

# 推荐系统：从个性化推荐到广告投放（上）

H 博士 & 袁从德 | 2026-03-22 | 强化学习快速入门与实战 · 应用篇

你好！欢迎来到强化学习的第十九讲。在前几讲中，我们系统学习了强化学习的基础框架、核心算法、若干进阶专题，以及其在机器人控制等场景的应用，而今天我们将转向一个更贴近日常生活、商业价值极高的领域——推荐系统。我们将用两节课来探索如何用技术逻辑连接"满足用户个性化需求"与"实现平台商业变现"，尤其聚焦强化学习如何解决推荐系统中的关键痛点。

通过本节课的学习，你将能够：

- 掌握推荐系统的核心逻辑：**理解用户画像、物品特征的构建方法，以及协同过滤、基于内容、矩阵分解等经典推荐算法的原理差异与适用场景。
- 理清推荐系统的完整架构：**明确数据层（数据收集与预处理）、算法层（单算法与混合推荐策略）、业务层（结果展示与用户反馈）的协同机制。

## 一、强化学习——破解推荐系统核心痛点的关键技术支撑

在信息爆炸的互联网时代，推荐系统既是缓解用户信息过载、提升体验的"导航仪"，也是平台提高用户留存、实现商业价值的"核心引擎"。而传统推荐系统受限于"离线训练 - 固定部署"的静态逻辑，长期面临滞后性与短视性两大核心痛点——既无法实时响应用户兴趣突变与环境变化，又易因聚焦短期点击率陷入"信息茧房"，损害长期用户价值。

强化学习的出现，为突破这些痛点提供了革命性的技术路径。它通过构建"环境 - 智能体 - 反馈"的动态交互闭环，让推荐系统从"被动匹配历史偏好"转向"主动预判动态需求"，既能实时捕捉用户状态与环境变化，即时调整推荐策略以打破滞后性陷阱，又能以"长期累积奖励最大化"为目标，平衡短期指标与长期体验，从根本上革新了推荐系统的决策逻辑，成为推动推荐技术从静态走向动态、从短视走向长效的关键力量。

### 1.1 从静态推荐到动态决策

在 2018 年之前，以协同过滤为代表的主流推荐系统，虽在特定场景下实现了基础的个性化推荐，但受限于技术框架，存在两大难以突破的本质缺陷，严重制约了推荐效果的进一步提升。

传统推荐系统采用"离线训练 - 固定部署"的静态决策逻辑，模型基于历史用户行为数据完成训练后，参数即处于固定状态，无法根据实时变化的环境与用户需求进行动态调整。这种模式导致系统对用户兴趣的捕捉存在显著滞后，尤其在用户偏好发生突变的场景下表现得更为突出。例如，在春节、双十一等重要节假日，用户的消费需求和内容偏好会出现剧烈波动——平日关注日常通勤穿搭的用户可能转而搜索年货礼品，专注于短视频娱乐的用户可能更倾向于观看返乡攻略类内容，但传统模型往往需要 2~3 天的时间才能通过积累新的历史数据完成迭代，期间推荐内容与用户真实需求严重脱节。

静态推荐系统的优化目标高度聚焦于即时点击率（CTR）等短期指标，通过持续推送用户当下可能点击的内容来提升短期数据表现。但这种"短视化"策略往往以牺牲长期用户体验为代价：系统会不断强化用户的短期兴趣，形成"信息茧房"，忽略用户潜在的长期兴趣探索。例如，若用户因偶然因素点击了一条八卦新闻，系统会持续推送同类内容，而不再推荐其原本长期关注的科技资讯，最终导致用户因内容同质化而产生厌倦感，严重影响平台的长期用户留存与活跃度。

为解决静态推荐系统的滞后性与短视性问题，以强化学习（Reinforcement Learning）为核心的动态决策框架逐渐成为推荐系统的革新方向。强化学习通过模拟"智能体与环境交互"的过程，实现了推荐决策的动态化与长期化，其核心逻辑可结合图 19-1 的强化学习应用框架拆解为以下关键维度。

