

文章编号: 2095-2163(2023)05-0001-09

中图分类号: TP391

文献标志码: A

面向新领域的推荐系统综述

让冉¹, 邢林林¹, 张龙波¹, 蔡红珍²

(1 山东理工大学 计算机科学与技术学院, 山东 淄博 255000;

2 山东理工大学 农业工程与食品科学学院, 山东 淄博 255000)

摘要: 互联网时代的快速演进带来了数据信息的海量增长, 推荐系统旨在海量数据中提升用户获取信息的有效性。同时推荐系统促进领域的发展并带来了新的机遇, 现阶段各行业在应用中涌现出大量新领域下为用户进行个性化推荐的需求。然而, 推荐系统在新的领域进行实际场景应用, 往往需要从零开始构架数据体系, 这依赖专家对领域特征进行分析, 总结获取新领域的数据关系, 并且存在推荐系统应用领域限定性强、数据稀疏、冷启动等阻碍。本文旨在面向新领域推荐系统的构建这一主题进行综述, 从新领域背景和挑战、新领域推荐方法、方法评估3个研究方向, 介绍了相关工作的研究现状, 给出了研究面向新领域技术应用的建议, 并对新领域推荐的发展趋势进行展望, 指出下一步需要开展的工作。

关键词: 推荐系统; 新领域; 数据稀疏; 冷启动; 协同过滤

Survey of recommendation systems for new domains

RANG Ran¹, XING Linlin¹, ZHANG Longbo¹, CAI Hongzhen²

(1 School of Computer Science and Technology, Shandong University of Technology, Zibo Shandong 255000, China;

2 School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo Shandong 255000, China)

[Abstract] Nowadays, the rapid evolution of the Internet era has brought about the massive growth of data and information, and the recommendation system aims to improve the effectiveness of users' access to information in the massive data. At the same time, recommendation systems promote the development of the domain and bring new opportunities, and at this stage, there are a lot of demands for personalized recommendations for users in new domains emerging from various industries in the application. However, recommendation systems for practical scenario applications in new domains often require framing the data system from scratch, which relies on experts to analyze domain features and summarize to obtain data relationships in new domains, and there are hindrances such as strong limitation of recommendation system application domains, sparse data, and cold start. This paper aims to review the topic of building recommendation systems for new domains. Specifically, it introduces the current research status of related work in three research directions: background and challenges of new domains, research on new domain recommendation methods, and evaluation of methods. This paper gives suggestions for researching the application of technologies oriented to new domains and gives an outlook on the development trend of new domain recommendation and points out the next work to be carried out.

[Key words] recommendation systems; new domains; data sparsity; cold-start; collaborative filtering

0 引言

推荐系统(Recommendation System, RS)^[1]主要是数据挖掘、预测算法、机器学习等多种学科结合而形成的一个新的研究领域, 旨在通过向用户提供相关建议, 帮助用户发现互联网上的用户兴趣。如今, 推荐系统已经成为在线用户体验中普遍存在的一部分, 充当信息过滤器的作用, 为用户提供符合需求的

个性化的信息。建立推荐系统的技术很多, 一般可以分为基于内容的推荐系统^[2]、基于协同过滤的推荐系统^[3]和基于混合的推荐系统^[4]。基于内容的推荐方法, 利用交互数据以及辅助信息, 推荐与用户过去偏好相似的项目; 基于协同过滤的方法是通过用户以往的交互信息, 直接预测用户的兴趣, 并从过去有类似兴趣和偏好的其他用户向用户推荐项目; 混合方法结合了基于内容和基于项目的方法。推荐系统能

基金项目: 国家重点研发计划项目(2018YFB1403302)。

作者简介: 让冉(1998-), 女, 硕士研究生, 主要研究方向: 推荐系统; 邢林林(1987-), 男, 博士, 讲师, 硕士生导师, CCF会员, 主要研究方向: 生物信息学、推荐系统、机器学习等; 张龙波(1968-), 男, 博士, 教授, 硕士生导师, 主要研究方向: 数据库理论与应用、数据挖掘; 蔡红珍(1972-), 女, 博士, 教授, 博士生导师, 主要研究方向: 生物基材料。

通讯作者: 邢林林 Email: xinglinlin@sdut.edu.cn

收稿日期: 2022-05-10

根据用户的兴趣偏好为用户进行个性化推荐,为用户提供新的、感兴趣的内容,并且能为用户推荐关注领域的最新进展、历史脉络以及有价值信息,这对于消除信息壁垒,增加信息价值起到重要作用。

最近,无论学术界还是工业界都对于推荐系统应用于实际场景中的研究兴趣大增。目前推荐系统模型的研究主要利用 Movielens^[5]、Yelp^[6]等公开数据集,但在实际应用场景中,推荐效果会受到数据规模^[7]、领域的独特性、上下文信息^[8]等多种因素的影响。因此,依照公开数据集开发的模型,使用在特定领域实际应用场景中,往往导致用户对于推荐结果满意度较低的现象。

