

Intelligent Recommendation Technology

智能推荐技术

潘微科

深圳大学计算机与软件学院

第1章 概述

- 推荐技术简介
- 推荐问题分类
- 常用数据集
- 常用验证方法
- 深度学习平台简介
- 本章小结
- 习题

推荐技术简介

- 定义：在大数据和人工智能时代，智能推荐系统和技术已经成为电商、资讯、娱乐、教育、旅游和招聘等众多在线服务平台的核心技术、标准配置和重要引擎，用于帮助用户从海量物品中快速地找到他们感兴趣的物品（例如商品、新闻、视频、课程、景点和岗位等），特别是用户不容易发现的长尾物品（tail items），有效地缓解了信息过载（information overload）问题。

推荐技术简介

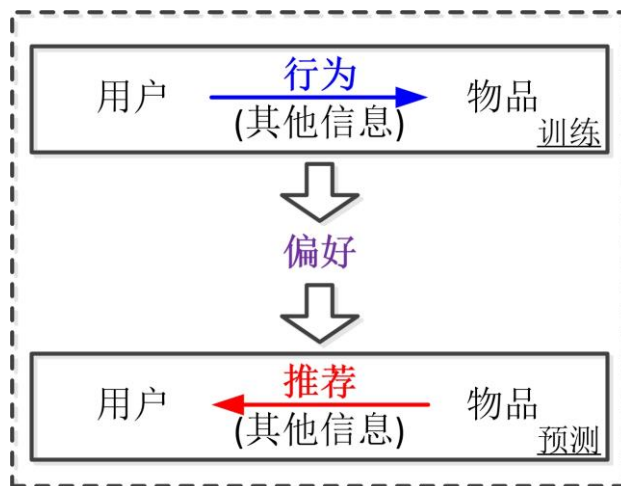
- 典型例子：抖音海外版TikTok推荐算法凭借其精准的视频推荐和有效的新领域拓展等功能，被《麻省理工科技评论》（MIT Technology Review）评为2021年全球十大突破性技术之一。

推荐技术简介

- **推荐系统中的用户与物品**：需要说明的是，物品可以是一组物品（例如旅游套餐推荐等），而用户也可以是一群用户（例如户外集体活动推荐等）。同时，有些场景中用户也可以作为“物品”，例如社交平台上的好友推荐等，即给用户推荐有相似兴趣的用户。

推荐技术简介

- 智能推荐技术示意图：



- 在推荐系统中，通过对用户**行为**和其他信息的建模，相关算法和技术能较为有效地挖掘和学习用户的真实**偏好**，进而为用户提供精准的个性化**推荐**服务，这一过程可以简单地概括为“**行为**→**偏好**→**推荐**”。

推荐技术简介

- 对比：
- 推荐系统 vs. 信息检索
- “行为→偏好→推荐” vs. “查询串→相关性→匹配”
- 两方面的重要区别：
- （1）行为 vs. 查询串，前者关注用户与系统的互动，后者侧重用户主动查询的内容。
- （2）偏好 vs. 相关性，前者关注用户的个性化需求，后者侧重查询串与候选文档（或网页）在内容方面的匹配程度。

推荐技术简介

- 用户行为建模的重要性：在电商、资讯等众多应用场景中，相较于**物品的描述**、**用户的画像**和**行为的上下文**等其他信息，用户行为往往是**与用户的偏好最直接相关的**，例如，用户购买商品和阅读新闻等的行为较能体现用户的喜好和兴趣。

推荐技术简介

- 部分相关的学术期刊论文（初步统计，截至**2021年7月13日**）
- **ACM Transactions on Interactive Intelligent Systems (ACM **TiIS**)**
- 2016(6), 2017(3), 2018(0), 2019(3), 2020(4), 2021(2)
- **ACM Transactions on Intelligent Systems and Technology (ACM **TIST**)**
- 2016(10), 2017(6), 2018(3), 2019(5), 2020(9), 2021(3)
- **ACM Transactions on Knowledge Discovery from Data (ACM **TKDD**)**
- 2016(4), 2017(3), 2018(6), 2019(3), 2020(5), 2021(4)
- **ACM Transactions on Management Information Systems (ACM **TMIS**)**
- 2016(1), 2017(0), 2018(2), 2019(1), 2020(2), 2021(0)
- **ACM Transactions on Information Systems (ACM **TOIS**)**
- 2016(4), 2017(8), 2018(5), 2019(20), 2020(14), 2021(5)
- **ACM Transactions on Internet Technology (ACM **TOIT**)**
- 2016(2), 2017(1), 2018(2), 2019(1), 2020(3), 2021(1)
- **ACM Transactions on the Web (ACM **TWEB**)**
- 2016(2), 2017(1), 2018(2), 2019(1), 2020(3), 2021(1)

