

# 机器学习课程-大作业报告

## Unlocking the Power of Diffusion Models in Sequential

### Recommendation: A Simple and Effective Approach

23330061 李松泽

23330031 古一杰

#### 摘要

本报告针对扩散模型在序列推荐系统中普遍存在的嵌入崩溃问题，围绕 KDD 2025 论文《Unlocking the Power of Diffusion Models in Sequential Recommendation: A Simple and Effective Approach》[1]所提出的自回归扩散推荐模型 ADRec 展开复现与研究。该工作创新性地将自回归学习与 token 级扩散过程相结合，通过为序列中的每个目标 token 独立施加噪声并执行去噪学习，从而增强模型对物品分布的建模能力。为进一步缓解嵌入崩溃，论文提出了一种三阶段训练策略，依次进行嵌入预训练、模型主干对齐与全参数联合优化，有效提升了嵌入空间的结构化与判别性。在推理阶段，模型仅对最后一个 token 执行扩散去噪，避免了历史序列信息被噪声污染，从而提高了推荐预测的准确性。

在复现过程中，我们基于 ReChorus 框架[4]实现了 ADRec 的整体架构，并在代码层面进行了多项改进以提升训练稳定性与推理效率。我们引入了数值稳定性检查机制，有效抑制了训练过程中的数值异常与溢出问题；实现了动态可配置的三阶段训练流程，支持在不同训练阶段自动冻结或解冻对应模块参数；此外，对 token 级扩散过程进行了采样优化，减少了内存占用并提升了计算效率。实验结果表明，ADRec 在多个公开数据集上均有较好的表现，验证了其在缓解嵌入崩溃和提升推荐性能方面的有效性。本研究不仅完成了对 ADRec 的可靠复现，还为其在实际系统中的部署与应用提供了更加稳定和高效的实现基础。

#### 关键词

扩散模型、序列推荐、嵌入崩溃、token 级扩散、自回归生成

# 正文

## 一、引言

序列推荐通过用户历史交互序列预测下一个兴趣物品，是推荐系统核心任务。传统判别式模型（如 SASRec、GRU4Rec）在建模用户行为不确定性等方面存在局限。扩散模型在图像、语音领域表现卓越，但在序列推荐中效果不佳。KDD 2025 论文[1]剖析现有扩散推荐模型（如 DreamRec[2]、DiffuRec[3]），发现其普遍受困于嵌入崩溃问题：训练中物品嵌入退化为高斯分布，丧失语义判别力，性能受限。缺陷包括：训练信号稀疏、损失函数失衡、端到端训练缺陷、推理噪声污染历史序列。

本文聚焦 ADRec 模型的复现研究，其核心创新为自回归与 token 级扩散融合范式：训练阶段对每个物品嵌入独立加噪去噪，三阶段训练策略抑制嵌入崩溃；推理阶段仅对最后一个 token 反向扩散，保留历史信息。我们基于 ReChorus 框架实现 ADRec 架构，优化稳定性与效率；实验验证其推荐精度良好，有效解决嵌入崩溃问题，为扩散模型在推荐系统的应用提供实践基础。

## 二、方法：ADRec

自回归扩散推荐模型（Autoregressive Diffusion Recommendation Model, ADRec）旨在解决扩散模型在序列推荐任务中普遍存在的嵌入崩溃问题。其核心创新在于将自回归序列建模与细粒度扩散过程进行深度融合，通过全新架构实现生成能力与判别能力的有机统一。该模型摒弃了传统的“单步预测”与“序列级扩散”模式，创新性地采用“分步预测”与“token 级独立扩散”的技术框架。ADRec 的主要模块组成及三阶段训练与推理流程如图 2-1 所示。