从近年来构建新领域实际应用场景中推荐系统的相关研究来看,新领域推荐系统的方法领域限定性强,并且推荐效果依赖于数据的质量以及领域特征,针对在新领域从零构建高质量推荐系统受到的多方面困难和阻碍进行总结分析^[9]。

在此背景下,本文通过讨论现阶段技术研究现状以及领域内的限制,详细阐述了现在新领域^[10]推荐系统实际应用场景中最紧迫的挑战。虽然在技术研究、领域限制等问题上,许多研究人员利用公开数据集进行了大量的研究,但是本文讨论的重点在于模型在新领域中实际场景的应用。由于无法全面包含所有实际应用场景中的局限与挑战,本文将重点放在新领域推荐系统实际应用场景中的冷启动问题以及数据稀疏性高等问题的研究上。

1 新领域背景和挑战分析

1.1 面向新领域推荐

新领域非特指某些领域的集合,而是强调在某

领域中需要从零构建一个有效的推荐系统^[10]。对于推荐系统应用而言,此领域为一个全新的领域。

1.1.1 新领域构建推荐系统特点

(1)新颖性。新颖性包括领域新颖以及数据新颖。领域的新颖性在一定程度上阻碍着推荐方法的应用,但是存在推荐模型如协同过滤^[11]、基于内容^[12]等模型无需领域知识,领域新颖并不能成为影响推荐效果的重要因素。数据新颖,即是在此领域内的数据没有构成完成推荐系统模型建立适合的数据形式,需要人工进行数据处理。

(2)分散性。分散性主要表现在两个方面:一是数据没有固定获取方式^[13]。新领域的数据获取需要多渠道收集数据,除了网络、图书馆书籍资料、新闻、杂志等媒体的信息,还需关注统计局、行业协会、研究机构的数据或者直接做市场调研;二是数据的零散分布,需要逻辑相互关联。初步获取到的数据,都是零散的,需要充分利用数据,寻找数据的关联性^[14]。

(3)再创性。再创性指的是新领域中能得到的信息只是一组组数据或是一种现象和启示。研究人员需要充分研究利用新领域推荐中的有限数据^[15],发现推荐目标与被推荐对象之间复杂的交互关系,进一步分析构造更为高效的推荐模型,从而获得更准确的推荐结果。

1.1.2 面向新领域构建推荐系统框架

如图1所示,面向新领域推荐系统的一般构建框架主要包括数据采集、数据处理、推荐方法、模型评估4个方面。

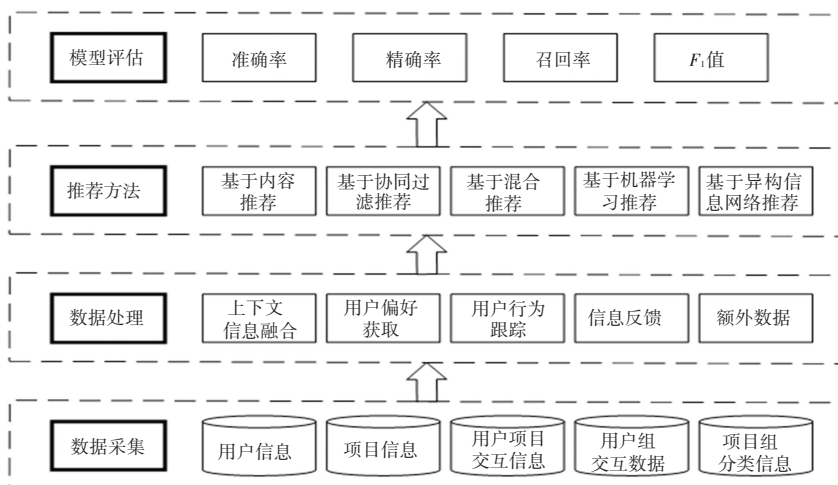


图1 面向新领域推荐实施框架

Fig. 1 Recommendation framework for new domains

(1)数据采集:面向新领域进行推荐所需要的数据,主要根据所研究领域的特殊性进行获取。若在此阶段领域内有较完备的数据库,则数据环节重点将放在数据处理上,否则,需要构建领域数据库体系,则需要多方面获取资源数据。

(2)数据处理:数据处理阶段是影响推荐效果最主要的环节。主要包括上下文信息融合、用户偏好获取、用户行为跟踪、信息反馈、额外信息等方面^[15]。

(3)推荐方法:推荐方法需要结合领域特点、交互数据、辅助信息,建立高效的推荐系统,旨在为用户进行个性化推荐。

(4)模型评估:推荐系统为用户生成推荐结果后,通过评价指标对推荐方法逐步优化,提高推荐结果质量。

1.2 新领域推荐面临的挑战

目前,推荐系统模型的建立及优化主要聚焦在公开数据集下进行研究,而且大部分模型在应用需求领域的限定性强,而本文则就新领域从零开始构建推荐系统将会受到的多方面困难和挑战进行分析。

