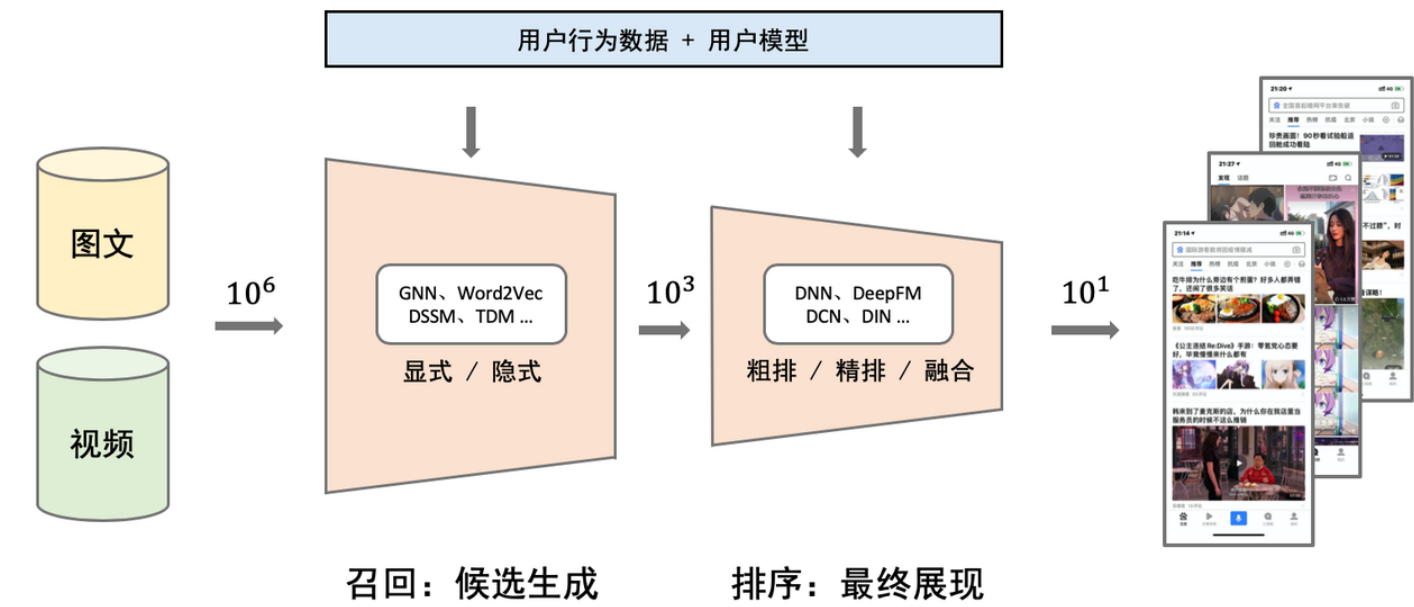


推荐系统

1. 题目背景导言:



推荐系统是一种基于用户行为和偏好，自动向用户推荐可能感兴趣的物品（如商品、电影、音乐等）的技术。它们在现代互联网应用中扮演着至关重要的角色，为用户提供个性化的体验，同时也帮助平台提升用户粘性和销售额。预计未来将朝着联邦学习、自适应推荐、可解释性推荐、社会化推荐、数据效率提升和多模态推荐等方向发展，以提升性能、增强用户体验和信任度、提高数据处理效率并丰富推荐内容。

- **召回:**

在推荐系统中，召回是指从庞大的物品库中筛选出一小部分可能符合用户兴趣的候选物品集合的过程，其目标是尽可能多地覆盖用户可能感兴趣的物品，以确保不会遗漏潜在的优质推荐选项。

