Modeling and Simulation 建模与仿真, 2025, 14(10), 235-247
Published Online October 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.1410620

Hans汉斯

基于集合表示学习的会话推荐

黄　晶

上海理工大学管理学院，上海

收稿日期：2025年9月10日；录用日期：2025年9月18日；发布日期：2025年10月22日

摘　要

会话推荐旨在预测用户在会话中的下一次交互。现有方法多依赖序列建模，假设交互存在严格的时间顺序，但实际场景中用户行为常呈现弱顺序或无序特征，易导致推荐不稳定。本文提出一种基于集合表示学习的会话推荐模型SetRec，将会话建模为无序集合，直接捕捉物品间的全局共现关系，降低对顺序噪声的依赖。为提升表征稳健性，模型引入自监督重构任务，通过随机遮蔽与恢复交互项，迫使编码器学习潜在依赖关系，并与推荐预测任务联合优化，实现“重构 + 预测”的协同增强。实验在Yoochoose与Diginetica两个公开数据集上进行，结果表明SetRec在准确率和排序质量上均优于主流方法，有效缓解了数据稀疏与顺序噪声带来的问题，验证了集合化建模与自监督机制在会话推荐中的优势。

关键词

集合表示学习，会话推荐，重构

Session-Based Recommendation with Set Representation Learning

Jing Huang

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

Received: September 10, 2025; accepted: September 18, 2025; published: October 22, 2025

Abstract

Session-based recommendation aims to predict the next interaction within a user session. Existing methods often rely on sequential modeling, assuming strict temporal dependencies among interactions. However, in real-world scenarios, user behaviors are usually weakly ordered or unordered, leading to unstable recommendations. This paper proposes SetRec, a session-based recommendation model based on set representation learning. By modeling sessions as unordered sets, SetRec directly captures global co-occurrence relationships among items and reduces sensitivity to sequential

文章引用: 黄晶. 基于集合表示学习的会话推荐[J]. 建模与仿真, 2025, 14(10): 235-247.
DOI: 10.12677/mos.2025.1410620

noise. To enhance representation robustness, the model introduces a self-supervised reconstruction task that randomly masks and recovers session items, forcing the encoder to learn latent dependencies. The reconstruction is jointly optimized with the recommendation prediction task, forming a "reconstruction + prediction" framework. Experiments on two public datasets, Yoochoose and Diginetica, demonstrate that SetRec outperforms mainstream methods in both accuracy and ranking quality, effectively mitigating the challenges of data sparsity and noisy order. These results verify the advantages of set-based modeling and self-supervised mechanisms for session-based recommendation.

Keywords

Session-Based Recommendation, Set Representation Learning, Reconstruction

1. 引言

随着互联网与移动设备的普及，用户在电商、短视频与社交平台上的交互行为呈现出海量化与碎片化特征，推荐系统在帮助用户快速发现感兴趣内容方面发挥着关键作用[1]。近年来，会话推荐(Session-Based Recommendation)逐渐成为推荐系统研究的重要方向，其核心目标是在缺乏长期历史信息的情况下，仅基于当前会话中的交互序列，预测用户的后续行为[2]。这类方法尤其适用于匿名用户或冷启动场景，因此在学术界与工业界均受到广泛关注。

早期方法主要依赖基于规则或统计的方法，如基于马尔可夫链[3]和 KNN [4]的方法，通过建模物品之间的转移关系来预测下一个交互。随着深度学习的发展，研究者提出了基于循环神经网络的序列建模方法，能够捕捉点击序列中的顺序依赖。随后，基于注意力机制的模型在增强短期兴趣表示方面取得了进展。近年来，图神经网络被引入会话推荐，通过构建会话图建模高阶依赖关系，进一步提升了推荐效果。Transformer [5]架构的引入则弥补了序列建模中长距离依赖捕捉的不足。尽管上述方法在不同程度上都推动了会话推荐的发展，但它们大多仍然依赖严格的交互顺序假设，难以充分刻画会话内部的全局共现关系。

针对上述问题，本文提出了一种基于集合表示学习的会话推荐模型 SetRec。与序列方法不同，SetRec 将会话建模为无序交互集合，从而更自然地捕捉物品之间的共现关系，降低对顺序噪声的依赖。同时，模型引入自监督重构机制：通过随机遮蔽与恢复部分交互项，迫使编码器学习潜在依赖关系，并在表示层面获得更鲁棒的全局特征。进一步地，本文将重构任务与推荐预测任务联合建模，实现"重构 + 预测"的协同优化，在提升表征能力的同时保证推荐效果。

本文的主要贡献包括：

1) 提出一种集合化的会话建模方法，突破了传统序列假设的限制，有效缓解顺序噪声问题；

2) 设计自监督集合重构任务，并与推荐预测联合优化，从而提升潜在表示的稳健性与泛化性；

3) 在 Yoochoose 与 Diginetica 两个公开数据集上的实验表明，所提方法在推荐准确率与排序指标上均优于多种主流模型，验证了集合化建模与自监督机制的有效性。

2. 相关研究

2.1. 集合表示学习

集合表示学习(Set Representation Learning)旨在对无序元素集合进行建模，并在保持置换不变性

(Permutation Invariance)的同时捕捉集合内元素之间的依赖关系[6]。其核心思想是将集合作为整体输入，通过编码函数映射为低维潜在空间中的表征，从而支持分类、生成或预测等下游任务。与传统的序列或图建模方式不同，集合表示学习强调输入的无序性，即集合中元素的排列顺序不应影响整体表征的输出。这一特性使其在点云处理、多实例学习以及会话推荐等任务中具有重要应用价值。然而，由于集合规模往往不固定，且集合内元素之间可能存在复杂的高阶交互关系，如何在保证置换不变性的同时高效捕捉局部与全局依赖，成为集合表示学习的核心问题。