推荐技术简介

- 部分相关的学术期刊论文（初步统计，截至**2021年7月13日**）
- **IEEE Transactions on Big Data (IEEE TBD)**
- 2016(3), 2017(2), 2018(1), 2019(1), 2020(2), 2021(1)
- **IEEE Transactions on Cybernetics (IEEE TCYB)**
- 2016(0), 2017(2), 2018(0), 2019(5), 2020(3), 2021(0)
- **IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)**
- 2016(12), 2017(8), 2018(10), 2019(9), 2020(19), 2021(22)
- **IEEE Transactions on Multimedia (IEEE TMM)**
- 2016(4), 2017(7), 2018(5), 2019(4), 2020(4), 2021(3)
- **IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI)**
- 2016(0), 2017(2), 2018(0), 2019(0), 2020(1), 2021(0)

推荐技术简介

- 部分相关的学术会议论文（初步统计，截至**2021年7月13日**）
- **ACM International Conference on Web Search and Data Mining (WSDM)**
- 2018(13), 2019(23), 2020(23), 2021(38)
- **AAAI Conference on Artificial Intelligence (AAAI)**
- 2018(15), 2019(33), 2020(27), 2021(33)
- **The Web Conference (WWW)**
- 2018(15), 2019(39), 2020(31), 2021(56)
- **International Joint Conference on Artificial Intelligence (IJCAI)**
- 2018(24), 2019(49), 2020(19), 2021(18)
- **ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)**
- 2018(22), 2019(41), 2020(30), 2021(39)
- **ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)**
- 2018(30), 2019(31), 2020(88), 2021(67)
- **ACM Conference on Recommender Systems (RecSys)**
- 2018(78), 2019(76), 2020(57), 2021(54+)

推荐技术简介

- 美国国防部（Department of Defense, DoD）也在2014年将一个类似的问题——对人类行为的计算机建模——列为六大颠覆性基础研究领域之一。

推荐技术简介

- 用户行为类型：
 - (1) 显式反馈，例如用户对物品的1~5分的数值评分（numerical rating）行为和喜欢/不喜欢的二值评分（binary rating）行为，以及点赞（like）等单值行为。
 - (2) 隐式反馈（又称为单类反馈），例如用户对物品的点击（click）、浏览（browse/view/examination）、收藏（collect/favorite）、加入购物车（add-to-cart）和购买（purchase）等行为。
 - (3) 异构反馈，例如同时包含两种或两种以上的显式反馈和/或隐式反馈。
 - (4) 序列反馈，例如在隐式反馈中包含时序信息。

推荐技术简介

- 其他行为信息：
- (1) 多方面行为 (multi-aspect behavior) ，如用户对物品的价格、质量和售后服务等不同方面的反馈。
- (2) 多粒度行为 (multi-granularity behavior) ，如用户对商品、品牌、商家和品类等不同粒度的反馈。
- (3) 多模态行为 (multi-modal behavior) ，如文本评论行为和数值反馈行为等。
- (4) 多领域行为 (multi-domain behavior) ，如用户在电商、资讯和社交等不同领域的行为。

推荐技术简介

- 说明：面向显式反馈、隐式反馈、异构反馈和序列反馈等不同类型的行为进行建模的技术和算法是较为基础和关键的，也是本课程的重点。

第1章 概述

- 推荐技术简介
- 推荐问题分类
- 常用数据集
- 常用验证方法
- 深度学习平台简介
- 本章小结
- 习题

推荐问题分类

- 根据推荐算法的输入和输出，也就是学习用户偏好时可以使用用户行为或用户反馈等数据的类型以及建模的目标，我们可以得到不同的推荐问题，包括单行为推荐、多行为推荐、序列推荐、联邦推荐和多源数据推荐等。

推荐问题分类

- **单行为推荐** (single-behavior recommendation) 是指算法的输入只有一种类型的行为或反馈数据，例如，显式反馈或隐式反馈。接着，根据评分预测 (rating prediction) 和物品排序 (item ranking) 等不同的建模目标，可以得到以下三类问题：
 - (1) 基于显式反馈的评分预测，又称为协同过滤 (collaborative filtering, CF) ；
 - (2) 基于显式反馈的物品排序，又称为协同排序 (collaborative ranking, CR) ；
 - (3) 基于隐式反馈的物品排序，又称为单类协同过滤 (one-class collaborative filtering, OCCF) 。

推荐问题分类

- **多行为推荐** (multi-behavior recommendation) 是指算法的输入至少包含两种不同类型的行为或反馈数据，例如同时包含数值评分和二值评分的异构显式反馈、同时包含购买和浏览的异构单类反馈等。类似地，根据评分预测和物品排序等不同的建模目标，一般有以下两类问题：
 - (1) **基于异构反馈的评分预测**，通常是指对异构显式反馈数据进行建模，因此又称为异构协同过滤 (heterogeneous collaborative filtering, HCF) ；
 - (2) **基于异构反馈的物品排序**，通常是指对异构隐式（或单类）反馈数据进行建模，因此又称为异构单类协同过滤 (heterogeneous OCCF, HOCCF) 。