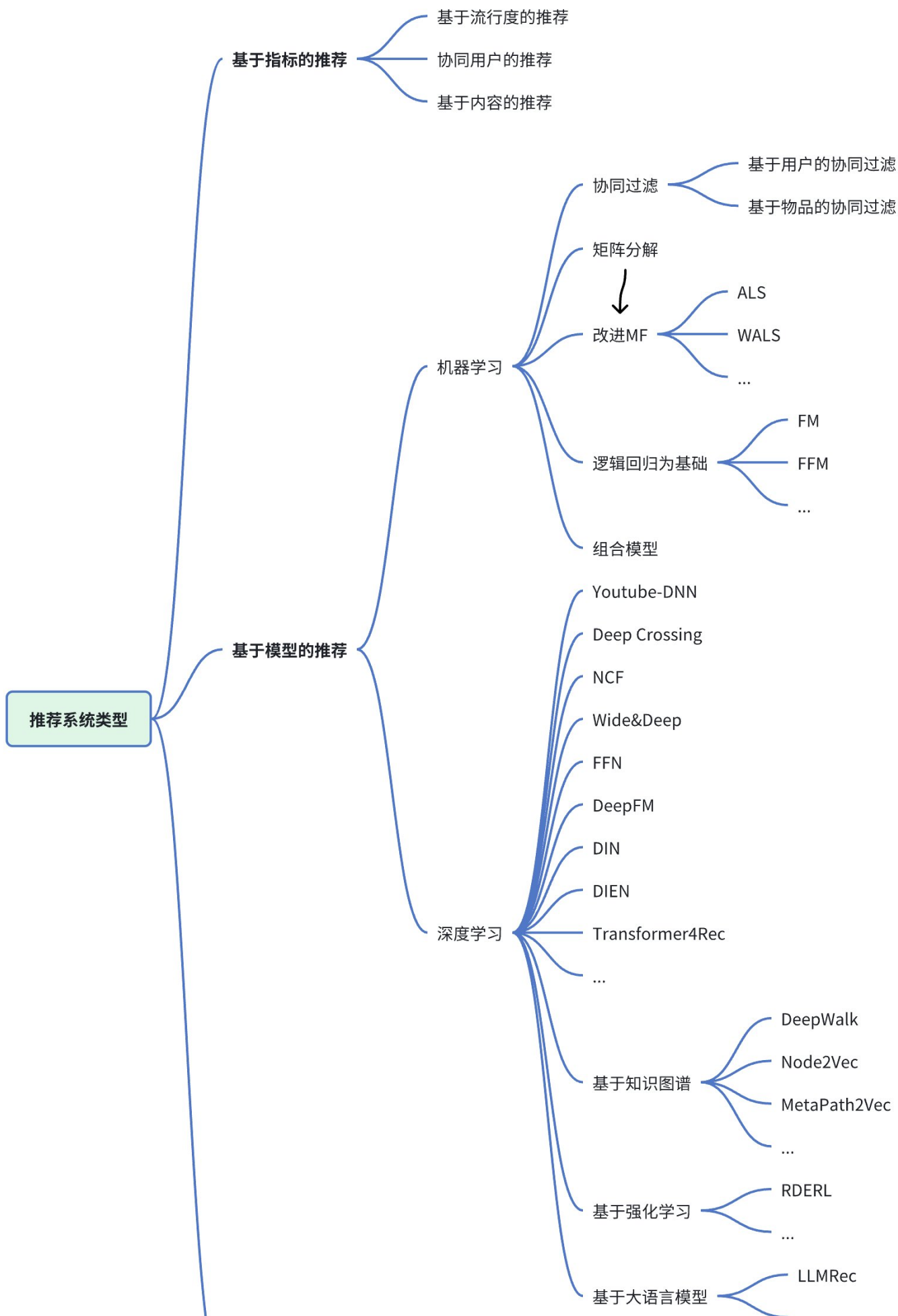
- **排序:**

排序则是在召回的基础上，对候选物品集合按照用户可能的喜爱程度进行优先级排序，通过评估用户对各个候选物品的偏好概率，将最有可能被用户喜欢的物品排在前面，从而为用户提供更精准、更具吸引力的推荐列表。

- **面对的问题:**

- 数据稀疏性：用户与物品之间的交互数据通常非常稀疏，绝大多数用户只与少数物品有过交互，导致推荐系统难以准确建模用户偏好。
- 冷启动问题：新用户或新物品缺乏足够的历史交互数据，使得推荐系统无法有效地为其提供推荐。

