

# 面向会话表示的多阶段自监督集合学习方法

龚晓宇

上海理工大学管理学院, 上海

收稿日期: 2025年9月7日; 录用日期: 2025年9月16日; 发布日期: 2025年10月14日

## 摘要

会话推荐旨在根据用户在会话单元中的交互组对用户下一项感兴趣的交互进行推荐。然而, 现有会话推荐方法通常将会话建模为序列形式, 使推荐结果过于依赖交互顺序。为了应对这一问题, 本文提出了一种面向会话表示的多阶段自监督集合学习方法SR-MSSL。该方法将会话建模为集合形式, 通过自注意力机制对集合中的交互项进行信息聚合, 并通过多阶段自监督学习挖掘元素间的隐含关联, 从而在不依赖交互顺序的情况下获取具有表达力的会话表示。实验在Yoochoose和Diginetica数据集上进行, 结果显示, 本文方法在多个指标上优于传统推荐方法与多种深度学习模型, 验证了集合建模在会话表示学习中的有效性。此外, 通过位置编码消融实验进一步证明, 去除交互的顺序信息能够提高会话模型的推荐性能。

## 关键词

会话推荐, 集合表示学习, 自监督学习

# Multi-Stage Self-Supervised Set Learning Approach for Session Representation

Xiaoyu Gong

Business School, University of Shanghai for Science and Technology, Shanghai

Received: September 7, 2025; accepted: September 16, 2025; published: October 14, 2025

## Abstract

Session-based recommendation seeks to predict the next interaction of interest for a user by leveraging the set of interactions within a session. However, current session-based recommendation approaches usually treat sessions as sequences, making the recommendations excessively dependent on the order of interactions. To mitigate this limitation, we propose SR-MSSL, a multi-stage self-supervised set learning framework for session representation. The approach

**models sessions as sets, employing a self-attention mechanism to aggregate information across interactions and leveraging multi-stage self-supervised learning to capture latent relationships among elements, thus deriving expressive session representations independent of interaction order. Experiments on the Yoochoose and Diginetica datasets demonstrate that our method surpasses traditional recommendation techniques and several deep learning models across multiple evaluation metrics, confirming the effectiveness of set-based modeling for session representation learning. Additionally, ablation studies on positional encoding reveal that discarding the sequential information of interactions can further improve the recommendation performance of session models.**

## Keywords

**Session-Based Recommendation, Set Representation Learning, Self-Supervised Learning**

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着信息过载与技术迭代，推荐系统已经成为提升用户体验不可或缺的工具，被广泛应用于各种信息化场景[1] [2]。当前用户与系统间的互动随着交互方式的演化，往往呈现为短时、连续且上下文依赖的行为组。该行为组即是特定时间段的相关行为组成的会话，并衍生了基于会话的推荐(会话推荐)这一推荐系统的新兴推荐范式[3]。然而，会话数据通常存在序列短与上下文稀疏等问题，使得如何学习高质量的会话表征成为了研究难点。

传统方法[4] [5]往往依赖于局部行为关系，例如马尔可夫链通过计算相邻交互间的转移概率进行预测，但这种方式难以体现用户长期兴趣。为解决这一局限，循环神经网络和长短时记忆网络被引入[6] [7]。此类方法会增强模型对时间依赖性的捕捉能力，从而更准确地预测用户行为。然而，过度依赖时间信息可能导致推荐结果对交互顺序敏感。另一方面，图神经网络方法[8] [9]通过建模项目间的复杂关联，为提取会话表征提供了新范式。但在静态图下，用户兴趣的动态演变仍难以充分反映，尤其面对长会话或大规模数据时效果有限[10]。

为了应对上述挑战，本文提出了面向会话表示的多阶段自监督集合学习方法 SR-MSSL。该方法基于集合形式建模，通过将会话数据建模为集合，可以依赖集合的无序特性有效缓解会话中的时间信息。此外，该方法从多阶段演进的视角去理解会话推荐任务。具体而言，该方法将完整会话划分为逐时间步推进的若干子会话。通过每一个阶段的子会话推荐，模型可以深刻理解用户兴趣的变化逻辑，从而提升模型所提取的会话表示的质量。

研究贡献如下：

- 1) 通过构建基于自注意力的置换不变特征提取层，模型在有效提取会话表征的同时，缓解会话表示对于交互顺序的过度依赖。
- 2) 通过引入多阶段自监督学习框架，模型能够有效学习会话的动态推荐过程。这使得模型能够捕捉用户兴趣的动态转化，从而提取更具有表达力的会话特征。
- 3) 本文表征学习模型与下游的会话推荐任务相结合，与其他先进会话推荐方法进行对比实验。实验通过多个推荐指标证明了本文模型的有效性。

## 2. 相关研究

### 2.1. 会话推荐

会话推荐旨在根据用户在单次会话中的交互行为预测其下一步可能的兴趣。现有研究者进行了多种会话表示方面的探索。序列模型如 GRU4Rec 通过门控循环单元对行为序列进行建模，实现了对用户短期兴趣的预测。图神经网络方法如 SR-GNN [8] 通过将会话表示为图结构，利用节点之间的依赖关系提取复杂的交互特征。尽管这些方法在会话推荐中取得了一定进展，但它们往往依赖交互顺序，且对用户兴趣的动态变化刻画不足。

