基于负反馈修正的多轮对话推荐系统

朱立玺 黄晓雯 赵梦媛 桑基韬

(北京交通大学计算机与信息技术学院 北京 100044)

(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)

摘 要 传统的推荐系统从交互历史中挖掘用户兴趣,面临着无法动态地获取用户实时偏好和细粒度偏好的问题,近年对话推荐系统领域的兴起为此问题提供了新的解决方案.对话推荐系统优势在于其可以动态地和用户进行交互,并在交互过程中获取用户的实时偏好,从而提高推荐系统准确率,提升用户体验.然而对话推荐系统相关研究工作中缺乏对负反馈的充分利用,难以对用户偏好表示进行细粒度的修正,即难以有效平衡用户长期偏好和实时偏好之间的关系,同时存在属性候选集过大导致交互轮次过多的问题.因此,本文基于经典的对话推荐框架 CPR(Conversational Path Reasoning)提出了一种能够有效利用用户负反馈的对话推荐模型 NCPR(Negative-feedback-guide Conversational Path Reasoning).不同于现有的对话推荐系统工作,NCPR 能够充分利用用户在交互过程中给出的属性粒度和物品粒度的负反馈对用户的偏好表示进行动态的修正.此外,CPR 将对话推荐建模为一个图上的路径推理问题,NCPR 使用协同过滤算法基于属性粒度的负反馈对属性候选集进行重排序,在利用图结构的自然优势限制属性候选集大小的同时,进一步减少候选属性空间大小.四个基准数据集上的实验结果表明,NCPR 在推荐准确率和平均交互轮次两个评价指标上的表现优于先进的基线模型.最后,我们设计并实现了一个网页端的对话推荐系统,与在线用户进行交互产生推荐结果,证明了 NCPR 在真实的对话推荐场景下的有效性.

关键词 对话推荐系统;强化学习;交互负反馈;知识图谱;协同过滤

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2023.01086

Multi-Round Conversational Recommendation System Based on Negative Feedback Correction

ZHU Li-Xi HUANG Xiao-Wen ZHAO Meng-Yuan SANG Ji-Tao

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

(*Beijing Key Lab of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing* 100044)

Abstract Traditional recommendation systems can only estimate user preferences and model user preferences on items from past interaction history, thus suffering from the limitations of obtaining real-time and fine-grained user preferences. Conversational recommendation systems (CRS), which introduce conversational technology into recommendation systems, provide a new solution to this problem in recent years. Unlike traditional recommendation systems which, due to their static way of working, cannot answer what the user's current preferences are and what the user's reasons are for buying an item, CRS are interactive recommendation systems, which have the advantage of dynamically interacting with users and obtaining their real-time preferences in the process of interaction, understanding what they currently like and what their reasons are for liking an item, thus allowing the recommendation system to quickly understand the user's intent and make recommendations that the user desires, enhancing the user's experience and trust in the

收稿日期:2022-02-10;在线发布日期:2022-10-08.本课题得到中央高校基本科研专项资金(2021RC217)、国家自然科学基金(62202041)资助.**朱立玺**,硕士研究生,主要研究方向为用户建模和推荐系统.E-mail:21120456@bjtu.edu.cn.**黄晓雯**(通信作者),博士,讲师,主要研究方向为多媒体计算、数据挖掘、用户建模、推荐系统等.E-mail:xwhuang@bjtu.edu.cn.**赵梦媛**,硕士研究生,主要研究方向为用户建模和推荐系统.**桑基韬**,博士,教授,博士生导师,主要研究领域为机器学习与认知计算、人工智能及应用.