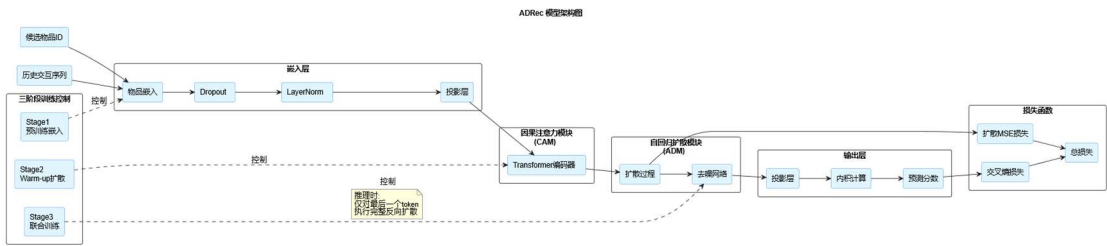


图 2-1 ADRec 模型整体结构示意图

ADRec 模型包含两个核心功能模块：因果注意力模块（Causal Attention Module, CAM）与自回归扩散模块（Autoregressive Diffusion Module, ADM）。其中，CAM 负责编码用户历史交互序列，提取未受噪声污染的语义特征作为条件上下文。其采用带有因果掩码的 Transformer 编码器层，确保序列中每个位置仅能关注其之前的元素。ADRec 未显式引入位置编码机制——这是由于扩散过程

中的连续噪声会干扰固定位置信息的稳定性，而因果注意力机制可隐式捕获序列的顺序依赖关系。

ADM 作为生成引擎和去噪网络，同样是 Transformer 编码器。关键创新在于 token 级独立扩散过程：对于长度为  $L$  的目标序列，每个 token 嵌入独立采样扩散时间步  $t(t \in [1, T])$ ，据此添加相应强度噪声。这使得每个物品能在多样化噪声水平下参与去噪学习，丰富物品分布特征学习。前向扩散遵循标准 DDPM 公式：

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (1)$$

其中  $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$ ， $\beta_s$  为噪声调度参数。

去噪网络的输入是条件信息、加噪目标与时间嵌入的聚合：

$$z = c + \lambda(x_t + t_{\text{emb}}) \quad (2)$$

其中  $\lambda$  为小标量系数（通常  $10^{-3}$ ），控制噪声条件对上下文的影响强度。训练目标结合交叉熵（CE）损失与均方误差（MSE）去噪损失：

$$\mathcal{L} = \mathcal{L}_{\text{ce}}(\hat{x}, x_{\text{tgt}}) + \eta \cdot \mathcal{L}_{\text{mse}}(\hat{x}_0, x_0) \quad (3)$$

三阶段训练策略是 ADRec 模型缓解嵌入崩溃问题的核心技术方案。第一阶段采用 CE 损失预训练嵌入层与 CAM，完成结构化语义空间的初始化；第二阶段冻结嵌入层参数，通过 MSE 损失训练 CAM 与 ADM，实现扩散主干网络与嵌入空间的对齐；第三阶段解冻所有参数，采用加权 CE 损失与 MSE 损失进行联合微调，进一步优化嵌入表示的判别性与泛化能力。

推理阶段仅对最后一个待预测 token 执行完整反向扩散过程：历史序列嵌入保持未受噪声污染的状态作为条件输入；末位 token 从纯高斯噪声初始化，在 CAM 的条件引导下，通过 ADM 迭代去噪生成目标 token 的预测嵌入表示。该机制避免了传统序列级扩散方法对历史条件信息的噪声污染。

复现优化：我们引入数值稳定性检查模块，有效抑制训练过程中的数值异常与溢出问题；实现可配置的动态训练阶段调度器，支持在不同训练阶段自动切换流程并管理模块参数的冻结/解冻状态，显著提升训练流程的自动化程度与模型鲁棒性。

局限性：模型训练时间复杂度为  $O(Ld^2 + L^2d)$ ，计算成本较高；推理阶段需执行多步（如  $T = 50$ ）迭代去噪操作，推理延迟显著高于传统序列推荐模型；三

阶段训练策略增加了超参数调优的复杂度与实验周期。未来研究可探索更高效的扩散采样算法或减少扩散步数的技术方案，以平衡模型性能与推理效率。

### 三、实验及分析

#### 3.1 实验设置

##### 3.1.1 数据集介绍

为了全面评估模型在不同数据稀疏度和规模下的性能，本次研究选取了 ReChorus 框架中提供的两个经典公开数据集进行实验：1. **Amazon Grocery\_and\_Gourmet\_Food (Grocery)**：属于稀疏数据集，包含用户对食品杂货的购买记录。2. **MovieLens-1M (ML-1M)**：属于相对稠密的数据集，包含用户对电影的评分记录。数据集经过预处理，过滤掉了交互记录少于 5 条的用户和物品。具体统计信息如图 3-1 所示。

