学习进度

本周学习任务：

学习《Recommender Systems Handbook (3rd Edition, 2022)》关于概述部分的内容。

下周学习任务：

学习Part I General Recommendation Techniques：A Comprehensive Survey of Neighborhood-Based Methods for Recommender Systems

1. 初识推荐系统

针对海量互联网信息，**推荐系统作为一种通过构建和利用用户模型来生成推荐的工具，它能够为用户提供所需的建议用于决策辅助，以缓解信息过载问题。**推荐任务可以划分为非个性化及个性化两大类，前者更适应于缺乏用户偏好等更精确信息的情形，后者则更适应于针对不同用户或群体的个性化推荐需求，相对更灵活实用。具体而言，**推荐系统的常见任务**如下：找到一些好项目、找到所有好项目、上下文注释（结合具体的上下文情景，强调用户的部分长期偏好）、推荐一系列项目（比如下一本感兴趣的书）、捆绑推荐（捆绑一组项目作为一个大项目推荐）、仅浏览（在特定的浏览会话中，推荐属于用户兴趣范围内的项目）、找到可靠的推荐者（给予用户能选择符合特定口味“推荐者”的主动性）、改善配置（用户直接向推荐系统提供自己的偏好信息）及表达自我（用户有时候可能更关心如何传播自己的观点或反馈）等。

同传统机器学习领域一样，在特定的任务情景中，想要精心挑选出合适的算法并调校出更实用的推荐模型，需要投喂大量的数据用于训练、验证及测试。而**构建推荐系统所需的数据，主要来源于项目、用户及用户与项目间的交互。**在一般情形中：项目是系统推荐的对象；用户是推荐系统的服务对象，不同用户使用系统的目的及自身表现出的行为特征可能差别很大；而存储在系统日志中的用户与项目间的交互数据，即包含可靠而又稀少的显式反馈（比如：用户点了一次外卖，给了五星好评），也包含根据用户行为隐含推断而来的隐式反馈（比如：用户完整观看了系统推荐的视频，表明了正反馈，然而用户如果没有看推荐的视频，只能说明互动的缺失，并不能简单地视为负反馈）。显而易见的是，对于整个系统而言，选择使用哪些数据以及如何使用它们，是综合考量任务情景、数据可用性及推荐算法等因素的权衡结果，更考验开发人员的经验。

此外，**推荐系统的主要应用领域**有：电子商务、娱乐推荐（推荐电影、音乐及游戏等）、中介服务（旅游、专家咨询、房地产及婚介等）、社交推荐（社交媒体内容及用户推荐）及个性化知识推送（内容个性化报纸、推荐文档及电子学习应用程序）等。

1. 推荐系统的常见方法

为了推荐符合用户需求的项目，推荐系统需要预测和比较某些项目对于用户的效用。**协同过滤（Collaborative Filtering）**是推荐系统领域较为常见的方法之一，可分为基于邻域和基于模型两类。其中，基于邻域的协同过滤方法可以根据用户行为或项目的相似性完成推荐，主要存在基于用户和基于项目两种实现方式。基于用户的协同过滤，通过分析用户间相似性的方式完成推荐，这种方法认为：如果两个用户的行为偏好越相似，那么某个用户感兴趣的项目，另一个用户也越可能更感兴趣。而基于项目的方法则关注项目间的相似性，如果用户喜欢某个项目，那么他可能也会喜欢与该项目相似的其它项目。显而易见的是，协同过滤不仅可以捕捉用户或项目间的复杂关系，还可无需显式的特征工程就能直接利用用户行为数据完成推荐。然而，当新用户或项目缺乏足够的历史数据时，此类方法可能无法提供准确的推荐，存在冷启动问题。此外，当用户或项目的数量过于庞大时，计算相似性矩阵的复杂程度显然也较高。