研究者提出了多种适用于集合建模的神经网络方法。例如，PointNet [7]最早在点云分析任务中提出，利用简单的逐元素映射与全局池化机制实现了集合级别的表示，其核心思想是保证输入顺序不变性。随后，DeepSets [8]正式提出了在任意排列下保持不变的集合函数建模框架，通过加和、平均或最大池化的方式对集合进行聚合，成为集合表示学习的经典方法。在此基础上，RepSet [9]引入可学习的参考集合，通过建模输入集合与参考集合之间的匹配关系来获取更具判别力的表示，从而克服了单一池化带来的表达受限问题。进一步地，Set Transformer [10]借助注意力机制对集合元素间的交互进行显式建模，在提升表达能力的同时保持了置换不变性，为集合学习提供了新的方向。这些方法为集合表示学习奠定了坚实基础，也为会话推荐中摆脱时间顺序依赖、建模全局共现关系提供了理论支撑。

2.2. 会话推荐

会话推荐(Session-Based Recommendation, SR)关注在没有长期用户画像的条件下，仅利用单次会话中的交互行为为用户提供个性化推荐[2]。与传统基于用户历史的推荐不同，会话推荐需要在短时间内捕捉用户即时兴趣，并在有限信息下输出准确预测。

早期方法多基于邻域与规则建模，如 Item-KNN 通过计算物品之间的相似度来生成推荐[3]，FPMC 将矩阵分解与马尔可夫链结合以建模短期转移关系[4]。这类方法依赖共现统计，难以捕捉复杂的序列依赖。随着深度学习的发展，循环神经网络(RNN)被引入会话推荐，例如 GRU4Rec [11]利用门控循环单元建模点击顺序，显著提升了预测效果。之后，STAMP [12]进一步地研究引入注意力机制，强调用户最近一次点击的重要性，从而更好地平衡短期与长期偏好。

随着图神经网络的发展，研究者提出基于图的会话推荐方法。典型代表 SR-GNN [13]将会话转化为有向图，通过图神经网络捕捉物品之间的高阶依赖，并结合注意力机制获得全局兴趣表示。这类方法在多个公开数据集上取得了领先效果。这些方法虽然借助深度神经网络能够刻画会话中的复杂依赖关系，但普遍依赖严格的交互顺序假设，对于弱顺序或无序场景的研究仍相对有限[2]。在此背景下，本文在上述工作的基础上，引入适用于集合数据的“编码器–解码器”架构，使模型能够在不依赖交互顺序的条件下完成会话推荐任务。

3. 模型

3.1. 问题描述

在推荐系统中，设物品集合为$V=\left\{v_1,v_2,\cdots,v_{|V|}\right\}$，其中每个元素$v_i$表示一个候选物品。会话推荐是指在一个会话场景中，用户通过点击、浏览等行为，逐步与系统交互形成一段交互序列，记为$\boldsymbol{I}=\left[v_1,v_2,\cdots,v_n\right]$，其中$v_i\in V$表示用户在当前会话中第$i$次选择的物品，且这些交互通常按照时间顺序排列。会话推荐的任务目标是基于历史交互序列，预测用户在下一步最可能选择的物品$\hat{v}_{n+1}$，其形式化定义为$\hat{v}_{n+1}=g_{rec}\left(I\right)$，其中$g_{rec}\left(\cdot\right)$为推荐函数。

传统方法普遍认为用户的下一次交互依赖于交互序列，因此推荐模型大多基于时序建模。但在实际应用中，这种假设未必成立。用户在一次会话中往往会在短时间内浏览或点击多个相关物品，这些物品

之间的关系并不严格受时间顺序约束。如果推荐模型过度依赖时间顺序，就可能忽略物品之间的共现关系，从而导致推荐结果出现偏差甚至错误。为克服这一问题，本文提出将会话视为一个无序交互集合，以消除对严格顺序的依赖。设会话交互集合为 $\mathbf{S}=\{e_1,e_2,\cdots,e_n\}$，其中每个交互项 e_i 表示用户在会话过程中的一次交互行为。此时，会话推荐可定义为：

$$\hat{e}_{n+1}=f_{rec}(S) \tag{1}$$

其中 $f_{rec}(\cdot)$ 为基于集合表示学习的会话推荐模型。

这一集合化的表述能够更好地刻画交互项之间的全局依赖关系，从而更符合用户的真实兴趣模式。更进一步，通过结合自监督学习方法，还能在集合表示上捕捉到更加稳健和泛化的潜在特征，为后续推荐预测提供坚实基础。因此，会话推荐的研究逐渐从序列化建模转向集合化建模，并在复杂交互场景中展现出显著优势。

3.2. 模型概述

本文提出了一种基于集合自监督重构的会话推荐模型 SetRec，其整体目标是在会话推荐任务中学习更加稳健和表达力更强的潜在表征，从而提升推荐的准确性。与传统方法依赖时间顺序建模不同，SetRec 将会话数据看作无序集合，通过随机掩蔽部分交互项，并利用编码器学习全局潜在表征，再借助解码器对被掩蔽的交互进行重构，以此强化模型对交互间潜在关系的捕捉。同时，在得到的潜在表征基础上，模型还直接执行会话推荐任务，预测用户可能的后续行为。通过这种“重构 + 预测”的联合学习框架，SetRec 不仅能够增强表示学习的质量，还能在复杂场景下综合建模用户短期和潜在长期兴趣，实现更精准的推荐输出。