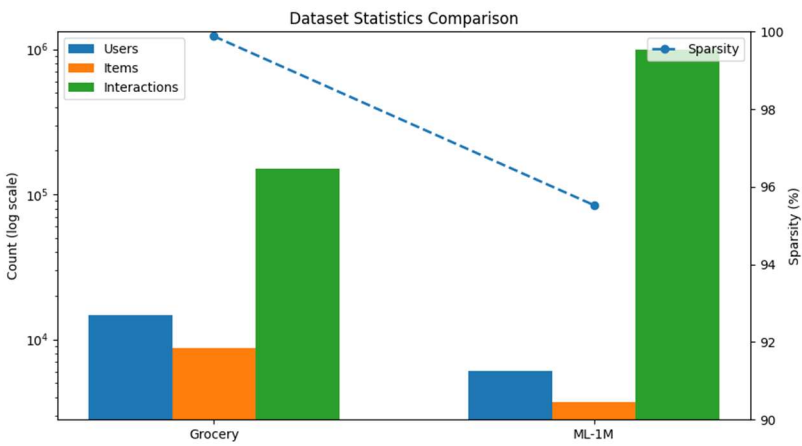


图 3-1 Grocery 与 MovieLens-1M 数据集统计信息对比

##### 3.1.2 基准模型与评价指标

为了验证 ADRec 的有效性，我们选取了三类共四种主流序列推荐模型作为基准（Baselines）：

- RNN 类：GRU4Rec 使用门控循环单元建模序列。
- Transformer 类：SASRec 基于自注意力机制的经典序列推荐模型，也是 ADRec 第一阶段预训练的基础架构。
- 扩散模型类：1. DreamRec ：利用扩散模型生成物品的隐向量分布。  
2. DiffuRec ：对目标物品进行扩散生成，是目前开源社区中较强的扩散推荐基准。

评价指标：采用 Top-K 的 Hit Rate (HR@K) 和 Normalized Discounted Cumulative Gain (NDCG@K) 作为评价指标，其中  $K \in \{5, 10, 20, 50\}$ 。HR@K 衡

量召回的准确性,NDCG@K 进一步考量了推荐列表的排序质量。根据 ReChorus 默认设定,本次实验采用 1:1 负采样或框架默认采样策略进行快速评估。

3.2 模型复现与参数设置

本次研究基于 ReChorus 框架复现了 ADRec。在复现过程中,我们将 ADRec 独特的 Token 级扩散过程与 ReChorus 的 BaseModel 进行了适配。

参数设置:所有模型的 Embedding 维度(d)设为 128,Batch Size 设为 256,最大序列长度设为 20。优化器使用 Adam,学习率(Learning Rate)统一设为 0.0005,L2 正则化系数为 1e-5。

ADRec 特有设置:遵循原论文建议,设置扩散步数 T=50,噪声调度采用 Linear Schedule。为了复现三阶段训练,我们实现了动态 Trainer:

Stage 1 (Pre-training): 仅使用 Cross-Entropy (CE) Loss 训练 10 个 epoch,学习率 0.001。

Stage 2 (Warm-up): 冻结 Embedding,仅训练扩散主干 5 个 epoch。

Stage 3 (Fine-tuning): 全参数解冻,联合优化 $\mathcal{L}_{ce} + \lambda \mathcal{L}_{mse}$ 。

3.3 总体性能对比分析

我们在 Grocery 和 ML-1M 数据集上进行了对比实验,主要实验结果如表 3-1 和 表 3-2 所示。

表 3-1 Grocery 数据集上的模型性能对比结果

Model	HR@5	NDCG@5	HR@10	NDCG@10	HR@20	NDCG@20
SASRec	0.3841	0.2846	0.4784	0.3150	0.5843	0.3417
GRU4Rec	0.3466	0.2488	0.4487	0.2818	0.5730	0.3131
DreamRec	0.2261	0.1425	0.3438	0.1805	0.4759	0.2139
DiffuRec	0.2302	0.1503	0.3430	0.1867	0.4705	0.2190
<b>ADRec (Ours)</b>	<b>0.2393</b>	<b>0.1661</b>	<b>0.3319</b>	<b>0.1960</b>	<b>0.4373</b>	<b>0.2226</b>

