

推荐系统研究进展与应用

黄 勃¹, 严非凡¹, 张 昊², 李佩佩¹, 王晨明¹, 张佳豪¹, 方志军^{1†}

1. 上海工程技术大学 电子电气工程学院, 上海 201620;
2. 江西中烟工业有限责任公司 信息中心 江西南昌 330096

收稿日期: 2021-08-01 †通信联系人 E-mail: zjfang@sues.edu.cn

基金项目: 国家自然科学基金青年基金(61802251), 科技创新2030—“新一代人工智能”重大项目(2020AAA0109300)

作者简介: 黄 勃, 男, 副教授, 现从事人工智能、软件工程研究。E-mail: huangbosues@sues.edu.cn

摘要: 推荐系统一直以来都是很活跃的研究方向。回顾了近5年来推荐系统涉及的数据集、算法、效果评估的研究发展状况, 给出了推荐系统在互联网应用场景中的系统架构, 从工业知识与经验以及工业生产工艺与工艺参数这两方面对推荐系统工业生产中的最新应用做了介绍。分析了推荐系统中“信息茧房”问题产生的原因, 展望了未来推荐系统在“信息茧房”、可解释性、隐私性等方面的研究。

关键词: 推荐系统; 推荐算法; 深度学习; 推荐系统的应用

中图分类号: TP181

文献标志码: A

文章编号: 1671-8836(2021)06-0503-14

Progress and Application of Recommendation System

HUANG Bo¹, YAN Feifan¹, ZHANG Hao², LI Peipei¹, WANG Chenming¹,

ZHANG Jiahao¹, FANG Zhijun^{1†}

1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2. Information Center, Jiangxi Tobacco Industry Co. Ltd., Nanchang 330096, Jiangxi, China

Abstract: Recommendation system has always been a very active research direction. This paper reviews the research and development of datasets, algorithms and effect evaluation involved in the recommendation system in recent 5 years, gives the system architecture of the recommendation system in the Internet application scenario, and introduces the latest application of the recommendation system in industrial production from two aspects: industrial knowledge and experience, industrial production process and process parameters. This paper analyzes the causes of the problem of “information cocoon room” in the recommendation system, and looks forward to the research on “information cocoon room”, interpretability and privacy of the recommendation system in the future.

Key words: recommendation system; recommendation algorithm; deep learning; application of recommendation system

0 引言

互联网时代, 人们对于信息的需求得到了极大满足。但是在海量的信息中找到自己需要的信息却又变得很困难, 产生了“信息过载”问题。推荐系统作为一种筛选信息的工具, 可以更加有效地解决“信息过载”问题, 以个性化的方式提供满足用户需求的内容^[1]。

文献[2]对于推荐系统作出了如下定义: 定义函数 s 来计算一个项目 $i \in I$ (I 为所有项目的集合) 推荐给某一位用户 $u \in U$ (U 为所有用户的集合) 的概率, 推荐算法通过计算概率为用户找到其最感兴趣的项目 $i' \in I$, 形式化的表示如下

$$\forall u \in U, i'_u = \operatorname{argmax} s(u, i), i \in I \quad (1)$$

推荐系统一般是被动触发的, 由系统根据用

引用格式: 黄勃, 严非凡, 张昊, 等. 推荐系统研究进展与应用[J]. 武汉大学学报(理学版), 2021, 67(6): 503-516. DOI: 10.14188/j.1671-8836.2021.1001.

HUANG Bo, YAN Feifan, ZHANG Hao, et al. Progress and Application of Recommendation System [J]. *J Wuhan Univ (Nat Sci Ed)*, 2021, 67(6): 503-516. DOI: 10.14188/j.1671-8836.2021.1001(Ch).

户过去的行为、用户的属性以及上下文等来猜测用户的兴趣,给用户推送物品(包括电商的商品、新闻、电影等),用户可以更便捷地得到个性化信息。它打通了内容和用户之间的壁垒,极大地提高了对信息的利用效率,广泛应用在电子商务、新闻资讯、视频音乐、社交等方面,在工业生产方面也有发展。

推荐系统的搭建一般包括准备数据、选择算法、模型训练、效果评估等步骤。本文对推荐系统的数据集、算法和效果评估等方面研究进行回顾,

总结了推荐系统在互联网应用场景中的系统架构,对推荐系统工业生产中的应用做了介绍。

1 推荐系统的数据集

本节对推荐系统中常用的标准数据集进行了归纳总结,见表1。数据是推荐系统的基础,包括被推荐的物品的信息、用户的行为日志等。离线训练时,一个优质的数据集可以对推荐算法的验证起到

表1 推荐系统常用数据集

Table 1 Common datasets for recommendation system

场景	数据集	简介
电子商务	Amazon	该数据集包含 Amazon 平台的评论和元数据,包括自 1996 年 5 月至 2014 年 7 月的 1 428 亿条评论、产品元数据(种类、价格、品牌等)以及链接数据
	Yelp	此数据集整合了 Businesses、Reviews 和 User data 信息
	DIGINETICA	该数据集包含了从电商搜索引擎日志中抽取的用户会话信息,包含匿名的 User IDs, 经过哈希处理的查询、产品说明和元数据、点击、购买等信息
	YOOCHOOSE	该数据集由 YOOCHOOSE GmbH 构建,支持 RecSys Challenge 2015
广告	Retailrocket	该数据集是从一个真实世界的电子商务网站中收集的。它经过了哈希处理,但未经过任何内容相关的处理
	Criteo	该数据集包含了 Criteo 几天内的流量数据
	iPinYou	该数据集包含了三季的 iPinYou 全球 RTB (Real-Time Bidding) 算法竞赛所使用的训练集和测试集
签到	Foursquare	该数据集包含了从纽约和东京收集的约 10 个月的签到。每次入住都附有时间戳、GPS 坐标及其语义信息等
	MoviesLens	该数据集由 GroupLens Research 收集,是其电影评价网站的评分数据
电影	Netflix	该数据集是 Netflix 大奖赛所使用的官方数据集
	Douban	豆瓣是一个允许互联网用户分享有关电影的评论与观点的网站。该数据集包含了 28 部电影的超过两百万条短评
音乐	LastFM	该数据集包含了收听来自 Last.fm 在线音乐系统的两千位用户的社交网络、标签和音乐艺术家信息
	LFM-1b	该数据集包含超过 12 万名 Last.fm 用户的十亿首音乐收听记录。每条记录都存在艺术家、专辑、曲目名称和时间戳这几个特征
	Yahoo Music	该数据集是 Yahoo 网站音乐社区中用户对各种音乐艺术家偏爱的快照
图书	Book-Crossing	该数据集是 Cai-Nicolas Ziegler 使用爬虫,从 Book-Crossing 社区中收集的四周的数据(2004 年 8 到 2004 年 9 月),包含 278 858 个用户对约 271 379 本书的 1 149 780 个评分
新闻	MIND	该数据集是一个用于新闻推荐研究的大型数据集。它是从微软新闻网站的匿名行为日志中收集的,包含了大约 16 万篇英语新闻文章和超过 1 500 万篇由 100 万用户产生的日志
游戏	Steam	该数据集为 Steam 平台的用户评论和游戏信息,包含 7 793 069 条评论,2 567 538 位用户和 32 135 个游戏。除评论文本外,数据还包括每个评论中用户的游戏时间
动漫	Anime	该数据集包含 myanimelist.net 的用户偏好信息。每个用户都可以将动漫添加到他们已看列表中并给它一个评分,而该数据集则是这些评分的汇总

积极的作用。从表1可见,推荐系统数据集所涉及的领域愈加广泛且形式丰富,助力了不同领域推荐算法的研究。新的数据集也在朝着更加符合真实场景数据的方向发展。

2 推荐算法

对于推荐系统来说,推荐算法是其核心,使用高效且准确的推荐算法是取得良好推荐效果的关键。

一般来说,根据推荐算法是否使用了深度学习技术,将其分为传统推荐算法与基于深度学习的推荐算法。

2.1 传统的推荐算法

根据推荐原理的不同,我们将传统推荐算法分为基于热度的推荐算法、基于社交网络的推荐算法、基于标签的推荐算法、基于人口统计学的推荐算法、基于内容的推荐算法、基于协同过滤的推荐算法、基于模型的推荐算法以及混合推荐算法,如图1所示。