推荐问题分类

- **序列推荐** (sequential recommendation) 是指算法的输入包括**用户的反馈数据** (例如购买、浏览等) 和**行为的先后时序信息**。一般有以下两类问题：
 - (1) **单行为序列推荐**，通常是指序列中只有一种类型的单类反馈 (例如购买行为)，因此又称为序列单类协同过滤 (sequential OCCF, SOCCF) ；
 - (2) **多行为序列推荐**，通常是指序列中有两种或两种以上不同类型的单类反馈 (例如浏览、加入购物车和购买等行为)，因此又称为序列异构单类协同过滤 (sequential HOCCF, SHOCCF) 。

推荐问题分类

- **联邦推荐**（federated recommendation）是指算法在对用户的行为数据进行建模时，还需要保护**用户隐私**（user privacy）和**数据安全**（data security）。根据不同的应用场景，以及不同程度的用户隐私和数据安全的保护要求，可以分为以下三个类别：
 - （1）跨用户联邦推荐（cross-user federated recommendation, CUFR） ,
 - （2）跨组织联邦推荐（cross-organization federated recommendation, COFR） ,
 - （3）双跨联邦推荐（dual-cross federated recommendation, DCFR） 。

推荐问题分类

- **跨用户联邦推荐**通常是指算法的输入是**同一组织内**不同用户的数据，例如用户的浏览和购买行为，其中重点关注**个体用户（或用户群组）间的隐私**，即每个用户（或用户群组）的数据不离开本地（即个体用户或用户群组）且隐私信息不被泄露。
- 跨用户联邦推荐与联邦学习中的**横向联邦学习**有一定的相似性，前者的**用户**对应后者的**样本**，前者的**物品**对应后者的**特征**。

推荐问题分类

- **跨组织联邦推荐**通常是指算法的输入包含来自**两个或两个以上不同组织**的用户的数据，例如来自电商平台的消费行为和来自资讯平台的阅读行为，其中重点关注**组织间的数据安全**，即每个组织的数据不离开本地（即组织），而不考虑同一组织内的个体用户（或用户群组）间的隐私。
- 当不同组织间**用户相同（或重合较多）、物品不同（或重合较少）**时，属于**纵向联邦学习**范式。
- 当不同组织间**用户不同（或重合较少）、物品相同（或重合较多）**时，属于**横向联邦学习**范式。

推荐问题分类

- **双跨联邦推荐**同时考虑组织间的数据安全和每个组织内的个体用户（或用户群组）间的隐私。

推荐问题分类

- 此外，结合评分预测和物品排序等不同的建模目标，每一类联邦推荐问题又可以进一步细分为面向评分预测的联邦推荐问题和面向物品排序的联邦推荐问题。

推荐问题分类

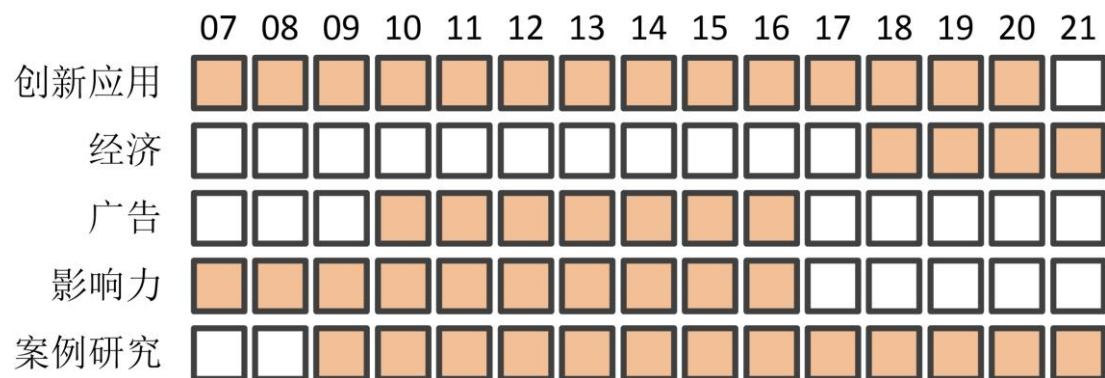
- **多源数据推荐** (multi-source recommendation) 是指算法可以使用用户行为数据和其他数据，包括**物品**的类别、内容描述和知识图谱数据，**用户**的社交关系数据，**行为**的时间、位置等上下文信息，等等。
- 为了有效地融合多种数据和知识，在工业界的应用中，通常通过与**特征工程** (feature engineering)、**知识图谱** (knowledge graph)、**深度学习** (deep learning)、**集成学习** (ensemble learning)、**检索重排**、**混合推荐**和**A/B测试**等相结合的方式进行建模，因为这种方式往往具有较好的推荐效果和可扩展性。
- 说明：多源数据推荐侧重于工程实践，本课程重点关注单行为推荐、多行为推荐、序列推荐和联邦推荐等问题中**关于用户行为建模的技术**。

推荐问题分类

- 推荐系统的研究领域较为宽泛，根据推荐系统领域的旗舰国际会议 **ACM RecSys 2007~2021** 征文通知中的研究话题和研讨会主题，我们总结得到其中的研究方向和问题，大致包括 **实践应用**、**问题设置**、**算法研究** 和 **用户研究** 四个方面。
- 在本书中，我们重点关注不同问题设置中的 **推荐技术和算法**，较少涉及实践应用和用户研究。