随着推荐场景的日益复杂，自监督学习与多阶段学习也逐渐成为模型的优化策略。自监督学习方法旨在通过精心设计的自监督任务，挖掘交互间的隐含联系。如 BERT4Rec [11] 采用基于 Transformer 的双向自注意力架构，通过掩蔽自预测的方式学习上下文信息，从而增强表征的表达力。GRec [12] 通过编码器-解码器对部分掩蔽的会话项目进行预测，在不直接暴露未来信息的情况下利用上下文关系，提升对用户行为模式的建模能力。多阶段方法则分阶段提取用户的会话特征，捕捉用户兴趣的演化过程。例如 Caser [13] 通过卷积操作在不同层次提取用户历史交互的局部模式，逐步捕获行为序列中的结构特征。A-PGNN [14] 将图神经网络与注意力机制结合，能够在用户历史和当前会话中连续建模复杂的物品关系，从而精细地刻画行为转移规律。本研究在结合了自监督与多阶段两个方案基础上，提出了多阶段自监督会话表示学习框架，从而更全面地捕捉用户兴趣变化。

### 2.2. 集合表示学习

集合表示学习旨在从无序且长度可变的元素集合中提取有意义的特征表示。传统方法多依赖核技巧，如 Fisher 核 [15]、Probability Product 核 [16] 及 Bhattacharyya 亲和度 [17] 等。这类方法通过计算集合向量的分布或相似性来进行建模，虽然能够在小规模数据上有效工作，但在计算效率和适用范围上存在明显局限。

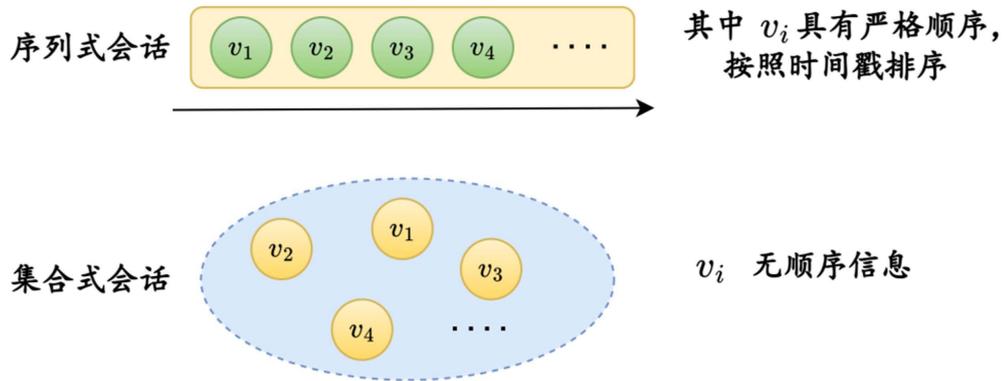
随着深度学习的发展，研究者提出了多种针对集合的神经网络方法。PointNet [18] 首次在深度网络中处理无序集合，通过池化操作实现顺序不敏感的特征聚合，完成下游任务。DeepSets [19] 从理论上证明了置换不变性的实现方式，并提出通用框架。Murphy 等 [20] 的研究表明，通过设计更复杂的池化算子，可以显著增强集合表示的学习能力和下游任务性能。Set Transformer [21] 引入的自注意力机制，使模型能够在处理集合无序性与不定长的同时捕捉元素间的关联信息。这类方法虽然不严格保证置换不变性，但置换等变性机制在实践中可以成为置换不变性的有效构成成分。

基于该研究思路，本研究将用户会话中的历史交互视为集合输入，以减少对交互顺序的依赖。自注意力机制在此框架中用于对集合元素进行编码，使模型能够在缓解顺序依赖的同时，充分捕捉交互间的潜在关联，从而提取具表达力的会话表征。

## 3. 模型

### 3.1. 问题表述

在基于会话的推荐中，设  $V = \{v_1, v_2, \dots, v_{|V|}\}$  表示所有项目（例如用户的点击、浏览与购买等操作）组成的总集合，其中  $|V|$  表示项目数量。与常规会话的序列形式不同，我们将会话  $S$  定义为集合形式，以  $S = \{v_1, v_2, \dots, v_n\}$  进行表示。会话推荐通常假设同一会话内的用户交互具有内在关联性和相似性 [22]，这使得整个会话可以被视为一个元素集合。基于这一视角，推荐任务可理解为对当前交互集合进行扩展，即识别并引入与现有交互特征高度相关的新项目，从而丰富会话的表示并捕捉用户潜在兴趣。



