计 算 机 研 究 与 发 展
Journal of Computer Research and Development

https://crad.ict.ac.cn
62(2): 285−313, 2025

本期推荐

多样化推荐综述

彭迎涛[1] 孟小峰[1] 杜治娟[2]

[1](中国人民大学信息学院 北京 100872)

[2](内蒙古大学计算机学院 呼和浩特 010021)

(pengyingtao@ruc.edu.cn)

Survey on Diversified Recommendation

Peng Yingtao[1], Meng Xiaofeng[1], and Du Zhijuan[2]

[1](*School of Information*, *Renmin University of China*, *Beijing* 100872)

[2](*College of Computer Science*, *Inner Mongolia University*, *Hohhot* 010021)

Abstract The recommender system has a significant role in alleviating information overload, allowing users to conveniently obtain products and services on various application platforms like Tmall, TikTok, and Xiaohongshu. However, most of the recommendation systems focus on the accuracy rate as the center, which leads to adverse effects such as the limitation of users' vision, fewer display opportunities for some merchants, a single content ecosystem of the platform, and an unbalanced allocation of resources and information, such as triggering the filter bubble and the Matthew effect. As a result, strengthening the diversity of the recommendation system has become a key research point to fulfill the increasingly diversified material demands in people's lives. In recent years, research on diversified recommendations has developed rapidly. However, this aspect needs to be more systematic in organization and summarization. This paper systematically reviews the issue of diversified recommendations within recommendation systems. Firstly, we put forward the problem definition, technical framework, classification, and application scenarios of diversified recommendations. Secondly, we make comparisons and analyses of models and algorithms from four perspectives. Subsequently, we summarize the commonly used datasets and metrics for diversified recommendations. Finally, we deliberate on the problems and challenges in this field to inspire future innovation and promote development.

Key words recommendation system; diversity; deep learning; balance; interest preference

摘 要 推荐系统对缓解信息过载问题起着重要的作用,它使得用户从繁杂网络信息(如天猫、TikTok、小红书等)中轻松获取产品和服务.然而多数推荐系统以准确率为中心,导致用户视野受限、部分商家展示机会少、平台内容生态单一且资源信息分配不均衡等不利影响,如引发过滤气泡和马太效应等.由此,提升推荐的多样性逐渐成为推荐系统研究领域的关注重点,其目标是满足人们日益增长的多元化物质生活需求.近年来,推荐系统在多样性方向的技术研究呈现迅速发展态势,然而,目前多样化推荐的研究缺乏系统的整理和归纳,缺少系统地对推荐的多样化问题进行梳理和综述.首先提出了多样化推荐的问题定义、技术框架、分类及其应用场景.其次从4个方面对模型和算法进行了比较和分析.然后总结了多样化推荐的常用数据集和评测指标.最后探讨了该领域中的问题和挑战,以期激发未来创新,推动多样化推荐的发展.

收稿日期: 2023-07-25; **修回日期**: 2024-08-20

基金项目: 国家自然科学基金项目(62172423, 91846204, 62162048)

This work was supported by the National Natural Science Foundation of China (62172423, 91846204, 62162048).

通信作者: 孟小峰(xfmeng@ruc.edu.cn)

关键词 推荐系统; 多样化; 深度学习; 平衡性; 兴趣偏好

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202330600 **CSTR:** 32373.14.issn1000-1239.202330600

近年来, 在大数据、互联网、人工智能技术的飞速发展下, 网上商城、线上影院、个性化服务等进入大众的视野并对人们的工作生活产生了深刻的影响. 随着生产和消费数据的爆炸式增长, 信息过载已经成为严峻的挑战. 因此, 推荐系统(recommendation system, RS)作为解决信息过载的重要方法之一, 受到了学术界和工业界的广泛关注.

推荐系统已经在电子商务[1]、在线新闻[2-3]和多媒体内容[4]等众多场景上取得了成功应用. 传统的推荐方式致力于项目的相关性建模, 即在现有大多数研究中, 准确性是主要目标, 甚至是唯一目标. 为了追求更高的准确性, 不同技术的推荐算法应运而生, 如基于属性[5]、基于复杂模型[6]的算法等. 然而, 高度相关推荐的结果并非总尽如人意[7], 高度相关的系统并不符合用户和商家的利益. 对用户而言, 冗余或相似的推荐项目可能会导致“过滤泡沫”, 从长远来看会损害客户的满意度. 例如, 如果只向观看过“雷神”的用户持续推荐漫威电影, 由于用户并非只喜欢一种类型影片, 这可能危及用户的长期体验. 对于平台而言, 极少数最受欢迎的商品占据多数的曝光机会, 这将致使新来的或非流行的商家感到失望, 不利于平台的发展. 面对多场景多渠道的数据源和日益暴增的数据规模, 传统基于准确性推荐方式往往使用户收到的是不如人意的同质的推荐结果. 因此, 为满足人们日益增长的美好生活需要, 多样化的兴趣推荐受到工业界和学者的关注和重视.

多样化推荐是通过挖掘用户历史行为数据学习用户多样化偏好表示, 获得多样化的推荐结果, 从而提升用户体验. 多样化推荐是更有价值的, 因为相对传统准确性的推荐技术, 它为该领域新的发展提供了很大的潜力. 具体来说, 在现实场景中, 用户对兴趣的感知是多样性的、异质性的, 这要求多样化推荐系统在保证推荐项目高准确性的同时, 更要兼顾推荐列表内项目之间的异质化. 基于此, 系统才会给平台吸引和留存更多的用户并带来更大的收益. 相反, 传统推荐方式下, 信息过载问题虽然得到缓解, 但重复地推荐某一类项目, 不仅给用户带来兴趣冗余问题, 还将对平台的用户留存率和商品转化率造成负面效应.

然而, 现有多样化推荐的综述并不完善, 其在多样化推荐的技术方法梳理上不够系统详细. 特别地, 它们缺少了对最新研究技术的分析和对比, 而且在评测方法和数据上缺少详细的介绍. 例如, Kunaver 等人[8]以一种有条理的方式展示大多数相关文献资源, 并就每个文献中的发现提供一些批判性的想法, 其主要根据关注点分为定义、评估和算法开发 3 个部分并进行了分析和比较. Wu 等人[9]首先回顾多样性的各种定义, 并生成一个分类法, 以阐明如何在推荐系统中对多样性进行建模或测量, 然后从分类的角度总结多样化推荐的主要优化方法, 最后展望了本领域的研究方向. Wu 等人[10]提出了统一的分类法, 用于对搜索和推荐中多样化的指标和方法进行分类梳理.

为了弥补当前工作的不足, 本文从问题概述、技术方法、评价指标和数据等方面对多样化推荐进行综述. 在问题概述方面, 我们从问题出发对多样化及多样化推荐的概念进行定义, 总结了多样化推荐方法的框架、分类及常见应用场景. 在技术方法方面, 我们从模型算法与技术特征出发, 将方法划分为基于 2 阶段的方式、学习多样化排序方式、基于信息感知的方式和基于联合模型的方式 4 类, 并在每个类别中详细分析介绍了技术特点、优点及不足等. 在评价指标和数据集方面, 我们归纳了多样化推荐任务中常用的评测指标和数据集. 最后, 对开放性问题和未来方向进行了展望和分析.

1 多样化推荐问题概述

从普适层面和算法层面 2 个层次介绍多样化的概念. 首先从普遍性出发, 介绍了普适层面的多样性的定义, 然后从算法层面定义了推荐系统的多样化的概念.

1.1 问题定义

1)普适层面的多样性. 多样性一般被理解为异质性, 即多样性是任何系统的一个属性, 系统元素可以按类别范围进行分配. Stirling[11]讨论了科学、技术和社会等领域内的多样性, 并进行综合性的一般分析, 认为多样性通常涵盖 3 个必要属性或个别不常见的属性, 通过考虑系统元素可以被分成的类别来

分析其多样性. 其中 3 个必要的属性为多样、平衡和差异.

①多样. 指系统元素被分配到的不同的类别数，即集合中存在类别的数量，每个独立的类别可表示为系统元素是多样性的一个信号.

②平衡. 指系统元素按类别分配模式的函数. 该函数通过系统元素的相对分布来评估类别平等的程度.

③差距. 指可以区分系统元素的方式和程度(即对元素异质性的判断). 这一属性评估每个类别的特异性(即如何能被轻易区分)，即被理解为多样性信号的异质性.

结合推荐系统的特征和应用特点，我们从普适层面的多样性衍生给出推荐系统多样化的定义.

2)推荐系统的多样化. 多样化推荐是指系统通过挖掘用户交互行为并为其推荐多样的且符合兴趣偏好的商品.

基于已有多样性的相关研究工作，我们将多样化推荐系统的一般性概念特征概括为准确性、多样性、平衡性、整体性.

①准确性. 指推荐列表项目符合用户的潜在需要. 生成的推荐项目首先要满足用户的兴趣偏好.

②多样性. 指推荐列表项目尽可能地被包含在不同的主题或类别中.

③平衡性. 指对准确性和多样性相对权衡表现. 即推荐的结果既要打破高度相似的同质化，又要符合用户的兴趣偏好.

④整体性. 指系统对所有商品推荐的能力. 对“长尾”或曝光不平衡的项目，系统有能力实现对所有项目的召回.

多样化推荐的项目应该根据目标用户的兴趣范围而多样化. 实际场景中，用户对于不同种类或不同属性的商品都会表现出不同程度的兴趣. 由此，多样化推荐的结果应当尽可能包含不同的商品来满足用户的各种需要，而非推荐用户最感兴趣的单一类别商品. 例如，Niemann 等人[12] 在研究中发现，具有高准确性的项目往往在内容和/或类型比较相似，单一地追求推荐的准确性可能会导致同质化的推荐. 然而，现有的推荐系统工作大多关注推荐的准确性，在推荐结果的多样性方面探索不足. 此外，增加推荐列表的多样性可能会带来准确性的损失，如何平衡推荐的准确性和多样性必然成为了一个关键挑战.

1.2 研究价值和意义

多样化推荐算法的研究对用户、商家、平台和社会等多个方面都具有重要的意义和影响：

1)用户方面. 多样化推荐算法能够更好地满足用户个性化的兴趣和需求，提供丰富多样的推荐内容，帮助用户发现新颖、有趣的信息，提高用户的满意度和体验. 用户在使用推荐系统时更容易发现跨足多个领域的内容，避免了信息的狭隘性.

2)商家方面. 多样化推荐有助于商家更准确地了解用户的兴趣，提高商品和服务的推广效果. 同时，拓展了长尾市场，为小型商家和创作者提供了更多曝光和商业机会. 商家可以更精准地向目标用户推荐产品，提高销售转化率. 对于小型商家和新兴领域，多样化推荐有助于扩大市场份额，推动创新和多元化发展.

3)平台方面. 多样化推荐算法提升了推荐平台的用户满意度和忠诚度，使平台更有吸引力. 同时，通过优化推荐质量，平台能够更好地吸引广告主，提高广告效益. 推荐平台可以吸引更多用户，增加用户粘性，提高平台的活跃度. 对于广告主而言，推荐算法的优化也意味着更精准的广告投放，提高广告点击率和投资回报率.

4)社会方面. 多样化推荐有助于打破信息孤岛，促进文化多样性和信息共享. 通过推动用户之间的社交连接，社会网络更加紧密，促进了社会的交流和合作. 社会中的信息传播更加广泛，不同文化和观点之间的交流增加，有助于促进社会的共识和理解. 多样化推荐也为社会中的小众群体提供了更多展示的机会，促进多元物质文化需要和社会的繁荣.

总体而言，多样化推荐算法的研究不仅对个体用户体验有积极影响，也为商家、平台和整个社会创造了更多的机会和潜力. 然而，在推动多样化的同时，也需要注意隐私保护、推荐偏见等问题，确保算法的实际应用是公正和可持续的.

1.3 一般框架

基于定义，我们提出多样化推荐系统的一般框架，框架可以划分成 4 个部分：原始数据模块、数据准备模块、多样化算法模块、验证和评测模块，其中多样化算法模块、验证和评测模块是框架的重要部分，如图 1 所示.

1)原始数据模块. 原始数据模块采集用户相关信息(如浏览、点击、属性、辅助信息等)，为多样化推荐系统提供数据来源. 收集的数据包括交互数据、属性信息、文本数据、图片数据、音视频数据等.

2)数据准备模块. 数据准备模块主要对收集到的数据进行提取等预处理操作. 主要包括交互数值

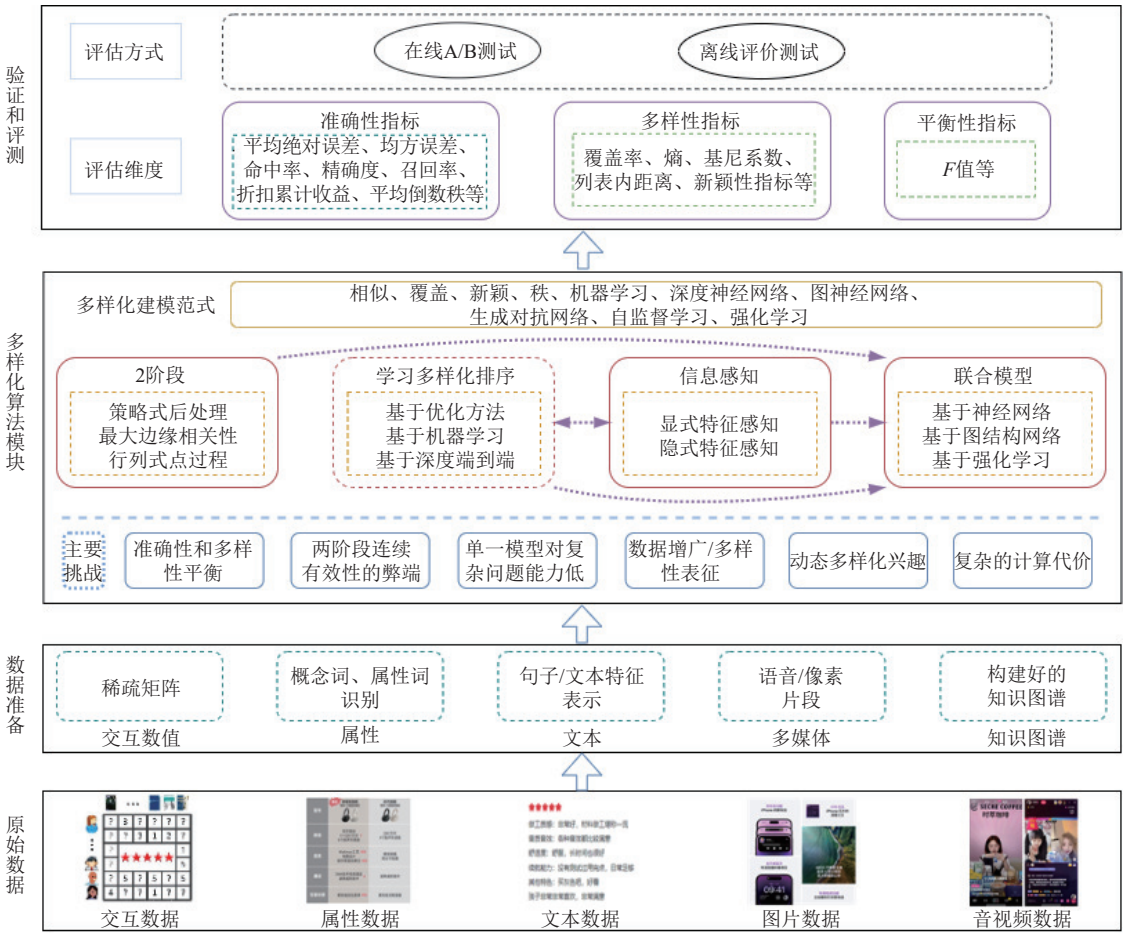


Fig. 1 General framework for diversified recommender systems

图 1 多样化推荐系统的一般框架

获取、属性词识别挖掘、句子和文本表征、图像和语音语义以及知识图谱利用等.

3)多样化算法模块. 多样化算法模块采用不同的多样化推荐算法生成推荐列表. 我们根据技术分类将其划分为基于 2 阶段的方式、基于学习多样化排序的方式、基于信息感知的方式和基于联合模型的方式. 首先, 不同方式解决的核心问题有所不同: ①基于 2 阶段的方式最早用来解决多样化推荐中准确性和多样性权衡问题. ②基于学习多样化排序的方法主要解决了 2 阶段方法的信息丢失、计算复杂性等问题. ③基于信息感知的方式主要解决了单一模型对复杂问题的能力低、数据增广/多样性表征等问题. ④基于联合模型的方式主要解决单一模型能力低、多样性表征增强困难、动态多样化兴趣表示差、复杂的计算代价等问题. 其次, 这些方法中的模块之间是可能存在包含关系的. 例如基于联合模型的方式中可能会涵盖 2 阶段、学习化排序、信息感知等方法模块; 基于学习多样化排序和信息感知的方法也可能存在相关模块的包含关系.

4)验证和评测模块. 验证和评测模块是衡量推荐系统的重要途径. 首先, 根据评估方式, 我们将其分为在线 A/B 测试和离线评价测试 2 种类型. 然后, 我们将评测的维度归纳为 3 种, 其包括准确性指标、多样性指标、平衡性指标.

1.4 分 类

多样化推荐系统的理想状态是能够向用户推荐符合偏好的多样化项目, 同时使系统满足推荐的结果是整体多样化的. 基于此, 多样化推荐可以分为个体多样性(用户内的多样性)和整体多样性(用户间的多样性)[13]. 个体多样性是针对某一用户推荐列表项目的多样性, 而整体多样性是针对所有用户推荐结果的多样性.

1)个体多样性是多样化推荐研究较多的领域. 个体多样性关注的是在生成推荐列表时如何最大限度地提高列表内项目间的多样性问题, 其理想状态是能够向用户推荐符合兴趣的多样化项目. 然而, 这也会增加推荐结果不符合用户期望的风险. 因此, 个体多样性推荐面临的挑战是如何在准确性和多样性

之间取得更好的平衡.

2)整体多样性是指在从系统层面避免出现推荐结果“长尾化”问题(长尾化问题是指只推荐少量的流行类别的项目).一般来说,整体多样性是个体多样性的组合体现,但较高的个体多样性并不一定意味着较高的整体多样性[14].例如,如果系统向所有用户推荐相同的5个不同类别的集合,则每个用户的推荐列表会非常多样化(即高度的个体多样性).然而,如果系统只能从项目池中推荐5个类别,此时整体多样性较低,这会给平台带来致命伤害.因此,一些工作致力于改善推荐系统的整体多样性[15].

鉴于目前大多数研究主要聚焦用户层面的个体多样性的工作,因此,本文主要侧重对兴趣推荐的个体多样性工作进行梳理和总结.