1.2.1 数据预处理。

众多推荐模型需要在已经形成完备结构的数据集基础上进行模型训练。但是对于新领域而言,没有较全面的推荐目标与被推荐目标的数据信息,更多的数据信息需要数据采集、数据处理。数据预处理方法需根据数据集自身特点分析、模型类型分析、以及问题种类分析^[15]等方面进行。

1.2.2 领域专业因素

现阶段大部分推荐系统的研究主要为:电影、音乐、新闻等领域,但更多的新领域实际应用推荐涉及新领域问题。电影领域中,用户通常不喜欢被重复推荐相同的电影;但是在音乐领域,用户可能在稍后的时间希望再次被推荐相同的音乐片段。不同的领域具有独特的属性^[16]。在不同领域中进行推荐,要结合领域相对应的特殊性^[17]。不同领域用户所处交互环境不同,影响推荐效果的因素也有所区别。

1.2.3 冷启动

冷启动问题是新领域推荐系统实现过程中面临的挑战之一。

冷启动主要分为3大类:

(1)用户冷启动:新的用户注册到系统,为新用户进行个性化推荐。

(2)项目冷启动:新项目被添加到系统中,将新项目推荐给可能对其感兴趣的用户。

(3)系统冷启动:向新用户个性化推荐新项目,

在新领域上进行推荐(没有用户,也没有用户行为,只有一些项目的信息),设计个性化推荐系统。

一般的推荐系统需要根据用户的历史交互信息来预测用户未来的行为,因此需要用户的大量历史交互数据信息。但是对新用户和新项目来说,没有相对应的历史交互,数据的缺失直接导致用户与用户、项目与项目之间无法计算相似性,无法预测评分,进而无法向新用户进行高质量推荐,新项目无法被推荐^[18]。

1.2.4 数据稀疏

数据稀疏是指用户与项目之间、用户与用户的交互信息在交互矩阵中所占的比例过小,即大多数事件/项目只有少数参与者,使用户-项目矩阵出现极端稀疏性,导致推荐效果不理想。为了减少数据稀疏带来的影响,一些算法仅应用评分或签到数据作为数据源进行推荐。为了提高推荐性能,改进算法考虑了添加额外信息(商品属性以及用户信息)或利用迁移学习,尝试从源域学习知识应用到目标域来实现更好的推荐^[19]。

2 面向新领域推荐方法

本节主要介绍面向新领域推荐方法,以及结合方法分析所适用的领域,并且将推荐系统中的主流方法进行对比。

2.1 基于内容推荐方法

基于内容的协同过滤,是根据项目或内容的元数据,发现项目或内容的相关性,然后基于用户以前的喜好记录给用户推荐相似的项目。图2是以电影推荐系统为例的基于内容推荐方法示意。

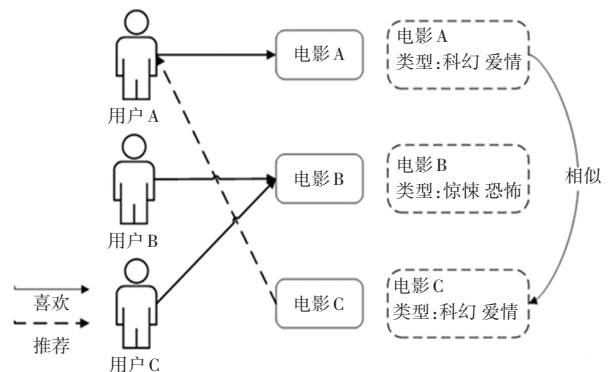


图2 基于内容推荐方法

Fig. 2 Content-based recommendation method

Guo等^[20]在2019年提出了采用TF-IDF结合LDA主题模型来解决用户兴趣和行为标签的缺失。此模型侧重于根据司法案件内容进行推荐的方法,

将中文自身的特殊性考虑进模型,使用"jieba"文本分割来预处理案例。

Chang^[21]等在 2021 年将推荐系统应用在寄宿家庭推荐中,提出利用大量用户对于寄宿家庭的评论数据,建立一个中文分词语料。使用 word2vec 培训语料库来获取寄宿家庭标签特征,然后利用 TF-IDF 对标签特征进行矢量化,通过余弦相似度计算特征向量的相似度值,得到新用户的推送值。

Ana 等人^[22]在相似性分析的上下文中比较距离测量值,并通过 Minkowski、Euclidean、Manhattan 距离、Bray-curtis 相似性等 4 个指标来衡量音乐的相似性,确定适合古典音乐相似性分析背景的特征提取和工程方法,并确定所选距离指标之间的性能差异,以考虑音乐的多维性质,为古典音乐构建推荐系统。基于内容推荐方法的优缺点分析详见表 1。