**Figure 1.** The distinction between sequentially modeled sessions and set-based sessions  
**图 1.** 序列式会话与集合式会话的区别

如图 1 所示，集合型会话要求模型在特征提取时避免顺序信息的引入，从而满足集合的无序性。基于上述表示，会话推荐问题转化为集合表示学习问题，可被定义为：

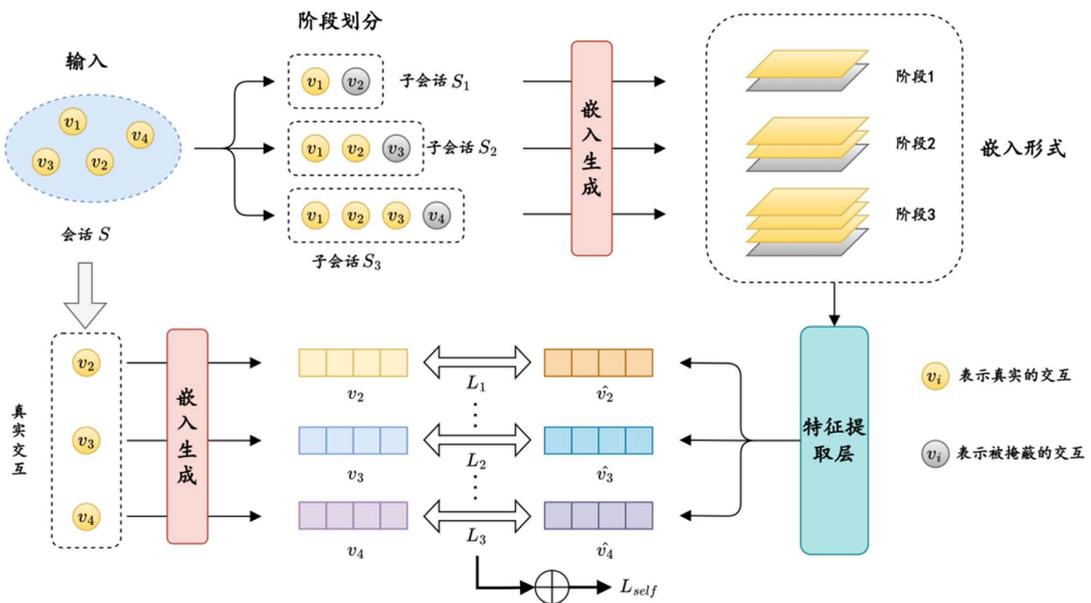
$$e_S = \mathcal{F}_{set}(S) \tag{1}$$

其中  $e_S$  为会话  $S$  的嵌入形式， $\mathcal{F}_{set}(\cdot)$  是指基于会话  $S$  的集合特征提取模型。

综上所述，本研究的关键在于构建一个满足集合特性的特征提取模型  $\mathcal{F}_{set}$ 。该模型结合自注意力机制以及多阶段自监督框架，提取会话表示，从而满足下游推荐任务需求。

### 3.2. 模型概述

在本节中，提出了一种面向会话表示的多阶段自监督集合学习方法 SR-MSSL，旨在缓解会话表征对交互顺序依赖的同时，更精细地刻画用户的兴趣变化。该模型将用户会话建模为集合，利用多阶段自监督框架来优化模型参数，从而提取更具有表达力的会话特征。



**Figure 2.** The full architecture of the SR-MSSL model  
**图 2.** SR-MSSL 模型的完整结构

SR-MSSL 的结构如图 2 所示。该模型的基本思想为将完整会话划分为代表不同推荐阶段的子会话，逐步往后进行下一项推荐。即让模型学习用户每一次交互的意图，以此来模拟每一次推荐，从而捕捉到用户兴趣的动态变化。

### 3.3. SR-MSSL 模型

#### 3.3.1. 特征提取层

从模型的结构上看，多阶段自监督框架会更为复杂，特征提取层会相对简单。但从整个表示学习的流程上看，特征提取却更为重要。多阶段自监督框架存在的目的是优化该提取结构的可学习参数。

在本文的设计中，特征提取层被构建为能够接收值输入的形式。具体而言，当集合作为输入数据时，对模型的置换不变性提出了要求。置换不变性指即便输入数据的相对顺序发生了变化，模型的输出不会发生改变。Zaheer 在 DeepSets 中给出了该模型设计的一般范式：

$$SL(\{x_1, x_2, \dots, x_n\}) = \sigma\left(\phi\left(\{\omega(x_1), \omega(x_2), \dots, \omega(x_n)\}\right)\right) \quad (2)$$

其中  $SL(\cdot)$  为集合学习函数， $\omega(\cdot)$  为嵌入提取函数， $\phi(\cdot)$  为求和函数， $\sigma(\cdot)$  为前馈输出。从该等式出发，可以拆解出一个基于“编码 - 解码”的通用框架。其中编码器用于交互信息的提取，解码器则用于信息聚合输出。

更进一步地，Zaheer 还发现如果编码器被设计为置换等变模型，该“编码 - 解码”框架同样表现出置换等变性。置换等变性与置换不变性相反，是指当模型输入的相对顺序发生变化时，输出发生同样置换的顺序变化。自注意力机制符合这一特性，证明如下：

设输入序列为  $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{n \times d}$ ，其中  $n$  是序列长度， $d$  为特征维度。通过可学习的线性变换矩阵  $W_Q$ ， $W_K$ ， $W_V \in \mathbb{R}^{d \times d}$  计算获得输入的查询  $Q = W_Q$ 、键  $K = W_K$  和值  $V = W_V$ 。自注意力输出即为：

$$At(X) = \text{soft max} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (3)$$

设任意置换矩阵  $P \in \{0, 1\}^{n \times n}$ ，对输入序列进行置换，可得  $X^* = PX$ 。对应的查询、键和值变为：

$$Q^* = X^*W_Q = PXW_Q = P(XW_Q) = PQ \quad (4)$$

$$K^* = X^*W_K = PK, V^* = X^*W_V = PV \quad (5)$$

则自注意力输出变为：

$$At(X^*) = \text{soft max} \left( \frac{Q^*K^{*T}}{\sqrt{d_k}} \right) V^* = P \text{soft max} \left( \frac{QK^T}{\sqrt{d_k}} \right) \cdot P^T PV \quad (6)$$

由于  $P$  为置换矩阵，满足  $P^T P = I$ ，注意力输出为：

$$At(X^*) = P \cdot \text{soft max} \left( \frac{QK^T}{\sqrt{d_k}} \right) V = P \cdot At(X) \quad (7)$$

证毕，即自注意力具有置换等变性。

综上，本研究选用了 Transformer 架构作为特征提取层。然而，传统的 Transformer 架构通过位置编

码引入了顺序信息，与本文集合建模的需求不符。因此本文选择使用去除了位置编码与 dropout 层的 Transformer 变体作为特征提取层的主体。

在解码器的选择上，本研究采用了直接将各交互提取的嵌入求和的方式。求和函数具有天然的置换不变性。由此搭建的“编码 - 解码”框架实现了从置换等变的输入到置换不变的输出，满足了集值输入问题的需求。最后，通过一个线性层对解码器的输出进行线性变换，得到所需维度的输出。综上，模型的特征提取“编码 - 解码”框架如图 3 所示：