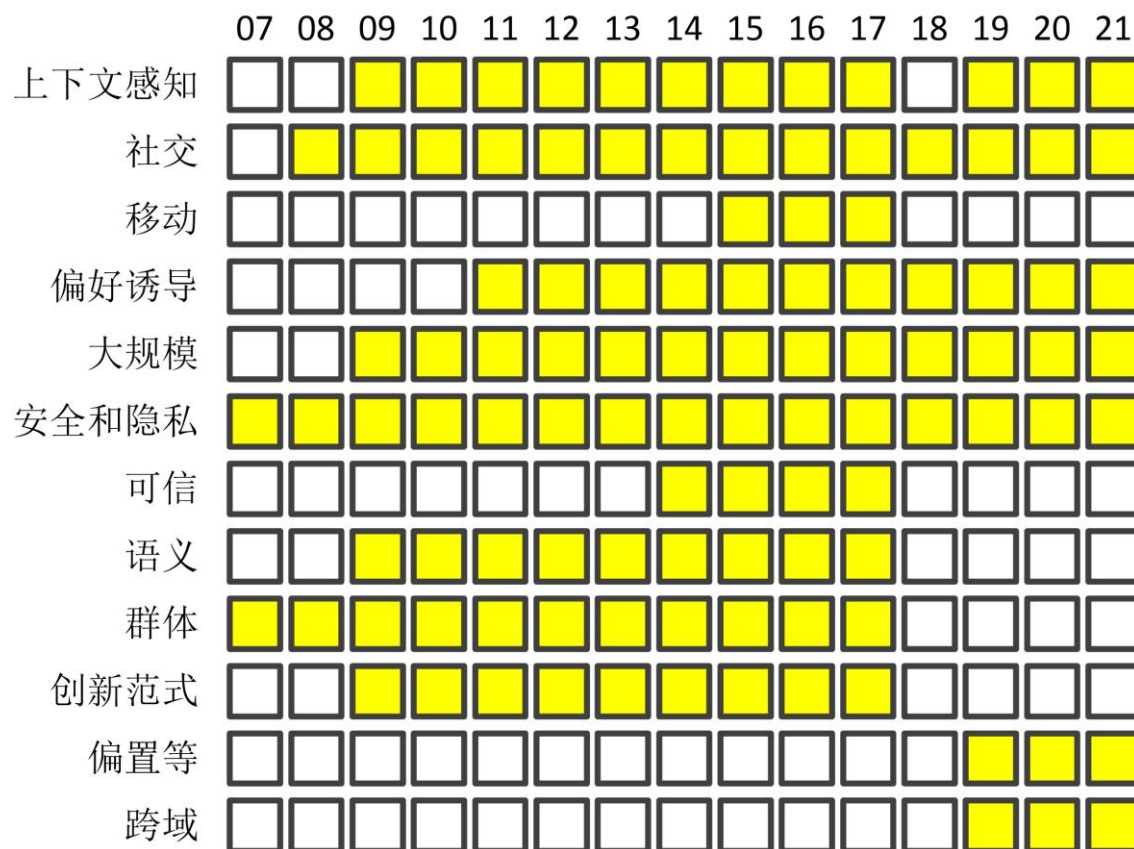
推荐问题分类

- 实践应用方面的研究话题的出现情况



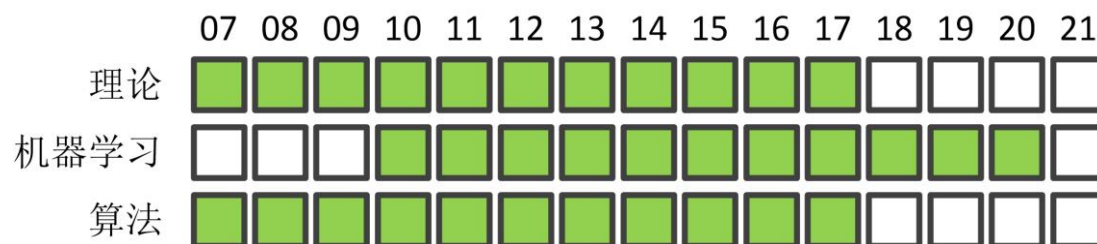
推荐问题分类

- 问题设置方面的研究话题的出现情况



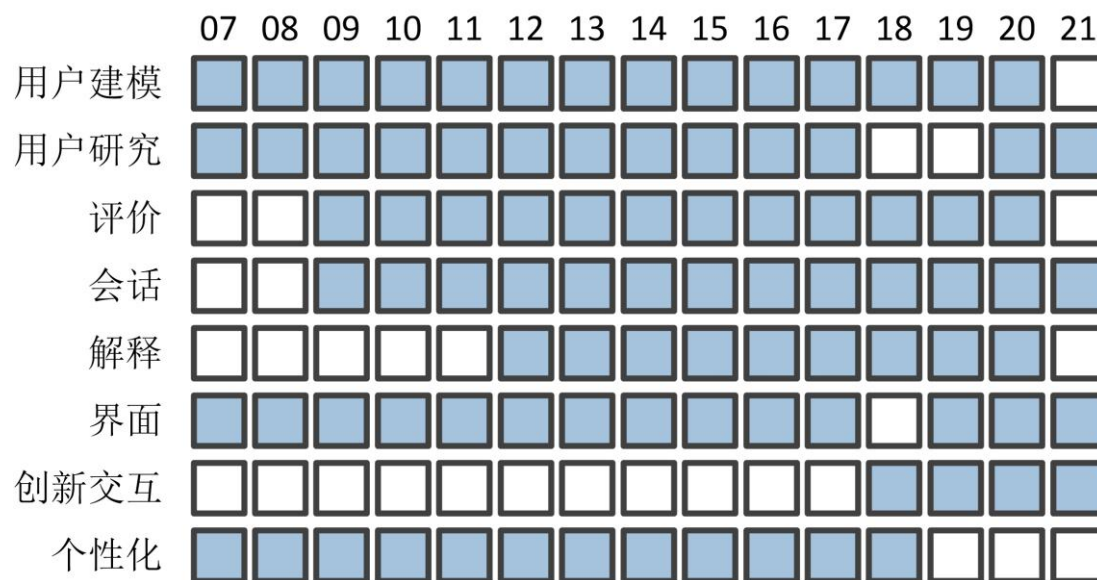
推荐问题分类

- 算法研究方面的研究话题的出现情况



推荐问题分类

- 用户研究方面的研究话题的出现情况

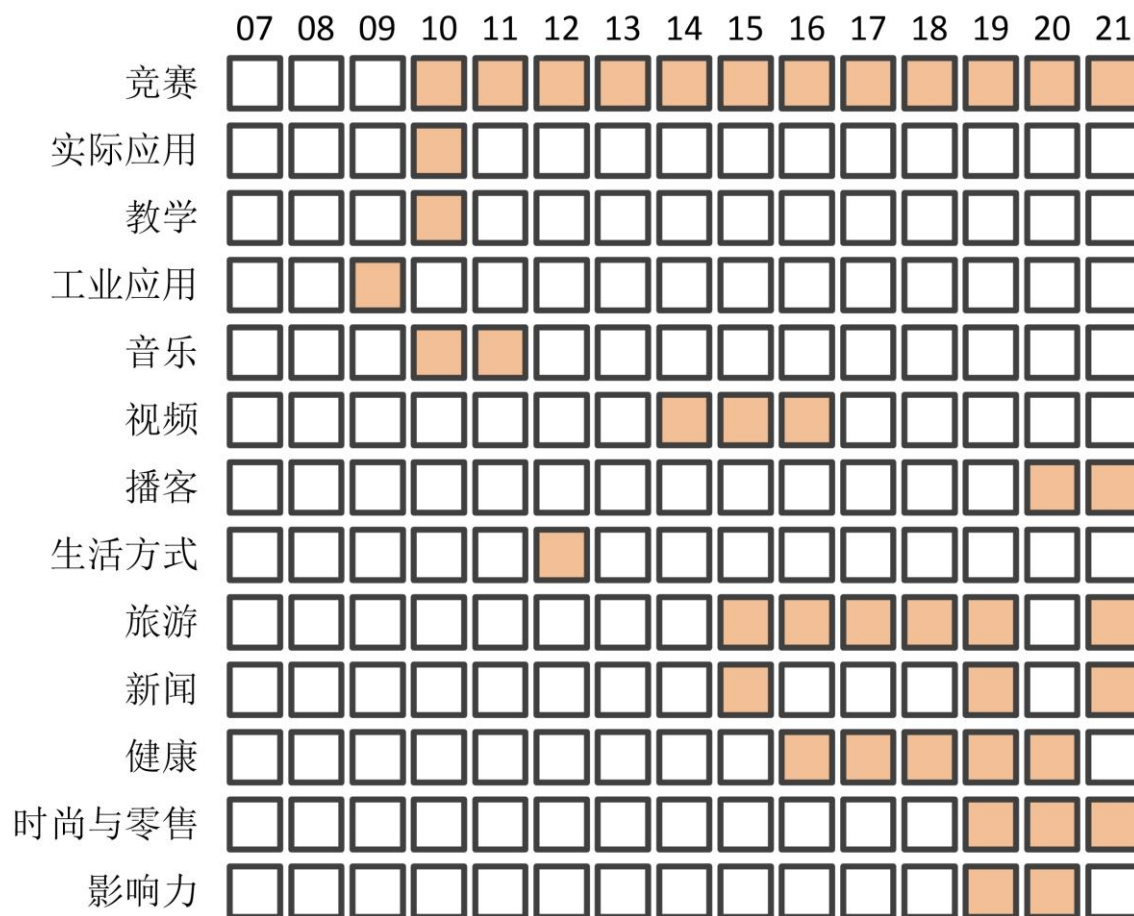


推荐问题分类

- 我们从中可以看到研究话题的历史变迁，特别是近几年新出现的话题，如经济、偏置、跨域和创新交互等。

推荐问题分类

- 实践应用方面的研讨会主题的出现情况



推荐问题分类

- 问题设置方面的
研讨会主题的出现情况



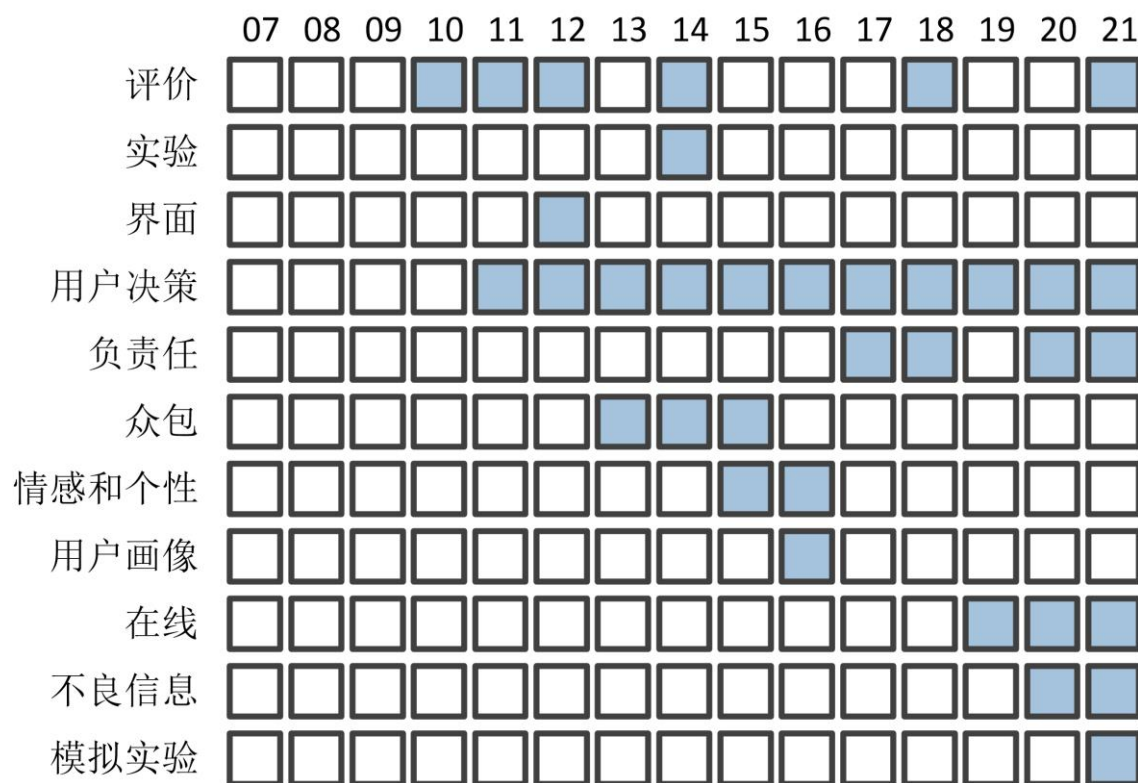
推荐问题分类

- 算法研究方面的研讨会主题的出现情况