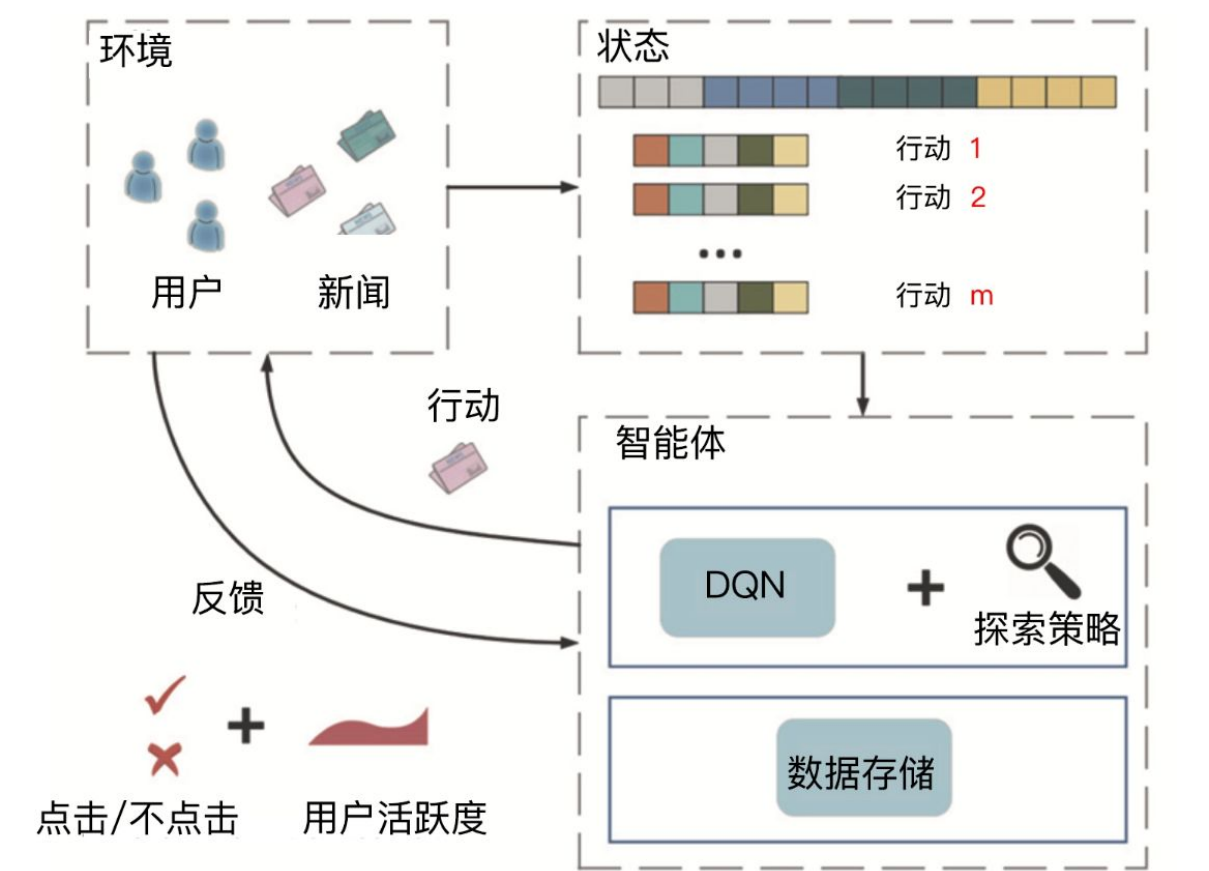


图 19-1 强化学习在推荐系统中的应用框架

## 1.2 强化学习驱动的推荐系统革新

强化学习与推荐系统的融合，已在技术演进中沉淀下成熟的实践路径。相较于依赖离线数据训练、决策逻辑固化的传统推荐模型，强化学习凭借其独特的技术特性，为推荐系统突破瓶颈提供了关键支撑，核心优势体现在：实时动态的在线学习能力和原生适配的探索与利用（EE）机制两大维度。

**实时动态的在线学习能力**，让推荐决策更贴合用户兴趣的动态变化。在推荐场景中，用户的实时上下文（如当前浏览场景、近期行为轨迹、设备与时间信息等）可被精准建模为“状态（state）”，而推荐系统作为与环境交互的“智能体（agent）”，能基于实时生成的状态信号动态调整推荐动作（action）——无论是用户兴趣从“日常穿搭”转向“节日礼品”，还是环境从“工作日通勤”切换为“周末休闲”，强化学习都能跳过传统模型依赖离线数据迭代的滞后环节，即时捕捉这些变化并更新推荐策略，从而更敏锐地跟上用户兴趣的演进节奏。