表 1 基于内容的推荐方法优缺点分析

Tab. 1 Analysis of the advantages and disadvantages of content-based recommendation methods

基于内容的推荐方法	
优势	1. 用户独立性高 2. 结果可解释性强 3. 解决新项目冷启动问题
劣势	1. 存在新用户冷启动 2. 无法挖掘用户的潜在兴趣 3. 缺乏多样性
新领域适用情况分析	适用于文本类领域,例如:新闻网站

2.2 协同过滤推荐方法

协同过滤推荐技术是推荐系统中应用最早、最为成功的技术之一。采用最近邻技术,利用用户的历史喜好信息计算用户之间的距离,通过目标用户的最近邻居用户对商品评价的加权评价,来预测目标用户对特定商品的喜好程度,系统则根据这一喜好程度对目标用户进行推荐。

2.2.1 基于内存的协同过滤

基于内存的协同过滤^[23],分为基于用户和基于项目。基于用户的协同过滤推荐算法首先输入项目的用户评分数据矩阵,使用统计技术计算用户之间的相似度,寻找与目标用户相似的邻居集,然后根据目标用户的邻居集中选择邻居集中得分最高的项目集合,生成目标用户的 TOP-N 推荐。基本原理就是利用用户访问行为的相似性来互相推荐用户可能感兴趣的资源,如图 3 所示。

基于项目的协同过滤推荐^[11]基本思想,是根据所有用户对项目或者信息的评价,发现项目和项目之间的相似度,然后根据用户的历史偏好信息,将相

似度较高的项目作为推荐结果推荐给该用户,如图 4 所示。

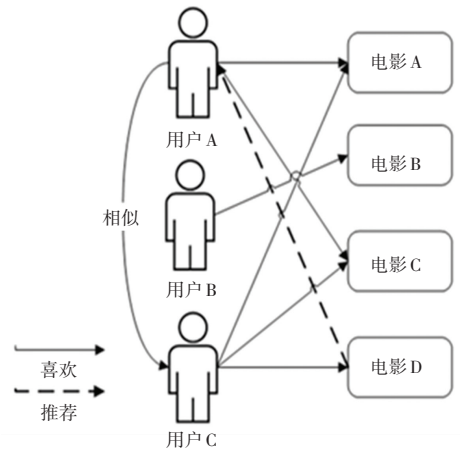


图 3 基于用户的协同过滤推荐方法

Fig. 3 User-based collaborative filtering recommendation method

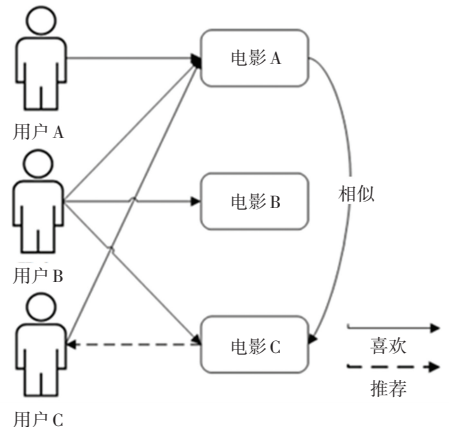


图 4 基于项目的协同过滤推荐方法

Fig. 4 Item-based collaborative filtering recommendation method

2020 年 Musa^[11]等提出了两种不同的方法来计算这些相似性权重的项目,并选择最好的项目来建立模型。为新领域进行基于协同过滤算法的推荐,不可避免存在数据稀疏、冷启动和推荐不准确问题。为了克服这些问题,可以参考 2020 年 Ullah 等^[23]提出的基于集成积分的矩阵分配协同过滤方法。

在新领域实际推荐应用中,用户的兴趣常随着时间而变化。Hui 等人^[24]在 2020 年提出了改进基于用户兴趣扩散的用户综合相似度计算方法,计算用户兴趣的直接相似度和用户兴趣扩散的相似度,通过参数调整,得到用户兴趣的综合相似度。针对用户兴趣随时间变化的情况,将时间相关函数应用于用户之间的相似度计算。然而该算法依然存在推荐效率低,以及启动时间长等问题。2021 年 Wu 等人^[25]根据用户的属性信息,采用基于用户属性特征优化的 K 平均聚类算法进行聚类,生成多个聚类,

结合每个聚类中的用户属性特征, 建立一个新的相似度计算模型, 通过聚类中搜索最近邻, 生成推荐列表来实现推荐。

若面向的是非社交的新领域进行推荐, 领域中内容之间的关联将是推荐的一个重要原则, 基于内

容将比基于相似用户的推荐原则更加有效。相反, 若是社交领域中构建推荐系统, 协同过滤方法将优于基于内容的方法, 同时协同过滤加上社会网络信息, 可以增加用户对推荐解释的信服程度。基于用户与基于项目的协同过滤方法对比结果见表 2。

