldsymbol{X}_T^{(k)}$ and $\boldsymbol{r}_C^{(k)}$ denote replicated inputs, ensuring dimensional compatibility with the latent variable z_T during concatenation.

For the *Classifier*, based on the obtained session representations $\boldsymbol{h'}_S$, the final recommendation probability for each candidate item is computed using their initial embeddings and the current session representation. This is formulated by first calculating the dot product between these representations, then applying the softmax function to obtain the output prediction $\hat{y}$:

$$\hat{y}_i = \text{Softmax}\left(\boldsymbol{h'}_S^{\mathrm{T}} \cdot h_{v_i}\right), \tag{5}$$

Dataset	# training	# test	# items	Avg.Len
Diginetica	719,410	60,858	43,097	5.12
RetailRocket	433,643	15,132	36,968	5.43
Nowplaying	825,304	89,824	60,417	7.42

Table 1: Statistics of the used datasets.

where $\hat{y}_i$ in $\hat{y}$ denotes the probability of item v_i appearing as the next-click in the current session.

The final learning objective during model training is formulated in Equation 2, where the likelihood term $p(\boldsymbol{Y}_T|\boldsymbol{X}_T, \boldsymbol{r}_C, z)$ is computed through the cross-entropy loss between the probability distribution derived from Equation 5 and the one-hot encoded ground-truth labels of the session. The $D_{\mathrm{KL}}(\cdot)$ term represents the Kullback-Leibler divergence between the prior and posterior distributions:

$$\begin{aligned}\mathcal{L} =&\mathbb{E}_{q(z|\boldsymbol{s}_T)}[-\log p(\boldsymbol{Y}_T|\boldsymbol{X}_T, \boldsymbol{r}_C, z)] \\ &+ \beta \cdot D_{KL}(q(z|\boldsymbol{s}_T)||q(z|\boldsymbol{s}_C)),\end{aligned} \tag{6}$$

where the likelihood evaluation is implemented as:

$$-\log p(\boldsymbol{Y}_T|\cdot) = \sum_{i=1}^{|\mathcal{V}|} y_i \cdot \log(\hat{y}_i) \tag{7}$$

Inference Mode. Given a session set $\mathcal{S}$, the *NP-MiSR* aims to predict the next interaction item V_{l+1}^S for each session $S \in \mathcal{S}$. To prevent data leakage and address the unavailability of V_{l+1}^S during prediction, we construct the target input $\boldsymbol{X}_T \in \mathbb{R}^{1\times d}$ using the aggregated representation of the current session's sequence $V_{1:l}^S$. For context construction, we utilize other sessions' subsequences $V_{1:l-1}$ as their input features $\boldsymbol{x}$, with their respective l-th item representations V_l serving as labels $\boldsymbol{y}$. These pairs are then organized into context session-label pairs $[\boldsymbol{X}_C, \boldsymbol{Y}_C] \in \mathbb{R}^{C\times 2d}$ following Equation 3

During inference, only the context information is processed through the encoders. The *Latent Encoder* transforms its output via MLPs $\boldsymbol{\psi}(\cdot)$ to generate mean vector and variance vector of the prior distribution $q(z|\boldsymbol{s}_C)$. Through the reparameterization trick, we sample K latent vectors $\{z_C^{(k)}\}_{k=1}^K$. The *Deterministic Encoder* maintains its training-phase procedure to produce $\boldsymbol{r}_C$. Following the decoding scheme in Equation 4, we process the concatenated inputs $[\boldsymbol{X}_T, \boldsymbol{r}_C, z_T]$ to obtain the final session representation $\boldsymbol{h'}_S$. The classifier subsequently computes the item probability distribution.

Experiments

In this section, we report our experimental setting, including datasets, baselines, evaluation metrics, and an analysis of experimental results. We aim to answer the following questions:

- **RQ 1:** Compared with state-of-the-art methods in session-based recommendation and multi-interest recommendation (SOTA), does our method demonstrate competitive or better performance?
- **RQ 2:** How do different modules of NP-MiSR impact the recommendation performance?
- **RQ 3:** How well does NP-MiSR perform when learning from only a partial dataset?
- **RQ 4:** How do the model's hyperparameters affect its effectiveness?