**原生适配的探索与利用机制**，有效破解了传统推荐模型的“多样性困境”。对传统推荐模型而言，如何在“利用已知兴趣提升短期效果”与“探索潜在偏好拓展推荐边界”之间找到平衡——过度“利用”易导致推荐内容同质化，形成“信息茧房”；盲目“探索”又可能因推荐偏离用户兴趣引发体验下降。而强化学习算法天然集成了系统化的探索策略（如  $\epsilon$ -greedy、Thompson 采样等），能在保证推荐准确性的基础上，有策略地向用户推送潜在兴趣相关内容，既避免了推荐结果的单调固化，又通过持续探索丰富了用户的兴趣维度，实现“精准性”与“多样性”的动态平衡。

强化学习在推荐系统中形成了完整的“环境 – 智能体 – 反馈”交互闭环，各要素与推荐场景深度适配：

- **环境（Environment）**：涵盖推荐场景中的所有动态因素，包括当前时间（如是否为节假日）、用户实时状态（如登录设备、浏览时长）、内容池更新情况等，为决策提供实时上下文信息。
- **智能体（Agent）**：即推荐系统的决策核心，替代了传统的固定参数模型，负责根据环境状态动态生成推荐策略（如行动 1、行动 2 至行动 m 对应的不同内容组合方案）。
- **行动（Action）**：智能体的输出结果，即向用户推送的具体内容组合，行动的选择直接影响用户反馈与后续状态。
- **反馈（Reward）**：打破单一短期指标局限，不仅包含即时点击率，更融入了用户停留时长、复购率、长期留存率等长期价值指标，形成多维度奖励信号。
- **数据存储与探索策略**：系统通过实时存储交互数据，支持智能体持续学习；同时引入“探索 – 利用”（Exploration–Exploitation）策略，在推送用户已知兴趣内容（利用）与探索潜在兴趣内容（探索）之间寻求平衡，避免“信息茧房”。

**强化学习在推荐系统的核心优势——破解静态推荐的两大难题：**

- **实时适配，打破滞后性陷阱：**强化学习的智能体无需依赖离线历史数据迭代，而是通过与环境实时交互不断更新决策逻辑。当环境（如节假日到来）或用户状态（如兴趣切换）发生变化时，智能体可立即捕捉到新的状态信号，调整推荐行动，实现"即时感知 – 即时响应"，从根本上解决了静态模型的滞后问题。
- **长期优化，规避短视性偏差：**强化学习的核心目标是最大化"累积奖励"，而非单一即时奖励。智能体会权衡当前点击率与未来留存率等长期收益，例如，即使某条潜在兴趣内容的即时点击率较低，但系统预判其可能激发用户的长期活跃，仍会选择推送该内容。这种"长期价值优先"的决策逻辑，有效平衡了短期指标与长期用户体验，显著提升用户活跃度与留存率。

## 二、个性化推荐基础

在信息爆炸的互联网时代，用户日均面临百万级的商品、内容与服务选择——电商平台的商品库动辄千万 SKU，内容平台的短视频日更新量超亿条。传统"人找信息"的搜索模式，已无法满足用户高效获取感兴趣内容的需求，而个性化推荐正是解决这一矛盾的核心技术：它通过深度洞察用户需求，将"信息找人"变为现实，既缓解了用户的信息过载焦虑，也成为平台提升用户留存、实现商业价值的关键引擎。

要理解个性化推荐的运作逻辑，需先掌握其三大核心支柱：刻画用户需求的用户画像、定义推荐对象的物品特征，以及连接二者的推荐算法。三者共同构成了"推荐给谁（用户）– 推荐什么（物品）– 怎么推荐（算法）"的技术闭环，也是后续理解强化学习如何优化推荐系统的基础。

### 2.1 核心概念：拆解个性化推荐的"三要素"

个性化推荐的本质，是让系统像"懂用户的顾问"一样，精准匹配"人的需求"与"物的价值"。这一过程的实现，依赖于对"用户、物品、匹配逻辑"三个核心要素的精准定义。

#### 2.1.1 用户画像：构建"可计算的用户需求"

用户画像是对用户需求的结构化、动态化抽象描述，其核心目标是将零散的用户数据（如点击、购买、浏览记录）转化为系统可理解的"特征向量"，让系统能实时感知"用户是谁、喜欢什么、需要什么"。它不是静态的标签堆砌，而是随用户行为持续迭代的"需求动态模型"，主要由静态属性与动态行为两大维度构成。