表 2 基于用户与基于项目的协同过滤方法对比

Tab. 2 Comparison of user-based and Item-based collaborative filtering methods

	基于用户的协同过滤方法	基于项目的协同过滤方法
算法思路	找到和目标用户兴趣相似的用户集合; 找到这个集合中的用户喜欢的, 且目标用户没有听说过的项目推荐给目标用户;	计算项目之间的相似度; 根据项目的相似度和用户的历史行为给用户生成推荐列表;
性能	用户数据增长, 导致用户相似度代价大, 推荐结果解释性不强;	项目数据增长, 导致项目相似度代价大, 矩阵推荐结果解释性强;
冷启动	用户相似度矩阵更新具有时间间隔, 在新用户对很少的项目产生交互时, 不能立即更新个性化推荐;	新用户和项目产生交互行为, 即可为新用户推荐相关产品;
实时性	新用户有新行为, 但推荐结果不一定立即变化;	新用户有新行为, 一定会导致推荐结果的实时变化;
领域	适用于用户较少且时效性较强, 用户个性化不明显的领域;	适用于项目数明显比用户数少的领域图书、电子商务、电影领域;
文献	[26-28]	[29-31]

2.2.2 基于模型的协同过滤

基于模型的协同过滤推荐方法, 是通过用户与项目的交互数据训练推荐模型, 预测未知项目评分进行结果推荐。主流方法包括: 关联算法^[32]、聚类算法^[33]、分类算法、回归算法、矩阵分解^[34]、神经网络^[35]、图模型^[36], 以及隐语义模型等等。在新领域推荐常伴随着在冷启动以及数据稀疏的情况下, 在现有的大多数工作中, 用户聚类被直接用来识别目标用户的相似用户, 并向其推荐。

Zhang 等^[37]利用用户聚类来重建用户-项目二分网络, 使网络密度显著提高。针对冷启动, 其中大部分的方法采用在数据层面^[38]上解决此问题。早期的方法通常将辅助信息^[39]作为用户或项目特征。

2020 年 Lu 等^[40]提出 MetaHIN 模型。该模型由语义增强任务构造器以及协同适应元学习器组成。语义增强任务构造器利用语义上下文的语义序列, 对任务进行多方面的语义编码, 以增加支持和查询集的用户任务与异构的语义上下文, 在一定程度上缓解了数据稀疏和冷启动带来的问题。基于模型的协同过滤方法分析结果详见表 3。

2021 年 Zhu^[36]利用 PV-DM 模型为每篇论文的标题生成数字特征, 通过结构化和文本表示相结合的“作者-文本查询”场景, 构建了一个双向的 RNN。以用户的身份为基础, 通过现场查询获取学者以前的写作和引用偏好, 将异构知识嵌入的注意

递归神经网络应用于科技论文推荐中。

表 3 基于模型的协同过滤方法分析

Tab. 3 Analysis of model-based collaborative filtering methods

	基于模型的推荐方法
优势	易解决数据稀疏性 结果可解释性强
劣势	建模成本高 数据信息降维导致部分信息丢失
新领域适用情况分析	适用于工业界应用

2.3 混合推荐方法

混合推荐方法是指同时使用多种推荐技术产生推荐, 从而通过结合加权、变换、混合、特征组合、层叠、特征扩充、元级别等混合策略、弥补算法不足, 获得更好的推荐效果。

Walek 与 Fojtik^[41]在 2020 年提出了预测单片混合推荐系统, 此系统利用 SVD 算法的合作式推荐系统、基于内容的系统来构建推荐模型, 模糊专家系统组成的推荐系统用于最终的排序建议。现有的大多数的协同过滤方法严重依赖显式反馈数据, 当数据稀疏性时, 这种协同过滤方法的效果不明显。Feng 等^[41]结合面向评分的概率矩阵分解和面向成对排序的贝叶斯个性化排序, 提出了新的 CF 排序模型。以上讨论的方法多是基于合作学习技术, 部分使用基于内容或混合过滤技术。Riyahi^[42]提出了基于内容、协作和混合过滤的新 CF, 使用 wordnet 词法数据

库提取标签的语义相关性,并根据标签的语义相关性,将标签组织成层次结构。基于内容的过滤部分,采用层次结构搜索相关帖子,使用相关语义标签扩

展用户查询。通过相似性度量,计算合作对象的隐含评分,最后将两部分结合到推荐系统的混合过滤部分。各种推荐系统方法对比结果详见表4。

表4 推荐系统方法对比

Tab. 4 Comparison of recommended system methods

	基于内容推荐	基于协同过滤推荐	混合推荐
优势	推荐结果解释性强 解决项目冷启动 适合小规模推荐	推荐个性化 自动化程度高 处理复杂非结构化对象 发现用户潜在偏好	不存在冷启动问题 发现用户潜在兴趣 克服数据稀疏问题 4.适合用户多的推荐
劣势	数据稀疏问题 新用户问题 复杂属性不好处理 需要足够数据构造分类器	数据稀疏问题 冷启动问题 推荐质量受数据集限制 系统初始推荐质量差	算法工作量大 推荐过程复杂
文献	[42]	[43]	[44]