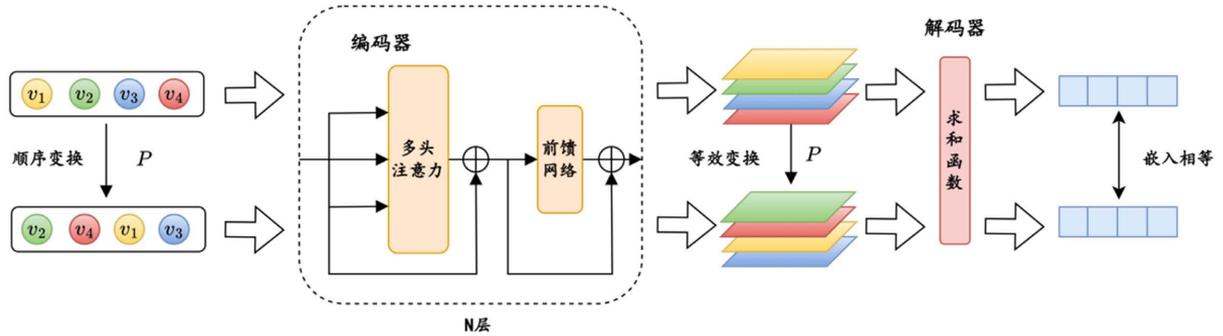


Figure 3. The “encoder-decoder” framework in the feature extraction layer  
图 3. 特征提取层的“编码 - 解码”框架

### 3.3.2. 多阶段自监督框架

多阶段自监督通过会话的多阶段自预测实现。该方法的主要思路在于将会话划分为多个逐步扩展的子序列，每个子序列代表一个逐渐积累的用户交互历史，从而模拟推荐过程中的不同阶段。在每一个阶段，利用当前的用户交互序列  $S_i$  作为输入，对下一项推荐  $\hat{v}_{i+1}$  进行预测。例如，首先使用包含初始交互的子序列  $S_1 = \{v_1\}$  预测下一项推荐  $\hat{v}_2$ ，接着使用包含更多交互的子序列  $S_2 = \{v_1, v_2\}$  预测进一步的推荐  $\hat{v}_3$ ，如此类推。这种方法通过逐步引入更多的用户历史数据，能够更精细地捕捉用户兴趣的动态变化。

自预测任务旨在恢复会话中目标交互  $v_i$  的嵌入  $e_i$ ，通过在推荐结果与真实交互间建立自监督信号，使模型能够捕捉全局偏好。该过程可形式化为一个回归任务：

$$\text{Reg}(S_i, v_i) = \frac{1}{|S|} \sum_{S_i \in S} \|f_\eta(S_i) - e_i\|^2 \quad (8)$$

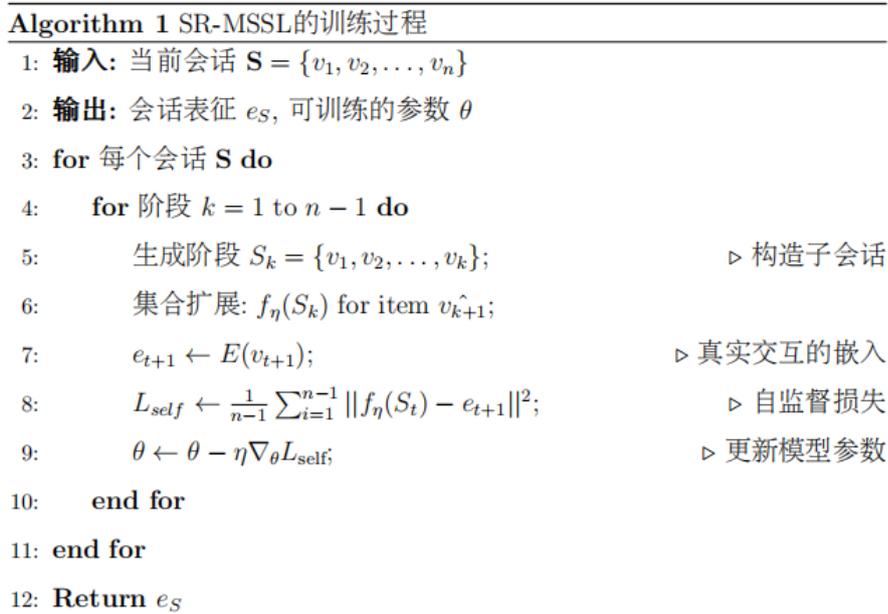
其中  $f_\eta(S_i)$  表示模型的推荐结果。基于上述表述，该自监督任务的损失函数可以写作：

$$L_{\text{self}} = \frac{1}{n-1} \sum_{i=2}^n (\hat{v}_i - E(v_i))^2 \quad (9)$$

其中  $n$  为交互总数， $n-1$  表示推荐阶段即子会话的数量， $\hat{v}_i$  表示模型输出的预测， $v_i$  表示用户真正的交互。

### 3.3.3. 模型训练

在本小节，本文将介绍模型的训练方式。针对长度为  $n$  的会话  $S$ ，本文将将其划分为逐步推进  $n-1$  个子会话，各自进行自预测推荐。最终，损失的加和作为模型的自监督损失用于优化模型参数。SR-MSSL 的伪实现如图 4 所示。



**Figure 4.** Pseudocode of the SR-MSSL model  
**图 4.** SR-MSSL 的伪代码

## 4. 实验

在本节实验中，我们通过与多种基线模型的对比，验证了所提方法在下游推荐任务中的有效性。同时，进一步考察是否引入位置编码对模型性能的影响，从而揭示集合建模与序列建模在会话表示学习中的差异，并凸显集合化表示在推荐场景下的优势。

### 4.1. 数据集及数据集预处理

本研究选择了两个具有代表性的真实数据集 Yoochoose 和 Diginetica 作为输入数据。其中 Yoochoose 来自 RecSys 2015 挑战赛，这是由真实电商平台的用户交互记录构成的会话数据集。Diginetica 则来自 CIKM Cup 2016，本文实验中仅保留其中的交易数据部分以保证研究的针对性。