**基于内容（Content-Based）**也属于较为常见的推荐方法，它主要通过将用户配置文件的属性同项目属性相匹配的方式，学会推荐与用户过去喜欢的项目相类似的项目。值得注意的是，此类方法还可以推荐具有相似特征的新项目，更适应项目历史数据较为缺乏的情形。此外，由于使用了明确的项目特征，使得此类方法的推荐结果更具解释性，用户可以清晰地看到推荐的逻辑。然而，基于内容的推荐方法通常需要有效的特征表示，这无疑需要完成大量的特征工程工作来提取和选择有意义的特征，这无疑增加了系统设计和维护的复杂性。此外，由于此类方法主要根据用户的历史行为完成推荐，可能会导致推荐的局限性，使得用户无法发现潜在的新兴趣，陷入“信息茧房”。值得补充的是，基于内容的推荐方法常结合其它方法来克服各自的局限性，以获得更好的推荐性能。比如，基于语义-内容的方法（Semantic and Content-Based Methods）通过引入语义索引技术的方式，使用概念而非关键字来表示项目和用户配置文件，消除了传统关键词方法关于对象特征表达缺乏一致性及歧义的问题。

在更现实的情景中，用户的兴趣在不同的情形下可能会发生变化，而前述基于内容这种依赖于对象静态特征的方法，显然难以捕捉用户兴趣的动态变化。**上下文感知技术（Context-Aware Techniques）**则应运而生，它通过将上下文信息引入推荐过程的方式，考虑了用户当前决策的具体情景。

相较于传统的推荐方法，**基于深度神经网络**的推荐方法提供了一种更灵活且端到端学习的框架。此类方法减少了对特征工程的需求，能够更轻易地融合多种类型的数据，更好地处理大规模数据用于学习更复杂的特征表示，从而更全面地捕捉潜在的用户兴趣及项目特征。当然，基于深度神经网络的推荐方法，仍存在数据稀疏性及冷启动等问题。

1. 推荐系统的特殊方法

伴随着推荐系统的飞速发展，诞生了许多应对特定问题的特殊推荐方法。比如，**基于会话的推荐系统（Session-Based RSs）**，主要用于应对会话期间匿名用户的短期建模场景。**群体推荐系统（Group RSs）**主要用于兼顾组内多个用户的偏好模型，以组为单位提供项目推荐的情形。**人文推荐系统（People-to-People RSs）**主要用于将用户推荐给其他用户的同时，又需要保证双方兼顾互惠的情形。**推荐系统中的对抗性机器学习方法（Adversarial ML in RSs）**，则通过抵御利用机器学习机制进行对抗攻击的方式，增强模型韧性并保护用户数据，从而保证推荐服务的可靠及健壮性。同样，还存在综合利用多个不同领域数据和知识来提供个性化推荐服务的**跨域推荐系统（Cross-Domain RSs）**,此类方法存在领域知识融合及迁移学习等实现方式，其关键点在于：领域相似性分析、特征工程融合及用户-物品的跨领域建模等问题。此外，值得注意的是，近年来自然语言处理技术在推荐系统领域得到了广泛的应用。

1. 推荐系统的价值与评估

对于用户、项目供应商及系统所有者（交互中介）这三个推荐系统的主要涉众而言，推荐系统的价值是不同的。具体而言，系统如果能更多地提高用户粘性及用户保真度，是系统所用者希望看到的；如果能更多地增加项目销量并销售更多样化的项目，是项目供应商希望看到的；当然系统如果能够更好地理解用户需求，是三方涉众都希望看到的。然而，推荐系统对于不同涉众的价值不仅存在共通或不同的情况，有时候是相互竞争的。因此，评价一个推荐系统的价值，需要综合兼顾多方涉众的利益，并额外考虑多样性、新颖性及偏见与公平。