如图 1 所示，SetRec 模型主要由两部分组成：集合重构任务与推荐预测任务，两者在框架中承担不同角色并协同发挥作用。集合重构任务是模型的核心，通过对会话集合进行随机掩蔽，再利用编码器获取全局潜在表征，并由解码器重构被掩蔽的交互项，从而促使模型捕捉交互之间的隐式关联，提升对用户偏好的整体理解。推荐预测任务则基于同一潜在表征，对候选物品进行打分，直接预测用户的下一步行为，实现会话推荐的目标。二者结合，一方面利用自监督重构强化表示学习，另一方面通过监督预测保证模型在实际推荐任务中的有效性，从而使 SetRec 在表达能力与推荐性能之间形成互补。

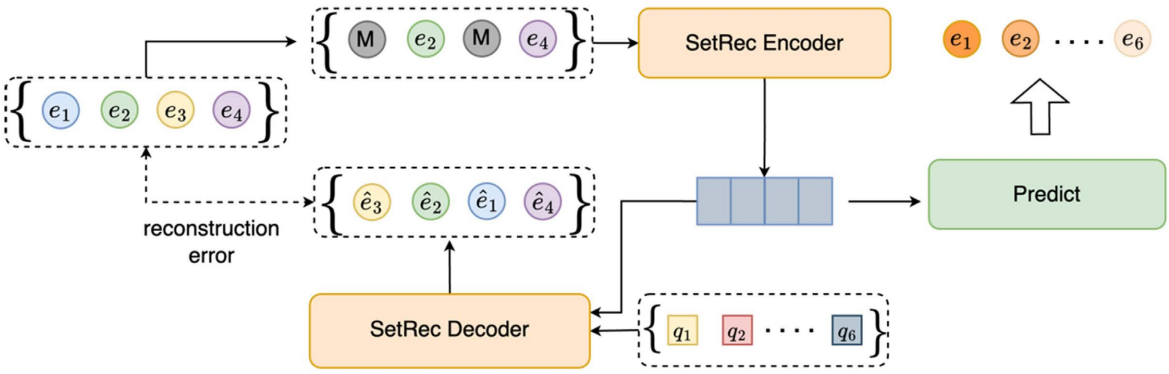

Figure 1. SetRec model
图 1. SetRec 模型

3.3. SetRec 模型

3.3.1. 会话输入与随机遮掩

在 SetRec 模型中，会话数据被建模为一个无序的交互集合 $\mathbf{S}=\{e_1,e_2,\cdots,e_n\}$，其中每个元素 e_i 表示用

户在会话过程中的一次交互行为。与传统依赖时间顺序的序列化表示不同，集合化的输入方式避免了对交互顺序的过度依赖，使模型能够更关注交互项之间的潜在共现关系与整体偏好结构。为了进一步增强表征的泛化能力，并为自监督任务提供学习信号，本研究在输入阶段引入了随机遮掩策略。遮掩比例的设定则在一定程度上影响模型的学习效果，遮掩过少会导致训练信号不足，遮掩过多则会削弱上下文信息，因此在实验中需要合理选择，具体情况会在第 4 节实验部分展开详细讨论。

通过这种随机遮掩机制，模型不仅能够模拟实际推荐场景中信息缺失或稀疏的情况，还能在训练过程中被迫学习交互项之间的潜在依赖关系与语义结构，从而提升对用户全局兴趣的建模能力。具体而言，从集合 $\mathbf{S}$ 中按照一定比例随机选取子集 $\mathbf{M}$，并用特殊标记符 M 替代这些位置，得到带有缺失信息的输入集合 $\tilde{\mathbf{S}}$：

$$\tilde{\mathbf{S}} = \left(\mathbf{S}/\mathbf{M}\right) \cup \{M\} \tag{2}$$

其中，$\mathbf{M}$ 表示被随机遮掩的元素子集，$\tilde{\mathbf{S}}$ 为最终输入。未被遮掩的元素提供已知信息，被替换的部分则作为模型需要恢复的目标。总体而言，会话输入与随机遮掩机制为后续的特征学习与集合重构奠定了坚实的基础，使 SetRec 能够更有效地捕捉用户偏好的多样性与潜在模式。

3.3.2. 集合型会话特征学习

在会话推荐中，用户交互往往以集合形式出现，即在短时间内点击多个项目。与序列推荐不同，这类会话数据不依赖交互顺序，而更关注集合内元素的共现模式与语义关系。

以购物场景为例，假设用户在一次会话中购买了{手机壳，充电器，耳机}。用户 A 的顺序可能是“手机壳→充电器→耳机”，而用户 B 的顺序可能是“耳机→手机壳→充电器”。从语义角度看，这两次会话的特征均为〈手机配件类〉。然而序列建模会将二者视为完全不同的输入，从而学习到不一致的表示；而集合建模则只关注集合中物品的共现关系，而非交互顺序，因此两种会话都能被压缩为相同的潜在语义〈手机配件类〉。这样既能避免顺序带来的噪声，又能更准确地刻画用户真实兴趣。因此，需要一种对顺序不敏感的集合特征学习方法，从整体会话中提取潜在表征。

在集合型会话特征学习中，模型需要满足置换不变性。这意味着集合型会话中的项目元素无论以何种排列顺序输入，最终学习到的潜在表征应该保持一致：

$$SL\left(\{e_1, e_2, \cdots, e_n\}\right) = SL\left(\left\{e_{\pi(1)}, e_{\pi(2)}, \cdots, e_{\pi(n)}\right\}\right) \tag{3}$$

其中 $\pi(\cdot)$ 表示任意排列。

为此，本研究引入基于注意力机制的集合型编码器。该编码器通过自注意力捕捉集合中元素的高阶交互关系，并通过全局聚合获取会话潜在表征。与传统 Transformer 编码器不同，本模型去除了位置编码与 dropout，以保证顺序不敏感性与信息完整性。同时，注意力机制可自适应分配不同项目间的权重，更好地建模集合的共现模式[14]。