为确保实验结果的可比性，本文在数据预处理阶段参考了已有工作的做法[3][4]。去除了仅包含单一项目的会话，因为这类样本无法有效提供序列或集合建模所需的上下文信息。此外，Yoochoose 数据集较大，有研究证明在最近的数据集上表现优于完整数据集。因此本研究选用 Yoochoose 1/64 版本，即最近的 1/64 会话数据。数据统计表如表 1 所示。

**Table 1.** Dataset statistics used in the experiments

**表 1.** 实验中所用数据集的统计

| Dataset        | Yoochoose1/64 | Diginetica |
|----------------|---------------|------------|
| clicks         | 557,248       | 982,961    |
| items          | 16,766        | 43,097     |
| train          | 369,859       | 719,470    |
| test           | 55,898        | 600,858    |
| average length | 6.16          | 5.12       |

## 4.2. 基线模型

为了验证本文所提 SR-MSSL 模型在会话表征提取上的有效性，我们在下游推荐任务中与多种代表性方法进行了对比。这些方法包括传统推荐基线 POP、Item-KNN 和 FPMC，深度学习模型 GRU4REC、SGNN-HN、SR-PredAO，以及同样采用集合型建模的 TSESRec。

- 1) POP: 一种最基础的会话推荐模型，通过提取训练集中最常被用户交互的项目来进行推荐。
- 2) Item-KNN: 利用余弦相似度，推荐与当前会话中已点击项目相似的其他项目。
- 3) FPMC: 基于马尔可夫链与矩阵分解相结合的序列预测模型，用于推荐下一个可能的项目。
- 4) GRU4REC: 利用循环神经网络对用户的交互序列进行建模，通过小批量并行训练将会话数据串联处理，并采用排序损失进行优化，从而捕捉序列中的行为模式。
- 5) SGNN-HN: 基于星型图神经网络构建项目之间的复杂转移关系，并结合高速网络动态选择合适的项目嵌入，以增强模型对会话结构的适应能力。
- 6) SR-PredAO: 这是一个通用的会话推荐框架，结合高效的预测附加模块(Predictor Add-On)，提升了推荐准确性和系统性能。本文将应用于 SGNN-HN 模型进行对比实验。
- 7) TSESRec: 以 Transformer 为核心驱动的集合扩展模型，将会话推荐任务重新表述为集合状态的建模问题。通过集合状态记忆模块和概率阈值门控策略，实现对历史偏好与新项目探索之间的平衡。

## 4.3. 评价指标

本文使用会话推荐研究中应用广泛的评估指标  $P@K$  和  $MRR@K$  来比较各模型的性能。

$P@K$  (Precision at  $K$ ): 衡量的是目标项目是否出现在预测排序的前  $K$  个位置。其定义为:

$$P@K = \frac{n_{hit}}{N} \quad (10)$$

其中,  $N$  表示测试样本数量,  $n_{hit}$  代表在前  $K$  个排名中包含目标项目的样本数, 若目标项目位于前  $K$  位, 则指标取 1, 否则为 0。在实验中, 本文主要采用  $P@10$  和  $P@20$ 。

$MRR@K$  (Mean Reciprocal Rank at  $K$ )通过计算目标项目排序位置的倒数来衡量整体推荐效果, 若排名超过  $K$ , 则贡献值为 0, 其定义为:

$$MRR@K = \frac{1}{N} \sum_{v_i \in V} \frac{1}{\text{Rank}(v_i)} \quad (11)$$

其中  $\text{Rank}(v_i)$  是项目  $v_i$  在总项目集  $V$  中的排名。

两个指标的取值范围均在  $[0,1]$ , 数值越大说明推荐系统能够更稳定地将目标项目置于前列, 从而表现出更优的推荐质量。

## 4.4. 实验设置

本文提出的 SR-MSSL 模型在下游推荐任务上通过网格搜索和早停策略确定了超参数配置。最终, 本文将交互项目的维度设置为 512, 模型层数为 1, 多头注意力机制中的头数为 16。在训练过程中, 本文采用小批量的 Adam 优化器进行参数更新。批次大小为 256, 训练轮次为 10。考虑到数据集规模较大且迭代次数较多, 初始学习率  $lr$  设置为  $10^{-4}$ 。

## 5. 实验结果与分析

### 5.1. 性能对比实验

为了全面评估 SR-MSSL 模型的表现, 我们在下游推荐任务中与多种代表性会话推荐方法进行了对

比。实验结果如表 2 所示, 采用 P@K 与 MRR@K 两类指标进行衡量, 其中数值越高代表模型性能越优。表中加粗部分标记了各指标下的最优结果, 所有数值均以百分比(%)形式呈现。

**Table 2.** Performance comparison results (SR-MSSL\* indicates the integration with recommendation tasks)

**表 2.** 性能对比实验结果(SR-MSSL\*表示 SR-MSSL 与下游推荐任务的结合)

| Method          | Yoochoose 1/64 |              |              |              | Diginetica   |              |              |              |
|-----------------|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                 | P@10           | MRR@10       | P@20         | MRR@20       | P@10         | MRR@10       | P@20         | MRR@20       |
| POP             | 4.98           | 2.03         | 6.71         | 1.65         | 0.66         | 0.25         | 1.11         | 0.28         |
| FPMC            | 38.87          | 16.59        | 45.62        | 15.01        | 18.07        | 7.13         | 26.53        | 6.95         |
| Item-KNN        | 33.54          | 11.42        | 51.60        | 21.81        | 25.07        | 10.77        | 35.75        | 11.57        |
| GRU4REC         | 47.11          | 21.67        | 60.64        | 22.89        | 17.93        | 7.33         | 29.45        | 8.33         |
| SGNN-HN         | 61.52          | 31.30        | 72.06        | 32.61        | 37.21        | 15.99        | 55.67        | 19.54        |
| SR-PredAO       | 62.94          | 31.34        | 72.62        | 32.47        | 38.27        | 16.63        | 55.91        | 19.06        |
| TSESTRec        | 61.77          | 32.13        | 69.54        | 33.37        | 41.04        | 19.53        | 53.61        | 21.47        |
| <b>SR-MSSL*</b> | <b>65.44</b>   | <b>34.03</b> | <b>73.12</b> | <b>33.69</b> | <b>43.34</b> | <b>22.97</b> | <b>57.66</b> | <b>24.18</b> |