1.5 常见场景

在当前生产和生活中,推荐技术已被应用于各个领域场景,比如电子商务、新闻、短视频、影视、音乐、社交等.多样化推荐的场景通常包含传统推荐的应用场景,但其不同的是多样化推荐结果的侧重点不同.多样化推荐的推荐结果既要使候选项目符合用户的兴趣偏好(即与用户偏好相似),同时又要满足推荐列表之间的项目是尽可能非同质化的.因此,依据传统推荐系统的常见场景,我们总结了多样化推荐的常见场景并举例,如表1所示.

Table 1 Application Scenarios for Diversified Recommenders

表1 多样化推荐的应用场景

场景	举例
电商	Amazon、淘宝、京东等
新闻	微软新闻、Google News等
短视频	抖音、快手、小红书等
影视	豆瓣电影、优酷、Youtube等
音乐	网易云音乐、QQ音乐、酷狗音乐等
社交	Facebook, Twitter, Pinterest等

2 研究方法分类与建模范式

2.1 研究方法分类

多样化推荐作为推荐系统的重要研究方向之一,是解决用户丰富多样的需求、提升商家收益等方面的关键途径,是当前业界的研究热点.我们将目前多样化推荐方法的研究分类总结为4个方面,如图2所示.

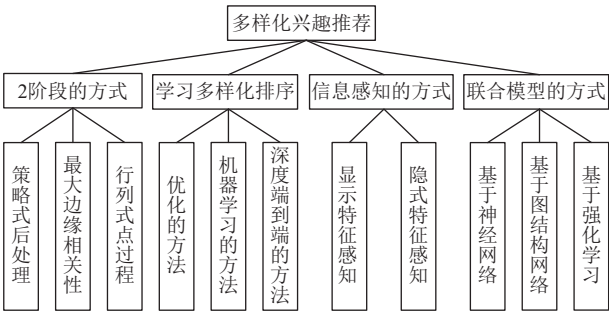


Fig. 2 Classification of diversified recommendation methods

图2 多样化推荐方法分类

从多样化推荐的核心问题出发,我们基于“多样化建模目标”,将当前的研究方法分类为基于2阶段的方式、基于学习多样化排序的方式、基于信息感知的方式和基于联合模型的方式.分类依据如下:1)2阶段方法的目标是通过策略规则等方法,针对有效地实现重新排序候选项目列表以此来提高多样性.该类方法的特点是针对性强、效果明显.2)学习多样化排序方法的目标是将多样化推荐提升到上游任务,消除2阶段方式在计算损失、连贯性等方面的弊端.其特点是端到端优化、泛化性强、减少特征工程等.3)信息感知方式旨在从数据增广方面利用属性数据、社交关系、地理位置等显式或隐式信息,感知用户的不同偏好进行多样化推荐.其特点是有效利用辅助数据、隐式信息等.4)联合模型方式的目标是为了解决大数据时代下数据、用户、商家等多方面的复杂挑战.它通过将多种技术算法和模型进行联合,以实现生成更加准确和多样化的推荐列表.该方法具有适用性强、易于处理复杂问题的特点.

虽然以往的研究[16]将多样化推荐方法划分为前处理方法、中处理方法和后处理方法3类,但随着数据、用户需求的变化,推荐系统所面临的挑战越来越复杂,这样的分类以经不足以充分概括当前的技术特性.由此,我们基于多样化推荐的核心问题和建模目标将研究方法划分为4个方面,并将前处理、中处理和后处理涵盖其中,旨在构建出更加科学合理的技术框架和分类体系.

2.2 多样化建模技术范式

我们对多样化推荐的4种研究方法的多样化建模技术范式进行归纳总结,由表2可知,在多样化推荐现有的4类方法中,兴趣的多样性建模机制主要包括相似、覆盖、新颖、秩、机器学习(machine learning, ML)、深度神经网络(deep neural network, DNN)、图神经网络(graph neural network, GNN)、生成对抗网络(generative adversarial network, GAN)、自监督学习

Table 2 Diversified Recommendation Modeling Technology Paradigm

表 2 多样化推荐建模技术范式

建模机制	2 阶段方法	学习多样化排序	信息感知方法	联合模型方法
相似	√	√	√	
覆盖	√	√	√	√
新颖	√			
秩	√			√
ML		√		
DNN		√	√	√
GNN			√	√
GAN			√	√
SSL				√
RL				√

(self-supervised learning, SSL)、强化学习(reinforcement learning, RL)等. 具体来说, 相似范式是指通过相似距离表示候选项目间的异质性, 并以算法实现表内候选列表的多样性. 覆盖范式是指算法通过提高生成的候选项目的类别覆盖范围, 来提升推荐列表的多样性. 新颖范式是指算法通过考虑新颖的物品纳入到候选列表来达到多样性高的目标. 秩具体是指在行列式点过程(determinal point process, DPP)的这类方法中, 算法将候选列表多样性问题转化为计算行列式的秩(秩越高表示多样性越大), 以此实现在满足准确性的同时使多样性尽可能地好. 机器学习、深度神经网络、图神经网络、生成对抗网络、自监督学习和强化学习 6 类技术范式一般从模型本身结构或特点出发, 结合具体的应用场景, 对多样化特征进行精细化建模, 以实现生成多样化的兴趣推荐列表. 此外, 从表 2 可以看出, 2 阶段方法、学习多样化排序和信息感知方法这 3 类技术主要表现为单一或少量的建模技术范式. 联合模型的方法往往是通过融合多种不同方法的特点, 实现对用户多样化兴趣特征的深层次、精细化的复杂建模, 这类方法通常会取得不错的效果, 但可能也会面临复杂性、通用性和扩展性的问题等.

3 多样化推荐方法对比与分析

由于推荐系统的多样化建模技术是当前研究的热点, 其相关文献主要集中在基于 2 阶段的方式、学习多样化排序的方式、基于信息感知的方式和基于联合模型的方式 4 类. 因此, 本节主要对每类别中具体的方法进行比较和分析.

3.1 2 阶段的方式

多样化推荐系统生成最终推荐列表中商品间的异质性应尽可能的大. 为了简单有效地实现, 传统技术大多是基于 2 阶段的方式. 2 阶段的方式包括预测生成候选的项目列表和权衡准确性以及多样性的项目排名. 具体地, 首先, 我们使用推荐算法获取用户对于商品的偏好分值. 然后, 我们计算候选项目之间的相似分值, 并根据这 2 项分值贪心地选择推荐的商品集合, 使其在满足商品分值高的同时集合内商品间的相似度尽量低. 根据 2 阶段技术的方法不同, 我们将其归纳为基于策略式后处理、最大化边缘相关性(maximal marginal relevance, MMR)和 DPP 三类方法.

3.1.1 基于策略式后处理的方法

早期, 为了更有效地对候选项目实现多样化排名, 策略式方法以简单有效而受到研究者的关注. 策略后处理方法按照求解处理的方式不同, 可划分为启发式策略和组合优化策略 2 种类型.

1)启发式策略

多样化推荐方法中, 启发式策略通常基于直观或经验(如用户满意度、包含主题、信任区域等)构造算法, 然后给出待排序的候选列表的一个可行解.

传统细化启发策略的解决方法一般通过后处理对预排序的候选列表进行重新排序. 例如, 已有的研究工作试图从权重分配[17]、阈值限制[18]等方面解决该任务, 其核心问题始终是选择哪些权重或阈值. 从问题优化的角度看, 这类启发策略并不能够解决 NP-hard 问题. 为了缓解上述问题, DiRec[19]从评级向量的角度出发缓解对传统语义信息的依赖. 首先, 将每个项目视为被评级的向量, 使用向量距离(如余弦距离)作为相似性/多样性的度量. 其次, 为了平衡可能相互冲突目标, 呈现排名最高的项目和选择高度多样化的项目, 引入优先覆盖树的启发式方法, 通过项目自然聚类和增量放大机制, 允许关注单个项目并识别其邻域项目, 为其选择适当的代表, 以平衡聚簇的排序和多样性.

一些文献从不同用户角度, 针对性地提出启发性策略来提升多样化推荐的效果. 如, 在用户满意度方面, 传统性推荐系统大多聚焦在推列表的准确性上. 然而, 平均精确度与用户满意度之间存在交互偏差, 其并不能代替用户满意度. 为此, Ziegler 等人[20]提出启发式主题多样化算法以增加 top-K 推荐列表的多样性. 此外, 他们还提出列表内相似度作为一种

新的度量手段，以弥补已有准确性度量在衡量用户满意度上的不足. 在多样性分布偏好方面，Meymandpour等人[21]解决了衡量推荐系统用户对多样性偏好分布的多样性问题. 具体地，采用纯粹内容相似性度量(PICSS)来评估多样性，克服了新项目评级有限或热门项目评级高相关的影响. 此外，提出新颖的多样性评估方法 RMSDE. 在时尚推荐场景方面，Sá等人[22]考虑独家和稀有物品，提出 3 种策略的多样化方法，对输出进行重新排序.

受信息检索技术的启发，Zhang 等人[23]注意到检索一组与用户匹配的项是信息检索应用中的常见问题，提出将最大化检索列表的多样性并保证将查询相似性的目标建模为二进制优化问题. 在求解方法上，他们将该优化问题转化为一个信任域问题来探索解决策略. 此外，还提出了一种评价度量项目新颖性的方法.

2)组合优化策略

启发式策略通常专注于影响用户多样化兴趣的单一因素，通过强制调整策略来实现预定目标，然而这种方法存在较大的局限性. 为了更有效地提升多样性效果，基于优化的方法受到学者广泛关注. 优化策略的目标是从有限的项目列表中找到“最佳”对象. 当前的基于优化策略的研究主要集中在多目标优化、商品流行度以及子模函数最大化求解等方面.

商品流行度方面，现实场景中流行的项目往往更容易被推荐给用户. 传统面向用户的方法存在 2 个问题：首先，没有清晰地将用户偏好与流行趋势分开. 其次，没有考虑项目流行程度的时间因素. 为了克服流行度的偏差，Nagatani 等人[24]提出基于矩阵分解的协同过滤(CF)，结合了个人用户对物品流行度的需求. 具体地，构建了结合用户的流行倾向和项目的时间流行度的算法模型，公式为：

$$\hat{x}_{ui} = (b_i^0 + b_i(t))(1 + g_u) + \boldsymbol{f}_u^{\mathrm{T}} \boldsymbol{f}_i, \tag{1}$$

其中g_u是流行倾向的用户特定参数，$b_i(t)$是时间t的时变项目偏差. 通过优化的方法学习了参数b_i^0和$b_i(t)$. g_u值作为项目的受欢迎程度和对项目的偏好之间的平衡参数. 用户u的g_u值大于 0 时，表示该用户更喜欢流行项目.

考虑流行商品新颖性对增强推荐兴趣多样化有一定促进作用. 为了克服传统协同过滤方法在多样化推荐上关于相似性、热门性、著名性等的局限性，新颖性作为一个扩大用户兴趣的指标受到领域学者的关注[25]. 然而，已有结合新颖性来促进多样化推荐的文献存在 2 点不足：1)推荐过多的流行项目；2)牺牲了准确性质量[26]. 为此缓解上述问题，Oh 等人[27]提出了个人流行趋势匹配(PPTM)的贪婪重排方法，该方法通过考虑个人用户对流行项目的倾向来增加推荐的新颖性. 具体地，方法引入地球移动距离(earth mover's distance, EMD)测量个人流行趋势之间的语义距离，然后基于 PPTM 算法解决多目标优化以尽可能地反映用户的偏好.

文献 [15] 将多样性定义为具有理论保证的子模函数最大化问题来实现求解，但都没有清晰地定义用户兴趣的覆盖范围，而是使用推荐项目评级熵改进兴趣的多样性，而这并不可靠. 因此，从用户兴趣的覆盖范围出发，Sha 等人[28]提出一种通用框架来推荐相关和多样的项目. 该框架将问题形式化为寻找 top-K 项，其目标是最大化项目相关性、用户兴趣覆盖范围及它们之间多样性的线性组合. 具体地，为了最优地提升推荐项目的相关性和差异性的总和并考虑用户兴趣，框架的组合优化方式为：

$$g(S) = \sum_{i \in S} \boldsymbol{p}_u \boldsymbol{q}_i + \alpha \sum_{j \in \Omega} w_j \max_{i \in S} \boldsymbol{q}_i \boldsymbol{q}_j + \beta \sum_{i,j \in S} \|\boldsymbol{q}_i - \boldsymbol{q}_j\|, \tag{2}$$

其中，α和β为控制参数，$\|\boldsymbol{q}_i - \boldsymbol{q}_j\|$为项目向量$\boldsymbol{q}_i$和$\boldsymbol{q}_j$之间的欧氏距离. 第$j$项的权重$w_j$可以设置为 1 或$\mathrm{e}^{\gamma-(r_{\max}-r_{uj})}$. 权重$w_j$用来量化项目$j$的重要性.

为实现组合求解最优，Ashkan 等人[29]提出多样性加权效用最大化(DUM)来保证结果多样性和效用之间的权衡. 其核心是将问题转化为在子模多样性计算上最大化一个模效用函数. 具体地，假设，$E = \{e_1, e_2, \cdots, e_L\}$是一组项目集合，$\boldsymbol{w} \in (\mathbb{R}^+)^L$是项目效用的向量表示，$\boldsymbol{w}(e)$是项目$e$的效用. $f: 2^E \to \mathbb{R}^+$是多样化的函数，其将E的任何子集映射到一个非负的实数. 多样性加权效用最大化问题的目标函数为

$$A^* = \arg\max_{A \in O} \sum_{k=1}^{L} [f(A_k) - f(A_{k-1})] \boldsymbol{w}(a_k), \tag{3}$$

其中，$A = \{a_1, a_2, \cdots, a_L\}$是来自项目的列表，$O$是项目列表的全排列，$A_k = \{a_1, a_2, \cdots, a_k\}$是$A$中的前$k$个项目的集合. $f(A_k) - f(A_{k-1})$是将a_k添加到A_{k-1}后的多样性增益. 解决方案是在列表$A^* = \{a_1^*, a_2^*, \cdots, a_L^*\}$中，最大限度地增加推荐项目的效用. 与之类似的是，Qin 等人[15]提出熵正则化器来捕获多样性的方法. 熵正则化器具有满足单调性和子模性的良好特征，因此将其与模块化评级集函数结合可以得到子模块化目标函数. 具体地，将目标表示为选择最好的项目集并选择“多样化”的项目，组合优化的公式为：

$$\arg\max_{S:|S|\leqslant K} f(S)\triangleq R(S)+\lambda g(S), \tag{4}$$

其中, $R(S)$是度量推荐的质量, $g(S)$是多样性促进正则化器, λ是正则化常数且$\lambda>0$.

此外, 在多元化排名问题上, He 等人[30]提出了通用的多元化排序方法, 并提出采用可扩展的贪心算法来寻找可证明的临近最优解. 具体来说, 该方法提出一个优化目标函数, 目标是找到k个例子中的一个子集$\mathcal{T}$, 该子集$\mathcal{T}$既与查询相关, 又在彼此之间进行多样化, 公式为:

$$\arg\max_{|\mathcal{T}|=k} g(\mathcal{T})=w\sum_{i\in\mathcal{T}} q_i r_i-\sum_{i,j\in\mathcal{T}} r_i S_{i,j} r_j, \tag{5}$$

其中, w是正则化权衡参数. 在优化函数$g(\mathcal{T})$中, 第 1 项衡量$\mathcal{T}$的加权整体相关性, 第 2 项衡量$\mathcal{T}$内示例之间的差异性.

3.1.2 基于最大边缘相关性的方法

早期的多样化推荐工作通常采用 2 阶段方法, 在候选列表生成后使用重排或后处理, 通过策略方法(如启发式等)得出项目列表, 以在准确性和多样性之间取得平衡. 然而, 这种方法中兴趣项目的产生独立于候选项目的生成, 导致重新排序策略与候选集生成分离. 这种分离使得多样化信号未反映在上游匹配模型中, 增加了最终排序列表不理想的风险.

MMR[31]是最大化边缘相关性的先驱工作, 其思想是基于贪婪算法对项目重新排序以实现最小化冗余. MMR 算法在 SIGIR1998 上被提出用于搜索领域, 目标是保持关联性的同时减小排序结果的冗余. 该方法从所有候选文档中逐一选择文档放入候选列表, 选择时同时考虑此文档与检索内容的匹配度和与列表中其他文档之间的差异(最大相似度). 具体地, 算法将文档与搜索的相关性与文档之间的相似性用一定的权值相减作为目标函数, 每次贪婪地选择目标函数值最大的文档添加到列表中, 来达到在检索结果中增加多样性的目的, 具体公式为:

$$f_{\mathrm{MMR}}\overset{\mathrm{def}}{=}\arg\max_{D_i\in\mathcal{R}\backslash\mathcal{S}}\left[\lambda f_{\mathrm{sim}_1}(\boldsymbol{D}_i,\boldsymbol{Q})-(1-\lambda)\max_{D_j\in\mathcal{S}} f_{\mathrm{sim}_2}\left(\boldsymbol{D}_i,\boldsymbol{D}_j\right)\right], \tag{6}$$

其中, $\boldsymbol{Q}$是用户查询向量, $\mathcal{R}$为推荐列表, $\mathcal{S}$是正在生成的候选集, $\mathcal{R}\backslash\mathcal{S}$为$\mathcal{R}$中没有被选中的列表集, λ为调节权重的超参数. 从式(6)可知, 前项$\lambda f_{\mathrm{sim}_1}(\boldsymbol{D}_i,\boldsymbol{Q})$为查询和当前待排序项目向量的相关性, 后项$(1-\lambda)\max_{D_j\in\mathcal{S}} f_{\mathrm{sim}_2}\left(\boldsymbol{D}_i,\boldsymbol{D}_j\right)$为当前项目与已生成候选集合中项目的最大相关性, 并采用λ来平衡相关性与多样性.

借鉴 MMR 的思想与方法, 在多样化推荐方向, 相关文献取得了一定的进展. 例如, 在旅游领域, 多样性是旅行计划中应考虑的重要因素[32]. 传统个性化旅行推荐(PTR)的文献 [33] 并没有将多样性融合到行程规划中, 这可能导致推荐的服务是不灵活的、不便捷的. 为避免这类问题, Luan 等人[34]提出基于最大边际相关性的个性化旅行推荐模型, 其在旅行规划中同时考虑旅行准确性和旅行多样性. 具体地, 采用蚁群优化方法快速解决了困难的组合优化问题. 此外, 还提出新的评估策略来计算行程多样性.

3.1.3 基于 DPP 的方法

MMR 的目标函数只考虑项目之间的二元相似性, 这与推荐项目集合的整体相似性还存在一定的偏差. DPP[35]是一种优雅的排斥概率模型, 可应用于各种机器学习任务, 其最大后验(maximum a posteriori, MAP)概率推理在许多应用中起着重要的作用. 推荐列表就是从候选商品集合中选择能够最大化后验概率的商品子集, 这一筛选过程就是 DPP 的 MAP 概率推断, 因此, DDP 算法被成功地应用到多样化推荐领域. DPP 用来提升推荐列表多样性的最大优势在于, 它直接估计整个列表的得分, 而不是两两之间进行比较, 因此它对所有物品间的相似关系考虑得更加全面.

