



图书分类号：TP183

学校代码：10615

西南石油大学

硕士学位论文

面向深度协同过滤的用户

论 文 题 目 组 构 建 与 选 择 偏 差 抑 制 方

学 科 专 业 法

软 件 工 程

研 究 方 向 推 荐 系 统

研 究 生 姓 名 彭 子 峰

学 号 202121000482

导 师 姓 名 卞 小 强

论 文 完 成 时 间：2024 年 3 月

西南石油大学研究生学位论文知识产权声明书及 学位论文版权使用授权书

本人完全了解学校有关保护知识产权的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属于西南石油大学。学校有权保留并向国家有关部门或机构送交论文的复印件和电子版。本人允许论文被查阅和借阅。学校可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。同时，本人保证，毕业后结合学位论文研究课题再撰写的文章一律注明作者单位为西南石油大学。

本学位论文属于

1、保密（），在 年解密后适用本授权书。

2、不保密（）

（请在以上相应括号内打“√”）

学位论文作者签名: _____ 指导教师签名: _____
年 月 日 年 月 日

西南石油大学研究生学位论文独创性声明

本人声明：所呈交的研究生学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，本论文不包含其他人已经发表或撰写过的研究成果，也不包含其他人为获得西南石油大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名:

年 月 日

西南石油大学研究生学位论文知识产权声明书及 学位论文版权使用授权书

本人完全了解学校有关保护知识产权的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属于西南石油大学。学校有权保留并向国家有关部门或机构送交论文的复印件和电子版。本人允许论文被查阅和借阅。学校可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。同时，本人保证，毕业后结合学位论文研究课题再撰写的文章一律注明作者单位为西南石油大学。

本学位论文属于

1、保密（），在 年解密后适用本授权书。

2、不保密（）

（请在以上相应括号内打“√”）

学位论文作者签名: _____ 指导教师签名: _____
年 月 日 年 月 日

西南石油大学研究生学位论文独创性声明

本人声明：所呈交的研究生学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，本论文不包含其他人已经发表或撰写过的研究成果，也不包含其他人为获得西南石油大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名:

年 月 日

摘要

协同过滤 (Collaborative Filtering, CF) 是推荐系统 (Recommender System, RS) 中最重要的技术之一，CF 利用用户行为数据来学习用户直接的相似性，并不依赖于项目或用户的自身信息。目前，已有多种协同过滤模型被提出，其目的都是为了改进推荐的效果，进而提升用户体验。但这些模型很少考虑到了用户行为数据中存在的选择偏差以及用户和项目间的适配关系，从而导致模型学习到了不完整的用户偏好，降低了推荐的精度。针对这些问题，本文提出了两个协同过滤模型，具体内容如下：

(1) 为了抑制选择偏差导致的协同过滤模型学习到不完整的用户偏好，提出了一种基于用户评分习惯的无趣项识别及填充方法 (Uninteresting Items Identify and Multiple Low-rating Filling Based on Users Rating Habits, UIIMLF-URH)，结合了显式反馈和隐式反馈的特点。该方法分为无趣项识别和无趣项填充两个阶段。在无趣项识别阶段，利用用户的显式反馈记录挖掘用户的评分习惯，结合隐式反馈数据使用加权矩阵分解模型预测使用前偏好矩阵，识别出用户的无趣项。在无趣项填充阶段，根据显示反馈数据推测项目的质量，将无趣项划分为高质量和低质量两部分，分别以不同的低评分填充用户对其的评分。两个真实的电影数据集上的实验结果表明，UIIMLF-URH 在无趣项的识别和填充阶段都优于已有方法，显著改善了协同过滤算法在 Top-K 推荐中的性能。

(2) 为了解决传统 RS 模型中没有建立用户和项目间的合理关联，导致未能做出最合适的推荐的问题，提出了一种基于建立理想用户组的深度协同过滤方法 (Ideal User Group-based Neural Collaborative Filtering, IUG-CF)。该方法分为理想用户组 (Ideal User Group, IUG) 构建、对用户匹配 IUG 进行推荐。在 IUG 构建阶段，将用户的人口统计信息与用户的评分行为结合，为每个项目构建出一个最适合推荐的用户群体。在推荐阶段，协同过滤与深度学习相结合，人口统计信息被映射为向量，计算 IUG 与用户的人口统计信息向量间的距离，生成 Top-K 推荐列表。在两个真实的电影数据集以及一个短视频数据集上的实验结果表明，IUG-CF 是一个能够有效提升推荐性能的模型。

关键词：协同过滤；用户特征；无趣项；深度学习

Abstract

Collaborative Filtering (CF) is one of the most important techniques in Recommender Systems (RS). CF utilizes user behavior data to learn direct similarities between users, without relying on the intrinsic information of items or users. Currently, various CF models have been proposed, all aiming to improve recommendation effectiveness and enhance user experience. However, these models seldom consider the selection bias in user behavior data and the compatibility relationship between users and items, leading to incomplete learning of user preferences and reduced recommendation accuracy. To address these issues, this thesis proposes two collaborative filtering models. The specific research content is as follows:

(1) To address the issue of incomplete user preference learning caused by selection bias in collaborative filtering models, a method called Uninteresting Items Identify and Multiple Low-rating Filling Based on Users Rating Habits (UIIMLF-URH) is proposed, which combines the characteristics of explicit and implicit feedback. The method consists of two stages: uninteresting item identification and uninteresting item filling. In the uninteresting item identification stage, users' rating habits are mined from explicit feedback records, and a weighted matrix factorization model is used with implicit feedback data to predict the pre-use preference matrix, identifying users' uninteresting items. In the uninteresting item filling stage, the quality of items is inferred from explicit feedback data, and uninteresting items are divided into high-quality and low-quality categories. Different low ratings are then used to fill users' ratings for these items. Experimental results on two real movie datasets demonstrate that UIIMLF-URH outperforms existing methods in both uninteresting item identification and filling stages, significantly improving the performance of collaborative filtering algorithms in top-N recommendations.

(2) To address the issue of traditional RS models failing to establish reasonable associations between users and items, resulting in suboptimal recommendations, a method called Ideal User Group-based Neural Collaborative Filtering (IUG-CF) is proposed. The method consists of two stages: Ideal User Group (IUG) construction and recommendation based on matching users to IUGs. In the IUG construction stage, demographic information and user rating behavior are combined to construct the most suitable user groups for each item. In the recommendation stage, collaborative filtering is combined with deep learning. Demographic information is mapped to vectors, and the distance between the demographic information vectors of IUGs and users is calculated to generate top-N recommendation lists. Experimental results on two real movie datasets and one short video dataset demonstrate that IUG-CF is an effective model for improving recommendation performance.

Key words: Collaborative Filtering; User Features; Uninteresting Items; Deep Learning

目 录

第1章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 无趣项的识别和填充	2
1.2.2 基于人口统计学的推荐	2
1.2.3 基于深度学习的推荐系统	3
1.3 研究内容	4
1.3.1 基于用户评分习惯识别并填充无趣项的选择偏差抑制方法	5
1.3.2 基于建立理想用户组的深度协同过滤方法	5
1.4 论文结构	5
第2章 相关工作	7
2.1 推荐系统	7
2.1.1 推荐系统概述	7
2.1.2 基于内容的推荐系统	8
2.1.3 协同过滤推荐系统	9
2.1.4 评估方法	10
2.1.5 评估指标	10
2.2 推荐系统相关技术	11
2.2.1 矩阵分解	11
2.2.2 无趣项相关技术	12
2.3 深度学习	13
2.3.1 理论基础	14
2.3.2 深度协同过滤	15
2.4 本章小节	16
第3章 基于用户评分习惯识别并填充无趣项的选择偏差抑制方法	17
3.1 引言	17
3.2 算法设计	18
3.2.1 使用前偏好建模算法	18
3.2.2 无趣项的识别	20
3.2.3 无趣项的填充	21
3.3 实验结果与分析	22
3.3.1 数据集	23

3.3.2 评价指标.....	23
3.3.3 参数实验.....	24
3.3.4 使用前偏好建模模型.....	24
3.3.5 假设验证.....	26
3.3.6 填充策略.....	26
3.3.7 结合深度协同过滤.....	28
3.4 本章小结	29
第 4 章 基于建立理想用户组的深度协同过滤方法	32
4.1 引言	32
4.2 算法设计	34
4.2.1 问题陈述.....	34
4.2.2 理想用户组构建.....	36
4.2.3 用户及项目建模.....	37
4.2.4 偏好评分计算.....	37
4.2.5 模型训练.....	38
4.3 实验结果与分析	38
4.3.1 数据集.....	39
4.3.2 评估指标.....	39
4.3.3 实验细节.....	40
4.3.4 特性验证.....	40
4.3.5 模型对比.....	43
4.3.6 填充实验.....	46
4.4 本章小结	48
第 5 章 总结与展望	49
5.1 总结	49
5.2 展望	49
致 谢	51
参考文献	52
攻读硕士学位期间发表的论文及科研成果	58

第 1 章 绪论

1.1 研究背景及意义

随着科学技术的不断革新，互联网技术也在不断发展。在全面进入大数据时代的当下，日常生活中产生的数据呈指数级增长。信息爆炸带来了信息的多样性的同时，也带来了人类难以消化海量数据的弊端，信息过载以及选择困难成为了大多互联网用户面临的两道难关，使个人和机构都更期望获取精简、定制化的信息。推荐系统作为帮助用户获取精简、期望的信息的重要工具，在工业界得到了广泛运用。当今，各大互联网公司纷纷推出了自己的推荐产品，运用于电商、社交、影视等多个领域，高效的推荐使用户数量和营收都快速增长。在计算机科学技术不断进化的同时，推荐系统也逐步引入了新的技术，推动推荐的思路和原理发生了翻天覆地的变化。推荐系统被分为基于内容的推荐系统、基于人口统计学的推荐系统、协同过滤和混合推荐系统^[1-2]。

协同过滤 (Collaborative Filtering, CF) 是推荐系统中最流行的技术之一^[3-4]，它利用用户行为数据来发现用户之间的相似性，根据用户特点来提供个性化的推荐。协同过滤不依赖于信息本身，而是依赖于用户的行为数据，例如点击、转发、评分等，这些数据成为协同过滤的重要信息源。然而，这一特点也给协同过滤带来了局限性。用户在评分时会更倾向于给感兴趣的项目评分，而忽略掉不感兴趣的项目，这就导致了用户评分数据的非随机缺失^[5-7]。这种行为使得协同过滤模型学习到的用户偏好不完整，从而降低了推荐精度^[8-9]。

“物以类聚，人以群分”，反映了一种基本的社会现象和认知规律，在推荐系统中也有类似的运用。“物以类聚”意味着相似的项目通过其属性、特征等关联在一起，推荐系统根据项目的相似性可以向用户推荐类似于他们之前喜欢的项目。而“人以群分”说明了用户有各自的喜好，他们倾向于和相似的人聚在一起，因此相似的用户会表现出相似的喜好^[1,10-12]。在推荐系统中，将用户归类时，用户的人口统计学信息是最常用的归类标准，通过年龄、性别等用户特征将用户划分为若干用户组。然而，仅依赖人口统计信息归类用户可能导致带偏见的推荐和对某些用户群的刻板印象，因此归类用户时考虑用户行为、偏好和反馈等其他因素也至关重要。

鉴于现有的方法的不足，基于以上背景，本文主要内容基于对协同过滤的方法的改进，提出两个方法从不同角度探索用户的完整偏好，以提升推荐性能。

1.2 国内外研究现状

1.2.1 无趣项的识别和填充

使用前偏好指的是用户在接触某个项目之前，对该项目的喜好程度，通常由项目的外部特征决定，如电影的演员、导演等。使用前偏好是隐性的，而使用后偏好一般指用户的评分，能够在数据集中明确反馈给模型。为了获取使用前偏好，可以通过分析用户行为、构建用户画像、进行数据挖掘等方式。

用户的使用前偏好决定了他们选择某个项目的可能性，挖掘使用前偏好可以帮助推荐系统模型更好地理解用户，从而提供更加符合用户兴趣的推荐结果。

Hwang 等人提出了“使用前偏好”的概念^[5]，用以表示用户在选择一个项目之前对该项目的兴趣程度。基于使用前偏好的概念，Hwang 等人提出了 Zero-Injection 方法，该方法通过加权矩阵分解^[13]来计算使用前偏好。在该方法中，使用前偏好最低的部分项目被视是用户的无趣项，对应的缺失评分会被填充为零分。这一方法缓解了数据稀疏问题，防止了无趣项被推荐，且易于扩展^[14-17]。在上述方法中，加权矩阵分解权重分配是该方法面临的挑战之一。Hu 等人认为，用户评分越高，置信度越高，因此把正样本的显式评分作为确定权重的依据^[9]；Pan 等人则认为，用户评分的数量越多，其缺失样本是负样本的可能性越大，因此将评分数量作为确定权重的依据^[18]；Shi 等人结合了用户活跃度、项目流行度和时间因素，提出了一种非均匀加权策略^[19]。

填充无趣项有两个主要目的：首先，它可以对用户曾经购买但未评分的项目进行评分补全，从而使模型能够更全面地学习用户的偏好。其次，对于用户不感兴趣且未购买过的项目，可以用低分填充，以确保这些项目不会出现在推荐列表中。此外，这种填充操作还可以增加模型的训练数据，降低数据集的稀疏性。

传统的数据填充方案的代表方法有 PureSVD^[20]以及 AllRank^[6]，它们分别用 0 和 2 分填充所有缺失项。然而，这种填充所有缺失样本的方法忽略了其中少量的正样本。此外，0 分低于合理评分的最低分 1 分，这夸大了用户的负面偏好。近年来，一些方法沿用了使用前偏好的概念，并提出了新的数值填充方案。Lee 等人将使用前偏好与深度学习相结合，利用全连接网络计算填充值，并将单个填充值扩展为多个^[21]。Bae 等人计算皮尔逊相关系数确定每个用户的邻居，填充无趣项后，根据用户的邻居信息对填充的评分进行聚合及更新^[22]。

1.2.2 基于人口统计学的推荐

基于人口统计学的推荐 (Demographic Recommender Systems, DRS) 是一种传统推荐模型，它将用户的人口统计信息，如年龄、性别、职业等，视为个性化推荐的重要因素。最早关于基于人口统计信息建模用户的兴趣的概念由 Grudy 于 1979 年提出^[10]，这为探索人口统计在推荐系统中的作用奠定了基础。在 1990 年代，DRS 变得更加普遍，比如

Lifestyle Finder^[11], 它利用用户的人口统计数据创建用户档案并推荐网页。Pazzani 提出了一种利用用户主页中的人口统计信息计算用户相似度的模型^[12], 这该模型激发了各种用户档案之间相似度计算指标的发展^[23-24]。

21 世纪头十年, 研究人员尝试将人口统计信息整合到混合推荐系统中^[25-26]。近年来, 人口统计信息在电子商务^[27]和社交平台^[28-29]的推荐算法中越来越被广泛使用。此外, 研究人员还探索了将人口统计信息整合到推荐系统中的各种方法。例如, Tahmasebi 等人^[30]结合了余弦相似度和人口统计相似度来增强用户的邻域集合, 而 Yassine 等人^[31]提出了一种基于物品特征选择的模型, 该模型整合了人口统计信息以对用户进行分割。Bobadilla 等人^[32]展示了提取人口统计信息的特征提取的重要性, 而 Liu 等人^[33]使用评分记录推断用户的人口统计信息, 从而缓解了获取此类信息的困难。

此外, 群体推荐系统专注于向一群人推荐项目^[34]。解释推荐原因和聚合策略是群体推荐领域中的关键子任务^[35-36]。人口统计信息在这些子任务中承担了重要作用, 为用户提供个性化建议^[37]。

冷启动问题是推荐系统中一个众所周知的问题。许多研究表明, 在解决冷启动问题时, DRS 是一个常见的选择^[30,38-39]。DRS 中常用的方法包括计算新用户与现有用户之间的相似度^[30,40]和共同聚类^[41]。

传统的 DRS 主要集中在基于人口统计信息对用户进行分组, 并探索不同相似度计算指标如何提高推荐性能。在以往工作的基础上, 本文第四章提出的算法将人口统计属性映射到高维空间, 并将每个物品与由特定人口统计属性组合而成的理想人口统计联系起来。

1.2.3 基于深度学习的推荐系统

近年来, 基于深度学习的推荐系统因其出色的性能而备受关注^[42-44]。深度学习算法被用于学习用户和项目之间复杂的映射关系, 并通过潜在因子表示和匹配函数实现。

目前国内外研究主要集中在以下几个方面:

- **混合推荐模型:** 将深度学习与传统推荐算法(如协同过滤、矩阵分解等)结合, 提升推荐效果。例如, 腾讯的 DIEN 模型结合了深度兴趣网络与注意力机制, 提高了推荐准确率。

- **图神经网络(GNN):** 利用图神经网络处理用户和项目之间复杂的关系, 增强推荐系统的性能。阿里巴巴的 PinSage 模型就是一个典型应用, 通过图卷积网络在大规模数据集上实现高效推荐。

- **自监督学习:** 利用自监督学习技术从未标记数据中学习特征, 缓解数据稀疏性问题。例如, 字节跳动的研究团队提出了基于对比学习的方法, 显著提升了推荐系统的性能。

- **用户行为建模:** 通过深度学习技术对用户的长短期行为进行建模, 提高推荐的个

个性化程度。京东的研究团队在这一领域做出了重要贡献，提出了多层次注意力机制来捕捉用户的不同兴趣。

NeuMF^[45]是一个广泛使用的深度协议过滤模型，它结合了 MF 和 MLP 来捕获用户和物品之间的线性和非线性交互作用，通过合并这两种方法，它能够从给定数据中学习任意函数。许多研究都围绕着 NeuMF 模型展开，如 NHF^[46]将项目和用户的辅助信息添加到模型以更深入地了解项目和用户之间的交互。另一方面，HybridNNMF^[47]将全局多层感知器和局部残差网络结合起来，有效地建模了用户-物品之间的交互。在基于深度学习的推荐模型中，通常继承了注意力机制，例如 LRML^[48]采用了基于记忆的注意网络来表示潜在的用户-关系，以及 IFDNAMF^[49]利用注意力机制识别预测结果中不同交叉特征的重要性。一些模型专注于探索和利用物品和用户特征，如，CFFNN^[50]利用交叉特征融合网络通过结合用户特征和物品特征来增强协同过滤。另一方面，Wide&Deep 模型^[51]以“Wide”和“Deep”两个组件结合线性模型和深度神经网络，学习层次化的特征表示。最后，CoupledCF^[52]模型同时学习了用户和物品之间的显式和隐式耦合关系，涉及属性和深度特征之间的内在关系。

基于深度学习的推荐模型在利用人口统计信息方面也十分普遍。举例来说，前述的四个模型，即 Wide&Deep、NHF、IFDNAMF 和 CoupleCF，都采用了人口统计信息来帮助模型更好地学习用户特征。通过整合人口统计信息，基于深度学习的推荐系统能够更好地理解用户需求，进而提升用户体验和推荐性能。

未来，基于深度学习的推荐系统研究将朝着以下几个方向发展：跨域推荐：利用不同领域的用户行为数据，提高推荐系统的泛化能力。隐私保护：在推荐过程中保护用户隐私，平衡个性化推荐与数据安全。在线学习：实时更新推荐模型，提升推荐的时效性和准确性。多模态推荐：结合文本、图像、视频等多模态数据，提高推荐系统的全面性和精确度。

1.3 研究内容

本节介绍了本文涉及的两个主要工作。首先，本文的第一个工作引入了“使用前偏好”概念，从用户的历史评分记录和项目的被评分记录中挖掘用户的评分习惯和项目的质量。在此基础上，提出了结合用户评分习惯的无趣项识别方法和结合项目质量的无趣项填充方法。其次，第二个工作提出了“理想用户组”的概念，通过将项目与用户通过用户的人口统计信息关联起来，建立了一种新的联系。并结合协同过滤方法与神经网络技术，提出了一种新颖的深度协同过滤推荐模型。这一模型在推荐系统中的应用，为提高推荐效果和用户满意度带来了新的可能性。

1.3.1 基于用户评分习惯识别并填充无趣项的选择偏差抑制方法

数据稀疏性和显式反馈数据中的选择偏差对协同过滤算法的推荐性能产生了负面影响。然而，现有方法通常只采用权重分配和统一的低值填充来缓解这些问题，却忽略了用户的评分习惯和项目质量差异。针对这一问题，本工作的主要贡献如下：

- (1) 提出了一种基于用户评分习惯的方法，充分结合了显式反馈和隐式反馈的优点，并采用多种低值填充策略。
- (2) 该方法综合考虑了评分数量和用户评分习惯的挖掘，使得显式反馈和隐式反馈的优势得以充分发挥。
- (3) 通过在公开数据集上的实验以及与其他方法的对比，验证了本方法的有效性和优越性。

1.3.2 基于建立理想用户组的深度协同过滤方法

大部分使用用户的人口统计信息的推荐系统仅单独考虑了用户侧特征，然而项目和用户之间存在着一种适配性，需要综合项目和用户的特征来进行更适合的推荐。面对更好地探索项目和用户之间的关系的挑战，本工作的主要内容和贡献如下：

- (1) 在人口统计信息推荐领域提出了理想用户组 (Ideal User Group, IUG) 的概念。通过使用理想用户组关联项目，提供了一个新的视角描述项目，并且避免了冷启动问题。
- (2) 提出了一种新的深度协同过滤模型，它结合 IUG 提高了推荐性能。该模型考虑了项目在不同人群中的流行度，并向具有与项目 IUG 相似的用户推荐项目。
- (3) 介绍了一种建立用户和项目之间联系的新技术，通过拆分和合并人口统计信息，使得推荐系统能够充分利用人口统计信息，提供更好的推荐。