system. However, current CRS studies more often uses positive feedback on the granularity of attributes given by the user to model the user's real-time preferences, ignoring the impact of negative user feedback on modeling the user's real-time preferences, but negative feedback is an important part of user feedback and can also indicate real-time user preferences, making it difficult to make fine-grained corrections to the user preference representation, which means it is difficult to effectively balance the relationship between users' long-term preference and real-time preference. At the same time, current state-of-the-art work in the field only takes advantage of the natural advantages of graph structure to limit the size of the attribute candidate set, which suffers from the problem of too many interaction rounds due to the large attribute candidate set. To address the problems mentioned above, we propose a conversational recommendation model NCPR based on the classical conversational recommendation framework CPR, which can make full use of all feedback given by the user during the interaction to correct the user's preference representation, including positive feedback at the attribute granularity, negative feedback at the attribute granularity, and negative feedback at the item granularity. In addition, CPR models conversation recommendation as a path inference problem on a graph, i. e. CRS can only select attribute nodes adjacent to the current node to ask the user. This approach helps CRS to limit the size of the attribute candidate set in a single round of decision making. NCPR uses an attribute-based collaborative filtering algorithm to reorder the attribute candidate set based on negative feedback from the attribute granularity, i. e. removing those attributes in the attribute candidate set that are similar to those rejected by the user, which can further reduce the candidate attribute space size while taking advantage of the natural advantages of graph structure to limit the size of the attribute candidate set. Experimental results on four benchmark CRS datasets show that the proposed method significantly outperforms state-of-the-art baselines in terms of both evaluation metrics. In addition, we design and implement a web-side conversational recommendation system that interacts with online users to generate recommendation results, demonstrating the effectiveness of NCPR in a real-world conversational recommendation scenario.

Keywords　conversational recommendation systems; reinforcement learning; interactive negative feedback; knowledge graph; collaborative filtering

1　引　言

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，近年来引起了学术界的高度关注，也为工业界带来了巨大收益. 然而传统推荐系统的工作方式是静态的，传统推荐系统中的推荐模型[1-4]基于离线的、历史的用户交互数据作为输入以不断优化线下的表现. 这种工作方式导致传统的推荐系统存在一定的缺陷，即传统的推荐系统很难回答两个问题[5]：(1) 用户当前的兴趣点(实时偏好)在哪里？因为用户的偏好具有多样性，并且会随着时间的推移不断发生变化，因而推荐系统即便有足够多的历史数据也很难拟合用户的线上行为并准确推测出用户当前的兴趣偏好；(2) 用户购买这个物品的理由是什么？例如，用户可能经过朋友的大力推荐购买了这个物品，也有可能是第一次见到这种物品，在好奇心的驱使下下了订单. 对于一个用户来说，不同理由下的购买行为，其表现出的对物品的喜爱程度是不同的.

学者们尝试引入社交网络等辅助信息来回答上面的两个问题，但受限于传统推荐系统固有的静态工作方式，同时引入的辅助信息也存在噪声，这两个问题并没有被有效地解决[5]. 问答系统中对话技术的发展[6-8]为这两个问题提供了新的解决思路. 推荐系统中引入对话技术后可以动态地和用户进行交互，获取用户的实时反馈，推断出用户的实时偏好并理解用户行为的动机[5]，进而向用户做出心仪的推荐. 学者们首先提出了交互式推荐系统的概念，即通过询问物

品的方式来获取用户的实时反馈,相关研究[9-11]取得了初步成功.然而物品的数量是非常多的,因此交互式推荐系统的效率很低,即使有相关工作[10,12]针对此问题提出了一些解决方案,比如询问物品集合,引入树结构等,但结果依旧不尽如人意.因此学者们进一步提出了对话推荐系统(Conversational Recommendation System,CRS),希望通过询问物品相关的属性来获取用户更细粒度的反馈.