Experimental Setup

The detailed experimental setup will be described below.

Dataset and Preprocessing We conducted experiments on three real-world datasets(Diginetica, RetailRocket, Nowplaying) to verify the effectiveness of our method. The statistical characteristics of the processed datasets are presented in Table 1. To ensure fair comparison with baseline methods, we adopted the same data preprocessing pipeline as SR-GNN.

Parameter Setup For a fair comparison, we keep consistent with the settings of GCE-GNN. We set both the embedding dimension and latent vector dimension to 100, train for 20 epochs with a batch size of 100. Context ratio T set to 10%, sampling times K set to 10 times. All experiments are conducted on an NVIDIA 3090Ti GPU with 24GB VRAM, and each experiment is repeated more than three times.

Evaluation Metrics We evaluate recommendation performance using three metrics: both **Hit Rate (HR@N)** and **Mean Reciprocal Rank (MRR@N)** are established metrics in recommendation research. In addition, we employ the **Coverage (COV@N)** metric, a diversity indicator that reflects the proportion of unique items appearing in top-N recommendation lists across all sessions. All metrics follow the *higher-is-better* principle. We set $N = \{5, 10, 20\}$ to examine performance at different recommendation lengths.

Baseline Our experiment compares NP-MiSR with the following conversational models: **GCE-GNN**, **CORE**(Hou et al. 2022), **TAGNN**(Yu et al. 2020), **A-Mixer**(Zhang et al. 2023), **MiaSRec**(Choi et al. 2024), **Link**(Choi et al. 2025), **DMI-GNN**(Lv, Liu, and Xu 2025).

Overall Comparison (RQ1)

As shown in Table 2, which presents complete results of 7 baseline methods and our proposed method across 9 metrics on three real-world datasets. It can be observed that DMI-GNN and MiaSRec achieve second-best results on the majority of metrics across two datasets, which is attributed to their simple yet effective multi-interest learning in session-based recommendation. CORE also demonstrates competitive performance on RetailRocket datasets, respectively.

Notably, our method consistently outperforms all baseline methods across all metrics on each dataset. This compelling evidence strongly validates the effectiveness of our proposed approach.

Ablation Study (RQ2)

Our framework primarily consists of two core modules: a session representation learning layer and an NP model. To validate the effectiveness of each component, as shown in Table 3, we design two variants: **w.o. NP** and **w.o. SR**. The

Dataset	Metric (%)	GCE-GNN	CORE	TAGNN	A-Mixer	MiaSRec	DMI	Link	**NP-MiSR**	**IMPV.**
Nowplaying	H@5	12.51	11.82	10.70	12.13	11.38	13.47	11.92	**18.97**	**40.83%**
	M@5	7.57	6.74	7.08	7.69	6.90	8.16	6.54	**12.04**	**47.55%**
	C@5	66.15	53.25	58.84	49.07	58.25	75.79	75.90	**94.02**	**23.87%**
	H@10	17.02	18.22	14.43	16.83	16.67	18.86	17.55	**24.19**	**28.26%**
	M@10	8.17	7.59	7.55	8.31	7.60	8.87	7.29	**15.99**	**80.27%**
	C@10	78.00	67.31	71.59	61.54	68.43	87.34	87.46	**97.59**	**11.58%**
	H@20	22.48	25.22	19.05	23.03	22.65	25.59	23.53	**30.01**	**17.27%**
	M@20	8.54	8.08	7.86	8.74	8.01	9.34	7.70	**20.48**	**119.27%**
	C@20	86.64	76.69	82.56	69.34	78.40	95.56	95.24	**99.20**	**3.81%**
Diginetica	H@5	29.37	28.59	26.11	27.38	29.51	29.97	29.38	**41.17**	**37.37%**
	M@5	16.61	16.17	15.37	15.13	16.97	16.96	16.36	**30.78**	**81.37%**
	C@5	67.15	65.22	47.44	66.17	64.09	69.76	72.01	**87.11**	**20.96%**
	H@10	41.14	39.91	36.04	39.39	41.15	41.69	41.01	**53.13**	**27.44%**
	M@10	18.17	16.67	16.82	16.72	18.51	18.51	17.90	**42.12**	**127.55%**
	C@10	81.62	80.55	60.97	79.54	78.18	83.56	85.25	**94.85**	**11.26%**
	H@20	54.14	52.88	49.90	53.21	54.21	54.95	54.33	**65.09**	**18.45%**
	M@20	19.07	18.57	16.71	17.68	19.42	19.44	18.82	**54.56**	**180.65%**
	C@20	92.54	89.37	74.90	90.01	90.26	94.03	94.60	**98.49**	**4.11%**
RetailRocket	H@5	40.29	47.15	36.77	40.55	42.80	42.35	41.52	**47.64**	**1.04%**
	M@5	27.36	37.15	25.30	26.95	29.37	29.78	29.00	**41.39**	**11.41%**
	C@5	40.80	43.20	37.11	39.15	38.82	41.99	41.44	**57.32**	**32.68%**
	H@10	48.19	54.08	44.19	49.33	50.52	50.00	48.69	**54.66**	**1.07%**
	M@10	28.43	33.08	26.30	28.14	30.41	30.81	29.95	**48.99**	**48.09%**
	C@10	57.16	60.37	51.52	53.84	53.67	58.76	58.45	**72.86**	**20.68%**
	H@20	55.96	61.67	51.89	57.18	57.48	57.59	56.12	**62.04**	**0.59%**
	M@20	28.97	38.54	26.84	28.69	30.90	31.34	30.46	**56.45**	**46.47%**
	C@20	73.61	82.59	66.04	69.65	70.52	76.18	77.25	**85.76**	**3.83%**