1.4 论文结构

本文结构总共分为五部分，具体如下：

第一章：绪论。本章介绍了协同过滤推荐系统的研究背景及意义，以及目前国内外关于无趣项识别及填充、基于人口统计学的推荐、基于深度学习的推荐系统的研究现状，最后阐述了本文研究的内容和结构。

第二章：相关工作。本章对推荐系统概念和分类进行了概述，详细介绍了推荐系统任务的评估方法和评估指标，以及传统推荐系统和基于深度学习的推荐系统的相关基础理论。

第三章：基于用户评分习惯识别并填充无趣项的选择偏差抑制方法。本章提出了一个基于用户评分习惯识别并填充无趣项的选择偏差抑制方法。该方法通过挖掘用户的评分习惯，更准确地识别用户的无趣项，并挖掘项目的质量，更准确、合理地填充无趣项，以此完善用户的偏好，减少选择偏差对协同过滤模型带来的负面影响。该方法在两个电

影数据集上的实验结果都取得了优秀的效果。

第四章：基于建立理想用户组的深度协同过滤方法。本章提出了一个基于建立理想用户组的深度协同过滤方法。该方法通过用户的人口统计学信息来建立用户与项目间的映射，为每个项目确定一个“理想用户组”，然后将其与深度协同过滤结合，提出了新的推荐模型 IUG-CF，该模型在两个电影数据集及一个短视频数据集上的实验展示了其优秀的推荐效果。此外，本章结合第三章方法进行了实验，证明了第三章算法的适用性。

第五章：总结与展望。本章总结了本文的研究内容，以及简单陈述了对未来研究方向的展望。

第 2 章 相关工作

本章首先介绍了推荐系统，随后介绍了其分支：协同过滤推荐系统、基于人口统计学的推荐系统以及基于深度学习的推荐系统。随后，详细介绍了对无趣项的识别和填充的相关理论。最后，阐述了深度学习的理论基础和深度协同过滤相关理论。

2.1 推荐系统

2.1.1 推荐系统概述

推荐系统是人工智能领域中的经典运用，也是一种信息过滤系统，旨在预测和提供用户可能感兴趣的项目，从而提升用户体验并增加营收。根据不同的技术和方法，推荐系统可以分为多种类型，其中包括基于内容的推荐系统和协同过滤推荐系统。

基于内容的推荐系统包括基于项目内容的推荐和基于人口统计信息的推荐。这些系统以“物以类聚，人以群分”的基本思想为基础，通过分析项目标签和用户的人口统计学信息，将项目和用户进行分类，从而向用户推荐与其喜好相似或与类似用户喜好相符的项目。

协同过滤推荐系统则分为基于用户的协同过滤和基于项目的协同过滤两种类型。这种推荐系统利用用户行为和偏好发现用户之间的相似性，以预测用户对项目的偏好。其优势在于无需预先了解项目或用户的基本信息，仅通过用户的历史行为即可发现其兴趣，因此具有更广泛的应用场景。

推荐系统的任务包括但不仅限于 Top-K 推荐和评分预测两种。

(1) Top-K 推荐：该系统旨在向用户提供最相关且最具吸引力的项目列表。从所有可供推荐的项目中，选取推荐系统模型预测出的用户最有可能喜欢的前 k 个项目，然后推荐给用户。Top-K 推荐通常用于隐式数据数据集，这类数据集中的用户行为不包括对项目的评分，仅包含用户的点赞、收藏等行为，交互的取值一般为 0 或 1。因此，这类任务常用的评估指标包括命中率 (Hit Ratio, HR)、归一化折现累积增益 (Normalized Discounted Cumulative Gain, NDCG)、平均倒数排名 (Mean Reciprocal Rank, MRR) 等。

(2) 评分预测：该系统着眼于预测用户对特定项目的评分。评分预测需要包含显式交互数据的数据集，训练数据为用户对项目准确的历史评分记录。该类任务最常用的方法是矩阵分解，常用的评估指标包括平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Square Error, RMSE)。

2.1.2 基于内容的推荐系统

基于内容的推荐系统 (Content-Based Recommender Systems) 是一类广泛应用于推荐技术领域的方法，其主要思想是通过分析项目的内容特征来进行个性化推荐。不同于协同过滤推荐系统，基于内容的方法无需依赖用户之间的相似性，而是直接利用项目的描述信息和用户的历史行为数据来预测用户对未见项目的偏好。

在基于内容的推荐系统中，每个项目通常用一组特征向量来表示，这些特征可以是显式的（如文本、标签、属性）或隐式的（如通过自然语言处理技术提取的特征）。系统首先对用户的历史行为进行分析，从用户已评价或已互动的项目中提取出用户的偏好特征向量。接着，通过计算目标项目与用户偏好特征向量之间的相似性，系统推荐那些与用户偏好最为匹配的项目。

首先，特征提取和表示是基于内容的推荐系统的关键步骤。对于文本类项目，如新闻文章或科学论文，自然语言处理技术被广泛应用以提取文本特征，包括关键词提取、主题建模等。然而，对于非文本类项目，如图片、音频和视频，特征提取则更加复杂，通常需要借助于计算机视觉和音频处理技术。

其次，基于内容的推荐系统容易产生“过拟合”问题，即过度依赖用户的历史偏好，导致推荐结果缺乏多样性和新颖性。为缓解这一问题，研究者们提出了多种策略，例如引入探索性推荐，通过适度推荐用户未曾接触过的新类别项目，提升推荐的多样性。

此外，基于内容的方法还存在难以捕捉隐含兴趣的问题。用户的偏好往往是多方面且隐性的，单纯依赖显性特征可能无法全面反映用户的真实兴趣。为此，混合推荐系统逐渐成为一种有效的解决方案，通过结合协同过滤和基于内容的方法，综合利用用户行为和项目特征，提升推荐的准确性和用户满意度。

基于内容的推荐系统可以分为基于项目内容的推荐和基于人口统计信息的推荐，其主要实现思路是根据项目的标签信息或用户的人口统计信息来计算项目间或用户间的相似度。以基于人口统计信息的推荐系统为例，首先需要收集用户的人口统计信息，如年龄、性别、职业等。然后把这些信息转换为可比较的形式，独热编码 (One-Hot Encoding) 是最常见的方法。独热编码将每个人口统计特征都创建一个唯一的二进制特征向量，其中每个维度代表一个可能的取值。如果该用户具有该特征值，则相应的维度为 1，否则为 0。然而，随着推荐系统的发展，独热编码存在一些缺点，如高维稀疏和缺乏语义信息等。嵌入技术 (Embedding) 应运而生，它可以将每个分类特征映射到一个低维稠密的实数向量空间中。这种方法降低了向量维度和稀疏性，同时也具备学习特征之间内在关系和语义信息的能力。因此，在新型的推荐系统中，Embedding 技术也常用于对特征进行编码。

基于编码技术，得以对用户的相似度进行计算，假设用户共有 n 个维度的特征，分别用向量 $\mathbf{u} = (u_1, u_2, \dots, u_n)$ 和表示 $\mathbf{v} = (v_1, v_2, \dots, v_n)$ ，常用的计算方法有欧氏距离

(Euclidean Distance), 和余弦相似度 (Cosine Similarity), 计算过程如下:

$$E(\mathbf{u}, \mathbf{v}) = \sqrt{\sum_{i=1}^n (u_i - v_i)^2} \quad (2-1)$$

$$\text{Cos}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|} \quad (2-2)$$

欧氏距离计算了两个用户在各个维度上的差值，并对其平方后求和再取平方根，欧氏距离越小表示两个用户越相似。

余弦相似度是通过计算用户向量间的点积和向量范数之积的比值来衡量用户间的相似度，其取值在-1 到 1 之间，值越接近 1 表示两个用户越相似。

2.1.3 协同过滤推荐系统

协同过滤利用用户的历史行为数据来发现用户或项目之间的相似性，从而进行个性化推荐。基于协同过滤的推荐方法不依赖于项目或用户的自身信息，而是通过分析用户与项目之间的交互行为来进行推荐。协同过滤一般被分为基于用户的协同过滤 (User-based Collaborative Filtering) 和基于项目的协同过滤 (Item-based Collaborative Filtering)，其区别在于根据用户的评分信息挖掘用户间的相似性还是根据项目的被评分记录来挖掘项目间的相似性。

基于用户的协同过滤方法利用用户行为相似性进行推荐。具体而言，对于目标用户，系统首先识别出与其偏好相似的用户群体，然后根据这些相似用户的行为数据推荐项目。这种方法的优势在于能够捕捉用户之间潜在的偏好关联，但其计算复杂度较高，且在用户数目庞大时容易产生“冷启动”问题和数据稀疏性问题。

基于项目的协同过滤方法则通过计算项目之间的相似性来实现推荐。对于目标用户，系统会根据其历史评分记录，推荐与其已评价高分项目相似的其他项目。该方法在处理大规模数据时较为高效，并且在数据稀疏的情况下表现相对稳定。然而，其推荐结果容易受限于已有项目的评价历史，难以推荐全新项目。

要根据评分记录计算两个用户之间的相似度，常用的是皮尔逊相关系数 (Pearson Correlation Coefficient)，公式如下：

$$\rho(u, v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - \bar{r}_v)^2}} \quad (2-3)$$

皮尔逊系数衡量了两个变量之间线性关系强度，其取值在-1 和 1 之间，当系数为 1 时表示完全正相关，当系数为-1 时表示完全负相关，而当系数为 0 时表示没有线性关系。计算出相似度后，可以进行用户 u 对项目 i 的评分 $p_{u,i}$ 预测，公式如下：

$$p_{u,i} = \bar{r}_u + \frac{\sum_{u' \in N} \rho(u, u')(r_{u',i} - \bar{r}_{u'})}{\sum_{u' \in N} |\rho(u, u')|} \quad (2-4)$$

其中， \bar{r}_u 表示用户 u 对所有已评分项目的平均评分， N 表示与用户 u 相似度高的用户集合， $\rho(u, u')$ 表示两个用户间的相似度， $\bar{r}'_{u'}$ 表示用户 u' 对所有已评分项目的平均分。

2.1.4 评估方法

推荐任务可分为 Top-K 推荐和评分预测，Top-K 推荐计算用户对项目的偏好评分，以此排序产生一个长度为 K 的推荐列表，项目在推荐列表中的位置越靠前，表示模型认为该项目更有可能被该用户喜欢，适用于任何有用户-项目间交互记录的数据集。评分预测则是输出用户对项目的评分，适用于具有显式评分数据的数据集。

传统的机器学习在评估模型时常用的方法有 K 折交叉验证法 (K-fold Cross-validation) 和留一法 (Leave-one-out)，这两种评估方法在推荐系统领域中任适用。K 折交叉验证法适用于评分预测和 Top-K 任务，该评估方法将数据集按百分比分为训练集和测试集两部分，进行 K 次实验取结果的平均值，并保证 K 次实验中的测试集无交集。

留一法用于 Top-K 推荐任务，评估时关注测试项在推荐列表中的顺序。不同的是，推荐系统任务中的留一法有多个分支，针对于每个用户，在用户打过分的项目中选取一个作为测试样本，其他都作为训练样本。将历史行为按时间升序排列后取最新的一次交互项目作为测试样本的时间排序 (Temporal Ordering, TO) 留一法，该方法更适用于时序推荐模型；而在用户历史行为中随机选取一个项目作为测试样本叫做随机排序 (Random Ordering, RO) 留一法。Top-K 推荐任务中，推荐列表的其他备选项的选择的方法不同，可以将评估方法分为全排序 (Full Ranking) 和随机采样两种。全排序指将数据集中的所有未交互过的项目都作为备选项并计算用户对其的偏好分数，而随机采样则是随机选取 N 个未交互过的项目作为备选项，这样做好处是在较大的数据集评估时节约时间成本。

2.1.5 评估指标

根据不同的推荐任务，评估指标的选择也各不相同。

采用 K-fold 交叉验证法的 Top-K 推荐中，常用的评估指标有：精确率 (Precision, P)、召回率 (Recall, R) 和归一化折损累计增益 (Normalize Discounted Cumulative Gain, NDCG)。

(1) P:

$$P = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (2-5)$$

(2) R:

$$R = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (2-6)$$

(3) NDCG:

$$DCG@N = \sum_{i=1}^N \frac{2^{rel_i} - 1}{log_2(i + 1)} \quad (2-7)$$

$$NDCG@N = \frac{DCG@N}{IDCG@N} \quad (2-8)$$

其中, $R(u)$ 是给用户 u 推荐的项目集合, 长度为 N ; $T(u)$ 是测试集中用户 u 的相关项; rel_i 是第 i 个推荐项目的相关性, 取值为 0 或 1, 仅当第 i 个项目在集合 $T(u)$ 中时取值为 1, 用户的相关项被推荐时, 位置越靠前, DCG 的值越大; $NDCG@N$ 是理想状态下的 $DCG@N$ 值。

采用留一法的 Top-K 推荐中, 常用的评估指标有命中率 (Hit Ratio, HR) 和 NDCG, 和 K 折法相比, NCGD 的计算公式有一定简化:

(1) HR:

$$HR@k = \frac{\#hits@k}{\#users} \quad (2-9)$$

(2) NDCG:

$$NDCG@k = \frac{1}{\#users} \sum_{l=1}^{\#users} \frac{1}{\log_2(p_{l,k} + 1)} \quad (2-10)$$

评分预测需要包含显式交互数据的数据集, 训练数据为用户对项目准确的历史评分记录。该类任务最常用的方法是矩阵分解, 常用的评估指标包括平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Square Error, RMSE), 公式如下:

(1) MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2-11)$$

(2) RMSE:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2-12)$$

其中 y_i 和 \hat{y}_i 分别表示项目 i 的真是评分和预测评分。

2.2 推荐系统相关技术

2.2.1 矩阵分解

矩阵分解是传统推荐系统中进行评分预测的经典技术, 它将稀疏的评分矩阵 $\mathbf{R} \in \mathbb{R}^{M \times N}$ 分解为 $\mathbf{P} \in \mathbb{R}^{M \times N}$ 和 $\mathbf{Q} \in \mathbb{R}^{K \times N}$ 两个低秩矩阵, 将用户和项目映射到 k 维空间中, 其向量被称为隐特征向量, 其优化目标可以表示为:

$$\arg \min_{\theta} L(\theta) = \frac{1}{2} \underbrace{\arg \min_{p_u, q_i}}_{p_u, q_i} \sum_{u,i} I_{ui} \left[(r_{ui} - p_u^T q_i)^2 \right] + \lambda (\|p_u\|_F^2 + \|q_i\|_F^2) \quad (2-13)$$

其中, p_u 和 q_i 分别为矩阵 \mathbf{P} 和 \mathbf{Q} 的隐特征向量, r_{ui} 为真实评分, θ 表示矩阵分解的参数, λ 表示正则项的系数。

由于有些用户会给出偏高的评分, 有些项目也会收到恶意的过低评分, 于是 Bias-SVD 算法将矩阵分解改进至考虑偏置的影响, 用户的偏置是独立于项目特征的因素, 表示了用户的打分习惯, 因为用户的性格各不相同, 有喜欢打低分的批判型用户, 也有倾

向于打高分的乐观用户。同理，项目的偏置也是独立于用户兴趣的因素，例如一部好电影获得的评分会普遍偏高，捕获了偏置的影响后，Bias-SVD 的优化目标为：

$$\begin{aligned} L(\theta) = & \underbrace{\arg \min_{P_u, q_i}}_{P_u, q_i} \sum_{P_u, q_i} (r_{ui} - \mu - b_u - b_i - p_u^T q_i)^2 \\ & + \lambda (\|p_u\|_2^2 + \|q_i\|_2^2 + \|b_u\|_2^2 + \|b_i\|_2^2) \end{aligned} \quad (2-14)$$

其中， b_u 和 b_i 分别表示用户和项目的偏置， μ 表示训练集中所有评分记录的均值，表示了训练数据整体的评分情况。

Bias-SVD 通过挖掘显式反馈数据优化了矩阵分解，显式反馈的优点就在于通过准确的评分反映了不同程度的喜好，但是当数据为隐式反馈时就显得乏力。在此基础上，SVD++ 算法进行了进一步改进，考虑了用户的隐式反馈，探索了项目和项目之间的关联，例如对复仇者联盟有高评价的观众更可能会对复仇者联盟 2 有高评价，因为这两部电影之间是续集的关系。其次，用户或许看过一部电影，但是没有进行评分，这也就造成了显式数据的反馈，而隐式反馈通过“观看”这一行为，也可以捕获用户的喜好。考虑隐式反馈的 SVD++ 的优化目标如下：

$$\begin{aligned} L(\theta) = & \underbrace{\arg \min_{P_u, q_i}}_{P_u, q_i} \sum_{u, j} \left(r_{ui} - \mu - b_u - b_i - p_u^T q_i - p_u^T \frac{1}{\sqrt{|N(u)|}} \sum_{j \in N(u)} y_j \right)^2 \\ & + \lambda (\|p_u\|_2^2 + \|q_i\|_2^2 + \|b_u\|_2^2 + \|b_i\|_2^2 + \sum_{j \in N(u)} \|y_j\|_2^2) \end{aligned} \quad (2-15)$$

其中， $N(u)$ 表示用户 u 提供了隐式反馈的项目集合， $\frac{1}{\sqrt{|N(u)|}}$ 是为了消除不同 $|N(u)|$ 引起的差异。

2.2.2 无趣项相关技术

使用前偏好 (Pre-use Preference) 是指用户在接触一个项目之前，对该项目的喜爱程度，其通常由项目的外部特征决定，如电影的演员、导演。使用前偏好是隐性的，相反，使用后偏好一般指评分，能够在数据集中明确反馈给模型。因此要获取使用前偏好，可以通过分析用户行为、构建用户画像分析、数据挖掘等方式。用户的使用前偏好决定了用户选择该项目的可能性，挖掘使用前偏好可以帮助推荐系统模型更好地理解用户，提供更加符合用户兴趣的推荐结果。

使用前偏好的取值设置为 [0,1]，其预测通常使用加权矩阵分解补全使用前偏好矩阵，使用前偏好的矩阵可以借助评分矩阵。在推测使用前偏好的模型中，通常将用户交互过的项目的使用前偏好设置为 1，因为用户选择该项目极大可能是因为对其感兴趣，反之，用户对其他项目的使用前偏好将被初始化为 0，完成了初始使用前偏好矩阵的构

建，随后用加权矩阵模型进行预测。加权矩阵分解的优化目标为：

$$L = \sum_{u=1}^m \sum_{i=1}^n w_{ui} (\hat{p}_{ui} - p_{ui})^2 + \lambda (\|\mathbf{X}_u\|_F^2 + \|\mathbf{Y}_i\|_F^2) \quad (2-16)$$

其中， p 表示使用前偏好， w_{ui} 表示用户 u 对项目 i 的权重。权重的设置可以根据用户的评分数量、评分平均值、或其他进行数据挖掘得到的数据来设置。当补全使用前偏好矩阵后，针对于每个用户，可以将该用户对每个项目的使用前偏好进行降序排序，设置一个百分比阈值，排名靠后的阈值以下的项目就被定义为该用户的无趣项 (Uninteresting Item)。

无趣项的识别对于推荐系统模型是至关重要的。对于 Top-K 推荐的协同过滤模型，可以通过识别出无趣项排除掉一些非必要的推荐，使得出现在推荐列表的推荐项被用户感兴趣的概率提高。由于无趣项的识别只针对于数据集，所以该技术对于任何协同过滤模型是正交的，处理后的数据集可以运用于任何模型。此外，推荐系统中也有成对训练的方法，在训练时为每个历史交互项目配对一个未交互过的项目作为负例，如下所示：

$$L = - \sum_{(i,j) \in \mathcal{W}} \log \tilde{y}_{ij} - \sum_{(i,k) \in \mathcal{W}^-} \log(1 - \tilde{y}_{ik}), \quad (2-17)$$

其中 \mathcal{W} 和 \mathcal{W}^- 分别表示正例和负例集合。通常，负例的选取是随机的，这其中或许会选取到一些用户可能会感兴趣的项目，通过无趣项识别技术，可以在无趣项中选取负例，这能让模型学习到更完整的用户喜好。

因为推荐系统中绝大部分数据集是高度稀疏的，其稀疏度达到 95% 以上，为了降低数据集的稀疏度，让模型能够有更多准确的训练样本，无趣项的填充也伴随着无趣项的识别一起运用。无趣项的填充是将识别出的无趣项用一个低分值去填充，代表用户不喜欢该项目。这样做将大幅度扩充训练样本的数量，让模型不仅能够从用户的高评分中学习到用户的喜好，也能从这些无趣项的评分中学习到用户不喜欢的项目特征，完善了用户的画像。无趣项的填充可以非常极端地用 0 分来填充，尽管 0 分已经低于了正常评分的最低值，在满分为 5 分的评分机制中，也可以用 1 分或者 2 分来填充无趣项，或是用其他更加科学的机制来确定不同的分值来填充。

2.3 深度学习

深度学习已是当前非常流行的一种机器学习技术，通过模拟人类神经系统的机制，来使用数据进行学习。深度学习模型中含有若干神经元，其数量可至百亿甚至千亿，每个神经元之间相互连接，以传递信息的方式学习输入和输出间的关系。随着计算机技术和深度学习技术的不断发展，深度学习已经在计算机视觉、自然语言处理、推荐系统等多个领域展示了其优越的性能。基于深度学习的推荐系统模型能够从结合显式反馈、隐式反馈、项目标签、用户人口统计数据等诸多特征进行推荐，解决了传统矩阵分解无法