对话推荐系统是近年来的一个新兴领域,可以根据问题的设置分为不同的研究方向,比如利用多臂老虎机解决用户冷启动,对话理解和对话生成,基于强化策略推荐等,本文侧重于基于强化策略进行物品推荐.在探索的过程中学者们提出了不同的场景设置,如单轮对话推荐(Single-Round Conversational Recommendation,SCR)和多轮对话推荐(Multi-Round Conversational Recommendation,MCR).在单轮对话推荐场景[13]中,CRS 可以向用户询问多次属性,但只能做出一次推荐,不论推荐是否被用户所接受,这轮对话都会结束.与单轮对话推荐场景设置相比,多轮对话推荐场景下的 CRS 可以对用户进行多次关于物品属性的询问并且可以做出多次推荐直到推荐列表中含有用户心仪的物品或者用户失去耐心选择退出.多轮对话推荐场景下 CRS 可以同时获取属性粒度和物品粒度的反馈,并利用这些在线反馈来建模用户实时偏好,从而做出符合用户个性化的推荐,更贴近现实场景.尽管当前的 CRS 工作已经取得了初步成功,但是它们对负反馈的使用方式比较简单[14-15].例如,EAR[14]中将属性粒度的负反馈编码成因子分解机的输入特征向量,并将物品粒度的负反馈作为训练的负样本用于在线更新推荐模型;在 SCPR[15]中,属性粒度的负反馈被编码为决策网络的状态向量,物品粒度负反馈的处理为仅仅将这些物品从物品候选集中移除.这些方法建模用户的实时偏好时仅考虑用户的正反馈,并没有充分挖掘负反馈中隐藏的有价值的信息.工业界的一些工作已经证明了用户的负反馈可以有效地帮助模型构建用户画像.例如,京东团队的 Zhao 等人[16]提出了一种强化学习模型 DEERS,该模型基于 Pairwise 将用户的负反馈运用到了推荐当中,并在真实的电商场景中取得了不错的效果;此外,微信团队的 Xie 等人[17]提出了一种可以有效融合显/隐式正反馈和显/隐式负反馈的模型 DFN,该模型可以学习到用户无偏的兴趣偏好.因此,如何充分利用交互过程中用户的负反馈是个值得关注的问题.相较于传统的推荐系统,对话推荐系统的一大优势在于能够获取用户的属性粒度和物品粒度的实时反馈,来建模用户的实时偏好.因此,属性粒度的负反馈和物品粒度的负反馈对于平衡用户长期偏好和实时偏好之间的关系有着重要作用.

同时我们注意到,当前 CRS 工作中存在属性候选集过大导致交互轮次过多的问题.EAR 将整个属性候选集作为强化模型的决策空间,当属性候选集较大时模型很难学习到有效的决策策略.SCPR 引入知识图谱并限制强化模型的决策空间为 2,将选择属性的任务交给属性打分模型,可以有效缓解强化模型难收敛的问题.但 SCPR 仅利用图结构的自然优势在单轮决策中删除了部分不相关的属性,在接收用户属性粒度的负反馈后并没有对属性候选集的大小进行限制,只是将询问的属性从属性候选集中移除.UNICORN[18]采用了 SCPR 的做法,并没有对属性候选集的大小进行额外的限制,由于图结构只能在单轮决策中对候选集大小进行限制,随着交互的进行,属性候选集的大小可能不降反增,这种做法增加了属性打分模型选择属性的难度.

我们将用户偏好修正[19]和协同过滤[2]的思想应用到 CPR 中,提出一种称为 NCPR(Negative-feedback-guide Conversational Path Reasoning)的对话推荐模型.具体而言,NCPR 与用户交互时,利用交互过程中属性的正反馈、属性和物品的负反馈对用户的偏好表示进行修正,目的是随着交互轮次的增加,让模型更加关注用户的实时偏好.NCPR 收到用户的属性粒度的负反馈时,会基于收到的属性负反馈进行协同过滤对属性候选集进行重排序,在利用图结构的自然优势删除部分不相关的候选属性的同时,进一步减少候选属性空间大小.我们在四个基准数据集上进行了对比实验、消融实验和分析实验,实验结果表明,相较于先进的基线模型,NCPR 在推荐准确率和平均交互轮次两个指标上均取得了最好表现.此外,我们设计并实现了一个网页端的对话推荐系统,可以实时地和用户进行交互并做出推荐,证明了 NCPR 在真实场景下的有效性.