集合型编码器由多层自注意力构成。设输入集合 $\tilde{\mathbf{S}}$ 经嵌入层得到序列 $\mathbf{X}$。单头自注意力对输入置换是对称的，多头注意力由单头拼接而成，因此也具备置换等变性。前馈神经网络(FFN)和层归一化(LN)均为逐行操作，不影响置换等变性。因此，经多层自注意力与聚合模块处理后，输入序列 $\mathbf{X}$ 的潜在特征表示 $\mathbf{M}$ 仍满足置换等变性，可表示为：

$$\mathbf{M} = AGG\left(\left(FFLN \circ MAttn\right)^{L}\left(\mathbf{X}, \mathbf{X}, \mathbf{X}\right)\right) \tag{4}$$

其中，$M\mathrm{Attn}\left(\mathbf{X}, \mathbf{X}, \mathbf{X}\right)$ 表示多头自注意力机制；$FFLN\left(\cdot\right)$ 表示前馈神经网络和层归一化的组合，$\left(\cdot\right)^{L}$ 表示堆叠 L 层操作，$AGG\left(\cdot\right)$ 为全局聚合函数。

3.3.3. 自监督强化会话特征

会话推荐的关键信号是共现而非顺序。常规方法通常将每次交互视为独立样本，仅保留离散的项目特征，忽略了项目之间的依赖关系，得到的表示偏向“点信息”，缺乏“组语义”。然而，用户的短期偏好往往表现为一段时间内的整体意图，这依赖于会话内部的共现模式。为弥补这一不足，本文引入自监督重构机制：在输入集合中随机遮蔽部分元素，要求模型仅依赖剩余上下文进行恢复。只有当潜在表征能够刻画项目间的协同关系时，重构才可能成功，因此该机制迫使表示学习内化共现结构，并在稀疏场景下获得更稳健的特征。

在实现上，首先在特征提取阶段得到会话潜在表征 $\mathbf{M}$ 。为避免引入顺序信息，解码器结构基于 Transformer，但移除了位置编码与 dropout 模块，以减少顺序干扰和信息丢失。解码器的输入由两部分构成：一是编码器输出的会话潜在表征 $\mathbf{M}$ ，承载集合级的上下文信息；二是与集合元素对应的可学习查询向量 $\mathbf{Q}$ ，用于在重构时引导模型定位输入集合 $\mathbf{S}$ 中的哪一个元素，这里采用基于匹配的方法[15]。解码器通过多头注意力机制建模查询与潜在表征的交互，其输出可表示为：

$$P = AGG\left(\left(FFLN \circ MAttn\right)^{L}\left(\mathbf{Q},\mathbf{M},\mathbf{M}\right)\right) \tag{5}$$

其中，$M\mathrm{Attn}\left(\cdot\right)$ 表示多头注意力机制；$FFLN\left(\cdot\right)$ 表示前馈神经网络和层归一化的组合，$\left(\cdot\right)^{L}$ 表示堆叠 L 层操作，$AGG\left(\cdot\right)$ 为全局聚合函数。最终输出集合 $\mathbf{P}=\left\{p_1,p_2,\cdots,p_k\right\}$。接下来需要将预测集合 $\mathbf{P}$ 与真实集合 $\mathbf{S}=\left\{e_1,e_2,\cdots,e_n\right\}$ 对齐。由于集合建模要求不依赖于元素顺序，本文采用最优匹配算法在预测集合 $\mathbf{P}$ 与真实集合 $\mathbf{S}$ 之间建立一一对应关系。定义：

$$z_i=\mathrm{softmax}\left(p_{\pi(i)}\right),\ z_i\left[v\right]\in\left[0,1\right],\ \sum_{v\in V}z_i\left[v\right]=1$$

其中，z_i 表示在匹配方式 π 下，第 i 个预测向量对应的概率分布；V 表示所有候选类别的全集。则最优匹配为：

$$\hat{\pi}=\arg\min_{\pi\in Perm(k)}\sum_{i=1}^{n}L_{CE}\left(z_i,e_i\right) \tag{6}$$

在最优匹配 $\hat{\pi}$ 确定后，从对应的分布中取概率最大的类别作为预测结果

$$\hat{e}_i=arg\ \max_{v\in V}\ z_i\left[v\right]$$

$$\hat{S}=\left\{\hat{e}_1,\hat{e}_2,\cdots,\hat{e}_n\right\}$$

最后，优化目标是重构损失：

$$L_{rec}=\sum_{i=1}^{n}\mathrm{L}_{CE}\left(z_i,e_i\right) \tag{7}$$

其中，$z_i=\mathrm{soft\,max}\left(p_{\hat{\pi}(i)}\right)$ 是在最优匹配下得到的第 i 个预测分布。

3.3.4. 推荐预测

在完成集合重构并获得稳定的潜在表示后，模型进一步执行推荐预测任务。具体来说，利用自监督强化训练得到的会话表示 $\mathbf{M}$ ，通过一层线性变换映射到候选物品的得分空间：

$$\mathbf{z}=\mathbf{W}\cdot\mathbf{M}+\mathbf{b} \tag{8}$$

其中，$\mathbf{W}$ 和 $\mathbf{b}$ 为可训练参数，则 $\mathbf{z}$ 刻画了候选物品与当前会话的相关性。随后采用 softmax 将得分归一化为概率分布：

$$\hat{\mathbf{y}} = \mathrm{softmax}(\mathbf{z}) \tag{9}$$

该分布体现了不同候选物品被推荐的相对优先级。

在训练阶段，预测部分采用 Top-1 损失以最小化排序误差。给定正样本 i 和负样本 j，其形式可写为：

$$L_{pred} = \frac{1}{n}\sum_{i=1}^{n}\left(\sigma\left(\hat{y}_j - \hat{y}_i\right) + \sigma\left(\hat{y}_j^2\right)\right) \tag{10}$$