如前所述，在推荐系统的设计阶段，需要评估验证哪些方法能够更好地兼顾涉众们各自的特定需求。**推荐系统的评估方式**有：离线评估、在线评估及重点用户研究（Focused User Studies）。其中，离线评估是在相同的用户交互数据集上运行几种算法，根据一组替代和互补的评价指标来比较它们的性能。离线评估虽然能够较准确地评估各种方法完成系统定义任务的质量，但是它不能提供关于用户满意度、接受度或系统体验的任何见解。而在线评估则在不同的用户组上运行不同版本的算法，通过比较分析系统日志的方式不断调校算法，以提高系统性能。相较于离线评估，在线评估通过系统与真实用户交互的方式，更多地考虑到了用户满意度等维度。值得注意的是，这种以用户为中心的评估方式已经成为新的趋势，推荐系统的设计目标不再局限于算法作为工具的准确性，更高的准则是提供有益、愉快且个性化的体验，从而进一步提高用户留存率及满意度。然而，如果在实际应用环境中在线评估不可行或者风险太大时，可以邀请一小组用户使用推荐系统的不同版本来完成各种任务，以展开重点用户研究评估。具体而言，需要分发问卷收集“重点用户”关于他们使用系统的经验和感受等定性信息，并收集任务完成时间及错误率等系统定量信息，从而完成更为全面的评估。

显而易见的是，当用户需要系统的推荐用于决策辅助时，用户同样也可能不容易评价推荐的好坏。因此，从用户的角度出发，一个行之有效的推荐系统必须激发用户对系统的信任，系统需要具备能够很好地向用户解释推荐结果的能力。而推荐系统说服用户接受推荐的能力，往往取决于系统在展示、比较和解释项目时所支撑的特定人机交互。至此，**评估推荐系统价值的主要维度**可以归结为：准确性、用户满意度、公平性及用户对于推荐的接受度等。

1. 推荐系统的挑战

令人担忧的是，**一个不断“优化”的推荐系统，或许会进一步限制推荐项目的范围，某种程度上否定了推荐系统关于帮助用户更好地按需发现潜在可用信息的目标，造成了“信息茧房”现象。**具体而言：推荐系统过度个性化，可能会导致过于严格的信息过滤；过度拟合用户兴趣，可能会导致系统过于在乎用户的历史偏好，从而忽视了用户变化或拓展兴趣的可能性；过度强调点击率和反馈，可能会导致系统更倾向于推荐容易引起用户点击的热点项目，将限制用户接触到一些非热点的高价值信息。因此，**在评估一个推荐系统的好坏时，还需要适当地额外考虑到推荐的多样性及探索性**，并通过适时的用户反馈引导获取用户的长期兴趣变化。

直觉上，获取更多关于用户偏好的可靠信息，或许有助于完成更高质量的用户画像。然而，显式反馈这类较可靠的信息往往数量稀少且获取成本高昂，而从用户行为推断而来的隐式反馈则相对更容易获得。由于，隐式反馈仅存在积极或缺失类，并不存在明确的负面信号。故此，管理隐式反馈还需要额外考虑可用数据的偏差。此外，隐式反馈并不能完全代表来自用户的全部评价，许多推荐系统应用程序仍然向用户提供明确输入对推荐项目意见的可能性。而且，对于社会（Social）、群体（Group）及人文（People-to-People）等推荐系统而言，显式反馈仍然是核心。至此，值得讨论的是：在兼顾多方涉众需求的同时，**如何简化获取用户偏好的成本？**令人兴奋的是，伴随着便携设备及新型传感器的快速普及，将能够更便捷高效地记录用户活动的细节。由此，从更轻易获得的海量用户行为数据中挖掘出用户的个性信息，并利用这些关于个性、情绪及情感方面的信息来简化偏好获取的过程，逐渐成为了一个值得探索的现实。

事实上，在推荐系统辅助用户完成决策的过程中，系统同用户间的人机交互方式，将极大影响推荐生成的质量与风格。故此，更人性化的人机交互模式，不仅能使得推荐系统轻易地获取更可靠的用户偏好信息，还能一定程度上提高用户借鉴推荐结果完成决策的效率。问题在于，**如果系统被设计成更全面地考虑到了不同类型用户的需求，那么在人机交互的过程中关于推荐的解释将变得更为复杂，这可能导致用户难以理解、分析和评估推荐结果，反而不利于用户更好地完成决策。**因此，推荐系统需要在个性化推荐和简化解释之间寻求一个平衡点，并考虑直接简化解释、仅提供关键信息、个性化控制和用户教育引导等措施。