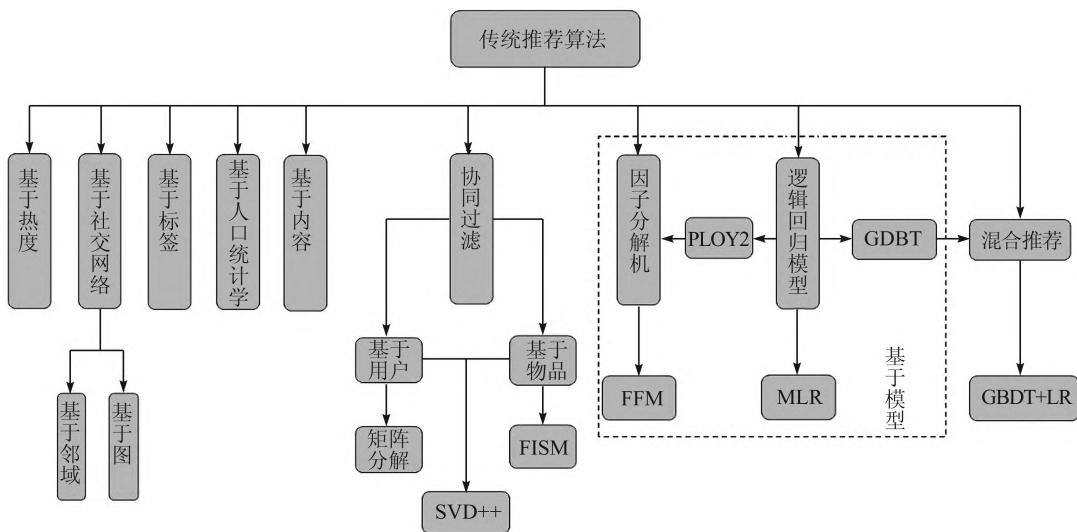


图1 传统推荐算法框架图

Fig. 1 Framework diagram of traditional recommended algorithms

基于热度的推荐是把热点的内容优先推荐给用户,能够覆盖到大部分的内容需求,且启动成本低。基于社交网络的推荐是利用社交网络数据来为用户提供推荐,包括基于邻域的社会化推荐算法^[3]和基于图的社会化推荐算法^[4]。网站从新用户的社交网络中获取好友列表,然后推荐其好友喜欢的物品。标签可以用来表达物品的语义以及用户对物品的兴趣。基于标签的推荐算法便是利用用户的标签数据来给用户提供推荐,分别统计用户常用标签以及被打过某标签次数最多的物品,然后根据用户常用标签找到具有这些标签的热门物品,将热门物品推荐给用户。基于人口统计学的推荐算法利用用户的一些基本信息(年龄、性别以及出生地等)来计算用户间的相关程度,从而向用户进行推荐。基于内容的推荐算法^[5]基本思想是基于物品自身的属性给用户推荐与其曾经喜爱的物品相似的物品。系统会事先对物品标记上相应的属性特征,然后提取用户感兴趣的物品的属性特征作为用户的偏好。当需要产生推荐时,系统会计算待推荐物品的属性特征与用户感兴趣物品的属性特征之

间的相似度,然后取相似度最高的N个物品进行推荐。基于内容的推荐算法流程如图2所示。

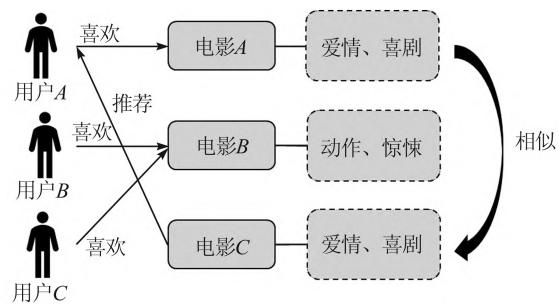


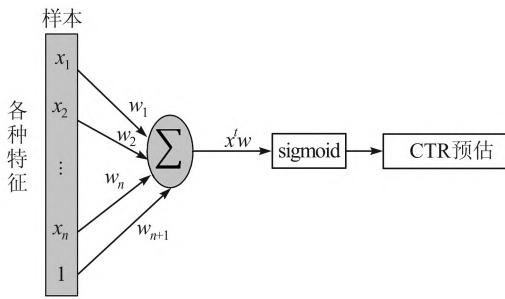
图2 基于内容的推荐

Fig. 2 Content-based recommendations

Goldberg等^[6]提出的协同过滤算法基于这样一个假设:若用户X和Y对t个项目进行相似的评分,或者有相似的行为,则用户就会对其他项目进行类似的评分或行为。它通过收集用户过去的行为以获取用户对产品显式或隐式的信息,获取产品或用户的相关性,然后基于相关性进行推荐。协同过滤本质上是要解决一个矩阵的填充问题,一般用奇异值分解^[7](singular value decomposition, SVD)的方法来对矩阵进行填充。

协同过滤算法可以细分为:基于用户的协同过滤、基于物品的协同过滤。矩阵分解模型(matrix decomposition, MF)^[8]是一种经典的基于用户的协同过滤算法。它将用户和物品都映射到联合潜在因子空间,将用户与物品表示为隐向量,然后用两者的内积来表达用户对物品的兴趣程度。Kabbur 等^[9]提出了进一步表达用户信息的因子项目相似性模型(factored item similarity models, FISM)。此模型本质上是基于物品的协同推荐算法,它将用户历史评分过的物品作为特征属性来得到用户表示,用户表示与物品表示的内积来表达用户对物品的偏好。SVD++模型^[10]则结合了基于用户的推荐算法和基于物品的推荐算法这两者的优势。

为了能利用用户、物品本身的特征以及上下文特征,Richardson 等^[11]提出了逻辑回归(logistic regression, LR)模型,将推荐问题转化成了CTR(click-through-rate)预估问题,如图3所示。LR 模型的迭代模式是固定的,关键在于特征的设计。在 LR 的基础上,Chang^[12]提出了 PLOY2 模型。此模型将所有的特征两两交叉,一定程度上解决了特征组合的问题,但是也导致数据稀疏,一些特征交叉权重缺乏有效数据来训练而难以收敛。针对 PLOY2 的问题,Rendle^[13]提出了因子分解机(factorization machine, FM)模型。为每个特征学习一个隐权重向量,在特征交叉时,使用两个特征隐向量的内积作为交叉特征的权重,降低了特征权重系数的数量。Juan 等^[14]在 FM 模型的基础上引入特征域的概念,提出了域感知因子分解机模型(field-aware factorization machine, FFM),进一步增强了特征交叉的能力,同时复杂度也高出一个数量级。基于对计算广告领域样本的观察,阿里团队提出了混合逻辑回归 MLR(mixed logistic regression)模型^[15]。MLR 模型在 LR 的基础上加入了聚类的思想,先对样本进行分片,再在样本分片中应用 LR 进行 CTR 预估。

图 3 LR 模型^[11]Fig. 3 LR model^[11]

混合推荐算法可以融合多种算法的优势来提高推荐系统的性能。混合推荐算法有加权型、切换型、交叉型、特征组合型、瀑布型、特征递增型、元层次型等^[16]。He 等^[17]提出了梯度下降树(gradient boosting decision tree, GBDT)与 LR 的混合推荐模型,其主要应用场景是预估 CTR 点击率,先使用 GBDT 对一些稠密的特征进行特征选择和组合,再利用 LR 模型进行二次训练来产生最后的预测结果。同样也有人将 GBDT 与 FM、FFM 等模型相结合,也取得了超越单一算法的性能表现。