表5整理了不同的推荐技术在音乐、社交、新闻、电影、视频、旅游6个典型领域的应用,并列出了

近年来相关领域较典型模型技术,以及需要的数据类型和数据集,并对模型的特点进行了归纳总结。

表5 各领域推荐方法对比

Tab. 5 Comparison of recommended methods by domain

推荐技术	代表性模型	数据类型	模型特点	常用数据集
音乐推荐	CB RNN	CoSeRNN ^[45] 音乐信息、歌手信息、项目持续时间、环境	CoSeRNN 学习用户的顺序倾听行为,结合捕获用户的长期音乐偏好的全局长期嵌入以及序列和上下文相关的偏移量,使模型适应当前的环境。	Spotify
社交网络	GNN	MEGCN ^[46] 用户、项目、交互链接、评分、链接密度、评分密度	基于通道稀疏性的图神经网络(MEGCN)利用 InfluenceNorm 和 ChannelNorm 来捕捉社交推荐任务的影响扩散过程中的自身兴趣和共同兴趣。	Yelp Flickr
新闻推荐	CB BERT	KRED ^[47] 实体频率、类别、位置	新闻文档的知识表示增强模型(KRED)聚焦知识图中邻域的信息来丰富实体的嵌入,主张采用多任务框架对模型进行优化,使不同的新闻推荐应用程序能够统一。	MSN News
电影推荐	CF MF DL Graph-based	RMIF ^[48] 用户、项目、评分、评分量表 GHRs ^[49] 用户、项目、评分、评分量表、数据密度、人口统计、位置信息	具有多重隐式反馈(RMIF)的推荐方法,模型分解了显式评分矩阵和隐式态度矩阵,引入用户相似度、项目的评分记录和用户的积极态度3个隐式反馈。 GHRs模型基于图的建模,结合用户的人口统计和位置信息。利用自动编码器特征提取对用户进行聚类,提高准确率,并通过用户在相似度图中作为节点的相似度来找到用户之间的关系,与用户的边信息相结合,解决冷启动问题。	Movielens
视频推荐	ML	TPM ^[50] 用户、交互信息、视频、会话次数、视频场次	此模型内容无关,需要顺序处理用户-项目交互可以应用于不同的推荐领域即社交媒体中的新闻推荐,甚至是基于患者电子健康记录(EHR)的药物推荐。	YOUTV
旅游推荐	MF 多辅助信息 DL 自注意力机制	PMF-MAI ^[51] 旅行出发(目的地)城市、财务成本、时间成本 TRKG ^[52] 用户、项目、时间间隔、点击流长度、购买的项目	具有多辅助信息的概率矩阵分解(PMF-MAI)模型,将用户-项目交互矩阵的概率矩阵分解与由多个辅助信息构建的一组特征的线性回归融合。 带有关键字生成的旅游推荐模型(Travel Recommendation with Keywords Generation, TRKG)对于项目文本内容建模而不是用户评论,在旅游推荐组件和关键词生成组件中整合了长期和短期的用户偏好。	Tuniu

3 新领域推荐模型性能评估

不同领域推荐任务使用的评价指标往往不同,就是希望推荐的结果都是用户感兴趣的结果,即用户感兴趣的信息比例要高。因为通常给用户展示的窗口有限,所以更要推荐给用户真正感兴趣的信息^[19]。现阶段模型的评级指标更多的侧重计算查准率、召回率等能体现模型优势的数值,而忽视了以用户角度评价模型推荐结果优劣的实验测试。

领域中评分预测任务需要预测准确度, TOP-N 推荐任务需要计算查准率、查全率等准确度指标,以及以用户角度为衡量标准的覆盖率、多样性、新颖性、惊喜度、实时性、健壮性等非准确度指标。

(1) 查准率:查准率是指所有预测为正类的结果中,真正的正类的比例。查准率可以表示推送给用户的内容用户是否感兴趣。计算公式如下:

$$Precision = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (1)$$

(2) 召回率:召回率也被称作查全率,是指所有正类中被分类器找出来的比例。公式如下:

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

(3) F1 分数:查准率和召回率是矛盾统一两个指标,为了提高精确率,分类器会把准确度更高的样本预测为正样本,但往往因为过于保守而漏掉正样本,导致召回率降低, F1 分数可以看作精确性和召回的结合。其计算公式如下:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