	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21
结果重现	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
深度学习	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
强化学习	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
图神经网络	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

推荐问题分类

- 用户研究方面的研讨会主题的出现情况



推荐问题分类

- 我们从中可以看到研讨会主题的历史变迁，特别是近几年新出现的主题，如播客、时尚与零售、跨市场、多目标、图神经网络和不良信息等。

推荐问题分类

- 特别值得一提的是，自21世纪初至今，国内学者在推荐系统领域取得了一系列重要成果，包括多篇中文综述论文，非常值得我们关注。
- 曾春,邢春晓,周立柱. 个性化服务技术综述[J]. 软件学报, 2002, 13(10):1952-1961.
- 刘建国,周涛,汪秉宏. 个性化推荐系统的研究进展[J]. 自然科学进展, 2009, 19(1):1-15.
- 许海玲,吴潇,李晓东,阎保平. 互联网推荐系统比较研究[J]. 软件学报, 2009, 20(2): 350-362.
- 王立才,孟祥武,张玉洁. 上下文感知推荐系统[J]. 软件学报, 2012, 23(1):1-20.
- 黄震华,张佳雯,田春岐,孙圣力,向阳. 基于排序学习的推荐算法研究综述[J]. 软件学报, 2016, 27(3):691-713.
- 黄际洲,孙雅铭,王海峰,刘挺. 面向搜索引擎的实体推荐综述[J]. 计算机学报, 2019, 42(07):1467-1494.
- 廖国琼,蓝天明,黄晓梅,陈辉,万常选,刘德喜,刘喜平. 基于事件社会网络推荐系统综述[J]. 软件学报, 2021, 32(2):424-444.
- 张玉洁,董政,孟祥武. 个性化广告推荐系统及其应用研究[J]. 计算机学报, 2021, 44(3):531-563.