DPP 的训练方法与传统推荐系统中使用的方法有很大的区别, 并且将 DPP 优化与现有的推荐器结合起来是复杂的. 由于 DPP 的 MAP 推断是 NP-hard, 而且即便是采用贪婪算法, 在计算上仍然过于昂贵而无法大规模使用于实时场景. 为了克服这一挑战, Chen 等人[36]提出一种新的算法来大大加速 DPP 的贪婪映射推理. 具体地, 首先在多样性量化上, 算法通过项目相似矩阵$\boldsymbol{S}$的行列式$\det(\boldsymbol{S})$表示商品集合多样性程度. 然后, 在选择商品子集时, 通过构建核矩阵$\boldsymbol{L}$同时考虑准确性和多样性指标. 具体地, 通过矩阵转换和计算, 核矩阵行列式可以表示为:

$$\lg(\det(\boldsymbol{L}))=\sum_{i\in R_u}\lg(r_{u,i}^2)+\lg\det(\boldsymbol{S}), \tag{7}$$

其中, $\sum_{i\in R_u}\lg(r_{u,i}^2)$指用户和商品的相关性, $\det(\boldsymbol{S})$指商品间的多样性. 此外, 在次模函数最大化优化上, 他们还提出贪婪次模最大化优化算法, 具体地, 算法每次从候选集中贪心选择一个使边际收益最大的项目加入到最终的结果子集中, 直到满足条件为止. 每次选择商品j添加到结果集Y_{g}中, Y_{g}初始化为空集, 商品j需要满足下面的等式:

$$j=\arg\max_{i\in Z\backslash Y_{\mathrm{g}}}\lg\det(\boldsymbol{L}_{Y_{\mathrm{g}}\cup\{i\}})-\lg\det(\boldsymbol{L}_{Y_{\mathrm{g}}}), \tag{8}$$

其中，Z是离散的内容集合，Y_g是内容集合Z的一个子集，即候选列表，$\boldsymbol{L}\in\mathbb{R}^{m\times m}$是其关联的内核矩阵.

然而，传统的 DPP 采样算法在候选结果的数量上至少是线性的，当需要根据用户的偏好对结果进行个性化时，这些算法成本可能会变得更加昂贵. 为了解决计算代价问题，Gillenwater 等人[37]提出了基于树的快速重复采样方法. 具体地，首先算法预先计算一个有N个叶子的特殊二叉树，其中每个节点都包含概率信息来决定是向左或向右前进以到达下一个采样项. 其次，利用树结构来开发一种近似采样算法，目标是有效地识别其子树的节点. 为了实现个性化，文献 [37] 提出一个简单的个性化模型，其中每个用户都与任意特征权重相关联，这些权重反映了它们参与不同类型内容的倾向. 与上面工作不同的是，从低秩因子分解的角度，Gartrell 等人[38]提出利用低秩因子分解来学习 DPP 核. 这种方法使学习算法比以前的方法在计算上快 1 个数量级，同时还提供了项目预测的方法.

为增强 DPP 在推荐上性能，Warlop 等人[39]提出了在购物篮推荐场景下的张量化 DPP，利用张量分解的思想来为下一项篮子完成任务定制模型，其目标是找到最有可能一起购买的物品. 文献 [39] 将其转化为一个分类问题，其目标是预测一组特定的项目. 此外，将复杂的多样性统计模型集成到大规模成熟系统中是一项挑战. 如果模型对多样性的定义与用户对多样性的感知之间没有很好地匹配，该模型很容易降低用户对建议的感知. 为了解决该问题，Wilhelm 等人[40]提出一个基于 DPP 的多样性统计模型. 该模型首先定义了 2 个耦合优化问题：参数化矩阵分解和结构学习. 然后提出一种多样化的协同过滤算法（DCF）来解决耦合问题. 另外，还引入了新的成对准确性度量和规范化的主题覆盖多样性度量来分别度量准确性和多样性的性能.

3.1.4 小 结

本节介绍了 2 阶段方式的多样化推荐方法，这些方法可以进一步细分为基于策略、基于 MMR 和基于 DPP 三类. 这些方法的比较分析如表 3 所示.

基于策略的方法可分为启发式和组合优化的范式. 启发式方法通常采用贪婪策略，实现对特定多样性目标的建模，例如对评级、主题、信任域等的多样化. 而组合优化方法更关注实现准确性和多样性的组合最优策略，采用正则器、子模函数优化算法、EMD 算法等技术. MMR 和 DPP 方法的核心在于平衡准确性和多样性，通过最大化边际相关性和行列

Table 3 Comparison of Two-Stage Based Methods

表 3 基于 2 阶段方式的方法比较

类别	策略	领域	多样性建模				关注点	主要挑战	核心方法	相关文献	优点					缺点				
			相似	覆盖	新颖	秩					通用性强	效率高	新度量	动态好	易扩展	通用性弱	效率低	冷启动	动态差	扩展性差
策略式	启发式	普适方法	√				评级	相似性-多样性平衡	评级的向量/距离度量	[19]	√									√
				√			主题	准确度局限性、满意度衡量	主题多样化算法	[20]		√				√				
			√				偏好	长尾效应	基于内容的 PICSS	[21]			√			√				
				√			信任区域	最佳子集匹配、多样性度量	信任区域松弛方法	[23]			√					√		
		时尚推荐			√		策略	相关性-多样性平衡	邻域余弦距离聚类、随机等策略	[22]		√							√	
	组合优化	普适方法			√		流行度	流行度/时间动态	流行与时间流行度建模	[24]				√			√			
					√		熵	用户偏好预测	熵正则化器	[15]		√						√		
					√		流行趋势	相关性-新奇性平衡	EMD 算法、PPTM	[27]		√								√
				√			兴趣覆盖	相关性-多样性平衡	子模函数优化算法	[28–29]		√				√				
			√				通用多样化	多元化排序	可扩展贪婪算法	[30]	√								√	
MMR		POI 推荐		√				组合优化问题求解	最大边际相关性	[34]		√						√		
DPP	组合优化	普适方法				√	DPP 优化	相关性-多样性平衡	DPP 算法	[36]		√							√	
		购物篮推荐				√	张量化 DPP	大篮子问题		[39]		√						√		
		视频推荐				√	多样性统计	DPP 耦合优化问题		[40]					√				√	

式点过程中的内核约束算法等进行优化,减轻了贪婪策略方法的可扩展性问题.

总的来说,2 阶段方法主要是根据给定的多样化目标,设计一个算法对生成的候选列表中的项目重新排序. 这类方法并非最佳的解决方案. 首先,有些方法假设用户的意图是静态的. 其次,该类技术大多属于策略式后处理范式,在 2 个独立的步骤中实现推荐的准确性和多样性,即:1)对项目进行评分,用推荐模型基于准确性度量生成一个有利的候选项目集;2)根据项目的多样性指标,选择一个多样化的推荐列表[8]. 然而,这类算法或使用有限的特征空间,或需要对不同的参数设置进行广泛的调整,存在较大的局限性. 而且,基于 MMR 和 DPP 的方法并不是端到端的,其中生成预测和多样性的优化分为 2 个阶段. 这类 2 阶段的方式不可避免地影响获得全局最优解的能力. 因此,解耦设计是否以可接受的准确性损失使兴趣推荐多样化仍然不确定.

3.2 学习多样化排序的方式

2 阶段方法主要是在排名阶段做出策略调整. 然而,这类方法是次优的,因为每一步的误差可能会产生最终效果上的累积损失,并且策略需要对权衡参数进行广泛调整. 由此,2 阶段的方式必然给阶段之间的连贯性、效率和计算损失带来弊端. 于是将多样化推荐提升到任务的上游,受到领域学者的关注. 与搜索领域中的学习化排序(learning to rank, LTR)[14] 类似,学习多样化排序将相关性和多样性作为学习的目标,直接为用户学习一个最佳的候选列表. 我们按照技术的不同将其归类为 3 个方面:基于优化的方法、基于机器学习的方法、基于深度端到端的方法.

3.2.1 基于优化的方法

基于优化的方法源于直观的目标驱动的思想. 该类方法在学习化排序中引入正则化器、增强约束以及自适应性等技术,以实现在保证准确性的同时提高兴趣推荐的多样性.

早期,为克服个性化推荐系统中准确性-多样性平衡的困境,Gan 等人[41] 提出了用户相似度的幂律调整的方法. 具体地,为了缓解用户的协同过滤框架中流行对象的不利影响,该方法给正则器增加一个额外的目标,通过引入幂函数调整用户相似度来实现个性化推荐分数. 后期,正则器用于矩阵分解(matrix factorization, MF)算法,其核心思想是通过限制潜在因子的大小来控制过度拟合. 例如,Jamali 等人[42] 提出社交正则化,其通过鼓励社交网络中临近的用户具有相似的潜在因素,将社交网络信息纳入优化. 受此启发,Wasilewski 等人[43] 采用多元化学习算法 ALS(alternating least squares)和最大化拉普拉斯规则化器来增强推荐结果的多样性. 对于给定项目的距离矩阵,算法提出使用距离矩阵来提高优化,以在推荐列表中产生不同项目集的因素. 具体地,其使用规则增强推荐的多样性,优化目标为:

$$\min_{P,Q} f_{\mathrm{acc}}(P,Q)+\lambda g_{\mathrm{reg}}(P,Q), \tag{9}$$

其中,$f_{\mathrm{acc}}(\cdot)$是准确性目标,$g_{\mathrm{reg}}(\cdot)$是正则化项.

不同于启发式选择参数和排序策略,在协同过滤的矩阵分解中引入多样性增强约束引起学者的关注. 如 Gogna 等人[44] 提出基于单阶段优化的解决方案. 具体地,在矩阵分解模型中设计了密集用户和稀疏项目潜在因子矩阵,并通过构建 DiABlO 算法来求解. 为了提供较高的多样性同时保持与用户偏好的相关性,文献 [44] 修改基本模型并加入额外的惩罚,以促进推荐中的多样性,公式表达为:

$$\min_{\boldsymbol{U},\boldsymbol{V}} \|\boldsymbol{Y}-M(\boldsymbol{U}\boldsymbol{V})\|_{\mathrm{F}}^2+\lambda_u\|\boldsymbol{U}\|_{\mathrm{F}}^2+\lambda_v\|\boldsymbol{V}\|_1+\lambda_{\mathrm{d}}\sum_{u\in Users} Var(\boldsymbol{U}_u), \tag{10}$$

$Var(\boldsymbol{U}_u)=\sum_{l=1}^{F}(\boldsymbol{U}_u(l)-m_u)^2$表示用户 u 的潜在因子向量的方差,F 是考虑的潜在因子的数量,$\boldsymbol{U}_u(l)$是用户 u 的潜在因子向量$\boldsymbol{U}_u$的第 l 个元素,$m_u=\dfrac{1}{F}\sum_{l=1}^{F}\boldsymbol{U}_u(l)$是潜在因子向量的平均值,$\lambda_{\mathrm{d}}$是正则化参数.

虽然许多学者认识到多元化推荐的重要性,但早期的解决方案要么需要额外的项目语义信息,要么计算过程中降低了准确性. 因此,从不同的方面考虑用户偏好的多视角、覆盖概念、偏好分类等对增强推荐兴趣的多样性具有借鉴意义.

Shi 等人[45] 研究了协同过滤推荐中的自适应多样化问题,并提出基于潜在因子组合(LFP)的推荐框架,将潜在因子模型与组合检索连接起来. 潜在因素模型从用户日志数据中构建潜在用户因素/主题,并使用它们来预测未知评级. 然后,模型通过预测用户评分来捕获潜在因素的不确定性,并以此推断多样化水平. Wu 等人[46] 描述如何从新的角度产生准确和多样化推荐. 文献 [46] 首先介绍覆盖率测量方法,使其用于量化整个集合的有用性. 然后,提出同时考虑相关性分数和基于覆盖度量的推荐框架,并利用子模态属性有效地解决最大化相关性和覆盖度的 NP-hard 问题. 具体来说,针对兴趣相关性和兴趣覆盖范围这 2 个标准,对每个用户的邻域集建模为:

$$F(S_u,u)=\alpha IRel(S_u,u)+(1-\alpha)ICov(S_u,u), \tag{11}$$

其中，$IRel(S_u,u)$度量用户u和所选邻居集S_u之间的相关性，$ICov(S_u,u)$是S_u的兴趣覆盖评分，α是平衡系数.

Zanitti 等人[47]提出以分类偏好为模型的解决多样化推荐问题的方案. 该方案以用户为中心从 4 个维度(全局覆盖、局部覆盖、新颖性和冗余性)进行建模，构建基于类别的用户模型，从不同的用户中找到他们喜欢的类别中的推荐物品. 具体地，多样化推荐的过程分为 2 步：1)将活跃用户转换为基于类别的用户资料，即基于 K-近邻的实现；2)形成远邻，即期望符合目标用户偏好的项目来自不同的类别. 受已有文献 [48] 的启发，为了控制邻居的多样化，模型引入外部参数α来确定成对多样性的显著性，对于给定用户u和属于同簇的另一个用户v，多样显著性s计算为：

$$s(u,v)=(1-\alpha)(1-d(u,v))+\alpha(d(u,v)), \tag{12}$$

多样性 s 和α的增长成正比，$d(u,v)$是活动用户u和用户v之间的距离，α是相似性和距离的权衡因子. 因此，优化的目标是提取k个最重要的用户，使户u和v之间的显著性最大化为$\overline{\mathcal{V}}=\arg\max_{v\in V}(s(u,v))$.

3.2.2 基于机器学习的方法

推荐系统中很多方法借鉴了检索任务的技术(如 MMR). 最近一些研究工作探索了使用机器学习技术来学习排序函数以实现优化推荐列表的多样性[49].

受检索问题的启发，多样化推荐问题可以转化为顺序文档选择的过程，在选择过程中可实现兼顾推荐项目的多样性. 此类方法模拟用户动态效用具有手工成本小、准确性高的特点. 例如，为建模用户在每个等级中感知效用，Xia 等人[50]提出基于连续状态马尔可夫决策过程(Markov decision process, MDP)的多样化排名模型，将用户感知效用建模为 MDP 状态的一部分. 算法核心思想是根据当前状态依次选择下一个文档，然后更新动作的选择状态，以循环方式对状态的转换进行建模，并通过策略梯度学习模型参数. 然而，通过学习决策过程促进多样性导致这类算法的通用性降低，不适合直接整合到现有的推荐系统中. 此外，现实世界的数据集是不平衡的，用户和项目的交互是稀疏的，因此难以捕捉用户的潜在多样化偏好. 随机游走是接近布朗运动的数学状态，在推荐系统中可用于捕获丰富交互网络增强特征学习以提升 top-K 推荐效果[51-52]. 改进的游走算法可以利用项目之间的间接关系，减轻稀疏性的影响，实现更广泛的项目空间覆盖，以增加推荐列表的多样性[53]. 因此，为了缓解相似物品发现工作中准确度高但多样性低的问题，Liu 等人[54]探索了随机游走对用户相似度测量的方向性影响，发现通过有向随机游走计算的用户相似度与初始节点的度数是反向相关的. 文献 [54] 通过调整从邻居到目标用户的相似度方向引入一种新的算法，以专门解决协同过滤的多样性挑战，并解决了准确性-多样性的困境. 类似地，在社交网络上，Paudel 等人[55]提出利用改进的用户项目图上的随机游走来提高信息多样性. 具体地，模型引入了用户项目反馈图的改进随机游走机制，其中系统性地删除某些节点的随机游走遍历，以降低它们相对于起始节点的重要性. 在核心技术上，该算法与传统随机游走区别有 2 点：一是包含特殊节点擦除概率的擦除矩阵. 二是擦除过程算法的改进. 此外，在音乐推荐上，为结合用户偏好的多种属性，Wang 等人[56]提出新的音乐推荐模型 RNDM，该模型包含 2 部分：特征挖掘和随机游走的关系构建. 具体地，RNDM 采用 3 个计算属性以描述用户的音乐品味：新颖性、多样性和主流. 基于此，他们通过改进的随机游走算法构建每个用户的虚拟关系，并利用目标用户与其虚拟朋友的偏好，实现更加个性化的多样化推荐.

与随机游走的图采样思想不同，一些工作将兴趣多样化建模为图计算问题(如子图选择、图流量等)来求解. 例如，为了弥补协同过滤的推荐系统中多样性的问题，Antikacioglu 等人[57]根据度属性定义了多样性概念，并将多样性的问题建模为项目目标度值的最大权重度约束的子图选择问题，然后找到尽可能满足这些约束的子图. 具体地，证明了寻找最大权重最小差异子图的问题可以简化为寻找最小成本流的问题. 其目标是使推荐系统与任何所需的度数分布之间的差异可以在多项式时间内最小化. 假设给定度数界限，向量$(\boldsymbol{c}_1,\boldsymbol{c}_2,\cdots,\boldsymbol{c}_l)$将被视为根据用户的显示约束完全满足的硬约束向量. 对所有的$u_i\in U$当且仅当$\deg_H^+(u_i)=\boldsymbol{c}_i$时，子图$H$是一个可行的解决方案. 为处理这个问题，定义给定的可行解H的目标为：

$$D(H)=\sum_{v_j\in R}\left|\deg_H^-(v_j)-\boldsymbol{a}_j\right|, \tag{13}$$

其中，向量$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_r$为目标度分布. 在整体多样性方面，Adomavicius 等人[58]将多样性最大化问题表述为图的最大流问题，并开发了基于最大流量或最大二分匹配计算来最大化整体推荐多样性，同时保持可接受的准确性水平.

3.2.3 基于深度端到端的方法

基于深度学习的推荐模型受到广泛关注，其采用嵌入和 MLP 范式[59-60]，将大规模稀疏输入映射到

低维嵌入向量. 这些方法通过在深度网络中捕获用户和项目在特征表示上的非线性关系, 并通过相似度计算获得推荐列表. 然而, 传统深度范式对用户特征的固定长度表示限制了模型捕获用户多样化兴趣的能力. 为了克服这一限制, 基于深度端到端的方法通过兴趣加权、解耦表征、意图挖掘、多头注意力机制等方式深入探讨多样化偏好的表示, 以提高推荐列表的多样性.

