计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(4):47−58 [doi: 10.15888/j.cnki.csa.008403]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

智能推荐系统研究综述①

胡　琪[1], 朱定局[1], 吴惠粦[2], 巫丽红[3]

[1](华南师范大学 计算机学院, 广州 510630)
[2](广州国家现代农业产业科技创新中心, 广州 510030)
[3](广东农工商职业技术学院, 广州 510507)
通信作者: 朱定局, E-mail: zhudingju@m.scnu.edu.cn

摘　要: 伴随着电子商务平台和新型数字媒体服务迅速发展, 网络数据规模持续增长, 数据类型呈现多样化, 如何从大规模数据中挖掘有价值的信息, 已经成为信息技术的一项巨大挑战. 推荐系统能够缓解“信息过载”问题, 挖掘数据潜在价值, 将个性化信息推送给有需要的用户, 提高信息利用率. 深度学习的表征能力与推荐系统相融合, 有助于深层次地挖掘用户需求, 提供精准的个性化推荐服务. 本文首先分析传统推荐算法的优缺点, 再总结深度学习技术在推荐系统中的研究进展. 最后, 分析和展望智能推荐系统未来发展方向.

关键词: 推荐系统; 深度学习; 信息过载; 推荐算法; 个性化; 协同过滤; 目标检测

引用格式: 胡琪,朱定局,吴惠粦,巫丽红.智能推荐系统研究综述.计算机系统应用,2022,31(4):47–58. http://www.c-s-a.org.cn/1003-3254/8403.html

Survey on Intelligent Recommendation System

HU Qi[1], ZHU Ding-Ju[1], WU Hui-Lin[2], WU Li-Hong[3]

[1](School of Computer Science, South China Normal University, Guangzhou 510630, China)
[2](Guangzhou National Modern Agricultural Industry Technology Innovation Center, Guangzhou 510030, China)
[3](Guangdong Vocational College of Agriculture, Industry and Commerce, Guangzhou 510507, China)

Abstract: With the rapid development of e-commerce platforms and new digital media services, the scale of network data continues to grow and data types are diversified. The mining of valuable information from large-scale data has become a huge challenge for information technology. Recommendation systems can alleviate the “information overload” problem, explore the potential value of data, push personalized information to users in need, and improve information utilization. The combination of the representational capabilities of deep learning and recommendation systems helps to dig deeper into user needs and provide accurate personalized recommendation services. This study analyzes the advantages and disadvantages of traditional recommendation algorithms, summarizes the research progress of deep learning technology in recommendation systems, and probes into the future development directions of intelligent recommendation systems.

Key words: recommendation system; deep learning; information overload; recommendation algorithm; personalization; collaborative filtering; target detection

1　引言

互联网信息服务不断扩展, 为用户提供更多的信息服务, 也加快数据规模的增长. 互联网数据包括用户个人信息, 浏览记录、消费历史、项目属性等数据, 如果不对这些数据加以利用, 会极大地浪费存储资源, 造成“信息过载”问题[1]. 推荐系统技术能够挖掘数据隐含价值, 协同用户数据和项目属性捕捉客户的需求, 提供个性化信息服务. 让用户获取所需要的信息, 从而提高

① 基金项目: 中国高等教育学会专项课题 (2020JXD01); 广东省普通高校“人工智能”重点领域专项 (2019KZDZX1027); 广东高校省级重点平台和重大科研项目 (2017KTSCX048); 广东省公益研究与能力建设 (2018B070714018); 广东省中医药局科研项目 (20191411); 广东省高等学校产业学院建设项目 (人工智能机器人教育产业学院)
收稿时间: 2021-06-11; 修改时间: 2021-07-14; 采用时间: 2021-07-20; csa 在线出版时间: 2022-03-22

数据的有效利用率. 推荐系统在缓解数据过载的问题中发挥着重要作用, 能够协助用户发现潜在的兴趣[2], 缓解数据过量导致用户无法发现自己需要的信息.

推荐系统已经成为许多电子商务和多媒体平台的内核, 个性化推荐服务能够帮助平台吸引用户的注意力, 提高用户访问量. 推荐系统为网络平台的发展提供源源不断的动力, 其商业价值也引起工业界和学术界的关注. 深度学习作为一项热门技术, 已经在计算机视觉、自然语言处理等多个领域展现出无限潜力, 也为推荐系统提供了新的方法[3]. 凭借深度学习技术的强大表征能力, 学习用户和项目的隐向量表示, 挖掘用户的历史行为数据、商品的多样化数据以及上下文场景信息, 捕获用户潜在偏好, 向用户生成更加精确的个性化推荐列表.

本文主要综述推荐系统的发展脉络, 总结传统的推荐算法的优缺点, 分析深度学习技术在推荐系统中的前沿应用, 并且展望深度学习推荐算法未来研究方向.

2 传统推荐算法

2.1 协同过滤

协同过滤[4] 是早期使用最为广泛的推荐算法, 核心思想是综合用户和项目显式反馈信息, 筛选出目标用户可能感兴趣的项目进行推荐. 协同过滤算法主要类型可分为基于用户的协同过滤和基于物品的协同过滤, 两种类型的算法都需要基于构建的用户和项目的二元共现矩阵, 协同整个矩阵数据去预测用户对项目的评分. 基于用户的协同过滤, 需要计算用户之间的相似度, 找到与目标用户类似的用户, 加权求和相似用户的评分作为目标用户对项目的预测评分, 对评分排序生成推荐项目列表. 2003 年, Amazon 团队[5] 发表关于协同过滤的论文, 介绍基于物品的协同过滤在商品推荐服务中的应用. 该算法基于共现矩阵, 找到目标用户评价高的物品, 利用物品向量计算物品之间的相似度, 最终将与高评分物品的类似物品作为推荐列表的结果. 协同过滤算法具备可解释性, 能够发掘出用户新的兴趣点, 但随着用户和物品的规模增大, 共现矩阵数据会变得更加稀疏, 计算相似度时准确率会降低, 影响算法实际效果. 且推荐结果的头部效应明显, 评分高的受欢迎物品会多次推荐, 而评分信息少的新物品较少推荐, 算法泛化能力较差.

2.2 矩阵分解

2006 年, 矩阵分解[6] 算法在用户评分预测任务中表现出色, 缩小预测评分与用户真实评分的误差. 算法主要思想是通过分解共现矩阵, 为用户和项目分别生成一个隐向量, 使用隐向量表示用户的兴趣和项目的属性, 用于挖掘用户与项目之间深层次潜在关系, 从而提高预测准确性. 矩阵分解算法通过使用奇异值分解 (SVD)、特征根结构分解 (ED) 等方法分解共现矩阵分别得到用户隐向量p_u和物品的隐向量q_i, 用户向量与项目向量间的点积为用户对项目的预测评分, 物品预测值和真实值之差作为损失函数, 如式 (1):

$$e_{ui} = r_{ui} - q_i^{\mathrm{T}} p_u \tag{1}$$

其中, r_{ui}为用户 u 对物品 i 真实评分标签, p_u与q_i分别为用户和物品向量, p_u与q_i的点积作为用户 u 对物品 i 的预测评分. 使用梯度下降算法训练模型, 加入正则化项防止过拟合. 相较于协同过滤算法, 矩阵分解泛化能力更强, 缓解数据稀疏问题. 空间复杂度更低, 只需保存用户和项目向量. 矩阵分解算法分解得出的向量隐含用户信息和项目信息, 但隐向量缺乏可解释性. 矩阵分解仅利用用户与项目的评分信息, 没有使用其他相关特征信息, 损失了有用信息, 且无法有效解决冷启动问题.