第1章 概述

- 推荐技术简介
- 推荐问题分类
- 常用数据集
- 常用验证方法
- 深度学习平台简介
- 本章小结
- 习题

常用数据集

- **MovieLens**电影评分数据集：是推荐算法研究中最常用的数据集之一，由美国明尼苏达大学（University of Minnesota）的GroupLens研究组收集和发布，包含了多个在不同时间段内收集到的规模各不相同的子数据集，例如，最经典的由943个用户对1682部电影的100000条评分记录组成的**MovieLens 100K**数据集（1998年4月发布）、由6040个用户对3952部电影的1000209条评分记录组成的**MovieLens 1M**数据集（2003年2月发布）和由71567个用户对10681部电影的10000054条评分记录组成的**MovieLens 10M**数据集（2009年1月发布），以及较新的在2019年12月发布的**MovieLens 25M**数据集，等等。考虑到推荐系统数据的稀疏性问题，MovieLens数据集中仅保留了评分记录数不小于20的用户的数据。并且，除了最基本的(用户ID,物品ID,评分,时间戳)信息外，MovieLens数据集还提供了一些用户的基本信息（性别、年龄和职业）和物品的描述信息（电影名、电影类型和上映日期）。
- <https://grouplens.org/datasets/movielens/>

常用数据集

- **Netflix电影评分数据集**：也是一个推荐算法研究中常用的历史比较悠久的历史数据集，由电影租赁网站Netflix在2005年底发布，用于**Netflix百万美元竞赛**，以激励人们设计和开发更好的推荐算法。Netflix数据集中包含了在1998年10月至2005年12月期间，**480189**个匿名的Netflix用户对**17770**部电影的**99072112**条评分记录（包含评分日期），并且包含了电影的标题和发行年份信息。

常用数据集

- **Amazon电商评论数据集**：美国加州大学圣迭戈分校（University of California, San Diego）的Julian McAuley教授在其个人主页上发布了他的研究团队收集的一些推荐系统数据集，其中有一个从知名电商平台Amazon上收集到的关于商品的评论数据集。Amazon数据集包含了1996年5月至2014年7月期间的约1.4亿条评论数据（每条评论由评分及文字评论等组成），并且给出了一些商品的描述、类别、价格、品牌和图像特征等信息。为了方便研究，研究人员还将原数据集按照商品类别分为24个较小的子数据集，例如，图书数据集、电子产品数据集和电子游戏数据集等。
- <http://jmcauley.ucsd.edu/data/amazon/>

常用数据集

- **天猫电商多行为数据集**：天猫（Tmall）电商多行为数据集是在IJCAI 2015竞赛中发布的，包含了一些天猫用户在“双十一”前6个月的匿名行为日志。一条行为记录对应的信息有：**用户ID、物品ID、时间戳和行为类型**等等。其中，记录的行为类型包括点击、加入购物车、购买和收藏。Tmall数据集还提供了用户的性别和年龄等信息，以及物品的类别、商家和品牌等信息。
- <https://tianchi.aliyun.com/dataset/dataDetail?dataId=42>

第1章 概述

- 推荐技术简介
- 推荐问题分类
- 常用数据集
- 常用验证方法
- 深度学习平台简介
- 本章小结
- 习题

常用验证方法

- 交叉验证法：
- 对于不考虑序列信息的推荐问题，我们可以使用交叉验证法（cross validation）来对模型进行训练和验证，具体方法为：对于一个数据集，将它随机分成 k （例如 $k = 5$ ）个大小相同的子集，每次取其中一个子集作为测试集，其余 $k - 1$ 个子集共同构成训练集，如此重复 k 次，最后取在 k 个不同测试集上的结果的平均值作为最终的评估结果。
- 注意，在大多数算法中，我们需要先使用一个较小的验证集来寻找最优的模型参数。

常用验证方法

- 留一法：
- 留一法 (leave-one-out) 是交叉验证法的一个特例，即将数据集中的每个样本都分到一个不同的子集中（即 k 等于数据集的大小），只选取一个样本用作测试，其余数据全部用于训练。在推荐算法研究中，留一法多用于引入序列信息的情况。通常，为了评估模型对用户行为序列的预测能力，我们将每个用户的最后一次行为记录划入测试集，其余记录划入训练集。
- 注意，在进行正式的模型测试前，为了调节模型参数，可以将每个用户的倒数第二次行为记录划入一个新增加的验证集，然后使用剩下的训练集数据来训练模型并观察在各参数下的表现。