2.2 传统推荐算法的特点

基于上述内容,对多种传统推荐算法原理以及特点总结如表2所示。传统推荐算法并未过时,考虑到推荐系统的构建成本以及上线难度,FM 模型、GBDT+LR 等仍然是许多中小型公司的良好选择,像爱奇艺这样的大型公司的推荐算法依旧融合了 GBDT、FM 等模型。同时,传统推荐算法与深度推荐算法之间也有很深的联系:矩阵分解模型中的隐向量思想与深度学习中的 embedding 技术思想相似;FM 模型中的特征交叉被深度学习模型进一步发展。GBDT+LR 中体现出的特征工程模型化的思想,成为了深度学习推荐模型的核心思想。

2.3 基于深度学习的推荐算法

相比于传统推荐算法,深度学习不但能捕捉到浅层的特征表示,更能学习到高阶交叉特征,增强模型的表达能力,更好地理解用户的需求,提升系统的泛化性能,解决一些传统模型无法解决的问题。深度学习的网络结构包括深度神经网络(deep neural networks, DNN)、卷积神经网络(convolutional neural networks, CNN)、循环神经网络(recurrent neural networks, RNN)、生成对抗网络(generative adversarial networks, GAN)、图神经网络(graph neural networks, GNN)等。基于深度学习的推荐算法主要围绕在提升表征学习能力、深度协同过滤、充分利用特征三个层面。

2.3.1 基于表征学习

基于表征学习的推荐算法重点在于通过深度神经网络学习到更好的用户与物品的表示,然后通过定义匹配函数来计算两者的匹配分数。匹配函数通常是用用户与物品向量的内积或余弦距离来表示。学习到的用户与物品的表示在排序之外的其他环节也可以发挥作用。在欧氏空间数据的表征上,基于深度自编码器的方法效果很好,而处理非欧氏空间的数据时,通常使用图神经网络。

AutoRec 模型^[18]借鉴了自编码器的思路,对输入做重建,来建立用户和项目的表示,如图4所示。

表2 传统推荐算法总结

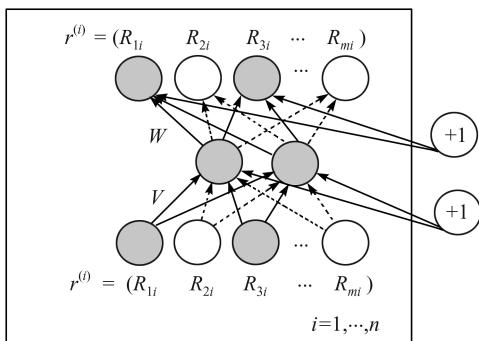
Table 2 Summary of Traditional Recommended Algorithm

传统推荐算法	原理	特点
基于热度推荐	把热点的内容优先推荐给用户	启动成本低;无法实现个性化推荐
基于社交网络的推荐	利用社交网络数据(好友等)提供推荐	可解释性强,能够缓解冷启动问题;存在隐私问题、系统响应时间长
基于标签的推荐	利用用户的标签数据来提供推荐	增加了推荐的多样性和可解释性;对新用户和长尾物品很难有推荐结果
基于人口统计学的推荐	分析用户的基本信息,考虑用户之间的相关程度,从而向用户进行推荐	不需要历史数据,没有冷启动问题;高度依赖数据的准确性且算法粗糙
基于内容的推荐	基于物品自身的属性,给用户推荐与其曾经喜爱物品相似的物品	建模方便,可满足小众爱好的用户,解释性良好;十分依赖人工特征提取,推荐的新颖度很低
协同过滤	CF 根据用户的历史行为构造用户-物品共现矩阵,基于相似用户或相似物品进行排序	简单、易用;长尾物品很难被推荐,泛化能力差
	MF 将CF中的共现矩阵分解为用户矩阵和物品矩阵,利用用户隐向量和物品隐向量的内积进行排序推荐	提升了对稀疏矩阵的处理能力,加强了泛化能力;仅利用用户和物品的交互信息
	FISM 将用户喜欢过的项目作为用户的表达来刻画用户	增强了用户信息的表达
	SVD++ 综合利用用户特征向量、物品特征向量以及用户对有过评分物品的隐式反馈来进行推荐	结合了基于物品的推荐与基于用户的推荐两者的优势;仅利用用户和物品的交互信息
基于模型的推荐	LR 将用户、物品、上下文信息输入逻辑回归模型后得到CTR,根据CTR进行推荐	综合利用各种特征,简单易用;无法进行特征交叉、特征筛选,表达能力较差
	PLOY2 在LR的基础上将所有的特征两两相交叉,得到CTR后,根据CTR进行推荐	一定程度上解决了特征组合的问题;令数据更加稀疏,难以收敛且特征权重增加,训练成本大
	FM 在LR的基础上加入二阶特征交叉,为每一个特征训练一个隐向量,通过隐向量的内积运算得到交叉特征的权重	具备二阶特征交叉能力,表达能力更强;受限于组合爆炸,不易做到高阶的特征组合
	FFM 在FM的基础上加入特征域的概念,每个特征在不同特征域交叉时采用不同的影响量	进一步加强了特征交叉的能力;模型训练开销大
协同去噪自编码器(auto-encoder)	GBDT+LR 利用GBDT自动进行多阶特征组合,将组合后得到离散特征向量输入LR,得到CTR,根据CTR进行排序推荐	具备更高阶的特征组合能力,特征工程模型化;GBDT不能并行训练,训练时间长
	MLR 先对样本分片,在每个分片上构建逻辑回归模型,最好对每个分片的概率和逻辑回归的得分进行加权平均,根据加权平均数排序推荐	具备较强的表达能力;相对于深度学习模型,模型结构还较为简单

通过自编码器将用户向量或者项目向量投射到低维向量空间,然后再将其投射到正常空间,利用自编码器中目标值和输入值相近的特性,重建出用户对于未交互过的项目的打分。与AutoRec模型相比协同去噪自编码器(collaborative denoising auto-en-

coders,CDAE)模型^[19],其用户侧输入同时使用了用户历史交互过的项目以及用户本身的偏好,思想与SVD++模型类似。

SR-GNN模型^[20]将用户的交互历史划分为一个个会话,为每个会话序列构建有向图,然后在会

图 4 AutoRec 模型示意图^[18]Fig. 4 Schematic diagram of AutoRec model^[18]

话图中使用图神经网络学习物品和会话的嵌入表示。这种方法可以较好地捕获用户动态兴趣，缺点是完全没有考虑用户信息。为了应对多层图卷积会使不同兴趣的用户具有相似的嵌入性这一问题，Liu 等^[21]提出了一种新的兴趣感知消息传递图卷积网络模型，该模型将具有相似兴趣的用户及其交互项分组到不同的子图中，在子图中进行高阶图卷积从而同时捕捉到相似兴趣的特征及图结构，减少了高阶噪声信号的影响，进而学习到更好的用户表示。但此方法并未完全消除噪声，用户与项目之间信息传播的噪声在子图中依旧存在。

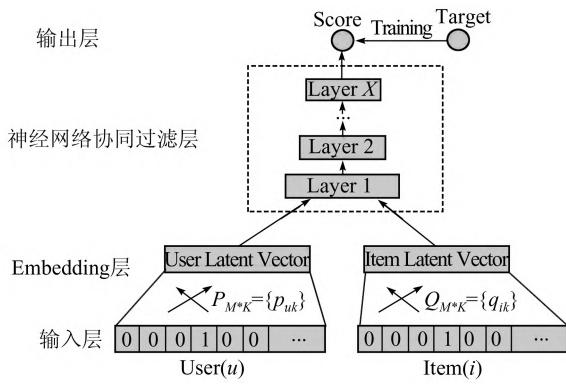
2.3.2 基于深度协同过滤

在深度学习的加持下，传统推荐算法中的协同过滤类算法又有了新的发展。深度矩阵分解模型^[22](deep matrix factorization models, DMF)在 MF 模型中增加了非线性的 MLP 网络来增强模型的表达能力。与 MF 模型相同，DMF 仅利用了用户与物品的交互信息，因此考虑通过引入辅助信息来提升推荐效果。深度协同过滤(deep collaborative filtering, DCF)模型^[23]的输入除了用户和物品的交互矩阵，还有用户特征和物品特征。用户和物品侧的特征各自通过一个自编码器来学习，用户与物品的交互矩阵依然做矩阵分解。