在点击预测方面, 为缓解多样化推荐的瓶颈问题, Zhou 等人[6]提出了新颖的深度兴趣网络(deep interest network, DIN). 该网络通过设计局部激活单元自适应地从历史行为中学习用户兴趣的表示以解决上述挑战. 具体来说, 模型引入局部激活单元, 采用软搜索关注相关的用户兴趣, 并利用加权和来获得用户对候选项的兴趣表示, 具体公式为:

$$\boldsymbol{v}_U(A)=f(\boldsymbol{v}_A,\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_H)=\sum_{j=1}^{H}\boldsymbol{a}(\boldsymbol{e}_j,\boldsymbol{v}_A)\boldsymbol{e}_j=\sum_{j=1}^{H}\boldsymbol{w}_j\boldsymbol{e}_j, \tag{14}$$

其中, $\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_H\}$为长度为$H$的用户$\boldsymbol{U}$的行为向量列表. $\boldsymbol{v}_A$为广告A的嵌入向量. $\boldsymbol{a}(\cdot)$是输出激活权重的前馈神经网络.

在视频推荐上, 传统方法通常基于启发式的规则来引入多样性, 这些规则无法捕获用户偏好, 也无法在多样性和项目准确性方面做出权衡. Mahajan 等人[61]提出通过建模低多样性对用户单个项目的参与度的影响来引入多样性, 从而能够考虑多样性和相关性来调整项目分数. 此外, 为了缓解冗余或嘈杂的影响, 以往工作只考虑用户偏好[62], 在效果上表现欠佳. 面对上述问题, Lin 等人[63]提出特征分离自平衡重排序框架(FDSB)来感知多样性. 该模型由 2 个核心模块组成: 分离注意力编码器(DAE)和自平衡多方面排序器. DAE 使用多头注意力机制从丰富的项目特征中学习分离特征, 排序器能够自适应地平衡准确性和多样性.

在序列推荐方面, 以往研究主要集中在优化推荐准确度上, 忽略了推荐项目的多样性. 而现有的提高推荐项目多样性的方法适用性差, 这是因为这些方法假设用户意图是静态的, 并且依赖于对推荐项目列表的后处理来促进多样性. 为了缓解这一问题, Chen 等人[64]将 SRs 重新定义为列表生成任务来兼顾准确性和多样性, 并提出端到端神经模型的综合方法, 称为意图感知的多样化顺序推荐(IDSR). 具体来说, IDSR 引入隐性意图挖掘(IIM)模块来捕获序列中多种用户意图, 并设计意图感知的推荐解码器以生成多样化的推荐列表. 此外, 模型还设计多样性促进(IDP)损失函数来监督 IIM 模块的学习, 并引导模型在训练中考虑多样性. 具体地, 模型最大化所有意图的满意度来生成推荐项目的列表, 具体公式为:

$$P(R_l|u,S_u)=\sum_{m=1}^{M}P(a_m|u)P(R_l|a_m,u,S_u), \tag{15}$$

其中, $P(R_l|a_m,u,S_u)$表示R_l对a_m的满意概率, $P(a_m|u)$表示意图对于用户u的重要性. 由于直接优化$P(R_l|u,S_u)$是困难的, 因此采用贪心策略生成R_l, 每次选择一个得分最高的项目的值$S(v)$, 公式为:

$$S(v)\leftarrow\lambda P(v|u,S_u)+(1-\lambda)\sum_{m=1}^{M}P(v|a_m)W(\bar{R}_{t-1},a_m), \tag{16}$$

R_{t-1}是在步骤$t-1$之前生成的推荐项目列表. 此外, 为了弥补文献 [64] 的不足, Chen 等人[65]将工作推广到多兴趣多样化方面, 提出一种统一的方法来兼顾准确性和多样性. 算法先使用隐式兴趣挖掘模块挖掘用户的多种兴趣, 然后设计兴趣感知和促进多样性的解码器来生成涵盖这些兴趣的推荐列表.

在新闻推荐方面, 大多数新闻推荐系统使用隐式反馈(如历史交互)来描述用户的兴趣, 这导致推荐结果往往是流行的. 为了增加推荐新闻的多样性, Choi 等人[66]提出以用户为中心的模型, 采用 2 个编码器来学习普遍流行的和以用户为中心的兴趣表示. 在训练过程中, 模型应用正则化方法分别对一般的兴趣和特殊的兴趣进行编码, 以此实现多样化推荐. 与之不同的是, 在异构网络下的论文推荐中, Wang 等人[67]提出基于动态补偿的论文推荐模型 COMC, 模型共同利用了异构网络中的多样性和动态性. 具体地, 模型利用节点的入度和出度来指导有向 HIN 中元路径采样的入出度方法, 有效地保证了元路径采样的多样性.

3.2.4 小 结

本节介绍了学习多样化排序的方法, 根据方法的复杂程度将其划分为基于优化、基于机器学习和基于深度学习的 3 类方法. 如表 4 所示, 我们对这些方法进行了比较和分析. 1)基于优化的方法采用正则器、分离类偏好等优化方式, 有针对性地实现多样化推荐, 但通常存在策略的局限性, 如通用性差、难以达到最优解等. 2)基于机器学习的方法利用连续状态 MDP、图随机游走、子图选择等方式, 将问题转化为应用特征的学习求解问题. 这类方法通常在通用性和效用性上表现更好, 但存在特征工程手工干预较多、局限性大的问题. 3)基于深度学习的方法利

Table 4 Comparison of Learning to Diverse Rank Methods

表 4 学习多样化排序方法的对比

类别	领域	优化方式	多样性建模				关注点	主要挑战	核心方法	相关文献	优点				缺点			
			相似	覆盖	ML	DNN					效率高	动态好	通用性强	冷启动	动态差	通用性弱	效率低	扩展性差
基于优化	普适方法	正则化	√				幂律调整	减少流行因素影响	幂律相似度调整	[41]	√			√				
			√				矩阵分解	准确性-多样性平衡	非多元化学习算法、最大化拉普拉斯规则化器	[43]	√				√			
		约束优化	√				增强约束	多样性的优化设计	增强约束、DiABIO 算法	[44]	√					√		
		组合理论		√			自适应多样化	准确性-方差平衡	潜在因素、投资组合理论	[45]	√				√			
	电影场景	覆盖偏好		√			覆盖度量	相关性-覆盖度平衡	覆盖特性算法	[46–47]	√					√		
机器学习	检索排序	连续状态马尔可夫决策过程			√		动态效用	多样化排名转为顺序文档选择	连续状态 MDP	[50]		√					√	
	普适方法	图采样		√	√		随机游走/多样化意图/虚拟关系构建	传统随机游走多样性低	改进随机游走算法	[54–56]	√							√
		图优化		√	√		子图选择	推荐多样化/子图优化	惩罚偏离、最小成本网络流	[57]	√					√		
深度学习	点击率预测	神经网络训练				√	深度兴趣	低维向量表示局限性	兴趣加权、数据自适应	[6]			√		√			
	短视频推荐	解耦表征学习				√	特征分离	相关性-多样性平衡	DAE、特征分解、自适应排名	[63]	√							√
	序列推荐	神经网络训练				√	意图挖掘	捕捉多样性	隐式意图挖掘、意图感知多样性	[64–65]		√					√	
	新闻推荐					√	多注意力聚合	流行和个人兴趣、多样化联合训练	新闻/用户注意力编码器、点击预测因子	[66]	√					√		
	论文推荐	异质信息网络				√	元路径语义	短期偏好被长期趋势掩盖、元路径的语义多样性	进出度抽样、元路径采样	[67]		√				√		

用深度神经网络将多样化推荐转化为对用户多兴趣建模的问题，通过向量表征、多意图挖掘、多注意力机制等手段来增强对用户多样化偏好的识别. 相比机器学习方法，这类模型更好地捕获用户兴趣表示，但需要克服固定向量对多样性的限制.

总体而言，学习多样化排序的方法将传统多样化推荐候选列表的生成和排序集成在一个统一的阶段，缓解了 2 阶段方法在连贯性、工作效率和计算损失上的弊端. 然而，这类方法大多只关注用户的交互数据，对辅助数据等利用不足.

3.3 基于信息感知的方式

随着信息技术的快速发展，越来越多的人可以在互联网上产生交互行为活动，这使数据呈现多元、丰富、海量等特征. 随着数据的形式和结构的日益丰富，推荐技术的研究不再局限于依赖单一数据. 因此，基于信息感知的方式来提升多样化推荐的效果受到相关学者的关注. 这类方法缓解了传统多兴趣推荐算法的单一性，探索融合辅助信息（如元数据、背景、语义语境、上下文内容、情感偏好等），进而从不同的维度和角度来提高兴趣推荐的多样化. 在本节中，我们按照信息的显隐型差异将基于信息感知的方法分为基于显式特征感知的方法和隐式特征感知的方法.

3.3.1 基于显式特征感知的方法

推荐系统中采用的数据一般可分为 2 类，即显式数据和隐式数据. 常见的显式数据包括评分、收藏、属性、主题等. 常见的隐式数据包含隐式反馈、上下文潜在特征、情感意图、隐式信号等. 一般地，现有的工作大多通过对显式特征挖掘完成对目标用户的多样化推荐. 为了提高兴趣推荐项目的多样性，避免模型生成同质的高相似结果，属性、元数据、评分等显式特征较早被用于挖掘用户多样化的偏好，以增强兴趣推荐的多样化效果.

虽然已有工作提出策略来增强多样化推荐效果，但是这类策略不适应用户个性化的需求[20,68]. 良好的多样化推荐系统并非要求单一的多样化，还应该结合属性特征等实现用户多样化的个性偏好[69]. 为了满足用户个性化的多样化需求，Di Noia 等人[70] 提出自适应的基于属性的多样化方法，能够满足 top-K 推荐列表的多样性程度，同时兼顾用户在不同的内容项目上的多样性倾向. 具体地，模型使用熵作为用户偏

好多样性程度的度量,将其与用户个人资料维度结合使用,以校准列表的多样性程度. Gogna 等人[71]则是利用项目元数据提出(凸)优化框架. 该框架将均匀分布约束与基于潜在因子模型(LFM)的协同过滤方法的原则相结合,采用基于拆分 Bregman 技术[72]的有效算法来解决多样化问题.

语境感知可以很好地学到潜在和丰富的语义语境,对提高下游任务起到一定的辅助作用. 在社交网络场景下,用户和项目上的主题分布是潜在的,并且用户的主题兴趣可能会随着时间而变化. 为了缓解这一问题, Huang 等人[73]提出了动态生成模型,该模型基于社区级主题影响来捕获潜在日志,以推断用户和项目的潜在主题分布. 具体地,模型能够跟踪社交网络中主题兴趣和主题流行度的时变分布,并利用 collapsed Gibbs 算法采样数据对模型进行训练. 为了解决时间敏感问题并用于动态模型,该模型提出自适应的 PM-2 多样化算法来获得多样化推荐列表. 在时间多样性上, Lathia 等人[74]通过调研 CF 数据随时间的变化证明了时间多样性的重要性. 技术上,文献 [74] 研究了用户评级模式的间隔时间特征是如何影响多样性的,然后提出并评估在不损害准确性的情况下最大限度地提高时间上的推荐多样性的混合机制方法. 算法包含时间混合切换、时间用户交换和频繁用户的重排序等技术模块.

为了增加推荐结果的多样化,以往工作大多考虑通过语义信息(包括电影的类型或导演、书籍的作者等)来定义多样性. 然而,这种语义多样化有 2 个缺点: 1)缺少语义信息; 2)基于给定语义信息来定义多样性可能是不可靠的. 为了弥补这类缺陷, Hao 等人[75]考虑点击多样性的动态因素(点击多样性会随着用户点击新项目而变化)并提出了结合点击和印象差异的动态多样化推荐模型(DDRM),旨在根据用户点击和展示历史生成多样化的兴趣列表.

知识图谱(knowledge graph, KG)作为辅助信息,可以将不同的信息集成到整个网络中,从而更好地学习嵌入表示[76]. 知识图谱中包含了结构化的丰富的知识和信息. 基于此, Zheng 等人[77]借助知识图谱利用图卷积网络(graph convolutional network, GCN)框架使多样化生成在上游阶段完成,并提出将多样化与匹配相结合的 DGCN,克服了现有工作中结构解耦问题. 具体地,模型首先构建 2 部图,其次提出重新平衡邻居发现来解决不同类别的不一致问题. 为了增强多样性,算法提出基于类别提升的负采样,在增加相似项目概率的同时采用基于项目类别的对抗学习来提升多样性.

此外,在序列推荐场景中,最近的工作从用户的行为序列中为每个用户提供一个整体嵌入[78-79]. 然而,统一的用户嵌入很难代表多种兴趣. Cen 等人[80]提出了一个新颖的可控多兴趣的序列推荐框架 ComiRec. ComiRec 利用多兴趣模块(动态路由、自注意力机制)从用户行为序列中捕捉多种兴趣,然后利用这些兴趣从大规模项目库中检索候选项目. 最后,生成的候选项目被送入聚合模块,以此获得整体推荐. 聚合模块利用可控因素来平衡推荐的准确性和多样性,公式为:

$$Q(u,\mathcal{S})=\sum_{i\in\mathcal{S}}f(u,i)+\lambda\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}}g(i,j), \tag{17}$$

其中, u是用户,项目i,j属于集合$\mathcal{S}$. $f(u,i)=\max\limits_{1\leqslant k\leqslant K}(\boldsymbol{e}^{\mathrm{T}}\boldsymbol{v}_u^{(k)})$是指项目的内部产品$\boldsymbol{e}_i$与用户兴趣$\boldsymbol{v}_u^{(k)}$的接近程度, $g(i,j)=\delta(CATE(i)\neq CATE(j))$是指多样性函数, $CATE(i)$表示项目i的类别.

3.3.2 基于隐式特征感知的方法

现实场景中,显式信息可能包含模糊、缺失、夸大的成分. 模型仅使用提取到的显式信息并非足够准确,较难准确地建模用户对项目的多样化偏好关系. 为了弥补仅使用显式特征的不足,许多工作考虑使用隐式特征作为辅助信息来提升多样化信息推荐.

在隐式反馈方面, Hurley[81]探讨使用隐式反馈将多样性纳入个性化排序的方法. 其目标是提供符合用户偏好的项目排名,同时也将不同的项目排列在一起. 受在 RankALS 和 RankSGD 方法[82]的启发,模型最小化其中均方误差目标,通过项目间的不同来加权 2 个项目的评分差异. 具体地,模型多样化排序策略的目标函数为:

$$f_R(\theta)=\sum_{u\in\mathcal{U}}\sum_{i\in\mathcal{I}}c_{ui}\sum_{j\in\mathcal{I}}s_j[(\hat{r}_{ui}-\hat{r}_{uj})-(r_{ui}-r_{uj})]^2, \tag{18}$$

其中, θ是预测模型的参数, c_{ui}和s_j是目标函数的参数, r_{ui}表示隐式评级, $\hat{r}_{ui}$表示预测评级.

随着深度学习技术的发展,多兴趣的注意机制使用户表示在不同的项目上有所不同,用来捕捉用户兴趣的多样性. 例如, DIN[6]通过设计局部激活单元自适应地从历史行为中学习用户多样化的兴趣表示; MIND[83]通过利用胶囊路由机制将用户的历史行为编码为用户的不同兴趣偏好. 然而,这些方法仅考虑用户过去的行为,忽略了未来潜在偏好与隐含特征,且未能捕捉到用户多样化偏好随时间演变的趋势. 为了缓解这一问题, Lu 等人[84]提出 FAT,将用户过去的行为历史和隐式的未来意识进行综合考量. 模

型的核心是基于相似用户有相似偏好，从与用户u相似的用户中构建未来序列. 然后，模型将获取的未来序列通过兴趣归类以捕获未来行为中的多样化意图. 最后，模型通过引入时间变量得到用户的未来意识嵌入表示，并将其与用户嵌入预测任务来生成多样化的推荐列表.

在上下文方面，基于会话上下文的推荐系统会生成更相关的推荐. 然而，用户往往在不断的变化中有不同的需求，为了使会话推荐的结果多样化，推荐系统需要学习候选项目与给定上下文之间的“关联关系”，而非纯粹的“相似性”[85]. 为了缓解这类不足，Hu 等人[86] 设计了双宽向量浅层网络，这种结构可以更有效地从用户会话配置文件中学习不同需求的多样化偏好.

在情感信息方面，新闻推荐场景中如果浏览的新闻以某种情绪为主，该模型可能会密集推荐具有相同情绪取向的新闻，从而使用户难以接收到不同的观点和新闻事件. Wu 等人[87] 提出情感多样性感知的神经新闻推荐方法，可以推荐具有多样化情感的新闻. 具体地，该方法提出一种情感感知新闻编码器，其与辅助情感预测任务联合训练，以学习情感感知的新闻表示. 此外，该方法还提出了一种情感多样性正则化方法，通过结合浏览新闻的整体情感方向以及候选新闻的点击和情感分数来惩罚模型. 与之相似，为了提高多样性和响应时间同时保持多标准推荐系统(MCRS)的准确性，Hong[88] 提出基于减少和征服的并行张量分解(DnCPTF)模型. 在 DnCPTF 中，情感分析缓解了稀疏问题，并且 2 阶段聚类将相似的用户评论分组到子模型中. 此外，可控的细分保证了高度的多样性和较短的响应时间，然后并行分解子模型以预测评级，并通过从子模型整合的评级推荐前N个项目.

在网络结构方面，社交推荐场景中正确的链接预测可能并不都具有同样的价值，如何使不同的链接推荐通过网络的信息流引起其相应多样性的增强从而减轻过滤气泡，是一个具有潜在影响的研究问题. 而迄今为止，预测和推荐链接主要被设想为一个准确定位问题[89]. 为了研究多样性的问题，Sanz-Cruzado 等人[90] 根据社会网络分析的概念研究了相关多样性指标的定义. 首先，从加强网络结构多样性的视角，阐述了弱关系推荐的概念. 其次，依据贪婪优化的策略对多样化水平上如何保持准确性进行权衡.

3.3.3 小 结

本节介绍了基于信息感知的多样化推荐方法，这些方法根据信息的特征分为显示特征和隐式特征 2 类. 我们列举了子类别代表的领域、挑战、方法和优缺点，如表 5 所示.

显示特征的方法主要利用评分/元数据、属性信息、时间信息、知识图谱、多元兴趣信息等特征实现多样化的兴趣推荐. 例如，在时间动态特征方面，为满足时间动态多样性的需求，采用基于时间混合表征、自适应算法和递归神经网络等方法，以捕获不同时间动态的多样性特征表示. 隐式特征的方法主要通过捕获隐式反馈、上下文、情感信息等并注入到模型中，以实现多样化表征的目标. 例如，隐式信号方面，相关研究通过动态路由、自注意机制捕获用户的隐式偏好机制；情感信息方面，相关工作通过情感感知编码器捕获用户的多样化情感表示.

总体而言，通过充分利用与用户和项目相关的属性、主题、语义隐式反馈等显式或隐式信息，系统有效地规避了单一模型在捕捉多样性特征方面的局限性，丰富了用户潜在的多样性表征，从而促进了多样化推荐的效果. 然而，现实场景中的数据信息通常不规范且存在噪声，这对模型的有效性提出了挑战. 特别是，单一模型在处理复杂数据时存在局限性，难以准确捕捉多样性表征.