常用验证方法

- 强泛化效果验证和弱泛化效果验证：
- 在推荐技术的验证中，比较主流的模型效果验证方法包括强泛化效果验证和弱泛化效果验证。

常用验证方法

- 强泛化效果验证先对用户进行划分从而得到训练集用户 \mathcal{U}^{tr} 、验证集用户 \mathcal{U}^{vad} 和测试集用户 \mathcal{U}^{te} ，使得 $\mathcal{U}^{\text{tr}} \cap \mathcal{U}^{\text{vad}} \cap \mathcal{U}^{\text{te}} = \emptyset$ 且 $\mathcal{U}^{\text{tr}} \cup \mathcal{U}^{\text{vad}} \cup \mathcal{U}^{\text{te}} = \mathcal{U}$ ，再将每一位用户的交互数据划分到对应的数据中。显然，验证集用户和测试集用户的交互数据没有出现在训练过程中，因此模型需要具备较强的泛化能力才能获得较好的推荐效果。

常用验证方法

- 弱泛化效果验证则将每一位用户的交互数据按照设定的阈值划分到训练集、验证集和测试集中，因此验证集用户和测试集用户在训练过程中会以训练集用户的角色出现并参与到训练过程中。
- 显而易见，弱泛化效果验证在验证和测试模型效果时，需要衡量模型在每一个用户上的推荐效果，当用户数目较多时，这种验证方式相较于强泛化效果验证而言效率较低，并且由于验证集用户和测试集用户在训练中已经出现并参与了训练，因此弱泛化验证对模型泛化能力要求相对较低。

第1章 概述

- 推荐技术简介
- 推荐问题分类
- 常用数据集
- 常用验证方法
- 深度学习平台简介
- 本章小结
- 习题

深度学习平台简介

- **TensorFlow**和**PyTorch**分别是由谷歌人工智能团队和Facebook人工智能研究院开发的两个常用的深度学习框架，两者都对Python提供了很好的支持，具有丰富的社区资源，非常适合新手入门。
- **Keras**是以TensorFlow或Theano作为后端进一步封装得到的高级API组件，模块化程度较高，用户交互比较友好。
- **PaddlePaddle**（飞桨）是百度自主研发的产业级深度学习平台，集成了推荐系统、计算机视觉和自然语言处理等领域中常见的模型，并配有详细的文档和教程，用户可以根据自己的需求快速实现相应的深度学习算法。
- 可以从以下地址找到这些深度学习平台的**中文版安装及使用指引**：
 - (1) TensorFlow: https://tensorflow.google.cn/overview/?hl=zh_cn
 - (2) PyTorch: <https://pytorch123.com/>
 - (3) Keras: <https://keras.io/zh/>
 - (4) PaddlePaddle: <https://www.paddlepaddle.org.cn/>

第1章 概述

- 推荐技术简介
- 推荐问题分类
- 常用数据集
- 常用验证方法
- 深度学习平台简介
- 本章小结
- 习题

本章小结

- 对推荐技术和用户行为等进行了介绍
- 对本课程中涉及的各类推荐问题进行了介绍
- 对推荐算法研究领域常用的数据集和验证方法（含数据集预处理方法）进行了介绍
- 列举了几种常用的深度学习平台
- 需要特别说明的是，智能推荐技术的研发和部署可能会涉及用户隐私和权益、网络秩序和安全等方面的问题，因此，研究者和实践者必须遵守有关法律法规，用好管好推荐技术，让技术给用户、企业和社会带来益处。

第1章 概述

- 推荐技术简介
- 推荐问题分类
- 常用数据集
- 常用验证方法
- 深度学习平台简介
- 本章小结
- 习题

习题

- 1. 请想一想，以某电商平台为例，从用户访问电商网站到完成下单，推荐技术扮演了什么角色？
- 2. 请结合实际例子分析智能推荐系统与信息检索系统（例如搜索引擎）的联系与区别。
- 3. 请举例说一说，智能推荐系统和技术给你的生活和学习带来了什么样的影响？
- 4. 请举例说一说，智能推荐技术可以应用在哪些场景（1.1节中提到的除外）？
- 5. 如何看待智能推荐技术应用广泛这一现象？
- 6. 请想一想，在哪些场景中适合使用以评分预测为目标的推荐算法，在哪些场景中适合使用以物品排序为目标的推荐算法？
- 7. 请用MATLAB或Python绘制sigmoid、tanh和ReLU函数的曲线图。
- 8. 关于推荐效果的评估，除了推荐的准确程度以外，你觉得还有哪些比较重要的指标？

习题

- 9. 请用示意图的方式展示精确率 (Precision) 和召回率 (Recall) 的区别。
- 10. 在什么情况下，召回率 (Recall) 和1-call的值是相等的？
- 11. 请实现面向评分预测的评价指标（包括不同用户段和物品段上的评价指标）。
- 12. 请实现面向物品排序的评价指标。
- 13. 请想一想，在对推荐列表进行评估时，与精确率 (Precision) 相比，归一化折损累计增益 (NDCG) 更加关注哪个方面？
- 14. 请想一想，自行搭建一个包含推荐系统的网络平台有哪些需要注意的事项？
- 15. 关于智能推荐技术，你认为什么样的智能才是真正的智能？
- 16. 请想一想，10年后的智能推荐技术会是怎么样的？
- 17. 请查阅相关资料，举例说一说TensorFlow与PaddlePaddle的异同点。