He 等^[24]提出的基于神经网络的协同过滤(neural collaborative filtering, NCF)框架，引入了 MLP(multilayer perceptron)去拟合用户和物品的非线性关系，提升了网络的拟合能力，如图 5 所示。用户和物品侧的双塔设计可以加入任意附加信息的特征，而 MLP 网络也可以进一步设计。因此，NCF 模型十分灵活。例如，NMF 模型便是在 NCF 框架的基础上同时利用了 MF 和神经网络 MLP 的能力来拟合匹配得分。ONCF 模型^[25]是在 NCF 框架上引入了外积的概念。CFGAN^[26]模型将协同过滤思想与对抗生成网络结合，用向量对的方式进行训练，以

用户的历史交互序列向量作为模型输入，避免了判别器的困惑，也可以保证平滑的传回梯度。

以上协同过滤方法的本质在于通过算法使用户与项目在映射空间的向量尽量接近，而基于翻译的框架则是另外一种思路，它重点在于找到一个关系向量，使得用户向量加上关系向量尽可能接近项目向量。He 等^[27]提出了 TransRec 模型，此模型是用来解决下一个项目的推荐问题，其基本思想是用户本身的向量，加上用户上一个交互的物品的向量，应该接近于用户下一个交互的项目的向量。LRML 模型^[28]也是基于翻译框架，它通过引入记忆网络来学习度量距离，从而进一步提升效果。

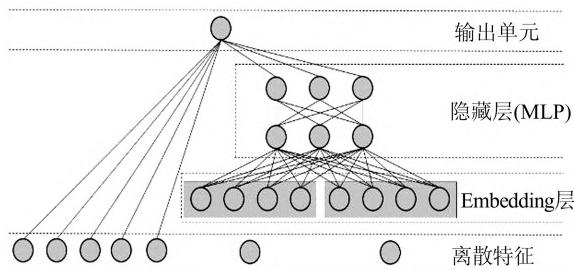
图 5 NCF 模型框架^[24]Fig. 5 NCF model framework^[24]

2.3.3 基于特征

基于深度学习的推荐算法大部分是在研究如何更好地利用特征向量，包括高阶特征交互、注意力机制以及融合时间序列等。Deep Crossing 模型^[29]引入了残差网络的概念，它是一个真正的把深度学习应用于推荐系统中的模型，完整地解决了特征工程、稀疏向量稠密化、多层次神经网络进行优化目标拟合等问题。

Google 提出的 Wide&Deep 模型^[30]框架如图 6 所示，Wide 部分是一个广义的线性模型，可以用较少的参数学习样本中的低阶特征，增强模型记忆性。Deep 部分是一个前馈神经网络，离散型特征会被嵌入到低维稠密向量，这些稠密向量会被送入多层次感知机(MLP)中训练。Deep 部分通过学习低维特征向量来探索用户与项目之间的潜在关系，增强泛化性能。Wide&Deep 模型平衡了模型历史记忆与推理泛化能力，缺点在于仍然要手动设计交叉特征。此模型框架对基于特征的深度学习推荐算法的发展产生了重大影响，许多推荐算法都是在此框架上进行改进的。

深度特征交叉网络(deep&cross network, DCN)^[31]无需进行额外的特征工程工作，其将 Wide&Deep 模型中的 Wide 侧替换为交叉网络，该交叉网络通过调整结

图 6 Wide&Deep 模型^[30]Fig. 6 Wide&Deep model^[30]

构层数能够构造出有限阶交叉特征,对特征进行显式交叉编码,既精简了模型参数,又提高了模型的表征能力。深度因子分解机(deep factorization machine, DeepFM)模型^[32]、xDeepFM模型^[33]与DCN模型一样,在Wide部分做了改进。DeepFM将Wide部分由LR模型替换了FM模型来避免人工特征工程,xDeepFM模型则是将Wide部分的LR模型替换了特征交叉(compressed interaction network,CIN)模块,旨在捕捉高阶特征交互。神经因子分解机(neural factorization machine,NFM)模型^[34]和基于注意力机制的因子分解机(attentional factorization machines,AFM)^[35]模型则是对Deep部分进行了改进。NFM模型在embedding层后进行Bi-interaction操作,将各向量两两相乘后直接相加输出一个低维向量,压缩了MLP输入的规模,加速了网络的训练。AFM模型引入了注意力机制,作者认为不同的特征交叉项对结果的贡献是不同的,因而对特征项进行了加权累加的操作。

阿里提出的深度兴趣网络(deep interest network,DIN)同样运用了注意力机制,DIN模型在embedding层与MLP之间加入了注意力网络,有效地利用了用户兴趣多样性以及当前候选商品仅与用户一部分兴趣有关这一特点。深层兴趣演化网络(deep interest evolution network,DIEN)^[37]在DIN模型的基础上融合了时间序列,考虑到了用户兴趣的发展变化,加入AUGRU序列模型对兴趣演化过程建模。Feng等^[38]发现用户的行为在每一个会话和异构交叉会话中是高度同质的,因此提出了深度会话兴趣网络(deep session interest network,DSIN),DSIN利用用户多个历史会话来模拟CTR预测任务中的用户系列行为,此模型也融入了时间序列。Smirnova等^[39]提出了基于RNN的上下文感知推荐系统,它在序列模型中融入了上下文信息,提高了推荐效果。

2.4 基于深度学习的推荐算法的特点

深度学习技术赋予了推荐算法强大的非线性表示能力,使得基于深度学习的推荐算法取得了更好的推荐效果。基于表征的推荐算法,以用户与物

品的嵌入向量作为媒介来实现推荐。基于深度协同过滤的推荐算法,本质上也是想获得更好的用户与物品表示,可以通过融入辅助信息来进一步提升推荐效果。基于特征的推荐算法是近年来的主流方向,各特征之间的深层交互能让模型学习到更准确的用户偏好。不同深度神经网络在推荐领域的应用中也有着各自的特点,如CNN可以高效地从非结构化的数据中提取特征,RNN可以捕捉用户的动态兴趣等。但深度神经网络也使得产生的推荐结果具有不可解释性。

3 推荐系统评估

3.1 在线评估

在线评估指设计一个在线实验,然后根据用户的在线反馈结果来衡量推荐系统的性能。ABTest是一种主流的在线实验方式^[40]。ABTest通过制定两个方案来实现相同的目标,让一部分用户使用A方案,另一部分用户使用B方案,记录下两部分用户的反馈情况,然后根据相应的评估指标,从而确定哪种方案更好。常见的在线评估指标包括点击率、转化率、网站成交额等。点击率表示的是商品的点击次数与曝光次数的比值。转化率指事物从状态A进入到状态B的概率。在电商推荐系统中,转化率通常是指发生目标行为(如购买)的商品与发生目标行为条件(如点击查看)的比值,转化率的提高是目前电商关注的重点。网站成交额是销售额、取消订单额、拒收订单额、退货订单额的总和,用来研究用户的购买意向。

3.2 离线评估

离线评估是指根据待评估推荐系统在实验数据集上的表现,基于一些离线评估指标来衡量推荐系统的效果。离线评估最主要的两个环节分别是拆分数据集和选择离线评估指标。数据集通常被拆分为训练数据集、验证数据集和测试数据集。为了划分这几种数据集,传统的方法有留出法、K-折交叉验证法、自助法等。区别于以上传统方法,离线Replay方法在离线状态下对线上更新过程进行仿真,评估过程更接近真实线上环境,因此评估结果更加可靠。常见的离线评估指标可以分为两大类:准确度指标和非准确度指标,如表3所示。一般会结合具体业务场景,选择2~4个离线评价指标。例如涉及排序评估时,近年来的文章中通常使用归一化折损累计收益(NDCG)、平均准确率(MAP)等指标。