Table 2: Overall recommendation performance. The best and second scores are marked with **boldface** and underline forms, separately. The last column "Impv." stands for the improvement of our method against the strongest baseline.

Dataset	Retailrocket					
Metric (%)	H@5	M@5	H@10	M@10	H@20	M@20
Original	47.01	41.39	54.66	48.99	62.04	56.45
w.o. NP	40.29	27.36	48.19	28.43	55.96	28.97
w.o. SR	3.88	1.50	5.01	1.93	6.75	2.60
Dataset	Diginetica					
Metric (%)	H@5	M@5	H@10	M@10	H@20	M@20
Original	41.17	30.78	53.13	42.12	65.09	54.56
w.o. NP	29.37	16.61	41.14	18.17	54.14	19.07
w.o. SR	2.61	1.32	4.47	1.55	6.85	1.69

Table 3: Ablation study.

variants that remove the neural process and session representation learning layer are tagged as "w.o.NP" and "w.o.SR", separately. w.o.SR replaces the session representation learning layer with a simple embedding layer with learnable weights for session representation aggregation. The experimental results reveal the following findings:

- **Observation 1**: The NP module substantially enhances recommendation performance. Comparative analysis between the full implementation and the **w.o. NP** variant demonstrates that incorporating neural processes yields an average performance improvement of 39.58% across all datasets. This enhancement stems from two synergistic mechanisms: (1) The multi-sampling strategy in the latent encoder extends the single-point output of the session representation layer to capture diverse interest distributions; (2) Subsequent averaging operations provide uncertainty estimation in final recommendations, effectively mitigating the risk of overconfidence in current preference predictions.

- **Observation 2**: The session representation learning layer proves essential for generating meaningful interest representations. The **w.o. SR** variant exhibits significant performance degradation, primarily due to the lack of effective session representation extraction(e.g., GNN). Simple weight aggregation fails to learn discriminative session representations, consequently hindering the establishment of meaningful session-item correlations. Our complete NP-MiSR framework successfully integrates the complementary strengths of both session representation learning and neural processes, thereby achieving state-of-the-art recommendation performance. Notably,as shown in Fig.3, by integrating the NP model with multiple dialogue models, the experimental results demonstrate that all hybrid models achieve significant performance improvements, which strongly validates the versatility and compatibility of the NP model.