2.3 逻辑回归模型

协同过滤和矩阵分解算法只利用用户与项目的交互信息, 而逻辑回归 (logistic regression, LR)[7] 模型能融合用户画像特征、物品属性、上下文信息, 将特征转化为数值向量, 输入到网络中训练, 学习各个特征的权重, 输出层预测样本为正的概率. 逻辑回归模型有益于并行化计算, 模型较为简单易于部署而广泛应用, 但表征能力有限, 没有进行多特征交叉组合, 特征筛选, 影响预测准确性. 2017 年, 阿里巴巴团队[8] 提出混合逻辑回归模型 (mixed logistic regression, MLR), 由于传统的逻辑回归模型表达能力有限, 无法拟合复杂非线性表达式, MLR 模型吸收“分而治之”的思想, 将特征空间分成几个区域, 在每个区域训练一个线性模型, 将不同区域的线性模型结果进行加权求和作为最终的输出结果. 只要 MLR 模型具有足够的分割区域, 可以拟合任意非线性函数. 相较于传统的 LR 模型, MLR 模型可以扩展到大量样本和高维特征, 在稀疏数据中学习数据的非线性表示.

2.4 因子分解机模型

2010 年, Rendle[9] 提出因子分解机模型, 在逻辑回归的基础上, 加入二阶交叉特征组合. FM 算法为每一个特征引入了一个具有低维稠密的隐向量特征, 并使

用向量特征的内积作为特征交叉的权重, 如式 (2). 即使两个特征共同存在的数据较少, 也可以衡量两者之间的相关性, 从而缓解了数据稀疏所导致的难以计算特征交互的问题. 相较于逻辑回归模型, FM 模型表达能力更强. 但受限于组合爆炸问题, 导致特征组合无法扩展到三阶及以上.

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=n+1}^{n} \left\langle v_i, v_j \right\rangle x_i x_j \qquad (2)$$

其中, w_0 为全局偏置, w_i 表示第 i 个特征的权重, $\left\langle v_i, v_j \right\rangle$ 为特征隐向量 $v_i v_j$ 的内积, 内积值作为特征交叉的权重, 最终预测值 $\hat{y}(x)$ 为一阶特征与二阶交叉特征求和. 在 FM 模型基础上, FFM 模型[10] 把相同性质的特征归为同一个域, 细化特征组合的表示. 每个隐向量对应一个域, 当两个特征 x_i 和 x_{i+1} 组合时, 用特征对应域的隐向量内积作为权重. FFM 模型精细化表示特征组合, 同时也扩大了训练参数量, 增加了过拟合的风险.

2.5 GBDT+LR 组合模型

2014 年, Facebook 团队[11] 将梯度提升决策树与逻辑回归结合起来, 使用组合模型完成推荐任务, 模型结构如图 1 所示. 该模型的主要思想是使用梯度提升决策树进行自动化特征工程, 提取重要特征和进行特征组合, 树的最后一层叶节点生成新的离散特征, 作为逻辑回归模型的输入据经过激活函数后输出预测结果. 该组合模型的提出, 推进了特征工程模型化进程, 能够减少人工进行特征组合和特征筛选的工作量, 实现模型端到端训练.

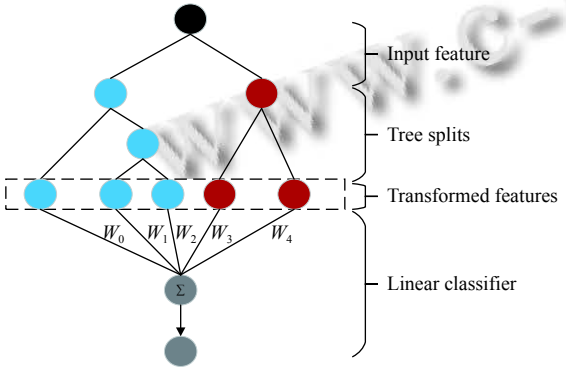

图 1 GBDT+LR 组合模型结构

总的来说, 传统推荐算法种类繁多, 具有不同的优势, 需要结合实际的推荐场景加以灵活运用, 表 1 列出各个算法的优势和存在的劣势.

表 1 传统推荐算法对比

传统推荐算法	优势	劣势
协同过滤算法	算法简单 不需要领域知识 能发掘新的兴趣点	数据稀疏 冷启动问题 头部效应明显
矩阵分解算法	泛化能力加强 缓解数据稀疏问题 泛化能力加强	损失其它相关用户物品和上下文信息 缺乏解释性 损失其它相关用户物品和上下文信息
逻辑回归模型	模型简单, 易于实现 学习各个特征权重, 具有可解释性 模型简单, 易于实现	表达能力较差 没有进行特征组合和特征筛选 表达能力较差
因子分解机模型	解决稀疏数据交叉特征组合问题 模型表达能力增强	模型参数多, 训练困难 容易过拟合 无法学习三阶及以上特征
梯度提升树+逻辑回归组合模型	自动化特征组合 端到端训练 减少手工特征组合	泛化能力差 容易过拟合

3 深度学习技术

深度学习技术已经在人工智能领域取得很多研究成果, 深度学习与推荐系统相结合, 能够缓解传统推荐模型表达能力不足的问题. 深度学习的表征能力更强, 需要大量数据训练模型, 能够缓解数据规模大和数据稀疏问题. 深度学习的基本结构有: 多层感知机、卷积神经网络、循环神经网络、注意力机制等.

3.1 多层感知机

多层感知机是前馈结构的神经网络, 数据通过输入层, 经过多个隐藏层, 汇入输出层计算最终结果, 网络结构如图 2 所示, 利用 BP 反向传播算法来监督训练神经网络, 调整每层神经元的权重, 拟合非线性函数, 缩小预测值与真实值的误差. 多层感知机在推荐系统中常用于挖掘高阶特征交叉[12], 学习潜在数据模式.

3.2 卷积神经网络

卷积神经网络 CNN 是模仿生物视觉系统构建的网络结构[13], 使用卷积操作处理二维数据特征, 在计算机视觉领域应用广泛. CNN 中的卷积运算的参数共享减少了模型中需要学习的参数数量, 相较于全连接神经网络计算效率更高. 在推荐系统中卷积神经网络主要用于提取视觉特征、文本特征, 融合用户画像特征, 从更多方面捕获用户偏好, 常应用于图片推荐、新闻推荐、多模态推荐等场景.

3.3 循环神经网络

循环神经网络 RNN 是一种常用于处理时间序列数据的深度网络结构, 结构如图 3 所示. RNN 不仅能够进行前馈计算, 且能够保持上个时刻的信息, 利用历史状态数据和当前状态预测输出[14], 因此可以处理文本和音频等序列数据.