表 3 离线评估指标
Table 3 Offline Evaluation Indicators

评估指标类别	具体评估指标
准确度指标	预测分类 AUC, 准确度(Accuracy), 精确度(Precision), 召回率(Recall), F-measure
	预测评分 平均绝对误差(MAE), 均方误差(MSE), 均方根误差(RMSE)
	相关性 皮尔逊积距相关系数, 斯皮尔曼等级相关系数, 肯德尔等级相关系数
	排序 平均排序分(ARS), 归一化折损累计收益(NDCG), 平均准确率(MAP), MRR
非准确度指标	多样性, 新颖性, 惊喜度, 覆盖率, 信任度, 实时性, 鲁棒性

3.3 用户调研

推荐系统的离线实验指标和实际商业指标之间存在差异,若是将不太有把握的推荐算法直接上线测试往往有较大的风险。因此在推荐算法上线前,一般会做一次用户调研。在进行用户调研的过程中,需要保证测试用户的分布与真实用户的分布尽量相同。

4 推荐系统的应用

4.1 推荐系统在互联网中的应用

推荐系统在互联网中应用得最为广泛,主要有以下应用场景:广告推荐、电商推荐、新闻资讯推荐、视频音乐推荐等,如表4所示。这些场景的推荐

系统对于内容处理有着很高的要求,通常会把要用于推荐的内容表示成神经网络可以识别的向量和标签,然后输入到相关的模型中使用。京东推荐是推荐系统在互联网中应用的典型。京东推荐系统起步于2012年,随着大数据时代的到来,京东推荐团队于2013年设计了新的推荐系统。接下来因为业务的快速增长以及移动互联网的到来,推荐类型从传统的商品推荐逐步扩展到活动、优惠券、入口图等其他类型的推荐。由于个性化推荐业务需求强烈,京东基于大数据和个性化算法,在2015年底再次升级了推荐系统,并于2016年“6·18”期间大放异彩,其开创的“智能卖场”实现了活动会场的个性化分发,不仅提升了网站成交额,也降低了人工成本,大大提高了流量效率和用户体验。

表 4 推荐系统在各个领域中的应用
Table 4 Recommends applications of the system in various fields

类别	国内	国外	典型算法
视频类	爱奇艺, 腾讯视频, 抖音	Netflix, Hulu, Youtube	DNN+GBDT+FM
资讯类	今日头条, 微博新闻	Google news, Digg	DKN, NMF
音乐类	QQ 音乐, 网易云音乐	Spotify, Last. fm	BST, DIEN
社交类	QQ, 新浪微博	Facebook, Twitter	GraphRec, DiffNet
电子商务类	淘宝, 京东, 拼多多	Amazon, eBay	DIEN, ESMM
服务类	美团, 饿了么, 携程	Agoda, Airbnb	SR-GNN, word2vec

具体应用场景中的推荐系统架构如图7所示,一般包括特征提取模块、存储单元模块、推荐算法模块以及接入调度模块。用户与内容模块主要完成日志处理、数据统计、CTR更新以及画像更新等操作;存储单元模块根据数据的类型与用途分别将其存在不同的存储单元中;推荐算法模块包含召回和排序两个组件,基于一定的逻辑与数据运算过程生成最终的推荐结果;接入调度模块执行用户请求的发送与推荐结果的回传等操作。

推荐算法仍然是核心,包括四个环节:召回、粗排、精排和重排。召回环节:通常采集到的候选物

品是海量的,若直接对其做精排序,模型复杂度太大,不利于模型上线。因此,为了降低候选物品的数量级,召回阶段会根据用户的兴趣和历史行为,从海量的物品库中快速筛选出用户潜在感兴趣小部分物品,送入排序环节,在实际的推荐系统中,会有多路召回。粗排环节:排序环节可以融入较多特征,使用复杂模型来精准地做个性化推荐。在某些场景中,召回阶段返回的物品数量级仍然很大,因此可以通过少量人工特征来对召回的结果进行粗略地排序,在保证一定精准的前提下,进一步对物品进行筛选。该阶段通常会对特征和模型复杂度

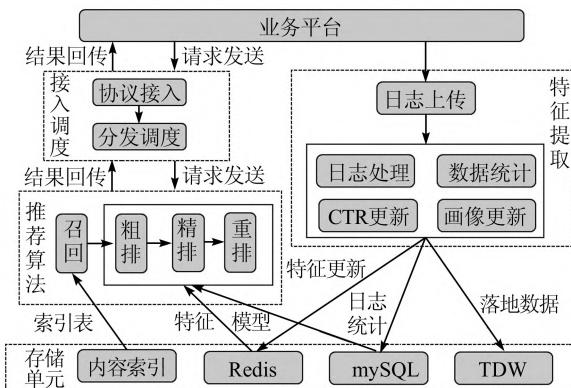


图7 互联网推荐系统应用架构

Fig. 7 Internet recommendation system application architecture

进行精简。精排环节：经过召回与粗排之后，候选物品的数量级已降至可接受范围。为了考虑用户体验、保证排序精度，该阶段可以使用丰富的人工特征以及各种复杂模型，在保证能上线的前提下，尽量精准地对物品进行个性化排序。重排环节：该环节往往是结合具体的业务场景，对精排的结果进行后续处理，如去掉用户已浏览的物品、去重保证多样性、打散相似的物品、固定类型物品插入等，也可进行一些探索，如使用汤普森采样、bandit 算法等。各环节常用的模型与策略如表5所示。

通常认为由推荐系统产生的最优结果理论上是真值，然而在实际的应用中，推荐系统中存在许多偏差，如图8所示。从用户的角度出发，其产生的数据是有偏差的。用户更倾向于对位置靠前的物品进行交互，所产生的交互物品可能并不是用户所偏好的即位置偏差；用户倾向于给自己很喜欢或者很不喜欢的物品进行打分，因此观测到的评分不是所有评分的代表性样本即选择偏差；用户的打分数据倾向于和群体观点保持一致，使得评分无法反应用户的真实偏好即一致性偏差；带标签的数据都是曝光过的，未曝光的数据无法确定其标签即曝光偏

差。各种模型的假设也会产生偏差，如 RNN 的时间依赖假设、注意力机制假设等。模型产生的推荐结果往往存在流行度偏差，热门物品会获得比预期更高的热度，而长尾物品曝光不足。若是数据不均衡，推荐的结果还会有歧视性偏差，某些群体的推荐结果有偏。以上的种种偏差在推荐系统的反馈循环中会不断被加剧，使推荐生态逐步恶化。

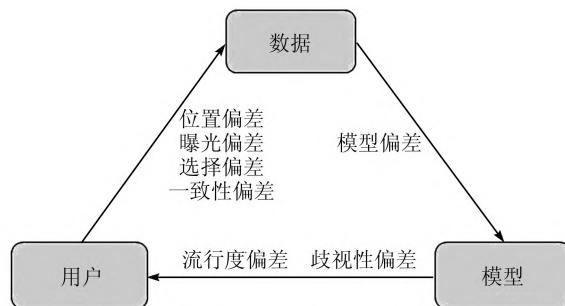


图8 推荐生态中的偏差

Fig. 8 Bias in recommended ecology

Guo 等^[41]、Zhao 等^[42]采用单独建模位置偏差，预测时用排序模型的思路来解决位置偏差问题。对于曝光偏差，常采用基于启发式的方法^[43]和基于采样的方法^[44]。正则化^[45]、对抗学习^[46]以及因果推断方法^[47]等是解决流行度偏差以及歧视性偏差的有效手段。基于倾向得分的模型^[48~50]是一种通用的方法，可以解决多种偏差，但是在复杂应用场景中的实际效果还需要进一步验证。表6对常见偏差的解决方法做了总结。如何更好地平衡偏差和推荐系统短期的表现是当前的一大挑战，若将所有的偏差消除，推荐系统短期业务指标大概率会下降。

