

对话推荐算法研究综述*

赵梦媛^{1,2,3}, 黄晓雯^{1,2,3}, 桑基韬^{1,2,3}, 于剑^{1,2,3}



¹(北京交通大学 计算机与信息技术学院, 北京 100044)

²(北京交通大学 人工智能研究院, 北京 100044)

³(交通数据分析与挖掘北京市重点实验室(北京交通大学), 北京 100044)

通信作者: 黄晓雯, E-mail: xwhuang@bjtu.edu.cn

摘要: 推荐系统是一种通过理解用户的兴趣和偏好帮助用户过滤大量无效信息并获取感兴趣的信息或者物品的信息过滤系统. 目前主流的推荐系统主要基于离线的、历史的用户数据, 不断训练和优化线下模型, 继而为在线的用户推荐物品, 这类训练方式主要存在 3 个问题: 基于稀疏且具有噪声的历史数据估计用户偏好的不可靠估计、对影响用户行为的在线上下文环境因素的忽略和默认用户清楚自身偏好的不可靠假设. 由于对话系统关注于用户的实时反馈数据, 获取用户当前交互的意图, 因此“对话推荐”通过结合对话形式与推荐任务成为解决传统推荐问题的有效手段. 对话推荐将对话系统实时交互的数据获取方式应用到推荐系统中, 采用了与传统推荐系统不同的推荐思路, 通过利用在线交互信息, 引导和捕捉用户当前的偏好兴趣, 并及时进行反馈和更新. 在过去的几年里, 越来越多的研究者开始关注对话推荐系统, 这一方面归功于自然语言处理领域中语音助手以及聊天机器人技术的广泛使用, 另一方面受益于强化学习、知识图谱等技术在推荐策略中的成熟应用. 将对话推荐系统的整体框架进行梳理, 将对话推荐算法研究所使用的数据集进行分类, 同时对评价对话推荐效果的相关指标进行讨论, 重点关注于对话推荐系统中的后台对话策略与推荐逻辑, 对近年来的对话推荐算法进行综述, 最后对话推荐领域的未来发展方向进行展望.

关键词: 对话推荐系统; 用户建模; 人机交互; 推荐系统

中图法分类号: TP18

中文引用格式: 赵梦媛, 黄晓雯, 桑基韬, 于剑. 对话推荐算法研究综述. 软件学报, 2022, 33(12): 4616-4643. <http://www.jos.org.cn/1000-9825/6521.htm>

英文引用格式: Zhao MY, Huang XW, Sang JT, Yu J. Survey on Conversational Recommendation Algorithms. Ruan Jian Xue Bao/Journal of Software, 2022, 33(12): 4616-4643 (in Chinese). <http://www.jos.org.cn/1000-9825/6521.htm>

Survey on Conversational Recommendation Algorithms

ZHAO Meng-Yuan^{1,2,3}, HUANG Xiao-Wen^{1,2,3}, SANG Ji-Tao^{1,2,3}, YU Jian^{1,2,3}

¹(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

²(Institute of Artificial Intelligence, Beijing Jiaotong University, Beijing 100044, China)

³(Beijing Key Laboratory of Traffic Data Analysis and Mining (Beijing Jiaotong University), Beijing 100044, China)

Abstract: Recommender system is an information filtering system that helps users filter a large number of invalid information to obtain information or items by estimating their interests and preferences. The mainstream traditional recommendation system mainly uses offline and historical user data to continuously train and optimize offline models, and then recommend items for online users. There are three main problems: the unreliable estimation of user preferences based on sparse and noisy historical data, the ignorance of online contextual factors that affect user behavior, and the unreliable assumption that users are aware of their preferences by default. Since the dialogue

* 基金项目: 国家重点研发计划 (2018AAA0100604); 中央高校基本科研专项 (2021RC217); 北京市自然科学基金 (JQ20023); 国家自然科学基金 (61632002, 61832004, 62036012, 61720106006)

收稿时间: 2021-04-22; 修改时间: 2021-06-28, 2021-09-22; 采用时间: 2021-10-25; jos 在线出版时间: 2021-11-24

system focuses on the user's real-time feedback data and obtains the user's current interaction intentions, "conversational recommendation" combines the interactive form of the dialogue system with the recommendation task, and becomes an effective means to solve the traditional recommendation problem. Through online interactive methods, conversational recommendation can guide and capture users' current preferences and interests, and provide timely feedback and updates. Thanks to the widespread use of voice assistants and chatbot technologies, as well as the mature application of technologies such as reinforcement learning and knowledge graphs in recommendation strategies, in the past few years, more and more researchers have paid attention to conversational recommendation systems. This survey combs the overall framework of the conversational recommendation system, classifies the datasets used in the conversational recommendation algorithm, and discusses the relevant metrics to evaluate the effect of the conversational recommendation. Focusing on the background interaction strategy and recommendation logic in conversational recommendation, this survey summarizes the existing research achievements of the domestic and foreign researchers in recent years. And finally, this survey also summarizes and prospects future works of conversational recommendation.

Key words: conversational recommendation system (CRS); user modeling; human-computer interaction; recommendation system

推荐系统是人工智能的一个重要应用领域, 能够帮助用户在没有明确需求或者环境信息过载时, 在庞大量级的数据中找到契合其偏好的信息. 一直以来, 推荐系统都是工业界和学术界的研究热点, 在各类应用场景下被广泛地讨论和研究. 推荐系统力求为用户提供精确且高效的业务信息 (包括但不限于如商品、音乐、电影、新闻、景点等). 从用户角度来看, 用户要得到推荐系统精确且高效的推荐, 就要求推荐系统能够理解用户的兴趣和偏好, 建模完整的用户画像; 从开发者角度来看, 开发者要通过推荐系统实现业务收益, 就要求推荐系统能够引导用户不断发现新的兴趣点, 以实现其用户的长期留存.

目前主流的推荐系统都是基于离线的、历史的用户数据, 不断优化线下的表现, 训练好模型后再为在线用户推荐物品. 这种离线更新的方式使得推荐系统具有内在的弱点, 即难以拟合用户线上的行为, 且用户的实时反馈无法被及时接收, 因此系统难以获取用户在使用推荐系统时当前的意图与兴趣点. 这样的推荐系统具有两个特点: 一是单向, 即系统依靠用户历史数据, 评估用户的兴趣点, 单向地向用户输出推荐物品; 二是静态, 即系统在线上使用时, 推荐模型在与用户建立的会话周期中不会及时更新和学习. 从这两个特点出发, 拓展分析, 可总结出传统单向、静态的推荐系统中存在的 3 个主要问题.

1) 首先, 推荐系统无法可靠地从历史交互数据中估计出用户的偏好^[1,2]. 历史数据通常具有噪声大、稀疏性等问题, 例如, 当前用户习惯性好评, 历史数据中该用户对所有物品均为五星好评, 推荐系统无法估计出该用户真实兴趣. 并且, 对于只有少量历史数据的新用户, 数据驱动的静态推荐系统更难以建模其偏好^[3].

2) 其次, 用户的行为是上下文敏感的, 用户的决策受各种上下文因素影响^[4,5]. 当前系统推荐的目标物品, 不仅与用户的历史数据相关, 很大程度上还取决于当前推荐的上下文环境信息, 例如时间因素: 历史数据显示用户曾经对手机感兴趣, 不代表用户会一直对手机感兴趣, 随着时间迁移和上下文环境变化, 用户兴趣会随之发生变化^[6].

3) 再者, 传统推荐系统具有内在假设, 假设用户在开始使用系统时就已经清楚自己的偏好, 推荐系统只需要根据用户历史数据猜测用户的兴趣点. 但这一假设不完全可靠, 用户可能是在决策过程中才产生自己的偏好^[7], 即当推荐系统给用户展示了一些他未知的物品选项空间, 用户才意识到自己的偏好^[8], 例如电商场景下, 用户看到华为手机出新款, 突然产生购买新款华为手机的兴趣.

即使是现在已有的面向实时推荐的在线推荐系统, 可基于用户当前的行为数据进行实时更新, 其本质上仍是传统推荐的推荐方式, 即系统被动地接受用户侧产生的数据. 随着对话系统 (dialogue system) 的发展, 对话技术 (dialogue technology) 为解决传统推荐中存在的问题提供了新的解决思路和方法, 即对话推荐 (conversational recommendation) 方式. “对话推荐”通过丰富的交互行为, 打破了静态推荐系统中系统与用户之间信息不对称的壁垒, 允许推荐系统在与用户的交互会话中动态捕捉用户偏好. 对话推荐, 一方面, 探索用户当前的兴趣偏好, 引导用户发现自己新的兴趣点, 实现用户的长期留存; 另一方面, 在交互中实时接受用户的反馈, 更新推荐模型的策略, 实现动态更新与学习. 基于对话推荐思想实现的推荐系统被称为对话推荐系统 (conversational recommendation system, CRS), CRS 是一种以推荐任务为导向去进行多轮用户交互的推荐系统. 近年有大量的研究者关注 CRS, 通过研究对话推荐, 致力于提高用户兴趣建模的效率, 用以提升在线推荐的满意度和系统中的用户长期留存率.

CRS 由于起源于对话系统, 其与用户的交互形式狭义上为自然语言类型的语音与文字, 但跳出 NLP 领域, 在更广泛的应用场景 (例如电商: 淘宝、Amazon、eBay、Airbnb 等; 多媒体平台: MovieLens、LastFM、Tiktok、YouTube 等) 下考虑, 其交互形式可以是用户在推荐系统中的一切行为, 如在电商场景下, 用户加购、关注物品等行为; 在音乐和视频内容提供商的场景下, 用户播放、快进等行为, 这些用户行为都可被定义为 CRS 与用户的交互形式. 对话推荐领域的研究工作面向不同的交互形式时, 关注点有所不同, 总体可分为两个方向, 一是关注于对用户交互形式的理解, 包括语音识别、自然语言语义理解等, 该方向与自然语言处理领域的经典研究课题紧密联系; 二是关注于对话策略的设计, 这方向的研究则涉及了更广泛的机器学习与深度学习研究领域, 包括强化学习、知识图谱、图神经网络等. 本文主要侧重于后者, 聚焦对话推荐策略方向的相关工作, 总结国内外对话推荐研究工作中对话推荐算法的特点, 为未来对话推荐实现更智能的推荐和更高效的交互提供研究思路与方向.

本文整体组织结构如下: (1) 概述对话推荐的发展历史, 介绍对话推荐的概念和特点 (第 1 节); (2) 归纳梳理对话推荐任务的通用数据集和评估标准 (第 2 节); (3) 介绍对话推荐系统框架中的 3 大基本功能模块 (第 3 节); (4) 对比分析对话推荐算法的设计思路与实现模型, 基于不同实现技术手段对现有工作进行归纳, 并介绍对话推荐算法的训练方式 (第 4 节); (5) 总结对话推荐算法的研究工作中存在的 4 大问题与挑战 (第 5 节); (6) 最后, 对对话推荐的未来研究方向进行展望 (第 6 节).

1 概述

用户使用推荐系统的过程, 从用户的角度来看, 是一个用户进行决策的过程. 传统推荐系统从用户离线数据中估计用户偏好进行推荐的实现机制, 并不直接支持用户进行在线决策, 且推荐系统处于被动地位. 对此, 对话推荐展现了不同于传统推荐的优越性, CRS 以对话交互的形式为用户提供了及时反馈的渠道, 通过对话交互去主动引导用户的决策和捕捉用户的动态偏好.

CRS 的雏形早在 20 世纪 70 年代晚期就已出现, Rich^[9]实现了一个有交互界面的图书管理系统, 该系统通过询问用户问题给用户阅读建议. 后来又有基于接口^[10]、基于表单^[11,12]的方式, 用户通过勾选目标物品属性项, 限制物品的推荐结果. 对话推荐的交互推荐形式, 使其天然具有实现用户个性化推荐的优势. 但早期的研究受限于技术水平, 没有太多的技术突破与创新. 近年以来, 一方面, 机器学习方法和深度学习方法的技术突破, 使得对话推荐策略不再依靠启发式的人工规则, 研究者从使用上下文老虎机到尝试使用简单的机器学习模型, 再到设计端到端的深度学习模型, 逐渐实现更智能的 CRS. 总之, 技术的成熟应用极大地促进了 CRS 的性能提升. 另一方面, 自然语言处理技术、语音识别技术以及图像处理技术的日渐成熟, 使得对话推荐可以适应各类应用场景下的各种交互形式, 因此, 具有灵活交互形式的 CRS 可被广泛应用到各类推荐场景中^[13-15], 极强的实用性吸引了越来越多的研究工作关注对话推荐领域的发展与革新. 本节从 CRS 的定义出发, 介绍对话推荐的特点, 为后文对话推荐算法的梳理奠定基础.

1.1 对话推荐系统的定义

对话推荐系统的交互推荐思路是将传统静态推荐系统与对话系统结合, 在与用户的交互中通过询问用户相关问题动态地更新用户的兴趣偏好, 这种方式可以总结为“system ask user respond (SAUR)”^[16]. 因此, 对话推荐系统有如下两种定义.

定义 1. 对话推荐系统是以推荐为目标导向的对话系统. 系统通过与用户的在线对话达到捕捉用户兴趣从而推荐物品的目的^[17].

定义 2. 对话推荐系统是具有多轮交互形式的推荐系统. 交互的形式狭义上为自然语言形式或语音形式, 广义上还包括推荐系统中用户的任何行为数据的形式 (如: 点击、浏览、加购、购买、评论等).

在诸多相关文献的论述中, CRS 极易与对话系统 (dialogue system, DS) 及交互推荐系统 (interactive recommender systems, IRS) 两个研究名词混淆. DS、IRS 二者与 CRS 的概念对比如表 1 所示. 以下两小节将通过说明 CRS 与其他两个概念 (DS, IRS) 的区别和联系, 进一步明确 CRS 的概念定义和研究内容, 并突出 CRS 的特点, 明确本文所要综述的对象内容是对话推荐系统, 从而避免读者产生概念上的困惑.

表1 DS、IRS 与 CRS 的概念对比

	DS	IRS
与CRS的联系	(1) 都与用户进行交互, 获取用户输入 (2) CRS可看作有推荐任务导向的DS	(1) 都提供用户交互接口, 获取用户偏好信息 (2) 都面向推荐场景, IRS可看作早期CRS研究的存在形式
与CRS的区别	(1) 交互形式不同: DS使用自然语言文字、语音, CRS还可使用用户行为数据 (2) 应用场景不完全相同: DS的应用场景更为广泛, CRS仅面向推荐任务	(1) 交互对象不完全相同: IRS仅限于交互推荐的物品, CRS还将物品相关属性作为交互对象 (2) 优化目标略有区别: IRS侧重优化序列推荐结果, CRS侧重用户兴趣建模结果

1.1.1 对话推荐系统与对话系统

对话系统大致可被分为两类: 任务导向型 (task-oriented) 对话系统和非任务导向型 (non-task-oriented) 对话系统, 后者也被称为聊天机器人. 任务导向的对话系统旨在帮助用户完成具体的任务, 例如帮助用户找寻商品, 预订酒店餐厅等, 其中一种面向推荐任务的任务导向型对话系统可被看作是以自然语言文字、语音为交互形式的CRS, 该种系统由于在推荐任务中具有很高的商业应用价值, 在近年常被作为单独的研究方向, 并被扩展到更通用的交互形式上. 在早期的诸多研究工作^[16,18,19]中, 研究者对推荐任务导向的对话系统 (recommendation-task-oriented DS) 与 CRS 的区分并不明确, 二者的研究工作常常互有交融.

1.1.2 对话推荐系统与交互推荐系统

IRS 不同于传统推荐系统将推荐看作一步预测任务 (one-step predication task) 的思路, 而是将推荐看作一个多步序列决策过程 (multi-step decision-making process)^[20-23]. 在交互过程的每一步, IRS 输出给用户一个交互物品对象, 并接收用户对于该物品的反馈, 接着继续推荐下一个物品, 直到当前用户的序列会话结束. IRS 的目标在于优化会话结束前的整个推荐序列的结果, 而 CRS 的目标有两阶段, 第 1 阶段是探索会话交互中的用户兴趣, 第 2 阶段是利用学习到的动态兴趣去提高推荐准确率^[24,25].