捕获用户和项目间的非线性特征的缺点，提升了推荐效果。本节将概述深度学习的基本架构和在推荐系统中的应用。

2.3.1 理论基础

深度学习的结构模仿了人类大脑的神经元。神经元接收多个输入，将每个输入乘上对应的权重并求和，然后经过激活函数将输出压缩至一个合理的范围。激活函数接收神经元的输入，产生相应的输出，使神经网络能够学习和表示复杂的非线性关系。激活函数的发展历史可以追溯到神经网络的早期阶段，最早的激活函数是阶跃函数 (Step Function)，如下所示：

$$\text{step}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2-18)$$

然而，阶跃函数不连续且不光滑，导数几乎处处为零，使得梯度下降等优化算法无法有效地更新参数。Sigmoid 函数也是最早的激活函数之一，它将输入压缩到 $(0,1)$ 之间，具有平滑的 S 形曲线图像，公式如下：

$$S(x) = \frac{1}{1 + e^{-x}} \quad (2-19)$$

由于 Sigmoid 函数存在梯度消失和输出不均匀分布的问题，深度学习模型逐渐更多地选择 ReLU(Rectified Linear Unit) 函数作为激活函数。ReLU 函数在大于 0 时输出输入值，否则输出 0，不但解决了梯度消失的问题，还提高了计算效率，公式如下：

$$\text{ReLU}(x) = \max(0, x) \quad (2-20)$$

三种激活函数的图像对比如图2-1所示。

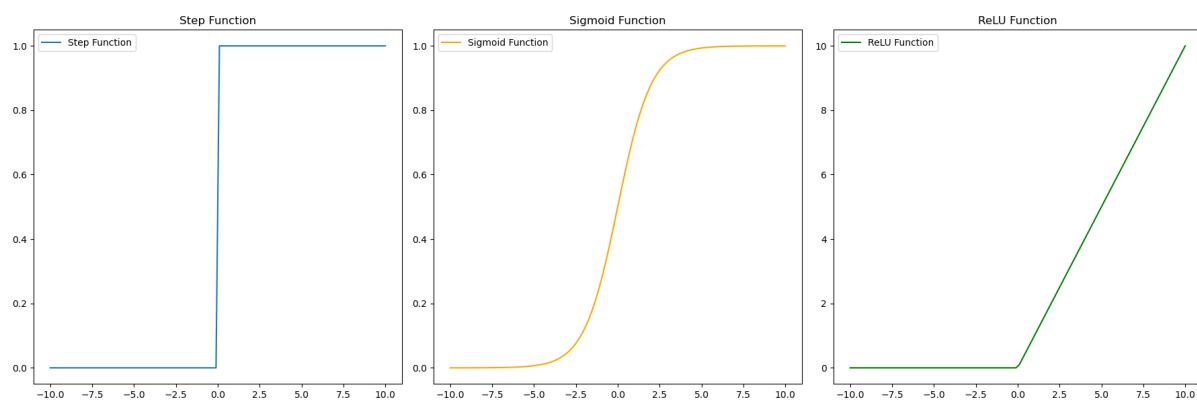


图 2-1 不同激活函数图像对比

多层感知机^[53](Multilayer Perceptron, MLP) 是一种常见的前馈神经网络模型。相较于单层感知机，MLP 具有更深的网络结构，由多个神经元层组成，其中除了输入层和输出层外，还包括一个或多个隐藏层 (Hidden Layer)，使得 MLP 能够学习到更复杂的线性关系，并且能够通过反向传播算法优化模型参数。MLP 的计算公式可以分为前向传播

和输出层激活两个阶段。在前向传播阶段，输入数据通过多个隐藏层，经过权重和激活函数的计算，最终到达输出层。每一层的输出作为下一层的输入，直至到达输出层。对于第 k 层，输出可以表示为：

$$z^k = \mathbf{W}^k \cdot a^{(k-1)} + b^k \quad (2-21)$$

其中， \mathbf{W}^k 是第 k 层的权重矩阵， $a^{(k-1)}$ 是第 $k-1$ 层的输出， b^k 是第 k 层的偏置项。随后，在输出层，使用激活函数将加权输入转换为最终的输出结果，如下所示：

$$\hat{y} = \sigma(z^K) \quad (2-22)$$

其中， $\sigma(\cdot)$ 表示选择的任意激活函数。从上述的计算过程中可以看出，神经网络通过加入权重参数、偏置项和激活函数来拟合特征与目标值之间的函数关系。

2.3.2 深度协同过滤

深度协同过滤是一种基于深度学习技术的推荐系统方法，旨在解决传统协同过滤方法中存在的稀疏性和冷启动等问题。随着互联网的发展和数据的爆炸性增长，个性化推荐成为了各大互联网平台的核心服务之一。然而，传统的基于矩阵分解的协同过滤方法在处理大规模稀疏数据时效果并不理想，而深度协同过滤技术的出现填补了这一空白。

深度协同过滤推荐系统模型是将深度学习与传统推荐任务结合，通过学习矩阵分解中无法学习到的非线性特征提升推荐效果。Wide&Deep 模型是 Google 公司提出的一种用于推荐系统的深度学习模型。该模型结合了线性模型 (Wide) 和深度神经网络模型 (Deep)，Wide 部分用于捕捉低阶的特征交互关系，Deep 部分用于学习高阶的非线性特征交互关系。在该模型的输入中包含了连续特征，如年龄，以及分类特征，如用户性别职业等，这些特征都由嵌入技术转换为 Embedding 进入神经网络进行学习，提升了模型的泛化性。

神经协同过滤模型 (Neural Collaborative Filtering, NCF) 随后被提出。该模型结合了神经网络和协同过滤的思想，更好地捕捉了用户和项目之间的复杂关系。与 Wide&Deep 相似，NCF 通过广义矩阵分解 (Generalized Matrix Factorization, GMF) 学习项目和用户间的线性关系，通过 MLP 学习数据中的非线性交互，随后这两部分的输出将被拼接起来再经过一次运算后通过激活函数输出，得到用户对当前项目的喜好分数。

深度协同过滤技术在电商、社交网络、音视频推荐等领域得到了广泛的应用。然而，深度协同过滤技术也面临着数据稀疏、冷启动、模型解释性等方面的挑战，如何提高模型的泛化能力、降低计算复杂度、增强模型的可解释性等是未来深度协同过滤研究的重要方向之一。

2.4 本章小节

本章首先介绍了推荐系统的相关概念及分类，并详细说明了本文会涉及到的评估方法和评估指标，然后对传统推荐系统的相关技术进行了详细分析，最后介绍了目前流行的基于深度学习的协同过滤相关技术。

第3章 基于用户评分习惯识别并填充无趣项的选择偏差抑制方法

3.1 引言

协同过滤是推荐系统中最有效、最流行的技术之一^[3-4]，它根据用户的偏好提供个性化的推荐。用户的历史数据可以分为显式反馈（如评分、评论等）和隐式反馈（如点击、搜索等）两种。显式反馈通过明确的数值大小反映用户的偏好，为评分预测提供了准确的信息。然而，用户可能更倾向于对自己喜欢的项目评分，从而导致低评分的历史数据缺失^[5-7]，这种非随机缺失极大地扭曲了完整的用户偏好，降低了Top-K推荐的准确率^[8-9]。与显式反馈相比，隐式反馈不仅关注观测数据，还关注了缺失数据。因此，结合这两种反馈的优点可以充分还原用户完整的偏好，更有效地进行Top-K推荐。

项目未被用户评分有三种可能性^[5]：(1) 用户不知道该项目的存在；(2) 用户购买了该项目，但没有对其进行评分；(3) 用户知道该项目但对其不感兴趣，因此没有购买或评分。第三种情况可以被理解为用户具有潜在的负面偏好，即用户对某些项目不感兴趣。为了提高推荐质量，可以将这部分项目定义为“无趣项”并避免推荐给用户。识别无趣项可以借用单类协同过滤 (One-Class Collaborative Filtering, OCCF) 模型^[18]，在该模型中，所有未评分项目都被视为负样本，通过赋予这些负样本权重来量化其置信度，并进行加权矩阵分解。现有的方法认为，当用户对更多的项目评分时，用户对未评分项目不感兴趣的可能性越大^[19,21]，因此给予这些用户的负样本更高的置信度。然而，这种以用户为导向的方法并没有考虑到用户的个性差异。

为了挖掘细粒度的用户个性并更合理地分配权重，本节提出了一种名为基于用户评分习惯的无趣项识别与多低值填充 (Uninteresting Items Identify and Multiple Low-rating Filling Based on Users Rating Habits, UIIMLF-URH) 方法，结合了隐式反馈和显式反馈的优点。本章认识到零值填充可能夸大了用户的负面偏好 [10]，因此通过显式反馈来量化项目的质量，并设定了一个阈值，将不同质量的项目用不同的非零低值填充。本节在两个数据集上进行了实验，将无趣项的识别与填充分为两个阶段，并对比了三种使用前偏好建模方法和两种无趣项填充方法。实验结果表明，UIIMLF-URH 方法在常用的 Top-K 推荐评价指标中，包括精确率、召回率和归一化折损累计增益等三个评价指标上的效果均占优势。实验结果验证了 UIIMLF-URH 方法在挖掘用户个性、合理分配权重方面的有效性。通过结合隐式反馈和显式反馈的优点，并考虑到项目质量和用户评分习惯，UIIMLF-URH 方法能够更准确地识别用户的无趣项，并采用合理的填充策略提高推荐系统的性能。这种方法提供了更个性化和准确的推荐结果，能够提高用户的满意度和使用体验。此外，该方法也能够为第四章提出的深度协同过滤模型提供帮助。

3.2 算法设计

本节提出的 UIIMLF-URH 方法旨在是计算使用前偏好并识别无趣项，然后用多种低值填充无趣项。填充后的评分矩阵可以作为任何协同过滤模型的输入，因此本节方法可以应用于协同过滤模型。图 3-1 展示了该方法的流程示意图。

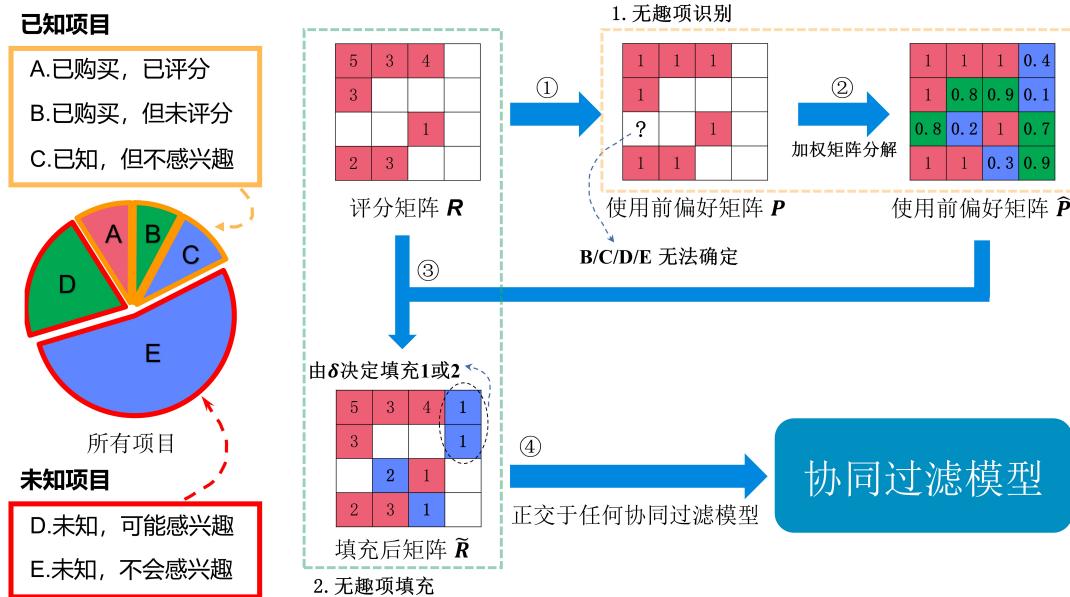


图 3-1 UIIMLF-URH 模型示意图

3.2.1 使用前偏好建模算法

未评分的项目可能包括用户已经购买但由于不喜欢而没有评分的项目。然而，在OCCF中，这些缺失评分的正样本仍然被当作负样本处理，这会增加模型的误差。因此，这类样本应该被赋予较低的权重。以用户为导向的加权方案假设用户评分的次数越多（即正样本越多），用户缺失数据为负样本的概率就越大[9]。因此赋予这些用户相关的负样本更高的权重。然而，这种方法只考虑了用户评分的数量，并未考虑到用户评分的特点。

定义原评分矩阵 $\mathbf{R} = (r_{ui})_{m \times n}$ ，其中 m 和 n 分别为用户和项目数量， r_{ui} 是用户 u 对项目 i 的评分。为了计算使用前偏好，引入初始使用前偏好矩阵 $\mathbf{P} = (p_{ui})_{m \times n}$ 。已评分的项目视作正样本，对应的使用前偏好 $p_{ui} \in [0, 1]$ 被设置为最高值 1。缺失样本的使用前偏好被初始化为 0，如式(3-1)所示。

$$p_{ui} = \begin{cases} 1, & r_{ui} \neq \text{null} \\ 0, & r_{ui} = \text{null} \end{cases} \quad (3-1)$$

定义用户 u 每条评分与对应项目的平均分之差的平均值为该用户的评分习惯分 α_u 。

由于评分波动较小的用户展现出更稳定的评分习惯，因此，选择用户所有评分的标准差来修正评分习惯，如式(3-2)所示。

$$\alpha_u = \frac{\sum_{i \in I_u} (r_{ui} - \bar{r}_i)}{|I_u|} \times \frac{1}{e^{s_u}} \quad (3-2)$$

其中， \bar{r}_i 表示项目 i 的平均得分， I_u 是用户 u 评过分的项目集合， s_u 是用户 u 的历史评分的标准差。根据 α_u 的定义以及 3.1 小节的分析，可以推断出 α_u 的值越大，说明用户 u 越倾向于给项目打高分。基于这个观察，本节提出第一条假设：

假设 1. 倾向于打高分的用户对未评分的项目不感兴趣的可能更大。

在 3.3 小节中，本节对这一假设进行了验证，通过对实验结果和分析的观察，发现倾向于打高分的用户更有可能对未评分的项目表现出较低的兴趣。根据假设 1，用户对应的 a_u 越大，对未评分的项目不感兴趣的可能更大，因此在 OOCF 模型中，这类用户的负样本具有更高的置信度，即缺失样本为负样本的概率更大。本节引用了 Pan 等人的研究 [9]，将用户评分数量与对未评分项目的喜好程度关联起来，使得评分数量越多的用户对未评分项目的不喜欢程度越高。定义 β_u 表示用户 u 的评分数量，并通过 log 函数进行平滑，以平缓极端值带来的影响，如式(3-3)所示。

$$\beta_u = \log\left(\sum_{i \in I} p_{ui}\right) \quad (3-3)$$

对 α 和 β 进行归一化处理，如式(3-4)和式(3-5)所示。

$$\hat{\alpha}_u = \frac{\alpha_u - \alpha_{min}}{\alpha_{max} - \alpha_{min}} \quad (3-4)$$

$$\hat{\beta}_u = \frac{\beta_u - \beta_{min}}{\beta_{max} - \beta_{min}} \quad (3-5)$$

定义权重矩阵 $\mathbf{W} = (w_{ui})_{m \times n}$ ，为了综合考虑用户评分习惯和评分数量，通过对 $\hat{\alpha}_u$ 和 $\hat{\beta}_u$ 进行线性加权来计算权重，如式(3-6)所示。

$$w_{ui} = \begin{cases} \gamma \hat{\alpha}_u + (1 - \gamma) \hat{\beta}_u, & p_{ui} = 0 \\ 1, & p_{ui} = 1 \end{cases} \quad (3-6)$$

其中， $\gamma \in (0, 1)$ 是调节系数。将 \mathbf{P} 分解为两个低秩矩阵 $\mathbf{X} \in \mathbf{R}^{m \times k}$ 和 $\mathbf{Y} \in \mathbf{R}^{n \times k}$ ，其中 k 是特征维数，分别表示用户和项目特征向量矩阵。结合权重矩阵 \mathbf{W} ，加权矩阵分解的目标函数 L 如式(3-7)所示。

$$L = \sum_{u=1}^m \sum_{i=1}^n w_{ui} (\hat{p}_{ui} - p_{ui})^2 + \lambda (\|\mathbf{X}_u\|_F^2 + \|\mathbf{Y}_i\|_F^2) \quad (3-7)$$

其中， γ 是正则化系数， $\hat{p}_{ui} = \mathbf{x}_u^\top \mathbf{y}_i$ ， $\|\cdot\|$ 表示 F 范数。在优化过程中，考虑到正则化项和损失函数的影响，本节采用了文献 [19] 的方法进行优化。首先，对 L 求 x_{uf} 的偏导

表 3-1 不同情况的原因分析

类别	使用前偏好	使用后偏好	场景	可能原因
A	高	高	已看, 已评分	观前就感兴趣, 且观后评价高
B	高	低	已看, 未评分	观前感兴趣, 但观后评价低
C	低	未知	已知, 不喜欢	观前不感兴趣, 不会去看
D	可能高	未知	未知, 会感兴趣	不知道有这个电影, 如果知道会感兴趣
E	可能低	未知	未知, 不会感兴趣	不知道有这个电影, 如果知道也不会感兴趣

数, 如式(3-8)所示。

$$\frac{1}{2} \frac{\partial L}{\partial x_{uf}} = - \sum_{i=1}^n (p_{ui} - \hat{p}_{ui}^f) w_{ui} y_{if} + x_{uf} \sum_{i=1}^n w_{ui} y_{if}^2 + \lambda x_{uf} \quad (3-8)$$

其中, $\hat{p}_{ui}^f = \hat{p}_{ui} - p_{uf} q_i f$ 。令 $\frac{\partial L}{\partial x_{uf}} = 0$, 得到 x_{uf} 的解, 如式(3-9)所示。

$$x_{uf} = \frac{\sum_{i=1}^n (p_{ui} - \hat{p}_{ui}^f) w_{ui} y_{if}}{\sum_{i=1}^n w_{ui} y_{if}^2 + \lambda} \quad (3-9)$$

同理, 得到 y_{if} 的解, 如式(3-10)所示。

$$y_{if} = \frac{\sum_{i=1}^n (p_{ui} - \hat{p}_{ui}^f) w_{ui} x_{uf}}{\sum_{i=1}^n w_{ui} x_{uf}^2 + \lambda} \quad (3-10)$$

迭代上述两步, 通过交替优化用户和项目特征向量, 逐渐逼近收敛至局部最优, 最终, 得到了近似于矩阵 \mathbf{P} 的使用前偏好矩阵 $\hat{\mathbf{P}} = \mathbf{X}^\top \mathbf{Y}$ 。

3.2.2 无趣项的识别

无趣项被定义为用户不会感兴趣的项目, 即使用前偏好较低的项目。以电影举例, 如表3-1所示, 对于用户来说有五种不同的场景。首先, 在已经观看并评价的电影中, 可以观察到了两种不同的情况。对于电影 A, 用户在观看前就表现出了浓厚的兴趣, 并且在观看后评价也较高, 这表明用户对该项目的兴趣是符合其偏好的, 这部分数据是用于训练的最重要数据。相反, 对于电影 B, 用户在观看前也有较高的偏好, 但在观看后评价却较低, 这可能意味着该项目并没有达到用户的期望, 或者用户的期待与实际呈现不符, 所以用户会不打分。这部分项目由于用户未打分, 所以在显式反馈中缺失了交互数据, 这部分数据不会用于训练, 且会在推荐的备选项中再次出现, 这就造成了数据中的选择偏差。其次, 未观看的项目中, 电影 C 和电影 E 可以归为一类, 用户在观看前已经表现出了低的偏好, 这部分电影就算推荐给用户, 用户也不会选择去看, 导致推荐质量的降低。这部分项目是本章工作中期望识别出的无趣项。最后, 对于电影 D, 用户会感兴趣, 但是用户并不知道这部电影, 这类项目即为推荐系统任务期望推荐给用户的项目。

通过前文对使用前偏好的建模和对不同场景的分析, 可以通过预测的使用前偏好对进行无趣项的识别。由于无趣项通常是使用前较低的项目, 因此针对每个用户, 可以

根据其对每个项目的使用前偏好进行降序排序，从而确定无趣项。不同的用户的活跃程度的不同，因此应当针对不同用户设定不同大小的无趣项集合。可以通过引入百分比参数 $\theta\%$ 来控制用户无趣项集合的大小，将用户使用前偏好列表中排在末尾的 $\theta\%$ 项目视为其无趣项。增大 θ 能筛选出更多的无趣项，但也可能将用户感兴趣的项目误判为无趣项。相反，减小 θ 会减少误判的可能性，但可能保留更多用户不感兴趣的项目。因此，确定适当的 θ 值成为实验中的重要的考量之一。

根据补全后的使用前偏好矩阵 $\hat{\mathbf{P}}$ ，无趣项的筛选过程如下式所示：

$$I_u^{un}(\theta) = \{i \notin I_u \mid \rho(\hat{p}_{ui}, P) \leq \theta\} \quad (3-11)$$

其中， $\rho(\hat{p}_{ui})$ 表示 \hat{p}_{ui} 的大小在集合 $P = \{\hat{p}_{ui} | r_{ui} = null\}$ 中的百分位排名， I_u 表示用户 u 评过分的项目集合。当 $\theta = 20\%$ 时，表示使用前偏好排名倒数 20% 的项目被视为当前用户的无趣项。