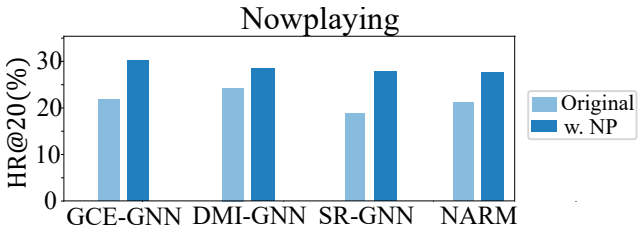

Figure 3: Results of applying the NP model to different session models.

Few-shot Learning (RQ3)

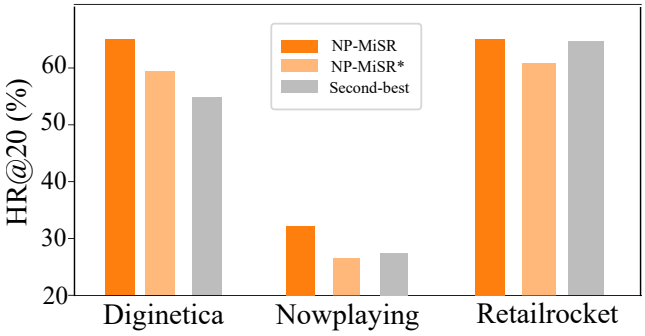

Figure 4: The performance of NP-MiSR on small datasets.

To verify that NP-MiSR has a good ability to balance efficiency and overhead, we designed the experiment as shown in Figure 4.

In the figure, "NP-MiSR*" represents results learned using only 10% of the data across three datasets; "NP-MiSR" corresponds to results learned using the full dataset; and "Second-best" indicates the best value achieved by the compared baseline models trained on the full dataset. It demonstrates that NP-MiSR using reduced data achieves results comparable to the best-performing baseline models, even surpassing them on the Diginetica dataset. This indicates NP-MiSR's exceptional efficiency in session data utilization.

Impact of Hyperparameters (RQ4)

To understand the impact of different values of hyperparameters sampling times K and context ratio T on the model performance, we designed the hyperparameter experiment as shown in the Figure 5.

As illustrated in Figure 5, we conducted comprehensive experiments to evaluate the impacts of hyperparameters K and T. The empirical results reveal that increasing the value of K generally leads to improved model performance. This phenomenon can be attributed to the enhanced capacity of larger K values to better capture the functional distribution learned by the model. Notably, the lower section of the figure demonstrates that varying context ratios exhibit differential effects across distinct datasets, which we hypothesize may stem from the inherent distributional differences across the datasets.

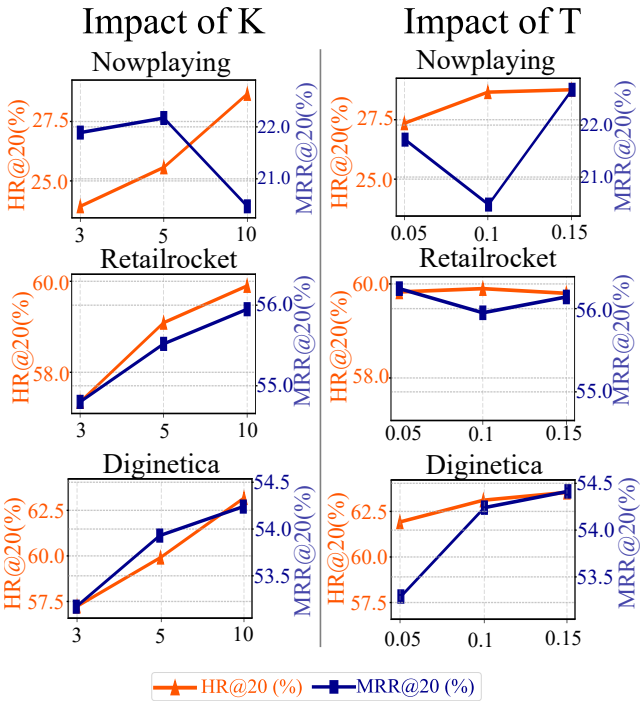

Figure 5: Results of hyperparameters T and K on different datasets.