- 长尾问题:大部分冷门的商品得不到暴露,然而它们的加和价值超过热门物品。



2. 任务

2.1 数据集：

- 推荐系统通常需要大量的训练数据。为了方便起手，建议使用 [ml-latest.zip](#) 数据集（这是 MovieLens 数据集的最新版本，包含用户对电影的评分数据）。
- 如果有余力，大家也可以选择其他感兴趣的数据集进行训练和测试。

数据集名称	推荐系统场景	链接
MovieLens	电影推荐	MovieLens 数据集 ：包含多个不同规模的电影评分数据集，适用于推荐系统研究。
Netflix	电影推荐	Netflix Prize data ：Netflix提供的电影评分数据集，曾用于 Netflix Prize 竞赛。
MicroLens	视频流推荐	MicroLens 数据集 ：大规模短视频推荐数据集，包含原始文本、音频、图像和视频数据。
GoodReads	书籍推荐	Goodreads Book Graph Datasets ：包含书籍评分和评论数据，适用于书籍推荐系统。
Anime	动漫推荐	Anime Recommendations Database ：动漫推荐数据库，提供动漫评分和用户偏好数据。
Steam	游戏推荐	GitHub - kang205/SASRec ：基于自注意力机制的顺序推荐系统，并包含 Steam 上的评论和游戏信息,适用于游戏推荐。
Amazon	电商推荐	Recommender Systems and Personalization Datasets ：亚马逊提供的电商推荐数据集，包含用户购买行为和评分数据。
NineRec	多领域推荐	NineRec 数据集 ：多模态多领域推荐数据集，涵盖9个不同领域的的数据，适用于跨领域推荐系统研究。

2.2 目标：

- 构建一个基础的推荐系统，能够预测用户的兴趣并提供个性化推荐。

- 不要求完整实现一个完整的推荐系统，但需要实现推荐系统中**排序**的一部分并了解推荐系统的运行原理。


2.3 模型选择：

- 可以选择实现以下几种推荐系统的**排序**模型：
 - **Wide&Deep**：是一种结合了线性模型（Wide）的记忆能力和深度神经网络（Deep）的泛化能力的推荐系统模型，能够同时处理稀疏特征组合和复杂特征交互，提升推荐效果。
 - **DeepFM**：DeepFM 是一种结合了因子分解机（FM）和深度学习技术的推荐系统算法，利用深度学习的非线性表示能力与 FM 的线性表示能力，同时捕捉低阶和高阶特征交互。
 - **Deep_Crossing**：Deep Crossing 是一种深度学习架构，用于解决推荐系统中的特征交叉问题。它通过结合深度神经网络和交叉网络来自动学习特征之间的交叉关系，而无需人工特征工程。
 - **GBDT+LR**：GBDT（Gradient Boosting Decision Trees）和 LR（Logistic Regression）的结合是一种经典的集成学习方法。GBDT 通过决策树的级联训练自动发现特征交互关系，生成的叶子节点作为新特征输入到 LR 中，进一步提升分类性能。
 - **DIN**：DIN 是一种针对电商领域点击率预估的推荐模型，利用注意力机制捕捉用户历史行为与候选广告之间的相关性，动态调整用户行为特征的权重，从而更精准地预测用户的点击行为。
 - **DCN**：是一种结合了交叉网络和深度神经网络的推荐系统模型，通过显式学习特征交叉和捕捉复杂特征交互，自动挖掘高阶特征关系，从而提高点击率预测的准确性和效率。
- 也可以尝试其他推荐系统模型。

2.4 项目资源：

- 由于网络问题，相关的参考项目代码托管在 **Gitee** 上,大家可以使用 `git clone` 命令或者直接点击下面的链接在项目中下载对应的zip文件。



 <https://gitee.com/tanke11223344/Recommender-System-Evaluation>
tkzzzzzz6/Recommender-System-Evaluation

本项目是推荐系统的实践项目，包含了召回和排序两个主要模块的多种算法实现。

- 大家可以通过命令行克隆这个项目，作为参考。

```
1 git clone https://gitee.com/tanke11223344/Recommender-System-Evaluation.git
```