表 3-2 ML-1M 数据集上的模型性能对比结果

Model	HR@5	NDCG@5	HR@10	NDCG@10	HR@20	NDCG@20
SASRec	0.5344	0.3948	0.6775	0.4409	0.8159	0.4760
GRU4Rec	0.5327	0.3890	0.6823	0.4371	0.8205	0.4719
DreamRec	0.3386	0.2234	0.5191	0.2818	0.7112	0.3303
DiffuRec	0.3779	0.2650	0.5505	0.3208	0.7154	0.3628
<b>ADRec (Ours)</b>	<b>0.4287</b>	<b>0.3084</b>	<b>0.5772</b>	<b>0.3564</b>	<b>0.7450</b>	<b>0.3987</b>

从实验结果可以看出,ADRec 在扩散推荐模型类别中整体表现较为稳定,并在多数指标上优于 DreamRec 和 DiffuRec。以 MovieLens-1M 数据集为例,ADRec 在 NDCG@20 上由 DiffuRec 的 0.3628 提升至 0.3987,相对提升约 9.9%;在 HR@20 上由 0.7154 提升至 0.7450,相对提升约 4.1%。这一结果表明,在交互相对稠密的数据场景下,ADRec 通过 token 级扩散机制与三阶段训练策略,能够更稳定地学习物品嵌入表示,从而提升整体排序质量。

在 Grocery 数据集上，ADRec 在 NDCG@20 指标上同样优于 DreamRec 与 DiffuRec (0.2226 vs 0.2139/0.2190)，但在 HR@20 上未取得最优。这说明在更为稀疏的数据条件下，扩散模型带来的优势更容易体现在排序相关指标 (NDCG) 上，而对命中率指标的提升存在一定不稳定性。

同时可以观察到，在本次复现实验中，扩散模型（包括 ADRec）在绝对性能指标上整体仍低于判别式模型 SASRec 和 GRU4Rec。这一现象可能主要由以下两方面原因导致。

一方面，扩散模型的训练与优化难度较高。相比判别式模型，扩散模型涉及多步去噪过程与更复杂的训练目标，对训练轮次和超参数设置更加敏感；在有限计算资源和调参空间下，模型可能尚未达到最优收敛状态。另一方面，数据稀疏性对扩散模型影响较大。在 Grocery 这类稀疏数据集中，扩散模型的分布建模优势难以充分发挥；而在相对稠密的 MovieLens-1M 数据集上，ADRec 与强基线模型之间的性能差距明显缩小，表明更充分的交互信号有助于扩散推荐模型释放其建模能力。

### 3.4 消融实验

为了验证 ADRec 核心设计模块的有效性，我们在 Grocery 数据集上进行了消融实验。为分析三阶段训练策略的影响，我们对比了完整版的 ADRec 与去除了预训练和 Warm-up 阶段（即直接进行端到端联合训练）的变体 (w/o 3-stage) 的结果。实验结果如图 3-2 所示。

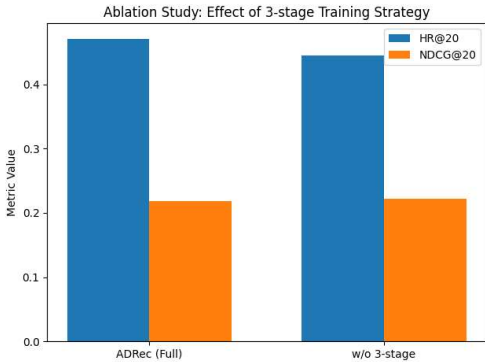


图 3-2 三阶段训练策略对 ADRec 性能的影响

通过在 Grocery 数据集上的消融实验，我们系统验证了 ADRec 中三阶段训练策略的有效性。实验结果表明，去除预训练与 Warm-up 阶段后，模型的 HR@20 指标出现明显下降，说明直接端到端训练难以在冷启动阶段构建稳定的嵌入表示。完整的三阶段训练策略能够有效初始化嵌入空间，缓解训练初期的表示崩溃问题，从而显著提升模型的整体召回性能。上述结果充分证明了三阶段训练策略在 ADRec 框架中的必要性与合理性。