Conclusion

In this paper, we propose a multi-interest session-based recommendation framework (NP-MiSR) that integrates neural processes with traditional session models. This framework addresses two key issues in multi-interest approaches for session-based recommendation: the lack of flexibility in interest modeling and the challenge of fusing information across sessions. We validate our results on three public datasets.

Acknowledgments

This work is supported by the Natural Science Foundation of China No. 62472196, Jilin Science and Technology Research Project 20230101067JC

References

Cen, Y.; Zhang, J.; Zou, X.; Zhou, C.; Yang, H.; and Tang, J. 2020. Controllable multi-interest framework for recommendation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2942–2951.

Chen, G.; Zhang, X.; Zhao, Y.; Xue, C.; and Xiang, J. 2021. Exploring periodicity and interactivity in multi-interest framework for sequential recommendation. *arXiv preprint arXiv:2106.04415*.

Choi, M.; Kim, H.-y.; Cho, H.; and Lee, J. 2024. Multi-intent-aware session-based recommendation. In *Proceedings of the 47th international ACM SIGIR conference on*

research and development in information retrieval, 2532–2536.

Choi, M.; Lee, S.; Park, S.; and Lee, J. 2025. Linear Item-Item Models with Neural Knowledge for Session-based Recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1666–1675.

Garnelo, M.; Schwarz, J.; Rosenbaum, D.; Viola, F.; Rezende, D. J.; Eslami, S.; and Teh, Y. W. 2018. Neural processes. *arXiv preprint arXiv:1807.01622*.

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.

Jiang, Y.; Xu, Y.; Yang, Y.; Yang, F.; Wang, P.; and Li, C. 2025. Auto Encoding Neural Process for Multi-interest Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 11879–11887.

Kang, W.-C.; and McAuley, J. 2018. Self-attentive sequential recommendation. In *2018 IEEE international conference on data mining (ICDM)*, 197–206. IEEE.

Kingma, D. P.; Welling, M.; et al. 2013. Auto-encoding variational bayes.

Li, C.; Liu, Z.; Wu, M.; Xu, Y.; Zhao, H.; Huang, P.; Kang, G.; Chen, Q.; Li, W.; and Lee, D. L. 2019. Multi-interest network with dynamic routing for recommendation at Tmall. In *Proceedings of the 28th ACM international conference on information and knowledge management*, 2615–2623.

Ludewig, M.; and Jannach, D. 2018. Evaluation of session-based recommendation algorithms. *User Modeling and User-Adapted Interaction*, 28(4): 331–390.

Lv, M.; Liu, X.; and Xu, Y. 2025. Dynamic Multi-Interest Graph Neural Network for Session-Based Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 12328–12336.

Shen, Q.; Zhu, S.; Pang, Y.; Zhang, Y.; and Wei, Z. 2023. Temporal aware multi-interest graph neural network for session-based recommendation. In *Asian Conference on Machine Learning*. PMLR.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, J.; Lukasiewicz, T.; Massiceti, D.; Hu, X.; Pavlovic, V.; and Neophytou, A. 2022. Np-match: When neural processes meet semi-supervised learning. In *International Conference on Machine Learning*, 22919–22934. PMLR.

Wang, T.-Y.; Chen, C.-T.; Huang, J.-C.; and Huang, S.-H. 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): 1–28.

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.

Weston, J.; Weiss, R. J.; and Yee, H. 2013. Nonlinear latent factorization by embedding multiple user interests. In *Proceedings of the 7th ACM conference on Recommender systems*, 65–68.

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.; Gao, J.; Zhou, P.; Ye, Q.; Hua, Y.; Kim, J. B.; Wu, F.; and Kim, S. 2023. Rethinking multi-interest learning for candidate matching in recommender systems. In *Proceedings of the 17th ACM conference on recommender systems*, 283–293.

Yu, F.; Zhu, Y.; Liu, Q.; Wu, S.; Wang, L.; and Tan, T. 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.

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, S.; Yang, L.; Yao, D.; Lu, Y.; Feng, F.; Zhao, Z.; Chua, T.-S.; and Wu, F. 2022. Re4: Learning to re-contrast, re-attend, re-construct for multi-interest recommendation. In *Proceedings of the ACM web conference 2022*, 2216–2226.