3.4 基于联合模型的方式

虽然改变单个模型的内部工作机制可以提高多样性，但单个模型缺乏捕获高度稀疏推荐系统数据集中包含的复杂关系的能力. 随着深度学习技术的发展，利用一组模型(也称为联合模型或混合模型)来克服这个问题越来越引起人们的广泛关注. 具体地，联合模型的设计具有更高的特征捕获能力，提升模型对用户多元兴趣偏好的挖掘能力，以实现在保证推荐准确性的同时更好地提升兴趣列表的多样性.

3.4.1 基于神经网络的联合模型

深度学习技术的飞速发展，使其相关研究应用到推荐系统中并取得显著的效果. 例如深度网络提取特征增强表征能力，更加准确地学习用户和项目的特征. 虽然利用深度神经模型可以捕获用户的多样化偏好表征，提升兴趣列表的多样性. 然而，算法在深度表征的同时如何兼顾准确性和多样性之间的平衡仍然是一个巨大的挑战.

现有的方法没有考虑领域级别和用户级别的多样性分布差异，这限制了模型对用户多元偏好的学习能力，此外，大多数研究对准确性和多样性的平衡依然依赖人为设定的权重因子. 如，Cheng 等人[91] 提出 DCF 来解决耦合的参数化矩阵分解和结构学习问

Table 5 Comparison of Information Awareness Based Methods

表 5 基于信息感知方法的对比

类别	显式信息	领域	方式	多样化建模				关注点	主要挑战	核心方法	相关文献	优点				缺点			
				相似	覆盖	DNN	GCN/GAN					效率高	动态好	通用性强	易扩展	通用性弱	冷启动	扩展性差	动态差
显示特征	元数据	普适方法	单个优化		√			top-K 列表多样化	多样性-准确性权衡	潜在因子模型、Bregman 算法	[71]	√				√			
	属性信息		自适应	√				（凸）优化	个性化多样性需求	自适应/属性的多样化方法	[70]	√				√			
	时间动态		混合机制		√			时间动态变化	时间动态多样性需求	时间混合切换/交换、重新排名策略	[74]		√				√		
		社交推荐	动态建模		√	√		时间多样性		自适应算法	[73]		√					√	
		会话推荐	递归神经网络			√		多时间节点		递归神经网络、Word2Vec 模型	[75]		√			√			
	知识图谱	知识推荐	采样生成对抗		√		√	多样性生成	多样性生成	类别负采样、对抗学习	[77]			√					√
	多元兴趣	序列推荐	聚合			√		多兴趣表示	多兴趣表示	动态路由算法、自注意力机制	[80]			√			√		
隐式特征	隐式信号	序列推荐	用户建模			√		未来意识	隐式信号不同趋势	动态路由算法、时间感知机制	[84]		√					√	
		个性化排序	最小化排名目标	√				隐式评分	列表多样化	矩阵分解、RankALS 和 RankSGD	[81]	√						√	
	上下文	会话推荐	训练分类器			√		会话推荐多样化	列表多样化	上下文嵌入、双宽向量浅层网络	[86]				√	√			
	情感信息	新闻推荐	神经网络			√		情感多样性感知	情感多样性感知	情感感知编码器、多样性正则化	[87]	√				√			
		多准则	减而治之策略			√		推荐多样性、实时性	情感多样性感知	情感分析、并行分解算法、可控细分策略	[88]	√	√			√			

题，但其生成训练阶段仍依赖提前设定的权衡参数. 为了缓解这类不足，Liang 等人[92] 基于领域级和用户级的自适应性的多样化偏好，提出带有自适应平衡策略和度量学习的双分支网络框架（EDUA）. 首先，双分支架构使得模型具备了自动调整准确性和多样性重要程度的能力；其次，自适应平衡策略通过自适应地确定双分支中的学习重点来建模领域级别的多样性；最后，在每个分支中，构建了双向自适应度量学习骨干网络，可以捕获用户对特定物品的兴趣（用户级别的多样性）和物品对目标用户的定位. 技术上，自适应学习分支采样策略结合了用户的多样性偏好. 假设用户 u 的多样性分数为 $d_u = \left|C_u^+\right| / \left|V_u^+\right|$. d_u 的值越高，说明用户 u 的兴趣多样性越高. 模型采用正向采样概率 P_i^{O} 和反向采样概率 P_i^{R}，根据用户的多样性偏好，动态地关注更广泛或更窄的类别范围.

$$P_i^{\mathrm{O}} = \frac{N_i}{\left|V_u^+\right|},\ P_i^{\mathrm{R}} = \frac{w_i}{\sum_{j=1}^{\left|C_u^+\right|} w_j}, \tag{19}$$

其中，$w_i = \left|V_u^+\right| / N_i$. 采样时，生成一个随机数 $z = Uniform(0,1)$. 如果 $z < d_u$，使用反向采样概率 P^{R} 采样得到一个类别. 反之，使用正常采样概率 P^{O} 采样得到一个类别.

在社交媒体方面，针对小红书 APP 中独特的滑动浏览特性，Huang 等人[93] 使用时间序列分析技术从项目序列的角度解决推荐多样性问题. 其设计了一种滑动频谱分解的方法（SSD）来实现长商品序列上的多样性. 用户的整个浏览序列可以根据固定的时间窗口进行切分得到一个列数固定的矩阵，其中每一行可以代表用户在一个窗口内的浏览序列，每一行对应的 2 维嵌入矩阵的行列式可以衡量其多样性. 由于行列式与矩阵的奇异值的关系，奇异值的乘积便可以近似作为多样性的表示. 因此，总体目标中的相关性和多样性便能够一起优化.

$$\max_{\{i_1,i_2,\cdots,i_T\}\subset Z} \sum_{t=1}^{T} \boldsymbol{r}_{i_t} + \gamma \prod_{\sigma_{ijk}>0} \boldsymbol{\sigma}_{ijk}, \tag{20}$$

γ 是调整质量和多样性的超参数. 其中关键的一步是获得商品的嵌入向量，文献 [93] 综合了基于内容和协同的技术提出一种融合的 CB2CF 模型，将文本、图片和商品间的共现关系一并编码到向量中.

在新闻推荐方面，其多样化推荐面临的挑战是动态新闻不断变化的多样化偏好. 多样性是新闻推荐系统让用户接触多元信息的必要保证. 受双塔模型[94] 的启发，Raza 等人[95] 提出新闻推荐系统的双塔架构，并且结合与新闻项目相关的丰富辅助信息（元

数据)来补充项目表征. 此外, 模型通过探索类别损失函数在拟议的架构中的作用引入了类别方面的多样性, 使模型在不均衡的新闻类别的选择中保持一致.

长期以来, 多样性兴趣的表达是由某些预选的项目属性来实现的, 即用户的多样化选择是由物品本身的特征或其属性所驱动的. 因此, Zhang 等人[96]专注于定义物品类别上的多样性, 并提出与推荐算法的选择无关的通用多样化框架. 具体地, 模型将推荐模块中学习到的用户表征分解为与类别无关的和与类别有关的部分, 以便从 2 个正交的角度区分用户对项目的偏好以实现多样化的目标. 此外, 与文献 [96] 类似, 文献 [97] 从休闲的角度研究兴趣推荐中缺乏多样性的问题, 并提出 DecRS 模型框架. 近期, Chen 等人[98]也试图在相关的推荐场景中实现推荐的多样化, 其定义的多样化是针对多个项目方面的, 实现在不同的项目方面适应性地平衡准确性和多样性. 然而, 这种方法的不足是当只考虑 1 个项目方面时(例如项目类别), 其算法就会退化为 MMR 算法.

最大边际相关性在解决多样化任务方面是有显著的缺点: 直接牺牲了准确性, 在所有的内容中进行多样化是低效的. 为了解决上述缺陷, Gao 等人[99]提出基于 VAE 的协同过滤(TD-VAE-CF)方法来实现目标的多元化推荐, 有效地减轻了过滤泡沫效应并保证了准确性. 具体地, 方法为目标多样化维度训练概念激活矢量(CAVs)[100], 并使用这些矢量来调节 VAE-CF 中用户偏好的潜在嵌入, 以沿着该目标维度进行多样化, 同时保持正交维度的主题相关性.

列表生成模型以优化整个列表的整体效用为目标, 激发了最近的列表重新排名和生成推荐方法, 为多样化推荐工作在技术方法上提供了思路. 受此启发, Wu 等人[101]提出了一种基于 DPP 的生成对抗网络模型(PD-GAN)用于多样化推荐. 其在模型层面将负责准确性和多样性的模型分开, 分别作为生成器和判别器. 生成器首先使用矩阵分解模型来建模用户的偏好和商品之间的相似度, 再将得到的向量构建为核矩阵输入到 DPP 模型中, 使得 DPP 模型的输出尽量满足多样性, 再将推荐结果列表交给判别器. 判别器同样使用一个矩阵分解模型训练, 其对得到的推荐列表进行打分, 让结果的准确性尽量高, 通过平衡 2 个目标实现多样化的推荐. 其将双目标平衡与生成对抗模型相结合的思路十分巧妙. 类似的, 为了给用户生成足够多样化的项目列表并保持较高的推荐质量, Liu 等人[102]提出了 GFN4Rec 的生成方法. 它利用流网络的洞察力来确保列表生成概率与其奖励之间的一致性. 模型的主要优点是对数尺度奖励匹配损失, 它本质上提高了生成多样性, 以及自回归项目选择模型, 可以捕获项目的相互影响, 同时捕获列表的未来奖励.

此外, 传统的多样化推荐一直专注于基于用户和兴趣偏好的匹配与平衡, 却忽略挖掘用户行为和特征之间的因果关系. 为了更好地理解用户的行为过程, 因果方法已经成为多样化推荐中的一种新兴技术. UCRS[103]从因果生成的角度建模多样化推荐过程, 利用反事实推理来缓解过时特征的影响, 并基于可控排名策略来优化推荐结果的多样性. 这表明使采用因果方法来消除用户偏见以增强多样性成为可能. 类似地, DecRS[104]对偏差放大的因果效应进行了探索, 并分析了物品分布不均衡对用户兴趣表示和预测结果的混淆效应的影响. 由此 DecRS 提出采用后门调整的近似算子来减轻过滤气泡问题.

3.4.2 基于图结构网络的联合模型

近年来, 图神经网络是解决兴趣推荐的多样性的重要技术手段之一. 图神经网络方法中的邻域聚合通常会导致学习到的用户的嵌入非常接近其之前交互的项目的嵌入. 然而, 这种方式使推荐的候选项目非常相似, 并抑制了多样性[77]. 基于扩展聚合表示的方式可以在一定程度上缓解这种情况, 但其容易学习到多余的虚假边缘和噪声关系, 从而导致性能下降.

为了缓解图神经网络技术在多样化推荐的局限性, Gan 等人[105]提出了 DivKG 模型, 其使用知识图谱嵌入和 DPP 来平衡准确率和多样性. 首先模型从知识图谱结构中捕获用户、物品、额外实体间的不同关系, 然后利用图神经网络 TransE 对历史交互来优化边际损失, 将实体和关系表征成k维的向量, 并将其用来构建 DPP 的核矩阵以实现 top-K 的多样化预测. 其核心是使用这些表示来构造 DPP 的核矩阵, 以做出更优的n个多元化的预测. 与之类似的, Zheng 等人[77]致力于在图卷积网络基础上, 构建端到端的多样化算法 DGCN. 模型分别构建了重新平衡的邻居发现、分类增强的负抽样和对抗性学习的模块, 提升了推荐结果的多样性. Sun 等人[106]提出一种基于贝叶斯图神经网络 BGNN 的新训练框架. 具体来说, BGNN 结合了基于节点复制的随机图生成模型[107]. 节点复制模型可用于生成与观察图相似的样本图, 同时在边方面包含了足够的多样性. 技术上, BGNN 完整的图采样包含 2 个阶段: 一是获取随机ζ, 二是获取随机ζ后, 对每个节点执行概率为$0 \leqslant \varepsilon \leqslant 1$的复

制, 就增加了第 2 层的随机性. 将节点 j 复制到节点 q 的事件使用指标函数 F 表示, 则生成模型表示为:

$$p(\boldsymbol{G}|\boldsymbol{G}_{\text{obs}},\zeta)=\prod_{i=1}^{N}\varepsilon^{F\{\boldsymbol{G}_i=\boldsymbol{G}_{\text{obs},i^{\zeta}}\}}(1-\varepsilon)^{F\{\boldsymbol{G}_i=\boldsymbol{G}_{\text{obs},i}\}}. \tag{21}$$

其次, 在考虑图观测中的不确定性上, BGNN 的个性化排名框架扩展到节点复制的多个图.

在互补推荐方面, 互补推荐的目标是尽可能召回多种类的互补商品. 然而, 互补关系不等于相似, 常用的基于相似度量的方法并不能得到互补关系. 而且在低资源场景下, 以共同浏览和共同购买数据作为判断相似商品和互补商品的依据, 其多样性的表现效果并不好[108-109]. Hao 等人[110] 提出了基于深度学习的互补推荐框架 P-Companion 来实现相关性和多样性建模. 具体来说, P-Companion 将查询产品项目 i 和候选互补项目 j 以及它们的类型 $\{(w_i,w_j)\}$ 和远程监督标签 $y_{i,j}$ 作为输入. 模型使用编码解码模块将 w_i 的查询嵌入向量 $\boldsymbol{q}_{w_i}$ 转换为其互补基向量 $\boldsymbol{c}_{w_i}$, 用于预测 w_i 的互补类型, 公式为:

$$\boldsymbol{h}=Dropout(RELU(\boldsymbol{q}_{w_i}\boldsymbol{W}'+\boldsymbol{b}')), \tag{22}$$

$$\boldsymbol{c}_{w_i}=\boldsymbol{h}\boldsymbol{W}''+\boldsymbol{b}'', \tag{23}$$

其中, $\boldsymbol{W}'$ 和 $\boldsymbol{W}''$ 是编码和解码类型权重矩阵.

大多数的多样化推荐任务通常利用用户—项目的交互记录来挖掘用户的明确偏好, 而很少探索用户—项目的非交互记录. 然而, 被忽视的非交互记录对于捕捉用户的潜在兴趣以提高推荐的多样性尤为重要. Ye 等人[111] 提出了端到端的动态多样化图框架 DDGraph, 该框架基于用户和项目嵌入动态地构建用户—项目图. 技术上, 模型首先初始化用户项目交互图, 并通过为每个用户选择一组不同的项目并在项目和用户之间建立链接来动态更新该图. 其次, 为了增加项目的多样化, 模型设计了分位数渐进候选选择 QPCS 算法. 具体地, 为了平衡精确度和多样性, QPCS 逐步选择从 C_u^1 到 C_u^q 的项目. 对于每个子区域, QPCS 贪婪地选择 r_{iu} 得分最高的项目, 并将这些项目添加到 N_u 集合中. r_{iu} 是相关性和多样性之间的权衡.

$$r_{iu}=\lambda Rel(i,u)+(1-\lambda)Dist(i,P_u\cup N_u), \tag{24}$$

其中, $Rel(i,u)$ 是项目 i 对用户 u 的相关性值, $Dist(i,P_u\cup N_u)$ 是项目 i 的多样性值, λ 是平衡系数.

在 API 推荐方面, 传统的 API 关键词搜索方法在搜索结果多样性上有局限性, 这导致开发效率降低. 虽然已有的文献 [112–113] 中提出的方法目标为调整推荐多个 API 的组合, 但由于设计缺陷, 这些 API 组合不可避免地共享许多 API, 不满足大多数开发人员所需要的多样性. Gong 等人[114] 设计了多样性感知网络的 API 推荐方法 DAWAR, 该方法结合了关键词搜索技术和 DPP 来寻找满足融合要求的 API 的多样化组合. 具体地, DAWAR 将 API 推荐问题建模为一个图搜索问题, 目的是找到 API 相关图中的最小组 Steiner 树. 此外, DAWAR 采用 DPP 来使推荐的 API 组合多样化来解决多样性的问题.

此外, 图神经网络在用户—项目图上的训练过程中包括固定数量的观察边, 这忽略了隐性反馈推荐的内在性质, 即未观察到的边与负反馈和未知的正反馈相混合[115]. 仅仅依靠观察到的边而忽略了未知的正反馈可能会导致较差的多样性. 受对比学习的启发[116], Ma 等人[117] 提出多样化推荐的对比联合训练, 在保证准确性的同时大大改善了多样性. 具体地, 多样化的对比学习不仅对嘈杂的伪边缘具有鲁棒性, 而且还通过重新平衡项目级的流行度和类别级的优势来缓解流行度和类别的偏差, 从而提高了多样性.

3.4.3 基于强化学习的联合模型

随着强化学习技术的发展, 推荐系统中使用强化学习的一些尝试也逐渐成熟. 以往基于强化学习的推荐方法主要是通过学习用户对项目相关性特征的偏好来优化推荐结果的准确性[118]. 上述的方法通常忽略了推荐结果的多样性, 可能导致类似的项目推荐. 因此, 增强模型结果的多样性引起学者的广泛关注.

在交互推荐方面, 以往方法主要集中在优化推荐准确性上, 通常忽略了多样化推荐的其他重要属性, 如, 用户交互反馈等. Liu 等人[119] 提出了基于用户隐式反馈的交互式推荐系统的学习框架 DC^{2}B, 目标是在推荐结果的准确性和多样性之间取得良好的平衡. 为了解决由隐式反馈引起的偏差问题, DC^{2}B 从 2 个角度对用户和推荐系统之间的交互进行建模: 1)多样化的项目暴露, 推荐系统选择一组相关但多样化的项目来暴露给用户; 2)用户参与, 用户最终参与了一些暴露的项目(如点击项目). 具体来说, 模型考虑到项目的质量和所选项目集的多样性, 采用 DPP 来选择一组不同的项目以向用户展示. DPP 的优点是它显式地模拟了一个项目集被选择显示给用户的概率, 因此可以帮助解决由隐式反馈引起的偏差问题[120]. 此外, 文献 [120] 提出汤普森抽样框架下的变分贝叶斯推理算法来学习模型参数, 并为提出的 DC^{2}B 方法提供理论分析.

尽管存在一些工作考虑到了多样性, 但其对用

户在项目集多样性属性上的个性化偏好的调查通常被忽视. 它们并没有学习用户对物品集的多样性属性的个性化偏好[119,121]. 文献 [121] 开发了一种上下文组合老虎机的原则性方法, 其学习算法可以动态地识别新用户感兴趣的不同项目. 根据学者观察反馈, 设计奖励函数来衡量所选选集的准确性和多样性的质量, 目标是使 N 轮的总奖励最大化.

