



推荐系统：工业架构与核心算法

作者：刘京欣



前言

在人工智能技术快速发展的今天，推荐系统已经成为连接用户与海量信息的重要桥梁。从短视频平台的个性化内容分发，到电商平台的商品推荐，再到新闻资讯、直播与搜索服务，推荐系统无处不在，并深刻影响着人们获取信息、消费内容以及进行决策的方式。与此同时，以大语言模型为代表的新一代人工智能技术迅速崛起，正在重新塑造推荐系统的技术范式与发展方向，为推荐领域带来了新的机遇与挑战。

近年来，推荐系统已经从早期基于规则和协同过滤的方法，逐步发展成为融合机器学习、深度学习、强化学习以及大模型技术的复杂智能系统。然而，对于许多初学者而言，推荐系统的学习过程往往并不轻松。面对网络上大量零散的算法资料和论文内容，很多人容易陷入具体模型与公式的细节之中，却难以建立对整个推荐系统的整体认知。

回顾自己的学习与工作经历，我深刻体会到，推荐系统最重要的并不仅仅是某一个算法模型，而是对整个系统架构和业务逻辑的理解。初学阶段，我们往往关注的是模型本身，例如协同过滤、双塔模型、DeepFM、DIN、Transformer 等算法；而真正进入工业界之后才会发现，一个推荐系统的效果往往并不完全取决于某个先进模型，而是取决于召回、排序、重排、混排等多个模块之间的协同，以及对业务目标、用户体验、内容生态和工程架构的综合把控。事实上，工业级推荐系统本质上是一个复杂的系统工程。它既需要算法能力，也需要工程能力；既需要关注模型效果，也需要考虑系统稳定性；既需要追求用户价值，也需要兼顾商业价值和内容生态。许多优秀的推荐系统并非依赖某一个“神奇模型”，而是在长期迭代过程中不断优化系统链路、完善模块设计、平衡多种业务目标所形成的结果。

因此，本书希望以工业级推荐系统的视角，为读者呈现推荐系统从整体到局部、从架构到算法、从经典到前沿的发展全貌。我们将从推荐系统的整体架构出发，介绍召回、粗排、精排、重排、混排等核心模块的职责与设计思想，帮助读者建立完整的模块化认知；随后进一步深入各个模块中的经典算法与工程实践，理解这些算法为何被设计出来、解决了什么问题，以及它们在工业系统中的真实应用方式；最后，我们还将介绍推荐领域近年来的重要发展方向，包括生成式推荐、推荐大模型、强化学习推荐、多模态推荐以及 Agent 驱动的智能推荐系统等前沿技术。

本书的写作遵循“整体架构 → 模块设计 → 核心算法 → 工程实践 → 前沿趋势”的知识组织方式。我们希望读者在学习每一个算法时，不仅能够理解其数学原理和实现细节，更能够清楚地知道它在推荐系统中的位置、作用以及与其他模块之间的协同关系。通过这种方式，帮助读者建立真正完整的推荐系统知识体系，而不仅仅是记住一系列模型名称和论文结论。

本书适合以下读者阅读：

- 希望系统学习推荐系统知识的本科生、研究生以及科研人员；
- 从事推荐、搜索、广告相关工作的算法工程师与研发工程师；
- 希望深入理解推荐系统运行机制的产品经理与技术管理者；
- 对生成式推荐、推荐大模型以及强化学习推荐等前沿方向感兴趣的研究者和实践者。

推荐系统的发展仍在持续演进。从协同过滤到深度学习，从强化学习到生成式推荐，再到大模型与智能 Agent，新的技术不断推动着推荐系统向更智能、更个性化的方向发展。希望本书能够帮助读者建立扎实的推荐系统基础，形成系统化的工程思维，并在未来的学习与实践过程中持续探索这一充满活力和挑战的领域。

谨以此书献给所有热爱推荐系统、热爱人工智能技术的读者。愿你不仅学会推荐算法，更能够理解推荐系统。

刘京欣
2026 年 6 月

版权声明

© 2026 刘京欣。保留所有权利。

本书当前为草稿版本 (**Draft**), 仅供个人学习与**非商业用途**。未经作者书面许可, 禁止用于商业目的、禁止大规模分发、禁止修改后重新发布。若引用本书内容, 请注明出处。所有插图使用系统默认字体 (如宋体、微软雅黑), 仅用于内部写作。若未来正式出版, 将统一替换为可商用字体。