可以看出 CRS 与 IRS 的研究内容非常相似, 重点都在于交互过程中向用户推荐用户感兴趣的物品, 因此在很多研究者的研究工作中并不对二者进行明确区分, 常常混为一谈. 但二者其实并不完全是同一研究问题, 相较于 DS 与 CRS 的关系, IRS 与 CRS 的关系更为复杂, 所以这里将对二者的区别与联系分别进行说明.

IRS 与 CRS 的区别: 首先, IRS 的动作空间仅包含物品, 其通过物品列表与用户进行交互; 而 CRS 的动作空间有两类, 一类是物品空间, 用于直接推荐, 达到收益的目的, 另一类是物品属性空间, 用于获取用户信息, 达到探索的目的. 可以看出, IRS 的动作空间能够更契合一般的推荐应用, 即用户的偏好数据都隐含在交互过的物品列表上, 但推荐系统早期的相关研究已表明, 直接使用物品列表来表示用户偏好的效果欠佳, 原因有两方面: 一是, 询问物品属性的效率更高, 因为用户喜欢或不喜欢一个物品属性可以更显著地减少推荐候选集物品; 二是, 相较于物品属性, 用户对某个物品的正反馈更大程度上是受到当时的环境影响, 某一特定环境下产生的交互物品无法体现用户本身的偏好. 所以, CRS 通过物品属性列表来表示用户偏好的方式, 能够更好地获取用户的偏好信息. 再者, 从系统实现优化目标上来看, 二者各有不同的侧重点. IRS 的目标是优化推荐序列的总收益, 即优化用户离开会话前的所有推荐行为效果总和, 希望最大化序列推荐的收益并提高用户长期的推荐体验; 而 CRS 的目标是更快地了解用户当前的兴趣偏好, 提高推荐的效率, 因此在 CRS 的交互常常在成功推荐 1 个物品时就结束, 希望提高用户实时的推荐体验.

IRS 与 CRS 的联系: 如果将 IRS 中交互对象空间由物品泛化到物品属性, CRS 与 IRS 在实现方案与技术上基本一致, 可以认为 CRS 是一种特殊的 IRS. 但从近年发表的相关研究工作数量^[26]来看, CRS 的研究热度要高于传统的 IRS, 原因在于, 相比较只将推荐的物品作为交互对象的 IRS, CRS 的交互对象更为灵活, 可以适应更多的应用场景与需求, 给研究者更多的创新空间和思路. 因此, 在很多文献中^[10,11,27-32], 研究者将 IRS 看作是 CRS 早期的存在形式, 并更多地从 CRS 的角度进行讨论.

1.2 对话推荐任务的特点

总的来说, 从 CRS 的定义, 以及其与 DS、IRS 的对比中, 可以看出对话推荐的两大特点: 多轮交互、目标导

向. 以下两小节将分别说明对话推荐任务的这两大特点的内涵.

1.2.1 多轮交互

传统的推荐系统中, 用户在寻找具有特定属性的物品时, 可以通过主动搜索来进行, 例如, 用户可以搜索“春季的短款夹克”, 其中关键词“春季”和“短款”是“夹克”的属性. 在这个场景中, 用户自己构造查询, 推荐效果不仅依赖于搜索引擎, 更多是在依赖用户在构造查询方面的专业知识. 传统推荐系统需要用户根据自己的先验知识输入可能的属性选项限制^[33-36], 才能够准确地定位到合适的推荐候选项集合, 最终帮助用户找到目标物品. 这种主动搜索方式要求用户熟悉自己想要的每一件物品, 然而在大多数实际情况中, 用户不具备这样的先验条件. 当用户并不熟悉自己想要物品的属性时, 推荐系统被期望能够主动向用户介绍他们可能喜欢的潜在物品, 然而, 传统的推荐系统无法做到这一点, 它只能使用静态的历史记录作为输入, 从而导致了前文中提到的 3 个存在于传统单向、静态的推荐系统中的主要问题.

对话推荐的“多轮交互”特点可以弥补传统推荐系统中用户主动搜索的不足. 在 CRS 与用户的实时互动中, CRS 可以通过主动向用户提问的行为方式, 向用户展示用户未知的物品属性空间, 并利用用户的反馈信息, 直接了解用户对某些属性的需求和态度, 构建用户兴趣画像, 从而做出正确的推荐. 并且, 当用户对推荐的项目不满意时, CRS 也有机会在接下来的交互轮次中调整推荐结果.

总而言之, 由于传统推荐系统没有给用户开放提供反馈的渠道, 它在实际应用中需要阶段性地下线调整, 根据静态的历史数据被动地重新训练或更新模型, 由于庞大的数据量级和复杂的用户行为, 这个过程非常耗时耗力, 还要面临数据过时和噪声干扰的风险. 而对话推荐方式提供了交互渠道, 可以在与用户的交互中时刻微调 (fine-tune) 用户兴趣画像, 通过持续学习 (continual learning)、增量学习 (incremental learning) 不断契合用户的动态兴趣偏好, 最终实现“成功推荐”的目标.

1.2.2 目标导向

对话推荐的另一个特点是“目标导向”. 对话推荐要实现的目标任务就是给用户推荐用户感兴趣的物品, 因此, CRS 以实现“成功推荐”为最终目标, 进行获取用户偏好信息的交互. CRS 与传统推荐系统有同样的“推荐”目标, 但是, 二者在系统的运作与实现上完全不同.

(1) 从系统的推荐方式上来看, 传统的推荐系统可以看作是系统单方面向用户输出推荐物品, 即单向推荐 (one-direction recommend), 它对于用户当前的反馈没有反应, 更不会实时地更新系统的推荐策略; CRS 则注重用户的实时反馈, 不断主动地试探用户兴趣点, 并更新接下来的推荐策略, 即 CRS 使用交互推荐 (interaction recommend) 的方式.

(2) 从系统的更新方式上来说, 传统推荐系统从用户的历史数据中推断用户的兴趣点, 即使用离线更新策略 (offline-update strategy); CRS 则是在与用户的交互过程中, 将用户每轮对推荐物品的反馈作为模型更新的重要信息, 即使用在线更新策略 (online-update strategy).

2 对话推荐任务的数据集与评估

本节将根据目前对话推荐研究工作中使用的数据集的特点, 将数据集划分为两大类, 然后基于两类不同的数据集, 对对话推荐的评价指标和评估方式进行梳理和介绍.

2.1 常用数据集

正如前文所述, 对话推荐的交互形式既可以是语音、自然语言语句, 也可以是更为一般的用户交互行为数据 (如点击、浏览、购买等). 因此在对话推荐的研究工作中, 常用的数据集也分为两类: 第 1 类是研究者将用户交互语音或语句与推荐数据结合, 人工构造的数据集; 第 2 类是更为一般的推荐系统通用数据集.

两类数据集最显著的区别在于: 交互数据形式和内容不同. 第 1 类构造数据集的数据形式为语音、自然语言的语句对话, 第 2 类通用数据集则为用户与物品的交互记录 (如 0/1 点击, 评分等). 在具体场景中, 数据集类型的选择有两方面的考虑, 一方面, 在学术界, 数据集的选择取决于研究者关注的研究内容, 不关注语义理解且只关注

对话策略的研究者, 倾向于使用处理过程较为简单的第 2 类数据集; 另一方面, 在工业界, 数据集的选择则取决于应用场景, 若将 CRS 应用在语音助手上, 第 1 类数据集是必要的, 而对于数据多为用户点击行为的推荐应用平台, 第 2 类数据即可满足其需求。

对于第 1 类数据集, 不少研究者专门做了相关的构造数据集工作, 为对话推荐领域在理解和生成交互语音或语句方面的研究工作铺垫基础, 相关的开源数据集如表 2 所示。Facebook Rec 是由 Dodge 等人 (2015)^[37]发布的一个综合对话数据集, 使用经典的 MovieLens 评分数据集构建自然语言模板。ReDiaL 是 Li 等人 (2018)^[17]创建的一个包含人与人之间对话的多轮推荐对话数据集, 结合了社交聊天和推荐对话。CMU DoG 是 Zhou 等人 (2018)^[38]挖掘基于维基百科文章的两种电影领域的对话场景来构造的数据集。与 CMU DoG 同年, IIT DoG 由 Moghe 等人 (2018)^[39]构造, 同样为电影领域的数据集。Wizard-of-wiki 是 Dinan 等人 (2019)^[40]基于维基百科主题收集整理多域多回合对话推荐数据集。GoRecDial 是 Kang 等人 (2019)^[41]收集的一个目标驱动的推荐对话数据集。OpenDialKG 是 Moon 等人 (2019)^[42]贡献的数据集, 其中融合了对话推荐语料库与知识图谱 (knowledge graph, KG), 应用到具体任务上时, 主要是通过遍历知识图谱 Freebase 对对话逻辑进行建模。DuConv 是 Wu 等人 (2019)^[43]提供的中文对话数据集, 数据集对话中的一方带着明确的目标主动引导对话。KdConv 也是一个中文对话数据集, 由 Zhou 等人 (2020)^[44]整理提出, 对话数据中包含对多个主题的深入讨论。DuRecDial 是 Liu 等人 (2020)^[45]通过利用 KG 拓展对话数据构造出的数据集, 其中包含比现有数据集更多样的对话类型, 具有多域用例和丰富的交互可变性。TG ReDial 是 Zhou 等人 (2020)^[46]提出的一个电影领域的对话推荐数据集, 包含 1 万个完整的对话和近 13 万条语句, 加入了话题线索以引导用户至推荐场景, 实现语义的自然转移, 并且采用半自动的构建方式, 保留了用户真实的个性化信息 (如交互历史、偏好主题), 使得人工标注过程更加合理和可控。MGConvRex 是 Xu 等人 (2020)^[47]建立的一个记忆图 (memory graph, MG) 会话推荐平行语料库, 基于真实场景下大规模用户的 MG, 语料库包含了 7 千多条角色扮演对话, 并具有不相交的用户训练集与测试集。INSPIRED 是 Hayati 等人 (2020)^[48]提出的一个包含 1 001 条人类对话的数据集, 不同于其他数据集, 其中还包含了评估成功推荐的方式。

表 2 对话推荐数据集 (第 1 类构造数据集)

数据集	对话数量	内容领域
Facebook Rec ^[37]	1M	电影
ReDiaL ^[17]	10k	电影
CMU DoG ^[38]	4k	电影
IIT DoG ^[39]	9k	电影
Wizard-of-wiki ^[40]	22k	Wikipedia主题
GoRecDial ^[41]	9k	电影
OpenDialKG ^[42]	12k	电影、书籍、体育、音乐
DuConv ^[43]	29k	电影
KdConv ^[44]	4.5k	电影、音乐、旅游
DuRecDial ^[45]	10.2k	电影、音乐人、食物、餐厅、新闻、天气
TG ReDial ^[46]	10k	电影等多类主题
MGConvRex ^[47]	8k	餐厅
INSPIRED ^[48]	1k	电影

对话推荐研究使用的第 2 类数据集为常见的传统推荐数据集, 近年来越来越多的对话推荐研究工作使用第 2 类更为通用的推荐数据集, 表 3 列举了部分第 2 类通用的推荐数据集。MovieLens 是由明尼苏达大学的 GroupLens 实验室于 1997 年从 MovieLens 电影网站上收集并公开, 包含用户评分、电影属性等; LastFM 是 2011 年 HetRec 会议发布的音乐推荐数据集, 来源于英国的音乐社 Last.fm, 包含用户的听歌记录、用户属性等; Jester 是一个用户对笑话的评分数据集, 由加州大学伯克利分校的 Ken Goldberg 等人于 2001 年发布; Book-Crossings

是 Cai-Nicolas Ziegler 等人从图书网站 bookcrossing.com 上收集整理图书评分公开数据集; Wikipedia 是一个基于维基百科的大型数据集, 该数据集并非为推荐任务专门设计, 在推荐任务中使用时需要额外的内容提取工作; OpenStreetMap 是一个类似于 Wikipedia 的公开地理数据集, 包含地理位置、POI 等数据; Flixster 电影推荐数据集来源于美国社交电影网站 flixster.com, 包含电影票房信息、用户间的朋友关系等数据; CiaoDVD 的数据为 2013 年从英国 DVD 网站 dvd.ciao.co.uk 上收集的 DVD 类别数据; Yelp 是由美国点评网站 Yelp 公开的内部数据集, 现已成为通用数据集, 包含用户点评数据、商家属性等; AmazonReviews 由康奈尔大学于 2013 年发布, 包含 Amazon 购物网站上各品类产品的用户购买、评分等行为数据; Douban 是来源于豆瓣网站的一个匿名的电影数据集, 包含用户评分、电影属性等数据; Epinions 是于 2011 年 5 月发布的商品评分数据集, 除用户的评分数据外, 还包含用户的社交关系。

表 3 对话推荐数据集 (第 2 类通用的推荐数据集)

数据集	网址	数量	内容领域
MovieLens	https://grouplens.org/datasets/movielens/	100k, 1M, 2M, 20M	电影
LastFM	https://grouplens.org/datasets/hetrec-2011/	2k	音乐
Jester	http://eigentaste.berkeley.edu/dataset/	6M	笑话
Book-Crossings	http://www2.informatik.uni-freiburg.de/~cziegler/BX/	1M	书籍
Wikipedia	https://en.wikipedia.org/wiki/Wikipedia:Database_download#English-language_Wikipedia	418M	Wikipedia主题
OpenStreetMap	http://planet.openstreetmap.org/planet/full-history/	206k	地理
Flixster	http://socialcomputing.asu.edu/datasets/Flixster	8M	电影
CiaoDVD	https://www.librec.net/datasets.html	278k	碟片
Yelp	https://www.yelp.com/dataset	4.7M	商户
AmazonReviews	http://jmcauley.ucsd.edu/data/amazon/	143M	商品
Douban	http://socialcomputing.asu.edu/datasets/Douban	400k	电影
Epinions	http://www.trustlet.org/epinions.html	510k, 665k, 13.7M	商品

2.2 评价指标

在两类不同类型数据集的应用场景下, 对话推荐的评价指标也不相同, 目前尚没有统一且成熟的评价指标体系来评价对话推荐系统的功能. 如表 4 所示, 在对话推荐相关的研究工作中, 研究者由于研究的出发点和关注的角度不同, 往往在实验中使用不同的评价指标, 从其关注的某几个方面评价对话推荐的效果. 例如, 在基于人类对话语料的对话推荐研究工作中, 研究者关注端到端的自然语言对话功能^[17,45,49,50]和深度语言模型^[46], 使用 BLEU、ROUGE、METEOR 等得分来评价系统交互中的反馈是否准确和流畅^[51,52]; 另一些研究工作中, 研究者关注对话推荐中交互策略模型的设计, 则使用推荐成功率 (success rate, SR) 来评价推荐的准确率^[19,53], 以及使用平均对话轮次 (average turns, AT) 来评价对话推荐的系统交互效率^[19,45,53].