2.5 任务要求

2.5.1 实现模型：

- 从开源代码库中选择一个或多个推荐系统模型进行实现。
- 重点关注推荐系统中的**排序部分**，即如何根据用户特征和物品特征预测用户的兴趣。

2.5.2 改进模型（可选）：

- 如果有能力，可以尝试对模型进行改进，例如：
 - 调整模型结构。
 - 添加新的特征。
 - 优化训练过程。
- 改进的目的是提高模型的推荐效果（如准确率、召回率等）。

3. 模型测试标准(任选其一即可)

测试标准类型	召回模型测试标准	排序模型测试标准
基础指标	召回率（Recall）：衡量从海量候选物品中召回相关物品的能力，召回率越高，说明模型越全面。	精准度（Precision）：衡量推荐给用户的内容中，有多少是用户真正感兴趣的，精准度越高，说明推荐系统越能准确满足用户需求。
综合指标	F1分数（F1 Score）：精准度和召回率的调和平均数，用于平衡两者之间的关系，F1分数越高，说明召回模型的整体性能越好。	F1分数（F1 Score）：同召回模型，用于衡量排序模型在精准度和召回率之间的平衡。
排序效果指标	----- -----	AUC：衡量模型对正负样本的区分能力，AUC越高，说明模型的排序效果越好。
排名质量指标	----- -----	MAP（Mean Average Precision）：衡量模型在多个查询上的平均精准率，MAP越高，说明模型的排名质量越好。
用户体验指标	覆盖率（Coverage）：衡量推荐系统能够覆盖的物品范围，覆盖率越高，说明模型越能提供多样化的推荐。	MRR（Mean Reciprocal Rank）：衡量模型将正确答案排在前面的能力，MRR越高，说明模型越能满足用户快速找到感兴趣内容的需求。
其他指标	新颖性（Novelty）：衡量推荐结果的新颖程度，避免总是推荐热门但用户已知的物品。	NDCG（Normalized Discounted Cumulative Gain）：衡量模型在排名列表中的累积增益，考虑了排名位置对用户价值的影响，NDCG越高，说明模型的排名效果越好。

4. 提交要求:

4.1 代码

提交包含所有源代码的文件夹，代码需有清晰的注释。包括数据预处理、模型训练和测试的脚本。如果使用了特殊库或框架，需包含一个 `requirements.txt` 文件，列明所有依赖。确保代码具备可重复性，以便评审人员可以运行并验证结果。

4.2 文档

提交一个 `README.md` 文件，其中详细说明如何运行代码，以及任何必要的安装步骤。包括模型架构的描述、训练过程和选择这些方法的理由。描述数据预处理的步骤和参数调优的过程。

4.3 模型文件

提交训练好的模型文件，以便评审人员可以直接加载模型进行测试。如果模型文件过大，可以提供下载链接

（如Google Drive/阿里云盘/百度网盘）。

4.4 评估报告

提交一个包含模型性能评估的详细报告。报告应包括模型在各项评估标准（如精确度、召回率、F1分数等）上的表现。

4.5 示例 `README.md` 文件

```
1  # 项目名称
2
3  ## 项目描述
4  简要描述项目的背景、目标和主要功能。
5
6  ## 环境依赖
7  列出项目运行所需的环境和依赖库。
8
9  ### 1.克隆项目仓库:
10 git clone https://github.com/[yourusername]/[yourproject].git
11 cd yourproject
12
13 ### 2.安装依赖:
14 pip install -r requirements.txt
```