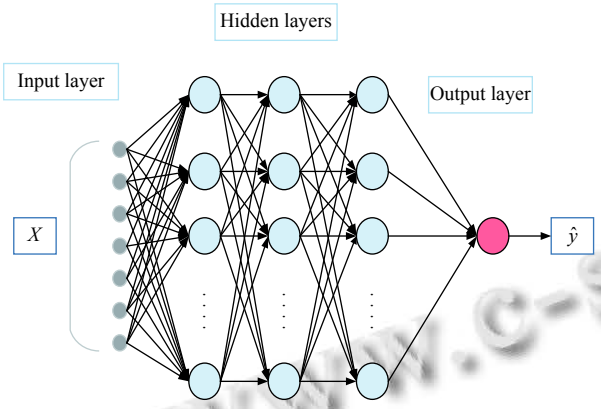

图 2 多层感知机结构

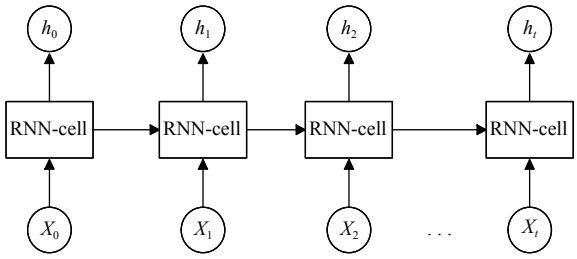

图 3 循环神经网络结构

为了解决时间间隔过长导致的信息流失问题和梯度消失与爆炸问题, 构建出新的变体长短期记忆网络 (LSTM)[15] 和门控循环单元 (GRU)[16]. 在推荐系统中, 循环神经网络可用在基于会话推荐, 基于用户当前会话行为, 学习用户的兴趣迁移过程, 预测用户下一个可能交互的项目.

3.4 注意力机制

注意力机制是一种模仿人类视觉的局部信号处理机制, 人在观察事物过程中通常关注于部分重要信息, 减少对无关信息的注意力, 从而快速做出判断. 注意力机制帮助推荐模型选择更有效的特征, 让模型关注于更重要信息, 减小数据噪声对结果的影响[17]. 深度学习的训练过程常被看做“黑盒”, 整个训练过程无法预知, 输出的结果无法提供很好的解释性. 深度推荐模型与注意力机制结合, 有利于增强模型的可解释性, 对各种特征赋予不同的注意力分数, 增强有效特征的影响力, 抑制数据噪声, 提升模型的推荐准确性.

4 深度学习在推荐系统中的应用

传统的推荐算法结构简单, 容易实现, 可以灵活运用于推荐任务. 然而在大数据的背景下, 传统推荐模型能力有限, 泛化能力较差, 无法很好应用在大规模数据和数据稀疏场景. 深度学习技术有助于推荐系统应用于大规模数据场景中, 深度学习深层复杂网络结构需要大量数据训练整个模型, 稀疏特征可以借由神经网络转换为蕴含丰富信息的低维度稠密向量. 复杂网络结构能够拟合任意非线性函数, 挖掘数据深层次的潜在模式. 深度学习模型可扩展性强, 能融合多种异构数据, 从多方面捕获用户兴趣, 提高模型的预测准确度. 本节主要分析深度学习在推荐场景中的应用.

4.1 嵌入技术在推荐系统中的应用

推荐系统通常使用嵌入 (embedding) 技术用低维度稠密向量去表征一个对象, 该对象可以是一个项目、一个用户等, 同时向量之间的距离隐含项目与项目之间、用户与用户之间、用户与项目之间的关系. 嵌入技术已经成为推荐系统中必不可少的环节, 主要处理稀疏特征, 融合大量信息形成一个有价值的低维向量, 输入到神经网络中训练模型. 也可以利用向量之间的关系, 作为召回策略, 筛选出与用户兴趣匹配的候选项目.

Grbovic 等人[18] 在房屋短租平台应用嵌入方法表征用户和推荐列表. 在 Skip-Gram 的基础上, 针对该平台在搜索排序和推荐实时个性化中设计了列表和用户的嵌入向量. 用户的搜索会话中的数据作为类似序列信息, 使用词向量[14] 方式学习每个房源的嵌入向量, 有效表征房源多个特征, 结合实际业务场景, 向用户精确推荐优质房源.

阿里巴巴团队[19] 利用嵌入技术用于学习 ID 类型数据的表示, 用于电商场景的推荐, 包括用户 ID、商品 ID、种类 ID 等, 传统的独热编码方式会导致数据过于稀疏, 且无法表示对象之间的潜在关系, 在电子商务平台中, ID 类数据非常稀疏, 动辄达到几亿维度, 需要使用低维度的向量高效表达 ID 数据. 该文基于 Item2Vec[20] 提出基于嵌入的框架, 通过采集用户行为的 ID 序列, 结合 ID 之间的结构化的联系, 能够为不同类型的 ID 学习一个低维向量用以表示. 在此基础上, 阿里巴巴团队[21] 提出基于图的嵌入方法用于推荐系

统, 为了解决阿里电商数亿规模的数据稀疏、数据量大以及存在的商品冷启动问题. 该方法首先基于会话构造一个商品有向图, 基于图构造与商品有交互的行为序列, 结合特征生成项目的图嵌入向量, 对每个向量进行特征加权. 该算法主要用于召回阶段, 基于与用户有过交互的商品, 召回相关候选项目.

Wu 等人[22] 提出 SR-GNN 模型, 考虑到物品转换成向量的复杂过程, 提出一种新的嵌入方式, 使用图数据结构对用户会话进行建模, 利用图神经网络学习图中节点的嵌入向量. 最后, 通过注意力机制把每个会话表征为当前会话的兴趣和全局兴趣的构成, 基于每个会话, 预测下一个项目交互概率. 该模型克服难以用隐向量表示项目的问题, 使用图结构模型生成精准项目嵌入向量, 为基于会话的推荐场景提供新的方法.

4.2 基于多层感知机的推荐模型

多层感知机模型在推荐系统中应用广泛, 通常原始数据经过嵌入层形成向量后, 会输入到多层感知机中, 学习数据非线性表示, 在进行低阶特征交叉后, 结合多层感知机进一步提取高阶特征交叉[23], 可应用在预测用对项目评分、精准排序任务和用户点击率预测.

2016 年, YouTube 团队[24] 将 DNN 应用在视频推荐服务中, 用神经网络对候选视频进行预测评分, 根据分数排序生成推荐列表. YouTube 平台的用户数量和视频规模庞大, 传统小规模数据集的算法并不适用, 同时平台的视频更新速度快, 需要平衡已有视频和新发布的视频所带来的冷启动问题, 追踪用户的实时行为. 推荐平台架构如图 4 所示.