表 4 对话推荐研究中常用的评价指标

评价指标	评价角度	评价内容	常用场景
BLEU, ROUGE, METEOR等	对话推荐的自然语言生成角度	生成的自然语言质量	基于第1类构造数据集的CRS
MSE, RMSE等	对话推荐的推荐性能角度	推荐的物品得分情况	基于第2类通用推荐数据集的CRS
Hits, Precision, Recall, F1-score, MRR, MAP, NDCG等	对话推荐的推荐性能角度	推荐的物品排序情况	基于第2类通用推荐数据集的CRS
SR@T	对话推荐的系统效率角度	在第T轮时, 系统的推荐成功率SR	基于第2类通用推荐数据集的CRS, 少数基于第1类构造数据集的CRS
AT	对话推荐的系统效率角度	系统在线实现一次成功推荐的平均交互轮次AT	基于第2类通用推荐数据集的CRS, 少数基于第1类构造数据集的CRS

本节根据各类指标的评估内容和角度,把目前对话推荐任务的相关评价指标分为3类:(1)评价对话推荐的语言质量的指标;(2)评价对话推荐的推荐性能的指标;(3)评价对话推荐的系统效率的指标。

2.2.1 评价语言质量的指标

使用第1类数据集的对话推荐研究工作中,CRS通过自然语言形式与用户进行交互,因此评价生成的交互语句的质量是非常必要的。研究者常常使用文本生成领域(text generation)的评价指标来评价对话推荐中语言模型生成的交互语句的质量,例如,使用BLEU(bilingual evaluation understudy)^[54]衡量CRS生成的词或n-gram词与ground-truth词的精确率(precision),即CRS生成的词语中有多少词出现在ground-truth参考话语中;使用ROUGE(recall-oriented understudy for gisting evaluation)^[55]衡量召回率(recall),即在ground-truth的语句中有多少单词或n-gram出现在CRS生成的话语中。

另外,由于CRS生成交互语句的最终目标是为了在交互过程中获取用户偏好信息,因此生成的语句被期望能够包含有效的用户偏好词^[17,56,57],所以仅使用BLUE、ROUGE等指标来评价生成语句的质量,是与对话推荐的最终目标有偏差的。但目前的研究工作对于这方面的考量和研究没有非常深入和全面,相关评价指标值得研究者未来更多的关注。

2.2.2 评价推荐性能的指标

对话推荐的推荐表现,多使用传统推荐领域的评价指标来评价,评价指标有两类:一是基于用户评分^[58]的评价指标,如MSE(mean squared error)和RMSE(root mean squared error)等,通过量化推荐物品的模型得分与用户真实评分之间的差值来评价,差异越小,说明推荐模型表现越好;二是基于排序顺序^[59]的评价指标,如Hits, Precision, Recall, F1-score, MRR(mean reciprocal rank), MAP(mean average precision)和NDCG(normalized discounted cumulative gain)等,用户评分高或者有正反馈的物品在推荐列表中的排序越靠前,说明推荐模型表现越好。

在评价对话推荐中推荐模型的表现时,存在的最大问题是,这些推荐系统的常用评价指标都是将系统的推荐结果与用户已有的行为数据对比,但用户的行为数据相对于物品空间是稀疏的,用户未被标记某物品为正样本,不代表用户对其不感兴趣,可能是该物品受制于资源或曝光机会而未被用户接触^[60,61]。因此,在对话推荐中使用这类推荐评价指标可能会造成CRS即使成功推荐出用户当前最感兴趣的物品,会因为该物品没有出现在用户的历史数据中,而被评价表现不佳。总而言之,使用传统推荐中的评价指标来评价对话推荐会将评价偏差(evaluation bias)^[60,62]引入到CRS中。对话推荐不同于传统静态推荐的最重要特点就是实时交互,即根据用户的在线反馈获取当前用户偏好,这个特点使其本身天然具有避免偏差的潜力。近年,研究者也逐渐地意识到,仅使用这类通用的推荐评价指标来评价对话推荐的推荐效果是片面的,并尝试提出可以评价CRS的整体推荐效果与交互效果的评价指标,接下来的小节将对这些指标进行介绍。

2.2.3 评价系统效率的指标

CRS的整体功能表现体现在两个方面,一是推荐的准确程度,二是用户偏好获取的速度。很显然,在一定的交互资源限制下,同样的交互轮次,实现成功推荐的概率越高,说明CRS的推荐功能更强大,是CRS高效性的体现之一。因此,研究者会使用固定交互轮次 T 下的推荐成功率 $SR@T$ 来整体评价对话推荐的系统效率^[19,53]。另一方面,在实现准确推荐的前提下,与用户的交互轮次越少,给用户带来的负担就越小,说明系统获取用户偏好的速度就越快,是CRS高效性的体现之二。因此,研究者使用平均对话轮次 AT 来评价系统的交互效率^[19,45,53],实现一次成功推荐使用的平均对话轮次越少,说明CRS的交互效率越高。总体看,目前可以比较全面的评价CRS的评价指标比较少,未来有待研究者们提出更多合理且全面的对话推荐评价指标。

2.3 评估方式

除了选择合适的评价指标对对话推荐的效果进行评价外,基于“多轮交互”的对话推荐还需要选择合适的评估方式,目前对话推荐的评估方式主要分为在线用户交互的评估方式和用户模拟器的评估方式。

2.3.1 在线用户评估

在线用户评估方式是让CRS上线与真实的用户交互,在真实的环境中验证系统的效果,评估的结果真实

且可信. 虽然在线用户评估的结果很优越, 但实际的研究工作大多没有条件进行线上用户评估, 在线用户的评估方式很难被广泛使用. 原因有二: 一是使用在线用户进行训练和评估, 不成熟的 CRS 会损害用户体验, 造成实际的损失; 二是在线用户评估数据的收集和整理需要额外的时间和精力, 投入的成本加大. 因此, 在目前的对话推荐相关研究工作中, 更好的方式是构建用户模拟器代替真实的用户, 利用模拟器模拟真实用户和 CRS 的交互过程.

2.3.2 用户模拟器评估

用户模拟器 (user simulator) 通过拟合用户的历史数据, 提供对话推荐的交互环境, 为 CRS 每轮交互的输出动作反馈用户可能的行为. 使用用户模拟器, 使得 CRS 的交互训练更加快捷, 也极大降低了对话推荐模型的评估成本. 但该方式也有明显的缺点, 即用户模拟器无法完全正确地预测在线用户的真实反馈. 此外, 使用模拟器代表真实用户与 CRS 交互时, 可能会将模拟器自身的误差传递给 CRS, 造成 CRS 虚假有效, 当 CRS 真正上线交互真实用户时, 推荐效率和准确率不能够达到预期. 因此, 如何衡量用户模拟器与真实用户环境的误差, 并对此提出合理的解决方法, 也是未来对话推荐研究工作的挑战.

3 对话推荐系统的基本功能流程

本节将对 CRS 的功能和处理流程进行整体介绍. 如图 1 所示, 一个标准的 CRS 框架包含了 3 大功能模块: 用户意图理解模块、对话策略模块和推荐模块. 从功能模块来看, CRS 与 DS 非常相似, 这进一步说明 CRS 的核心思路来源于 DS.

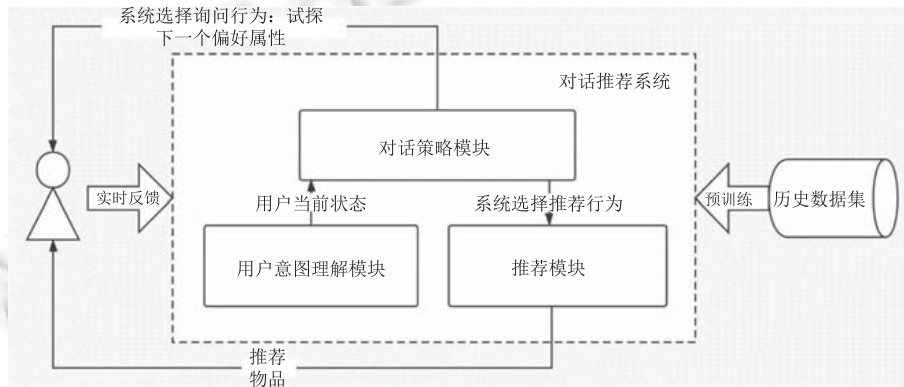


图 1 对话推荐系统的基本功能框架图

CRS 的 3 个模块有各自的任务, 模块间通过合作实现 CRS 的“多轮交互”和“推荐任务”. 用户意图理解模块是用户与系统交互的“翻译器”, 将用户的输入 (文本、图像、评分、浏览等交互行为) 处理成会话的状态表示, 作为对话策略模块的输入, 如果是基于自然语言交互的对话推荐, 该模块还需要生成流畅的自然语言反馈. 对话策略模块是 CRS 的核心模块, 也是用户意图模块和推荐模块的连接桥梁, 该模块处理和记录用户偏好, 并根据用户当前的交互历史数据, 决定系统当前轮次反馈给用户的交互内容, 引导用户决策, 最终实现符合用户偏好的一次成功推荐. 推荐模块负责建模用户与物品的关系, 在被对话策略模块调用时输出当前轮次下契合该用户偏好的推荐物品列表.

接下来 3 小节, 分别从这 3 个功能模块梳理当前研究工作.

3.1 用户意图理解模块

用户意图理解模块作为与用户直接进行信息交换的模块, 处理用户数据进行编码表示, 输入到 CRS 中. 用户意图理解模块的输入数据, 传统的形式是用户的对话文本, 在早年的研究工作中, 这一模块注重对自然语言的处理和理解. 近几年随着 CRS 的应用推广, 用户意图理解模块的输入可以是多模态数据 (multimodal data) 和用户行为数据^[51,63]. 这里将用户意图理解模块的研究工作分为两类: 一是对自然语言、语音和图像数据的处理, 二是对广泛

用户行为数据的处理.

3.1.1 面向自然语言、语音和图像的用户意图理解

基于第 1 类数据集的对话推荐的研究难题之一是如何像人类一样沟通交互, 这个问题的研究落在用户意图理解模块. 基于第 1 类数据集的 CRS 的用户意图理解模块的主要功能是处理自然语言、语音和图像, 其中包括理解输入数据和生成输出数据两部分.

(1) 理解输入的自然语言、语音和图像

为了解自然语言、语音和图像中包含的用户偏好, 简单的做法是设计一些预定义的输入标签直接获取输入数据的用户偏好语义信息^[53,64-66]. 复杂一些的研究工作使用深度学习的方法来处理用户输入的文本、语音和图像数据, 例如, 先通过循环神经网络 (recurrent neural network, RNN)^[17,51,52,67]或卷积神经网络 (convolutional neural network, CNN)^[51,63]的方式将当前轮次的会话进行编码, 编码成对话状态向量, 然后模块将获得的用户偏好表示向量填补人工事先定义好的用户兴趣槽 (槽值对表示用户当前兴趣偏好)^[16,19,68], 或者将用户偏好记录到记忆网络 (memory network) 中^[16,51].

其中基于 RNN 的处理方式, 主要面向文本和语音的数据形式. 这类方式最流行的模型主要有 3 种: 层次递归编码器解码器 (hierarchical recurrent encoder decoder, HRED)^[17,51,67]、具有注意力机制的记忆网络^[16,51]和 GRU 编码器^[52].

基于 CNN 的处理方式, 主要面向图像数据. 研究工作中使用的 CNN 模型有 ResNet^[63]、EI-tree^[51]等, 通过 CNN 提取图像的语义特征和语音等其他特征拼接起来, 对对话状态进行编码.

(2) 生成输出的自然语言、语音和图像

由于交互形式是自然语言, CRS 在该模块除了理解自然语言输入, 还被要求生成人类可理解的、流畅自然的文本或语音反馈数据^[51], 引导用户下一轮交互. 为了获取用户的偏好信息, 生成的语句除了有流畅、合理等文本生成的基本要求, 还被要求能够包含尽可能多的用户偏好引导信息. 有的研究工作实现该目标的方式比较简单, 例如, 生成推荐列表^[16,66]后, 按照既定规则将推荐的物品信息直接融合到自然语言模板中^[19,53,65]. 此外, 一些研究者提出了更为复杂但灵活的端到端的框架, 使 CRS 能够从原始的自然语言中准确地理解用户的情感和意图, 并生成可读的、流畅的、一致的、有意义的自然语言响应^[17,45,49,50,68].

3.1.2 面向用户行为数据的用户意图理解

基于第 2 类用户行为数据集的对话推荐中, 用户的兴趣偏好由推荐物品的属性组成, 相关研究工作通常直接将用户行为数据处理成用户当前的属性偏好集合^[53,64], 即用户属性偏好集合为物品的属性集合的子集. 基于这类数据的 CRS 中, 用户意图理解模块的任务是简单且明确的. 在处理输入数据方面, 常见的反馈为二值反馈 (binary response), 即接受或者拒绝, 该模块只需要将正负信号传递给其他模块即可; 在输出数据方面, 当 CRS 的对话策略选择下一轮继续获取用户偏好信息时, 该模块输出策略关注的偏好属性值, 而当 CRS 的对话策略选择向用户推荐物品时, 该模块输出推荐的物品.

3.2 对话策略模块

推荐系统存在的一个问题是每个用户只能与整个数据集中的少数项目进行交互, 有正反馈的数据是非常稀疏的, 特别是对于没有或只有少量的交互数据的冷启动用户 (cold-start users), 这个问题格外严重. 除历史数据稀疏问题外, 推荐系统由于拥有庞大的候选物品空间, 大量的用户可能感兴趣的物品难有机会被用户看到. 面对推荐系统的这些问题, CRS 的“多轮交互”特点使得对话推荐相比于传统推荐具有明显的优越性. 对话推荐的交互性使得用户有机会及时表达自己的意图, 以获得更好的个性化推荐服务, 并且由于具有交互性, 对话推荐还可以主动探索用户看不见的物品空间, 从而引导用户发掘自己的偏好. 然而对话推荐的探索过程是有代价的, 与利用已经获得的偏好推荐高置信度的物品相比, 暴露不相关的物品空间的信息可能会伤害用户的偏好, 失去做出准确推荐的机会. 此外, 由于用户与系统互动的时间和精力有限, 一次失败的探索将浪费用户的时间与精力, 可能造成用户的流失. 因此, 追求探索和收益的平衡 (exploration and exploitation, E&E) 是对话推荐研究中的一个关键问题, 对话策略模块

主要解决的问题就是权衡探索与收益,它是 CRS 的关键模块。

对话推荐的 E&E 问题表现为,CRS 在交互过程中需要知道当前是需要继续询问用户,从用户处获取更多信息,还是已经获取到足够的用户偏好信息,到达推荐目标物品的时机。对话策略模块的策略设计,非常重要且困难,是一个典型的博弈问题:首先,过多的询问确实可以掌握用户更多的偏好信息,提高推荐的成功率,但同时也会给用户造成较大的负担,这与推荐系统减轻用户搜索信息负担的目标初衷相悖,导致用户在系统决定采取推荐行动前就离开;但是,过少的询问,会造成用户偏好信息的匮乏,因而系统推荐的准确率过低,这又与推荐系统提高推荐物品准确率的目标相悖。因此,好的对话策略需要智能地平衡对话轮次与推荐准确率两个指标。

该模块的研究工作根据目前研究者所使用的核心策略技术,主要分成以下几类:基于人工规则的、基于协同过滤的、基于神经网络的、基于强化学习的、基于元学习的。基于人工规则的对话策略模块每轮的运作非常简单,根据人的先验领域知识,设置对话规则,例如,设置限制,规定固定轮次交互后都要提供一个推荐物品^[64],或者根据置信度推荐前 k 个物品^[16]。基于神经网络的对话策略模块,利用神经网络模型输出的动作概率分布决定对话策略^[17,51],例如,通过 GRU 将当前对话的上下文编码成隐状态向量输入,Softmax 层输出概率分布^[69],根据概率分布决定是否推荐以及推荐什么物品。基于强化学习的对话策略模块,将对话上下文、用户偏好等编码为状态向量,输入到强化策略网络中,策略网络基于奖励函数进行训练,学习对话策略,输出系统动作向量^[19,52,53]。在第 4.1 节中,本文将对基于不同模型机制的对话推荐算法进行细致梳理。

3.3 推荐模块

推荐模块是 CRS 中实现推荐功能的模块,根据已经捕捉到的用户偏好推荐用户当前最感兴趣的目标物品。大部分的研究工作中,推荐模块都采用简单的推荐模型,例如矩阵分解(matrix factorization)^[19,53,69],这是因为简单的推荐模型已经能够满足对话推荐系统的推荐需求,使用过于复杂的推荐模型,反而会使系统整体复杂度上升,使对话推荐模型的训练变得困难。

但也有少部分工作中尝试采用复杂的深度神经网络模型,主要有两大类。一类是基于自编码器(autoencoder)的模型,自编码器本身就常被用于推荐系统中,可以克服冷启动问题(cold-start problem)^[70],例如, Li 等人^[17]的工作中尝试使用了自编码器网络来训练 CRS 的推荐模块,预测在训练集中未观察到的用户对物品的评分。另一类是基于图卷积神经网络(graph convolutional networks, GCN)的模型,图卷积神经网络面向应用场景中的图结构数据,如社交网络或推荐项目关系^[71],通过卷积操作,同时考虑节点的属性以及邻居节点的属性,让模型学习到节点更好的局部表示,从而提高基于图结构数据的模型性能^[72],例如, Liao 等人^[67]在构建旅游内容的 CRS 时,将酒店、餐厅和其他旅游设施连接起来构建图结构数据,使用 GCN 模型编码旅游设施之间的结构关系,为旅行的用户提供更好的服务; Chen 等人^[49]引入知识图谱中的用户偏好信息到 CRS 框架中,通过 GCN 端到端的学习来提高推荐性能; Zhou 等人^[50]也引入了外部的知识图谱,通过 GCN 得到实体表征和词表征,从而利用互信息解决语义鸿沟问题,提高 CRS 推荐准确率。