5. 相关资料

<div><div></div><div>A= 文本</div></div>	<div>分类</div>
<div><div>▼</div><div>项目</div><div>3 条记录</div></div>	
<div><div>1</div><div>GitHub - hwwang55/DKN: A tensorflow implementation of DKN (Deep Knowledge-aware...</div></div>	<div>项目</div>
<div><div>2</div><div>GitHub - recommenders-team/recommenders: Best Practices on Recommendation Syst...</div></div>	<div>项目</div>
<div><div>3</div><div>PaddlePaddle/PaddleRec: Recommendation Algorithm大规模推荐算法库，包含推荐系统...</div></div>	<div>项目</div>
<div><div>▼</div><div>介绍</div><div>6 条记录</div></div>	
<div><div>1</div><div>[推荐系统vol.1] 聊一聊推荐系统是如何一步一步发展到今天的_哔哩哔哩_bilibili</div></div>	<div>介绍</div>
<div><div>2</div><div>【总结】推荐系统——精排篇 【3】 DIN/DIEN/BST/DSIN/MIMN/SIM/CAN 绝密伏击 绝密伏击...</div></div>	<div>介绍</div>
<div><div>3</div><div>回顾Google经典CTR预估模型WDL</div></div>	<div>介绍</div>
<div><div>4</div><div>推荐系统——精排篇 【2】 WDL/DCN/DCN-v2</div></div>	<div>介绍</div>
<div><div>5</div><div>一文了解推荐系统全貌</div></div>	<div>介绍</div>
<div><div>6</div><div>推荐系统[一]: 超详细知识介绍，一份完整的入门指南，解答推荐系统相关算法流程、衡量指...</div></div>	<div>介绍</div>
<div><div>▼</div><div>教程</div><div>3 条记录</div></div>	
<div><div>1</div><div>十分钟！全流程！从零搭建推荐系统_副本 - 以电影推荐为例，教大家怎么使用PaddleRec一...</div></div>	<div>教程</div>
<div><div>2</div><div>推荐算法在线实战，看完这些，你就能搞懂推荐系统！ - 飞桨AI Studio星河社区</div></div>	<div>教程</div>
<div><div>3</div><div>FunRec</div></div>	<div>教程</div>
<div><div>▼</div><div>数据集</div><div>6 条记录</div></div>	
<div><div>1</div><div>CTR预估数据集汇总</div></div>	<div>数据!</div>
<div><div>2</div><div>MovieLens GroupLens</div></div>	<div>数据!</div>
<div><div>3</div><div>天池数据集_阿里系唯一对外开放数据分享平台-阿里云天池</div></div>	<div>数据!</div>
<div><div>4</div><div>推荐系统数据集汇总_数据集-阿里云天池</div></div>	<div>数据!</div>
<div><div>5</div><div>RUCAIBox/RecSysDatasets: 这是一个推荐系统（RS）公共数据源的仓库。 --- RUCAIBox/R...</div></div>	<div>数据!</div>
<div><div>6</div><div>AaronHeee/LLMs-as-Zero-Shot-Conversational-RecSys: 评估数据，LLMs CIKM 2023 “大...</div></div>	<div>数据!</div>
<div><div>▼</div><div>论文</div><div>10 条记录</div></div>	
<div><div>1</div><div>WLiK/LLM4Rec-Awesome-Papers: A list of awesome papers and resources of recommen...</div></div>	<div>论文</div>
<div><div>2</div><div>Neural Collaborative Filtering</div></div>	<div>论文</div>
<div><div>3</div><div>Distributed Representations of Words and Phrases and their Compositionality</div></div>	<div>论文</div>
<div><div>4</div><div>Deep Neural Networks for YouTube Recommendations</div></div>	<div>论文</div>
<div><div>5</div><div>Field-aware Factorization Machines for CTR Prediction</div></div>	<div>论文</div>

5	Field-aware Factorization Machines for CTR Prediction	论文
6	Deep Crossing: Web-Scale Modeling without Manually Crafted Combinatorial Features	论文
7	Deep & Cross Network for Ad Click Predictions	论文
8	Deep Interest Network for Click-Through Rate Prediction	论文
9	Deep Interest Evolution Network for Click-Through Rate Prediction	论文
10	CHIANGEL/Awesome-LLM-for-RecSys: Survey: A collection of AWESOME papers and reso...	论文
▼	可能遇到的问题	2 条记录
1	Kaggle手机验证解决办法_your account cannot be phone verified. please file-CSDN博客	可能
2	超详细Windows10/Windows11 子系统（WSL2）安装Ubuntu20.04（带桌面环境）_wsl安...	可能

30 条记录