其中 $\sigma(\cdot)$ 是 sigmoid 函数，$\hat{y}_i$ 和 $\hat{y}_j$ 分别表示正、负样本的预测结果。该损失通过提升正样本得分、抑制负样本得分来优化推荐性能。最终，推荐预测与集合重构任务联合优化，整体目标为：

$$L = L_{rec} + \lambda L_{pred} \tag{11}$$

其中，λ 为平衡两部分损失的权重系数。这种多任务学习方式既能增强潜在表征的能力，又能提升推荐输出的准确度。

3.3.5. 模型训练

在模型训练阶段，SetRec 通过联合优化集合重构任务与推荐预测任务来学习稳健且具备泛化能力的潜在表征。具体而言，集合重构任务通过随机遮掩并重建会话集合，促使模型捕捉交互项之间的潜在关系与上下文信息；推荐预测任务则直接面向用户的下一步行为，确保模型在实际推荐场景中的有效性。二者结合能够实现表征学习与推荐预测的互补增强，从而提升整体性能。最终的损失函数由集合重构损失 L_{rec} 与推荐损失 L_{pred} 组成，其联合优化目标为 $L = L_{rec} + \lambda L_{pred}$。通过反向传播不断更新参数，模型能够在保持集合重构能力的同时，提升对下一交互物品的预测精度。整体训练算法如图 2 所示。

Algorithm 1 SetRec 训练流程

输入: 会话 $S = \{e_1, \ldots, e_n\}$，遮掩比率 ρ，权重 λ
输出: 下一物品预测 e_{n+1}
for 每个会话 S **do**
  $S \leftarrow \{e_1, \ldots, e_{n-1}\}$ ▷ 将会话分割为集合
  $\tilde{S} \leftarrow \mathrm{RandomMask}(S, \rho)$ ▷ 按比例 ρ 对会话 S 进行掩码
  $\mathbf{X} \leftarrow \mathrm{Embed}(\tilde{S})$
  $\mathbf{M} \leftarrow \mathrm{AGG}((\mathrm{FFLN} \circ \mathrm{MAttn})^L(\mathbf{X}, \mathbf{X}, \mathbf{X}))$ ▷ 编码器
  $\mathbf{P} \leftarrow \mathrm{AGG}((\mathrm{FFLN} \circ \mathrm{MAttn})^L(\mathbf{Q}, \mathbf{M}, \mathbf{M}))$ ▷ 解码器
  $L_{\mathrm{rec}} \leftarrow \mathrm{SetMatch}(S, \hat{S})$ ▷ 重构损失
  $\hat{\mathbf{y}} \leftarrow \mathrm{softmax}(W\mathbf{M} + b)$ ▷ 预测
  $L_{\mathrm{pred}} \leftarrow \mathrm{Top1Loss}(\hat{y}, e_n)$ ▷ 预测损失
  $L \leftarrow L_{\mathrm{rec}} + \lambda L_{\mathrm{pred}}$ ▷ 模型优化总损失
  $\psi \leftarrow \mathrm{Opt}(\psi, \nabla_\psi L)$ ▷ 反向传播优化
end for
return e_{n+1}

Figure 2. SetRec model training algorithm
图 2. SetRec 模型训练算法

4. 实验

4.1. 数据集及数据集预处理

本实验基于 Yoochoose 与 Diginetica 两个公开的会话推荐数据集进行。Yoochoose 数据集来源于

RecSys Challenge 2015，记录了用户在电商平台上的点击与购买行为，规模较大。参考 STAMP 的做法，本文直接使用社区整理后的 Yoochoose (1/64)数据集作为训练与测试数据。Diginetica 数据集则来自 CIKM Cup 2016，包含用户的搜索与点击记录，相比 Yoochoose 规模较小，但同样广泛用于会话推荐任务。

在预处理阶段，首先按照 SessionID 与时间戳顺序，将用户的交互划分为会话序列，并去除长度小于 2 的序列。随后，对出现过的所有商品建立索引字典(0 预留为填充符)，将会话序列映射为对应的索引序列。为了增强训练效果，模型输入会随机遮蔽一定比例的商品(默认 20%)，并要求模型同时完成序列重建与下一个商品预测两项任务。此外，所有序列均被截断或补齐至固定长度 $L = 20$，保证批量训练的统一性。经过处理后的数据集统计信息如表 1 所示，包括点击数、商品数、训练与测试会话数、平均会话长度以及平均物品出现频率。

Table 1. Experimental dataset statistics
表 1. 实验数据集统计

数据集	Click	Item	Train	Test	Session Length	Item Frequency
Yoochoose (1/64)	557,248	16,766	369,859	55,898	3.94	29.48
Diginetica	982,961	43,097	719,470	600,858	5.12	10.04

4.2. 实验设置

为了验证所提出模型的有效性，本文在统一的硬件环境下进行实验。模型训练与评测基于 Python 3.8 与 PyTorch 框架实现。首先在超参数设定方面，输入集合的最大长度为 20，批量大小为 128。编码器与解码器均为两层 Transformer 结构，项目交互维度设置为 128，注意力头数为 8。训练采用小批量 Adam 优化器，学习率初始化为10^{-4}，训练 30 个 epoch，并在训练过程中进行梯度裁剪(阈值 1.0)以保证稳定性。为了增强表征能力，训练时对输入序列随机遮蔽 20%的商品 ID，模型需同时完成两个任务：其一为集合重构任务，用于恢复被遮蔽的商品；其二为推荐任务，用于预测当前会话的下一个商品。测试阶段仅进行推荐任务，不再执行遮蔽操作。

4.3. 评价指标

为了全面评估模型在会话推荐任务中的表现，本文选取 Precision at K (P@K)和 Mean Reciprocal Rank at K (MRR@K)作为主要评价指标。