3.2.3 无趣项的填充

填充无趣项的目的有两个方面：首先，它可以对用户曾购买但未评分的项目进行评分补全，从而使模型能更全面地学习用户的偏好。其次，对于用户不感兴趣且未购买过的项目，可以用低分填充以确保这些项目不会出现在推荐列表中。同时，这种填充操作还可以增加模型的训练数据，降低数据集的稀疏性。

为了更全面地探索用户的偏好，对于缺失评分的项目，采用低分值进行填充是一种合理的选择^[5]。然而，需要注意的是，使用低于正常评分范围（1-5 分）的 0 分进行填充可能会夸大用户的负面偏好^[21]。由于显式反馈的核心优势在于评分的大小反应了明确的用户偏好，更细粒度的填充将更有利与模型对用户偏好的准确捕捉。因此，本节选择了正常评分范围内的两个最低分值，即 1 分和 2 分作为填充值。这样的选择不仅可以有效补全缺失的评分数据，还能够避免夸大用户的负面偏好。

在基于项目的协同过滤模型中，对项目的评分是决定用户之间相似度的重要依据，以图3-2为例进行说明。假设项目 i 和 j 的相似度为 $Sim(i, j)$ 。在原始评分矩阵中，基于项目的协同过滤方法可能会得出 $Sim(i_2, i_1) > Sim(i_2, i_3)$ 的结论。然而，如果只用 0 分来填充无趣项，那么会导致 $Sim(i_2, i_1) = Sim(i_2, i_3)$ 的结论。相比之下，如果在填充无趣项时选择不同的低分值进行替换，可以得出 $Sim(i_2, i_1) < Sim(i_2, i_3)$ 。因此，通过使用不同的低分值进行填充，可以让协同过滤模型找到真正相似的项目，并且提供更多的学习样本。

根据上述分析，无趣项的填充所面临的挑战在于确定一个标准，以确定哪些项目应该被分配相对较高的分值。项目本身的质量可被看作是一个有效的标准，其中项目的平均得分是描述项目质量最常用的指标之一，然而，这些评分受到用户主观意识的影响。为了推断项目本身的质量，本节将用户的评分与 3.2.1 小节中挖掘的评分习惯相结合，将用户评分由主观性转变为客观性，并消除评分习惯带来的偏差。因此本节提出第二条假

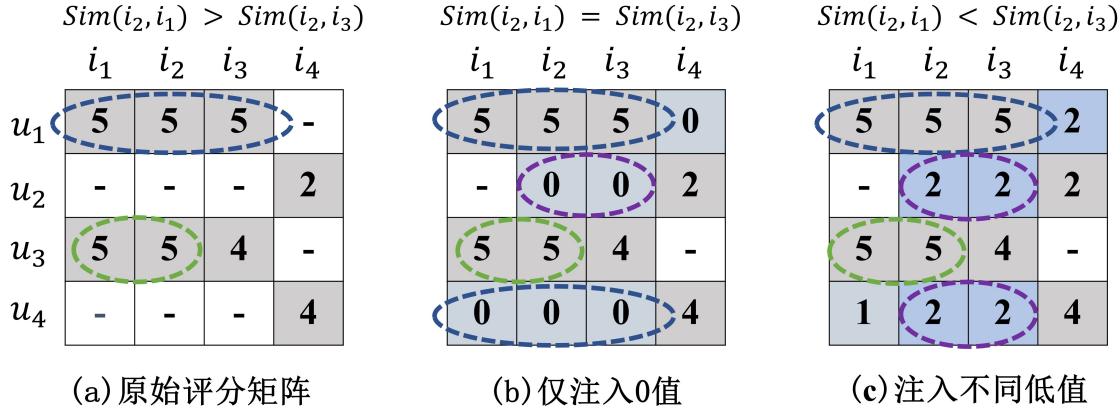


图 3-2 不同填充策略对比

设：

假设 2. 用户对项目的客观评分与用户的平均评分之间的差值决定了项目的质量分，差值越大，质量越高。

根据假设 2，通过计算用户评分与平均评分之间的差值，可以估计项目的质量分。在 3.3 小节中，本节对该假设进行了验证，进一步确认了用户评分与平均评分之间的差值与项目质量之间的关系。根据假设 2，可以定义项目的质量分 δ_i 。首先，将每条评分减去对应用户的平均评分，再减去该用户对应的评分习惯分 α_u ，消除用户评分习惯带来的偏差。然后将这些差值求平均，差值越大表示该项目的质量越高，如式3-14所示。

$$\delta_i = \frac{\sum_{u \in U_i} (r_{ui} - \bar{r}_u - \alpha_u)}{|U_i|} \quad (3-12)$$

其中， U_i 是对项目 i 评分的用户集合， \bar{r}_u 是用户 u 的平均评分。为了确定填充值 \tilde{r}_{ui} 的大小，引入阈值 $\xi \in [0, 1]$ ，决定了填充值的取值，如式(3-13)所示。

$$\tilde{r}_{ui} = \begin{cases} 1, & \hat{\delta}_i < \xi \\ 2, & \hat{\delta}_i \geq \xi \end{cases} \quad (3-13)$$

$$\hat{\delta}_i = \frac{\delta_i - \delta_{min}}{\delta_{max} - \delta_{min}} \quad (3-14)$$

其中， \tilde{r}_{ui} 是 δ_i 归一化处理后的结果，如式(3-14)所示， δ_{max} 和 δ_{min} 分别是所有项目分 δ_i 的最大值和最小值。阈值 ξ 对实验结果的影响在 3.4 小节中展示与分析。

3.3 实验结果与分析

本节验证以下三个问题：

- (1) 本节提出的使用前偏好建模方法与现有的其他方法相比，其效果如何？
- (2) 本节提出的两个假设是否合理？
- (3) 本节提出的填充策略与现有的填充策略相比，其效果如何？

表 3-2 数据集信息

数据集	用户数量	项目数量	评分数量	稀疏度	分值范围
Movielens-100k	943	1682	100,000	93.7%	[1,5]
Movielens-latest	610	9724	100,836	98.3%	[1,5]

实验共分为五个部分。第一部分确定了模型线性加权部分的最佳参数。第二部分验证了使用前偏好建模的效果，选取了三个对比模型在三个指标上比较。第三部分是对假设一的验证，实验以 α 的大小进行分组，比较不同组别中不同使用前偏好的占比。第四部分是对填充策略的验证，实验在两个协同过滤模型中与两个对比方法进行了对比。第五部分是对假设二的验证。

3.3.1 数据集

本节实验使用了两个公开的电影数据集^[54]：Movielens-100k 和 Movielens-latest，其信息如表 3-2 所示。实验采用 5 折交叉验证，对每个数据集随机选取 80% 的评分数据作为训练集，剩下的 20% 作为测试集。

需要注意的是，在 3.2 节中，由于数据转化为隐式反馈，不对用户的相关项进行筛选。由于预测高评分的项目更有意义^[6]，因此在 3.4 节中，用户的相关项仅包括评分在 4 分及以上的项目。

3.3.2 评价指标

本节实验使用了 Top-K 推荐中常见的三个评价指标进行评估，分别是精确率 (Precision, P)、召回率 (Recall, R) 和归一化折损累计增益 (Normalize Discounted Cumulative Gain, NDCG)。各评价指标的计算公式如式(3-15)- 式(3-18)所示。

精确率衡量的是检索结果中相关项的比例。它定义为从所有被检索出来的结果中有多少是相关的。计算公式为：

$$P = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (3-15)$$

召回率衡量的是所有相关项中有多少被成功检索出来。计算公式为：

$$R = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (3-16)$$

NDCG 是一种用于评估排序结果的指标，尤其在推荐系统和信息检索领域。它考虑了结果的相关性和位置，较高相关性的结果如果出现在列表的较前位置会获得较高的评分。计算公式分为两步：

$$DCG@N = \sum_{i=1}^N \frac{2^{rel_i} - 1}{log_2(i + 1)} \quad (3-17)$$

$$NDCG@N = \frac{DCG@N}{IDCG@N} \quad (3-18)$$

其中, $R(u)$ 是给用户 u 推荐的项目集合, 长度为 N ; $T(u)$ 是测试集中用户 u 的相关项; rel_i 是第 i 个推荐项目的相关性, 取值为 0 或 1, 仅当第 i 个项目在集合 $T(u)$ 中时取值为 1, 用户的相关项被推荐时, 位置越靠前, DCG 的值越大; $NDCG@N$ 是理想状态下的 $DCG@N$ 值。

3.3.3 参数实验

本节提出的 UIIMLF-URH 模型中, 参数 γ 是线性加权的权重系数, 以 0.1 为步长逐渐增加, 并评估其结果。

图3-3和图3-4的结果显示, 随着参数 γ 的增加, 推荐的准确率也随之上升, 达到最大值后逐渐下降。这表明相较于单独考虑评分数量和评分习惯, 将二者结合起来可以提高推荐的效果。特别是在用户评分习惯的权重较大时, 推荐的效果更好, 这表明用户评分习惯对于无趣项的识别具有较大的贡献。在后续的实验中, 为了进一步评估模型的性能, 在 MovieLens-100k 中固定 $\gamma = 0.6$, 在 MovieLens-latest 中固定 $\gamma = 0.9$ 。

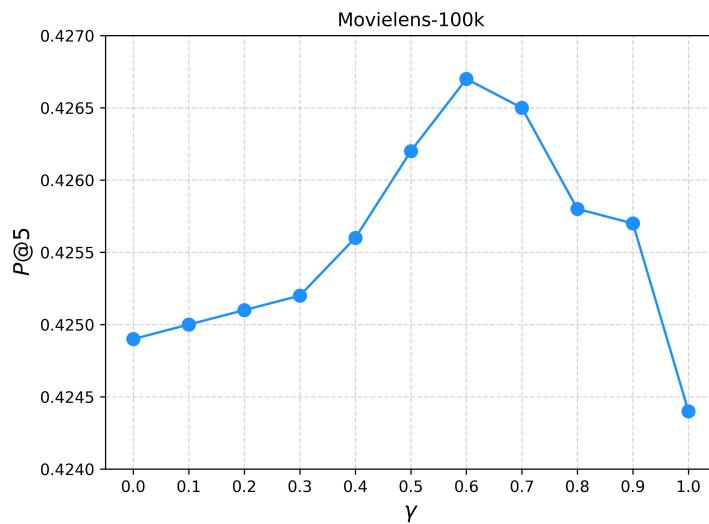


图 3-3 MovieLens-100k 中 γ 的影响

3.3.4 使用前偏好建模模型

本小节中, UIIMLF-URH 模型将与三个不同加权策略的方法进行比较。分别是 WRMF[8]: 经典的单类协同过滤方法, 其权重由该条数据的评分决定; eALS[19]: 改进了效率的加权矩阵分解方法, 其权重由项目的流行度决定; UTV[11]: 最新的无趣项识别方法, 其权重由用户活跃期间项目的流行度决定。

所有模型统一设置隐特征数量 $k = 20$, 正则项系数 $\lambda = 0.01$ 。三个对比模型都使用其对应文献中相同的参数。

表3-3 和表3-4 列出了 UIIMLF-URH 与另外 3 个算法对比的结果。可见, UIIMLF-

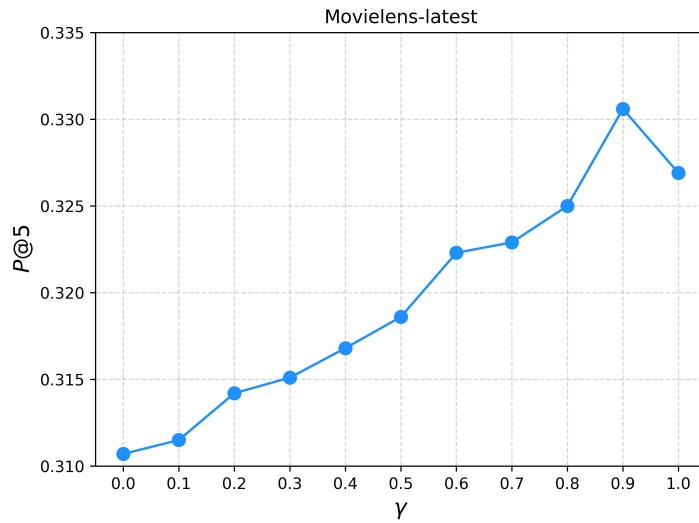
图 3-4 Movielens-latest 中 γ 的影响

表 3-3 MovieLens-100k 数据集使用前偏好建模对比

指标	WRMF	eALS	UTV	UIIMLF-URH
P@5	0.3469	0.3556	0.421	0.4267
P@10	0.3065	0.3043	0.3577	0.359
P@20	0.2574	0.2521	0.2867	0.288
R@5	0.1325	0.1328	0.1464	0.1489
R@10	0.2192	0.2192	0.2339	0.2341
R@20	0.3431	0.3305	0.3571	0.3569
NDCG@5	0.3681	0.3809	0.4529	0.4584
NDCG@10	0.3611	0.3648	0.4266	0.4282
NDCG@20	0.3774	0.375	0.4249	0.4277

表 3-4 MovieLens-latest 数据集使用前偏好建模对比

指标	WRMF	eALS	UTV	UIIMLF-URH
P@5	0.3013	0.3080	0.3283	0.3306
P@10	0.2803	0.2863	0.2713	0.2764
P@20	0.2092	0.2075	0.2166	0.2221
R@5	0.0996	0.1074	0.0927	0.0952
R@10	0.1617	0.1620	0.1511	0.1519
R@20	0.2124	0.2180	0.2245	0.2288
NDCG@5	0.3309	0.3373	0.3482	0.3513
NDCG@10	0.3028	0.3049	0.3206	0.3232
NDCG@20	0.3003	0.3085	0.3102	0.3142

表 3-5 未评分项的使用前偏好分布

使用前偏好	Movielens-100k	Movielens-latest
[0.0,0.2)	92.81%	98.48%
[0.2,0.4)	5.15%	1.19%
[0.4,0.6)	1.49%	0.25%
[0.6,0.8)	0.42%	0.06%
[0.8,1.0)	0.13%	0.02%

URH 算法在大多数情况下优于其他方法。这表明通过对用户个性进行更细粒度的探索，可以更准确地识别用户的无趣项，从而提高推荐效果。

表3-5的使用前偏好分布展示了在两个数据集中对未评分项目的用户兴趣情况。可以观察到以下情况：

(1) 绝大多数未评分项目的使用前偏好在 [0,0.2] 区间，这表示用户对大多数未评分的项目并不感兴趣，或者说没有明确的偏好倾向。这与假设 1 相符，即倾向于打高分的用户对未评分的项目不感兴趣的可能性更大。

(2) 少数未评分项目的使用前偏好在 [0.8,1.0] 区间，这说明一些用户对少数未评分的项目可能有一定的兴趣。这可能是因为这些项目是用户未知或不愿评分的项目，或者是近期推出的新项目。这部分项目可能具有一定的潜在吸引力，值得推荐给用户。

3.3.5 假设验证

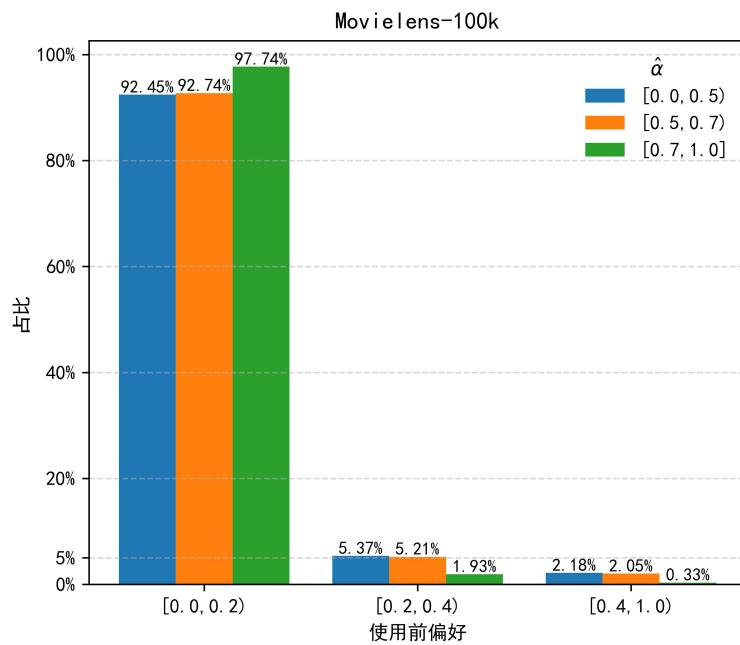
本节于3.2.1与3.2.2小节中分别提出了 2 个假设。本小节以 Movielens-100k 数据集为例，验证这两个假设。

根据假设 1，用户 u 未评分的项目中不感兴趣的项目占比随着 α_u 的增大而增大。为了验证该假设，将 α 的数值范围分成三组：[0,0.5)，[0.5,0.7)，[0.7,1]。每组中未评分项目的使用前偏好占比如图3-5所示。从结果中可以得知，随着 α_u 的增大，低使用前偏好的占比逐渐增加，而高使用前偏好的占比则逐渐减少。这表明随着 α_u 的增大，用户 u 对未评分项目不感兴趣的可能性增加，即缺失数据更有可能是负样本，因此，验证了假设 1。

为了验证假设 2，所有项目根据 δ 的数值大小被分为五组，用 100,000 条评分数据统计了每组项目的平均得分。图 3-6 展示了统计结果，可以观察到随着 δ 的增加，项目的平均得分也随之递增。这表明当项目收到的评分相对于用户的历史平均评分的差值越高时，项目本身的平均分更高，即项目本身的质量更高，假设 2 得到验证。

3.3.6 填充策略

本小节首先利用 BiasSVD^[55]模型对参数 θ 和 ξ 对填充结果的影响进行了评估。其中， θ 用于控制填充的数量， ξ 用于控制填充值的比例。确定两个参数后，本节在 BiasSVD

图 3-5 α 和使用前偏好占比的分组对比

和 ItemCF^[56]两个经典推荐模型上，采用了不同填充策略对评分矩阵进行了填充，并对比了不同填充策略对推荐效果的影响。BiasSVD 是一种经典的矩阵分解方法，它考虑了评分的偏置项；ItemCF 则是基于项目相似性的经典协同过滤算法。这两个模型使用了 Librec^[57]中提供的实现。

参数 θ 用于控制填充的数量。如果取值太高，可能会将用户可能感兴趣的项目误判为无趣项。以 MovieLens-100k 数据集为例，首先固定 $\xi = 0.8$ ，以 10% 为步长将 θ 从 0 增长到 90%，并观察 P@5 指标的变化。根据图 7 的结果可知，随着 θ 的增加，精确率呈平缓增加的趋势，说明排除更多的无趣项可以提升推荐效果。参数 ξ 用于控制填充值的比例。以 0.1 为步长将 ξ 从 0 增加到 1，结果如图 8 所示。其中，红色虚线表示填充较低值的无趣项在所有无趣项中的占比。由于在 MovieLens-latest 数据集中，推荐的准确率在参数变化时呈现了与 MovieLens-100k 几乎相同的趋势。因此在后续实验中将固定 $\theta = 90\%$, $\xi = 0.6$ 。

本节对比了两种填充策略，分别是 AllRank^[6]和 Zero-Injection^[5]。AllRank 用 2 分来填充所有缺失的项目；而 Zero-Injection 使用 0 分来填充所有无趣项，且无趣项识别过程与本节方法一致。表 3-6展示了这两种填充策略与本节方法的区别。

表 3-6 填充策略的对比

填充模型	填充范围	填充值大小
AllRank	所有缺失项	2
Zero-Injection	无趣项	0
UIIMLF-URH	无趣项	1, 2

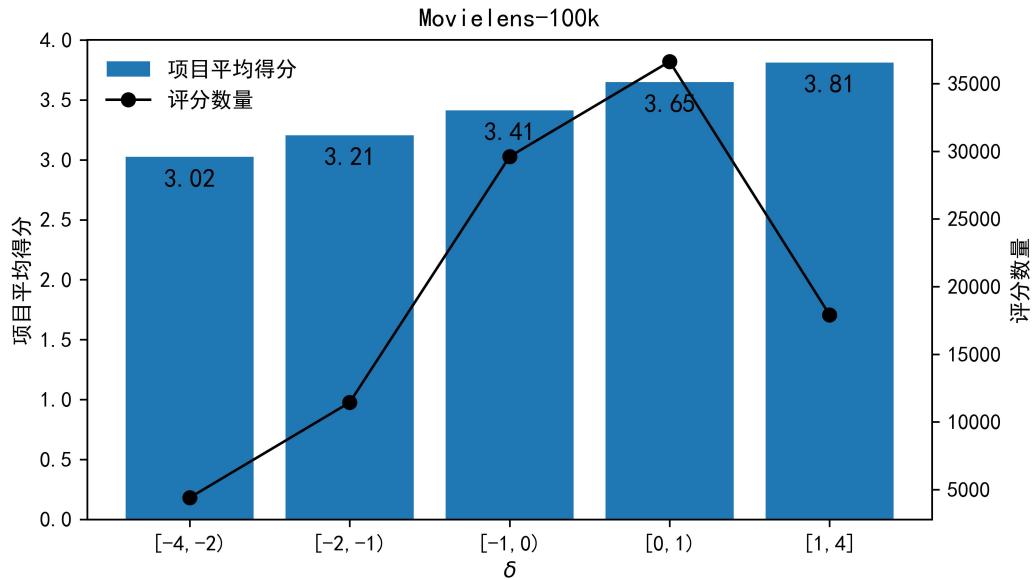
图 3-6 根据 δ 分组的项目评分统计

表 7 和表 8 分别展示了在两个数据集上进行实验的结果。实验结果表明，在 BiasSVD 和 ItemCF 两个模型中，不填充任何缺失项目的推荐结果在所有评估指标上表现较差。相比之下，执行了填充策略的方法都有更好的推荐效果。其中，UIIMLF-URH 方法在每个指标上都优于 AllRank 和 Zero-Injection 方法，说明采用不同低值填充无趣项的方法更好地平衡了用户的负面偏好。