4 对话推荐算法

对话策略模块是 CRS 中最核心的模块,如第 3.2 节所总结的,对话推荐的最大挑战就是平衡探索与收益的 E&E 问题。研究对话推荐算法主要研究内容为设计算法去输出最佳交互策略,以快速减小用户兴趣的不确定性,并实现对于当前用户的“成功推荐”。研究内容体现出了对话推荐的两大特点(“多轮交互”与“目标导向”),对话推荐算法的设计是实现高质量 CRS 的关键点。本章将详细阐述对话推荐算法的设计工作,总结目前的研究工作的思路和技术。

4.1 对话推荐算法的设计

对话推荐算法的研究工作,如表 5 所列举,每个研究工作在模型的设计和技术的使用上都各有特点。本节分为 5 小节,分类梳理。

表 5 对话推荐算法的研究工作总结

时间(年)	代表性研究工作	对话策略模块类型	技术手段	系统ASK动作空间	数据集的类型		交互轮次	
					第1类构造数据集	第1类通用推荐数据集	单轮	多轮
2012	Critique ^[27]	基于人工规则	人机交互界面	属性	√	—	—	√
2013	Interactive CF ^[25]	基于人工规则和协同过滤	协同过滤	物品	—	√	—	√
2016	Abs Pos ^[13]	基于人工规则	EE算法	属性	√	—	—	√
2017	CEI ^[52]	基于强化学习	RNN, 强化学习	对话文本	—	√	—	√
2018	Q&R ^[64]	基于强化学习	RNN	对话文本	√	—	√	—
2018	CRM ^[19]	基于强化学习	LSTM, 强化学习	属性	—	√	√	—
2018	SAUR ^[16]	基于人工规则	记忆网络	属性	—	√	—	√
2018	DeepCR ^[17]	基于神经网络	HRED, 自动编码器	对话文本	√	—	—	√
2018	KMD ^[51]	基于强化学习	HRED, EITree, 强化学习	对话文本	√	—	—	√
2019	DCR ^[72]	基于神经网络	HRED, GCN	对话文本	√	—	√	—
2019	VisualDialog ^[67]	基于人工规则	RNN, CNN	对话文本	—	√	—	√
2020	EAR ^[53]	基于强化学习	强化学习	属性	—	√	—	√
2020	KGSF ^[50]	基于神经网络	GNN, MIM	对话文本	√	—	—	√
2020	SCPR ^[65]	基于强化学习	GNN, 强化学习	属性	—	√	—	√
2020	KGQR ^[22]	基于强化学习	GNN, 强化学习	属性	—	√	—	√
2020	UMGR ^[47]	基于神经网络	图推理	对话文本	√	—	—	√
2020	NICF ^[73]	基于元学习和强化学习	元学习, 强化学习, 协同过滤	物品	—	√	—	√
2020	MeLU ^[3]	基于元学习	元学习	物品	—	√	√	—
2021	FPAN ^[74]	基于强化学习	Gating Mechanism, 强化学习	属性	—	√	—	√

4.1.1 基于人工规则的对话推荐算法

对话推荐的早期研究中, 研究者试图通过设计直接的人工逻辑和规则减小用户兴趣画像的不确定性和推荐的物品候选集大小. 这种基于人工规则的对话推荐算法, 在早期的文献中被称为 Critiquing-based 方法. Critiquing-based 方法是基于启发式策略获取用户对属性的偏好信息^[27-29,75], 启发式策略主要形式为接受用户对物品属性值的要求来限制候选集, 例如, “16 GB”的内存、“13.3 英寸”的屏幕大小等, 通过删除属性不满足的物品来重建候选集^[11,27,32,76-78]. Critiquing-based 方法的优缺点如表 6 所总结, 早期的 Unit Critiques 方法仅在一个特征属性上获取用户偏好去约束推荐结果, 后期研究者为了克服单个 Unit Critiques 方法无法获取用户组合偏好的缺点, 提出了组合多个 Unit Critiques 的 Compound Critiques 方法, 去获取用户多个物品属性的偏好.

表 6 基于人工规则的对话推荐算法的优缺点

方法	优点	缺点
Unit Critiques	可显式获取用户某一属性偏好对推荐结果的影响	无法获取用户的组合偏好
Compound Critiques	可显式获取用户的多个属性偏好对推荐结果的影响	特征组合规则设计复杂

4.1.2 基于协同过滤的对话推荐算法

协同过滤 (collaborative filtering, CF) 是推荐系统的重要算法, 与早期推荐系统直接基于内容过滤进行推荐不同, 协同过滤分析用户兴趣, 在用户群中找到与指定用户具有相似兴趣的用户群, 综合这些相似用户对某一物品信息的评分, 推测该用户对此物品的喜好程度^[79-82]. 在实际应用中, 协同过滤算法虽然简单, 但十分有效, 著名的商业应用案例有 Amazon、CDNow、MovieFinder 等. 将协同过滤算法思想引入到对话推荐算法中, 一方面简单却有效

的 CF 可以帮助对话推荐提高推荐的准确率,另一方面,对话推荐的“多轮交互”特征可以有效弥补 CF 的冷启动、评分稀疏、物品空间过大等不足。

由于协同过滤与对话推荐性能互补,不少研究者研究基于协同过滤的对话推荐算法,例如,Zhao 等人^[25]研究了交互式协同过滤问题,提出了扩展概率矩阵分解(PMF)方法^[83]的交互式 CF 框架,在该框架中,利用 PMF 模型来捕获用户和物品特征向量的分布,在此基础上使用线性 bandits 算法(Thompson 采样和 UCB-based 算法)选择需要反馈的项目,从而平衡交互式 CF 模型的 E&E 问题,最后将评分合并回 PMF 输出;Li 等人^[17]设计的 CRS 根据与新用户交互中学习到的新用户偏好,将新用户与具有相似偏好的老用户相关联,使用基于用户的协同过滤,从而减轻冷启动用户的负担.基于协同过滤的对话推荐算法的优缺点如表 7 所总结,CF 可通过建模历史数据,解决新用户的冷启动问题,在短期内实现对话推荐的收益最大化,但无法仅依赖当前有限的交互数据去建模对话推荐的长期收益。

表 7 基于协同过滤的对话推荐算法的优缺点

方法	优点	缺点
CF+bandits	一定程度上平衡短期的E&E问题	无法最大化长期的E&E收益
User-based CF	可解决用户冷启动问题	无法应对物品冷启动问题

4.1.3 基于神经网络的对话推荐算法

前期的对话推荐研究中,基于人工规则和协同过滤的对话推荐算法虽然具有很好的解释性,但无法全面地交叉组合用户的偏好表示和挖掘用户的偏好空间.随着后期神经网络的兴起与广泛应用,使用神经网络将用户、物品、评分表示为隐向量,为对话推荐策略生成推荐物品和询问属性打开了新的思路.研究者开始研究和设计基于神经网络的对话推荐算法,开始尝试在原来基于人工规则和协同过滤的对话推荐算法中加入神经网络的向量表示.Wu 等人^[75]提出了一种 Critiquing-based 的可解释神经协同过滤模型(CE-NCF),使用神经协同过滤模型(neural collaborative filtering, NCF)^[84]编码用户的偏好,NCF 是一个端到端的深度学习框架,框架主要包含 Explanation 模块和 Critiquing 模块,其中 Explanation 模块将用户对物品的偏好表示为一个潜在的向量,然后用该隐向量去计算用户对物品的评分以及生成用户对属性的偏好向量,偏好向量各维度由关键属性值组成,当用户在实时反馈中表示不喜欢某个属性时,NCF 框架的 Critiquing 模块通过将相应的维数设置为零来更新偏好向量,从而更好地适应用户当前的偏好.之后,Luo 等人^[29]为了解决基于概率生成 CE-VNCF 模型训练不稳定且低效的问题,引入生成模型变分自编码器(variational auto-encoder, VAE)辅助原模型,使得 Critiquing-based 对话推荐算法能够更快收敛、更有效率地计算、更加稳定。

在近期的对话推荐研究中,神经网络已经被普遍使用.除了最简单的多层感知机网络(multi-layer perceptron, MLP),其他网络模型也被引入到 CRS 算法设计中,其中 RNN 和 CNN 常被用来建模 CRS 中会话的序列数据^[52,56,64];图神经网络(graph neural network, GNN)则常被用来建模 CRS 中的图类型数据^[22,50,56,65,67].总之,如表 8 所示,基于神经网络的对话推荐算法的优缺点与神经网络本身特点紧紧关联,优点点在于算法对复杂数据的强大建模能力和特征组合能力;而缺点在于算法的高复杂度和不可解释性。

表 8 基于神经网络的对话推荐算法的优缺点

方法	优点	缺点
MLP	可学习特征组合,并将稀疏特征映射为稠密特征	
RNN/CNN	可建模具有时序特征的会话数据	模型复杂度相对较高,缺乏解释性
GNN	可建模图数据,学习交互数据间复杂的转移关系	

4.1.4 基于强化学习的对话推荐算法

实现高效的对话推荐的一个关键难题就是 E&E 问题,对话推荐可以利用系统已经获取的用户偏好信息,输出

当前的最佳推荐选项, 这是收益 (exploitation); 也可以冒着用户交互负担过重导致满意率下降的风险, 选择与用户继续交互关于用户偏好信息的未知选项, 这是探索 (exploration). 传统推荐系统在处理只有很少或没有历史数据的冷启动用户时, 面临的也是 E&E 难题, 研究者常常使用多臂老虎机算法 (multi armed bandit, MAB) 来解决推荐系统的冷启动问题. 面对话推荐中同样的 E&E 问题, 研究者类比传统推荐, 将 MAB 算法应用到对话推荐中^[21,85,86], 使用如 GLM-UCB 和 Thompson 采样^[85,87]等探索策略尝试达到对话推荐中合理的 E&E 平衡. 为了实现长期收益最大化, CRS 可能会做出短期牺牲, 在 E&E 的早期阶段, 进行的探索可能会失败, 但失败的经验也警示了模型在未来不要太频繁地采取这种探索行动. 从整体效果来看, 解决 E&E 平衡问题不仅可以改善推荐中的冷启动场景, 同时可以提高推荐系统中面向任何用户 (包括了冷启动用户和非冷启动用户) 的推荐性能.

基于 MAB 的对话推荐算法中, 用户偏好通常由一个线性函数来建模, 该函数通过适当的探索与收益折中来不断学习. 然而, 这些基于 MAB 的对话推荐算法预先假设在推荐过程中用户偏好保持不变, 为了达到可证明的下限, 算法在最坏的情况下优化对话推荐的策略, 因而倾向于过度悲观的策略, 无法实现 CRS 最优的整体性能. 由于基于 MBA 的模型未对用户偏好的动态变化进行建模^[25], 而对话推荐的主要特点为获取用户偏好可能的动态变化, 因此 MAB 算法无法满足对话推荐的要求. 此外, 基于 MAB 的对话推荐算法难以计算非线性的模型, 这极大地限制了神经网络模型在 CRS 中的应用^[84,88].

近年来, 强化学习 (reinforcement learning, RL) 因其对复杂环境优秀的建模潜力而被许多交互式推荐模型所采用, 且应用了神经网络模型的深度强化学习 (deep reinforcement learning, DRL) 更展现出在动态环境下进行决策和长期规划的巨大潜力, 一些研究者开始将 DRL 算法引入对话推荐算法设计中^[23,24,89-91].

RL 中有模型无关的强化学习方法 (model-free RL) 和基于模型的强化学习方法 (model-based RL). 模型无关的 RL 如图 2(a) 所示, 策略网络直接从与环境交互的数据中学习策略, 而基于模型的 RL 如图 2(b) 所示, 常被应用在有显式规则的世界模型 (world model) 的场景中, 如国际象棋、电子游戏等, 两类 RL 框架下的常用算法如图 3 所列举. 在对话推荐领域的研究工作中, Mahmood 等人^[92]首先提出在交互式推荐中使用基于模型的 RL, 使用了动态规划进行策略的迭代. 但基于模型 RL 所依赖的显式规则在推荐任务中并不存在, 即推荐系统中无固定的推荐规则使每种推荐行为有固定的收益, 于是后来又有更多的研究工作使用如 DQN (deep q-network)^[24]、DDPG (deep deterministic policy gradient)^[89]等模型无关的 RL 框架来处理对话推荐任务, 其中具有代表性的研究工作是 Sun 等人^[19]提出的一种会话推荐模型 (CRM) 和后续 Lei 等人^[53]提出的 EAR 框架. CRM 使用了面向任务的对话系统的体系结构, 利用对话状态跟踪潜在向量空间, 潜在向量表示对话的当前状态和用户迄今为止的偏好, 深度策略网络根据输入的潜在向量决定是推荐一个商品还是继续询问用户偏好, 并通过策略梯度法调整对话策略, 更新潜在向量空间, 这些更新类似于面向任务的 DS 都是短期的; EAR 框架则提出使用交叉熵信息、历史反馈信息、对话轮次信息等构造状态向量, 作为策略网络的输入, 使得策略模型能够更好地适应动态环境, 学习到更加灵活的对话策略.

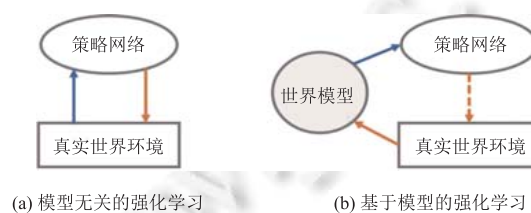


图 2 基于模型与模型无关的强化学习示意图

由于强化模型对于环境信息的依赖, 在对话推荐中合理引入更多的上下文信息, 能够帮助强化策略网络模型学习到更加智能、有效的策略. 这样的研究思路下, 知识图谱 (knowledge graph, KG) 作为可以承载物品、用户、物品属性之间结构关系的数据形式, 自然而然可以为学习对话推荐策略提供帮助. 此外, 传统推荐算法研究的发展路线中, 基于知识图谱的推荐系统是极具潜力的研究方向, 对于同是面向“推荐”任务的对话推荐, 引入知识图谱到

对话推荐算法框架中对对话推荐的研究也具有重大意义. 知识图谱的引入, 一方面使得 DRL 中不可解释的潜在向量具有解释性; 另一方面, 丰富了交互中的上下文信息, 提高了模型的置信度. 根据图谱在强化对话推荐算法中的应用方式, 这里将基于强化学习的对话推荐算法的相关研究工作分为基于知识增强的强化对话推荐算法和基于图推理的强化对话推荐算法.

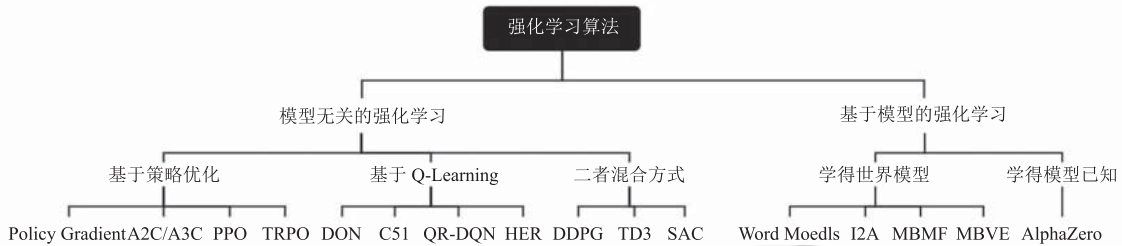


图 3 常用的强化学习算法

(1) 基于知识增强的强化对话推荐算法

目前将开源的知识库 (如 DBpedia, NELL, Microsoft Satori 等) 应用到传统推荐中的工作已经非常成熟^[93-96], 对话推荐虽然能够解决传统推荐中的一些问题, 但也存在与传统推荐相同的难以解决的问题, 如数据稀疏、缺乏可解释性等. 目前相关研究工作中^[97], 引入知识主要有以下 3 点作用: 1) 一定程度上解决数据稀疏问题, 通过关注数据的结构特征, 丰富实体表征的信息, 学习有效的用户偏好特征, 研究者应用图神经网络 (graph neural network, GNN) 学习物品、属性的高质量向量表示, 表征的语义信息能帮助端到端的 CRS 生成合适的推荐物品与交互属性, 此外, 近年图卷积网络及其变体也被广泛作为将异构图引入到对话推荐中的技术手段^[47,50,98,99], 提高对话推荐的系统性能; 2) 通过基于知识的剪枝方式和注意力机制, 减小推荐候选集, 易于模型训练, 一定程度上减轻了强化对话推荐算法动作空间过大的问题. 例如, 在 Zhou 等人的工作^[22]中, 通过图谱来筛选候选集, 极大地减小了强化推荐的动作空间; 3) 增加推荐的可解释性, 也更好地增强了对话推荐显式地引导用户交互的特点和优势.