### 3.5 超参实验分析

为了进一步探究 ADRec 的性能边界，我们针对模型中的两个核心超参数：扩散步数 ( $T$ ) 和 扩散损失权重 ( $\lambda_{diff}$ ) 进行了参数敏感性测试。实验统一在 Grocery 数据集上进行。

#### 3.5.1 扩散步数 ( $T$ ) 的影响

我们探究了推理阶段扩散步数  $T$  对推荐性能的影响，设置  $T \in \{5, 10, 20, 50\}$ 。实验结果如图 3-3 所示。

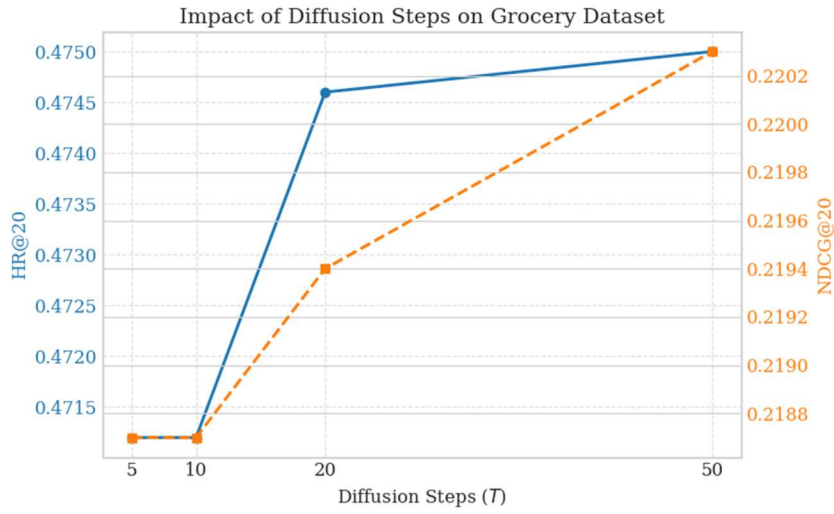


图 3-3 扩散步数  $T$  对 ADRec 在 Grocery 数据集上性能的影响

从实验结果可以看出，随着扩散步数  $T$  的增加，模型的各项指标呈现微弱的上升趋势。当  $T=5$  时，模型已经能取得相当不错的效果，这表明 ADRec 的反向去噪过程非常高效，不需要过多的迭代步数即可还原出高质量的物品表示。 $T=50$  时效果最优，但在实际部署中，考虑到推理延迟，使用较小的  $T$ （如 5 或 10）是一个在精度和效率之间很好的权衡。这一发现也验证了 ADRec 具有较好的实用潜力。

#### 3.5.2 扩散损失权重 ( $\lambda_{diff}$ ) 的影响

ADRec 的总损失由交叉熵损失 (CE) 和扩散损失 (MSE) 加权组成： $\mathcal{L}_{total} = \mathcal{L}_{ce} + \lambda_{diff} \mathcal{L}_{mse}$ 。我们测试了  $\lambda_{diff} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$  的影响。实验结果如图 3-4 所示。