式中: $R(u)$ 表示推荐给 u 的项目列表, $T(u)$ 表示真实用户 u 的项目列表, U 表示所有用户合集。

(4) 多样性:推荐多样性类型包括个体多样性、系统多样性、时序多样性。个体多样性从单个用户的视角衡量推荐的多样性,衡量系统能够找到用户喜欢的冷门项目的能力;系统多样性强调不同用户推荐的不同;时序多样性指用户兴趣的动态变化,即与过去推荐相比,新的推荐所体现的多样性。设 $Sim(i, j) \in [0, 1]$ 为项目 i 和项目 j 的相似性。则用户 u 的推荐列表 $R(u)$ 的多样性可定义为

$$Diversity(R(u)) = \frac{\sum_{i, j \in R(u), i \neq j} (1 - Sim(i, j))}{\frac{1}{2} \cdot |R(u)|(|R(u)| - 1)} \quad (4)$$

(5) 新颖性:新颖性指推荐系统向用户推荐与其相似度低的项目的能力。一般来说,系统所推荐的项目中,用户之间的相似度越低,新颖度则越高。

(6) 稳定性:稳定性指推荐系统的预测结果在短期内变化较小,这样才能被用户所信任。假设已知用户评分集合 $R1$, 对应的预测评分集合为 $P1$, 经过一段时间,用户对未评分的项目进行评分,再对 $P1$ 中的项目进行预测评分,得到预测评分集合 $P2$, 则系统稳定性可以表示为

$$stability = \frac{1}{|P_2|} \sum_{(u,i) \in p_2} |P_2(u,i) - P_1(u,i)| \quad (5)$$

4 结束语

在当前的互联网时代,大量新领域下的推荐需求应运而生。如何在新领域下应用推荐技术已经引起研究人员的关注。涌现出大量新领域下为用户进行个性化推荐的需求,也会成为此领域研究动力。如何在新领域中进行高质量推荐,减少数据稀疏、冷启动的影响以及模型迁移领域限制问题研究解决都将成为未来继续的研究点。

通过讨论面向新领域推荐系统现阶段技术研究现状以及领域内的限制,详细的阐述了现在新领域推荐系统实际应用场景中最紧迫的挑战。并且对比了现有推荐系统模型的优缺点,并融入了不同模型能够克服新领域推荐系统出现的问题以及进行了总结,最后本文讨论了不同领域实际应用场景中模型评估方法,并对新领域推荐系统的未来发展做了展望,希望能对在面向新领域推荐感兴趣的研究人员提供帮助。

参考文献

- [1] LÜ L, MEDO M, YEUNG C H, et al. Recommender systems[J]. Physics reports, 2012, 519(1): 1-49.
- [2] PAZZANI M J, BILLSUS D. Content-based recommendation systems[J]. The adaptive web: methods and strategies of web personalization, 2007: 325-341.
- [3] GIRSANG A S, EDWIN A W. Song recommendation system using collaborative filtering methods[C]// Proceedings of the 2019 The 3rd International Conference on Digital Technology in Education, Yamanashi, Japan; 2019 of Conference. 2019: 160-162.
- [4] WALEK B, FOJTIK V. A hybrid recommender system for recommending relevant movies using an expert system[J]. Expert Systems with Applications, 2020, 158: 113452.
- [5] JI Y, SUN A, ZHANG J, et al. A revisit of the popularity baseline in recommender systems[C]// Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 1749-1752.