本文的主要贡献如下:

(1)利用用户的在线反馈对用户的偏好表示进行动态的修正,随着交互轮次的增加,让模型更加关注于用户的实时偏好,帮助模型在知识图谱上更好地推理,进而做出用户心仪的推荐.

(2)使用基于属性的协同过滤模型实现对候选集属性的重排序，在利用图结构的自然优势删除部分不相关候选属性的同时，进一步减少候选属性空间大小.

(3)在四个对话推荐系统领域的基准数据集上进行了多组实验，实验结果表明所提方法在推荐准确率及平均交互轮次两种评价指标上显著优于先进的基线模型. 同时设计并实现了一个网页端的对话推荐系统，证明了 NCPR 在真实的对话推荐场景下的有效性. 访问网址 https://github.com/zlxxlz1026/NCPR 可查看系统实现和视频展示.

2　相关工作

2.1　基于知识图谱的推荐系统

近年来，知识图谱作为一种辅助信息被广泛用于个性化推荐中[20]. 一种将知识图谱用于推荐系统的方式为基于嵌入的方法. Wang 等人[21]提出了利用 KCNN 融合新闻语义层和知识层嵌入向量表示的 DKN 模型；Huang 等人[22]提出了从交互序列中捕获用户细粒度偏好的知识增强序列推荐模型 KSR；Socher 等人[23]提出了推理两实体间关系的神经张量网络模型 NTN 等. 另一种将知识图谱用于推荐系统的方式为基于路径的方法. 这类工作在初期将知识图谱视为异构信息网络(Heterogeneous Information Network，HIN)[24]. Huang等人提出的 MASR 模型[25]、EIUM 模型[26]；Yu 等人提出的 Hete-MF 模型[27]、HeteRec 模型[28]、HeteRec-p 模型[29]；Shi等人[30]提出的 SemRec 模型；Hu 等人[31]提出的 MCRec 模型；Zhao 等人[32]提出的 FMG 模型等，这类模型通过构造元路径(meta-path)或元图(meta-graph)的方式帮助模型学习更高级的语义信息[24].

2.2　对话推荐系统

学者们在 CRS 领域的不同方向中做了不同的探索和尝试.

2.2.1　多臂老虎机解决用户冷启动

对话推荐系统的一个优势在于可以解决用户冷启动的问题，但是研究发现多臂老虎机在用户冷启动场景下效果显著，越来越多的工作基于多臂老虎机期望达到探索和收益的平衡(E&E). 比如，Zhang 等人[33]认为传统的上下文多臂老虎机算法[34]将物品建模为臂进行探索实现 E&E，由于物品空间很大，因此需要大量探索，导致其学习速度很慢. 为了提高模型学习效率，他们提出对话多臂老虎机，(Conversational UCB algorithm，ConUCB)将属性询问整合进系统；Li 等人[35]认为简单的启发式函数不能找到属性询问的最佳时机，因此他们提出对话汤姆森采样(Conversation Thompson Sampling)，将属性和物品建模为无差别的臂，这种无缝统一属性和物品的方式，可以以一种智能的方式决定何时询问或推荐，同时利用汤姆森采样自然地达到探索和收益相平衡的目标；Xie 等人[36]认为现有的对话推荐系统要求用户给出确切的反馈，这种情况下用户给出的反馈往往带有偏差，因此他们提出了一种新的多臂老虎机算法 RelativeConUCB，来构建一个基于相对偏好的对话推荐系统.