P@K：用于衡量测试样本中，前 K 个推荐项目中命中真实目标的比例，该指标能够直观反映模型的召回能力。本文采用 P@20，其数学形式为：

$$P@K = \frac{n_{hit}}{N}$$

其中，N 表示测试样本数，n_{hit} 表示预测结果的前 K 个推荐中包含目标商品的样本数。$P@K$ 值越高，说明模型在有限推荐空间内覆盖用户真实需求的能力越强。

MRR@K：主要考察推荐结果的排序质量，是指所有测试样本中，目标项目在前 K 个推荐中的倒数排名的平均值。其数学式子为：

$$MRR@K = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{Rank_i}$$

其中，Rank_i 表示第 i 个测试样本中目标项目在推荐列表中的排名，若排名超过 K，则该样本的贡献值为

0。该指标能够衡量目标商品在推荐列表中出现的优先程度。较高的 MRR@K 值表明系统不仅能够成功召回目标商品，而且能够在靠前位置展示给用户，从而提升用户的实际体验。

4.4. 基线模型

为了验证 SetRec 模型的性能，本文选取以下具有代表性的会话推荐模型作为基线进行对比。

1) POP：最基础的会话推荐模型，按训练集中物品的交互计数或点击率由高到低生成推荐。

2) S-POP：会话级流行度模型。仅汇总当前会话内的出现频次或局部共现强度，以刻画即时兴趣并输出结果。

3) Item-KNN：物品邻域协同过滤。用共现统计或相似度度量(如余弦、Jaccard)寻找与会话中已交互物品最相近的候选，并对邻居得分进行聚合得到推荐。

4) FPMC：将矩阵分解建模的长期偏好与一阶马尔可夫转移的序列依赖合并，估计从上一物品到下一物品的转移概率以进行下一步预测。

5) GRU4REC：基于 GRU 的会话序列模型。以点击序列为输入，采用会话并行的小批量训练，并用 TOP-1 或 BPR 等排序损失直接优化下一物品预测。

6) STAMP：注意力与记忆优先模型。对会话内各交互做注意力聚合形成一般偏好，同时强调末次点击的短期兴趣。

7) SR-GNN：把会话转为有向加权图。用门控 GNN 学习物品表示，再以注意力结合整体与当前兴趣得到会话向量并进行下一物品打分。

4.5. 实验结果

4.5.1. 性能对比实验

为了验证本文提出的 SetRec 模型的整体性能，本实验将 SetRec 模型与常见的会话推荐方法进行了比较。表 2 展示了 SetRec 模型与其他常见会话推荐模型分别在 P@20 和 MRR@20 指标上的表现对比，其中下划线显示最佳结果。

Table 2. System resulting data of standard experiment
表 2. 标准试验系统结果数据

模型	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
S-POP	12.14	11.71	30.44	18.35	23.89	13.92	21.06	13.68
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
STAMP	54.40	25.03	68.74	29.67	33.98	14.26	45.64	14.32
SR-GNN	60.34	30.04	70.57	30.94	36.86	15.52	50.57	17.59
SetRec	**61.50**	**30.23**	**68.48**	**30.93**	**41.32**	**18.27**	**54.34**	**19.01**

由上述对比实验结果分析发现，本文提出的模型 SetRec 优于传统基线模型(POP、S-POP、FPMC、

Item-KNN)和序列模型 GRU4Rec，证明了本文所提出的集合会话推荐模型的有效性。然而，在指标 Precision 上与当前模型 SR-GNN 还存在一定的差距，但基本上保持同一水平。这是因为 Yoochoose 数据集中的会话交互顺序较为显著，而 SR-GNN 通过图结构建模了严格的序列依赖与高阶转移关系，因此在基于召回率的指标上具有一定优势。而在 MRR 指标上，本文提出的模型 SetRec 则均优于现有模型。这充分说明了 SetRec 在推荐结果排序的整体质量上更具优势。

从模型机制上分析，SetRec 的优越性主要来源于两方面。一是，SetRec 不依赖于严格的时间顺序建模，而是通过集合表示学习充分挖掘交互项之间的潜在关系，从而有效缓解了传统序列模型在处理噪声和长依赖时的不足。二是，模型引入集合重构任务，通过随机掩蔽与解码的方式增强了表示学习的鲁棒性，使潜在特征能够更全面地捕捉上下文信息。这一自监督任务不仅提升了模型的泛化能力，还在一定程度上避免了过拟合。同时，预测任务与重构任务的联合优化使模型能够在保持表达能力的同时，直接服务于推荐目标，从而兼顾了特征学习与任务效果。这表明，突破传统序列依赖并引入集合视角进行会话建模是提升推荐质量的可行且高效的方向。

4.5.2. 超参数敏感性分析

本节探索模型 SetRec 的参数对模型性能的影响。为对比模型在不同超参数下的性能，本实验以数据集 Yoochoose 1/64 和数据集 Diginetica 作为实验对象，通过调整模型的 mask 比例、层数、嵌入维度和样本批量大小来探讨它们对准确率(P@20)的影响。