(2) 基于图推理的强化对话推荐算法

在 Misu 等人的工作中^[100], 用户记忆 (也可称为知识) 被表示为一个概率状态, 该状态具有固定的马尔可夫概率模型层次结构, 可预测对话动作, 然而该模型在编码更丰富且更细粒度的知识以及积累用户的新信息两方面都缺乏灵活性. 同时, 后续研究工作中, 研究者尝试使用更灵活的图推理 (graph reasoning) 方法, 例如, Xu 等人^[47]结合实时的更新信息与历史数据, 构造用户记忆网络积累用户偏好, 通过在图上进行图推理, 预测用户当前和未来兴趣; Lei 等人^[100]也将图推理与 CRS 结合起来, 将用户交互数据与物品属性数据构建成异构图, 将用户交互的过程建模成图上的路径游走的过程, 通过在图上游走和推理的方式选择具体的属性节点和物品节点, 而策略网络只需要输出询问 (ASK) 与推荐 (REC) 两个动作, 使得策略网络的训练变得更加容易. 基于图推理的强化对话推荐算法, 一方面通过显式图谱路径推理, 增加了对话推荐的可解释性; 另一方面通过增加图的信息辅助, 减小了对话推荐的动作搜索空间.

总结基于强化学习的对话推荐算法的优缺点如表 9 所示, 优点集中在对话推荐中 E&E 问题的解决, 这正是强化学习框架在近年的对话推荐研究中广为使用的重要原因, 而缺点多为强化学习框架的常见应用难点, 如决策空间大、环境建模难等.

4.1.5 基于元学习的对话推荐算法

前文反复提到了对话推荐算法中关键的 E&E 问题, 除了多臂老虎机算法和强化学习算法外, 还有研究者尝试使用元学习 (meta-learning) 平衡 E&E, 核心思想是“学习如何学习一个推荐模型”, 将针对不同用户的个性化推荐视为不同的任务, 通过少量的支持集数据 (support set), 快速适应用户的偏好. 将元学习应用到对话推荐算法中是个值得研究的思路, 目标是学习一种学习算法, 该算法将用户的历史交互作为输入, 并输出一个可应用于新用户的模型, 该输出模型也就是学习到的策略函数, 元学习的思路解决了人工设计复杂的探索策略的困难. 目前基于元学

习的对话推荐算法工作已取得较为显著的成果, 例如, Zou 等人^[73]将的工作遵循了元强化学习 (meta-reinforcement learning, MRL)^[101]的思想, 提出了一个神经交互协同过滤 (neural interactive collaborative filtering, NICF) 框架, 将交互式协同过滤视为一个元学习问题, 使用 Q-Learning^[102]来学习推荐策略, 探索策略使用了 epsilon-greedy 贪心算法, 即模型以 $1 - \epsilon$ 的概率选择 Q 值最大的项目, 以 ϵ 的概率选择随机项目, NICF 通过学习一种神经探索策略去自适应地选择推荐方案, 以实现对不同用户的均衡探索和利用, 从而解决元学习中启发式选择的支持集质量不高的问题; 此外, Lee 等人^[3]提出模型无关元学习 (model agnostic meta-learning, MAML) 算法 MeLU^[103]解决推荐中的冷启动问题, 学习得到的推荐模型可以在微调 (fine-tuning) 阶段快速适应冷启动用户的偏好, 方法为向冷启动用户询问一些关于特定项目的问题 (该工作中被称为“候选证据”), 并将此策略扩展到 CRS 中, 研究一个动态的、多轮的查询策略来进一步提高对话推荐效果. 目前基于元学习的对话推荐算法的优缺点可被总结为表 10.

表 9 基于强化学习的对话推荐算法的优缺点

方法	优点	缺点
MAB	使用线性函数模型, 简单、易收敛, 可有效解决短期E&E问题	无法动态建模用户偏好
Model-based RL	基于显式规则迭代策略, 收敛快	推荐任务无显式的奖励规则, 因而建模环境依赖人工先验知识
Model-free RL	动态建模环境信息, 可学习复杂环境中的策略	动作空间大, 模型复杂、难收敛
KG+RL	具有一定解释性, 策略模型决策空间较小, 收敛较快	模型设计依赖具体任务目标, 无统一准则

表 10 基于元学习的对话推荐算法的优缺点

方法	优点	缺点
MRL	可不依赖启发式的策略设计去解决E&E问题	计算量大, 内存需求大
MAML	可快速地学习对话策略以及适应冷启动用户	

4.2 对话推荐算法的模型训练方式

对话推荐算法的实现方法非常多样, 但对话推荐算法的训练方式与具体实现方法无关, 如图 4 所示, 大体分为 3 类: 模型无关 (model-free)、基于模型 (model-based)、人参与回环 (human-in-loop). 对话推荐算法的模型训练方式的不同, 主要取决于模型训练过程中交互数据的获取与组织方式的不同. 模型无关的训练方式直接使用用户的历史行为数据, 训练数据均在历史数据中有记录; 基于模型的训练方式使用用户模拟器拟合用户历史数据, 训练数据由用户模拟器生成, 在原始的历史数据中不一定存在对应记录; 人参与回环的训练方式除使用用户模拟器生成训练数据外, 其部分关键性的训练数据来自真实人类的干预. 对于这 3 类训练方式, 以下将以基于强化学习的对话推荐算法为例进行具体说明.

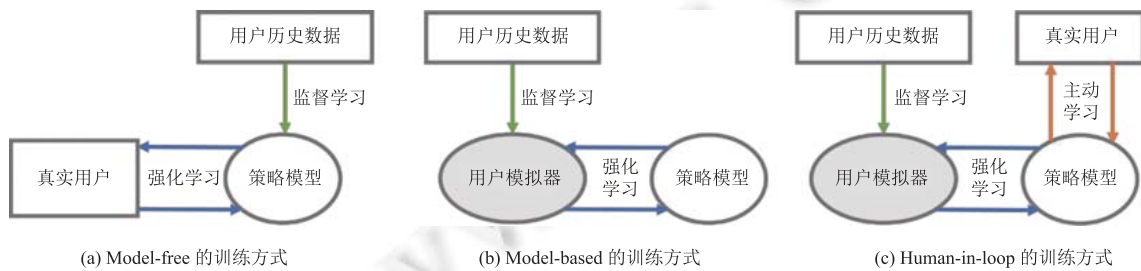


图 4 对话推荐算法的 3 种模型训练方式

4.2.1 模型无关方式

对话推荐中模型无关的训练方式, 如图 4(a) 所示, 是指对话推荐算法在交互训练中直接从大量的弱监督数据中学习规则与经验. 研究者通常是将数据直接喂给对话推荐算法, 整个训练过程独立于训练数据的环境, 不依赖于

世界模型的环境信息,可类比 RL 中的模型无关的算法训练流程,因此这类方式称为模型无关的对话推荐训练方式.

4.2.2 基于模型方式

与模型无关的训练方式完全不同,基于模型的对话推荐训练方式,类似基于模型的 RL 算法训练方式,在与环境的交互训练过程中,模型依赖于世界模型反馈的环境信息学习状态与奖励的转移概率.因此,研究者需要预先训练或构造用户模拟器,使用用户模拟器作为对话推荐算法的交互环境,模型的训练效果高度依赖用户模拟器的反馈,因此称为基于模型的方式.

如图 4(b) 所示,用户模拟器模拟用户的反馈,与 CRS 进行交互,为对话推荐算法进行强化学习训练提供足够的交互数据.由于 CRS 与用户模拟器的交互训练过程中产生了大量的新数据,因此基于模型的方式比模型无关的方式更有效率,也减轻了与真实环境交互的训练成本.由于基于模型的方式可以有效降低人工训练的成本,工业界中一直广泛地研究与应用这类训练方式^[104].虽然基于模型的方式更有效率且更低成本,但这个方式同时存在隐患,即它的训练效果高度依赖于用户模拟器的质量,低质量或者不合适的用户模拟器会忽略数据中的重要特征或者是增大数据中的噪声,造成对话推荐算法无法学习到正确的策略,最终导致 CRS 在与真实环境交互中的准确率较低.因此,在使用基于模型的方式时,如何衡量并减小真实世界与模拟环境的语义鸿沟,仍是一个巨大挑战.目前针对这一问题的研究工作较少,Zhou 等人^[50]尝试通过互信息最大化的多知识图谱语义融合技术,解决对话推荐中自然语言的表示和商品级别的用户偏好之间的语义鸿沟,这个工作或许为解决用户模拟器的不可靠问题提供了一个新思路,总之,对话推荐的模拟器问题未来值得更深入的研究.

4.2.3 人参与回环方式

早期,谷歌的研究者将人的知识和经验引入到强化学习的训练中^[105],让人标注部分重要数据^[106],从而减轻预训练中的监督学习所造成的协方差漂移问题(covariate shift)^[107].在如今的对话推荐的研究中,基于模型的方式使得模型的准确率深受模拟器生成的数据质量影响,基于相同解决思路,如图 4(c) 所示,对话推荐的研究者也通过在对话推荐算法训练中加入人为标注干预的方式^[105,108,109],去尝试减小模拟器的偏差和错误,这样的训练方式为对话推荐算法的人参与回环训练方式.

5 对话推荐研究存在的问题

近年对话推荐算法的研究工作非常丰富,解决了 CRS 实现中的部分问题,但对话推荐中仍有很多问题未被重视或者未被解决,本章将对对话推荐算法研究中仍存在的问题进行阐述.

5.1 评价指标不统一

目前对话推荐研究工作中,尚未形成统一的评价指标体系.根据 Jannach 等人的结论^[56],人类评估和专家分析对 CRS 是必要的,因为目前没有完美的指标可以来评估其所有方面,对话推荐模型的评价方法的研究还有很长的路要走.正如第 2.2 节和第 2.3 节所述,评估对话推荐需要实时反馈,这在现实情况下是高投入低回报的工作^[6],过高的成本使得在线人工评价的方法不具有普适性.因此,大多数对话推荐的研究工作采用用户模拟器方法来构建一个训练和评估 CRS 的环境^[110].然而,用户模拟器无法完全代替真实人类的反馈,模拟器与真实人类的反馈之间存在或大或小的偏差.对此,目前的研究工作中实现可靠仿真的可行方法有设计系统的仿真议程(simulation agenda)^[110,111]或者构建密集的用户交互数据等,如何最大程度模拟用户仍需进一步研究.

此外,如第 2.1 节所述,对话推荐研究工作在不同的数据集上有不同的假设和设置,评价指标也不尽相同.近期,Zhou 等人^[112]尝试实现了一个开源的 CRS 工具包,用于评价不同的对话推荐模型,然而它的实现模型主要基于端到端对话系统^[17,49,50]或深层语言模型^[46],缺乏基于第 2 类通用推荐数据集的、具有显示用户偏好的对话推荐算法^[53,65].因此,制定综合且全面的评价指标和评价方法来评估对话推荐的效果仍是个待解决的问题.

5.2 探索与收益的平衡

对话推荐算法的一大研究难点是平衡用户兴趣的探索(ASK)与推荐行为(REC)带来的收益.过长的交互能

够获取更多用户的信息从而提高推荐准确率,但同时也为用户带来额外的负担,使得用户可能在推荐前就流失,造成极高的用户流失风险。早期,研究者通过专家经验制定人工规则,尽可能地在用户负担与系统准确率中取得平衡,例如,设置最高对话轮次、设置一定数量的用户兴趣槽位等,当满足这些条件时,系统就进行推荐。但这样的人工规则受制于专家在领域的先验知识,在跨领域应用时,CRS的表现会大幅下降,研究者又将多臂老虎机算法应用到对话推荐算法中,一定程度上缓解了E&E平衡问题,但远达不到令人满意的程度。后期,具有探索与收益的强化学习思路被引入进来,强化学习算法确实是解决平衡用户负担与推荐准确率的有效方法,但如何设计有效的强化学习模型,包括但不限于进行函数塑形(reward shaping)等方案,以实现高效捕捉用户偏好并进行推荐,仍有待研究者不断尝试和创新。目前的对话推荐研究工作中,研究者做了各种尝试,例如,以用户交互图谱形式构造用户当前会话数据的图表示;使用注意力机制高效匹配用户的兴趣点;引入外部物品属性图谱丰富用户当前兴趣表示的信息等。这些工作的切入点非常丰富,但目前仍没有一个公认有效的针对性解决方案。在探索与收益问题上,目前的研究工作存在以下3个具体的问题:1)如何在保证推荐准确率的前提下,减轻用户的交互负担;2)如何将会话数据与用户历史数据结合,推断用户具有时序性特点的偏好;3)如何利用用户历史兴趣引导用户发掘未来偏好。

5.3 数据稀疏性

数据稀疏性问题一方面是传统推荐中的冷启动问题。冷启动是推荐中的常见问题,常被研究者作为一个专门的课题来讨论,又分为物品冷启动和用户冷启动,指当有新用户或新物品进入系统中,由于没有历史交互数据,导致推荐系统无法学习到物品或用户的特征,从而影响推荐效果。对话推荐天然关注于当前会话数据,所以对于没有历史交互数据的新用户有更好的表现,一定程度上可以解决用户冷启动问题,但在物品冷启动问题上没有非常好的解决方式。

数据稀疏性问题另一方面是对话推荐中的用户反馈数据稀疏问题。用户反馈数据稀疏问题,一般是指推荐系统中用户的正反馈数据相对于庞大的物品空间是非常稀少的,而在对话推荐中指向的是,CRS受制于有限的对话轮次,获取到的用户反馈仅为有限轮次数量的数据。因此如何从少量的数据中推断用户偏好是对话推荐的一大研究难点,虽然现在已经有通过强化学习、外部知识图谱等方式来解决该问题的研究工作,但效果还远达不到预期。

5.4 对话策略模型低效

深度强化学习已经被广泛应用到对话推荐中用于学习对话策略,但标准的强化模型被应用到对话推荐中后,研究者面临着模型难收敛、有效性低等问题,而导致对话策略模型低效的原因有以下两个。

1) CRS的环境缺少明确的规则,从而导致使用强化模型时没有明确的奖励函数以及清晰的动作空间,这里归结为模型训练方式问题,这个问题使得对话推荐算法在训练交互中的参数更新有延迟,造成CRS需要更多的交互轮次来获取到用户的偏好信息。

2) 强化对话推荐算法中的强化模型设计完全依赖于研究者的直觉与经验,与任务不合适的强化模型会导致对话策略模型的训练低效,甚至无效,这里把这个问题归结为模型面向任务的设计问题。其实,对于对话推荐算法训练低效问题,很多研究工作有尝试讨论,但都由于拘泥于各自工作的场景和数据,不具有普适性,例如,在Su等人的工作^[113]中,作者讨论得出与Q-Learning的强化方法相比,基于策略的强化方法具有更强的收敛性,因为基于值的方法在使用函数逼近时经常会无法收敛,它们在价值空间中进行了优化,价值计算的细微变化会导致策略空间的巨大变化;但在Lei等人的工作^[65]中,根据作者的实验,作者认为Q-Learning的强化模型相对于基于策略的强化模型,虽然不能得到更好的训练模型,但在训练中更易收敛。这样不同的研究结论说明,对话推荐算法的训练在面对不同数据任务时,对于训练效率的提高没有可靠的统一方法。

6 未来展望

总的来看,在过去几年中,对话推荐领域迎来了新的研究热潮,其中最新的方法依赖于机器学习和深度学习,

特别是强化学习, 预计在未来几年会有更多关于对话推荐的研究工作. 目前, 对话推荐的研究虽然在很多方面已取得重大进展, 但正如第 5 节所总结的目前研究工作中的诸多问题, 对话推荐领域中仍有诸多的研究方向值得关注, 本节概述了该领域的一些开放问题和有价值的未来研究方向.

6.1 协同优化的端到端对话推荐

如第 3 节所述, CRS 中的用户意图理解任务、对话策略任务和推荐任务通常在 3 个模块中分别研究, 这 3 个部分相互之间共享某些目标和数据^[49,50,53,114], 例如, 用户意图理解模块将提取的用户偏好值输入给推荐模块, 然后将推荐模块生成的推荐集合输入到对话策略模块中, 由对话策略模块决定当前是否推荐给用户. 然而, 3 个模块也拥有部分独立的数据, 例如, 用户意图理解模块拥有用户反馈的丰富语义信息, 但推荐模块不会直接知晓^[17]. 因此, 在 CRS 框架下, 3 个模块在多回合的交互过程中缺乏协调性^[17,49], 这也是目前很多 CRS 工作在在线用户评估中表现不理想的原因之一^[56], 更理想的方式是实现这 3 个任务在系统中的协同学习, 以达到共同的对话推荐目标.