2.2.2　对话理解和对话生成(NLP-based)

该类对话推荐系统希望通过自然语言来理解用户意图并和用户进行动态地交互，通常会引入 NLP 领域的一些指标来评价模型效果，比如引入 BLEU 来判断对话生成的流畅程度. Zhang 等人[37]尝试为对话推荐问题定义一个标准形式，提出了一个对话搜索引擎 SAUR；Li 等人[38]针对电影推荐的冷启动场景设计了一个可以与用户通过自然语言交互的对话系统，但该系统只考虑用户当前的对话状态而忽略了用户的交互历史，因此无法学习到用户长期的偏好；Zhou 等人[39]认为自然语言的表示和物品级别的用户偏好之间存在语义差异，因此采用互信息最大化(Mutual Information Maximization)手段来实现词级别和实体级别语义空间的对齐，在推荐和对话两个任务上均取得了不错的表现；Lu 等人[40]认为当前的对话推荐系统无法从简短的对话历史中获得充分的项目信息，因此提出了一个融入评论信息的对话推荐系统(Review-augmented Conversational Recommender，RevCore)，可以利用评论来丰富物品信息，并协助生成流畅的回复；Liang 等人[41]认为已有的工作无法将待推荐物品合适地融入生成的回复中，因此提出名为 NTRD 的对话推荐系统，将对话生成和物品推荐分离；Zhou 等人[42]认为对话数据与外部数据之间存在差异，使得利用多类型数据变得困难. 因此提出一个由粗到细的对比学习框架 C2CRS 来促进多类型数据间的语义融合.

2.2.3　基于强化策略的推荐(RL-based)

用户通常需要遵循固定的形式来和该类对话推荐系统进行交互，比如九宫格选择、是或否等. 这类

工作重点关注利用用户的反馈来生成更加符合用户个性化的推荐列表. 为了帮助模型学习到用户的长期偏好,Christakopoulou 等人[43]将用户的交互历史作为系统的输入,设计了一个用户引导页形式的对话推荐系统 Q&R,然而该系统只能和用户进行单轮交互;Sun 等人[13]提出了可以和用户多轮交互的对话推荐系统 CRM,但是 CRM 适用于单轮对话推荐场景,即 CRM 向用户做出推荐后,不论用户是否满意这次推荐,CRM 都会退出,这种场景是不贴近现实的[14];Lei 等人[14]提出了一种适用于多轮对话推荐场景的对话推荐系统 EAR. EAR 解决了和用户进行交互的三个基本问题:(1)每轮对话询问用户什么属性?(2)什么时候进行物品的推荐?(3)如何接受用户的反馈?虽然 EAR 取得了不错的表现,但仍存在物品候选集过大及决策空间大的问题. 为此,Lei 等人[15]将知识图谱引入对话推荐系统,提出了一种新的对话推荐框架 CPR,并在此框架的基础上实现了一个称为 SCPR 的对话推荐系统,他们的工作证明了引入知识图谱可以有效地提升对话推荐系统的准确率,并使得推荐结果更具解释性;Ren 等人[44]认为对话推荐系统的决策模型需要大量的语料进行训练,以保证涵盖所有情况. 基于此问题,他们提出了基于知识的问题生成系统(Knowledge-Based Question Generation System, KBQG);Deng 等人[18]认为 EAR、SCPR 等工作将上面提到的三个决策问题分到多个模块完成,对模型的可扩展性有影响. 因此他们基于动态权重图提出了一个新的对话推荐系统 UNICORN(UNIfied COnversational RecommeNder),使用统一的对话策略来解决三个决策问题.