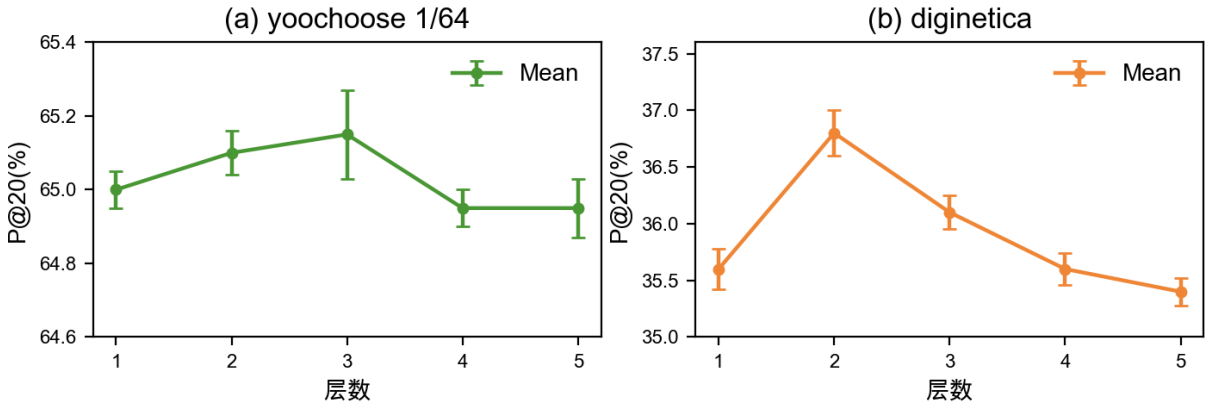

Figure 3. Impact of the number of layers on model performance
图 3. 层数对模型性能的影响

层数变化对模型 SetRec 性能的影响如图 3 所示。在两组数据集 Yoochoose 1/64 和 Diginetica 上都呈现出“先升后降”的趋势。在 Yoochoose 1/64 上，性能在 3 层时达到最高；而在 Diginetica 上，最佳结果出现在 2 层。超过这一深度后，模型效果下降，说明过深结构并未带来额外收益，反而引入了噪声与过拟合风险。该现象表明，SetRec 并不依赖过深堆叠，而是需要适度的层数来充分建模会话集合特征。

图 4 展示了掩码比例对模型 SetRec 性能的影响趋势。可以发现，在两个数据集上，性能随掩码比例的变化均呈现“适度最优”的规律。在 Yoochoose 1/64 数据集上，掩码比例为 0.2 时 P@20 指标最高，超过 72.2%，之后随着比例的增加，性能逐渐下降。在 Diginetica 数据集上，掩码比例同样在 0.2 时达到峰值。因此，选择合适的掩码比例能够在信息保留和特征扰动之间取得平衡，既能避免过拟合，又不会因过度掩盖而丢失重要上下文。