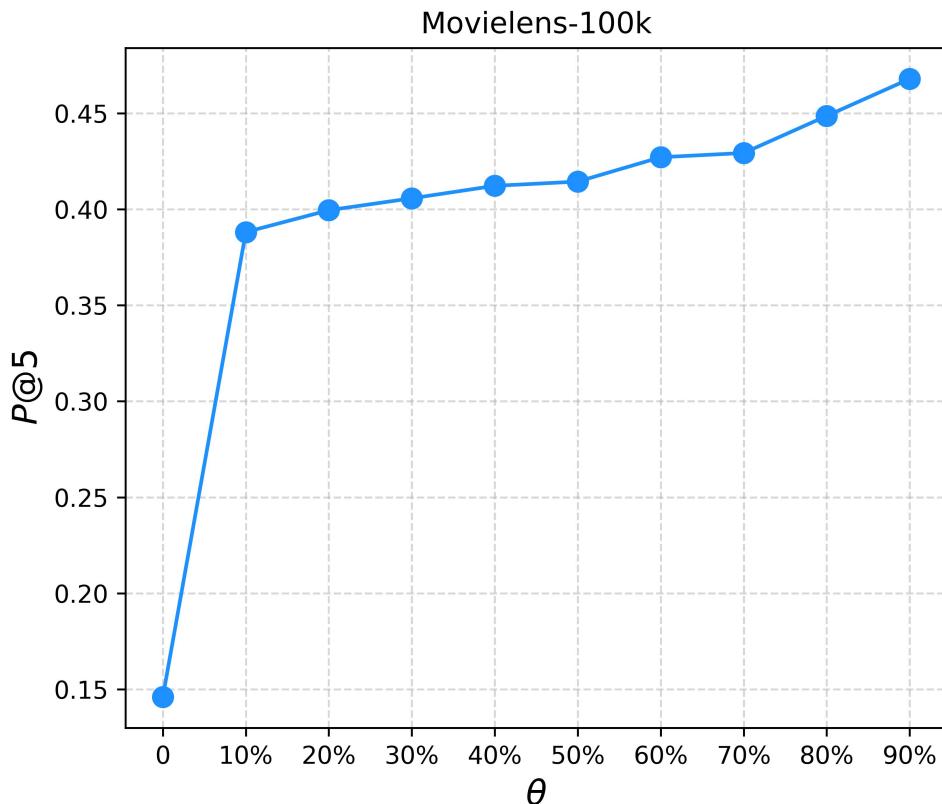
表 3-7 MovieLens-100k 中各填充方法比较

模型	BiasSVD				ItemCF			
	填充方法	不填充	AllRank	Zero-Inj.	本节方法	不填充	AllRank	Zero-Inj.
P@5	0.146	0.443	0.4507	0.4565	0.107	0.3421	0.3561	0.3671
P@10	0.1379	0.3944	0.3998	0.4059	0.0978	0.3065	0.325	0.3328
P@20	0.116	0.3248	0.3324	0.3401	0.0865	0.2689	0.2794	0.2896
R@5	0.0251	0.141	0.1437	0.1532	0.0283	0.095	0.1078	0.1321
R@10	0.0501	0.2219	0.2262	0.2584	0.0498	0.1641	0.1853	0.2319
R@20	0.0895	0.368	0.3746	0.4109	0.0829	0.2747	0.3064	0.3812
NDCG@5	0.1449	0.4779	0.4807	0.4911	0.1144	0.3626	0.3767	0.3979
NDCG@10	0.1422	0.4564	0.4615	0.4778	0.1112	0.3482	0.3684	0.3963
NDCG@20	0.1351	0.4539	0.4603	0.4852	0.1123	0.3529	0.38	0.4188

3.3.7 结合深度协同过滤

为了进一步验证该方法的可行性，本小节将引入最经典的深度协同过滤模型 NeuMF 进行实验：

- NeuMF^[45]: 结合了 MF 和 MLP 的协同过滤模型，用于建模线性和非线性用户-项

图 3-7 θ 的影响

目交互，它能够从数据中学习任意函数。

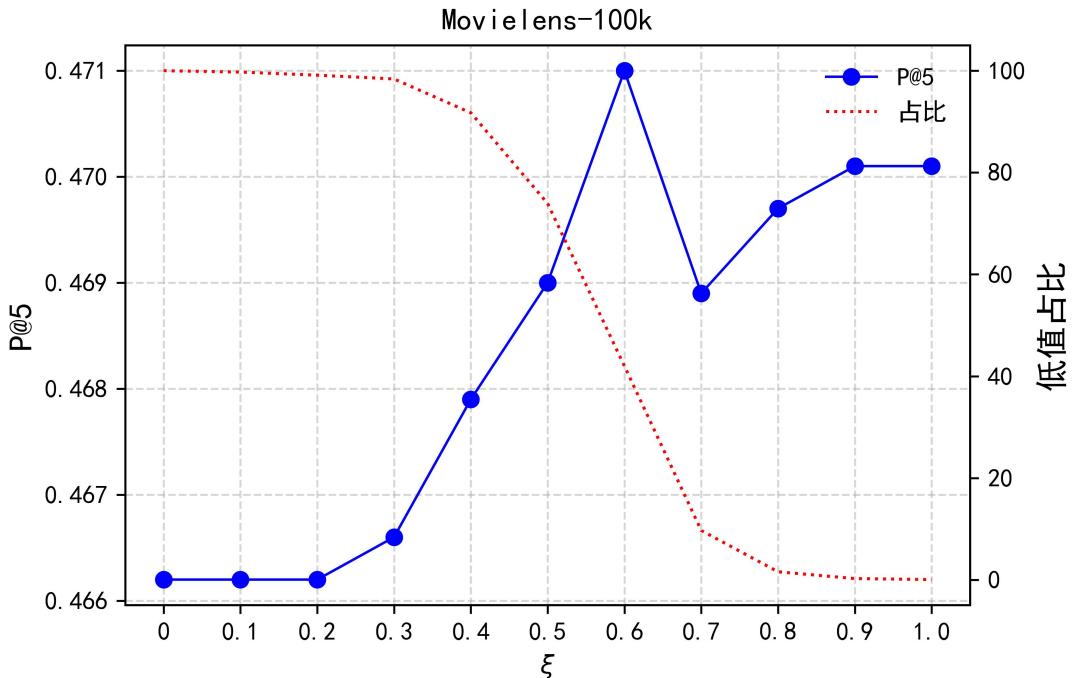
表3-9展示了实验的结果，结果表明，本章提出的方法在神经协同过滤模型上仍然有优异的表现。

3.4 本章小结

在推荐系统中，显式反馈数据（如用户评分）通常面临选择偏差和数据稀疏性的问题。这些问题会导致推荐结果的准确性和多样性受限。为了有效应对这些挑战，本研究提出了一种综合考虑用户前偏好、评分数量和评分习惯的方法，通过无趣项识别和填充，显著提升了推荐性能。

无趣项识别是识别出用户可能不感兴趣的项目，并通过填充策略处理这些项目。在本研究中，通过结合用户评分数量和评分习惯，识别出那些缺失的评分数据作为潜在的负样本。具体而言，评分数量反映了用户对某类项目的整体兴趣水平，而评分习惯则反映了用户对不同项目的偏好模式。通过将这两者结合起来，可以更准确地识别出无趣项。

通过计算使用前偏好并进行无趣项识别和填充，能够有效地缓解显式反馈数据中的选择偏差，并解决数据稀疏问题，从而更有效地利用显式反馈数据进行 Top-K 推荐。在无趣项的识别中，通常使用用户的评分大小或数量来度量缺失数据作为负样本的可信

图 3-8 ξ 的影响

度。在无趣项填充阶段，本研究引入了项目质量分的概念，将项目分为两个部分：高质量项目和低质量项目。对于无趣项，分别使用零值和低值进行填充，从而在推荐系统中反映出这些项目的低受欢迎程度。零值填充代表用户对该项目完全不感兴趣，而低值填充则表示用户对该项目的兴趣较低但并非完全没有兴趣。通过实验证，本章提出的方法在前偏好建模和无趣项填充两个阶段均表现出色。与传统方法相比，本方法在多个评估指标上均有显著提升，包括精确率、召回率和规范化折损累计增益。这些结果表明，通过综合考虑显式反馈和隐式反馈，充分挖掘用户评分习惯和评分数量，可以更有效地缓解选择偏差和数据稀疏性问题，从而提升推荐系统的整体性能。

表 3-8 MovieLens-latest 中各填充方法比较

模型 填充方法	BiasSVD				ItemCF			
	不填充	AllRank	Zero-Inj.	本节方法	不填充	AllRank	Zero-Inj.	本节方法
P@5	0.0584	0.1943	0.2046	0.2103	0.0894	0.1347	0.1435	0.1442
P@10	0.0459	0.1586	0.1682	0.1782	0.0784	0.1193	0.1225	0.1305
P@20	0.0393	0.1257	0.1355	0.1461	0.0679	0.0963	0.0992	0.1045
R@5	0.0202	0.0842	0.0902	0.0958	0.0367	0.0524	0.0587	0.0604
R@10	0.0294	0.1316	0.1391	0.1527	0.0610	0.1001	0.1033	0.1069
R@20	0.0475	0.1938	0.2065	0.2300	0.1091	0.1479	0.1536	0.1641
NDCG@5	0.0646	0.2101	0.2173	0.2309	0.0948	0.1434	0.1472	0.1503
NDCG@10	0.0580	0.2007	0.2071	0.2261	0.0942	0.1397	0.1433	0.1550
NDCG@20	0.0572	0.2023	0.2096	0.2345	0.1041	0.1442	0.1498	0.1604

表 3-9 MovieLens-100k 中各填充方法比较

模型 填充方法	NeuMF			
	不填充	AllRank	Zero-Inj.	本章方法
P@5	0.257	0.312	0.303	0.332
P@10	0.350	0.382	0.371	0.400
P@20	0.250	0.281	0.278	0.290
R@5	0.344	0.383	0.391	0.409
R@10	0.201	0.253	0.235	0.258
R@20	0.250	0.310	0.298	0.331
NDCG@5	0.377	0.421	0.410	0.431
NDCG@10	0.382	0.421	0.411	0.428
NDCG@20	0.361	0.387	0.396	0.410

第4章 基于建立理想用户组的深度协同过滤方法

4.1 引言

在信息过载的时代，推荐系统在提供更合适的电影、音乐、书籍等方面发挥着重要作用。有各种类型的推荐系统，包括基于内容的推荐系统、基于人口统计学的推荐系统 (Demographic-based Recommender System, DRS)、协同过滤 (Collaborative Filtering, CF) 和混合推荐系统^[1-2]。基于内容的 RS 根据历史评分记录和提取的文本或图像信息进行推荐^[58]。DRS 通过人口统计相似性连接用户，其合理性基于相似的用户群体具有相似的偏好^[10-12]。CF 通过用户-项目交互和用户或项目之间的相似性挖掘协同信息^[59-63]。混合推荐系统将用户-项目交互信息与人口统计和文本等其他信息相结合^[23,46]。

尽管人口统计信息，如年龄、性别和职业，有助于个性化推荐的指定，但必须认识到仅依赖人口统计信息可能导致偏见推荐和刻板印象的持续存在。传统的 DRS 通常使用人口统计信息的相似性来计算用户之间的相似性^[2]，然而除了人口统计信息之外，考虑到用户的行为、偏好和反馈等其他因素也至关重要。此外，现实世界中，每个项目都有一个理想的目标客户群体。例如，服装和鞋类等可以被划分为“休闲”或“运动”等针对不同用户群体的类别，电影和游戏爱好者也根据电影和游戏的标签而被分成多个组别。本节利用用户人口统计信息来构建一个项目的动态标签，用这个新标签来描述、定义项目，也不仅仅根据项目本身的相似性来寻找目标用户群体。

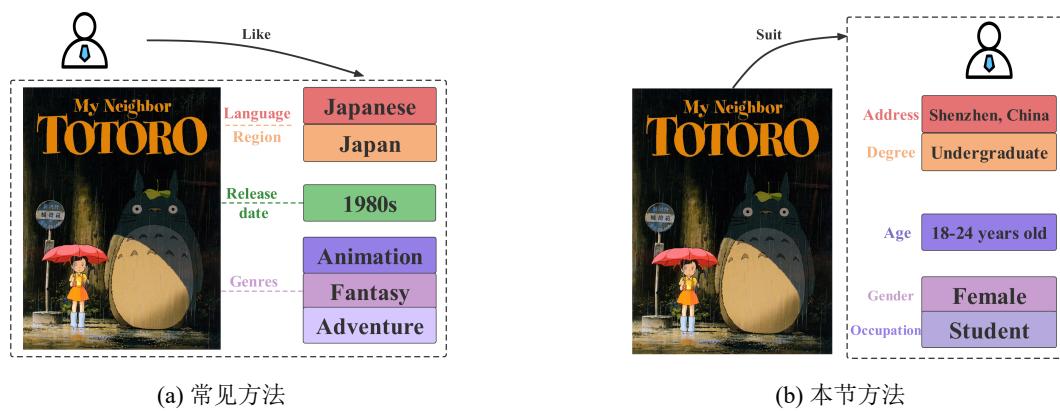


图 4-1 本节方法与常见方法在描述和推荐电影方面的比较

具有相似人口统计信息的人可能有相似的偏好，这是 DRS 的基本原则^[1,10-12]。同样，如果项目的目标用户特征更类似于某个用户，那么该项目更适合向该用户推荐。在本章方法中，推荐的依据从模型认为用户可能喜欢某个项目转变到模型认为该项目可能更适合于某用户。如图4-1所示，传统的推荐模型中，《龙猫》根据它自身的固有标签，会被描述为一部上世纪 80 年代在日本上映的奇幻冒险动画电影，这些标签是项目自身的物理特性，不会再更改。本章的方法会用一种新的方式来描述电影，《龙猫》的描述会转换

成适合于深圳居住的 18-24 岁女性本科生的观看的电影，这些出自于用户自身的特征将组合成为理想用户群的特征。与之前不同的是，新的描述方法是动态的，因为随着《龙猫》观众的增加，模型会根据不同人群的评分动态地更新这些标签。

理想用户组 (Ideal User Group, IUG) 是与项目一一对应的动态标签，IUG 的构建基于对项目打分的用户的人口统计信息的分割和组合。为了构建 IUG，首先按照人口统计信息的属性对用户进行分组，如性别、职业、年龄分别是不同的维度。随后，根据用户对某项目的评分记录计算每个组的平均评分，采用贝叶斯平均来解决某些用户群体成员较少但偏好较强的问题^[64]。最后，比较每个项目在不同属性下的平均分，组合不同维度平均评分最高的组相对应的属性，以获得 IUG 的人口统计信息。在此过程中，IUG 展示出了“组合”和“动态”两个特征，如图4-2所示。“组合”是指对 IUG 人口统计信息特征的评估不是以用户特征的整体为单位进行的，而是用一种更细粒度的方式进行评估，有效地解决了不同用户群的不同基数带来的影响。“动态”是指 IUG 是随着项目的评分人数增长而动态维护的，它会实时地反映当前最适合推荐的用户。

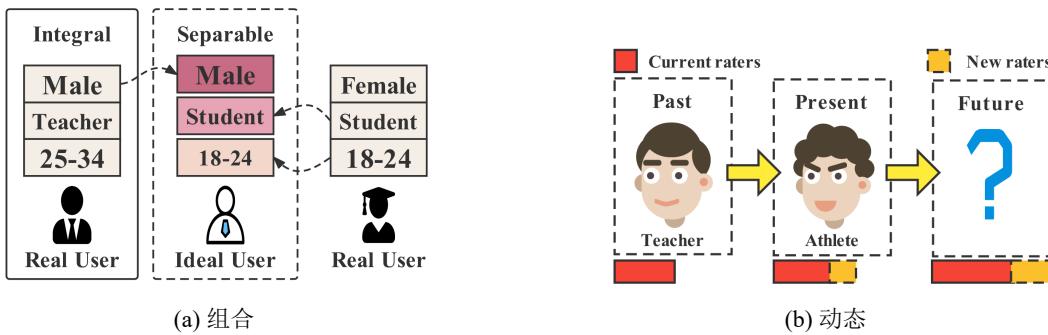


图 4-2 理想用户组的两个特征

本章提出了 IUG-CF 模型来验证本章方法的有效性，该模型将 IUG 的概念与深度协同过滤技术相结合。IUG-CF 模型包括两个部分。第一部分为所有项目构建 IUG；第二部分通过将用户人口统计信息映射到向量并计算它们之间的距离来生成推荐列表。4-3 是推荐过程的一个示例。

本章贡献如下：

- 在人口统计信息推荐领域提出了理想用户组 (IUG) 的概念。通过使用理想用户组关联项目，提供了一个新的视角描述项目，并且避免了冷启动问题。
- 提出了一种新的深度协同过滤模型，它结合 IUG 提高了推荐性能。该模型考虑了项目在不同人群中的流行度，并向具有与项目 IUG 相似的用户推荐项目。
- 介绍了一种建立用户和项目之间联系的新技术，通过拆分和合并人口统计信息，使得推荐系统能够充分利用人口统计信息，提供更好的推荐。

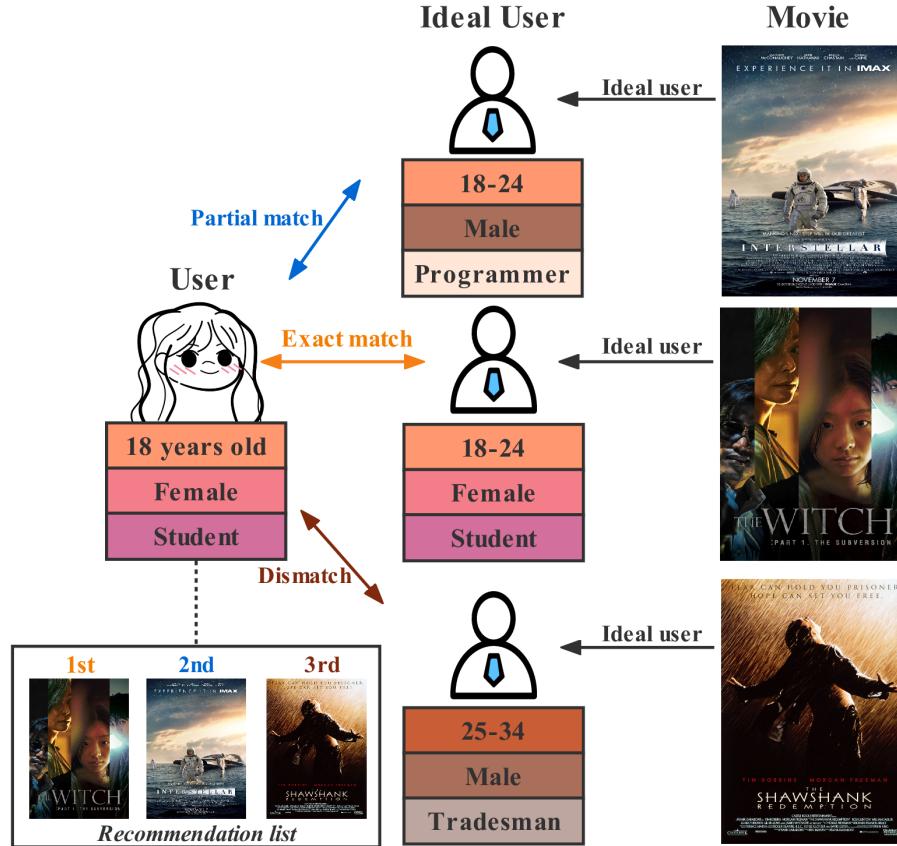


图 4-3 为用户生成推荐列表过程的示例

4.2 算法设计

本节介绍 IUG-CF 模型的总体流程，如图4-4所示。该模型由两部分组成：(1) 使用用户-项目的历史交互为所有项目构建理想用户组；(2) 根据 IUG 和用户的人口统计信息的匹配程度生成个性化的推荐。表4-1列出了本章用到的符号系统。

4.2.1 问题陈述

在 MovieLens-1M 数据集^[54]中，用户人口统计信息包括年龄、性别和职业等属性。每个属性都有特定的值，如性别属性的值有“男”和“女”。因此，本文对用户特征属性及其取值定义如下：

定义 4.1： 用户特征属性及其对应的取值分别被定义为 $q \in Q$ 和 $p_q \in \mathcal{P}_q$ 。且有：

$\forall q_g, q_h \in Q$, 使得 $\mathcal{P}_{q_g} \cap \mathcal{P}_{q_h} = \emptyset$, 以及 $\mathcal{P}_{q_1} \cup \mathcal{P}_{q_2} \cup \dots \cup \mathcal{P}_{q_n} = \mathcal{P}$ 。

定义 4.2： 用户 i 被定义为一个集合 $u_i = \{p_1, p_2, \dots\}$, 其中 p 代表其特征属性的值。

例如， $u_i = \{p_{q_1}, p_{q_2}, p_{q_3}\}$ 表示该用户拥有 p_{q_1} , p_{q_2} 以及 p_{q_3} 三种特征。 q_1 , q_2 , 和 q_3 分别代表不同的用户特征属性，如年龄、性别以及职业， p_{q_k} 表示第 k 个特征对应的取值。

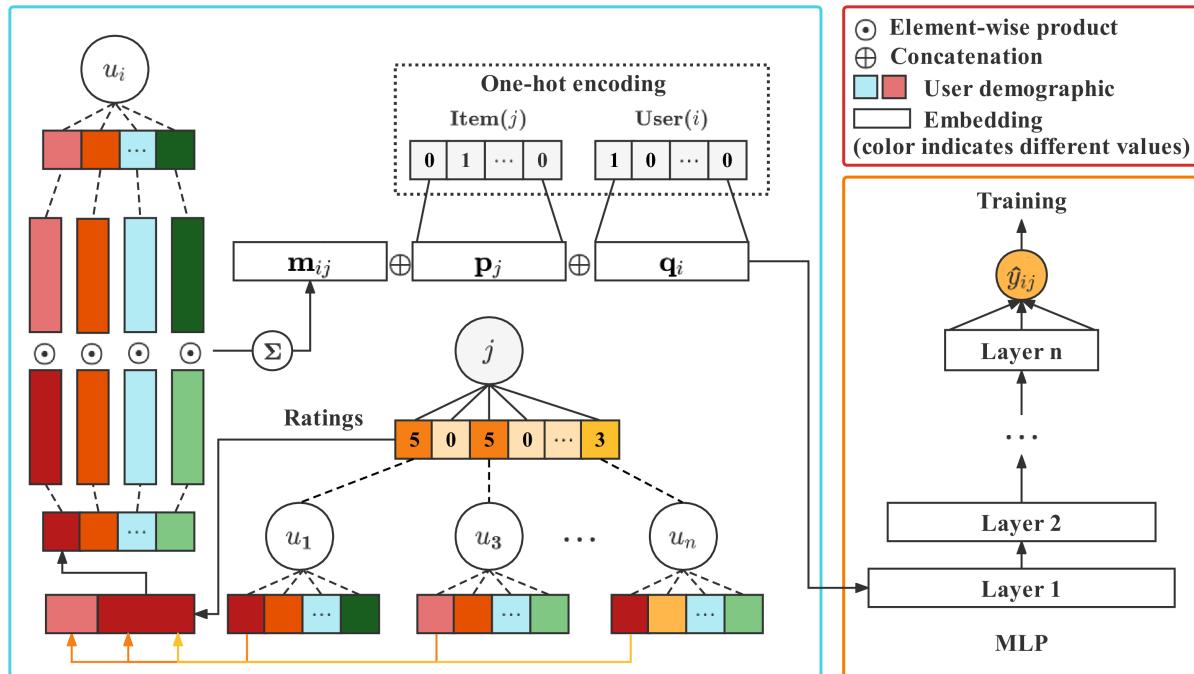


图 4-4 Overview of the IUG-CF model.