试验结果表明, 所提方法在下游任务上的表现全面超越基线模型。具体而言, 传统方法 POP、FPMC 与 Item-KNN 整体表现较低。这表明传统的基于马尔科夫链的方法在序列依赖性假设上可能并不可靠。

基于神经网络的方法在性能表现上整体优于传统算法, 这表明神经网络在会话推荐领域具有强大的应用潜力。作为第一个将 RNN 引入会话推荐领域的方法, GRU4REC 相较于传统算法实现了显著的性能提升。而 SGNN-HN 与 SR-PredAO 通过图建模方法在序列建模的基础上, 更进一步提升性能。这些方法标志着会话推荐模型表达能力的突破。本文研究方法的下游任务表现优于上述基线方法, 这证明了集合型会话建模在会话推荐领域具有探索前景。

值得注意的是, 本文模型还优于同样属于集合型建模的 TSESTRec 方法。据分析, TSESTRec 通过会话的单次自监督重构实现模型优化, 从而提升会话嵌入的表达能力。相比而言, SR-MSSL\*的多阶段自监督框架则更精细地刻画用户兴趣的动态变化, 从而达到了更加优秀的表现。

### 5.2. 位置编码消融

本文方法的核心在于将会话表示为集合。虽然前文实验已验证了其有效性, 但其在缓解顺序依赖方面的作用仍需进一步检验。为此, 我们设计了位置编码对比实验: 在相同参数下, 将加入位置编码的 SR-MSSL\*作为有序模型, 与原始无序模型进行性能比较。实验结果如表 3 所示, 加粗部分为最佳表现。

**Table 3.** Comparison of performance between unordered and ordered sessions in the SR-MSSL\* model

**表 3.** SR-MSSL\*模型下无序和有序性能对比

| Method         |         | SR-MSSL* (集合) | SR-MSSL* (序列) |
|----------------|---------|---------------|---------------|
| Dataset        | Metrics |               |               |
| Yoochoose 1/64 | P@10    | <b>73.14</b>  | 70.26         |
|                | MRR@10  | <b>31.45</b>  | 29.57         |
|                | P@20    | <b>74.58</b>  | 71.06         |
|                | MRR@20  | <b>32.76</b>  | 31.85         |

续表

|            |        |              |       |
|------------|--------|--------------|-------|
| Diginetica | P@10   | <b>58.45</b> | 52.14 |
|            | MRR@10 | <b>26.42</b> | 21.8  |
|            | P@20   | <b>60.28</b> | 56.73 |
|            | MRR@20 | <b>30.21</b> | 28.19 |

实验结果表明，基于集合建模的 SR-MSSL\*在四个评价指标上均优于加入位置编码的序列化版本。在 Yoochoose 1/64 数据集上，集合模型在 P@10 与 P@20 上分别比序列模型提升了 2.88%和 3.52%，MRR@10 与 MRR@20 也分别提高了 1.88%与 0.91%。同样地，在 Diginetica 数据集上，集合建模同样展现出显著优势，其中 P@10 与 P@20 分别提升了 6.31%和 3.55%，MRR@10 与 MRR@20 的提升幅度则分别为 4.62%与 2.02%。

这些结果充分说明，集合建模能够在保持推荐准确性的同时，有效缓解模型对交互顺序的过度依赖，提升模型推荐性能。进一步分析其原因，可以认为集合范式带来了两方面优势：一方面，它在会话内部不依赖局部顺序约束，更容易捕捉到相隔较远的项目之间的高阶关联，从而提升了对用户潜在兴趣模式的建模能力；另一方面，它对顺序噪声(如偶发点击或误操作)的敏感性显著降低，使得模型能够更稳健地刻画用户的核心偏好。尤其是在数据规模较小、交互行为稀疏的 Diginetica 数据集中，这两类因素的影响更为突出，因此集合方法相较序列方法展现出更明显的性能优势，验证了其在复杂场景下的稳健性与泛化能力。

### 5.3. 参数敏感性分析

为了考察模型在不同设置下的稳定性与表现，我们进一步进行了参数敏感性分析。实验仍在 Yoochoose 1/64 和 Diginetica 两个数据集上进行，且保持与前文一致的预处理方式。在具体实验中，我们主要关注两类关键超参数。一是批量大小，影响模型的训练效率与优化效果。二是嵌入维度，用于控制会话表示的表达能力。通过系统调整这两个参数，并在 P@20 与 MRR@20 指标下进行比较，可以更直观地观察超参数对模型性能的影响。