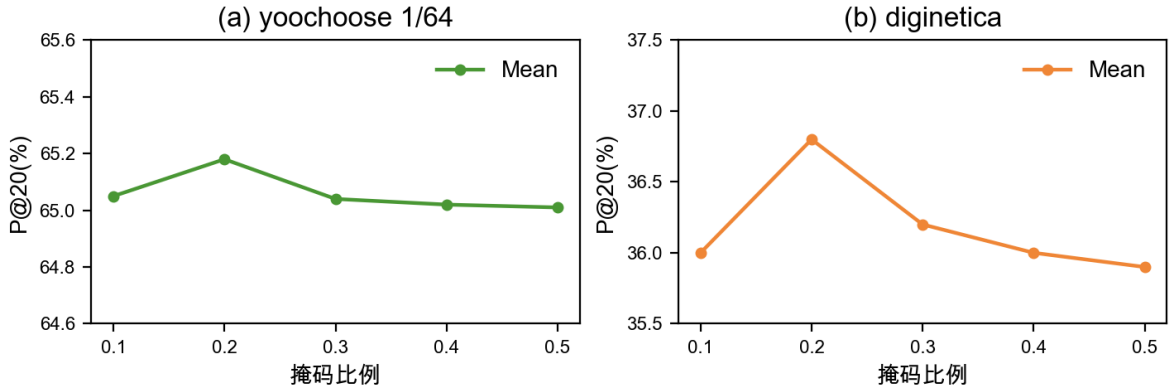

Figure 4. Impact of masking ratio on model performance
图 4. 掩码比例对模型性能的影响

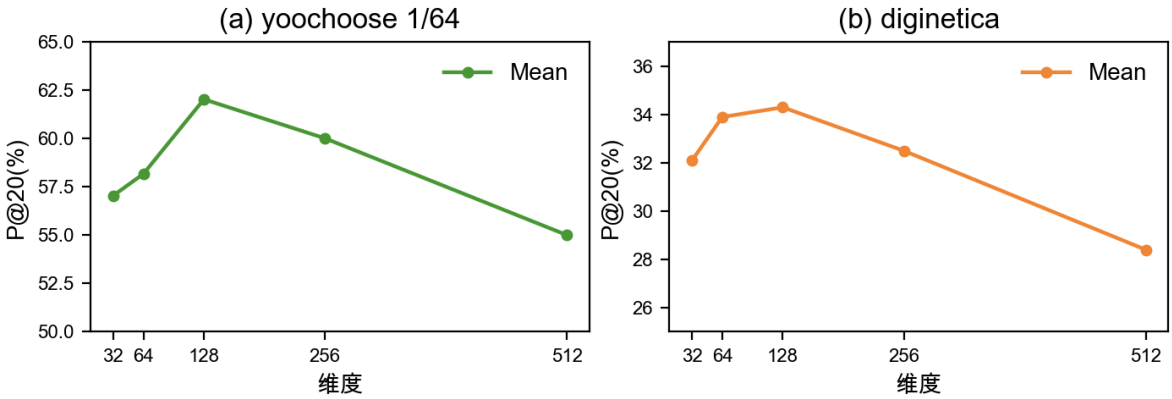

Figure 5. Impact of dimensionality on model performance
图 5. 维度对模型性能的影响

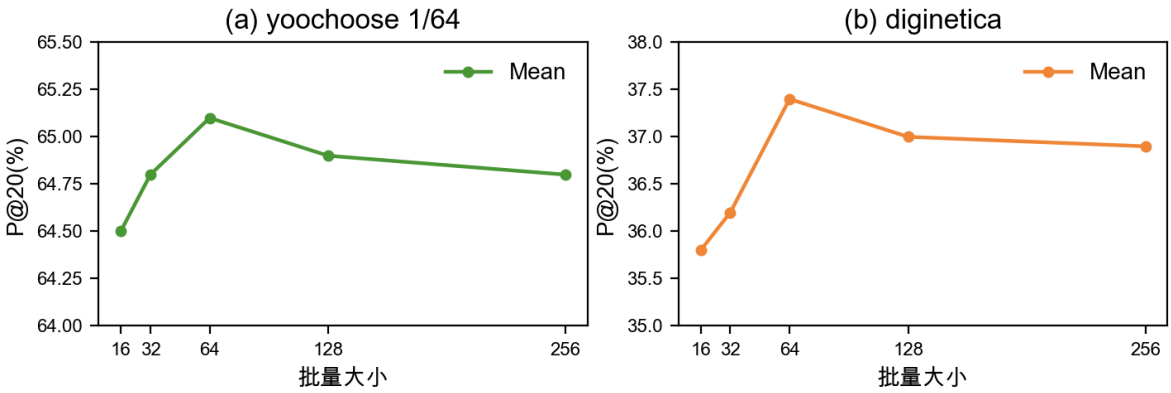

Figure 6. Impact of batch size on model performance
图 6. 批量大小对模型性能的影响

图 5 展示了模型 SetRec 在不同维度下的性能变化。就 Yoochoose 1/64 数据集而言，随着维度从 32 增加到 128，P@20 指标逐渐提升，但当维度继续增加至 256 和 512 时性能反而下降。Diginetica 数据集也表现出相同趋势，维度从 32 增至 128 时性能逐渐提高，在 128 达到峰值，而后维度进一步增加，性能持续下降。这表明模型对维度选择较为敏感，过低的维度限制了特征表达能力，过高的维度则增加了训

练复杂度和过拟合风险。综合两组结果发现 128 维是较为合适的折中选择，能保证模型具备较强的表示能力。

从图 6 可以观察到，批量大小对模型性能的影响呈现“先提升再趋缓”的变化规律。在 Yoochoose 1/64 数据集上，当 batch size 从 16 增加到 64 时，模型效果明显提升，但继续增大至 128 和 256 后，性能出现轻微下降，整体趋于稳定。在 Diginetica 数据集上，batch size 从 16 增长到 64 的过程中性能改善最为显著，而从 128 开始便出现下滑趋势，说明过大的 batch size 可能削弱了模型的泛化能力。综合结果表明，中等规模的 batch size (如 64)能在训练效率与推荐性能之间实现较好的平衡。

5. 结论

本研究提出的 SetRec 模型以集合表示学习为核心，将会话建模为无序集合，避免对序列顺序的过度依赖。通过引入随机遮掩机制提升输入鲁棒性，并结合集合编码与解码过程，模型能够有效捕捉交互项的全局依赖关系与潜在语义结构。同时，集合重构与推荐预测的联合优化框架不仅增强了潜在表示能力，还保证了推荐性能，为会话推荐提供了一种新颖且实用的建模思路。

在公开数据集上的实验结果表明，SetRec 在推荐准确率、召回率等指标上均优于主流方法，能够更好地捕捉用户短期兴趣与长期偏好。在无序条件下，模型依然保持稳定表现，并在数据稀疏和冷启动场景中展现出较强的适应性和泛化能力。这些实验充分验证了 SetRec 在不同设置下的有效性和稳健性，凸显了其在实际推荐系统中的应用潜力。

参考文献

[1] Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., *et al*. (2023) A Survey of Graph Neural Networks for Recommender Systems: Challenges, Methods, and Directions. *ACM Transactions on Recommender Systems*, **1**, 1-51. https://doi.org/10.1145/3568022

[2] 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

[3] Garcin, F., Dimitrakakis, C. and Faltings, B. (2013) Personalized News Recommendation with Context Trees. *Proceedings of the 7th ACM Conference on Recommender Systems*, Hong Kong, 12-16 October 2013, 105-112. https://doi.org/10.1145/2507157.2507166

[4] Hariri, N., Mobasher, B. and Burke, R. (2012) Context-Aware Music Recommendation Based on Latenttopic Sequential Patterns. *Proceedings of the Sixth ACM Conference on Recommender Systems*, Dublin Ireland, 9-13 September 2012, 131-138. https://doi.org/10.1145/2365952.2365979

[5] Vaswani, A., Shazeer, N., Parmar, N., *et al*. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **30**, 1-11.

[6] Wagstaff, E., Fuchs, F.B., Engelcke, M., *et al*. (2022) Universal Approximation of Functions on Sets. *The Journal of Machine Learning Research*, **23**, 6762-6817.

[7] Qi, C.R., Su, H., Mo, K., *et al*. (2017) Point-Net: 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.

[8] Zaheer, M., Kottur, S., Ravanbakhsh, S., *et al*. (2017) Deep Sets. *Advances in Neural Information Processing Systems*, **30**, 3391-3401.

[9] Skianis, K., Nikolentzos, G., Limnios, S., *et al*. (2020) Rep the Set: Neural Networks for Learning Set Representations. 2020 *International Conference on Artificial Intelligence and Statistics*, Wednesday, 26-28 August 2020, 1410-1420.

[10] Lee, J., Lee, Y., Kim, J., *et al*. (2019) Set Transformer: A Framework for Attention-Based Permutation-Invariant Neural Networks. 2019 *International Conference on Machine Learning*, Long Beach, 9-15 June 2019, 3744-3753.

[11] Hidasi, B., Karatzoglou, A., Baltrunas, L., *et al*. (2015) Session-Based Recommendations with Recurrent Neural Networks. arXiv:1511.06939.

[12] 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

[13] 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

[14] Romero, D.W. and Cordonnier, J.B. (2020) Group Equivariant Stand-Alone Self-Attention for Vision. arXiv:2010.00977.

[15] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S. (2020) End-to-End Object Detection with Transformers. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 213-229. https://doi.org/10.1007/978-3-030-58452-8_13