##### 静态属性，用户的"基础身份标签"

这类数据相对稳定，是构建用户画像的"初始骨架"，主要包括：

- **基础信息：**如年龄（需避免直接采集，可通过消费能力、内容偏好间接推断）、性别、地域（精确到城市或区县，影响本地化推荐，如地方特产、同城服务）、设备类型（iOS 用户与

Android 用户的消费习惯可能存在差异，如 iOS 用户更倾向于高客单价商品）、注册时间（新用户与老用户的推荐策略需区分，新用户需"探索兴趣"，老用户需"深化偏好"）。

- **社会属性：**如职业（学生群体关注学习用品、考研课程，职场人士关注办公软件、商务服饰）、收入层级（通过历史消费金额、支付方式间接判断，影响商品价格带的推荐范围）、家庭结构（有孩子的用户可能关注母婴用品、亲子内容，独居用户更关注小家电、便捷食品）。

静态属性的价值在于快速建立用户的"基础认知"，尤其在用户行为数据不足时（如冷启动阶段），能为推荐提供初步方向。

### 动态行为，用户的"兴趣动态轨迹"

相较于静态属性的稳定性，动态行为数据更能反映用户实时、变化的偏好，是用户画像的"鲜活血肉"，主要包括：

- **浏览行为：**如浏览时长（对某类商品停留超过 3 分钟，可能表示深度兴趣）、浏览路径（从"手机"页面跳转至"手机配件"，说明存在关联需求）、重复浏览次数（多次查看某件商品，可能有购买意向或决策犹豫）。
- **交互行为：**如点击（对推荐内容的"初步认可"，是兴趣的直接信号）、收藏/加购（对商品的"主动留存"，表明明确的潜在需求，优先级高于普通点击）、评论/点赞（对内容的情感反馈，如点赞"环保主题文章"，可强化"环保意识"标签）、分享（对内容的"主动传播"，说明高度认同，可作为核心兴趣标签）。
- **交易行为：**如购买记录（最直接的兴趣证明，购买"运动跑鞋"后，可关联推荐"运动袜""运动手环"）、复购频率（复购 3 次以上的商品，说明用户对品牌或品类的忠诚度，需重点维护）、退货/售后原因（如因"尺寸不符"退货，可调整后续推荐的尺码适配建议；因"质量问题"退货，可降低同类商品的推荐权重）。

动态行为数据的采集需注重"实时性"，例如用户在电商平台搜索"春节礼品"，系统需立即捕捉这一临时需求，调整推荐方向，避免因数据滞后导致推荐与需求脱节。此外，用户画像的构建还需遵循"动态迭代"原则——随着用户行为的积累（如用户从"学生"转变为"职场人"，兴趣从"考研资料"转向"职场技能课程"），画像需实时更新特征权重，确保始终贴合用户当前的真实需求，避免因"标签固化"导致推荐偏差。

### 2.1.2 物品特征：定义"可匹配的物品价值"

物品特征是对推荐对象（商品、文章、视频等）的结构化属性描述，其核心目标是将"非结构化的物品"（如一段视频、一件衣服）转化为系统可计算的"特征语言"，以便与用户画像的需求特征进行匹配。不同类型的物品，其特征维度差异显著，但核心逻辑均是"提取用户决策时关注的核心属性"。

#### (1) 实物商品：聚焦"决策关键因素"

以电商平台的实物商品（如手机、服装、家电）为例，特征设计需覆盖用户购买决策的全链路关注点：

- **基础属性：**品牌（如"苹果""华为"，影响用户对品质与价格的预期）、品类（如"智能手机""连衣裙"，用于归类推荐）、价格带（如"2000–3000 元""500 元以下"，匹配用户消费能力）；
- **功能属性：**手机的"处理器型号（骁龙 8 Gen2）""电池容量（5000mAh）""摄像头像素（1 亿像素）"（对应性能、续航、拍照需求），服装的"材质（纯棉）""尺码（M 码）""风格（通勤风）"（对应舒适度、合身度、审美需求）；
- **场景属性：**如"通勤背包"（适配日常上班场景）、"户外冲锋衣"（适配登山露营场景），场景属性能让推荐更贴合用户的使用场景（如向计划旅行的用户推"便携行李箱"）。