表 4-1 本章符号

符号	含义
\mathcal{I}, \mathcal{J}	用户和项目的集合
\mathcal{I}_j	对项目 j 评过分过的用户集合
\mathcal{I}_j^p	the set of users with value p who have rated item j
\mathcal{J}_i	the set of items rated by user i
\mathcal{P}	the set of all demographic values
\mathcal{P}_q	the set of demographic values of attribute $q \in Q$
\mathcal{P}_j	the set of demographic values that belong to the ideal user group of j
Q	the set of demographic information attributes
\mathcal{R}	the set of user-item interactions
\mathcal{U}'	the set of all IUs
\mathbf{p}_j	the embedding for item $j \in \mathcal{J}$
\mathbf{q}_i	the embedding for user $i \in \mathcal{I}$
\mathbf{d}^p	the embedding for value $p \in \mathcal{P}$
c_j	the average rating of item j
c_j^p	the average rating of item j on demographic value $p \in Q$
r_{ij}	the user i 's rating of item j
s_{ij}	the similarity of user i and the IU of item j
u_i	the demographics of user i
u'_j	the user member of item j 's IUG

4.2.2 理想用户组构建

IUG-CF 模型的首要任务是构建 IUG，其目标是在每个项目与其 IUG 之间建立准确的一对一映射。每个用户都具有独特的用户特征。为了构建 IUG，首先需要探索项目在不同用户特征属性上的受欢迎程度。通常衡量受欢迎的程度的指标是项目的平均评分，然而，不同人口特征群体的项目消费者分布可能是不均匀的。某些项目的受众群体人数可能远远多于其他群体。因此，本节引入了贝叶斯平均来解决这个问题。贝叶斯平均是 IMDB 网站用于确保每个电影的评分被填充足够的评分数的方法，它借鉴了贝叶斯推断的思想：既然不知道计分结果，就先估计一个值然后不断用新的信息修正，使其越来越接近正确的值。在本节中，贝叶斯平均可以让每个用户特征都对项目进行了足够的评分，使得用于比较的分值更有代表性，让比较更公平。于是，用户特征值 $p \in \mathcal{P}_q$ 对于项目 j 的平均分可以表示为：

$$c_j^p = \frac{a_j^q \times c_j + \sum_{i \in \mathcal{I}_j} r_{ij} \cdot B(p \in u_i)}{|\mathcal{I}_p^j| + a_j^q}, \quad (4-1)$$

其中， $a_j^q = \gamma \frac{|\mathcal{I}_j|}{|\mathcal{P}_q|}$ 是评分数量要扩展的数量； γ 是控制扩展数量的参数； c_j 是所有项目的平均评分； \mathcal{I}_p^j 是有 p 特征值并且对项目 j 评过分的用户集合；函数 $B(\cdot)$ 定义为：

$$B(x) = \begin{cases} 1 & (x = True); \\ 0 & (x = False). \end{cases} \quad (4-2)$$

为了评估项目在不同用户群之间的受欢迎程度，本节将每种人口统计属性视作单独的一个维度。集合 C_j 表示项目 j 在所有用户属性特征下的流行程度：

$$C_j = \{\mathbf{c}_{(j,q_1)}, \mathbf{c}_{(j,q_2)}, \dots, \mathbf{c}_{(j,q_n)}\}, \quad (4-3)$$

其中， $q \in Q$ ， $n = |Q|$ ，以及 $\mathbf{c}_{(j,q_i)}$ 是一个可以被以下方式表示的向量：

$$\mathbf{c}_{(j,q_i)} = (c_j^{p_1}, c_j^{p_2}, \dots, c_j^{p_m}), \quad (4-4)$$

其中， $p \in \mathcal{P}_q$ ， $m = \max_{q \in Q} |\mathcal{P}_q|$ 。为了得到对项目有最高评价的用户特征值，本节定义了一个函数 $M(\cdot)$ ：

$$p'_i = M(\mathbf{c}_{(j,q_i)}) = \max_{p \in \mathcal{P}_{q_i}} c_j^p. \quad (4-5)$$

通过这个方法，可以构建项目 j 的 IUG。和这个 IUG 具有同样人口统计特征的用户根据定义 1 可以表示为：

$$u'_j = M(C_j) = \{p'_1, p'_2, \dots, p'_n\}, \quad (4-6)$$

其中 n 是集合 Q 中元素的个数。

4.2.3 用户及项目建模

为了解决数据的稀疏性问题，可以通过嵌入 (Embedding) 技术取代传统的 One-Hot 编码，将所有对象嵌入到共享的低维空间中，同时保持它们的邻域邻近性^[65]。因此，本节使用 Embedding 技术对用户和项目进行编码，将用户 i 和项目 j 分别表示为 \mathbf{p}_i 和 \mathbf{q}_j 。

每个用户都有其独特的特征，根据式(4-6)，用户 i 可以被表示为由多个人口统计特征值组成的集合 $u_i = \{p_1, p_2, \dots, p_n\}$ 。特征值 $p \in \mathcal{P}$ 用对应的 embedding 向量 \mathbf{d}^p 表示。

4.2.4 偏好评分计算

为了基于用户人口统计信息建模当前用户与 IUG 之间的潜在关系，本节采用向量对应位置相乘的方法，并应用 sigmoid 激活函数进行处理再将激活函数的输出结果累加整合，如下式所示：

$$\mathbf{m}_{ij} = \sum_{g=1}^{|Q|} \sigma(\mathbf{d}^{p_g} \odot \mathbf{d}^{p'_g}), \quad (4-7)$$

其中 \odot 表示对应位置相乘， σ 表示 sigmoid 激活函数。向量 \mathbf{d}^p 对应用户特征 p 的特征 embedding，且有 $p_g \in u_i$ and $p'_g \in u'_j$ 。

在融合用户和项目的 embedding 时，通常采用向量拼接操作^[66]。然而，简单的拼接操作无法完全捕获用户与项目之间的所有潜在特征^[45]。为了更好地提取用户和项目之间的隐特征并探究用户与项目之间的关系，本节采用标准多层感知机 (Multilayer Perceptron, MLP) 来学习非线性交互函数。因此，用户 i 对项目 j 的偏好得分的计算方法如下式所示：

$$\hat{y}_{ij} = MLP(\mathbf{p}_j \oplus \mathbf{q}_i \oplus \mathbf{m}_{ij}), \quad (4-8)$$

其中 \oplus 表示向量连接， \mathbf{p}_j 和 \mathbf{q}_i 分别是项目 j 和用户 i 的特征 embedding。

MLP 的隐藏层采用了 ReLU 函数作为激活函数。在输出层，应用 sigmoid 函数将偏好得分映射至 $[0,1]$ 区间内。这一做法不仅提供了用户选择某个项目的概率估计，而且更便于将其应用于损失函数中进行优化计算。

本节的目标是赋予与 IUG 匹配度较高的用户更高的权重。为了实现这一目标，首先通过曼哈顿距离计算用户与 IUG 之间的相似性。接着，从性别、年龄和职业等用户特征属性维度汇总得到相似度得分 s_{ij} ，以获得一个整体相似度衡量指标：

$$s_{ij} = \sum_{q \in Q} \sum_{k=1}^m |\mathbf{d}_k^{p_{iq}} - \mathbf{d}_k^{p_{jq}}|, \quad (4-9)$$

其中 n 表示用户人口统计特征属性的数量， m 表示用户特征 embedding 的维度。这个相似度的指标能够衡量用户和 IUG 之间的特征的匹配程度。最后，将相似度 s 投影到适当的区间，用于调整偏好评分，如下所示：

$$\tilde{y}_{ij} = \frac{\hat{y}_{ij}}{\exp(s_{ij})}. \quad (4-10)$$

通过对比多个用户对同一个项目的偏好评分 \tilde{y} , 分值更高的用户与项目的 IUG 在人口统计信息方面更加匹配, 更接近于该项目最适合推荐的理想用户。

4.2.5 模型训练

为了突出隐式数据的二元属性, 本章采用成对 (pair-wise) 训练的方法学习模型^[45]。成对训练将用户给予高评分的项目视为正例, 而没有被用户未评价的项目被视作负例。训练数据以成对的样本的形式出现, 每对样本包括一个正例和一个负例, 模型通过学习正负例之间的差异来提升模型准确性。训练模型的损失函数为:

$$L = - \sum_{(i,j) \in \mathcal{W}} \log \tilde{y}_{ij} - \sum_{(i,k) \in \mathcal{W}^-} \log(1 - \tilde{y}_{ik}), \quad (4-11)$$

其中, i , j , k 分别代表用户 i , 正例 j 以及负例 k 。 \mathcal{W} 和 \mathcal{W}^- 分别是正例和负例的集合, 定义为:

$$\mathcal{W} = \{(i, j) | i \in \mathcal{I} \wedge j \in \mathcal{J}_i\}, \quad (4-12)$$

$$\mathcal{W}^- = \{(i, k) | i \in \mathcal{I} \wedge k \in \mathcal{J} \setminus \mathcal{J}_i\}, \quad (4-13)$$

其中, \mathcal{J}_i 是用户 i 评分过的项目的集合, \mathcal{J} 是所有项目的集合。在训练过程中, 通过最小化损失函数来提升模型对用户-项目的偏好评分的准确预测能力。为进一步强调高评分用户特征 embedding 与项目的 IUG 之间的相似性, 本节在损失函数中添加了正则化项。该正则化项衡量了当前用户特征嵌入与项目 j 的 IUG 之间的差异, 具体定义如下:

$$w_{ij} = \sum_{g=1}^n (\mathbf{d}^{p_{ig}} - \mathbf{d}^{p_{Jg}}). \quad (4-14)$$

通过结合式(4-11) 和式(4-14), 得到新的损失函数:

$$\begin{aligned} L = & - \sum_{(i,j) \in \mathcal{W}} (\log \tilde{y}_{ij} - \lambda ||w_{ij}||_1) \\ & - \sum_{(i,k) \in \mathcal{W}^-} (\log(1 - \tilde{y}_{ik}) + \lambda ||w_{ik}||_1) \end{aligned} \quad (4-15)$$

λ 是正则化系数。

算法 1 将展示了本章提出的模型的执行流程。

4.3 实验结果与分析

本节展示了 IUG-CF 模型在三个数据集上的实验结果。关于实验采用的数据集特征、评估指标体系以及实验参数的具体设定, 将在本章后续的不同部分分别详细阐述, 具体为第4.3.1节、第4.3.2节和第4.3.3节。同时, 对于该模型涉及的 IUG 的两个关键特性的验证, 将在第4.3.4节进行论述。随后, 在第4.3.5节内, 将对 IUG-CF 模型内部性能表现进行对比, 并与当前领域内的先进方法进行比较分析。

Algorithm 1 IUG-CF 算法流程

Input: \mathcal{R} : 评分记录; \mathcal{D} : 所有用户特征; η : 学习率; λ : 正则项系数; γ : 扩张系数; Epochs: 训练轮次;

Output: Θ : 模型中的所有参数;

```

1: 初始化  $\Theta$  中的所有参数;
2: for  $j \in \mathcal{J}$  do ▷ 构建 IUG
3:   for  $q \in \mathcal{Q}$  do
4:     for  $p \in \mathcal{P}_q$  do
5:        $c_j^p \leftarrow$  use (4-1) with  $\mathcal{R}, \mathcal{D}$ , and  $\gamma$ ;
6:     end for
7:   end for
8:   Generate the IUG via (4-3) - (4-6);
9: end for
10:  $\mathcal{U}' \leftarrow$  all IUs;
11: for epoch in range(Epochs): do
12:    $\tilde{y}_{ij} \leftarrow$  use (4-7) - (4-10) with  $\mathcal{U}'$  and  $\mathcal{D}$ ;
13:    $w_{ij} \leftarrow$  use (4-14) with  $\mathcal{U}'$  and  $\mathcal{D}$ ;
14:    $L \leftarrow$  use (4-15) with  $\tilde{y}_{ij}, w_{ij}$ , and  $\lambda$ ;
15:   使用学习率为  $\eta$  的 Adam 优化参数;
16: end for
17: return  $\Theta$  中的所有参数;

```

4.3.1 数据集

本章采用了三个不同规模的真实世界数据集: MovieLens-100k、MovieLens-1M 以及 KuaiRec。表4-2汇总了这些数据集的基本统计信息, 其中 $|\mathcal{I}_j|$ 和 $|\mathcal{J}_i|$ 分别表示用户的平均评分数量和项目被评分的平均数量。MovieLens 数据集在推荐系统的研究领域得到了广泛应用, 提供了丰富的用户对电影的显式评分数据以及其他附加用户信息, 如性别、年龄、职业等^[54]。而 KuaiRec 则是一个源自短视频分享移动应用快手的推荐日志的真实数据集^[67]。

在 MovieLens 数据集中, 本节选择了用户的性别、年龄和职业等用户特征信息来构建 IUG。在 MovieLens-1M 数据集中, 用户年龄以年龄段形式展示, 为保证一致性, MovieLens-100k 数据集中用户的年龄被转换为与 MovieLens-1M 数据集相同的年龄段表示。而在 KuaiRec 数据集中, 用户的特征信息已经过脱敏处理, 并被标记为 One-hot 编码。因此, 在本节中, 考虑到这 One-hot 编码的取值范围与 MovieLens 数据集中的性别、年龄和职业特征相似, 选择了数据集中的前三项编码作为分析对象。

4.3.2 评估指标

本章采用推荐算法常用的“留一法”评估方式来评估模型的性能^[52,68-70]。具体而言, 对每位用户随机选取一项积极项目作为测试项目, 并将余下的评过分的项目作为训练数

表 4-2 数据集基本信息

数据集	用户数	项目数	评分数	稀疏度	用户平均打分数量	项目平均被打分数量	用户特征属性
Movielens-100k	943	1682	100,000	93.70%	106.04	59.45	年龄, 性别, 职业
Movielens-1M	6040	3706	1,000,209	95.53%	165.60	269.89	年龄, 性别, 职业
KuaiRec	7176	10728	12,530,806	83.70%	1746.20	1168.05	脱敏信息

据。为了确保对高评分交互预测的有效性和显著性^[6], 本节设定测试项目的门槛为评分必须高于 4 分 (满分为 5 分)。对于 KuaiRec 数据集, 由于缺少显式评分, 本节将数据集中 “watch_ratio ≥ 3 ” 的交互记录中的项目作为测试项目。随后, 从用户未交互过的项目中随机抽取 100 个项目。采用所提出的模型根据计算出的偏好评分对这 101 项 (测试项以及 100 个随机项) 进行排序。为了确保所有模型间的公平比较, 所有对比模型也同样采用了相同的评估策略。

本节用命中率 (Hit Ratio, HR) 和归一化折损累计增益 (Normalized Discounted Cumulative Gain, NDCG) 两个指标来评估不同推荐模型的推荐结果。HR@ k 表示在推荐列表的前 k 项中, 测试项出现的次数在所有测试项中的百分比。NDCG@ k 则更关注测试项在推荐列表中的排名, 越靠前的测试项有更高的权重。两个指标的计算公式如下:

$$\text{HR}@k = \frac{\#hits@k}{\#users}, \quad (4-16)$$

$$\text{NDCG}@k = \frac{1}{\#users} \sum_{l=1}^{\#users} \frac{1}{\log_2(p_{l,k} + 1)}, \quad (4-17)$$

其中 $p_{l,k}$ 表示第 l 个用户的测试项在其对应的推荐列表中的位置。

4.3.3 实验细节

IUG-CF 模型使用 3.7 版本 Python 编写, 依赖于 keras 和 Tensorflow 2.5.0 版本。实验在搭载 Intel(R) Core(TM) i7-11800 CPU @ 2.30 GHz 处理器的 Windows 11 操作系统环境下进行。所提出的模型参数采用均值为 0、标准差为 0.01 的正态分布进行随机初始化。本节采用成对损失函数 (式(4-11)) 训练模型, 并使用 mini-batch Adam 优化器^[71]进行优化, 学习率设定为 0.001, 批次大小设为 256。用户, 项目以及用户特征的 embedding 维度分别设置为 512 和 128, 还配置了一个维度为 512 的单层多层感知机, l_1 -范数正则化系数 λ 被设定为 0.001。对于 Movielens-1M、Movielens-100k 及 KuaiRec 数据集, 其扩展系数 γ 分别设定为 0.3, 0.8 和 0.5。

4.3.4 特性验证

本节验证 IUG 的两个特性: “组合”, “动态”。

1) 组合

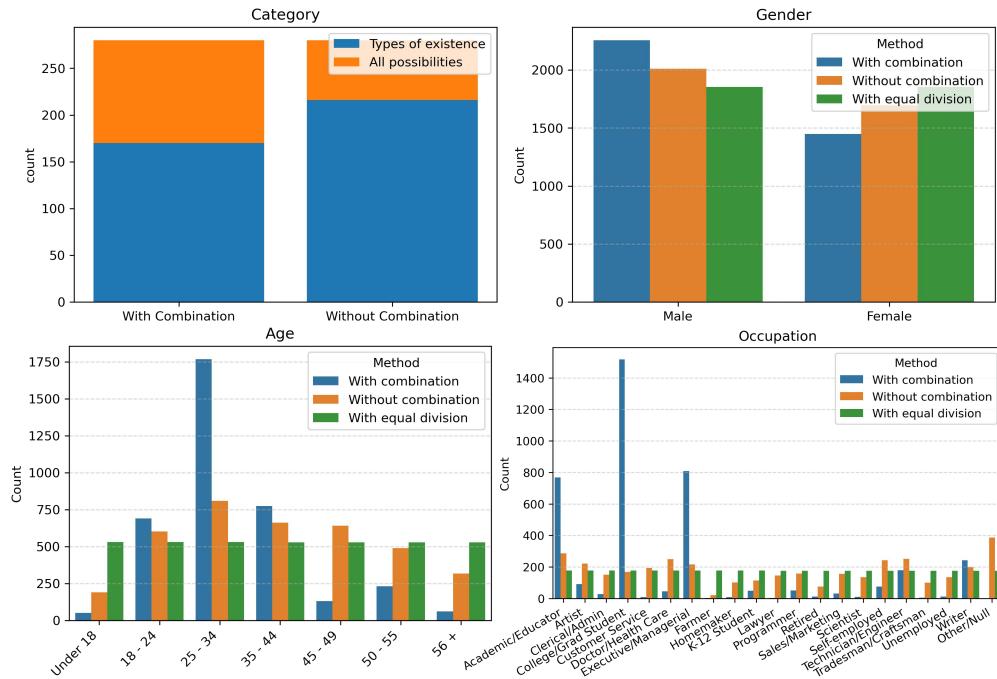


图 4-5 MovieLens-1M 数据集中采用不同构建方法的用户特征属性的分布

为了验证第一个特性：“组合”，本节将实验设置为比较不同的方法下，IUG 的种类以及性别、年龄、职业三种属性的计数。IUG 的种类是指数据集中所有项目的 IUG 汇总并去重后的数量，另外三种属性的计数是所有 IUG 的对应属性的数量求和。本节用于比较的三种方法如下：本节提出的将用户特征分离并组合的方法、随机平均分布 IUG 的方法以及将用户特征视为一个不可分割的整体来建立 IUG 的方法。下文分别将三种方法简称为组合法，随机法，整体法。

图4-5展示了比较结果。在数据集中，共选取了三种人口特征信息属性，年龄、性别、职业取值的可能性分别是 7 种，2 种，21 种，所以 IUG 的取值可能性是其乘积 294 种。从图中可以观察到，组合法相较于整体法，IUG 的种类有所减少，其原因是组合法通过更细粒度的方法构建 IUG，使得模型学习到的不同人口统计信息特征上的喜好更加精确。后续各属性的分布情况进一步证明了该结论。

随机法是将各属性平均分配，从三个属性的分布图可以观察到，整体法和随机分配的差异并不大，而组合法会有比较明显的某属性比较突出。说明整体法并不能准确地学习到各属性单独的喜好，无法从项目和用户之间的联系中学习到偏好，所以其结果与随机得到的比较接近。而组合法成功地识别出了某些反映出较强偏好的属性。

该实验可以得到以下三个结论：（1）组合法相较于整体法降低了 IUG 的多样性，使得个性化推荐更加集中于某些活跃用户；（2）整体法构建的 IUG 在每个特征属性上的分布接近于随机法，而组合法扩大了每个属性的分布差异；使得推荐结果更加准确且有针对性；（3）组合法排除了职业分布中的模糊因素，例如“其他”，从而更准确地确定了项目的最匹配用户。

2) 动态

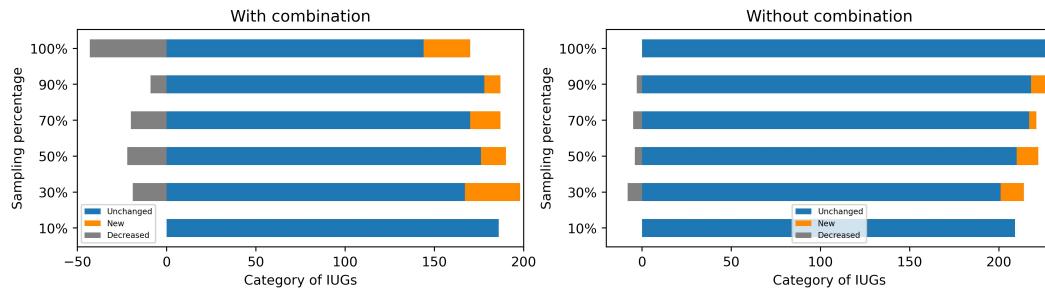


图 4-6 Variation chart of IUG categories with sampling percentage.