此外, 为了缓解多属性的问题, Ding 等人[122] 提出了集合模函数和分布函数的多臂老虎机框架. 其将推荐集合的生成建模为一个多步骤的决策过程, 以使用强化学习的策略最大化累计奖励, 最终实现推荐的商品集合在贪心情况下满足相关性和多样性的最大化. 详细地, 模型的目标是同时考虑项目集的相关性和多样性属性来吸引用户, 并将该任务表述为一个组合优化问题. 假设 $E=\{e_1,e_2,\cdots,e_L\}$是一组 L 项的集合, 称地面集. $\mathcal{A}\subseteq\{A\subseteq E:|A|\leqslant K\}$是 E的一个子集族, 最多有 K个项目. 给定用户的偏好 $\boldsymbol{\eta}=\left[\boldsymbol{\theta}^{\mathrm{T}},\boldsymbol{\beta}^{\mathrm{T}}\right]^{\mathrm{T}}\in\mathbb{R}^{d+m}$, 其中 $\boldsymbol{\theta}\in\mathbb{R}^d$是用户对相关性属性的偏好, $\boldsymbol{\beta}\in\mathbb{R}^m$是用户对多样性属性的偏好.

$$F(A|\boldsymbol{\eta})=\sum_{i=1}^{d}\theta_i R_i(A)+\sum_{i=1}^{m}\beta_i V_i(A), \tag{25}$$

$R_i(A)$是模块化函数, 用于捕获项目的第 i个相关性属性(如项目的类别), $V_i(A)$是分散函数, 用于捕获项目集的第 i个多样性属性(如集合中项目的余弦距离).

“过滤泡沫”大大降低了用户体验质量, 特别是会话推荐场景下多个目标相互冲突, 现有的方法解决该问题收效甚微. Stamenkovic 等人[123] 为推荐系统设置引入标量化多目标强化学习(SMORL), 有效地解决多目标推荐任务的多样性. 具体地, SMORL 使用强化学习代理同时满足 3 个可能相互冲突的目标: 1)促进点击率; 2)使建议集多样化; 3)引入新项目, 并优化相关性. 技术上, 强化学习可以被看作是引入多样化和新颖的推荐的正则器, 而基于排名的监督机制可以为参数更新提供更稳健的学习信号(包括负面信号). Zheng 等人[124] 专注于促进多样性的探索开发策略, 随机选择当前推荐项目附近的随机项目候选人. Hansen 等人[125] 提出基于强化学习抽样的排名器产生多样化项目的排名列表, 该模型基于排序的方法, 并不直接学习产生多样化的项目集, 而是在学习过程中利用强化学习算法[126] 实现多样化目标.

在序列推荐方面, 由于行为多态性的原因, 具有类似交互行为的用户在下一个动作中可能会点击完全不同的项目. 为了增强推荐系统的多样化效果, Gao 等人[127] 提出将准确性与多样性的平衡考虑的推理过程系统. 具体来说, 系统在强化学习和 DPP 方法的支持下, 在考虑长期回报的同时平衡了多样性问题. 其中, 强化学习被用来平衡互动项目的长期和短期相关性, 而 DPP 则作为多样性的描述, 因为决定因素可以表达项目之间的相似程度.

3.4.4 小 结

本节介绍了基于联合模型的方法, 按照模型的类型不同, 我们将这类方法划分为基于神经网络的联合模型、基于图结构网络的联合模型和基于强化学习的联合模型 3 类, 具体的方法对比与总结如表 6 所示.

基于联合模型的方式有效地缓解了传统单一模型在建模复杂表征时的局限性. 具体来说, 基于神经网络的联合模型的方法中, 采用邻域自适应、生成对抗学习、表征分解、因果反事实推理等不同的方式联合学习, 增强模型捕获多样化特征的能力, 提升了多样化推荐的效果. 然而, 现实场景中的很多数据是图结构的, 基于图结构网络的方法不仅可以学习语义上的嵌入, 而且可以更好地捕获数据的结构关系, 使得模型的多样化学习能力进一步提升. 另一方面, 强化学习的技术特点使得在解决交互推荐场景上, 可以更好地在用户与系统的交互反馈中, 动态学到用户的特征表示, 有利于捕获用户兴趣的多样性的变化, 提升模型学习能力.

基于联合模型的方式在实现准确性和多样性的权衡以及克服重排序的继承偏差方面突破了后处理方法中的“跷跷板”效应. 具体地, 本节将其概括分析为 3 个方面: 1)权衡准确性和多样性, 文献 [92–93] 等研究如何在推荐系统中有效地权衡准确性和多样性, 以避免二者之间的相互制约. 其中包括设计新的评价指标, 考虑同时优化准确性和多样性, 以及制定联合优化策略等. 2)克服重排序的继承偏差, 文献 [101–102] 等以优化整个列表的整体效用为目标, 通过研究如何在生成式重排序阶段中引入更多的随机性, 以减少对先前阶段推荐结果的依赖性, 从而提高多样性. 3)动态权衡策略, 文献 [95, 99, 119] 等研究动态权衡策略灵活调整准确性和多样性的权重, 以适应不同需求和情境. 总体来看, 基于联合模型的方法优点是通过使用复杂的模型结构更好地学习了用户的多样化特征表示. 特别地, 基于图结构网络和强化学习的方法, 在动态多样性的偏好学习上能力的提高, 这使得模型更好地处理复杂的问题和任务. 然而, 复杂的模型往往也带来高额的计算成本开销, 未来如

Table 6 Comparison of Joint Model Based Methods

表 6 基于联合模型方法的对比

方法	分类	领域	方式	多样性建模						关注点	问题挑战	技术方法	相关文献	优点					缺点			
				覆盖	DNN	GNN	GAN	SSL	RL					效率高	动态好	通用性强	易扩展	新指标	通用性弱	冷启动	扩展性差	动态差
基于神经网络	神经网络	新闻推荐	双塔模型	√	√					元数据	新闻偏好的动态变化	类别损失引入多样性	[95]	√	√					√		
	神经网络	媒体推荐	嵌入表征	√	√					最大边际相关性	新闻偏好的动态变化	VAE、协同过滤	[99]	√	√					√		
	神经网络	流推荐	时间序列分析		√					项目序列多样性	准确性-多样性平衡	滑动频谱分解（SSD）方法	[93]		√				√			
	神经网络	普适方法	领域自适应	√	√					领域级和用户级的多样性分布差异	准确性-多样性平衡	自适应平衡算法、双向自适应度量学习	[92]			√				√		
	神经网络	普适方法	分解表征	√	√					提高多样性	偏好区分	表征分解	[96]	√					√			
	生成对抗网络	普适方法	DPP 生成对抗网络	√	√		√			个性化多样性	流行偏差/多样化生成	PD-GAN 生成器、成对学习、对抗性学习	[101–102]	√							√	
图结构网络	图网络	图推荐	多样化选择动态图	√		√				动态丰富图形结构	相关性和多样性权衡	QPCS 算法	[111]		√						√	
	图网络	图推荐	随机图生成			√	√			节点生成模型	图采样不确定性	随机图生成模型、贝叶斯个性化排名	[106]	√								√
	图网络	电商推荐	联合训练学习	√		√		√		多样化与准确性	相关性和多样性权衡	自监督增强表征、模型训练	[110, 117]	√		√						√
	图搜索	APIs 推荐	图搜索优化	√		√				Web APIs 推荐的多样性	图搜索问题组合多样化	关键词搜索算法（最小群斯坦纳树）、DPP	[114]	√							√	
	KG 增强	KG 推荐	嵌入/预测	√		√				知识图谱嵌入的多样性	知识图谱嵌入的多样性	TransE、DPP	[105]	√			√					√
	KG 增强	KG 推荐	生成对抗	√		√	√			项目生成多样性	项目生成多样性	重新平衡的邻居发现、类别采样、对抗学习	[77]	√								√
强化学习	强化学习	交互推荐	老虎机方法	√					√	隐式反馈的偏差问题	准确性-多样性平衡	老虎机、强化学习、贪心优化	[119, 122]			√			√			
	联合强化	序列推荐	多目标强化学习	√				√	√	过滤泡沫		多目标强化学习、SDQL 损失和交叉熵损失联合训练	[123]			√		√			√	
			强化学习	√					√	用户在下一个动作		强化学习、DPP	[127]		√						√	

何在进一步提升模型的计算能力和减少代价等方面也将是一个值得思考和探索的问题.

4 常见数据集

在本节中, 我们对不同场景的数据集进行概述和梳理. 通过梳理总结, 我们将常见的多样化推荐的数据集按业务场景分为电影、商品、音乐、图书、新闻、游戏和社交平台等类型. 具体如表 7 所示.

4.1 电影数据集

1) MovieLens 是从 MovieLens 网站收集维护的数据集, 其中涵盖 3 个常用的基准数据集[128], 包含评分、电影属性和标签等, 分别是 MovieLens-100K 数据集、MovieLens-1M 数据集和 MovieLens-20M 数据集.

2) Douban-Movie 是通过豆瓣网站[129] 获得的, 豆瓣网站是中国流行的社交媒体网站之一, 其允许用户分享自己的观点和评分等. 该数据集包含用户间的社会关系、用户和电影属性等.

3) Netflix Prize 数据集是由电影网站 Netflix[130] 维护并发布的, 其中涵盖电影属性和 48 万个匿名用户对 1.7 万部电影的约 1 亿条评分信息.

4) Amazon-Movie 数据集[131] 来自亚马逊上 1.428 亿产品的评论和元数据. 该数据为其中的 Movie 模块.

5) Anime 数据集是通过 MyAnimeList 网站[132] 获得的, 其中包含用户偏好信息和浏览项目评分等.

4.2 商品数据集

1) Yelp 数据集[133] 是由美国商户点评网站 Yelp 公开发布的, 其中约涵盖 16 万个商户、863 万条评论

Table 7 Datasets and Papers of Different Application Scenarios

表 7 不同应用场景的数据集及论文

场景	数据集	文献
电影	MovieLens-100K	[15, 28, 44, 47, 54, 65, 70−71, 91, 101, 105, 111, 119, 122]
	MovieLens-1M	[21, 23, 29, 44−46, 65, 71, 84, 91−92, 96, 119, 122]
	MovieLens-10M	[57, 96]
	MovieLens-20M	[6, 37, 43]
	Douban-Movie	[21, 70]
	Netflix Prize	[36, 41, 43, 54, 57, 74, 81]
	Amazon-Movie	[24, 106]
	Anime	[101, 119, 122]
电商	Yelp	[73, 99]
	Amazon-Product	[6, 106, 110, 117]
	Tmall	[6, 65, 77, 80]
音乐	Last.FM	[56]
	Amazon-Music	[92, 106, 117]
书籍	Amazon-Book	[80, 84, 92, 96]
	Douban-Book	[46]
新闻	MIND	[66, 87]
游戏	Steam	[84]
社交平台	Weibo	[34]
	Kuaishou	[63]

等. 此外, 数据集还包含了餐馆、购物、酒店、旅游等信息以及用户提交的评论数据等.

2) Amazon-Product 评论数据是 Julian McAuley 团队发布的 Amazon 电商数据集, 其中涵盖产品数据信息(描述、类别、价格、品牌和图像等)和评论(评级、文本和投票)等, 此外, 数据集对商品按类别(图书、电子产品等)分为 24 个小的子集.

3) Tmall 数据集[134]是 IJCAI 2016 竞赛发布的天猫的多行为数据集, 其中包括用户与商品的交互行为(如点击、加入购物车、购买、收藏等)、用户的性别和年龄以及物品的品类、品牌等信息.

4.3 音乐数据集

1) Last.FM 数据集[135]是 Last.fm 在线音乐网站的 2 000 用户的社交网络、标签和音乐艺术家收听信息.

2) Amazon-Music 数据集是来自亚马逊 Music 模块的 1.428 亿评论数据.

4.4 图书数据集

1) Amazon-Book 数据集是自亚马逊 Book 模块的 1.428 亿产品的评论数据.

2) Douban-Book 数据集是来自豆瓣网的 38 万个用户对 9 万本书籍的 135 万个图书评分数据集, 此外, 数据集中还涵盖了用户与图书的属性数据等.

4.5 新闻数据集

1) MIND 数据集[136](microsoft news dataset, MIND)是微软公司发布的新闻用户的匿名行为数据集. 具体地, 数据集中包括了 16 万篇新闻、100 万个用户和 1 500 万条交互数据. 此外, 数据集中还涵盖了新闻的标题、摘要和类别以及用户的性别、属性等数据.

4.6 游戏数据集

1) Steam Video Games 数据集[137]是来自 Steam 平台的用户评论和游戏属性数据, 其中包括 256 万个用户、3 万个游戏项目和 779 万条评论. 此外. 除了文本评论, 其中还涵盖用户游戏的时间数据.

4.7 社交平台数据集

1) Weibo 数据集来自中国社交网站微博, 该网站类比于 Facebook 和 Twitter. 具体地, 数据集其中包含用户属性及其对事件的评论数据等.

2) Kuaishou 数据集是从快手软件中获取的数据集, 包含了 30 万活跃用户及其在相关视频上的行为日志(包含点赞、评论等).

5 常用评测指标

评测指标是评估推荐系统效果好坏的必要依据. 科学合理的评测方法和指标会直接影响推荐系统的可用性和改进方向. 一般来说, 传统推荐系统的评估方式包括离线测试和在线测试, 测评维度重点关注准确性方面. 然而, 在多样化推荐系统中, 仅关注推荐准确性的测量并不全面, 多样性指标和平衡性指标同样起到关键的作用, 如新颖性、偶然性和平衡性指标等. 本节中, 我们将已有常用多样化推荐指标按照准确性、多样性和平衡性 3 方面进行分类介绍.

5.1 准确性指标

早期的文献对准确性指标提供了广泛的调研和设计[138]. 准确性指标一般可以分别定义为 2 种主要的任务: 1)预测精度指标, 即判断单一产品预测的准确性与实际评分误差. 2)决策支持性指标, 即评价用户从所有产品项目中选择到的高质量的项目的有效性[20].

1)预测精度指标. 评估预测评分与用户真实评分的误差的程度. 平均绝对误差(mean absolute error, *MAE*)是一种有效地测量产品集合 B_i 的预测 $p_i(b_k)$ 的统计准确性[71,139]指标, 公式表示为:

$$MAE=\frac{\sum_{b_k\in B_i}|r_i(b_k)-p_i(b_k)|}{|B_i|},\tag{26}$$

其中, r_i是真实值, p_i是预测值. 与 *MAE* 相关的评价指标是均方误差(mean square error, *MSE*), 它是 *MAE* 的变体, 是在其求和之前将误差平方.

2)决策支持性指标. 该指标评价的是用户与其推荐列表的相关性程度, 如精确度(*precision*)和召回率(*recall*). 该类指标的目的是判断活跃用户与其候选列表的相关性, 而与预测精度指标不同, 不是去预测与实际的偏差值[20,92]. 具体地, 精确度指预测正确的正样本数占模型被判定为正样本的数量的比例.

$$precision = \frac{\sum_{u\in U}|R(u)\cap T(u)|}{\sum_{u\in U}|R(u)|}. \tag{27}$$

召回率(*recall*)是指模型预测正确的正样本数占真正的正样本数量的比例.

$$recall = \frac{\sum_{u\in U}|R(u)\cap T(u)|}{\sum_{u\in U}|T(u)|}, \tag{28}$$

其中, $R(u)$是按照用户在训练集中的行为给出的推荐列表, $T(u)$是按照用户在测试集中的行为给出的推荐列表.

3)命中率(hits ratio, *HR*). 是评价项目出现在推荐列表中的数值[80,105].

$$HR = \frac{1}{N}\sum_{i=1}^{N} hits(i), \tag{29}$$

其中, N为用户的总数量. $hits(i)$表示第i个用户访问的值是否在列表中, 如果是, $hits(i)=1$, 否则$hits(i)=0$.

4)折扣累计收益(discounted cumulative gain, *DCG*). 该指标表示用户喜欢的商品被排在推荐列表前面比排在后面可以更大程度上增强用户体验效果[29].

$$DCG(b,L) = \sum_{i=1}^{b} r_i + \sum_{i=b+1}^{L}\frac{r_i}{\log_b i}, \tag{30}$$

其中, r_i表示排在第i位的商品是否是用户喜欢的. 若$r_i=1$, 代表用户喜欢该商品; 若$r_i=0$, 则指用户不喜欢该商品. b是自由参数. L表示推荐列表长度. 然后, 由于在用户之间 *DCG* 没有直接的可比性, 需要对它们进行归一化. 具体来说, 首先将测试集中所有的项目以理想的次序排序, 取前 K 项并计算其 *DCG*. 然后, 将原 *DCG* 除以理想次序下的 *DCG*, 这样可以得到归一化的折扣累计收益(normalized discounted cumulative gain, *NDCG*).

$$NDCG@K = \frac{DCG}{iDCG}. \tag{31}$$

5)平均倒数秩(mean reciprocal rank, *MRR*). 是指正确检索结果值在检索结果中的排名, 用来评价系统的能力[66,70].

$$MRR = \frac{1}{Q}\sum_{i=1}^{|Q|}\frac{1}{rank_i}, \tag{32}$$

其中, $|Q|$表示用户数, $rank_i$是指对第i个用户推荐候选列表中在实际结果中的项目所处排列中的位置.

5.2 多样性指标

尽管准确性指标是多样化推荐中衡量有用性的一个主要方面, 但其无法衡量用户对多样性的满意程度. 我们梳理了多样化推荐研究中的多样化指标的内容, 并对常用的多样性指标做了总结.

1)覆盖率(*coverage*). 表示模型预测的项目在总物品集合中所占的百分比, 它可以用来反映用户候选列表中项目的多样化情况. 此外, 还可以用来反应整体系统层面物品长尾问题, 以此来刻画推荐系统的整体多样性[140]. 具体公式如下:

$$coverage = \frac{|R(u)|}{|I|}, \tag{33}$$

其中, I指整个物品集合, $R(u)$表示给用户u推荐列表中出现的唯一物品的集合.

物品流行度一般表示有多少用户与此项目进行了交互行为, 即物品在推荐列表中的出现的概率. 实际场景中通常采用计算方式有 2 类: 信息熵和基尼系数.

2)熵(*entropy*). 表示不同类别项目的分布. 信息熵用来表示信息的不确定性度量. 在推荐系统中, 信息熵越大, 表示系统越混乱, 项目的流行度越趋于相等[77].

$$entropy = -\sum_{i=1}^{n} p(i)\lg p(i), \tag{34}$$

其中, $p(i)$是第i个项目的流行度除以所有项目流行度之和.