6.2 知识增强的对话推荐

由于对话推荐对上下文信息的依赖以及其自身存在的数据稀疏性问题, 引入外界知识能够帮助 CRS 提高交互推荐的性能. 在对话推荐发展的早期阶段, 只有推荐物品本身被认为是知识^[13]. 随后, 研究者又引入了物品的属性信息, 帮助 CRS 建模用户偏好^[64]. 而在近期, 研究者更全面地考虑了知识图谱中丰富的语义信息^[42,47,50,65], 例如, Liao 等人^[51]提出了一种利用外部领域知识改进对话推荐的独特方法, 作者将外部知识域嵌入到记忆网络中, 根据提取的知识域和当前对话语境生成 CRS 的响应行为; Zhou 等人^[50]为了更好地理解句子中的概念, 提出在 CRS 框架中融合两个外部知识图谱, 一个是单词间关系的图谱 (包括同义词、反义词等) 用于理解句子中关键词的概念, 另一个是物品和物品属性关联的知识图谱. 研究如何将外部知识更好地引入到对话推荐中是对话推荐今后的一个重要研究方向.

6.3 基于多模态交互数据的对话推荐

除了知识图谱数据外, 多模态数据在对话推荐中的应用也有可观的发展前景, 其可从新的数据维度拓展 CRS 的交互数据形式. 对此, 可以横向对比 DS 的研究, 目前多数 DS 允许输入文本、图像、音频等多种类型数据. 处理多模态交互数据在 CRS 研究中是一个较大的挑战, 它要求 CRS 必须能够理解各种类型的输入数据的语义, 例如, 用户想通过一个物品图片找到相似的物品, CRS 需要了解该物品图片的各种语义特征, 从而理解用户的语义偏好, 才能找到相似的物品, 而不同类型的数据需要不同的方法进行语义提取, 这就需要 CRS 的用户意图理解模型进行额外的训练, 使得 CRS 的训练变得复杂.

目前已有少量对话推荐研究工作利用了图像数据^[51,63,115,116], 例如, Liao 等人^[51]的对话推荐工作中尝试通过 E1-tree 神经网络以及外部领域知识整合图像语义意义; Yu 等人^[63]提出了一种视觉对话增强 CRS 模型, 在他们的工作中, 模型输出推荐物品的图片列表, 而用户给出自然语言形式的反馈. 在这类工作中, 图像通过帮助模型学习物品更具信息量的表征, 使得系统更好地向用户传递信息. 除视觉图像数据外, 其他类型的数据也可以提升对话推荐整体效果, 例如语音数据, 用户的语音反馈包含语气, 可以传达用户的情感以及对某些实体的情感^[117]. 在未来, 也许能够看到更多的对话推荐研究工作关注处理多模态数据.

6.4 无偏的对话推荐

推荐系统在实际应用中不可避免地会遇到各种类型的偏差 (bias)^[60], 例如流行度偏差 (popularity bias)^[118,119]、一致性偏差 (conformity bias)^[120,121]、选择偏差 (selection bias, 也被称为积极偏见)^[122-126]、曝光偏差 (exposure bias)^[127]等. 由于 CRS 引入了用户的交互, 传统推荐系统中有些偏差类型在 CRS 中是可以避免的, 例如, 流行度偏差、选择偏差等.

对于流行度偏差, 由于静态推荐系统不确定用户是否会跟随人群并喜欢流行的物品, 受欢迎的产品物品被推荐的概率更高. 但是 CRS 则不同, 它可以实时查询用户对热门物品的态度, 并在用户给出负面反馈时, 在接下来的推荐中避免推荐这类商品, 因此, 流行度偏差在对话推荐中是可以避免的. 对于选择偏差, 由于推荐系统中用户只

会对感兴趣的物品有交互行为, 忽视不喜欢的物品, 这导致训练数据中用户对物品的评分数据的经常性丢失不是随机发生的, 更有可能是用户对不喜欢物品的选择性忽略^[125]. 选择偏差极易在训练回路中被放大, 从而损害推荐模型^[128,129]. 简单来说, 一个被有偏差数据所污染的推荐模型可能会反复生成相同的商品, 即使用户已经暗示了他们喜欢其他商品. 然而, CRS 不仅将用户的历史数据作为信息来源, 更会关注用户在当前交互中所表现出的偏好, 所以可以很大程度上避免依赖有偏差的历史交互数据, 从而避免了选择偏差.

但是, 某些类型的偏差在对话推荐中仍然存在, 例如曝光偏差, 这是由于 CRS 的交互轮次是有限的, 只能向用户展示部分物品, 用户也只能与其中的小部分物品进行交互. 与传统推荐相同, 在对话推荐中, 这部分有机会曝光给用户的物品一般是由某个模型或某个机制所选择, 所以用户在不知道其他物品存在的情况下, 只能在有限的曝光物品中选择.

目前研究对话推荐中去偏问题的研究工作相对较少, 有 Huang 等人^[122]在用户模拟阶段尝试去除交互推荐的选择偏差; 此外, Chen 等人^[60]对不同类型的偏差进行了全面的调查, 并梳理了推荐领域的一些去偏方法, 为对话推荐的去偏提供了一些思路.

6.5 用户反馈的改进

在现有的 CRS 研究工作中, 考虑用户反馈来改善系统性能的工作极少, 大多数工作的算法在处理用户的非正反馈方面存在问题, 例如, Lei 等人的 EAR 模型中^[53], 通过设计“Reflection”阶段来尝试解决这一问题, 在“Reflection”阶段, 处理用户非正反馈时, 简单地将被拒绝的物品作为负样本, 用来训练和更新下一轮模型参数; 除此之外, Cai 等人的工作中^[130]则是尝试采用了通过意图分类来学习用户的反馈. 这些处理用户反馈方法非常简单, 也有一定效果, 但是进一步思考后会发现其中的问题, 用户拒绝某个物品可能是因为他们之前通过别的途径购买过, 但这并不意味着他们不喜欢这个物品. 为了改进对用户反馈的粗暴处理, 近期 Xu 等人在提出的 CRS 框架 FPAN^[74]中尝试关注到用户属性级别反馈与物品级别反馈的改进, 基于门控机理^[131,132]对用户交互过程中的偏好表征进行动态微调, 实验效果取得大幅提升.

未来对话推荐研究可考虑更复杂的用户反馈处理策略, 如识别可靠的负样本 (negative sample)^[133,134], 或是从因果学习 (causal learning) 角度分析对话推荐模型, 将用户对物品和属性的偏好解耦合, 学习用户偏好的分离式表征 (disentangled representation)^[135,136]等.

6.6 更智能的对话推荐策略设计

目前的对话推荐的研究工作中所考虑的多轮次交互策略仍不够成熟, 部分工作是使用人工制定的规则函数来确定询问属性或做出推荐的时机, 例如, 完成固定轮次的询问后进行推荐^[137], 且在基于端到端对话系统或深度神经网络模型的 CRS 研究中, 甚至没有使用明确的策略来控制多轮次交互^[17,49]. 对于大多数的基于强化学习的对话推荐研究工作, 研究者设计对话推荐模型通过强化策略网络输出 CRS 的行为, 决定是否询问或推荐^[19,53,65], 但是在设计强化学习中的状态、行动和奖励方面还有很多改进的空间. 例如, 利用物品之间的关系 (如互补性^[138]和替代性)^[139-141]来规划未来的对话行为, 或考虑为 CRS 设计更复杂的系统行为, 如回答用户提出的开放领域问题或出于娱乐目的、非任务导向的话题聊天^[45,142]. 此外, 还可以考虑引入其他类型的强化技术到对话推荐的研究工作中, 解决 CRS 由于使用强化模型导致的难训练等问题^[143], 例如, 在某些对话推荐场景中有大量的用户行为模式, 难以定义奖励, 可利用逆强化学习 (inverse reinforcement learning, IRL)^[144]方法, 通过观察这些场景中的行为数据, 学习一个合适的奖励函数; 对于交互数据稀疏且类型多样的 CRS, 可利用元强化学习^[101,145]加速其训练过程, 提高对后续新任务的学习效率.

7 结束语

作为解决目前传统推荐系统中存在问题的有效手段, 对话推荐的研究具有重大意义. 近两年, 在推荐领域和数据挖掘领域的顶级会议与期刊中, 大量的对话推荐相关工作被发表, 可见对话推荐的研究进入蓬勃发展阶段. 本文详细讨论了对话推荐的研究内容与特点, 重点关注对话推荐中的算法设计, 对目前的对话推荐算法进行了归类和

总结,并介绍了对话推荐的评价指标及当前研究的挑战与未来的研究方向.本文有助于人们理解对话推荐的研究内容、研究难点与研究思路,引发人们对未来对话推荐研究的思考.

References:

- [1] Wang WJ, Feng FL, He XN, Nie LQ, Chua TS. Denoising implicit feedback for recommendation. In: Proc. of the 14th ACM Int'l Conf. on Web Search and Data Mining. ACM, 2020. 373–381. [doi: 10.1145/3437963.3441800]
- [2] Wang WJ, Feng FL, He XN, Zhang HW, Chua TS. Clicks can be Cheating: Counterfactual recommendation for mitigating clickbait issue. In: Proc. the 44th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. ACM, 2020. 1288–1297. [doi: 10.1145/3404835.3462962]
- [3] Lee H, Im J, Jang S, Cho H, Chung S. MeLU: Meta-learned user preference estimator for cold-start recommendation. In: Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 1073–1082. [doi: 10.1145/3292500.3330859]
- [4] Cen YK, Zhang JW, Zou X, Zhou C, Yang HX, Tang J. Controllable multi-interest framework for recommendation. In: Proc. of the 26th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. ACM, 2020. 2942–2951. [doi: 10.1145/3394486.3403344]
- [5] Gao CM, Yuan S, Zhang Z, Yin HZ, Shao JM. BLOMA: Explain collaborative filtering via boosted local rank-one matrix approximation. In: Proc. of the 2019 Int'l Workshops on Database Systems for Advanced Applications. Chiang Mai: Springer, 2019. 487–490. [doi: 10.1007/978-3-030-18590-9_72]
- [6] Jagerman R, Markov I, de Rijke M. When people change their mind: Off-policy evaluation in non-stationary recommendation environments. In: Proc. of the 12th ACM Int'l Conf. on Web Search and Data Mining. Melbourne: ACM, 2019. 447–455. [doi: 10.1145/3289600.3290958]
- [7] Wang WQ, Benbasat I. Research Note—A contingency approach to investigating the effects of user-system interaction modes of online decision aids. *Information Systems Research*, 2013, 24(3): 861–876. [doi: 10.1287/isre.1120.0445]
- [8] Wärnestål P. User evaluation of a conversational recommender system. In: Proc. of the 4th Workshop on Knowledge and Reasoning in Practical Dialogue Systems. 2005.
- [9] Rich E. Scene I Scene II User Modeling via Stereotypes. 2008. <https://core.ac.uk/display/24754407>
- [10] Tou FN, Williams MD, Fikes R, Henderson A, Malone T. RABBIT: An intelligent database assistant. In: Proc. of the 1982 National Conf. on Artificial Intelligence. Pittsburgh: AAAI, 1982. 314–318.
- [11] Burke RD, Hammond KJ, Young BC. The FindMe approach to assisted browsing. *IEEE Expert: Intelligent Systems and Their Applications*, 1997, 12(4): 32–40. [doi: 10.1109/64.608186]
- [12] Hammond KJ, Burke R, Martin C, Lytinen SL. A case-based approach to knowledge navigation. In: Proc. of the 1995 Working Notes of the AAAI Spring Symp. : Information Gathering from Heterogeneous, Distributed Environments. Montreal: AAAI, 1995. 2071–2072.
- [13] Christakopoulou K, Radlinski F, Hofmann K. Towards conversational recommender systems. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 815–824. [doi: 10.1145/2939672.2939746]
- [14] Yang SM. Overview of the interaction design for recommender systems. *Software*, 2013, 34(8): 154–158 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-6970.2013.08.059]
- [15] Yang ZQ, Yin Z, Wang H. Providing resource recommendation based on interactive behavior and resource content. *Journal of Computer-aided Design & Computer Graphics*, 2014, 26(5): 747–754 (in Chinese with English abstract).
- [16] Zhang YF, Chen X, Ai QY, Yang L, Croft WB. Towards conversational search and recommendation: System ask, user respond. In: Proc. of the 27th ACM Int'l Conf. on Information and Knowledge Management. Torino: ACM, 2018. 177–186. [doi: 10.1145/3269206.3271776]
- [17] Li R, Kahou SE, Schulz H, Michalski V, Charlin L, Pal C. Towards deep conversational recommendations. In: Proc. of the 32nd Conf. on Neural Information Processing Systems. Montreal, 2018.
- [18] Mahmood T, Ricci F. Improving recommender systems with adaptive conversational strategies. In: Proc. of the 20th ACM Conf. on Hypertext and Hypermedia. Torino: ACM, 2009. 73–82. [doi: 10.1145/1557914.1557930]
- [19] Sun YM, Zhang Y. Conversational recommender system. In: Proc. of the 41st Int'l ACM SIGIR Conf. on Research & Development in Information Retrieval. Ann Arbor: ACM, 2018. 235–244. [doi: 10.1145/3209978.3210002]
- [20] He C, Parra D, Verbert K. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 2016, 56: 9–27. [doi: 10.1016/j.eswa.2016.02.013]
- [21] Wang HZ, Wu QY, Wang HN. Factorization bandits for interactive recommendation. In: Proc. of the 31st AAAI Conf. on Artificial

- Intelligence. San Francisco: AAAI, 2017. 2695–2702.
- [22] Zhou SJ, Dai XY, Chen HK, Zhang WN, Ren K, Tang RM, He XQ, Yu Y. Interactive recommender system via knowledge graph-enhanced reinforcement learning. In: Proc. of the 43rd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. ACM, 2020. 179–188. [doi: 10.1145/3397271.3401174]
- [23] Chen HK, Dai XY, Cai H, Zhang WN, Wang XJ, Tang RM, Zhang YZ, Yu Y. Large-scale interactive recommendation with tree-structured policy gradient. In: Proc. of the 33rd AAAI Conf. on Artificial Intelligence. AAAI, 2019. 3312–3320. [doi: 10.1609/aaai.v33i01.33013312]
- [24] Zhao XY, Zhang L, Ding ZY, Xia L, Tang JL, Yin DW. Recommendations with negative feedback via pairwise deep reinforcement learning. In: Proc. of the 24th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. London: ACM, 2018. 1040–1048. [doi: 10.1145/3219819.3219886]
- [25] Zhao XX, Zhang WN, Wang J. Interactive collaborative filtering. In: Proc. of the 22nd ACM Int'l Conf. on Information & Knowledge Management. San Francisco: ACM, 2013. 1411–1420. [doi: 10.1145/2505515.2505690]
- [26] Gao CM, Lei WQ, He XN, de Rijke M, Chua TS. Advances and challenges in conversational recommender systems: A Survey. *AI Open*, 2021, 2: 100–126. [doi: 10.1016/j.aiopen.2021.06.002]
- [27] Chen L, Pu P. Critiquing-based recommenders: Survey and emerging trends. *User Modeling and User-Adapted Interaction*, 2012, 22(1-2): 125–150. [doi: 10.1007/s11257-011-9108-6]
- [28] Luo K, Sanner S, Wu G, Li HZ, Yang HJ. Latent linear critiquing for conversational recommender systems. In: Proc. of the Web Conf. 2020. Taipei: ACM, 2020. 2535–2541. [doi: 10.1145/3366423.3380003]
- [29] Luo K, Yang HJ, Wu G, Sanner S. Deep critiquing for VAE-based recommender systems. In: Proc. of the 43rd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. ACM, 2020. 1269–1278. [doi: 10.1145/3397271.3401091]
- [30] Pu P, Faltings B. Decision tradeoff using example-critiquing and constraint programming. *Constraints*, 2004, 9(4): 289–310. [doi: 10.1023/B:CONS.0000049205.05581.24]
- [31] Tversky A, Simonson I. Context-dependent preferences. *Management science*, 1993, 39(10): 1179–1189.
- [32] Smyth B, McGinty L. An analysis of feedback strategies in conversational recommenders. In: Proc. of the 14th Irish Conf. on Artificial Intelligence and Cognitive Science. 2003.
- [33] Ma H, Yang HX, King I, Lyu MR. Learning latent semantic relations from clickthrough data for query suggestion. In: Proc. of the 17th ACM Conf. on Information and Knowledge Management. Napa Valley: ACM, 2008. 709–718. [doi: 10.1145/1458082.1458177]
- [34] Bar-Yossef Z, Kraus N. Context-sensitive query auto-completion. In: Proc. of the 20th Int'l Conf. on World Wide Web. Hyderabad: ACM, 2011. 107–116. [doi: 10.1145/1963405.1963424]
- [35] Cai F, de Rijke M. A survey of query auto completion in information retrieval. *Foundations and Trends® in Information Retrieval*, 2016, 10(4): 273–363. [doi: 10.1561/15000000055]
- [36] Covington P, Adams J, Sargin E. Deep neural networks for Youtube recommendations. In: Proc. of the 2016 ACM Conf. on Recommender Systems. Boston: ACM, 2016. 191–198. [doi: 10.1145/2959100.2959190]
- [37] Dodge J, Gane A, Zhang X, Bordes A, Chopra S, Miller AH, Szlam AD, Weston J. Evaluating prerequisite qualities for learning end-to-end dialog systems. arXiv:1511.06931, 2016.
- [38] Zhou KY, Prabhume S, Black AW. A dataset for document grounded conversations. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 708–713. [doi: 10.18653/v1/D18-1076]
- [39] Moghe N, Arora S, Banerjee S, Khapra MM. Towards exploiting background knowledge for building conversation systems. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 2322–2332. [doi: 10.18653/v1/D18-1255]
- [40] Dinan E, Roller S, Shuster K, Fan A, Auli M, Weston J. Wizard of Wikipedia: Knowledge-powered conversational agents. In: Proc. of the 2019 ICLR Conf. Blind Submission. 2018.
- [41] Kang D, Balakrishnan A, Shah P, Crook P, Boureau YL, Weston J. Recommendation as a communication game: Self-supervised bot-play for goal-oriented dialogue. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 1951–1961. [doi: 10.18653/v1/D19-1203]
- [42] Moon S, Shah P, Kumar A, Subba R. OpenDialKG: Explainable conversational reasoning with attention-based walks over knowledge graphs. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 845–854. [doi: 10.18653/v1/P19-1081]