### (2) 内容产品：聚焦"兴趣匹配维度"

以内容平台的文章、视频为例，特征设计需围绕"内容价值与用户兴趣的契合度"：

- **内容属性：**文章的"关键词（人工智能、新能源）""主题分类（科技、财经、娱乐）""情感倾向（正面、负面，如向焦虑用户推积极生活类内容）"；视频的"时长（15 秒短视频、5 分钟科普）""创作者（用户关注的 UP 主，可作为信任背书）""标签（#美食教程、#旅行 vlog）"；
- **互动属性：**文章的"阅读完成率（超 80%→内容吸引力强）""评论热词（如'干货满满'→强化'实用型内容'标签）"；视频的"完播率（短视频超 50%→优质信号）""弹幕关键词（如'太治愈了'→提取'治愈系'标签）"。

### (3) 服务产品：聚焦"体验与时效性"

以外卖、旅游、本地生活服务为例，特征设计需突出"服务质量与即时需求匹配"：

- **服务属性：**外卖的"配送时长（30 分钟内）""商家评分（4.8 分以上）""菜品类型（川菜、轻食）"；旅游套餐的"行程天数（3 天 2 晚）""目的地（三亚、大理）""包含服务（含机票、含导游）"；
- **动态属性：**外卖的"实时库存（仅剩 5 份→制造稀缺感）""优惠活动（满 30 减 10→匹配性价比需求）"；旅游套餐的"剩余名额（仅剩 2 位→促进行动）""出行日期（国庆档期→匹配假期计划）"。

物品特征的完整性直接决定推荐精度——若推荐服装时缺失"尺码"特征，可能向需要大码的用户推均码商品；若推荐手机时缺失"5G 支持"特征，可能向需要 5G 的用户推 4G 机型，最终导致"推荐错位"。

### 2.1.3 推荐算法：搭建"需求与价值的匹配桥梁"

推荐算法是连接"用户画像（需求）"与"物品特征（价值）"的核心技术手段，其本质是通过数学模型计算"用户对物品的偏好程度（推荐得分）"，并按得分高低生成推荐列表。无论后续技术如何升

级（如引入强化学习），推荐算法的核心工作流程始终不变。

1. **特征输入**：将用户画像的特征向量（如"25 岁女性、上海、喜欢户外、消费能力中等"）与物品特征的特征向量（如"户外冲锋衣、品牌 A、价格 800 元、防水材质"）输入模型；
2. **偏好计算**：通过算法模型（如协同过滤、基于内容推荐）计算两者的匹配度——例如"用户喜欢户外"与"物品为户外冲锋衣"匹配度高，"用户消费能力中等"与"物品价格 800 元"匹配度高，综合得出推荐得分；
3. **结果输出**：根据推荐得分对候选物品排序，将得分最高的前 N 个物品（如前 20 个）作为推荐列表展示给用户（如电商首页的"猜你喜欢"、内容平台的"首页推荐"）。

推荐算法的选择，需结合"数据情况""业务场景"与"推荐目标"综合判断——例如用户 – 物品交互数据丰富时用协同过滤，物品内容特征清晰但交互数据稀疏时用基于内容推荐，需挖掘潜在关联时用矩阵分解。

## 2.2 主流推荐算法：从"相似匹配"到"潜在关联"

个性化推荐算法历经多年发展，已形成三类核心技术路径：基于协同过滤的"相似匹配"、基于内容的"兴趣延续"、基于矩阵分解的"潜在关联挖掘"。三者各有优劣，适用于不同的数据场景与业务需求，也是后续强化学习优化的"传统基准"。

### 2.2.1 协同过滤算法："物以类聚，人以群分"

协同过滤算法是推荐系统中最经典的算法之一，它的基本假设是：相似的用户会喜欢相似的物品。该算法主要分为基于用户的协同过滤和基于物品的协同过滤。

**基于用户的协同过滤**，首先要找到与目标用户兴趣相似的用户群体，然后根据这个群体对物品的喜好，为目标用户进行推荐。比如，用户 A 和用户 B 都喜欢电影《复仇者联盟》《钢铁侠》，而用户 A 还喜欢《美国队长》，那么系统就可能会把《美国队长》推荐给用户 B。在实际计算中，通常会使用余弦相似度、皮尔逊相关系数等方法来衡量用户之间的相似度。