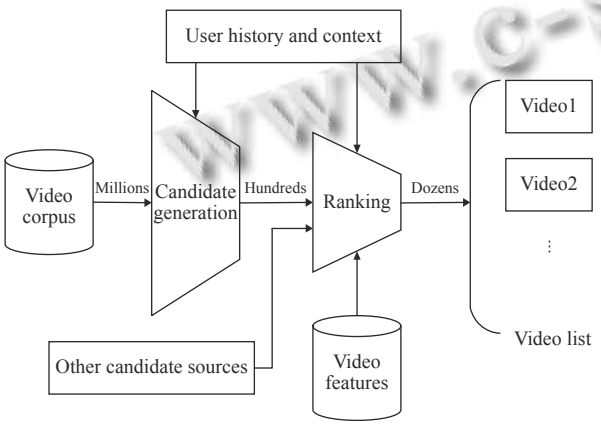

图 4 YouTube 平台架构

整个系统分为匹配阶段和排序阶段, 匹配阶段利用高效召回策略从百万级规模的视频库中召回用户可能感兴趣的候选项, 该阶段要求搜索效率高, 并且检索出的视频与用户的历史行为和偏好相关. 排序阶段主要是对召回的视频进行精粒度的打分排序, 将神经网络融合用户特征、视频属性和场景信息输入到模型中, 对候选视频进行评分预测, 依据分数进行排序, 选取高评分视频作为推荐列表.

Cheng 等人[25] 提出深广 (Wide & Deep) 模型, 模型由 Wide 部分和 Deep 部分构成, 模型结构如图 5 所示, 其中 Wide 部分使用线性模型, 提取数据的一阶特征, Deep 部分使用神经网络自动学习高阶特征提高泛化能力, 最终将两个部分的结果整合通过 Sigmoid 激活函数后输出预测结果. 模型中的 Wide 部分对应于模型的记忆能力, 从用户的数据中发现特征之间的相关性, 偏向于推荐和用户历史行为相关的内容. Deep 部分对应于模型的泛化能力, 稀疏特征经嵌入层形成低维稠密向量输入到隐藏层中, 利用神经网络的学习能力捕获新的潜在高阶特征组合, 泛化能力有利于推荐结果个性化, 让推荐结果具有多样性.

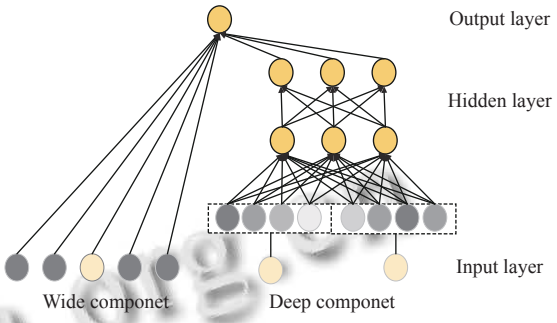

图 5 Wide & Deep 模型结构

多层感知机模型广泛应用于点击率预估任务, 能够充分利用用户画像特征、项目属性特征和上下文信息, 进行特征提取, 且能缓解数据稀疏、高阶特征组合等问题. 多层感知机与因子分解机进行结合, 可以弥补 FM 和 FFM 模型中特征组合无法扩展到三阶及以上的劣势. 通常在稀疏特征经过嵌入层转化为低维向量后, 进行低阶交叉特征组合并且利用 DNN 提取高阶特征组合, 经过 Sigmoid 函数输出点击概率, 例如 DeepFM[26]、FAT-DeepFFM[27]、NFM[28] 等模型. 协同过滤与神经网络相结合, 缓解稀疏特征导致的训练困难问题, He 等人[29] 提出神经协同过滤模型, 将矩阵分解的处理方式和深度学习融合, 模型结构如图 6. 神经协同过滤模型主要对隐式反馈数据进行建模, 用嵌入向量表征用户

和物品, 输入到多层神经网络中, 输出层预测用户评分, 采用平方损失函数训练模型. 利用神经网络学习隐向量表示用户和项目之间潜在关系, 将用户和项目映射到隐向量空间, 向量之间的距离反映出用户和项目的潜在关系, 可用于召回阶段, 计算相关性召回与目标用户相关项目候选集合.

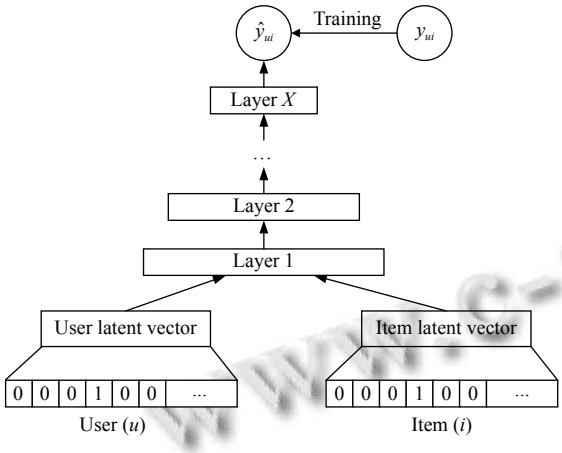

图 6 NCF 模型结构

Wang 等人[30] 针对点击率预估任务, 讨论归一化操作对于点击率预估效果的影响, 如层归一化、批次归一化、仅有方差的归一化. 通过对比实验, 将对向量化之后的特征进行归一化, 连续数值型特征使用层归一化, 稀疏分类性特征做批次归一化, 在多层感知机中使用偏差的归一化能够提升点击率预测准确性.

Huang 等人[31] 借鉴计算机视觉和自然语言处理中的门机制, 提升非凸神经网络的可训练性, 在嵌入层增加门机制, 用于从特征选择更重要的特征, 在隐藏层加入门机制, 用于筛选更加重要的特征交互传递到更深层的网络. 门机制的思想类似于注意力机制, 增强有效特征, 抑制数据噪声.

4.3 基于卷积神经网络的推荐系统

卷积神经网络的卷积和池化计算主要学习数据局部特征[32], 可以提取非结构化多媒体数据, 对多源异构数据进行表征学习. 网络可以融合多样化信息, 如物品图像、评论文本等, 挖掘用户视觉兴趣或从文本信息中提取用户偏好. 卷积神经网络提高了模型的可扩展性, 融合更多信息能够让模型从更多方面捕捉用户兴趣. 在推荐系统中, 卷积神经网络适用于多模态推荐、图片推荐和文本推荐任务.

通常用户的行为容易受到图像的影响, 光鲜的商品图片往往能够吸引到用户的注意力. Zhou 等人[33] 尝试用卷积神经网络捕捉分析用户喜欢的图像来提取用户视觉兴趣画像, 该系统通过计算视觉兴趣向量的余弦相似性, 找到符合用户视觉兴趣的住房. 该模型应用于酒店预订系统, 用图像特征预测用户喜欢的住房风格, 实现个性化推荐. Tang 等人[34] 提出卷积序列嵌入推荐模型, 将用户过去交互的商品看成序列, 预测用户未来可能交互的项目, 其思想是将在时间和空间上最近的序列形成一个“图像”, 使用卷积滤波器学习序列模式作为图像的局部特征. 该模型使用卷积神经网络学习序列特征, 用隐因子模型学习用户特征.

有相关研究利用卷积神经网络抽取文本特征, 使模型融入文本信息, 向用户推荐相关感兴趣的文字内容. Shen 等人[35] 将 CNN 用于在线学习资源推荐中, 模型使用卷积神经从学习资源的介绍、内容等文本信息中提取项目特征, 对于输入采用语言模型, 对于输出采用 L1 范数正则化的潜在因子模型, 在此基础上引入分裂 Bregman 迭代法求解该模型, 给学生推荐正确的学习资源. Gong 等人[36] 采用带注意力的 CNN 处理标签推荐问题, 整个模型由两部分组成前一部分用于获取文本特征, 后一部分对各个文本的表示进行 Softmax 多标签分类. 将 CNN 模型卷积层应用在预训练词向量上, 加入注意力机制, 利用注意力层来产生一个单词相对于它周围的单词的权重.