- [43] Wu WQ, Guo Z, Zhou XY, Wu H, Zhang XY, Lian RZ, Wang HF. Proactive human-machine conversation with explicit conversation goal. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 3794–3804. [doi: 10.18653/v1/P19-1369]
- [44] Zhou H, Zheng CJ, Huang KL, Huang ML, Zhu XY. KdConv: A Chinese multi-domain dialogue dataset towards multi-turn knowledge-driven conversation. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 7098–7108. [doi: 10.18653/v1/2020.acl-main.635]
- [45] Liu ZM, Wang HF, Niu ZY, Wu H, Che WX, Liu T. Towards conversational recommendation over multi-type dialogs. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 1036–1049. [doi: 10.18653/v1/2020.acl-main.98]
- [46] Zhou K, Zhou YH, Zhao WX, Wang XK, Wen JR. Towards topic-guided conversational recommender system. In: Proc. of the 28th Int'l Conf. on Computational Linguistics. Barcelona: Int'l Committee on Computational Linguistics, 2020. 4128–4139. [doi: 10.18653/v1/2020.coling-main.365]
- [47] Xu H, Moon S, Liu HL, Liu B, Shah P, Liu B, Yu P. User memory reasoning for conversational recommendation. In: Proc. of the 28th Int'l Conf. on Computational Linguistics. Barcelona: Int'l Committee on Computational Linguistics, 2020. 5288–5308. [doi: 10.18653/v1/2020.coling-main.463]
- [48] Hayati SA, Kang D, Zhu QXY, Shi WY, Yu Z. INSPIRED: Toward sociable recommendation dialog systems. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, 2020. 8142–8152. [doi: 10.18653/v1/2020.emnlp-main.654]
- [49] Chen QB, Lin JY, Zhang YC, Ding M, Cen YK, Yang HX, Tang J. Towards knowledge-based recommender dialog system. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 1803–1813. [doi: 10.18653/v1/D19-1189]
- [50] Zhou K, Zhao WX, Bian SQ, Zhou YH, Wen JR, Yu JS. Improving conversational recommender systems via knowledge graph based semantic fusion. In: Proc. of the 26th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. ACM, 2020. 1006–1014. [doi: 10.1145/3394486.3403143]
- [51] Liao L, Ma YS, He XN, Hong RC, Chua TS. Knowledge-aware multimodal dialogue systems. In: Proc. of the 26th ACM Int'l Conf. on Multimedia. Seoul: Association for Computing Machinery, 2018. 801–809. [doi: 10.1145/3240508.3240605]
- [52] Greco C, Suglia A, Basile P, Semeraro G. Converse-et-impera: Exploiting deep learning and hierarchical reinforcement learning for conversational recommender systems. In: Proc. of the 16th Int'l Conf. of the Italian Association for Artificial Intelligence on AI*IA 2017 Advances in Artificial Intelligence. Bari: Springer, 2017. 372–386. [doi: 10.1007/978-3-319-70169-1_28]
- [53] Lei WQ, He XN, Miao YS, Wu QY, Hong RC, Kan MY, Chua TS. Estimation-action-reflection: Towards deep interaction between conversational and recommender systems. In: Proc. of the 13th ACM Int'l Conf. on Web Search and Data Mining. Houston: ACM, 2020. 304–312. [doi: 10.1145/3336191.3371769]
- [54] Papineni K, Roukos S, Ward T, Zhu WJ. BLEU: A method for automatic evaluation of machine translation. In: Proc. of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2002. 311–318. [doi: 10.3115/1073083.1073135]
- [55] Szpakowicz S. Text Summarization Branches Out. Barcelona: Association for Computational Linguistics, 2004.
- [56] Jannach D, Manzoor A. End-to-end learning for conversational recommendation: A long way to go? In: Proc. of the IntRS Workshop at ACM RecSys2020. New York: ACM, 2020. 72–76.
- [57] Chen MM, Beutel A, Covington P, Jain S, Chi EH. Top-K off-policy correction for a REINFORCE recommender system. In: Proc. of the 12th ACM Int'l Conf. on Web Search and Data Mining. Melbourne: Association for Computing Machinery, 2019. 456–464. [doi: 10.1145/3289600.3290999]
- [58] Deng AL, Zhu YY, Shu BL. A collaborative filtering recommendation algorithm based on item rating prediction. Ruan Jian Xue Bao/Journal of Software, 2003, 14(9): 1621–1628 (in Chinese with English abstract). http://jos.org.cn/jos/article/abstract/20030917?st=article_issue [doi: 1000-9825/2003/14(09)1621]
- [59] Huang ZH, Zhang JW, Tian CQ, Sun SL, Xiang Y. Survey on learning-to-rank based recommendation Algorithms. Ruan Jian Xue Bao/Journal of Software, 2016, 27(3): 691–713 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4948.htm> [doi: 10.13328/j.cnki.jos.004948]
- [60] Chen JW, Dong HD, Wang X, Feng FL, Wang M, He XN. Bias and debias in recommender system: A survey and future directions. arXiv:2010.03240v2, 2020.
- [61] Liu DG, Cheng PX, Dong ZH, He XQ, Pan WK, Ming Z. A general knowledge distillation framework for counterfactual recommendation via uniform data. In: Proc. of the 43rd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval.

- ACM, 2020. 831–840. [doi: 10.1145/3397271.3401083]
- [62] Yang LQ, Yin C, Yuan X, Wang CY, Belongie S, Estrin D. Unbiased offline recommender evaluation for missing-not-at-random implicit feedback. *Proc. of the 12th ACM Conf. on Recommender Systems*. Vancouver: ACM, 2018. 279–287. [doi: 10.1145/3240323.3240355]
- [63] Yu T, Shen YL, Jin HX. A visual dialog augmented interactive recommender system. In: *Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining*. Anchorage: ACM, 2019. 157–165. [doi: 10.1145/3292500.3330991]
- [64] Christakopoulou K, Beutel A, Li R, Jain S, Chi EH. Q&R: A two-stage approach toward interactive recommendation. In: *Proc. of the 24th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining*. London: ACM, 2018. 139–148. [doi: 10.1145/3219819.3219894]
- [65] Lei WQ, Zhang GY, He XN, Miao YS, Wang X, Chen L, Chua TS. Interactive path reasoning on graph for conversational recommendation. In: *Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining*. ACM, 2020. 2073–2083. [doi: 10.1145/3394486.3403258]
- [66] Zou J, Chen YF, Kanoulas E. Towards question-based recommender systems. In: *Proc. of the 43rd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*. ACM, 2020. 881–890. [doi: 10.1145/3397271.3401180]
- [67] Liao LZ, Takanobu R, Ma YS, Yang X, Huang ML, Chua TS. Deep conversational recommender in travel. arXiv:1907.00710v1, 2019.
- [68] Ren XH, Yin HZ, Chen T, Wang H, Hung NQV, Huang Z, Zhang XL. CRSAL: Conversational recommender systems with adversarial learning. *ACM Trans. on Information Systems*, 2020, 38(4): 1–40. [doi: 10.1145/3394592]
- [69] Gulcehre C, Ahn S, Nallapati R, Zhou BW, Bengio Y. Pointing the unknown words. In: *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers)*. Berlin: Association for Computational Linguistics, 2016. 140–149. [doi: 10.18653/v1/P16-1014]
- [70] Sedhain S, Menon AK, Sanner S, Xie LX. AutoRec: Autoencoders meet collaborative filtering. In: *Proc. of the 2015 Int'l Conf. on World Wide Web*. Florence: ACM, 2015. 111–112. [doi: 10.1145/2740908.2742726]
- [71] Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. In: *Proc. of the 31st Int'l Conf. on Neural Information Processing Systems*. Long Beach: ACM, 2017. 1025–1035. [doi: 10.5555/3294771.3294869]
- [72] Ying R, He RN, Chen KF, Eksombatchai P, Hamilton WL, Leskovec J. Graph convolutional neural networks for web-scale recommender systems. In: *Proc. of the 24th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining*. London: ACM, 2018. 974–983. [doi: 10.1145/3219819.3219890]
- [73] Zou LX, Xia L, Gu YL, Zhao XY, Liu WD, Huang JX, Yin DW. Neural interactive collaborative filtering. In: *Proc. of the 43rd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval*. ACM, 2020. 749–758. [doi: 10.1145/3397271.3401181]
- [74] Xu KR, Yang JX, Xu J, Gao S, Guo J, Wen JR. Adapting user preference to online feedback in multi-round conversational recommendation. In: *Proc. of the 14th ACM Int'l Conf. on Web Search and Data Mining*. ACM, 2021. 364–372. [doi: 10.1145/3437963.3441791]
- [75] Wu G, Luo K, Sanner S, Soh H. Deep language-based critiquing for recommender systems. In: *Proc. of the 13th ACM Conf. on Recommender Systems*. Copenhagen: ACM, 2019. 137–145. [doi: 10.1145/3298689.3347009]
- [76] McCarthy K, Reilly J, McGinty L, Smyth B. Thinking positively-explanatory feedback for conversational recommender systems. In: *Proc. of the ECCBR 2004 Workshops*. 2004.
- [77] Smyth B, McGinty L, Reilly J, McCarthy K. Compound critiques for conversational recommender systems. In: *Proc. of the 2004 IEEE/WIC/ACM Int'l Conf. on Web Intelligence*. Beijing: IEEE, 2004. 145–151. [doi: 10.1109/WI.2004.10098]
- [78] Viappiani P, Pu P, Faltings B. Conversational recommenders with adaptive suggestions. In: *Proc. of the 2007 ACM Conf. on Recommender Systems*. Minneapolis: ACM, 2007. 89–96. [doi: 10.1145/1297231.1297246]
- [79] Chen BY, Huang L, Wang CD, Jing LP. Explicit and implicit feedback based collaborative filtering algorithm. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(3): 794–805 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5897.htm> [doi: 10.13328/j.cnki.jos.005897]
- [80] Zhang QB, Wang B, Cui NN, Song XX, Qin Q. Attention-based regularized matrix factorization for recommendation. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(3): 778–793 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5913.htm> [doi: 10.13328/j.cnki.jos.005913]
- [81] Du DF, Xu T, Lu YN, Guan C, Liu Q, Chen EH. User rating prediction based on trust-driven probabilistic matrix factorization. *Ruan Jian Xue Bao/Journal of Software*, 2018, 29(12): 3747–3763 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5322.htm> [doi: 10.13328/j.cnki.jos.005322]
- [82] Liu HF, Jing LP, Yu J. Survey of matrix factorization based recommendation methods by integrating social information. *Ruan Jian Xue*

- Bao/Journal of Software, 2018, 29(2): 340–362 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5391.htm> [doi: 10.13328/j.cnki.jos.005391]
- [83] Salakhutdinov R, Mnih A. Probabilistic matrix factorization. In: Proc. of the 20th Int'l Conf. on Neural Information Processing Systems. Vancouver: ACM, 2008. 1257–1264. [doi: 10.5555/2981562.2981720]
- [84] He XN, Liao LZ, Zhang HW, Nie LQ, Chua TS. Neural collaborative filtering. In: Proc. of the 26th Int'l Conf. on World Wide Web. Perth: ACM, 2017. 173–182. [doi: 10.1145/3038912.3052569]
- [85] Li LH, Chu W, Langford J, Schapire RE. A contextual-bandit approach to personalized news article recommendation. In: Proc. of the 19th Int'l Conf. on World Wide Web. Raleigh: ACM, 2010. 661–670. [doi: 10.1145/1772690.1772758]
- [86] Zeng CQ, Wang Q, Mokhtari S, Tao L. Online context-aware recommendation with time varying multi-armed bandit. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. San Francisco: ACM, 2016. 2025–2034. [doi: 10.1145/2939672.2939878]
- [87] Chapelle O, Li LH. An empirical evaluation of thompson sampling. In: Proc. of the 24th Int'l Conf. on Neural Information Processing Systems. Granada: Curran Associates Inc., 2011. 2249–2257. [doi: 10.5555/2986459.2986710]
- [88] Cheng HT, Koc L, Harmsen J, Shaked T, Chandra T, Aradhye H, Anderson G, Corrado G, Chai W, Ispir M, Anil R, Haque Z, Hong LC, Jain V, Liu XB, Shah H. Wide & deep learning for recommender systems. In: Proc. of the 1st Workshop on Deep Learning for Recommender Systems. Boston: ACM, 2016. 7–10. [doi: 10.1145/2988450.2988454]
- [89] Hu YJ, Da Q, Zeng AX, Yu Y, Xu YH. Reinforcement learning to rank in E-commerce search engine: Formalization, analysis, and application. In: Proc. of the 24th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. London: ACM, 2018. 368–377. [doi: 10.1145/3219819.3219846]
- [90] Zheng GJ, Zhang FZ, Zheng ZH, Yang X, Yuan NJ, Xie X, Li ZH. DRN: A deep reinforcement learning framework for news recommendation. In: Proc. of the 2018 World Wide Web Conf. Lyon: Republic and Canton of Geneva, 2018. 167–176. [doi: 10.1145/3178876.3185994]
- [91] Zhao XY, Xia L, Zhang L, Ding ZY, Yin DW, Tang JL. Deep reinforcement learning for page-wise recommendations. In: Proc. of the 12th ACM Conf. on Recommender Systems. Vancouver: ACM, 2018. 95–103. [doi: 10.1145/3240323.3240374]
- [92] Mahmood T, Ricci F. Learning and adaptivity in interactive recommender systems. In: Proc. of the 9th Int'l Conf. on Electronic Commerce: The Wireless World of Electronic Commerce. Minneapolis: ACM, 2007. 75–84. [doi: 10.1145/1282100.1282114]
- [93] Yu X, Ren X, Sun YZ, Gu QQ, Sturt B, Khandelwal U, Norick B, Han JW. Personalized entity recommendation: A heterogeneous information network approach. In: Proc. of the 7th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM, 2014. 283–292. [doi: 10.1145/2556195.2556259]
- [94] Zhang FZ, Yuan NJ, Lian DF, Xie X, Ma WY. Collaborative knowledge base embedding for recommender systems. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data. San Francisco: ACM, 2016. 353–362. [doi: 10.1145/2939672.2939673]
- [95] Wang HW, Zhang FZ, Wang JL, Miao Z, Guo M. RippleNet: Propagating user preferences on the knowledge graph for recommender systems. In: Proc. of the 27th ACM Int'l Conf. on Information and Knowledge Management. Torino: ACM, 2018. 417–426. [doi: 10.1145/3269206.3271739]
- [96] Ge Y, Chen SC. Graph convolutional network for recommender systems. Ruan Jian Xue Bao/Journal of Software, 2020, 31(4): 1101–1112 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4923.htm> [doi: 10.13328/j.cnki.jos.005928]
- [97] Ai L, Zhao H. A user interaction model for knowledge-based recommender system. Software Guide, 2015, 14(3): 15–17 (in Chinese with English abstract). [doi: 10.11907/tjdk.143999]
- [98] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907, 2016.
- [99] Schlichtkrull M, Kipf TN, Bloem P, van den Berg R, Titov I, Welling M. Modeling relational data with graph convolutional networks. arXiv:1703.06103, 2018.
- [100] Misu T, Sugiura K, Ohtake K, Hori C, Kashioka H, Kawai H, Nakamura S. Modeling spoken decision making dialogue and optimization of its dialogue strategy. In: Proc. of the 2010 SIGDIAL Conf. Tokyo: Association for Computational Linguistics, 2010. 221–224.
- [101] Duan Y, Schulman J, Chen X, Bartlett PL, Sutskever I, Abbeel P. RL²: Fast reinforcement learning via slow reinforcement learning. arXiv:1611.02779v2, 2016.
- [102] Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, Riedmiller M. Playing Atari with deep reinforcement learning. arXiv:1312.5602, 2013.
- [103] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proc. of the 34th Int'l Conf. on