4.2 推荐系统在工业生产中的应用

我国正大力发展高端制造业，并且倡导各领域的企业进行信息化、数字化升级。在这一趋势下，利用推荐系统来帮助企业与个人获取合适的信息、提高效率显得很有必要，尤其是在一些复杂度高、信息量大的工业领域。工业生产中的推荐系统与

表5 各环节常用模型与策略

Table 5 Common models and strategies for each link

具体环节	常用模型与策略
召回	CF、MF、LR、FM、双塔召回模型(Deep Structured Semantic Models, DSSM)、Youtube DNN、用户多兴趣网络、RNN 序列召回、TDM 深度树匹配召回、热度召回、基于地理位置召回、基于标签召回、历史行为重定向
粗排	简化的精排模型
精排	GBDT+LR、FM、FFM、FNN、PNN、Wide&Deep、Deep&Cross、DeepFM、DCF、xDeepFM、NCF 框架、DIN、DIEN、DSIN、BERT、RNN、多任务学习、多目标学习
重排	最大边界相关法(maximal marginal relevance, MMR)、基于行列式点过程(determinantal point process, DPP)、汤普森采样、bandit 算法、相关业务需求

电子商务等方面的推荐系统有一些基本的区别。电子商务等方面的推荐系统大多关注用户的兴趣、偏好以及历史行为。然而在工业生产中,这些信息不足以探索相关的知识和信息。Mehrpoor 等^[51]认为用户的项目、任务、角色、工作地点、工作时间以及相关规定等更多关键本体的组合对于工业生产中的推荐系统更加有效,并提出工业知识表示的语

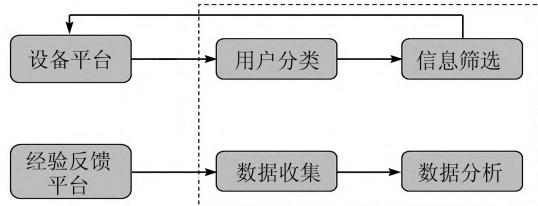
义推荐系统有以下必不可少的组件:用户的位置浏览器、推理和推断用户情况的上下文推理机、将用户情况转换为数据库可读查询的上下文解释器、知识和信息检索器、本体库以及储存用户信息的数据库。目前有关推荐系统在工业生产领域中应用的文章数量相对较少,本节将从工业知识与经验以及生产工艺与工艺参数这两方面来进行介绍。

表 6 常见偏差的解决方法

Table 6 Solutions to common bias

偏差类型	解决方法
位置偏差	将位置特征加入训练数据,预估时将该特征统一设置为某一定值;计算 CTR 时考虑位置效应;单独建模位置误差,但预测时候还是只用排序模型;利用反倾向得分来纠正
曝光偏差	倾向得分;对比学习;通过启发式方法补充未观测的交互反馈,将未观测的交互看作负样本,并赋予较低权重或根据用户和物品的活跃程度设置;通过采样来控制指定数据用于更新模型,从而纠正数据分布
选择偏差	对确实的数据进行数据补全;用相反的倾向分数加权;元学习
一致性偏差	将用户评分作为用户喜好和社会影响的综合结果
流行度偏差	设计正则项;利用迁移学习,将从热门物品中学到的知识迁移到长尾物品上;利用对抗学习以及因果推断的方法;结合业务场景,在重排阶段设计一些策略进行调整
歧视性偏差	设计正则项;在因果图上应用反事实干预;在粗排和精排阶段,对不同人群的训练数据重打标签或是重采样;在重排阶段根据业务上的公平性要求进行替换和排序

工业知识与经验的推荐:Benkaddour 等^[52]提出了一个基于协同过滤的推荐系统。此推荐系统可以向工业操作人员推荐相关的故障诊断文档,大大提高工业操作人员的效率。赵栋等^[53]提出了适用于设备设计的工程经验反馈系统,该系统在设备设计过程中能够主动为设计人员提供有价值的经验反馈,有效地提高了设备设计的效率。该推荐系统流程图如 9 所示。首先从经验反馈平台获取数据,按照预定义规则在离线状态下进行分类建模,然后通过分析设计人员信息以及设计相关的项目、机组和内容,实现在设备设计过程中的经验反馈个性化推送。也有越来越多的研究在利用知识图谱来实现工业知识与经验的推荐,完成设备故障诊断推荐、智能会话推荐等推荐任务。

图 9 工程经验反馈推荐系统流程图^[53]Fig. 9 flow chart of engineering experience feedback recommendation system^[53]

生产工艺与工艺参数的推荐:彭泽普^[54]提出了一个基于统计特征识别算法的冒口工艺推荐系统。该系统首先将铸件按照形状大类进行整理,然后基于特征提取形成特征数据库。在特征数据库的基础上,基于统计特征识别算法完成相似模型及其工艺的检索与推荐,系统框架如图 10 所示。牛中伟^[55]引入本体技术,提出了一个基于本体的难加工材料高性能加工工艺参数推荐系统。Kwon 等^[56]为了减少逻辑合成和物理设计(logic synthesis and physical design, LSPD)工具流参数的调整,提高整体设计效果,提出了一个 LSPD 参数推荐系统。

推荐系统在企业运营与决策的过程中也可以

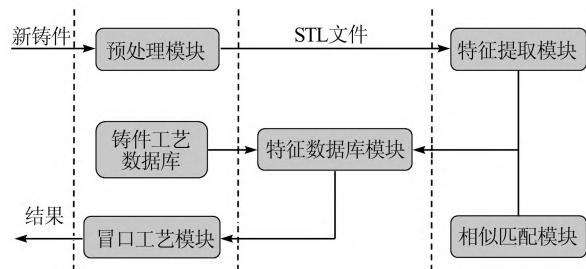


图 10 冒口工艺推荐系统框架示意图

Fig. 10 Framework diagram of riser process recommendation system

发挥作用。例如,孙妍^[57]提出了基于信誉评估的供应商推荐系统,以智能化的筛选和排序为制造企业提供决策支持,高效率地为企业选择出最合适的合作伙伴。随着产业升级,推荐系统在工业生产活动中必将有更加广泛、更深层次的应用。

推荐系统在工业生产中应用所面临的挑战主要在于数据端。各种设备数据分散在不同数据库、电子文件、线下文档中,且有些数据有保密要求,因而整合数据较为困难。工业生产领域中数据的专业程度较高,需要结合专家意见进行数据处理,十分不便。最后,工业数据的分布很不平衡,以设备数据为例,设备数据中大部分是设备正常数据,设备异常数据很少。

5 展 望

推荐系统最初应用的目的是快速匹配用户最需要的内容,提升检索效率,但如今却逐渐成为增强用户黏性,延长使用时间的机制保障,由此导致的“信息茧房”问题也愈发严重。为了保持较好的商业指标,推荐系统会不断重复推荐用户感兴趣的内容来取悦用户,减少甚至不推荐其他方面的内容。这使得用户以自身兴趣为砖瓦构筑了“信息茧房”,收窄了用户的信息接收范围,并减少了接受差异化信息的可能性。在推荐算法的影响下,用户会在高度同质化的内容中逐渐极端化。因此,在整个推荐生态中,应强化用户的主导地位,让推荐算法服务用户而非控制用户。

可以预见,推荐系统在未来会是大多数互联网企业的基础业务,具有广阔的前景。为了获得更好的推荐效果与用户体验,推荐系统中的“信息茧房”问题、可解释性问题、隐私性问题等亟待解决。针对“信息茧房”问题,基于深度强化学习的推荐(deep reinforcement learning,DRL)^[58,59]可以作为一种有效的解决方法,它将推荐过程当作是动态的,从而更容易地捕捉用户的长期兴趣并及时响应系统中新的变化。基于深度强化学习的推荐算法在当前面临着通用性等挑战,有待在将来解决。推荐结果具有可解释性能够提升用户对推荐系统的信任度与满意度。目前有较多学者利用知识图谱作为辅助信息来增强推荐结果的可解释性与准确性,基于知识图谱的推荐算法有望在将来取得较大的发展。大数据时代,人们对于隐私问题更为关注,如今国家有关部门也加强了相关方面的监管。因此,推荐系统中有关隐私性问题的研究显得愈加重要。当前,结合联邦学习的推荐系统^[60]