Zheng 等人[37] 构建 DeepCoNN 模型使用文本评论对用户行为和商品属性进行联合建模. 两个神经网络顶部的额外共享层连接了两个并行网络, 因此用户和项目表示可以相互交互以预测点击率. 整个模型有 3 层组成 Lookup 层, CNN 层, 输出层, Lookup 层将用户评论和商品评论转化为对应词向量, 输入到 CNN 中, 最后的输出预测结果, 训练模型缩小误差.

Liu 等人[38] 提出 FGCNN 模型, 使用卷积神经网络提取局部模式并且组合生成新的特征, 为防止全局信息的丢失, 引入多层感知机提取全局特征交互, 最终在 Criteo 数据集中 AUC 达到 80.22% 的效果. 模型通过 CNN 与 MLP 相结合的方式学习有用的局部特征和全局特征, 既减少手工特征量, 又缓解因特征稀疏导致的神经网络训练困难问题.

2020 年, 京东团队提出 CSCNN 模型[39], 有效利用电商平台中丰富的商品类目信息, 使用卷积神经网络提取图像信息, 创新性地将商品信息和商品主图作为

图像特征提取模块的输入, 提取商品主图中丰富的视觉特征, 有效挖掘商品视觉属性, 学习商品图像对于用户行为的影响, 提高点击概率预测的准确性.

4.4 基于循环神经网络的推荐系统

多层感知机和卷积神经网络是前馈结构的网络, 层与层之间全连接, 但每层神经元节点之间无任何连接, 不利于建模文本或者音频等时序数据, 因此提出循环神经网络 RNN, 处理时序数据. RNN 的最大特点在于神经网络具有记忆性并且能够参数共享, 它能够获取某一时刻的输入数据和前一时刻的隐层状态来预测当前时刻的输出. 近年来, 循环神经网络已经在机器翻译、自然语言处理领域中取得很多研究进展. 在推荐系统中, 主要使用循环神经网络具有记忆性的特点对用户的历史会话序列建模, 学习用户偏好演变过程以及用户上下文相关兴趣, 应用于会话推荐任务.

在基于会话的推荐中, 用户的行为和兴趣随着时间推移不断改变, 用户当前行为与历史浏览和搜索行为具有较强联系. Hidasi 等人[40] 在短会话推荐任务中使用循环神经网络, 把用户与物品交互行为组成行为序列, 输入模型中训练, 预测下一项目交互概率. 该模型采用 GRU 模型作为基本单元, 引入会话并行小批量数据, 该模型采用 GRU 模型作为基础结构单元, 对小批量的输出采样, 使用排序损失函数训练模型, 拟合目标任务, 捕获用户兴趣随着时间推移的演变过程. Hidasi 等人[41] 在此基础上, 进一步优化采样方法和损失函数, 为解决训练过程中存在的梯度消失问题, 提出新的损失函数 Top-Max 和 BPR-Max, 进一步提升模型训练效果.

Devooght 等人[42] 将协同过滤视为时间序列的预测问题, 应用 LSTM 捕捉用户的喜好演变过程. 每个项目用独热编码表示, 采样用户的历史行为作为时间序列, 将项目的向量输入到 RNN 模型中, 输出为每个项目对应神经元的 Softmax 值, 推荐输出层概率最大的若干个项目. Donkers 等人[43] 首次提出将用户编码信息融入 GRU 的网络结构中, 通过深度集成用户信息, 能够更有效的对用户行为序列建模. 通过改造 GRU 的门控结构, 整合用户信息以及行为序列到模型中, 完成个性化序列预测任务, 有效学习用户行为事件之间的隐藏关系, 预测用户兴趣进行下一项推荐.

考虑到过去的研究大多利用用户短期行为, 而没有考虑顾客长期稳定的偏好和演化过程. Li 等人[44] 提出 BINN 模型, 通过结合用户的偏好和当前消费动机来进行下一项推荐. 模型挖掘大量用户行为日志如浏览、点击、收藏等历史记录, 形成随时间推移的行为序列, 这些丰富信息有利于学习用户潜在兴趣. 使用新的神经物品嵌入方法, 获取统一的物品表示空间, 学习物品的潜在向量, 捕获物品之间的序列相关性. 开发出基于长短期记忆神经网络学习个人偏好和当前消费动机, 进行序列化推荐.

Feng 等人[45] 提出 DSIN 模型, 应用循环神经网络从用户行为序列中捕获动态不断变化的用户兴趣, DSIN 模型能够有效对用户对个会话进行建模, 用于点击率估计预测. 用户的连续行为由多个历史会话组成, 用户在每个会话和异构交叉会话中的行为是高度同构的, 加入自注意力机制提取用户在每个会话中的兴趣, 应用双向 LSTM 来捕获上下文会话兴趣的顺序关系, 最后使用本地激活单元来聚合用户对目标项的不同会话兴趣表示, 完成基于会话的推荐.

4.5 基于注意力机制的推荐系统

用户的兴趣具有多样性, 并且会随时间不断变迁, 用户点击行为具有局部活跃性, 某一时刻点击行为仅仅和过去的部分历史数据有关, 而不是所有历史记录. Zhou 等人[46] 提出 DIN 模型, 模型中引入注意力机制, 对用户行为序列数据建模, 将用户行为基于注意力机制进行加权求和, 使模型更加关注有益信息, 预测下一次点击动作. 并且提出小批量正则方法和自适应激活函数辅助模型训练. 2018 年, 在 DIN 模型基础上, 又设计出 DIEN 模型[47], 该模型使用 GRU 结构构建模型用户行为序列. DIEN 设计了兴趣提取层, 捕获用户随时间改变的兴趣演变过程. 同时设计了兴趣演化层来捕获与目标项相关的兴趣演化过程. GRU 每一步的局部激活都能增强相对兴趣的影响, 减弱用户兴趣迁移的干扰, 有助于充分学习相对于目标项目的兴趣演化过程.

阿里巴巴团队提出 ATRank 模型[48], 该模型基于注意力机制对用户异构行为序列建模, 融合用户不同的行为记录, 更好地理解用户兴趣, 提供更优质的个性化服务. 整个模型包括原始特征、语义映射层、自注意力层和目标网络. 语义映射层能让不同的行为可以在不同的语义空间下进行比较和相互作用. 自注意力层让单个的行为本身变成考虑到其他行为影响的记录. 目标网络则通过 Vanilla Attention 可以准确的找到相关的用户行为进行预测任务. 使用类似 Google 的自注

意力机制去除 CNN、LSTM 的限制, 加快网络训练速度, 提升预测效果.

Xiao 等人[49] 提出 AFM 模型, 该模型将注意力机制与 FM 算法融合. 考虑到 FM 算法虽然高效, 但它对所有特征交互的建模具有相同权重, 可能会影响预测的准确性, 并不是所有特征交互都对预测结果有益且具有预测性. 无用特征之间的交互可能引入噪声, 从而降低模型性能. 所以在 FM 算法中加入注意力机制, 为每个交叉特征计算一个权重表示对预测结果的影响大小.