3)基尼系数(Gini index, *GI*). 该指标在经济学中广泛使用, 其目标是衡量收入不平等. 在推荐系统中, 特定类别的项目数量可以理解为该类别的财富. 由此, 推荐系统中的基尼系数越低越表示项目间的多样性越好[7,95].

$$GI = \frac{1}{n-1}\sum_{j=1}^{n}(2j-n-1)p(i_j). \tag{35}$$

4)列表内距离(intra-list distance, *ILD*). 评价推荐列表中的项目间的平均距离, 是较早提出的评估推荐系统多样性的指标[20,122]. 其定义如下:

$$ILD = \frac{2}{|R_L|(|R_L|-1)}\sum_{(i,j)\in R_L} d_{ij}. \tag{36}$$

2个项目间的异质性d_{ij}可以使用欧氏距离表示. 此外, 类似的扩展指标还有列表内平均距离(intra-list average distance, *ILAD*), $ILAD=\underset{u\in U}{\text{mean}}\ \underset{i,j\in R_u,i\neq j}{\text{mean}}(1-S_{ij})$和列表内最小距离(intra-list minimal distance, *ILMD*) $ILMD=\underset{u\in U}{\text{mean}}\ \underset{i,j\in R_u,i\neq j}{\min}(1-S_{ij})$, 其中, $\boldsymbol{S}$表示项目之间的相似度矩阵.

5)新颖性指标(*serendipity*). 表示为用户和推荐项目之间的距离, 新颖性的值越高, 表示推荐项目与用户过去的交互差距越大, 因此可以被视为新颖性的推荐[141], 并以此表示多样性, 公式如下:

$$serendipity=\frac{\sum_{j=1}^{N}d(\boldsymbol{U},\boldsymbol{E}_j)}{N}, \tag{37}$$

其中, $d(\boldsymbol{U},\boldsymbol{E}_j)$表示用户的嵌入$\boldsymbol{U}$和项目的嵌入$\boldsymbol{E}_j$之间的余弦距离.

5.3 平衡性指标

平衡性指标旨在评价模型在平衡推荐的精准度和多样性的效果. 常用的平衡性指标为[92,119]:

$$F_{\text{score}}=\frac{2\times precision\times g_{\text{diversity}}}{precision+g_{\text{diversity}}}, \tag{38}$$

$g_{\text{diversity}}$代表多样性指标. 此外, Ding 等人[122] 提出了类似的变体指标$\beta_{-\text{score}}$, 如式(39)所示:

$$F_{\beta-\text{score}}=\frac{(1+\beta^2)\times recall\times g_{\text{diversity}}}{\beta^2 g_{\text{diversity}}+recall}, \tag{39}$$

其中, β是一个常数, 并设置$\beta\in(1,2)$.

6 挑战与展望

虽然领域学者已认识到提高推荐系统兴趣多样性的重要性并开始进行探究. 然而, 我们认为该领域存在一些未解决的挑战. 我们讨论了这些挑战, 并指出未来研究的方向以促进该领域的深入发展.

6.1 模型层面的挑战

1)多样化推荐的时序性. 即用户兴趣的多样性与兴趣变化的时序性. 目前, 关于多样性兴趣推荐系统的现有研究主要关注单个时间点, 忽略了连续时间跨度带来的影响. 在实际推荐系统中, 时间连续性对挖掘用户行为和潜在偏好起着关键的作用. 用户与推荐系统交互的不同阶段的多样化偏好的程度是差异化的. 由此, 我们认为在未来研究方向中, 以数据的时间序列为关注视角, 探索如何在连续交互环境下实现时序性的多样化推荐是一个有意义的话题. 如, 当用户在系统中有新的交互时, 模型应该推荐更多的符合时序偏好的多样化列表, 以帮助用户更好地探索自己的兴趣. 随着收集到有关用户与系统交互的更多数据, 系统能够自适应地平衡准确性和多样性, 不仅基于用户过去的偏好提供用户喜欢的物品, 而且还可以向用户展示不同的新颖的商品, 以吸引并留住用户, 提升商家的收益.

2)多样化推荐的个性化. 通过综合考虑用户的个人偏好、系统环境等因素, 为用户提供符合其个性化偏好的多样性推荐列表, 从而提高推荐的满意度. 由于不同用户喜好的差异, 相似的2个用户也会存在个性化的差异, 这具体表现在多样性的个性化上. 例如, 即便是同样喜欢篮球新闻的相似用户, 也会因为各自的个性化偏好不同, 使他们关注的球员与队伍不同. 因此, 多样化推荐的个性化可以为用户提供更加准确、多样化的推荐内容, 具有可扩展性和适应性, 能够适应不同的推荐场景和用户群体, 具有较高的实用性和应用性价值.

3)多样化推荐的可解释性. 对推荐结果进行解释, 使用户能够理解推荐系统产生的原因. 目前, 存在的多样化推荐工作研究仅仅关注生成多样化的项目列表, 大多工作忽视了模型的可解释性. 而实际场景中, 用户对系统推荐的物品可能会存在怀疑——为什么给我推荐这些商品. 因此, 为了让得推荐列表的商品更容易让用户接纳, 给用户推荐的可解释性对用户更为重要. 然而, 往往根据“用户的历史记录”和“相似用户的项目”的简单解释并不理想. 总体来说, 相比传统的多样化推荐系统, 提升系统的可解释性不但能够增强系统透明度, 还可以提高消费者对系统的信任度和接受度、用户满意程度以及为系统提升用户留存率和效益.

4)大语言模型在多样化推荐中的应用. 近期, 大语言模型(ChatGPT)在文本、图像等领域取得的显著效果引起了人们的重点关注. 大语言模型因其丰富的训练数据和深层复杂的模型能力, 在许多文本和图像任务上取得了显著的效果. 然而, 对大语言模型来说, 其训练数据是文本、图像等内容数据的结构形式, 与传统推荐系统的行为交互的数据有所不同, 因而不能够直接用于推荐系统中. 因此, 未来如何将传统推荐系统中的交互特征与大语言模型融合是面临的核心问题. 例如, Li 等人[142] 提出受搜索引擎启发的生成框架 GPT4Rec 以捕捉不同方面和粒度的用户兴趣来提高相关性和多样性. 此外, 将推荐系统的交互行为转化为大语言模型的输入范式问题可能是未来值研究和探索的方向. 将推荐系统中用户多样化需求和潜在偏好采用自然语言的形式表示, 让大模

型进一步理解和挖掘用户的深层表征,以此来提高多样化推荐的效果.

6.2 数据层面的挑战

1)面向多模态数据多样化推荐. 指利用不同模态的数据可以提供不同的信息,从而提高推荐的效果. 特别是随着技术的不断革新与发展,多种形式的数据和服务提供给用户使用. 例如,图片和视频可以提供更直观的视觉信息,文本可以提供更详细的语义信息等. 但如何有效地融合不同的数据形式则是多模态兴趣多样化推荐的核心问题. 在融合不同的数据形式时,考虑如何表征其中的联系才能达到更好的效果,同时还需要考虑数据的稀疏性问题和不同数据形式的融合问题,以提高推荐的准确性和多样性.

2)多样化推荐的冷启动问题. 传统推荐系统的冷启动问题表现为目标用户与目标系统的交互数据过少,无法对新用户进行推荐生成,一般可分为用户冷启动问题和物品冷启动问题. 在传统推荐系统中,解决冷启动问题通常采用先验数据、兴趣迁移、跨平台知识融合等手段和方法. 与传统高准确性的推荐系统不同,多样化推荐重点面临的是用户的冷启动问题(即系统无法给没有历史行为的新用户表示兴趣偏好). 而就新商品来说,由于目标系统中新商品从新颖性层面与多样化的建模需求相吻合,因此,这从系统需求层面稀释了该新物品冷启动问题的重要性. 在未来,多样化推荐系统在用户冷启动问题上更应该关注实时的用户兴趣偏好的变化. 其次,解决多样化推荐的用户冷启动问题,需要关注社会资源共享的协同效用. 随着AI的进步,未来用户的不同APP的不同信息可以实现加密共享,这使得公司之间达成信息同盟,利用共享的用户行为信息实现用户体验的协同优化. 此外,在大语言模型迅猛发展下,如何更好地利用大模型的能力缓解冷启动场景同样值得探索.

6.3 测评层面的挑战

1)平衡性指标. 推荐系统的核心是为用户提供满足其个性化偏好的项目,即准确性. 同样,推荐系统的多样性对提升商家收益和满足用户丰富的生活需求起着重要的作用. 因此,多样化推荐系统的研究重在准确性和多样性的平衡. 定义平衡性的综合指标.已有研究定义F值[92]及改进的α−NDCG[122]来实现准确性和多样性的平衡,然而,这些工作不能完全实现二者的权衡. 在将来的工作中,如何权衡准确性和多样性权重大小的同时促进二者值的提高是相关学习者和工业研究者应该关注的重点.

2)惊喜性指标. 已有的多样化推荐系统将新颖性作为评测的指标. 新颖性指标是指给用户尽可能地推荐其没有听说过的商品,一般采用空间表征距离来衡量. 然而我们不难发现,新颖性指标可以表示项目之间的异质性,但其在同时满足用户兴趣偏好上表现欠缺. 由此,未来我们可以尝试构建一种新的惊喜性衡量指标,使其满足推荐的商品与用户历史交互的商品异质的同时又符合用户的兴趣偏好.

7 总 结

随着物质水平明显提高,人民日益增长的美好生活需要呈现多样化、多层次、多方面的特点. 为了满足用户的多样化需求,建模和设计兴趣推荐模型,实现多样化的兴趣推荐系统是有效途径和重要举措. 本文对多样化推荐进行梳理与总结,首先介绍了推荐系统中多样性的问题描述、定义、分类和常见场景. 其次从4个角度阐述多样化推荐的研究方向,分析多样化推荐系统的研究进展,并讨论未来多样化推荐中可探索和发展的方向,旨在启发未来的创新,给领域研究者们带来一定的启发和帮助.

作者贡献声明:彭迎涛负责论文的总结和撰写;孟小峰指导论文总体架构;杜治娟提出指导意见和修改论文.

参 考 文 献

[1] He Ruining, McAuley J. VBPR: Visual bayesian personalized ranking from implicit feedback[C]//Proc of the 30th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 144−150

[2] Wu Chuhan, Wu Fangzhao, An Mingxiao, et al. NPA: Neural news recommendation with personalized attention[C]//Proc of the 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2019: 2576−2584

[3] Shi Cunhui, Hu Yaokang, Feng Bin, et al. A hierarchical knowledge based topic recommendation method in public opinion scenario[J]. Journal of Computer Research and Development, 2021, 58(8): 1811−1819 (in Chinese)
(史存会, 胡耀康, 冯彬, 等. 舆情场景下基于层次知识的话题推荐方法[J]. 计算机研究与发展, 2021, 58(8): 1811−1819)

[4] Chen Jingyuan, Zhang Hanwang, He Xiangnan, et al. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention[C]//Proc of the 40th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New

York: ACM, 2017: 335−344

[5] Lian Jianxun, Zhou Xiaohuan, Zhang Fuzheng, et al. xdeepfm: Combining explicit and implicit feature interactions for recommender systems[C]//Proc of the 24th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1754−1763

[6] Zhou Guorui, Zhu Xiaoqiang, Song Chengru, et al. Deep interest network for click-through rate prediction[C]//Proc of the 24th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1059−1068

[7] Zhou Tao, Kuscsik Z, Liu Jianguo, et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. Proceedings of the National Academy of Sciences, 2010, 107(10): 4511−4515

[8] Kunaver M, Požrl T. Diversity in recommender systems: A survey[J]. Knowledge-based systems, 2017, 123: 154−162

[9] Wu Qiong, Liu Yong, Miao Chunyan, et al. Recent advances in diversified recommendation[J]. arXiv preprint, arXiv: 1905.06589, 2019

[10] Wu Haolun, Zhang Yansen, Ma Chen, et al. A survey of diversification techniques in search and recommendation[J]. arXiv preprint, arXiv: 2212.14464, 2022

[11] Stirling A. A general framework for analysing diversity in science, technology and society[J]. Journal of the Royal Society Interface, 2007, 4(15): 707−719

[12] Niemann K, Wolpers M. A new collaborative filtering approach for increasing the aggregate diversity of recommender systems[C]//Proc of the 19th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2013: 955−963

[13] Chen Li, Wu Wen, He Liang. Personality and Recommendation Diversity[M]//Emotions and Personality in Personalized Services. Berlin: Springer, 2016

[14] Adomavicius G, Kwon Y O. Improving aggregate recommendation diversity using ranking-based techniques[J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 24(5): 896−911

[15] Qin Lijing, Zhu Xiaoyan. Promoting diversity in recommendation by entropy regularizer[C]//Proc of the 23rd Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2013: 2698−2704

[16] AI Box. An overview of the current status and development of diversified recommendations[EB/OL]. [2023-07-14]. https://zhuanlan.zhihu.com/p/634476644 (in Chinese)
(AI Box. 一文纵览多样化推荐的现状与发展[EB/OL]. [2023-07-14]. https://zhuanlan.zhihu.com/p/634476644)

[17] Drosou M, Pitoura E. Search result diversification[J]. ACM SIGMOD Record, 2010, 39(1): 41−47

[18] Yu Cong, Lakshmanan L, Amer-Yahia S. It takes variety to make a world: Diversification in recommender systems[C]//Proc of the 12th Int Conf on Extending Database Technology: Advances in Database Technology. New York: ACM, 2009: 368−378

[19] Boim R, Milo T, Novgorodov S. Diversification and refinement in collaborative filtering recommender[C]//Proc of the 20th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2011: 739−744

[20] Ziegler C N, McNee S M, Konstan J A, et al. Improving recommendation lists through topic diversification[C]//Proc of the 14th Int Conf on World Wide Web. New York: ACM, 2005: 22−32

[21] Meymandpour R, Davis J G. Measuring the diversity of recommendations: A preference-aware approach for evaluating and adjusting diversity[J]. Knowledge and Information Systems, 2020, 62(2): 787−811

[22] Sá J, Queiroz Marinho V, Magalhães A R, et al. Diversity vs relevance: A practical multi-objective study in luxury fashion recommendations[C]//Proc of the 45th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2022: 2405−2409

[23] Zhang Mi, Hurley N. Avoiding monotony: Improving the diversity of recommendation lists[C]//Proc of the 2008 ACM Conf on Recommender Systems. New York: ACM, 2008: 123−130

[24] Nagatani K, Sato M. Accurate and diverse recommendation based on users' tendencies toward temporal item popularity[C]//Proc of the 1st Workshop on Temporal Reasoning in Recommender Systems Co-located with the 11th Int Conf on Recommender Systems (RecSys 2017). New York: ACM, 2017: 35−39

[25] Onuma K, Tong Hanghang, Faloutsos C. Tangent: A novel, 'surprise me', recommendation algorithm[C]//Proc of the 15th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2009: 657−666

[26] Nakatsuji M, Fujiwara Y, Tanaka A, et al. Classical music for rock fans? Novel recommendations for expanding user interests[C]//Proc of the 19th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2010: 949−958

[27] Oh J, Park S, Yu H, et al. Novel recommendation based on personal popularity tendency[C]//Proc of the 11th IEEE Int Conf on Data Mining. Piscataway, NJ: IEEE, 2011: 507−516

[28] Sha Chaofeng, Wu Xiaowei, Niu Junyu. A framework for recommending relevant and diverse items[C]//Proc of the 25th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2016: 3868−3874

[29] Ashkan A, Kveton B, Berkovsky S, et al. Optimal greedy diversity for recommendation[C]//Proc of the 24th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2015: 1742−1748

[30] He Jingrui, Tong Hanghang, Mei Qiaozhu, et al. Gender: A generic diversified ranking algorithm[C]//Proc of the 26th Annual Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2012: 1151−1159

[31] Carbonell J, Goldstein J. The use of MMR, diversity-based reranking for reordering documents and producing summaries[C]//Proc of the 21st Annual Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 1998: 335−336

[32] Cho E, Myers S A, Leskovec J. Friendship and mobility: User movement in location-based social networks[C]//Proc of the 17th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2011: 1082−1090

[33] Lu E H C, Lin C Y, Tseng V S. Trip-mine: An efficient trip planning

approach with travel time constraints[C]//Proc of the 12th IEEE Int Conf on Mobile Data Management. Piscataway, NJ: IEEE, 2011: 152−161

[34] Luan Wenjing, Liu Guanjun, Jiang Changjun, et al. MPTR: A maximal-marginal-relevance-based personalized trip recommendation method[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(11): 3461−3474

[35] Kulesza A, Taskar B. Determinantal point processes for machine learning[J]. Foundations and Trends in Machine Learning, 2012, 5(2/3): 123−286

[36] Chen Laming, Zhang Guoxin, Zhou Eric. Fast greedy map inference for determinantal point process to improve recommendation diversity[C]//Proc of the 32nd Advances in Neural Information Processing Systems. Cambridge, MA: MIT, 2018: 5627−5638

[37] Gillenwater J, Kulesza A, Mariet Z, et al. A tree-based method for fast repeated sampling of determinantal point processes[C]//Proc of the 36th Int Conf on Machine Learning. New York: PMLR, 2019: 2260−2268

[38] Gartrell M, Paquet U, Koenigstein N. Low-rank factorization of determinantal point processes[C]//Proc of the 31st AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2017: 1912−1918

[39] Warlop R, Mary J, Gartrell M. Tensorized determinantal point processes for recommendation[C]//Proc of the 25th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2019: 1605−1615

[40] Wilhelm M, Ramanathan A, Bonomo A, et al. Practical diversified recommendations on Youtube with determinantal point processes[C]//Proc of the 27th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2018: 2165−2173

[41] Gan Mingxin, Jiang Rui. Improving accuracy and diversity of personalized recommendation through power law adjustments of user similarities[J]. Decision Support Systems, 2013, 55(3): 811−821

[42] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks[C]//Proc of the 4th ACM Conf on Recommender Systems. New York: ACM, 2010: 135−142

[43] Wasilewski J, Hurley N. Incorporating diversity in a learning to rank recommender system[C]//Proc of the 29th Int Florida Artificial Intelligence Research Society Conf. Palo Alto, CA: AAAI, 2016: 572−578

[44] Gogna A, Majumdar A. DiABlO: Optimization based design for improving diversity in recommender system[J]. Information Sciences, 2017, 378: 59−74

[45] Shi Yue, Zhao Xiaoxue, Wang Jun, et al. Adaptive diversification of recommendation results via latent factor portfolio[C]//Proc of the 35th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2012: 175−184

[46] Wu Le, Liu Qi, Chen Enhong, et al. Relevance meets coverage: A unified framework to generate diversified recommendations[J]. ACM Transactions on Intelligent Systems and Technology, 2016, 7(3): 1−30

[47] Zanitti M, Kosta S, Sørensen J. A user-centric diversity by design recommender system for the movie application domain[C]//Proc of the 27th Web Conf. New York: ACM, 2018: 1381−1389

[48] Yang Chao, Ai Congcong, Li Renfa. Neighbor diversification-based collaborative filtering for improving recommendation lists[C]//Proc of the 10th IEEE Int Conf on High Performance Computing and Communications & IEEE Int Conf on Embedded and Ubiquitous Computing. Piscataway, NJ: IEEE, 2013: 1658−1664