- [6] HE X, DENG K, WANG X, et al. Lightgcn: Simplifying and powering graph convolution network for recommendation [C]// Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. 2020; 639–648.
- [7] TÓTH B, SACHIDANANDAN S, JØRGENSEN E S. Balancing relevance and discovery to inspire customers in the IKEA app [C]//Proceedings of the 14th ACM Conference on Recommender Systems. 2020; 563.
- [8] RAMACHANDRAN L. Behavior – based popularity ranking on amazon video [C]//Proceedings of the 14th ACM Conference on Recommender Systems. 2020; 564–565.
- [9] SAHU A K, DWIVEDI P. User profile as a bridge in cross – domain recommender systems for sparsity reduction [J]. Applied Intelligence, 2019, 49; 2461–2481.
- [10] 黄河燕 刘啸. 面向新领域的事件抽取研究综述 [J]. 智能系统学报, 2022, 17(1); 201–221.
- [11] MUSA J M, ZHIHONG X. Item based collaborative filtering approach in movie recommendation system using different similarity measures [C]//Proceedings of the 2020 6th International Conference on Computer and Technology Applications. 2020; 31–34.
- [12] GUO Z, HE T, QIN Z, et al. A content – based recommendation framework for judicial cases [C]//Data Science; 5th International Conference of Pioneering Computer Scientists, Engineers and Educators, ICPCSEE 2019, Guilin, China, September 20–23, 2019, Proceedings, Part I 5. Springer Singapore, 2019; 76–88.
- [13] HU G, ZHANG Y, YANG Q. Conet: Collaborative cross networks for cross – domain recommendation [C]//Proceedings of the 27th ACM international conference on information and knowledge management. 2018; 667–676.
- [14] 闫广平. 云计算技术在计算机数据处理中的应用及其发展对策 [J]. 数字通信世界, 2022(5); 79–81, 84.
- [15] 李颜平, 吴刚. 基于典型数据集的数据预处理方法对比分析 [J]. 沈阳工业大学学报, 2022, 44(2); 185–192.
- [16] SCHEDL M, ZAMANI H, CHEN C W, et al. Current challenges and visions in music recommender systems research [J]. International Journal of Multimedia Information Retrieval, 2018, 7(2); 95–116.
- [17] 郭蕾, 刘文菊, 王贇, 等. 融合谱聚类和多因素影响的兴趣点推荐方法 [J]. 数据分析与知识发现, 2022, 6(5); 77–88.
- [18] HERCE – ZELAYA J, PORCEL C, BERNABÉ – MORENO J, et al. New technique to alleviate the cold start problem in recommender systems using information from social media and random decision forests [J]. Information Sciences, 2020, 536; 156–170.
- [19] 赵俊逸, 庄福振, 敖翔, 等. 协同过滤推荐系统综述 [J]. 信息安全学报, 2021, 6(5); 17–34.
- [20] GUO Z, HE T, QIN Z, et al. A content – based recommendation framework for judicial cases [C]//Data Science; 5th International Conference of Pioneering Computer Scientists, Engineers and Educators, ICPCSEE 2019, Guilin, China, September 20–23, 2019, Proceedings, Part I 5. Springer Singapore, 2019; 76–88.
- [21] CHANG Hao, WU Wei, SHU Xiaodong, et al. Application of content based recommendation system in homestay [C]// 2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS); 2021; 493–496.
- [22] CRUZ A F T, CORONEL A D. Towards developing a content – based recommendation system for classical music [C]// Information Science and Applications; ICISA 2019. Springer Singapore, 2020; 451–462.
- [23] SOVILJ D, SANNER S, SOH H, et al. Collaborative filtering with behavioral models [C]//Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization. 2018; 91–99.
- [24] HUI K, HOU H, XUE S. An improved collaborative filtering algorithm based on user interest diffusion and time correlation [C]//Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence. 2020; 38–42.
- [25] WU Q, CHENG X, SUN E, et al. Collaborative filtering algorithm based on optimized clustering and fusion of user attribute features [C]//2021 4th International Conference on Data Science and Information Technology. 2021; 125–129.
- [26] SOVILJ D, SANNER S, SOH H, et al. Collaborative filtering with behavioral models [C]//Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization. 2018; 91–99.
- [27] IFADA N, SOPHAN M K, PUTRI N F D, et al. A user – based normalization multi – criteria rating approach for hotel recommendation system [C]//6th International Conference on Sustainable Information Engineering and Technology 2021. 2021; 157–161.
- [28] SEYEDHOSEINZADEH K, RAHMANI H A, AFSHARCHI M, et al. Leveraging social influence based on users activity centers for point – of – interest recommendation [J]. Information Processing & Management, 2022, 59(2); 102858.
- [29] DA'U A, SALIM N, IDRIS R. An adaptive deep learning method for item recommendation system [J]. Knowledge – Based Systems, 2021, 213; 106681.
- [30] LIN Yuan, XU Bo, FENG Jiaojiao, et al. Knowledge – enhanced recommendation using item embedding and path attention [J]. Knowledge – Based Systems, 2021, 233; 107484.
- [31] SRILAKSHMI M, CHOWDHURY G, SARKAR S. Two – stage system using item features for next – item recommendation [J]. Intelligent Systems with Applications, 2022, 14; 200070.
- [32] PENG B, REN Z, PARTHASARATHY S, et al. HAM: Hybrid Associations Models for Sequential Recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 34(10); 4838–4853.
- [33] LU Y, KOKI T, CHAKRABORTY G, et al. Performance Comparison of Clustering Algorithm Based Collaborative Filtering Recommendation System [C]//Proceedings of the 7th Multidisciplinary in International Social Networks Conference and The 3rd International Conference on Economics, Management and Technology. 2020; 1–6.
- [34] ORTEGA F, LARA – CABRERA R, GONZÁLEZ – PRIETO Á, et al. Providing reliability in recommender systems through Bernoulli Matrix Factorization [J]. Information Sciences, 2021, 553; 110–128.
- [35] ZHOU W, YANG K, ZENG J, et al. FordNet: recommending traditional Chinese medicine formula via deep neural network integrating phenotype and molecule [J]. Pharmacological research, 2021, 173; 105752.
- [36] ZHU Y, LIN Q, LU H, et al. Recommending scientific paper via heterogeneous knowledge embedding based attentive recurrent neural networks [J]. Knowledge – Based Systems, 2021, 215; 106744.