Song 等人[50] 提出 AutoInt 模型利用多头自注意力机制来完成自动特征提取. 高阶特征组合有利于提升点击率估计准确度, 但依靠经验进行人工特征组合工作量非常大. 该模型通过自注意力机制构建特征交互层, 交互层的层数可作为超参数调整, 交互层叠加可以学习二阶、三阶及以上高阶组合. 在第一层的交互中, 通过注意力映射可以学习不同特征的相关性, 以加权求和的方式进行组合. 同时使用残差连接防止交互层加深导致神经网络退化问题, 防止梯度弥散问题.

依据用历史行为记录, 建模用户偏好动态渐变过程, 是对推荐系统的巨大挑战. 现有算法使用序列神经网络, 遵从从左到右的顺序, 利用单向信息建模, 这种严格的顺序降低了历史序列的表示能力, 影响准确性. 2019 年, Sun 等人[51] 提出 Bert4Rec 模型, 首次将 BERT 模型用于推荐系统, 由于深度双向信息会造成信息的泄露, 为了解决这个问题, 使用 Cloze Task 训练模型, 利用上下文信息预测 Masked Item, 在预测过程中, 将 Mask 加入到输入序列的最后, 然后利用 Mask 的嵌入向量进行推荐.

多数点击概率估计模型只考虑某一广告的信息而忽略其他相关广告对预测结果, Ouyang 等人[52] 提出 DSTN 模型, 将不同类型的广告作为辅助信息融入到模型之中, 如用户历史点击或者曝光未点击的广告和当前上下文已经出现过的广告等, 加入注意力提取对目标广告有用的辅助信息, 减少噪声数据影响, 利用上下文信息提高模型准确度.

大多数深度学习推荐模型将原始稀疏特征嵌入到低维向量, 输入到神经网络中获得最终的推荐预测概率, 这些工作只是连接不同的特征, 忽略用户行为的连续性. DIN 模型[41] 提出使用注意力机制来捕获候选项与用户先前点击商品之间的相似性, 但未考虑用户行为序列背后的序列性质. 阿里巴巴团队[53] 提出 BST 模型, 将 Transformer 技术应用于推荐系统中学习用户历史行为序列信息. 该模型相比于之前所提出的 DIN 模型准确率有较大提升. 数据经过嵌入层后, 输入到 Transformer 层, 该层用来捕获用户历史行为序列, 再与其他特征拼接输入到神经网络中进行训练.

4.6 树模型与推荐系统的结合

由于神经网络的训练过程不可预知, 推荐结果缺乏解释性, 目前很多的深度推荐模型如 Wide & Deep[25]、DeepFM[26] 等模型都是隐式地学习交叉特征, 可能引入数据噪声. 有相关研究将树模型用于有效的交叉特征, 将特征放入基于嵌入技术的注意力模型中, 不仅保障预测准确性, 也提高模型可解释性.

Wang 等人[54] 提出 TEM 模型, 使用树模型增强向量嵌入方法, 将嵌入技术和树模型的可解释的优点相结合. 该方法受到 GBDT+LR 组合模型的启发, 根据用户和物品的历史信息, 建立一个决策树来自动提取有效交叉特征, 将交叉特征输入到一个基于嵌入技术的神经注意力网络, 学习交叉特征的权重, 权重代表特征重要程度. 由于决策树提取的交叉特征明确, 而且注意力网络学习各个特征的权重, 增强模型可解释性.

阿里巴巴团队[55] 提出一种基于树结构的 TDM 模型, 解决很多模型不能调节用户和商品向量之间的内在乘积形式以利用高效搜索算法, 因此不能用于大规模推荐系统中召回候选集. 其主要思想是通过海量商品信息构建兴趣树, 自顶向下遍历兴趣树的节点并为每个用户生成推荐项, 从粗到细地预测用户的兴趣. 该方法可以从大量商品中快速检索出用户感兴趣的若干商品, 常用于推荐系统中的召回阶段.

总而言之, 深度学习技术推动了推荐系统的发展, 扩展了推荐系统特征提取能力, 增强模型表达能力, 融合更多类型特征, 学习用户多方面兴趣, 提供更多个性化推荐方法, 表 2 总结各项深度学习技术与推荐系统融合的特点描述以及优缺点分析.

5 基于深度学习的推荐系统未来发展方向

深度学习模型具备强大的表达能力, 已经证明深度神经网络能够拟合复杂的非线性函数, 深度学习技术与推荐系统的融合. 传统的机器学习模型 (如逻辑回归模型), 需要人工进行特征选择和特征交叉, 耗费大量人力, 利用神经网络对于高阶特征的自动提取和筛选, 捕获更有益的组合特征[56], 训练端到端模型, 减少

人工特征工程的工作量, 节省投入成本. 深度学习模型可拓展性强, 可以在模型中融合多种异构数据, 让模型获得更多的信息, 提高预测结果的精确性. 以下总结了几个推荐系统未来的发展方向.

表 2 深度学习技术在推荐系统中应用对比

深度学习技术	特点描述	优点	缺点
嵌入技术[18–22]	将高维离散特征转换为低维度稠密向量.	缓解独热编码造成的特征稀疏问题, 获取用户和项目的隐向量表示.	不同特征的嵌入维度难以确认, 嵌入向量的训练耗费大量时间和开销.
多层感知机[23–31]	前馈结构的神经网络, 数据通过输入层, 经过多个隐藏层, 汇入输出层计算最终结果.	结构灵活, 能够学习数据的非线性表达, 可以对高阶特征进行建模.	无法处理序列数据, 图像数据.
卷积神经网络[32–39]	具有卷积计算和池化操作的网络结构, 适用于二维图像数据局部特征的抽取.	能够提取图像特征、文本特征, 让模型能融入多样化数据, 捕捉用户更多兴趣.	CNN用于特征交互建模, 只能学习相邻特征之间部分特征交互.
循环神经网络[40–45]	具有记忆性、参数共享, 隐藏层节点之间有连接, 获取上一时刻信息并且结合当前状态预测输出.	能够对用户交互序列数据建模, 捕获用户偏好演变过程, 适用于基于会话的推荐任务.	短期记忆影响较大, 长期记忆影响小, 时序间隔变大会丢失远距离信息, 可能出现梯度消失和梯度爆炸问题.
注意力机制[46–53]	模仿人类视觉的局部信号处理机制, 对不同信息赋予不同权重, 使模型关注于更重要的信息.	计算注意力分数, 学习每个特征对于输出的影响程度, 协助模型捕获重要信息.	不能捕捉位置信息和学习序列中的顺序关系.
树模型[54,55]	一种基于特征空间划分的具有树形分支结构的模型.	时间复杂度低, 能够处理数值型和类别性数据, 有较高可解释性.	抗干扰能力弱、最优决策划分是NP难问题、对数据不均衡类别倾向数据多的类别.