[49] Chen H, Karger D R. Less is more: Probabilistic models for retrieving fewer relevant documents[C]//Proc of the 29th Annual Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2006: 429−436

[50] Xia Long, Xu Jun, Lan Yanyan, et al. Adapting Markov decision process for search result diversification[C]//Proc of the 40th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2017: 535−544

[51] Cooper C, Lee S H, Radzik T, et al. Random walks in recommender systems: Exact computation and simulations[C]//Proc of the 23rd Int Conf on World Wide Web. New York: ACM, 2014: 811−816

[52] Nikolakopoulos A N, Karypis G. Recwalk: Nearly uncoupled random walks for top-*n* recommendation[C]//Proc of the 25th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2019: 150−158

[53] Christoffel F, Paudel B, Newell C, et al. Blockbusters and wallflowers: Accurate, diverse, and scalable recommendations with random walks[C]//Proc of the 9th ACM Conf on Recommender Systems. New York: ACM, 2015: 163−170

[54] Liu Jianguo, Shi Kerui, Guo Qiang. Solving the accuracy-diversity dilemma via directed random walks[J]. arXiv preprint, arXiv: 1201.6278, 2012

[55] Paudel B, Bernstein A. Random walks with erasure: Diversifying personalized recommendations on social and information networks[C]//Proc of the 30th Web Conf. New York: ACM, 2021: 2046−2057

[56] Wang Mengsha, Xiao Yingyuan, Zheng Wenguang, et al. RNDM: A random walk method for music recommendation by considering novelty, diversity, and mainstream[C]//Proc of the 30th Int Conf on Tools with Artificial Intelligence (ICTAI). Piscataway, NJ: IEEE, 2018: 177−183

[57] Antikacioglu A, Ravi R. Post processing recommender systems for diversity[C]//Proc of the 23rd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2017: 707−716

[58] Adomavicius G, Kwon Y O. Optimization-based approaches for maximizing aggregate recommendation diversity[J]. INFORMS Journal on Computing, 2014, 26(2): 351−369

[59] Zhai Shuangfei, Chang Kenghao, Zhang Ruofei, et al. Deepintent: Learning attentions for online advertising with recurrent neural networks[C]//Proc of the 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1295−1304

[60] Shan Ying, Hoens T R, Jiao Jian, et al. Deep crossing: Web-scale modeling without manually crafted combinatorial features[C]//Proc of the 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 255−262

[61] Mahajan K C, Palnitkar A, Raul A, et al. CAViaR: Context aware

video recommendations[C]//Proc of the 32nd Web Conf. New York: ACM, 2023: 518−522

[62] He Yifan, Zou Haitao, Yu Hualong, et al. Diversity-aware recommendation by user interest domain coverage maximization[C]// Proc of the 19th Int Conf on Data Mining (ICDM). Piscataway, NJ: IEEE, 2019: 1084−1089

[63] Lin Zihan, Wang Hui, Mao Jingshu, et al. Feature-aware diversified re-ranking with disentangled representations for relevant recommendation[C]//Proc of the 28th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2022: 3327−3335

[64] Chen Wanyu, Ren Pengjie, Cai Fei, et al. Improving end-to-end sequential recommendations with intent-aware diversification[C]// Proc of the 29th ACM Int Conf on Information & Knowledge Management. New York: ACM, 2020: 175−184

[65] Chen Wanyu, Ren Pengjie, Cai Fei, et al. Multi-interest diversification for end-to-end sequential recommendation[J]. ACM Transactions on Information Systems, 2021, 40(1): 1−30

[66] Choi S, Kim H, Gim M. Do not read the same news! Enhancing diversity and personalization of news recommendation[C]//Proc of the 31st Web Conf. New York: ACM, 2022: 1211−1215

[67] Wang Jie, Zhou Jinya, Wu Zhen, et al. Toward paper recommendation by jointly exploiting diversity and dynamics in heterogeneous information networks[C]//Proc of the 27th Int Conf on Database Systems for Advanced Applications. Berlin: Springer, 2022: 272−280

[68] Zhang Mi. Enhancing diversity in top-*n* recommendation[C]//Proc of the 3rd ACM Conf on Recommender Systems. New York: ACM, 2009: 397−400

[69] Vargas S, Castells P. Exploiting the diversity of user preferences for recommendation[C]//Proc of the 10th Conf on Open Research Areas in Information Retrieval. New York: ACM, 2013: 129−136

[70] Di Noia T, Ostuni V C, Rosati J, et al. An analysis of users' propensity toward diversity in recommendations[C]//Proc of the 8th ACM Conf on Recommender Systems. New York: ACM, 2014: 285−288

[71] Gogna A, Majumdar A. Balancing accuracy and diversity in recommendations using matrix completion framework[J]. Knowledge-Based Systems, 2017, 125: 83−95

[72] Goldstein T, Osher S. The split Bregman method for L1-regularized problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(2): 323−343

[73] Huang Huimin, Shen Hong, Meng Zaiqiao. Item diversified recommendation based on influence diffusion[J]. Information Processing & Management, 2019, 56(3): 939−954

[74] Lathia N, Hailes S, Capra L, et al. Temporal diversity in recommender systems[C]//Proc of the 33rd Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2010: 210−217

[75] Hao Bin, Zhang Min, Guo Cheng, et al. Diversify or not: Dynamic diversification for personalized recommendation[C]//Proc of the 25th Pacific-Asia Conf on Knowledge Discovery and Data Mining. Berlin: Springer, 2021: 461−472

[76] Chen Yankai, Yang Yaming, Wang Yujing, et al. Attentive knowledge-aware graph convolutional networks with collaborative guidance for personalized recommendation[C]//Proc of the 38th IEEE Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2022: 299−311

[77] Zheng Yu, Gao Chen, Chen Liang, et al. DGCN: Diversified recommendation with graph convolutional networks[C]//Proc of the 30th Web Conf. New York: ACM, 2021: 401−412

[78] Kang Wang-Cheng, McAuley J. Self-attentive sequential recommendation[C]//Proc of the 18th Int Conf on Data Mining (ICDM). Piscataway, NJ: IEEE, 2018: 197−206

[79] Lv Fuyu, Jin Taiwei, Yu Changlong, et al. SDM: Sequential deep matching model for online large-scale recommender system[C]// Proc of the 28th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2019: 2635−2643

[80] Cen Yukuo, Zhang Jianwei, Zou Xu, et al. Controllable multi-interest framework for recommendation[C]//Proc of the 26th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2020: 2942−2951

[81] Hurley N J. Personalised ranking with diversity[C]//Proc of the 7th ACM Conf on Recommender Systems. New York: ACM, 2013: 379−382

[82] Takács G, Tikk D. Alternating least squares for personalized ranking[C]//Proc of the 6th ACM Conf on Recommender Systems. New York: ACM, 2012: 83−90

[83] Li Chao, Liu Zhiyuan, Wu Mengmeng, et al. Multi-interest network with dynamic routing for recommendation at Tmall[C]//Proc of the 28th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2019: 2615−2623

[84] Lu Yujie, Zhang Shengyu, Huang Yingxuan, et al. Future-aware diverse trends framework for recommendation[C]//Proc of the 30th Web Conf. New York: ACM, 2021: 2992−3001

[85] Gu Wanrong, Dong Shoubin, Zeng Zhizhao. Increasing recommended effectiveness with Markov chains and purchase intervals[J]. Neural Computing and Applications, 2014, 25(5): 1153−1162

[86] Hu Liang, Cao Longbing, Wang Shoujin, et al. Diversifying personalized recommendation with user-session context[C]//Proc of the 26th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2017: 1858−1864

[87] Wu Chuhan, Wu Fangzhao, Qi Tao, et al. SentiRec: Sentiment diversity-aware neural news recommendation[C]//Proc of the 1st Conf of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th Int Joint Conf on Natural Language Processing. Stroudsburg, PA: ACL, 2020: 44−53

[88] Hong Minsung. Decrease and conquer-based parallel tensor factorization for diversity and real-time of multi-criteria recommendation[J]. Information Sciences, 2021, 562: 259−278

[89] Lü Linyuan, Zhou Tao. Link prediction in complex networks: A survey[J]. Physica A: Statistical Mechanics and Its Applications, 2011, 390(6): 1150−1170

[90] Sanz-Cruzado J, Castells P. Enhancing structural diversity in social networks by recommending weak ties[C]//Proc of the 12th ACM

Conf on Recommender Systems. New York: ACM, 2018: 233−241

[91] Cheng Peizhe, Wang Shuaiqiang, Ma Jun, et al. Learning to recommend accurate and diverse items[C]//Proc of the 26th Int Conf on World Wide Web. New York: ACM, 2017: 183−192

[92] Liang Yile, Qian Tieyun, Li Qing, et al. Enhancing domain-level and user-level adaptivity in diversified recommendation[C]//Proc of the 44th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2021: 747−756

[93] Huang Yanhua, Wang Weikun, Zhang Lei, et al. Sliding spectrum decomposition for diversified recommendation[C]//Proc of the 27th ACM SIGKDD Conf on Knowledge Discovery & Data Mining. New York: ACM, 2021: 3041−3049

[94] Yang Ji, Yi Xinyang, Zhiyuan Cheng D, et al. Mixed negative sampling for learning two-tower neural networks in recommendations[C]//Proc of the 29th Web Conf. New York: ACM, 2020: 441−447

[95] Raza S, Bashir S R, Naseem U. Accuracy meets diversity in a news recommender system[C]//Proc of the 29th Int Conf on Computational Linguistics. New York: ACM, 2022: 3778−3787

[96] Zhang Xiaoying, Wang Hongning, Li Hang. Disentangled representation for diversified recommendations[C]//Proc of the 16th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2023: 490−498

[97] Shi Xiaoyu, Liu Quanliang, Xie Hong, et al. Relieving popularity bias in interactive recommendation: A diversity-novelty-aware reinforcement learning approach[J]. ACM Transactions on Information Systems, 2023, 42(2): 1−30

[98] Chen Zhihong, Wu Jiawei, Li Chenliang, et al. Co-training disentangled domain adaptation network for leveraging popularity bias in recommenders[C]//Proc of the 45th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2022: 60−69

[99] Gao Zhaolin, Shen Tianshu, Mai Zheda, et al. Mitigating the filter bubble while maintaining relevance: Targeted diversification with VAE-based recommender systems[C]//Proc of the 45th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2022: 2524−2531

[100] Kim B, Wattenberg M, Gilmer J, et al. Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav)[C]//Proc of the 35th Int Conf on Machine Learning. New York: PMLR, 2018: 2668−2677

[101] Wu Qiong, Liu Yong, Miao Chunyan, et al. PD-GAN: Adversarial learning for personalized diversity-promoting recommendation[C]// Proc of the 28th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2019: 3870−3876

[102] Liu Shuchang, Cai Qingpeng, He Zhankui, et al. Generative flow network for listwise recommendation[J]. arXiv preprint, arXiv: 2306.02239, 2023

[103] Wang Wenjie, Feng Fuli, Nie Liqiang, et al. User-controllable recommendation against filter bubbles[C]//Proc of the 45th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2022: 1251−1261

[104] Wang Wenjie, Feng Fuli, He Xiangnan, et al. Deconfounded recommendation for alleviating bias amplification[C]//Proc of the 27th ACM SIGKDD Conf on Knowledge Discovery & Data Mining. New York: ACM, 2021: 1717−1725

[105] Gan Lu, Nurbakova D, Laporte L, et al. Enhancing recommendation diversity using determinantal point processes on knowledge graphs[C]//Proc of the 43rd Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2020: 2001−2004

[106] Sun Jianing, Guo Wei, Zhang Dengcheng, et al. A framework for recommending accurate and diverse items using Bayesian graph convolutional neural networks[C]//Proc of the 26th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2020: 2030−2039

[107] Pal S, Regol F, Coates M. Bayesian graph convolutional neural networks using node copying[J]. arXiv preprint, arXiv: 1911.04965, 2019

[108] McAuley J, Pandey R, Leskovec J. Inferring networks of substitutable and complementary products[C]//Proc of the 21st ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2015: 785−794

[109] Wang Zihan, Jiang Ziheng, Ren Zhaochun, et al. A path-constrained framework for discriminating substitutable and complementary products in e-commerce[C]//Proc of the 11th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2018: 619−627

[110] Hao Junheng, Zhao Tong, Li Jin, et al. P-companion: A principled framework for diversified complementary product recommendation[C]//Proc of the 29th ACM Int Conf on Information & Knowledge Management. New York: ACM, 2020: 2517−2524

[111] Ye Rui, Hou Yuqing, Lei Te, et al. Dynamic graph construction for improving diversity of recommendation[C]//Proc of the 15th ACM Conf on Recommender Systems. New York: ACM, 2021: 651−655

[112] He Qiang, Zhou Rui, Zhang Xuyun, et al. Keyword search for building service-based systems[J]. IEEE Transactions on Software Engineering, 2016, 43(7): 658−674

[113] He Qiang, Zhou Rui, Zhang Xuyun, et al. Efficient keyword search for building service-based systems based on dynamic programming[C]//Proc of the 15th Int Conf on Service-Oriented Computing. Berlin: Springer, 2017: 462−470

[114] Gong Wenwen, Zhang Xuyun, Chen Yifei, et al. DAWAR: Diversity-aware web APIs recommendation for mashup creation based on correlation graph[C]//Proc of the 45th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2022: 395−404

[115] Yang Yonghui, Wu Le, Hong Richang, et al. Enhanced graph learning for collaborative filtering via mutual information maximization[C]//Proc of the 44th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2021: 71−80

[116] Wu Jiancan, Wang Xiang, Feng Fuli, et al. Self-supervised graph learning for recommendation[C]//Proc of the 44th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2021: 726−735

[117] Ma Xiyao, Hu Qian, Gao Zheng, et al. Contrastive co-training for

diversified recommendation[C]//Proc of the 20th Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2022: 4422−4430

[118] Chu Wei, Li Lihong, Reyzin L, et al. Contextual bandits with linear payoff functions[C]//Proc of the 14th Int Conf on Artificial Intelligence and Statistics. New York: PMLR, 2011: 208−214

[119] Liu Yong, Xiao Yingtai, Wu Qiong, et al. Diversified interactive recommendation with implicit feedback[C]//Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 4932−4939

[120] Liang Dawen, Charlin L, McInerney J, et al. Modeling user exposure in recommendation[C]//Proc of the 25th Int Conf on World Wide Web. New York: ACM, 2016: 951−961

[121] Qin Lijing, Chen Shouyuan, Zhu Xiaoyan. Contextual combinatorial bandit and its application on diversified online recommendation[C]//Proc of the 14th Int Conf on Data Mining. Philadelphia, PA: SIAM, 2014: 461−469

[122] Ding Qinxu, Liu Yong, Miao Chunyan, et al. A hybrid bandit framework for diversified recommendation[C]//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 4036−4044

[123] Stamenkovic D, Karatzoglou A, Arapakis I, et al. Choosing the best of both worlds: Diverse and novel recommendations through multi-objective reinforcement learning[C]//Proc of the 15th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2022: 957−965

[124] Zheng Guanjie, Zhang Fuzheng, Zheng Zihan, et al. DRN: A deep reinforcement learning framework for news recommendation[C]//Proc of the 27th World Wide Web Conf. New York: ACM, 2018: 167−176

[125] Hansen C, Mehrotra R, Hansen C, et al. Shifting consumption towards diverse content on music streaming platforms[C]//Proc of the 14th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2021: 238−246

[126] Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double q-learning[C]//Proc of the 30th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 2094−2100

[127] Gao Yixu, Shao Kun, Duan Zhijian, et al. Efficient dual-process cognitive recommender balancing accuracy and diversity[C]//Proc of the 27th Int Conf on Database Systems for Advanced Applications. Berlin: Springer, 2022: 389−400

[128] GroupLens Research. MovieLens[EB/OL]. [2023-07-14]. https://movielens.org/

[129] Douban. Douban[EB/OL]. [2024-07-25]. https://www.douban.com/ (in Chinese)
(豆瓣. 豆瓣[EB/OL]. [2024-07-25]. https://www.douban.com/)

[130] Netflix. Netflix[EB/OL]. [2024-07-25]. https://www.netflix.com

[131] McAuley J. Amazon Dataset[EB/OL]. [2024-07-25]. https://jmcauley.ucsd.edu/data/amazon/

[132] MyAnimeList. MyAnimeList[EB/OL]. [2024-07-25]. https://myanimelist.net/

[133] Yelp. Yelp Dataset[EB/OL]. [2024-07-25]. https://www.yelp.com/

[134] Aliyun Tianchi. Tmail Dataset[EB/OL]. [2024-07-25]. https://tianchi.aliyun.com/dataset/dataDetail?dataId=53 (in Chinese)
(阿里云天池. 天猫数据集[EB/OL]. [2024-07-25]. https://tianchi.aliyun.com/dataset/dataDetail?dataId=53)

[135] Last. fm. Last. fm[EB/OL]. [2024-07-25]. https://www.last.fm/

[136] MSNews. MSNews[EB/OL]. [2024-07-25]. https://msnews.github.io/

[137] Tamber. Steam video games dataset[EB/OL]. [2024-07-25]. https://www.kaggle.com/datasets/tamber/steam-video-games

[138] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems[J]. ACM Transactions on Information Systems, 2004, 22(1): 5−53

[139] Herlocker J L, Konstan J A, Borchers A, et al. An algorithmic framework for performing collaborative filtering[C]//Proc of the 22nd Annual Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 1999: 230−237

[140] Middleton S E, Shadbolt N R, De Roure D C. Ontological user profiling in recommender systems[J]. ACM Transactions on Information Systems, 2004, 22(1): 54−88

[141] Zhang Yuancao, Séaghdha D Ó, Quercia D, et al. Auralist: Introducing serendipity into music recommendation[C]//Proc of the 5th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2012: 13−22

[142] Li Jinming, Zhang Wentao, Wang Tian, et al. GPT4Rec: A generative framework for personalized recommendation and user interests interpretation[J]. arXiv preprint, arXiv: 2304.03879, 2023



Peng Yingtao, born in 1993. PhD candidate. Student member of CCF. His main research interests include recommender system, and knowledge graph and large language model.

彭迎涛, 1993 年生. 博士研究生. CCF 学生会员. 主要研究方向为推荐系统、知识图谱与大语言模型.



Meng Xiaofeng, born in 1964. PhD, professor, PhD supervisor. Fellow of CCF. His main research interests include database systems, data intelligence, data privacy and governance, and social computing.

孟小峰, 1964 年生. 博士, 教授, 博士生导师, CCF 会士. 主要研究方向为数据库系统、数据智能、数据隐私与治理、社会计算.



Du Zhijuan, born in 1986. PhD, associate professor. Member of CCF. Her main research interests include knowledge graph and big data management.

杜治娟, 1986 年生. 博士, 副教授. CCF 会员. 主要研究方向为知识图谱、大数据管理.