在此方面取得了一些创新性的成果,但仍有许多空白需要填补。

参 考 文 献:

- [1] 周万珍,曹迪,许云峰,等. 推荐系统研究综述[J]. 河北科技大学学报, 2020, 41(1): 76-87. DOI: 10.7535/hbkd.2020yx01009.
- [2] ZHOU W Z, CAO D, XU Y F, et al. A survey of recommendation systems [J]. *Journal of Hebei University of Science and Technology*, 2020, 41(1): 76-87. DOI: 10.7535/hbkd.2020yx01009 (Ch).
- [3] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6): 734-749. DOI: 10.1109/TKDE.2005.99.
- [4] 李善涛,肖波. 基于社交网络的信息推荐系统[J]. 软件, 2013, 34(12): 41-45. DOI: 10.3969/j.issn.1003-6970.2013.12.010.
- [5] LI S T, XIAO B. Recommendation system based on social network [J]. *Computer Engineering & Software*, 2013, 34(12): 41-45. DOI: 10.3969/j.issn.1003-6970.2013.12.010 (Ch).
- [6] 蔡永嘉,李冠宇,关皓元. 信任社交网络中基于图熵的个性化推荐算法[J]. 计算机应用, 2019, 39(1): 176-180. DOI: 10.11772/j.issn.1001-9081.2018061202.
- [7] CAI Y J, LI G Y, GUAN H Y. Personalized recommendation algorithm based on graph entropy in trust social network [J]. *Journal of Computer Applications*, 2019, 39(1): 176-180. DOI: 10.11772/j.issn.1001-9081.2018061202 (Ch).
- [8] 刘明昌. 基于内容的推荐技术研究[J]. 现代营销, 2016(6): 243. DOI: 10.3969/j.issn.1009-2994.2016.06.186.
- [9] LIU M C. Research on content-based recommendation technology [J]. *Modern Marketing*, 2016(6): 243. DOI: 10.3969/j.issn.1009-2994.2016.06.186(Ch).
- [10] GOLDBERG D, NICHOLS D, OKI B M, et al. Using collaborative filtering to weave an information tapestry [J]. *Communications of the ACM*, 1992, 35(12): 61-70. DOI: 10.1145/138859.138867.
- [11] ZHOU X, HE J, HUANG G Y, et al. SVD-based incremental approaches for recommender systems [J]. *Journal of Computer and System Sciences*, 2015, 81(4): 717-733. DOI: 10.1016/j.jcss.2014.11.016.
- [12] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems [J]. *Computer*, 2009, 42(8): 30-37. DOI: 10.1109/MC.2009.263.
- [13] KABBUR S, NING X, KARYPIS G. FISM: Factored item similarity models for top- N recommender systems [C]//

- Proceedings of the 19th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining.* New York: ACM, 2013: 659–667. DOI: 10.1145/2487575.2487589.
- [10] KOREN Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model [C]// *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* New York: ACM Press, 2008: 426–434. DOI: 10.1145/1401890.1401944.
- [11] RICHARDSON M, DOMINOWSKA E, RAGNO R. Predicting clicks: Estimating the click-through rate for new ads[C]// *Proceedings of the 16th International Conference on World Wide Web.* New York: ACM Press, 2007: 521–530. DOI: 10.1145/1242572.1242643.
- [12] CHANG Y W. Training and testing low-degree polynomial data mappings via linear SVM [J]. *Journal of Machine Research*, 2010, **11**(4): 1471–1490.
- [13] RENDLE S. Factorization machines [C]// *The 10th IEEE International Conference on Data Mining.* New York: IEEE Press, 2010: 995–1000. DOI: 10.1109/ICDM.2010.127.
- [14] JUAN Y, ZHUANG Y, CHIN W S, et al. Field-aware factorization machines for CTR prediction[C]// *Proceedings of the 10th ACM Conference on Recommender Systems.* New York: ACM, 2016: 43–50. DOI: 10.1145/2959100.2959134.
- [15] GAI K, ZHU X, LI H, et al. Learning piece-wise linear models from large scale data for ad click prediction[EB/OL]. [2017-04-19]. <https://arxiv.org/pdf/1704.05194v1.pdf>.
- [16] 周春华, 沈建京, 李艳, 等. 经典推荐算法研究综述[J]. 计算机科学与应用, 2019, **9**(9): 1803–1813. DOI: 10.12677/CSA.2019.99202.
- ZHOU C H, SHEN J J, LI Y, et al. Review of classical recommendation algorithms [J]. *Computer Science and Applications*, 2019, **9**(9) : 1803–1813. DOI: 10.12677/CSA.2019.99202(Ch).
- [17] HE X R, PAN J F, JIN O, et al. Practical lessons from predicting clicks on ads at Facebook [C]// *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* New York: ACM Press, 2014: 1–9. DOI: 10.1145/2648584.2648589.
- [18] SEDHAIN S, MENON A K, SANNER S, et al. AutoRec: Autoencoders meet collaborative filtering [C]// *Proceedings of the 24th International Conference on World Wide Web.* New York: ACM, 2015: 111–112. DOI: 10.1145/2740908.2742726.
- [19] WU Y, DUBOIS C, ZHENG A X, et al. Collaborative denoising auto-encoders for Top-N recommender systems [C]// *Proceedings of the 19th ACM International Conference on Web Search and Data Mining.* New York: ACM, 2016: 153–162. DOI: 10.1145/2835776.2835837.
- [20] WU S, TANG Y Y, ZHU Y Q, et al. Session-based recommendation with graph neural networks[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, **33**: 346–353. DOI: 10.1609/aaai.v33i01.3301346.
- [21] LIU F, CHENG Z Y, ZHU L, et al. Interest-aware message-passing GCN for recommendation [C]// *Proceedings of the Web Conference 2021.* New York: ACM, 2021: 1296–1305. DOI: 10.1145/3442381.3449986.
- [22] XUE H J, DAI X, ZHANG J, et al. Deep matrix factorization models for recommender systems[C]// *26th International Joint Conference on Artificial Intelligence.* Menlo Park: AAAI Press, 2017: 3203–3209. DOI: 10.24963/ijcai.2017/447.
- [23] LI S, KAWALE J, FU Y. Deep collaborative filtering via marginalized denoising auto-encoder[C]// *Proceedings of the 24th ACM International Conference on Information and Knowledge Management.* New York: ACM, 2015: 811–820. DOI: 10.1145/2806416.2806527.
- [24] HE X N, LIAO L Z, ZHANG H W, et al. Neural collaborative filtering[C]// *Proceedings of the 26th International Conference on World Wide Web.* Switzerland: International World Wide Web Conferences Steering Committee, 2017: 173–182. DOI: 10.1145/3038912.3052569.
- [25] HE X, DU X, XIANG W, et al. Outer product-based neural collaborative filtering [C]// *27th International Joint Conference on Artificial Intelligence.* Menlo Park: AAAI Press, 2018: 2227–2233. DOI: 10.24963/ijcai.2018/308.
- [26] CHAE D K, KANG J S, KIM S W, et al. CFGAN: A generic collaborative filtering framework based on generative adversarial networks[C]// *Proceedings of the 27th ACM International Conference on Information and Knowledge Management.* New York: ACM, 2018: 137–146. DOI: 10.1145/3269206.3271743.
- [27] HE R N, KANG W C, MCAULEY J. Translation-based recommendation [C]// *Proceedings of the 11th ACM Conference on Recommender Systems.* New York: ACM, 2017: 161–169. DOI: 10.1145/3109859.3109882.
- [28] TAY Y, TUAN A L, HUI S C. Latent relational metric learning via memory-based attention for collaborative ranking[C]// *Proceedings of the 2018 World Wide Web Conference on World Wide Web.* New York: ACM, 2017: 729–739. DOI: 10.1145/3178876.3186154.
- [29] SHAN Y, HOENS T R, JIAO J, et al. Deep crossing: Web-scale modeling without manually crafted combinatorial features [C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery*