5.1 深度学习与传统推荐算法的结合

传统的推荐模型结构简单, 应用广泛, 但表示能力有限, 无法挖掘深层次的用户和项目隐向量表示和高阶特征. 传统推荐算法融合深度学习技术, 弥补传统算法的不足, 利用深度推荐模型融合多种类型异构数据, 让模型吸收更多信息提高准确率, 更好捕获用户和项目的特征. 因此, 深度学习与传统推荐算法的融合, 可以充分利用二者的特点. 虽然目前已有相关的研究成果, 如 DeepFM[21]、NCF[24]、AFM[44] 等模型, 但这个方向依然具有很大的发展空间, 未来可以探索传统推荐算法与更多的深度学习模型的结合, 提出新的深度推荐模型.

5.2 深度学习推荐系统的可解释性

基于深度学习的推荐模型可以利用多源异构数据预测用户的喜好, 但模型的训练过程就像一个“黑盒”, 模型的大规模权重参数根据目标任务自动调整, 很难对模型输出的结果给予合理解释, 因此深层的神经网络高度不可解释. 如何做出可解释的推荐还是一项艰难挑战. 然而向用户提供有价值的推荐解释是非常重要的方面, 让用户明白推荐的理由, 能够加强用户对产品的理解和信任, 提升用户体验感. 注意力模型在推荐系统上的应用从一定程度上缓解了推荐模型的不可解释性[57], 因此构建解释性更强的推荐模型, 让用户理解推荐理由, 也是未来的研究热点之一.

5.3 融合更多类型数据的新网络结构

随着互联网上的信息量不断增加, 数据类型也更加多样化, 可扩展性对于模型在实际应用中的可用性十分重要. 深度推荐模型的可扩展性, 可以帮助在模型中整合多种辅助信息, 更多异构数据输入模型可以让模型从更多方面学习用户的偏好, 给出准确的预测. 未来新的深度推荐模型可以融合多模态数据, 除了使用用户与项目之间的交互数据, 还可以利用用户的时空序列数据、图像信息、项目数据的动态变化等, 对多样化数据建模能够发掘用户新兴趣点. 因此, 研究新的深度推荐模型结构去融合多种数据源也是重要的研究领域.

5.4 跨领域信息融合

许多互联网平台提供各种信息资源和网络服务, 搜索引擎提供信息搜索服务、电子商务平台提供购物服务、新闻平台推送实时热点新闻. 然而单领域推荐系统只注重某一特定领域, 而忽略了用户在其他领域的兴趣, 加剧了数据稀疏性和冷启动问题[3]. 融合多个平台的用户信息可以进行跨领域推荐, 克服单一领域的数据稀疏, 利用多个领域的信息可以挖掘用户个性化偏好. 通过深度学习技术, 可以将各类数据以向量表示作为模型输入数据, 利用源领域学习到的数据协助

目标领域作推荐[2]. 目前, 跨领域信息融合推荐方向的研究较少, 还有很大的探索空间和潜力.

5.5 新的训练优化方法与网络架构

推荐系统领域的研究, 深度学习技术在实际应用中通常会遇到两个挑战, 一是模型训练所耗费的资源多; 二是神经网络训练问题. 深度学习模型依赖大量数据进行训练, 需要足够的硬件资源来计算, 而且训练时间耗费的时间长, 同时参数调整难度大, 可能会出现收敛慢、易波动问题. 随着层数的增多, 可能出现过拟合问题. 在评估数据稀疏的情况下将导致推荐推荐模型训练不充分问题. 在之后的研究中可以尝试新的深度学习训练方法和优化策略, 在不损失预测准确度的情况下减少训练参数, 让模型变得更轻便[58], 使用负采样和剪枝算法对神经网络训练进行优化加速[59], 使用正则化方法[60] 防止过拟合, 增强深度神经网络训练效果.

6 结语

在信息技术迅速发展的时代, 互联网中的数据量也呈爆炸式增长的趋势, 随之而来的“信息过载”问题无法避免, 推荐系统在缓解信息过载问题中发挥重要作用. 深度学习技术与推荐系统相融合, 构建贴合用户兴趣的模型, 产生个性化推荐列表. 相较于传统推荐算法, 深度学习增强了模型的可扩展能力和表征能力, 让模型能够融入更多样的特征, 捕获用户的兴趣, 提高模型预测准确度. 本文分析了传统推荐算法的优缺点, 在此基础上进一步分析深度学习推荐模型的研究进展, 讨论和分析了推荐算法的研究现状和未来发展. 希望本文能够为推荐算法领域的研究人员理清脉络, 提供有益的帮助.

参考文献

1 黄立威, 江碧涛, 吕守业, 等. 基于深度学习的推荐系统研究综述. 计算机学报, 2018, 41(7): 1619–1647. [doi: 10.11897/SP.J.1016.2018.01619]

2 Zhang S, Yao L, Sun AX, *et al*. Deep learning based recommender system: A survey and new perspectives. ACM Computing Surveys, 2019, 52(1): 5.

3 Mu RH. A survey of recommender systems based on deep learning. IEEE Access, 2018, 6: 69009–69022. [doi: 10.1109/ACCESS.2018.2880197]

4 Sarwar B, Karypis G, Konstan J, *et al*. Item-based collaborative filtering recommendation algorithms. Proceedings of the 10th International Conference on World Wide Web. Hong Kong: ACM, 2001. 285–295.

5 Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 2003, 7(1): 76–80.

6 Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 30–37. [doi: 10.1109/MC.2009.263]

7 Richardson M, Dominowska E, Ragno R. Predicting clicks: Estimating the click-through rate for new ads. Proceedings of the 16th International Conference on World Wide Web. Banff: ACM, 2007. 521–530.

8 Gai K, Zhu XQ, Li H, *et al*. Learning piece-wise linear models from large scale data for ad click prediction. arXiv: 1704.05194, 2017.

9 Rendle S. Factorization machines. Proceedings of 2010 IEEE International Conference on Data Mining. Sydney: IEEE, 2010. 995–1000.

10 Juan YC, Zhuang Y, Chin WS, *et al*. Field-aware factorization machines for CTR prediction. Proceedings of the 10th ACM Conference on Recommender Systems. Boston: ACM, 2016. 43–50.

11 He XR, Pan JF, Jin O, *et al*. Practical lessons from predicting clicks on ads at Facebook. Proceedings of the Eighth International Workshop on Data Mining for Online Advertising. New York: ACM, 2014. 1–9.

12 Ruck DW, Rogers SK, Kabrisky M. Feature selection using a multilayer perceptron. Neural Network Computing, 1990, 2(2): 40–48.

13 de Andrade A. Best practices for convolutional neural networks applied to object recognition in images. arXiv: 1910.13029, 2019.

14 Medsker LR, Jain LC. Recurrent Neural Networks: Design and Applications. New York: CRC Press, 2001. 64–67.

15 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]

16 Cho K, van Merriënboer B, Gulcehre C, *et al*. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv: 1406.1078, 2014.

17 陈海涵, 吴国栋, 李景霞, 等. 基于注意力机制的深度学习推荐研究进展. 计算机工程与科学, 2021, 43(2): 370–380.