**基于物品的协同过滤**则是先计算物品之间的相似度，找到与目标物品相似的物品集合，然后根据用户对目标物品的行为（如购买、浏览等），向用户推荐相似物品。例如，用户购买了一款华为手机，而这款手机与另一款华为手机在配置、价格等方面非常相似，系统就可能会将另一款手机推荐给该用户。基于物品的协同过滤在电商领域应用较为广泛，因为物品的特征相对稳定，计算物品之间的相似度也更为高效。

### 2.2.2 基于内容的推荐算法："喜欢即相似"

基于内容的推荐算法主要依据物品的内容特征来进行推荐。它通过对物品的文本、图像、音频等内容进行分析，提取出物品的特征向量，然后根据用户的历史行为，找到用户感兴趣的内容特征，再

将具有相似特征的物品推荐给用户。

以新闻推荐为例，系统会对每篇新闻文章进行文本分析，提取关键词、主题、情感倾向等特征。如果用户经常阅读科技类新闻，且对人工智能相关内容表现出浓厚兴趣，那么系统就会从新发布的新闻中筛选出具有相似科技、人工智能特征的文章推荐给该用户。基于内容的推荐算法的优点是能够很好地建模用户的兴趣，推荐结果具有较强的可解释性。但它也存在一些局限性，比如对于新出现的物品，如果缺乏足够的内容信息，可能无法准确推荐；而且对于用户兴趣的拓展能力相对较弱，容易导致推荐内容过于单一。

### 2.2.3 矩阵分解算法："挖掘潜在的兴趣关联"

矩阵分解算法是在协同过滤算法的基础上发展而来的，它主要用于解决协同过滤算法中的数据稀疏性问题。在实际的推荐场景中，用户 - 物品交互矩阵往往非常稀疏，即大部分用户对大部分物品都没有产生过行为，这会导致基于相似度计算的协同过滤算法效果不佳。

矩阵分解算法的基本思想是将用户 - 物品交互矩阵分解为两个低秩矩阵的乘积，一个是用户特征矩阵，另一个是物品特征矩阵。通过这种方式，将高维稀疏的用户 - 物品交互关系转化为低维稠密的特征关系，从而能够更好地挖掘用户和物品之间的潜在联系。在实际应用中，常用的矩阵分解方法有奇异值分解（SVD）、非负矩阵分解（NMF）等。矩阵分解算法能够有效提高推荐的准确性，并且在大规模数据场景下具有较好的扩展性。