- and Data Mining. New York: ACM Press, 2016: 255–262. DOI: 10.1145/2939672.2939704.
- [30] CHENG H T, KOC L, HARMSEN J, et al. Wide & Deep learning for recommender systems [C]// Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, New York: ACM, 2016: 7–10. DOI: 10.1145/2988450.2988454.
- [31] WANG R X, FU B, FU G, et al. Deep & cross network for ad click predictions [C]// Proceedings of the ADKDD'17. New York: ACM, 2017: 1–7. DOI: 10.1145/3124749.3124754.
- [32] GUO H F, TANG R M, YE Y M, et al. DeepFM: A factorization-machine based neural network for CTR prediction [C]// 26th International Joint Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2017: 1725–1731. DOI: 10.24963/ijcai.2017/239.
- [33] LIAN J X, ZHOU X H, ZHANG F Z, et al. xDeepFM: combining explicit and implicit feature interactions for recommender systems [C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1754–1763.
- [34] HE X N, CHUA T S. Neural factorization machines for sparse predictive analytics [C]// Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2017: 355–364. DOI: 10.1145/3077136.3080777.
- [35] XIAO J, YE H, HE X N, et al. Attentional factorization machines: Learning the weight of feature interactions via attention networks [C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2017: 3119–3125. DOI: 10.24963/ijcai.2017/435.
- [36] ZHOU G R, ZHU X Q, SONG C R, et al. Deep interest network for click-through rate prediction [C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1059–1068. DOI: 10.1145/3219819.3219823.
- [37] ZHOU G R, MOU N, FAN Y, et al. Deep interest evolution network for click-through rate prediction [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 5941–5948. DOI: 10.1609/aaai.v33i01.33015941.
- [38] FENG Y, LÜ F, SHEN W, et al. Deep session interest network for click-through rate prediction [C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2019: 2301–2307. DOI: 10.24963/ijcai.2019/319.
- [39] SMIRNOVA E, VASILE F. Contextual sequence modeling for recommendation with recurrent neural networks [C]// Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems. New York: ACM, 2017: 2–9. DOI: 10.1145/3125486.3125488.
- [40] TANG D, AGARWAL A, O'BRIEN D, et al. Overlapping experiment infrastructure: more, better, faster experimentation [C]// Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2010: 17–26. DOI: 10.1145/1835804.1835810.
- [41] GUO H F, YU J K, LIU Q, et al. PAL: A position-bias aware learning framework for CTR prediction in live recommender systems [C]// The 13th ACM Conference. New York: ACM, 2019: 452–456. DOI: 10.1145/3298689.3347033.
- [42] ZHAO Z, HONG L C, WEI L, et al. Recommending what video to watch next: A multitask ranking system [C]// Proceedings of the 13th ACM Conference on Recommender Systems. New York: ACM, 2019: 43–51. DOI: 10.1145/3298689.3346997.
- [43] SAITO Y, YAGINUMA S, NISHINO Y, et al. Unbiased recommender learning from missing-not-at-random implicit feedback [C]// Proceedings of the 13th International Conference on Web Search and Data Mining. New York: ACM, 2020: 501–509. DOI: 10.1145/3336191.3371783.
- [44] DING J T, QUAN Y H, HE X N, et al. Reinforced negative sampling for recommendation with exposure data [C]// 28th International Joint Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2019: 2230–2236. DOI: 10.24963/ijcai.2019/309.
- [45] YAO S R, HUANG B. Beyond parity: Fairness Objectives for Collaborative Filtering [EB/OL].[2017-05-24]. <https://arxiv.org/pdf/1705.08804.pdf>.
- [46] BEIGI G, MOSALLANEZHAD A, GUO R C, et al. Privacy-aware recommendation with private-attribute protection using adversarial learning [C]// Proceedings of the 13th International Conference on Web Search and Data Mining. Menlo Park: AAAI Press, 2020: 34–42. DOI: 10.1145/3336191.3371832.
- [47] WU Y K, ZHANG L, WU X T. On discrimination discovery and removal in ranked data using causal graph [C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 2536–2544. DOI: 10.1145/3219819.3220087.
- [48] YANG L Q, CUI Y, XUAN Y, et al. Unbiased offline recommender evaluation for missing-not-at-random implicit feedback [C]// Proceedings of the 12th ACM Conference on Recommender Systems. New York: ACM, 2018: 279–287. DOI: 10.1145/3240323.3240355.

- [49] AGARWAL A, TAKATSU K, ZAITSEV I, et al. A general framework for counterfactual learning-to-rank [C]// Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval . New York: ACM, 2019: 5–14. DOI: 10.1145/3331184.3331202.
- [50] AI Q Y, BI K P, LUO C, et al. Unbiased learning to rank with unbiased propensity estimation [C]// The 41st International ACM SIGIR Conference. New York: ACM, 2018: 385–394. DOI: 10.1145/3209978.3209986.
- [51] MEHRPOOR M, GJÆRDE A, SIVERTSEN O I, et al. Intelligent services: A semantic recommender system for knowledge representation in industry [C]// 2014 International Conference on Engineering, Technology and Innovation. New York: IEEE Press, 2014: 1–6. DOI: 10.1109/ICE.2014.6871539.
- [52] BENKADDOUR F Z, TAGHEZOUT N, KADDOUR-AHMED F Z, et al. An adapted approach for user profiling in a recommendation system: Application to industrial diagnosis [J]. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2018, 5(3): 118. DOI: 10.9781/ijimai.2018.06.003.
- [53] 赵栋, 景晓冬. 工程经验反馈推荐系统在核电设备设计中的应用研究[J]. 制造业自动化, 2017, 39(10): 123–127. DOI: 10.3969/j.issn.1009-0134.2017.10.028.
ZHAO D, JING X D. Research of the engineering experience feedback recommendation system in the equipment design of nuclear power [J]. *Manufacturing Automation*, 2017, 39(10): 123–127. DOI: 10.3969/j.issn.1009-0134.2017.10.028(Ch).
- [54] 彭泽普. 基于统计特征识别算法的冒口工艺推荐设计系统的开发与应用[D]. 武汉: 华中科技大学, 2019.
PENG Z P. *Development and Application of Riser Process Recommendation Design System Based on Statistical Feature Recognition Algorithm* [D]. Wuhan: Huazhong University of Science and Technology, 2019(Ch).
- [55] 牛中伟. 基于本体的难加工材料高性能加工工艺参数推荐系统[D]. 南京: 南京航空航天大学, 2012.
NIU Z W. *The Process Parameter Recommender System of High-Performance Machining for Difficult-to-Cut Materials Based on Ontology* [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2012(Ch).
- [56] KWON J, ZIEGLER M M, CARLONI L P. A learning-based recommender system for autotuning design flows of industrial high-performance processors [C]// 2019 56th ACM/IEEE Design Automation Conference (DAC). New York: IEEE Press, 2019: 1–6.
- [57] 孙妍. 制造协作网中基于信誉评估的供应商推荐系统研究与实现[D]. 武汉: 武汉理工大学, 2017.
SUN Y. *Research and Development on Supplier Recommendation System Based on Reputation Evaluation in CMN* [D]. Wuhan: Wuhan University of Technology, 2017(Ch).
- [58] ZHANG J, HAO B W, CHEN B, et al. Hierarchical reinforcement learning for course recommendation in MOOCs [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33: 435–442. DOI: 10.1609/aaai.v33i01.3301435.
- [59] ZOU L X, XIA L, DING Z Y, et al. Reinforcement learning to optimize long-term user engagement in recommender systems [C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 2810–2818. DOI: 10.1145/3292500.3330668.
- [60] MUHAMMAD K, WANG Q Q, O'REILLY-MORGAN D, et al. FedFast: Going beyond average for faster training of federated recommender systems [C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2020: 1234–1242. DOI: 10.1145/3394486.3403176.

□