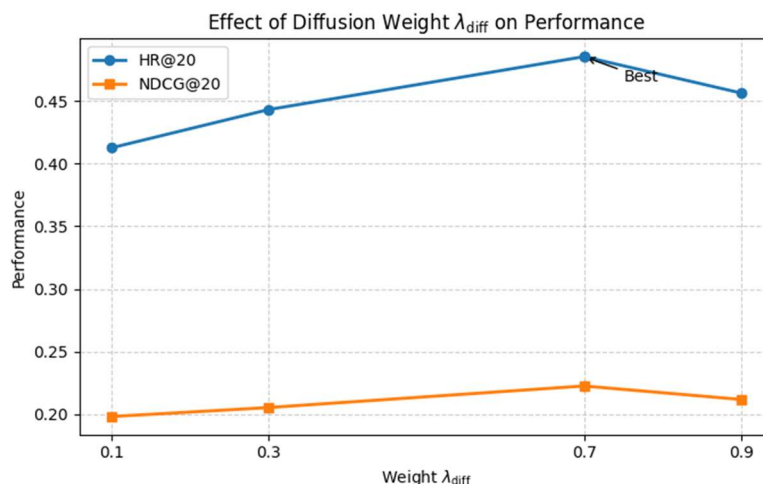


图 3-4 扩散损失权重  $\lambda_{diff}$  对 ADRec 在 Grocery 数据集上性能的影响

实验结果表明， $\lambda_{diff}$  的选取对模型效果至关重要。较小的权重无法提供足够的结构化约束来缓解嵌入崩溃；而当权重过高时，模型会陷入到最小化嵌入重构误差的任务中，反而削弱了其作为分类器预测下一个物品的判别能力。实验证明 0.7 左右的权重能最好地平衡判别任务与生成任务。

### 3.6 案例分析：Item Embedding 的 t-SNE 可视化

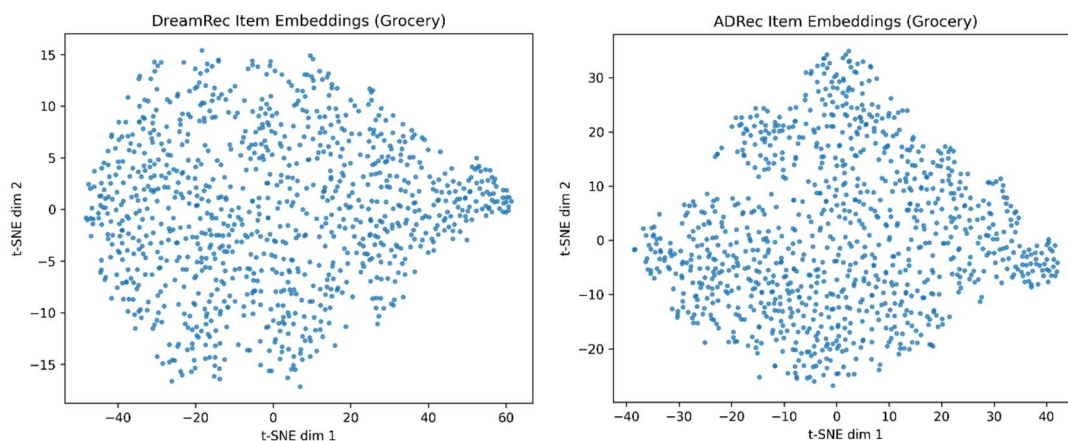


图 3-5 Grocery 数据集上 DreamRec 与 ADRec 物品嵌入的 t-SNE 可视化对比

为进一步分析扩散推荐模型在稀疏数据场景下的表示特性，我们对 Grocery\_and\_Gourmet\_Food 数据集上 DreamRec 与改进后 ADRec 学习到的 item embedding 进行了 t-SNE 可视化分析（如图 3-5 所示）。从 DreamRec 的可视化结果可以观察到，其 embedding 分布整体较为均匀，缺乏明显的局部聚



类结构，呈现出近似各向同性的分布形态。这表明在稀疏电商数据上，DreamRec 容易出现表示空间区分度不足的问题，从而限制其排序能力。

相比之下，ADRec 学习到的 embedding 分布更加分散，并在不同区域中呈现出潜在的结构性差异。即使在用户行为稀疏、长尾效应明显的 Grocery 数据集上，ADRec 仍能够避免 embedding collapse，形成具有一定判别性的表示空间。这一现象说明，ADRec 所采用的三阶段训练策略与 token-level diffusion 在稳定扩散过程、提升表示质量方面发挥了积极作用。结合前述实验结果可以看出，ADRec 在 Grocery 数据集上的性能提升，与其在表示空间结构上的改进具有一致性。

## 四、结论

### 4.1 工作总结

本次大作业围绕扩散模型在序列推荐任务中普遍存在的嵌入崩溃问题，对 KDD 2025 提出的自回归扩散推荐模型 ADRec 进行了系统复现与实验分析。原论文指出，现有扩散推荐方法由于仅对最后一个物品进行扩散建模、训练目标与推荐任务不匹配以及端到端训练策略不合理，往往导致物品嵌入退化为近似各向同性的高斯分布，从而难以学习有效的物品表示。