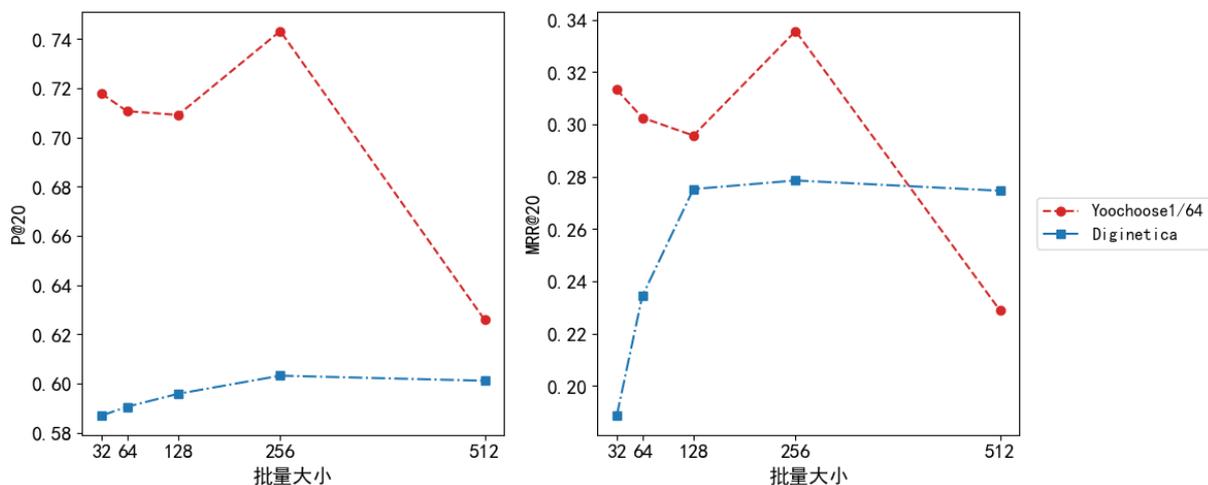


Figure 5. Batch size sensitivity analysis results  
图 5. 批量大小参数敏感性分析结果

从图 5 中可以看出,随着批量大小的增加,模型在两个数据集上的 P@20 和 MRR@20 指标整体呈现先上升后下降的趋势。在 Yoochoose1/64 数据集上,当批量大小为 256 时,模型在 P@20 和 MRR@20 上均达到峰值,分别为 74.58% 和 32.76%,而较小或过大的批量大小会导致性能下降。在 Diginetica 数据集上,随着批量大小的增加,指标变化相对平缓,但在 128 到 256 之间也出现了小幅提升,显示模型对批量大小具有一定敏感性。据分析,过小的批量可能导致模型难以充分捕捉会话中的全局偏好,过大的批量则可能使训练收敛不够精细,降低性能。

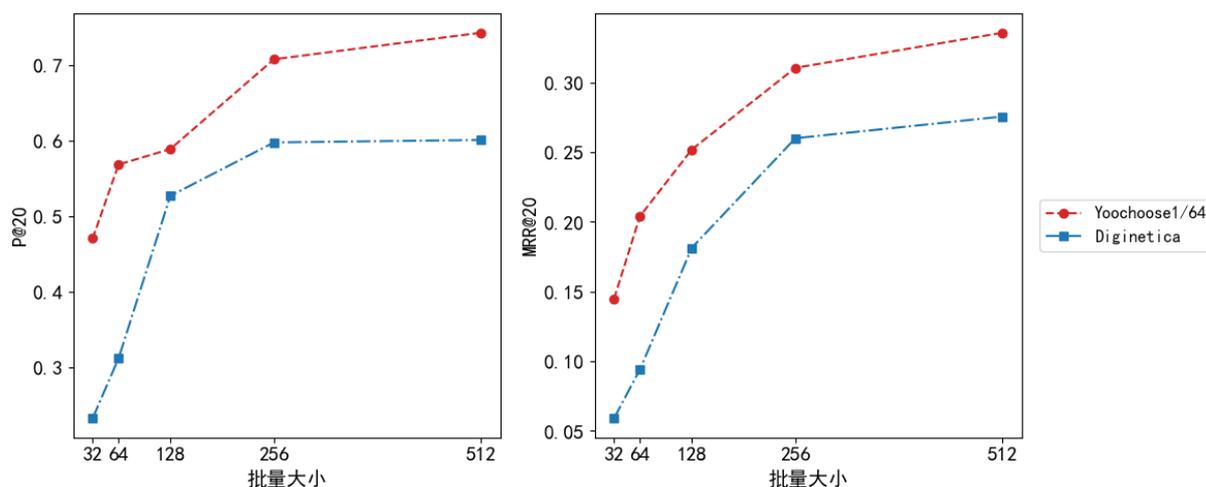


Figure 6. Sensitivity analysis results for embedding dimension

图 6. 嵌入维度参数敏感性分析结果

如图 6 所示,随着嵌入维度从 32 逐步增加到 512,模型在两个数据集上的 P@20 和 MRR@20 指标均呈明显上升趋势。在 Yoochoose1/64 数据集上, P@20 从 47.1% 提升至 74.58%, MRR@20 从 14.5% 提升至 33.6%; 在 Diginetica 数据集上, P@20 从 23.3% 提升至 60.1%, MRR@20 从 5.9% 提升至 27.6%。这表明,增大嵌入维度能够显著提升会话表示的表达能力,使模型更充分地捕捉会话中交互项的潜在关系。此外,模型性能提升在 256 至 512 时上升放缓,提示模型在维度过大时的边际收益减小,且带来计算成本增加。

## 6. 总结与讨论

本文提出了一种面向会话表示的多阶段自监督集合学习方法 SR-MSSL,该方法将单次会话视作无序集合进行建模,从而有效避免推荐结果对交互顺序的过度依赖。在 SR-MSSL 模型设计中,通过多阶段自监督学习逐步挖掘用户兴趣的动态变化,实现对会话信息的深度捕捉与表征学习。为验证该方法的可行性,本文在实验中将会话表示为集合,并通过注意力机制对集合元素进行特征聚合,在不依赖交互顺序的条件下学习集合的内在关系。