## 三、推荐系统的架构与流程

推荐系统架构与流程，如图19-2所示，主要分为数据层、算法层以及业务层。

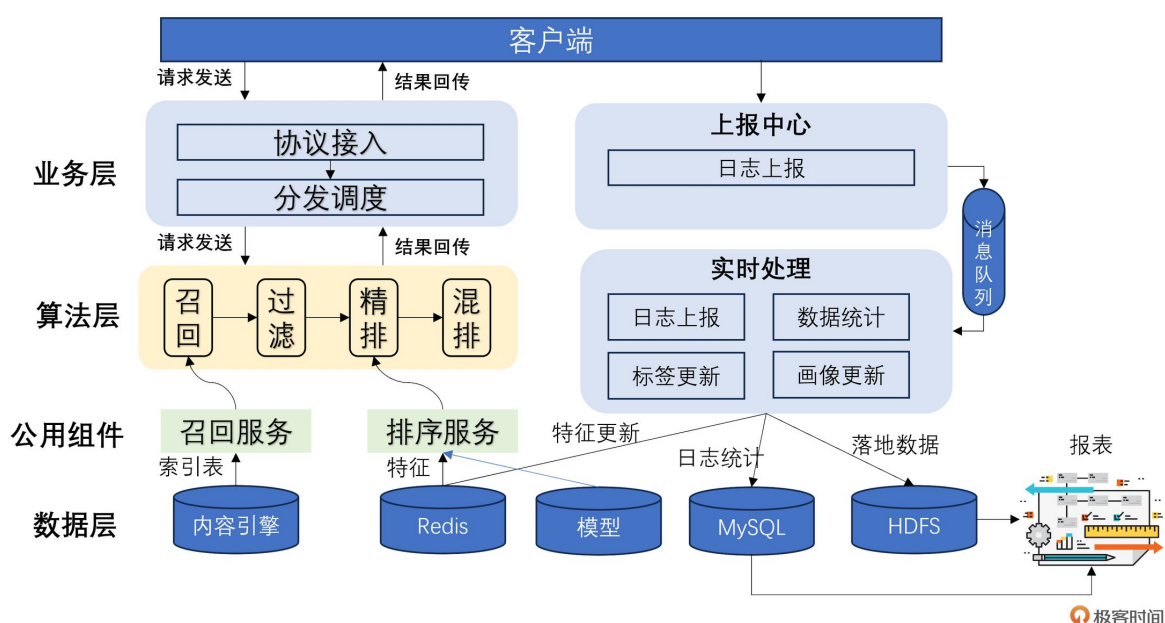


图 19-2 推荐系统的架构与流程

### 3.1 数据层

数据是推荐系统的基础，数据层负责收集、存储和预处理各种与用户和物品相关的数据。这些数据来源广泛，包括用户在平台上的注册信息、浏览行为、购买记录、评价反馈等，以及物品的基本属性、内容描述、图片视频等。

在数据收集阶段，需要确保数据的完整性和准确性，采用可靠的数据采集技术和工具，如日志采集系统、数据库同步工具等。收集到的数据会存储在关系型数据库（如 MySQL、Oracle）、非关系型数据库（如 MongoDB、Redis）或分布式文件系统（如 Hadoop HDFS）中。

数据预处理是数据层的关键环节，它包括数据清洗、去重、归一化、特征工程等操作。数据清洗用于去除数据中的噪声、缺失值和异常值，保证数据质量；去重可以避免重复数据对推荐结果的干扰；归一化则是将不同特征的数据统一到相同的尺度，便于后续的计算和分析；特征工程通过对原始数据进行提取、组合、转换等操作，生成更具代表性和预测性的特征，为推荐算法提供更好的数据支持。

### 3.2 算法层

算法层是推荐系统的核心，它包含了各种推荐算法和模型，负责根据数据层提供的数据，计算出用户对物品的推荐得分，并生成推荐列表。前面我们介绍了协同过滤算法、基于内容的推荐算法和矩阵分解算法等，这些算法在算法层中都有广泛的应用。

在实际应用中，为了提高推荐系统的性能和效果，通常会采用混合推荐的方式，即将多种推荐算法进行融合。例如，可以将协同过滤算法的结果和基于内容的推荐算法的结果按照一定的权重进行加权求和，得到最终的推荐列表。这样可以充分发挥不同算法的优势，弥补单一算法的不足。

随着深度学习技术的发展，深度学习算法在推荐系统中的应用也越来越广泛。深度学习模型能够自动学习数据中的复杂特征和模式，在处理大规模、高维度的数据时表现出明显的优势。常见的深度学习推荐模型有多层感知机（MLP）、卷积神经网络（CNN）、循环神经网络（RNN）及其变体（如 LSTM、GRU）等。这些模型可以用于构建用户画像、物品特征表示以及计算用户与物品之间的匹配度。

### 3.3 业务层

业务层主要负责将算法层生成的推荐结果展示给用户，并与用户进行交互，同时根据用户的反馈对推荐系统进行优化。在这一层，需要考虑推荐结果的展示形式、推荐场景和用户反馈机制等问题。

推荐结果的展示形式多种多样，常见的有列表式、瀑布流式、网格状等。不同的平台和业务场景会根据自身的特点和用户需求选择合适的展示形式。例如，电商平台通常会采用列表式展示商品推荐

结果，方便用户快速浏览和比较；而内容平台如抖音则采用瀑布流式展示视频推荐，让用户能够持续沉浸在内容消费中。

推荐场景也是业务层需要重点考虑的因素。不同的场景下，用户的需求和行为模式可能会有所不同，因此需要针对性地调整推荐策略。比如，在用户刚注册平台时，由于缺乏用户的历史行为数据，可能会采用基于热门物品或用户基本信息的推荐策略；而在用户浏览商品详情页时，根据用户当前浏览的商品，推荐与之相关的其他商品，如配件、替代品等。

用户反馈机制对于推荐系统的优化至关重要。业务层需要收集用户对推荐结果的反馈信息，如是否点击推荐内容、是否购买推荐商品、是否对推荐内容进行评价等。这些反馈信息会作为训练数据，用于更新和优化推荐算法模型，使得推荐系统能够不断适应用户的兴趣变化，提供更加精准的推荐服务。

---

下节课我们将继续学习推荐系统，会更加聚焦强化学习在推荐系统中的应用，总结在完整学习后进行。敬请期待！

希望你通过今天的学习，对强化学习有了清晰的认识。如果有任何疑问，欢迎在讨论区留言交流！