为了验证第二个特性：“动态”对推荐性能的影响，本节将实验设置为将训练样本按评分时间升序排序后，先用 10% 的训练样本来统计 IUG 的种类，随后逐步增加使用训练样本的比例，在此过程中观察每一次增加样本后 IUG 种类的变化情况。增加训练样本后 IUG 的会被分为三类，分别是保持不变、新种类、不再存在的种类。由于随机法会固定 IUG 的种类，所以在此不作比较。

图4-6展示了实验的结果。可以从图中观察到，每一次增加训练样本，组合法构建的 IUG 的种类变化的幅度最大会接近 50 种，且当训练样本达到一定数量后，IUG 的种类逐步减少。而组合法的 IUG 种类并未呈现出较大的变动趋势，且种类不断增多。这说明相较于组合法，整体法无法消化越来越多的用户评分信息，进而无法从打分用户的人口统计信息中归纳出最适合项目的用户特征。而组合法在此过程中不断地精简 IUG 的种类，在细粒度的构建方式中动态地更新最适合项目的用户群体的特征。

为了进一步证明上述观点，本节也在模型中用三种方法构建的 IUG 进行了实验，并用 HR 和 NDCG 两个指标比较了结果。根据图4-7展示的结果，组合法显示出更好的推荐性能，其次是整体法。这也证明了组合法在不断优化每个项目的最佳用户，并且越多的训练样本就能得到越准确的结果。随着训练样本的增加，组合法可以更准确地确定项目的最理想的用户群，因此采用组合法的 IUG-CF 模型的推荐结果更加准确。

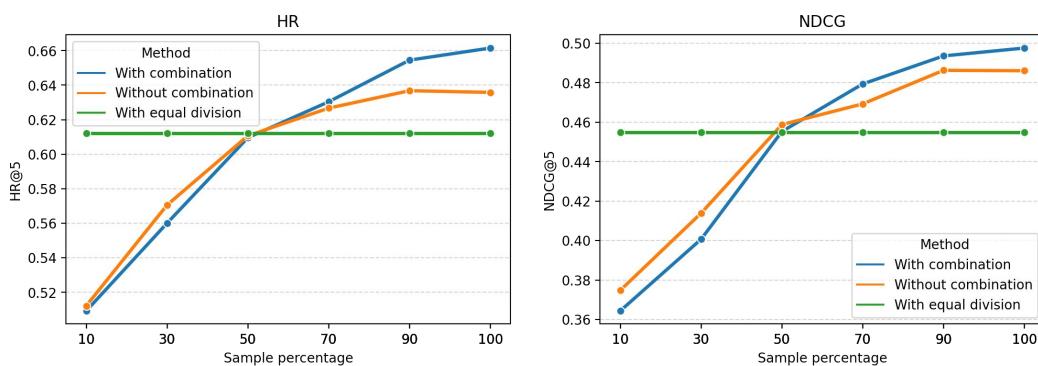


图 4-7 MovieLens-1M 数据集中三种方法在不同采样百分比下评估指标的变化趋势

4.3.5 模型对比

本节提出了另外六个版本的 IUG-CF 子模型，这些子模型在内部比较实验中用于探索最佳的模型结构。同时，与这些子模型的比较也将作为该模型的消融实验。随后本节介绍了六种基线模型，这些模型将和 IUG-CF 进行比较。

1) IUG-CF 子模型

本节进行了内部比较，通过替换或删除模型的部分结构来对比性能。内部比较包括以下模型：

- IUG-CF: 本章提出的完整模型。
- IUG-CF-W: 构建 IUG 时采用整体法的版本。
- IUG-CF-S: 构建 IUG 时仅使用部分用户特征的版本（本节仅使用年龄属性）。
- IUG-CF-A: 删除了基于 IUG 和用户相似度的预测分数调整的版本（式4-9,4-10）。
- IUG-CF-L: 删除了基于 IUG 和用户之间的特征嵌入距离缩放损失函数的部分的版本（式4-14,4-15）。
- IUG-CF-M: 加入用户和项目的特征 embedding 的点积 \mathbf{n} ，修改了式4-8为 $\hat{y}_{ij} = MLP(\mathbf{p}_j \oplus \mathbf{q}_i \oplus \mathbf{m}_{ij} \oplus \mathbf{n}_{ij})$ 的版本。
- IUG-CF-B: 在式4-1中将 a_j^q 设置为 0，即删除贝叶斯平均的版本。

2) 基线模型

本节主要比较基于用户特征的模型和代表性的深度学习模型。

- ItemAve^[38]: 这是一个基于用户特征的模型，旨在解决冷启动问题。它计算每个用户特征属性的平均评分，然后推荐相同属性上最高评分的项目。
- NeuMF^[45]: 结合了 MF 和 MLP 的协同过滤模型，用于建模线性和非线性用户-项目交互，它能够从数据中学习任意函数。
- LRML^[48]: 具有隐式反馈的 DL 模型，引入了一种隐式关系度量学习方法，学习用户-项目交互的独特和最佳关系转换。
- Wide&Deep^[51]: 将用户特征之类的多种特征映射到特征 embedding 中，并同时训练线性模型和深度模型，以提高模型的泛化能力的模型。
- CFFNN^[50]: 结合了用户和项目特征来理解用户对项目的偏好的 DL 模型。
- HybridNNMF^[47]: 将全局神经网络与本地神经块相结合，以捕获用户和项目之间的线性和非线性关系的混合模型。
- IFDNAMF^[49]: 引入了信息融合过程中的属性信息，并使用深度神经网络学习用户和项目之间的高阶交互的模型。
- NHF^[46]: 将 MF 和 MLP 相结合，并整合了用户人口统计信息和项目辅助信息的混合模型。
- KTUP^[72]: 同时学习推荐和知识图谱，将关系信息转移以增强对用户偏好的理解的图神经网络模型。

- CoupledCF^[52]: 联合学习用户和项目之间和内部的显式和隐式耦合，将用户人口统计信息与项目额外信息结合，使用卷积神经网络学习用户-项目耦合的模型。

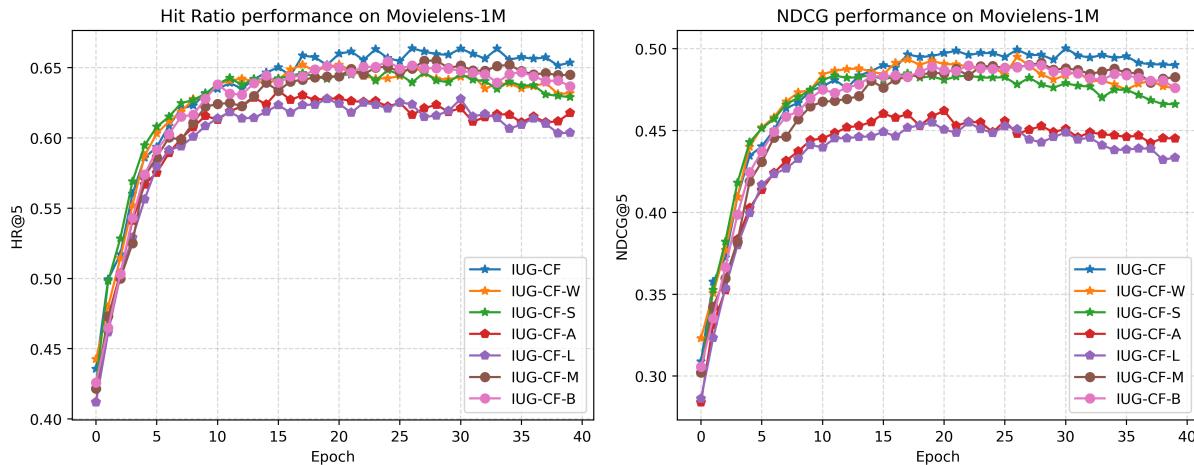


图 4-8 在 MovieLens-1M 数据集上使用 HR@5 和 NDCG@5 指标进行的内部比较

3) 结果分析

如前文所述，本节介绍了六个 IUG-CF 模型的分支版本。IUG-CF-W 和 IUG-CF-S 在构建 IUG 的方法上有所不同，IUG-CF-A，IUG-CF-L 和 IUG-CF-B 分别删除了完整模型中的某步骤，IUG-CF-M 增加了用户和项目的嵌入点积作为 MLP 层的输入。图4-8展示了在 MovieLens-1M 数据集上各版本模型的表现，评估的指标是 HR 和 NDCG。

从图可以看出，IUG-CF，IUG-CF-W，IUG-CF-S 的性能是逐步降低的，进一步证明了在 IUG 的构建方面，本章提出的细粒度方法优于粗粒度的整体法，同时，整体法也优于仅选择部分特征的构建方法。这也证明了越多的人口统计信息特征可以为模型提供更多有价值的信息。

通过观察 IUG-CF、IUG-CF-A、IUG-CF-L、IUG-CF-B 的比较结果，可以得出以下结论：首先，在本章方法中，最重要的组成部分之一是通过 IUG 与用户之间的特征 embedding 距离来调整损失函数的部分，删除这部分的 IUG-CF-L 表现出了最差的性能。这也说明了模型学习用户偏好依赖于用户与 IUG 之间的差别，构建 IUG 来建立项目与用户之间的联系是合理的。其次，IUG-CF 之外表现最好的 IUG-CF-B 模型删除了贝叶斯平均的部分。贝叶斯平均在本章用于建立一个公平的比较环境，因为项目的评分人群的基数不同，过少的样本会造成不公平的比较。而随着数据集的规模扩大，项目的各种评分人群的数量已经逐渐趋于平均值。最后，IUG-CF 的优异表现证明了本章提出的平滑预测分数、调整损失方法、建立公平的评分比较环境方法是有效的。

此外，通过比较 IUG-CF 与 IUG-CF-M 比较的结果可知，使用 GMF^[45]结构建模更多的线性交互并不会提高推荐性能。一个可能的原因是在 IUG-CF 模型中，用户特征嵌入之间的内积已经捕捉到足够的线性交互。

IUG-CF 模型与4.3.5节中列出的基线模型也进行了比较。表4-3展示了包括 IUG-CF 在内的所有模型生成前 10 个推荐列表后，用 HR、NDCG 两个指标评估的结果，其中最

表 4-3 各模型在 MovieLens-100k 和 MovieLens-1M 上的性能比较

Datasets	MovieLens-100k		MovieLens-1M	
	Metrics	HR@10	NDCG@10	HR@10
ItemAve	0.3278	0.1632	0.4822	0.1820
NeuMF	0.6807	0.4201	0.7260	0.4415
LRML	0.6363	0.3665	0.7281	0.4303
HybridNNMF	0.7260	0.4200	0.7380	0.4560
Wide&Deep	0.7400	0.4509	0.7387	0.4490
CFFNN	0.7012	0.4077	0.7511	0.4582
IFDNAMF	0.7481	0.4591	0.7558	0.4680
NHF	0.7784	0.4915	0.7715	0.4950
KTUP	0.7614	0.5138	0.7818	0.5156
CoupledCF	0.7895	0.5117	0.8003	0.5388
IUG-CF	0.7931 (.0017)	0.5399 (.0040)	0.8022 (.0013)	0.5408 (.0017)

佳的表现用粗体进行了标记，括号内是标准差。

显然，IUG-CF 在两个数据集上的表现均优于所有基线模型，与深度协同过滤的经典模型 NeuMF 相比，NDCG@10 指标增加了 22.79%，表明将用户特征纳入神经网络是有效的。通常，将用户特征与深度学习相结合的模型会将用户特征嵌入化，并将这些嵌入与用户或物品嵌入连接起来作为 MLP 层输入。Wide&Deep、IFDNACF、NHF 和 CoupledCF 等模型会将用户特征映射为嵌入，并将其整合到基于深度学习的推荐模型中。然而，仅简单拼接用户特征 embedding 然后将其作为输入限制了其价值，这样并不能学习到用户特征与项目的关联，只能利用到用户间的相似性。因此，充分挖掘用户特征 embedding 的 IUG-CF 表现出了优于这些模型的效果。

IUG-CF 模型在将用户特征与深度学习相结合之前，使用用户特征进行数据挖掘，并学习用户-IUG 之间的关联以进行基于用户特征的推荐。比较的结果证明了上述方法的合理性。

KuaiRec 是一个大规模且密集的短视频数据集，为了评估 IUG-CF 在不同领域和不同大小的数据集上的有效性，本节使用了 KuaiRec 数据集进行了进一步实验。由于 NeuMF 是神经协同过滤中最经典的基线模型的代表，CoupledCF 模型在之前的比较中性能最接近 IUF-CF，所以本节选择这两个模型作为对比模型。此外，也将评估的粒度细化，将 K 设置为从 2 到 10 递增，以观察性能的变化。表4-4显示了比较的结果。结果表明，随着 K 的减少，IUG-CF，CoupledCF 与 NeuMF 相比，两个指标差距越来越大。说明利用用户特征信息在生成更个性化和准确的推荐方面是有效的。IUG-CF 和 CoupledCF 的比较说明在神经协同过滤模型中先对用户特征信息进行数据挖掘是有效的，可以获取更多用户与项目间的关联信息。因此，对于视频推荐，将物品与用户特征进行映射以构建 IUG 也是有效的。基于以上的实验结果，验证了本章提出的 IUG-CF 模型在基于用户特征推荐的有效性和合理性。

表 4-4 三个模型在 KuaiRec 上的性能比较

K	HR@K			NDCG@K		
	NeuMF	CoupledCF	IUG-CF	NeuMF	CoupledCF	IUG-CF
2	0.3948 (.0042)	0.4612 (.0019)	0.4809 (.0027)	0.3487 (.0017)	0.4174 (.0017)	0.4273 (.0005)
4	0.5578 (.0014)	0.6288 (.0031)	0.6391 (.0009)	0.4235 (.0035)	0.4931 (.0015)	0.5027 (.0004)
6	0.6650 (.0018)	0.7265 (.0019)	0.7310 (.0015)	0.4635 (.0012)	0.5238 (.0013)	0.5374 (.0005)
8	0.7351 (.0015)	0.7922 (.0022)	0.7943 (.0019)	0.4851 (.0019)	0.5452 (.0010)	0.5580 (.0006)
10	0.7918 (.0023)	0.8405 (.0011)	0.8418 (.0009)	0.5015 (.0020)	0.5602 (.0010)	0.5713 (.0006)

表 4-5 MovieLens 数据集不同比例增量与原有数据的对比

数据集	用户量	项目量	原有交互量	总交互量	1% 增量	5% 增量	10% 增量
MovieLens-100k	943	1682	100,000	1,586,126	15,861	79,306	158,613
MovieLens-1M	6040	3706	1,000,000	22,384,240	223,842	1,119,210	2,238,424

4.3.6 填充实验

第三章介绍了数据集存在的选择偏差。选择偏差是指用户在使用推荐系统时，可能更倾向于与系统推荐的某些项目进行交互，而忽视其他物品，并且不愿对不喜欢的项目打分。选择偏差导致了显式反馈数据集中数据的非随机缺失，模型通过数据集学习时将会学习到不完整的用户偏好，因为训练样本中缺失了用户不喜欢的，本应该打低分的交互记录。

在本章提出的方法中，IUG 的建立依赖于用户的显式历史评分记录。缺失掉的低评分数据将使得某些用户群体对特定项目的平均评分比真实的更高，因此，选择偏差的存在也会影响 IUG 的准确性。本节的实验将结合第三章中提出的方法，结合本章提出的 IUG 的概念，进一步验证第三章提出的方法对选择偏差的抑制是否有效以及是否能够将其运用于协同过滤模型。

本章使用的 MovieLens 数据集的稀疏度高达 90% 以上，因此按照比例进行无趣项评分的填充时应注意填充量与原有的比例。表4-5展示了不同程度增量与原有交互数量的对比。可以从表中观察到：体量较小的 MovieLens-100k 数据集在 5% 的增量下就已经接近与原有体量，而 MovieLens-1M 数据集在 5% 增量时已经超出了原有体量。然而，由于无趣项的填充都是低值，在本章的 IUG 构建中，大量的低值会使得评分的权重淡化，导致无法学习到真实偏好。因此，考虑到实验环境与数据集的大小，本小节实验将选用 MovieLens-100k 数据集。本小节的第一个实验设置为：将增量从 0 至 5% 按 1% 为步长递增，观察 IUG 种类的变化情况。结果如图4-9所示。

从图中可以观察到，IUG 的种类随着训练数据的小幅增加呈逐渐减少的趋势。根据4.3.4小节的结论，IUG 的种类数减少是因为模型学习到更细粒度、更准确的用户偏好，动态地维护了与项目最匹配的用户群体。并且，随着数据量的增加，IUG 种类的变化程度也逐渐稳定，这是因为数据集被填充了大量低分值，当数据增量达到 10% 时，低分值已经超过了整个训练样本的 60%，这对于模型的学习可能是有害的。为了验证该想法，

本节设置了多种级别的增量，通过 HR 和 NDCG 两个评价指标评估 IUG-CF 模型在填充后的数据集上的效果。结果如图4-10所示。可以观察到当数据少量填充时，IUG-CF 的模型性能有所提升，随后开始下降，达到 5% 增量时模型的性能已经急剧下降，因为此时填充的数量已经达到 79306，已经达到原有数据量的 79%，过多的低分值影响了 IUG 的构建质量。