在实验过程中,通过对比是否加入位置编码的模型,进一步验证了集合建模在处理会话数据时的优势。实验结果表明,对于会话数据,不考虑项目间顺序反而有助于提高推荐性能。此外,本文方法与传统及深度学习基线模型进行了性能对比,在 P@K 和 MRR@K 指标上均展现出明显优势。值得注意的是,尽管基于注意力机制的集合学习方法在特定数据集上取得良好表现,但没有与推荐领域内表现优异的基于 GNN 的方法实现有效结合。未来的研究工作可以在集合表示与图结构建模之间进行更深入的探索,从而进一步提升会话推荐的性能。

## 参考文献

- [1] Lei, C., Liu, Y., Zhang, L., Wang, G., Tang, H., Li, H., *et al.* (2021) SEMI: A Sequential Multi-Modal Information Transfer Network for E-Commerce Micro-Video Recommendations. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Virtual Event Singapore, 14-18 August 2021, 3161-3171. <https://doi.org/10.1145/3447548.3467189f>
- [2] Deldjoo, Y., Schedl, M., Cremonesi, P. and Pasi, G. (2020) Recommender Systems Leveraging Multimedia Content. *ACM Computing Surveys*, **53**, 1-38. <https://doi.org/10.1145/3407190>
- [3] 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*, **54**, 1-38. <https://doi.org/10.1145/3465401>
- [4] Rendle, S., Freudenthaler, C. and Schmidt-Thieme, L. (2010) Factorizing Personalized Markov Chains for Next-Basket Recommendation. *Proceedings of the 19th International Conference on World Wide Web*, Raleigh, 26-30 April 2010, 811-820. <https://doi.org/10.1145/1772690.1772773>
- [5] Wu, X., Liu, Q., Chen, E., He, L., Lv, J., Cao, C., *et al.* (2013) Personalized Next-Song Recommendation in Online Karaoke. *Proceedings of the 7th ACM conference on Recommender Systems*, Hong Kong, 12-16 October 2013, 137-140. <https://doi.org/10.1145/2507157.2507215>
- [6] Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T. and Ma, J. (2017) Neural Attentive Session-Based Recommendation. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore, 6-10 November 2017, 1419-1428. <https://doi.org/10.1145/3132847.3132926>
- [7] Liu, Q., Zeng, Y., Mokhosi, R. and Zhang, H. (2018) STAMP: Short-Term Attention/Memory Priority Model for Session-Based Recommendation. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 1831-1839. <https://doi.org/10.1145/3219819.3219950>
- [8] Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X. and Tan, T. (2019) Session-Based Recommendation with Graph Neural Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 346-353. <https://doi.org/10.1609/aaai.v33i01.3301346>
- [9] Wang, D., Du, R., Yang, Q., Yu, D., Wan, F., Gong, X., *et al.* (2024) Category-Aware Self-Supervised Graph Neural Network for Session-Based Recommendation. *World Wide Web*, **27**, Article No. 61. <https://doi.org/10.1007/s11280-024-01299-8>
- [10] Li, Y., Gao, C., Luo, H., Jin, D. and Li, Y. (2022) Enhancing Hypergraph Neural Networks with Intent Disentanglement for Session-Based Recommendation. *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Madrid, 11-15 July 2022, 1997-2002. <https://doi.org/10.1145/3477495.3531794>
- [11] Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., *et al.* (2019) BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Trans-Former. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, 3-7 November 2019, 1441-1450. <https://doi.org/10.1145/3357384.3357895>
- [12] Yuan, F., He, X., Jiang, H., Guo, G., Xiong, J., Xu, Z., *et al.* (2020) Future Data Helps Training: Modeling Future Contexts for Session-Based Recommendation. *Proceedings of The Web Conference 2020*, 20-24 April 2020, 303-313. <https://doi.org/10.1145/3366423.3380116>
- [13] Tang, J. and Wang, K. (2018) Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, Marina Del Rey, 5-9 February 2018, 565-573. <https://doi.org/10.1145/3159652.3159656>
- [14] Zhang, M., Wu, S., Gao, M., Jiang, X., Xu, K. and Wang, L. (2022) Personalized Graph Neural Networks with Attention Mechanism for Session-Aware Recommendation. *IEEE Transactions on Knowledge and Data Engineering*, **34**, 3946-3957. <https://doi.org/10.1109/tkde.2020.3031329>
- [15] Jaakkola, T.S. and Haussler, D. (1998) Exploiting Generative Models in Discriminative Classifiers. *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II*, Cambridge, 20 July 1999, 487-493.
- [16] Jebara, T., Kondor, R. and Howard, A. (2004) Probability Product Kernels. *The Journal of Machine Learning Research*, **5**, 819-844.
- [17] Kondor, R. and Jebara, T. (2003) A Kernel between Sets of Vectors. *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, Washington, 21-24 August 2003, 361-368.
- [18] Qi, C.R., Su, H., Mo, K., *et al.* (2017) PointNet: DEEP Learning on Point Sets for 3D Classification and Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 652-660.
- [19] Zaheer, M., Kottur, S., Ravanbakhsh, S., *et al.* (2017) Deep Sets. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 3394-3404.
- [20] Radford, A., Narasimhan, K., Salimans, T. and Sutskever, I. (2018) Improving Language Understanding by Generative

---

Pre-Training. <https://www.mikecaptain.com/resources/pdf/GPT-1.pdf>

- [21] Lee, J., Lee, Y., Kim, J., *et al.* (2019) Set Transformer: A Framework for Attention-Based Permutation-Invariant Neural Networks. *International Conference on Machine Learning*, Long Beach, 10-15 June 2019, 3744-3753.
- [22] Ou, Z., Xu, T., Su, Q., *et al.* (2022) Learning Neural Set Functions under the Optimal Subset Oracle. *Advances in Neural Information Processing Systems*, **35**, 35021-35034.