18 Grbovic M, Cheng HB. Real-time personalization using embeddings for search ranking at airbnb. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge

Discovery & Data Mining. London: ACM, 2018. 311–320.
19 Zhao K, Li YC, Shuai ZQ, *et al*. Learning and transferring IDs representation in E-commerce. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018. 1031–1039.
20 Barkan O, Koenigstein N. ITEM2VEC: Neural item embedding for collaborative filtering. Proceedings of 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing. Vietri sul Mare: IEEE, 2016. 1–6.
21 Wang JZ, Huang PP, Zhao H, *et al*. Billion-scale commodity embedding for E-commerce recommendation in Alibaba. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018. 839–848.
22 Wu S, Tang YY, Zhu YQ, *et al*. Session-based recommendation with graph neural networks. arXiv: 1811.00855, 2019.
23 Wang RX, Fu B, Fu G, *et al*. Deep & cross network for ad click predictions. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax: ACM, 2017. 12.
24 Covington P, Adams J, Sargin E. Deep neural networks for YouTube recommendations. Proceedings of the 10th ACM Conference on Recommender Systems. Boston: ACM, 2016. 191–198.
25 Cheng HT, Koc L, Harmsen J, *et al*. Wide & deep learning for recommender systems. Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston: ACM, 2016. 7–10.
26 Guo HF, Tang RM, Ye YM, *et al*. DeepFM: A factorization-machine based neural network for CTR prediction. arXiv: 1703.04247, 2017.
27 Zhang JL, Huang TW, Zhang ZQ. FAT-DeepFFM: Field attentive deep field-aware factorization machine. arXiv: 1905.06336, 2019.
28 He XN, Chua TS. Neural factorization machines for sparse predictive analytics. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Shinjuku: ACM, 2017. 355–364.
29 He XN, Liao LZ, Zhang HW, *et al*. Neural collaborative filtering. Proceedings of the 26th International Conference on World Wide Web. Perth: International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2017. 173–182
30 Wang ZQ, She QY, Zhang PT, *et al*. Correct normalization matters: Understanding the effect of normalization on deep neural network models for click-through rate prediction. arXiv: 2006.12753, 2020.
31 Huang TW, She QQ, Wang ZQ, *et al*. GateNet: Gating-enhanced deep network for click-through rate prediction. arXiv: 2007.03519, 2020.
32 Wang JY, Chen YB, Chakraborty R, *et al*. Orthogonal convolutional neural networks. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 11502–11512.
33 Zhou J, Albatal R, Gurrin C. Applying visual user interest profiles for recommendation and personalisation. Proceedings of the 22nd International Conference on Multimedia Modeling. Miami: Springer International Publishing, 2016. 361–366.
34 Tang JX, Wang K. Personalized top-N sequential recommendation via convolutional sequence embedding. Proceedings of 11th ACM International Conference on Web Search and Data Mining. Marina: ACM, 2018. 565–573.
35 Shen XX, Yi BL, Zhang ZL, *et al*. Automatic recommendation technology for learning resources with convolutional neural network. Proceedings of 2016 International Symposium on Educational Technology (ISET). Beijing: IEEE, 2016. 30–34.
36 Gong YY, Zhang Q. Hashtag recommendation using attention-based convolutional neural network. Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York: AAAI Press, 2016. 2782–2788.
37 Zheng L, Noroozi V, Yu PS. Joint deep modeling of users and items using reviews for recommendation. Proceedings of the 10th ACM International Conference on Web Search and Data Mining. Cambridge: ACM, 2017. 425–434.
38 Liu B, Tang RM, Chen YZ, *et al*. Feature generation by convolutional neural network for click-through rate prediction. Proceedings of the World Wide Web Conference. San Francisco: ACM, 2019. 1119–1129.
39 Liu H, Lu J, Yang H, *et al*. Category-specific CNN for visual-aware CTR prediction at JD. com. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2020. 2686–2696.
40 Hidasi B, Karatzoglou A, Baltrunas L, *et al*. Session-based recommendations with recurrent neural networks. arXiv: 1511.06939, 2016.
41 Hidasi B, Karatzoglou A. Recurrent neural networks with top-k gains for session-based recommendations. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino: ACM, 2018. 843–852.

42 Devooght R, Bersini H. Collaborative filtering with recurrent neural networks. arXiv: 1608.07400, 2017.
43 Donkers T, Loepp B, Ziegler J. Sequential user-based recurrent neural network recommendations. Proceedings of the 11th ACM Conference on Recommender Systems. Como: ACM, 2017. 152–160.
44 Li Z, Zhao HK, Liu Q, *et al*. Learning from history and present: Next-item recommendation via discriminatively exploiting user behaviors. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018. 1734–1743.
45 Feng YF, Lv FY, Shen WC, *et al*. Deep session interest network for click-through rate prediction. Proceedings of the 28h International Joint Conference on Artificial Intelligence. Macao: IJCAI.org, 2019. 2301–2307.
46 Zhou GR, Zhu XQ, Song CR, *et al*. Deep interest network for click-through rate prediction. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018. 1059–1068.
47 Zhou GR, Mou N, Fan Y, *et al*. Deep interest evolution network for click-through rate prediction. arXiv: 1809.03672, 2018.
48 Zhou C, Bai JZ, Song JS, *et al*. ATRank: An attention-based user behavior modeling framework for recommendation. arXiv: 1711.06632, 2017.
49 Xiao J, Ye H, He XN, *et al*. Attentional factorization machines: Learning the weight of feature interactions via attention networks. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, 2017. 3119–3125.
50 Song WP, Shi CC, Xiao ZP, *et al*. AutoInt: Automatic feature interaction learning via self-attentive neural networks. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019. 1161–1170.
51 Sun F, Liu J, Wu J, *et al*. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019. 1441–1450.
52 Ouyang WT, Zhang XW, Li L, *et al*. Deep spatio-temporal neural networks for click-through rate prediction. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 2078–2086.
53 Chen QW, Zhao H, Li W, *et al*. Behavior sequence transformer for e-commerce recommendation in Alibaba. Proceedings of the 1st International Workshop on Deep Learning Practice for High-dimensional Sparse Data. Anchorage: ACM, 2019. 12.
54 Wang X, He XN, Feng FL, *et al*. TEM: Tree-enhanced embedding model for explainable recommendation. Proceedings of the 2018 World Wide Web Conference. Lyon, 2018. 1543–1552.
55 Zhu H, Li X, Zhang PY, *et al*. Learning tree-based deep model for recommender systems. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018. 1079–1088.
56 Su YX, Zhang R, Erfani S, *et al*. Detecting beneficial feature interactions for recommender systems. arXiv: 2008.00404, 2021.
57 Zhang K, Qian H, Cui Q, *et al*. Multi-interactive attention network for fine-grained feature learning in CTR prediction. Proceedings of the 14th ACM International Conference on Web Search and Data Mining. Virtual Event: ACM, 2021. 984–992.
58 Deng W, Pan JW, Zhou T, *et al*. DeepLight: Deep lightweight feature interactions for accelerating CTR predictions in Ad serving. Proceedings of the 14th ACM International Conference on Web Search and Data Mining. Virtual Event: ACM, 2021. 922–930.
59 Murugan P, Durairaj S. Regularization and optimization strategies in deep convolutional neural network. arXiv: 1712.04711, 2017.
60 Srivastava N, Hinton G, Krizhevsky A, *et al*. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 2014, 15(1): 1929–1958.