- Machine Learning. Sydney: ACM, 2017. 1126–1135. [doi: 10.5555/3305381.3305498]
- [104] Su SY, Li XJ, Gao JF, Liu JJ, Chen YN. Discriminative deep dyna-Q: Robust planning for dialogue policy learning. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing. 2018.
- [105] Abel D, Salvatier J, Stuhlmüller A, Evans O. Agent-agnostic human-in-the-loop reinforcement learning. arXiv:1701.04079, 2017.
- [106] Zhou FF, Gao F, Liu YG, Liang X, Zhao Y. Interactive volume data classification based on density-distance graph. Ruan Jian Xue Bao/Journal of Software, 2016, 27(5): 1061–1073 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4961.htm> [doi: 10.13328/j.cnki.jos.004961]
- [107] Ross S, Gordon GJ, Bagnell JA. A reduction of imitation learning and structured prediction to no-regret online learning. arXiv:1011.0686v3, 2011.
- [108] Wu YX, Li XJ, Liu JJ, Gao JF, Yang YM. Switch-based active deep dyna-Q: Efficient adaptive planning for task-completion dialogue policy learning. Proc. of the AAAI Conf. on Artificial Intelligence, 2019, 33(1): 7289–7296. [doi: 10.1609/aaai.v33i01.33017289]
- [109] Zhang ZR, Li XJ, Gao JF, Chen EH. Budgeted policy learning for task-oriented dialogue systems. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 3742–3751. [doi: 10.18653/v1/P19-1364]
- [110] Zhang S, Balog K. Evaluating conversational recommender systems via user simulation. In: Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. ACM, 2020. 1512–1520. [doi: 10.1145/3394486.3403202]
- [111] Schatzmann J, Thomson B, Weilhammer K, Ye H, Young S. Agenda-based user simulation for bootstrapping a POMDP dialogue system. In: Proc. of the Human Language Technologies 2007: The Conf. of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers. Rochester: NAACL, 2007. 149–152.
- [112] Zhou K, Wang XL, Zhou YH, Shang CZ, Cheng Y, Zhao WX, Li YL, Wen JR. CRSLab: An open-source toolkit for building conversational recommender system. In: Proc. of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th Int'l Joint Conf. on Natural Language Processing: System Demonstrations. Association for Computational Linguistics, 2021. 185–193. [doi: 10.18653/v1/2021.acl-demo.22]
- [113] Su PH, Budzianowski P, Ultes S, Gašić M, Young S. Sample-efficient actor-critic reinforcement learning with supervised data for dialogue management. In: Proc. of the 18th Annual SIGdial Meeting on Discourse and Dialogue. Saarbrücken: Association for Computational Linguistics, 2017. 147–157. [doi: 10.18653/v1/W17-5518]
- [114] Ma WC, Takanobu R, Tu MH, Huang ML. Bridging the gap between conversational reasoning and interactive recommendation. arXiv:2010.10333v1, 2020.
- [115] Cui C, Wang WJ, Song XM, Huang ML, Xu XS, Nie LQ. User attention-guided multimodal dialog systems. In: Proc. of the 42nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Paris: ACM, 2019. 445–454. [doi: 10.1145/3331184.3331226]
- [116] Zhang Z, Liao LZ, Huang ML, Zhu XY, Chua TS. Neural multimodal belief tracker with adaptive attention for dialogue systems. In: Proc. of the World Wide Web Conf. San Francisco: ACM, 2019. 2401–2412. [doi: 10.1145/3308558.3313598]
- [117] Pittermann J, Pittermann A, Minker W. Emotion recognition and adaptation in spoken dialogue systems. Int'l Journal of Speech Technology, 2010, 13(1): 49–60. [doi: 10.1007/s10772-010-9068-y]
- [118] Steck H. Item popularity and recommendation accuracy. In: Proc. of the 5th ACM Conf. on Recommender Systems. Chicago: ACM, 2011. 125–132. [doi: 10.1145/2043932.2043957]
- [119] Abdollahpouri H, Mansoury M. Multi-sided exposure bias in recommendation. In: Proc. of the 2020 ACM KDD Int'l Workshop on Industrial Recommendation Systems. Chicago: ACM, 2020.
- [120] Zhang YF, Zhang HC, Zhang M, Liu YQ, Ma SP. Do users rate or review?: Boost phrase-level sentiment labeling with review-level sentiment classification. In: Proc. of the 37th Int'l ACM SIGIR Conf. on Research & Development in Information Retrieval. Gold Coast: ACM, 2014. 1027–1030. [doi: 10.1145/2600428.2609501]
- [121] Liu YM, Cao XZ, Yong Y. Are you influenced by others when rating?: Improve rating prediction by conformity modeling. In: Proc. of the 10th ACM Conf. on Recommender Systems. Boston: ACM, 2016. 269–272. [doi: 10.1145/2959100.2959141]
- [122] Huang J, Oosterhuis H, de Rijke M, van Hoof H. Keeping dataset biases out of the simulation: A debiased simulator for reinforcement learning based recommender systems. In: Proc. of the 14th ACM Conf. on Recommender Systems. ACM, 2020. 190–199. [doi: 10.1145/3383313.3412252]
- [123] Pradel B, Usunier N, Gallinari P. Ranking with non-random missing ratings: Influence of popularity and positivity on evaluation metrics. In: Proc. of the 6th ACM Conf. on Recommender Systems. Dublin: ACM, 2012. 147–154. [doi: 10.1145/2365952.2365982]
- [124] Steck H. Evaluation of recommendations: Rating-prediction and ranking. In: Proc. of the 7th ACM Conf. on Recommender Systems. Hong Kong: ACM, 2013. 213–220. [doi: 10.1145/2507157.2507160]

- [125] Hernández-Lobato JM, Houlby N, Ghahramani Z. Probabilistic matrix factorization with non-random missing data. In: Proc. of the 31st Int'l Conf. on Machine Learning. Beijing: PMLR, 2014. 1512–1520.
- [126] Marlin BM, Zemel RS, Roweis S, Slaney M. Collaborative filtering and the missing at random assumption. In: Proc. of the 23rd Conf. on Uncertainty in Artificial Intelligence. Vancouver: AUAI, 2007. 267–275. [doi: 10.5555/3020488.3020521]
- [127] Aman A, Kenta T, Ivan Z, Thorsten J. A general framework for counterfactual learning-to-rank. In: Proc. of the 42nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2019. 5–14.
- [128] Sun WL, Khenissi S, Nasraoui O, Shafto P. Debiasing the human-recommender system feedback loop in collaborative filtering. In: Proc. of the 2019 World Wide Web Conf. San Francisco: ACM, 2019. 645–651. [doi: 10.1145/3308560.3317303]
- [129] Sinha A, Gleich DF, Ramani K. Deconvolving feedback loops in recommender systems. In: Proc. of the 30th Conf. on Neural Information Processing Systems 29 (NIPS 2016). Barcelona: NIPS, 2016.
- [130] Cai WL, Chen L. Predicting user intents and satisfaction with dialogue-based conversational recommendations. In: Proc. of the 28th ACM Conf. on User Modeling, Adaptation and Personalization. Genoa: ACM, 2020. 33–42. [doi: 10.1145/3340631.3394856]
- [131] Ma C, Kang P, Liu X. Hierarchical gating networks for sequential recommendation. In: Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. Anchorage: Association for Computing Machinery, 2019. 825–833. [doi: 10.1145/3292500.3330984]
- [132] Wang W, Zhang W, Liu SK, Liu Q, Zhang B, Lin LY, Zha HY. Beyond clicks: Modeling multi-relational item graph for session-based target behavior prediction. In: Proc. of the 2020 Web Conf. Taipei: ACM, 2020. 3056–3062. [doi: 10.1145/3366423.3380077]
- [133] Wang X, Xu YK, He XN, Cao YX, Wang M, Chua TS. Reinforced negative sampling over knowledge graph for recommendation. In: Proc. of the 2020 Web Conf. Taipei: ACM, 2020. 99–109. [doi: 10.1145/3366423.3380098]
- [134] Chen JW, Wang C, Zhou S, Shi QH, Feng Y, Chen C. Samwalker: Social recommendation with informative sampling strategy. In: Proc. of the 2019 World Wide Web Conf. San Francisco: ACM, 2019. 228–239. [doi: 10.1145/3308558.3313582]
- [135] Wang X, Jin HY, Zhang A, He XN, Xu T, Chua TS. Disentangled graph collaborative filtering. In: Proc. of the 43rd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. ACM, 2020. 1001–1010. [doi: 10.1145/3397271.3401137]
- [136] Ma JX, Zhou C, Cui P, Yang HX, Zhu WW. Learning disentangled representations for recommendation. arXiv:1910.14238v1, 2019.
- [137] Zhang XY, Xie H, Li H, Lui JCS. Conversational contextual bandit: Algorithm and application. In: Proc. of the Web Conf. 2020. Taipei: ACM, 2020. 662–672. [doi: 10.1145/3366423.3380148]
- [138] Shao YW, Zhang M, Ma WZ, Wang CY, Liu YQ, Ma SP. Integrating latent item-item complementarity with personalized recommendation systems. Ruan Jian Xue Bao/Journal of Software, 2020, 31(4): 1090–1100 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5925.htm> [doi: 10.13328/j.cnki.jos.005925]
- [139] Liu WW, Liu Q, Tang RM, Chen JY, He XQ, Heng PA. Personalized Re-ranking with Item Relationships for E-commerce. In: Proc. of the 29th ACM Int'l Conf. on Information & Knowledge Management. ACM, 2020. 925–934. [doi: 10.1145/3340531.3412332]
- [140] McAuley J, Pandey R, Leskovec J. Inferring networks of substitutable and complementary products. In: Proc. of the 21st ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Sydney: ACM, 2015. 785–794. [doi: 10.1145/2783258.2783381]
- [141] Wan MT, Wang D, Liu J, Bennett P, McAuley J. Representing and recommending shopping baskets with complementarity, compatibility and loyalty. In: Proc. of the 27th ACM Int'l Conf. on Information and Knowledge Management. Torino: ACM, 2018. 1133–1142. [doi: 10.1145/3269206.3271786]
- [142] Wu W, Yan R. Deep chit-chat: Deep learning for chatbots. In: Proc. of the 42nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Melbourne: Association for Computational Linguistics, 2019. 1413–1414. [doi: 10.1145/3331184.3331388]
- [143] Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 2017, 34(6): 26–38. [doi: 10.1109/MSP.2017.2743240]
- [144] Ng AY, Russell SJ. Algorithms for inverse reinforcement learning. In: Proc. of the 17th Int'l Conf. on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc., 2000. 663–670. [doi: 10.5555/645529.657801]
- [145] Wang JX, Kurth-Nelson Z, Tirumala D, Soyer H, Leibo JZ, Munos R, Blundell C, Kumaran D, Botvinick M. Learning to reinforcement learn. arXiv:1611.05763, 2016.

附中文参考文献:

- [14] 杨淑梅. 推荐系统交互性研究综述. 软件, 2013, 34(8): 154–158. [doi: 10.3969/j.issn.1003-6970.2013.08.059]
- [15] 杨智强, 殷钊, 王衡. 结合用户交互行为和资源内容的资源推荐. 计算机辅助设计与图形学学报, 2014, 26(5): 747–754.
- [58] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法. 软件学报, 2003, 14(9): 1621–1628. <http://jos.org.cn/jos/article/>

- abstract/20030917?st=article_issue [doi: 1000-9825/2003/14(09)1621]
- [59] 黄震华, 张佳雯, 田春岐, 孙圣力, 向阳. 基于排序学习的推荐算法研究综述. 软件学报, 2016, 27(3): 691–713. <http://www.jos.org.cn/1000-9825/4948.htm> [doi: 10.13328/j.cnki.jos.004948]
- [79] 陈碧毅, 黄玲, 王昌栋, 景丽萍. 融合显式反馈与隐式反馈的协同过滤推荐算法. 软件学报, 2020, 31(3): 794–805. <http://www.jos.org.cn/1000-9825/5897.htm> [doi: 10.13328/j.cnki.jos.005897]
- [80] 张青博, 王斌, 崔宁宁, 宋晓旭, 秦婧. 基于注意力机制的规范化矩阵分解推荐算法. 软件学报, 2020, 31(3): 778–793. <http://www.jos.org.cn/1000-9825/5913.htm> [doi: 10.13328/j.cnki.jos.005913]
- [81] 杜东航, 徐童, 鲁亚男, 管楚, 刘淇, 陈恩红. 基于信任机制下概率矩阵分解的用户评分预测. 软件学报, 2018, 29(12): 3747–3763. <http://www.jos.org.cn/1000-9825/5322.htm> [doi: 10.13328/j.cnki.jos.005322]
- [82] 刘华锋, 景丽萍, 于剑. 融合社交信息的矩阵分解推荐方法研究综述. 软件学报, 2018, 29(2): 340–362. <http://www.jos.org.cn/1000-9825/5391.htm> [doi: 10.13328/j.cnki.jos.005391]
- [96] 葛尧, 陈松灿. 面向推荐系统的图卷积网络. 软件学报, 2020, 31(4): 1101–1112. <http://www.jos.org.cn/1000-9825/5928.htm> [doi: 10.13328/j.cnki.jos.005928]
- [97] 艾磊, 赵辉. 基于知识的推荐系统用户交互模型研究. 软件导刊, 2015, 14(3): 15–17. [doi: 10.11907/rjdk.143999]
- [106] 周芳芳, 高飞, 刘勇刚, 梁兴, 赵颖. 基于密度-距离图的交互式体数据分类方法. 软件学报, 2016, 27(5): 1061–1073. <http://www.jos.org.cn/1000-9825/4961.htm> [doi: 10.13328/j.cnki.jos.004961]
- [138] 邵英玮, 张敏, 马为之, 王晨阳, 刘奕群, 马少平. 融合商品潜在互补性发现的个性化推荐方法. 软件学报, 2020, 31(4): 1090–1100. <http://www.jos.org.cn/1000-9825/5925.htm> [doi: 10.13328/j.cnki.jos.005925]



赵梦媛(1997—), 女, 硕士, 主要研究领域为数据挖掘, 推荐系统, 用户建模.



桑基韬(1985—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为机器学习与认知计算, 人工智能及应用.



黄晓雯(1993—), 女, 博士, 讲师, CCF 专业会员, 主要研究领域为多媒体计算, 数据挖掘, 用户建模, 推荐系统.



于剑(1969—), 男, 博士, 教授, 博士生导师, CCF 会士, 主要研究领域为机器学习, 计算智能, 图像分析, 数据挖掘.