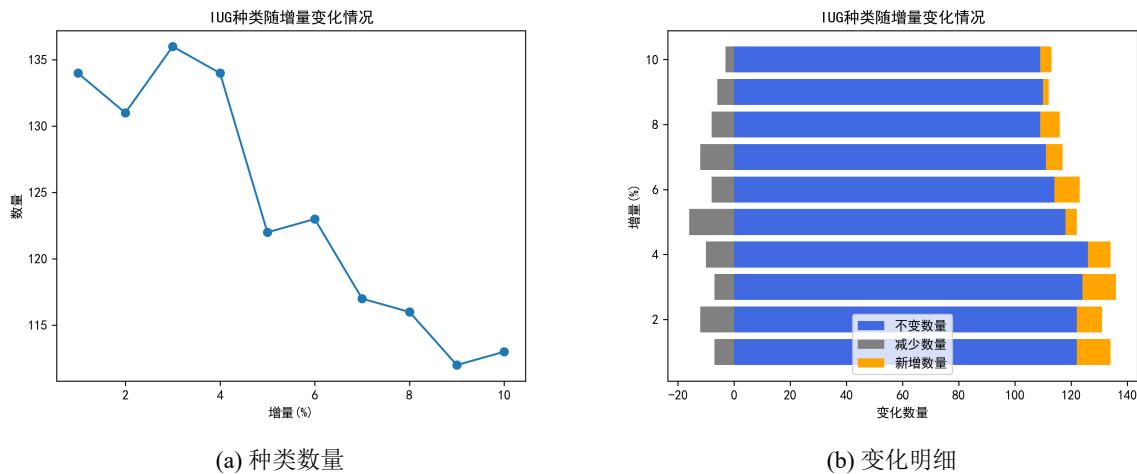


图 4-9 MovieLens-100k 中 IUG 种类及明细随增量变化情况

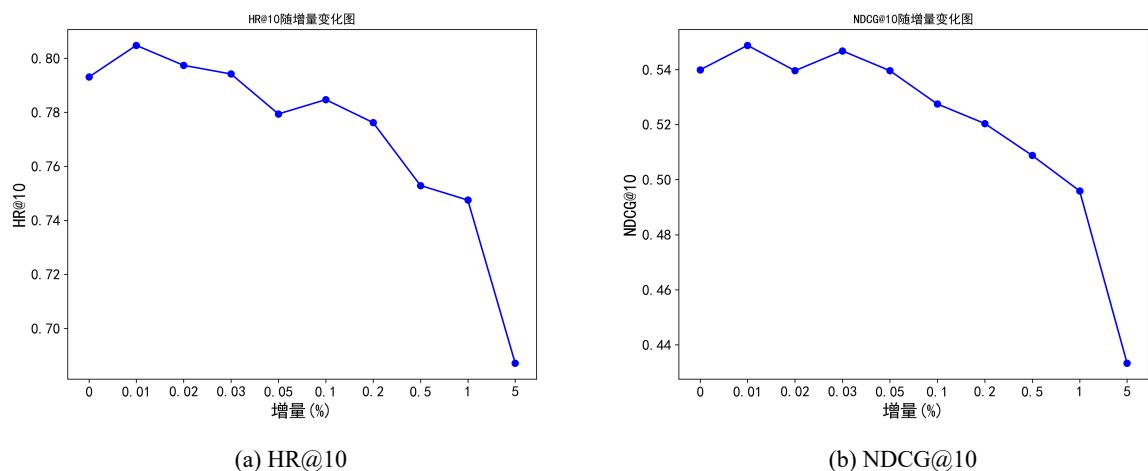


图 4-10 MovieLens-100k 中 IUG-CF 性能随增量变化情况

此外，第三章提出的算法还有另一个作用。本章的模型采用的是成对训练，负例是在用户未交互过的项目中随机选取，因为这些项目都被假定为用户不喜欢的项目。然而，同样是数据集中存在的选择偏差会影响该过程。从图3-1可知，数据集中存在着部分项目是用户可能喜欢的，但是由于用户不知道其存在，所以还未对其进行交互。如果这部分项目被随机选为负例，用户的喜好将会被扭曲。因此，将选择负例的过程与第三章的方法进行结合，仅在识别出的无趣项中选择负例，将提升负例的质量，这部分负例是用户不喜欢的项目的可能性大幅增加，模型学习到的用户偏好也将更加准确。

表 4-6 MovieLens-100k 中不同的负例选择方法的对比

模型	随机选取负例		无趣项中选取负例	
	HR@10	NDCG@10	HR@10	NDCG@10
NeuMF	0.6807(.0021)	0.4201(.0023)	0.7100(.0023)	0.4361(.0036)
IUG-CF	0.7931(.0017)	0.5399(.0040)	0.8101(.0015)	0.5512(.0047)

经过在 NeuMF 和 IUG-CF 上的实验，证明了此结论是正确的，其结果如表4-6所示。

4.4 本章小结

本章介绍了当前神经系统过滤面临的问题，并基于当前问题提出了理想用户组(IUG)的概念，IUG 是根据给项目打分的用户群体的人口统计信息构建的一个项目的动态标签。随后，基于 IUG 的概念，本章提出了一个基于理想用户组的深度协同过滤模型(IUG-CF)。该模型首先读取数据集中的用户-项目交互，基于评分用户的人口统计信息为所有项目构建对应的理想用户组，随后将用户与项目的 IUG 的 embedding 的距离融入深度协同过滤，以学习用户与 IUG 之间的非线性关系，最后通过模型计算用户与项目之间的偏好分数，排序后生成用户的推荐列表。

本章的实验证明了提出模型的有效性。

第 5 章 总结与展望

5.1 总结

随着互联网的迅速发展，获取高质量的个性化信息已经成为人们日常生活的必备需求，推动着推荐系统领域不断向前发展。然而，推荐系统在实践中面临着一些挑战和问题。其中之一是数据的非随机缺失所带来的选择偏差，这导致推荐模型只能学习到用户偏好的部分信息，无法完整反映用户的真实需求。另一个问题是现有的推荐模型通常只考虑了项目和用户的自身属性，缺乏对二者之间准确关联关系的建模，导致推荐结果与用户需求不够匹配。针对这些问题，本文提出了两个的工作：

(1) 针对数据的非随机缺失和选择偏差问题，本文提出了一种基于用户评分习惯的方法，结合了显式反馈和隐式反馈的优点，以及多种低值填充策略，以更好地补全用户偏好信息，提高推荐准确性。具体来说，本工作提出的方法分为两个部分，分别是识别无趣项和填充无趣项。首先，在识别无趣项部分，本工作提出了挖掘用户评分习惯的公式，将用户的评分习惯作为加权矩阵分解的权重的重要组成部分，提高了识别无趣项的准确性。其次，在填充无趣项部分，本工作引入了项目质量分的概念，将项目分为两个部分，分别使用零值和低值进行填充，相较于传统的填充方法，该方法更能让模型学习到准确、完整的用户特征。在公开数据集上的实验结果表明，在使用前偏好建模和无趣项填充两个阶段，本文提出的方法均优于传统方法，说明该工作提出的结合显式反馈和隐式反馈的优点的方法是有效的。

(2) 针对当前推荐系统中缺乏项目与用户之间准确关联关系的问题，本工作提出了理想用户组的概念，并将其融入深度协同过滤模型，以提高推荐性能。具体来说，本工作首先提出理想用户组(IUG)的概念，该概念是项目的动态项目特征，指出了最适合被推荐该项目的特定用户群，通过用户的评分记录和人口统计信息共同构建每个项目的IUG。随后，基于IUG的特点，提出了基于深度学习的推荐系统模型，命名为IUG-CF。为了验证IUG的有效性和合理性，以及IUG-CF模型的优越性，本工作在三个真实数据集上展开了大量的实验。实现首先确认了IUG的“组合”和“动态”两个特性，随后提出了多个IUG-CF的子模型进行内部比较，以确定IUG-CF模型的最佳结构。最后，将IUG-CF模型与当前先进模型进行了比较，实验结果表明，本工作提出的IUG和基于IUG的推荐模型IUG-CF能有效生成高质量的推荐结果。

5.2 展望

本文的工作虽然在一定程度上解决了前文提到的已有问题，但仍存在着不足和改进空间，如：

- (1) 本文第三章的工作尽管使用了多个低值来填充无趣项，但是仍是选取的整数值，并没有用更精确的小数评分。未来的工作中，将进一步使用更加精确、细化的评分来填充无趣项，更准确地学习到用户的偏好。
- (2) 本文第四章的工作将使用用户的人口统计信息来构建理想用户组，然而随着用户对于数据隐私的重视，用户的信息也将会进一步受到保护。基于此场景，未来的工作将探索不依赖用户特征的理想用户组构建方法，在保护用户敏感数据的前提下提供更准确的推荐。

致 谢

这是一篇普通的硕士学位论文，感谢三年间遇到的所有人。

感谢我的父母，感谢所有家人。

感谢我的导师张恒汝教授，多次在讨论时提出让我豁然开朗的观点，助我发表了一篇不错的论文。学术之外，在球场作为队友互相配合，作为对手互相喷垃圾话的时光，也格外有趣，希望毕业以后还能一起打球。

感谢我的室友。一起吃过的饭，打过的麻将，无数次的夜谈，都是这三年间我宝贵的回忆。没有你们，我的读研生活会少很多乐趣。

感谢在我生命的不同时期结识的各位挚友，以及三年间互相作为精神支柱的兄弟们。

感谢实验室的各位老师同学，感谢我的师兄师姐师弟师妹。

感谢五月天，《顽固》这首歌陪伴我度过了从备考到毕业的所有情绪低谷。

感谢顽固的自己。在读研的三年里，数百次凌晨离开实验室，多次在崩溃边缘擦干眼泪继续干。论文，竞赛，求职，都让我遇到了许多挫折，实验室的工位成了我的避风港，我总是靠进校时贴在工位上的便利贴“*I can do all things*”激励自己。

最后，感谢我的爷爷。我从小跟着爷爷长大，在建立三观的过程中，爷爷告诉我，要能吃苦，有担当，有诚信，懂得感恩，让我受益无穷。

2021年7月1日，我拿到了研究生录取通知书，恰逢爷爷被诊断出胃癌后第一次出院。我的录取通知书是爷爷拆开的，录取通知书只有一张纸，他看了好久。

2023年9月15日，爷爷病逝，还没能看到我的毕业证书。灵堂只有一张他的照片，我看了好久。

死亡不是生命的终点，遗忘才是。如果硕士学位论文真的能够永远地留存于某个数据库，那这便是我读书的意义。

谨以此文纪念在我上大学前每天为我做饭的爷爷。

参考文献

- [1] BOBADILLA J, ORTEGA F, HERNANDO A, et al. Recommender systems survey[J]. Knowledge-Based Systems, 2013, 46: 109-132.
- [2] AL-SHAMRI M Y H. User profiling approaches for demographic recommender systems [J]. Knowledge-Based Systems, 2016, 100: 175-187.
- [3] 秦琴, 张恒汝. 基于信任传递机制的三支推荐[J]. 模式识别与人工智能, 2020, 33(7): 600-609.
- [4] 吕亚兰, 徐媛媛, 张恒汝. 一种可解释性泛化矩阵分解推荐算法[J]. 南京大学学报(自然科学版), 2022, 58(1): 135-142.
- [5] HWANG W S, PARC J, KIM S W, et al. “Told you i didn’t like it” : Exploiting uninteresting items for effective collaborative filtering[C]// 2016 IEEE 32nd International Conference on Data Engineering (ICDE). 2016: 349-360.
- [6] STECK H. Training and testing of recommender systems on data missing not at random [C]// SIGKDD. 2010: 713-722.
- [7] CHEN J, DONG H, WANG X, et al. Bias and debias in recommender system: A survey and future directions[J]. ACM Transactions on Information Systems, 2023, 41(3): 1-39.
- [8] MARLIN B M, ZEMEL R S, ROWEIS S, et al. Collaborative filtering and the missing at random assumption[C]// . 2007: 267-275.
- [9] HU Y, KOREN Y, VOLINSKY C. Collaborative filtering for implicit feedback datasets [C]// 2008 Eighth IEEE international conference on data mining. 2008: 263-272.
- [10] RICH E. User modeling via stereotypes[J]. Cognitive science, 1979, 3(4): 329-354.
- [11] KRULWICH B. Lifestyle finder: Intelligent user profiling using large-scale demographic data[J]. AI Magazine, 1997, 18(2): 37-37.
- [12] PAZZANI M J. A framework for collaborative, content-based and demographic filtering [J]. Artificial Intelligence Review, 1999, 13(5): 393-408.
- [13] SREBRO N, JAAKKOLA T. Weighted low-rank approximations[C]// Proceedings of the 20th international conference on machine learning (ICML-03). 2003: 720-727.
- [14] LEE Y C, KIM S W, LEE D. gOCCF: Graph-theoretic one-class collaborative filtering based on uninteresting items[C]// Proceedings of the AAAI Conference on Artificial Intelligence: vol. 32: 1. 2018.

- [15] LEE Y C, PARK D, SON J, et al. Graph-theoretic one-class collaborative filtering using signed random walk with restart[C]//2020 IEEE International Conference on Big Data and Smart Computing (BigComp). 2020: 98-101.
- [16] LEE J, HWANG W S, PARC J, et al. *l*-Injection: toward effective collaborative filtering using uninteresting items[J]. IEEE transactions on knowledge and data engineering, 2017, 31(1): 3-16.
- [17] KUMAR C, CHOWDARY C R. A study on the role of uninterested items in group recommendations[J]. Electronic Commerce Research, 2023, 23(4): 2073-2099.
- [18] PAN R, ZHOU Y, CAO B, et al. One-class collaborative filtering[C]//2008 Eighth IEEE international conference on data mining. 2008: 502-511.
- [19] SHI L, LI S, DING X, et al. Selection bias mitigation in recommender system using uninteresting items based on temporal visibility[J]. Expert Systems with Applications, 2023, 213: 118932.
- [20] CREMONESI P, KOREN Y, TURRIN R. Performance of recommender algorithms on top-n recommendation tasks[C]//Proceedings of the fourth ACM conference on Recommender systems. 2010: 39-46.
- [21] LEE Y, KIM S W, PARK S, et al. How to impute missing ratings? Claims, solution, and its application to collaborative filtering[C]//Proceedings of the 2018 World Wide Web Conference. 2018: 783-792.
- [22] HE X, ZHANG H, KAN M Y, et al. Fast matrix factorization for online recommendation with implicit feedback[C]//Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. 2016: 549-558.
- [23] VOZALIS M G, MARGARITIS K G. Using SVD and demographic data for the enhancement of generalized collaborative filtering[J]. Information Sciences, 2007, 177(15): 3017-3037.
- [24] SAHU A K, DWIVEDI P. User profile as a bridge in cross-domain recommender systems for sparsity reduction[J]. Applied Intelligence, 2019, 49(7): 2461-2481.
- [25] BURKE R. Hybrid recommender systems: Survey and experiments[J]. User modeling and User-adapted Interaction, 2002, 12(4): 331-370.
- [26] LEKAKOS G, GIAGLIS G M. A hybrid approach for improving predictive accuracy of collaborative filtering algorithms[J]. User Modeling and User-Adapted Interaction, 2007, 17(1): 5-40.

- [27] SOUALI K, EL AFIA A, FAIZI R. An automatic ethical-based recommender system for e-commerce[C]//ICMCS. 2011: 1-4.
- [28] ZHAO X W, GUO Y, HE Y, et al. We know what you want to buy: a demographic-based system for product recommendation on microblogs[C]//SIGKDD. 2014: 1935-1944.
- [29] ZHAO W X, LI S, HE Y, et al. Exploring demographic information in social media for product recommendation[J]. Knowledge and Information Systems, 2016, 49: 61-89.
- [30] TAHMASEBI F, MEGHDADI M, AHMADIAN S, et al. A hybrid recommendation system based on profile expansion technique to alleviate cold start problem[J]. Multimedia Tools and Applications, 2021, 80: 2339-2354.
- [31] YASSINE A, MOHAMED L, AL ACHHAB M. Intelligent recommender system based on unsupervised machine learning and demographic attributes[J]. Simulation Modelling Practice and Theory, 2021, 107: 102198.
- [32] BOBADILLA J, GONZÁLEZ-PRIETO Á, ORTEGA F, et al. Deep learning feature selection to unhide demographic recommender systems factors[J]. Neural Computing and Applications, 2021, 33(12): 7291-7308.
- [33] LIU Y, QU H, CHEN W, et al. An efficient deep learning model to infer user demographic information from ratings[J]. IEEE Access, 2019, 7: 53125-53135.
- [34] SACHARIDIS D. Top-n group recommendations with fairness[C]//SAC. 2019: 1663-1670.
- [35] JAMESON A. More than the sum of its members: challenges for group recommender systems[C]//AVI. 2004: 48-54.
- [36] MASTHOFF J. Group recommender systems: aggregation, satisfaction and group attributes[J]. Recommender Systems Handbook, 2015: 743-776.
- [37] RENJITH S, SREEKUMAR A, JATHAVEDAN M. An extensive study on the evolution of context-aware personalized travel recommender systems[J]. Information Processing & Management, 2020, 57(1): 102078.
- [38] SAFOURY L, SALAH A. Exploiting user demographic attributes for solving cold-start problem in recommender system[J]. Lecture Notes on Software Engineering, 2013, 1(3): 303-307.
- [39] PANDEY A K, RAJPOOT D S. Resolving cold start problem in recommendation system using demographic approach[C]//ICSC. 2016: 213-218.
- [40] AL-SHAMRI M Y H, BHARADWAJ K K. A compact user model for hybrid movie recommender system[C]//ICCIMA: vol. 1. 2007: 519-524.

- [41] PEREIRA A L V, HRUSCHKA E R. Simultaneous co-clustering and learning to address the cold start problem in recommender systems[J]. Knowledge-Based Systems, 2015, 82: 11-19.
- [42] COVINGTON P, ADAMS J, SARGIN E. Deep neural networks for youtube recommendations[C]//RecSys. 2016: 191-198.
- [43] KIM D, PARK C, OH J, et al. Convolutional matrix factorization for document context-aware recommendation[C]//RecSys. 2016: 233-240.
- [44] XUE F, HE X, WANG X, et al. Deep item-based collaborative filtering for top-n recommendation[J]. ACM Transactions on Information Systems, 2019, 37.
- [45] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering[C]//WWW. 2017: 173-182.
- [46] SARKER M R I, MATIN A. A hybrid collaborative recommendation system based on matrix factorization and deep neural network[C]//ICICT4SD. 2021: 371-374.
- [47] ZENG W, FAN G, SUN S, et al. Collaborative filtering via heterogeneous neural networks[J]. Applied Soft Computing, 2021, 109: 107516.
- [48] TAY Y, ANH TUAN L, HUI S C. Latent relational metric learning via memory-based attention for collaborative ranking[C]//WWW. 2018: 729-739.
- [49] TIAN Z, PAN L, YIN P, et al. Information Fusion-Based Deep Neural Attentive Matrix Factorization Recommendation[J]. Algorithms, 2021, 14(10): 281.
- [50] YU R, YE D, WANG Z, et al. CFFNN: Cross feature fusion neural network for collaborative filtering[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 34(10): 4650-4662.
- [51] CHENG H T, KOC L, HARMSEN J, et al. Wide & deep learning for recommender systems[C]//DLRS. 2016: 7-10.
- [52] ZHANG Q, CAO L, ZHU C, et al. Coupledcf: Learning explicit and implicit user-item couplings in recommendation for deep collaborative filtering[C]//IJCAI. 2018: 3662-3668.
- [53] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural computation, 2006, 18(7): 1527-1554.
- [54] HARPER F M, KONSTAN J A. The movielens datasets: History and context[J]. Acm Transactions on Interactive Intelligent Systems, 2015, 5(4): 1-19.

- [55] KOREN Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]// Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. 2008: 426-434.
- [56] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]// Proceedings of the 10th international conference on World Wide Web. 2001: 285-295.
- [57] GUO G, ZHANG J, SUN Z, et al. Librec: A java library for recommender systems[C] // UMAP workshops: vol. 4. 2015: 38-45.
- [58] VAN METEREN R, VAN SOMEREN M. Using content-based filtering for recommendation[C]// Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 workshop: vol. 30. 2000: 47-56.
- [59] RESNICK P, IACOVOU N, SUCHAK M, et al. GroupLens: An open architecture for collaborative filtering of netnews[C]// CSCW. 1994: 175-186.
- [60] KONSTAN J A, MILLER B N, MALTZ D, et al. GroupLens: Applying collaborative filtering to usenet news[J]. Communications of the ACM, 1997, 40(3): 77-87.
- [61] ZHANG H R, MIN F, ZHANG Z H, et al. Efficient collaborative filtering recommendations with multi-channel feature vectors[J]. International Journal of Machine Learning and Cybernetics, 2019, 10(5): 1165-1172.
- [62] SHEN R P, ZHANG H R, YU H, et al. Sentiment based matrix factorization with reliability for recommendation[J]. Expert Systems with Applications, 2019, 135: 249-258.
- [63] KHOJAMLI H, RAZMARA J. Survey of similarity functions on neighborhood-based collaborative filtering[J]. Expert Systems with Applications, 2021, 185: 115482.
- [64] YANG X, ZHANG Z. Combining prestige and relevance ranking for personalized recommendation[C]// CIKM. 2013: 1877-1880.
- [65] TANG J, QU M, WANG M, et al. Line: Large-scale information network embedding[C] // WWW. 2015: 1067-1077.
- [66] ZHANG H, YANG Y, LUAN H, et al. Start from scratch: Towards automatically identifying, modeling, and naming visual attributes[C]// MM. 2014: 187-196.
- [67] GAO C, LI S, LEI W, et al. KuaiRec: A Fully-Observed Dataset and Insights for Evaluating Recommender Systems[C]// CIKM. 2022: 540-550.
- [68] BAIT, WEN J R, ZHANG J, et al. A neural collaborative filtering model with interaction-based neighborhood[C]// CIKM. 2017: 1979-1982.

- [69] ZHAO W X, MU S, HOU Y, et al. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms[C]// proceedings of the 30th acm international conference on information & knowledge management. 2021: 4653-4664.
- [70] ZANGERLE E, BAUER C. Evaluating recommender systems: survey and framework [J]. ACM Computing Surveys, 2022, 55(8): 1-38.
- [71] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [72] CAO Y, WANG X, HE X, et al. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences[C]// WWW. 2019: 151-161.

攻读硕士学位期间发表的论文及科研成果

- [1] 一作; Expert Systems with Applications; SCI 一区; 影响因子 8.5; Volume 238, Part B; 2024 年 3 月;
- [2] 一作; 南京大学学报; 北大核心; 在审;
- [3] 研究生数学建模竞赛国家三等奖, 队长;
- [4] 泰迪杯数学建模竞赛国家三等奖, 队长;
- [5] 中国大学生大数据挑战赛三等奖, 队长;