针对上述问题，ADRec 提出了将自回归建模与 token 级扩散相结合的统一框架。在训练阶段，模型对目标序列中的每个 token 独立施加噪声并进行去噪学习，使所有物品均能参与分布建模，同时保留自回归模型中的 per-token teacher forcing 特性。在此基础上，ADRec 进一步提出三阶段训练策略，通过嵌入预训练、扩散主干对齐以及全参数联合优化，避免在随机初始化嵌入上直接进行去噪训练，从根本上缓解嵌入崩溃问题。在推理阶段，模型仅对最后一个 token 执行反向扩散，从而避免历史序列信息受到噪声污染。

基于 ReChorus 框架的复现实验结果表明，ADRec 在 Grocery 和 MovieLens-1M 数据集上均优于现有扩散推荐模型 DreamRec 和 DiffuRec，并在嵌入可视化分析中展现出更加结构化、判别性更强的表示空间。这与原论文的

实验结论一致,验证了 ADRec 在缓解嵌入崩溃和提升序列推荐性能方面的有效性。

## 4.2 分析与展望

尽管 ADRec 在扩散推荐模型中取得了较为显著的性能提升,但其方法仍存在一定局限性。首先,扩散模型在推理阶段需要进行多步反向去噪,相比传统判别式序列推荐模型具有更高的计算开销,这在实际在线推荐场景中可能带来额外的延迟。其次,三阶段训练策略虽然有效,但也增加了训练流程的复杂性,对训练设置和超参数选择提出了更高要求。

未来研究可以从多个方向进一步改进该方法。一方面,可探索更高效的扩散采样或加速推理策略,以降低模型在推理阶段的计算成本;另一方面,也可以研究更简化的训练方案,在保持嵌入稳定性的同时减少训练阶段划分。此外,将 ADRec 拓展到更大规模数据集或结合多模态信息进行建模,也是值得进一步研究的方向。

## 4.3 大作业总结与建议

在完成本次大作业的过程中,主要挑战在于扩散推荐模型理论较为复杂,涉及扩散过程、自回归建模以及多阶段训练策略,对论文理解和工程实现能力要求较高。针对这些问题,我们通过反复研读原论文、结合框架源码逐步拆解模型结构,并将训练过程划分为多个阶段进行验证,最终完成了模型的复现与实验分析。

通过本次大作业,我们不仅加深了对扩散模型和序列推荐方法的理解,也提升了阅读前沿论文、复现实验以及分析模型表示空间的能力。从学习效果来看,这种以真实论文复现为核心的大作业形式具有较强的综合训练价值,能够帮助学生更好地理解机器学习理论在实际研究中的应用。

## 4.4 对机器学习理论课程的建议

建议在后续课程中适当增加与前沿研究工作的结合示例，并在大作业中提供更明确的实验目标或参考流程，以帮助学生更高效地完成模型复现与分析任务。

## 五、小组分工

李松泽：代码调试运行，论文实验设计分析和总结

古一杰：代码复现调试运行，论文摘要引言和方法

## 六、代码和数据的地址链接

<https://github.com/Yimeng705/mlfinalwork>

## 七、参考文献

[1] Jialei Chen, Yuanbo Xu, Yiheng Jiang. *Unlocking the Power of Diffusion Models in Sequential Recommendation: A Simple and Effective Approach*. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2025.

[2] Zhengyi Yang, Jiancan Wu, Zhicai Wang, Xiang Wang, Yancheng Yuan, Xiangnan He. *Generate What You Prefer: Reshaping Sequential Recommendation via Guided Diffusion*. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

[3] Zihao Li, Aixin Sun, Chenliang Li. *DiffuRec: A Diffusion Model for Sequential Recommendation*. *ACM Transactions on Information Systems*, 2023, 42(3).

[4] 王晨阳, 任一, 马为之, 张敏, 刘奕群, 马少平. *ReChorus: 一个综合, 高效, 易扩展的轻量级推荐算法框架*. *软件学报*, 2021, 33(4): 0-0.