3 NCPR

3.1 交互流程

本文主要关注多轮对话推荐场景. 具体来说,每个物品 v 都有一组由数据集提供的属性集合 $\mathcal{P}$,$\mathcal{P}$ 中的属性 p 可以是对任意一个物品 v 的描述(标签). 在用于歌手推荐的 LastFM 数据集中,物品 v 表示一个歌手/乐队,属性 p 就是这个歌手/乐队的标签,比如流行音乐、古典音乐、爵士乐等;在用于商户推荐的 Yelp 数据集中,物品 v 表示一家商户,属性 p 就是这家商户的标签,比如快餐店、星级饭店等. 用户通过指定一个心仪的属性 p_0 来开启对话,比如我想找一家五星级饭店,则指定五星级饭店为 p_0,此时用户正反馈集合初始化为 $\mathcal{P}_u^+=\{p_0\}$,属性候选集初始化为 $\mathcal{P}_{\text{cand}}=\mathcal{P}\setminus p_0$,物品候选集初始化为 $\mathcal{V}_{\text{cand}}=\mathcal{V}_{p_0}$,即包含属性 p_0 的物品集合. 接下来的每轮交互中,NCPR 从动作集合 $\{a_{\text{ask}},a_{\text{rec}}\}$ 中选择一个动作:(1)当选择的动作为 a_{ask},NCPR 从属性候选集 $\mathcal{P}_{\text{cand}}$ 中选择得分最高的属性向用户询问,用户给出属性粒度的正/负反馈. 对于属性粒度的正反馈,NCPR 将该属性加入用户正反馈属性集 $\mathcal{P}_u^+$,然后更新属性候选集和物品候选集. 对于属性粒度的负反馈,NCPR 将该属性加入用户负反馈属性集 $\mathcal{P}_u^-$,然后基于该属性对属性候选集进行协同过滤,删除属性候选集中与该属性相似度较高的那些属性,缩小属性候选集的大小,相当于对属性候选集做一个重排序;(2)当选择的动作为 a_{rec},NCPR 利用交互过程中用户给出的所有反馈对用户的偏好表示进行一个细粒度的修正,即平衡用户长期偏好和实时偏好之间的关系. 然后利用修正后的用户偏好对物品候选集进行打分并按降序排序,从物品候选集中选择 top-k(k 为预设值)进行推荐,用户会给出物品粒度的正/负反馈. 对于物品粒度的负反馈,SCPR 仅仅将其从物品候选集中移除,NCPR 则将其保存在物品负反馈集合 $\mathcal{V}_u^-$ 中,用于模型再次选择 a_{rec} 动作时修正用户的偏好表示. 模型收到物品粒度的正反馈或者交互轮次达到预设的最大值 T 时,对话结束. 算法 1 详细说明了上述过程.

算法 1. NCPR 交互流程.

输入:用户 u,属性集合 $\mathcal{P}$,物品集合 $\mathcal{V}$,
推荐物品列表长度 k,最大交互轮次 T

输出:推荐成功或失败

1. 用户 u 初始化偏好属性 p_0;
2. 更新:$\mathcal{P}_u^+=\{p_0\}$;$\mathcal{P}_u^-=\{\}$;$\mathcal{P}_{\text{cand}}=\mathcal{P}\setminus p_0$;$\mathcal{V}_{\text{cand}}=\mathcal{V}_{p_0}$;$\mathcal{V}_u^-=\{\}$;
3. FOR $t=1,2,\cdots,T$ DO
4. 选择一个动作 a
5. IF $a==a_{\text{ask}}$ THEN
6. 利用属性打分模型对属性候选集进行打分并按降序排序
7. 从 $\mathcal{P}_{\text{cand}}$ 中选择得分最高的属性 p
8. IF u accepts p THEN
9. 更新:$\mathcal{P}_u^+=\mathcal{P}_u^+\cup p$;$\mathcal{V}_{\text{cand}}=\mathcal{V}_{\text{cand}}\cap V_p$;
10. ELSE
11. 计算 p 和其它属性之间的相似度,得到 $\mathcal{P}_t$
12. 更新:$\mathcal{P}_u^-=\mathcal{P}_u^-\cup p\cup\mathcal{P}_t$;

13.　END IF
14.　更新：$\mathcal{P}_{\text{cand}}=\mathcal{P}_{\text{cand}}\backslash(\mathcal{P}_u^-\cup\mathcal{P}_u^+)$
15.　ELSE IF $a==a_{\text{rec}}$ THEN
16.　利用$\mathcal{P}_u^+,\mathcal{P}_u^-,\mathcal{V}_u^-$对用户的偏好表示进行修正
17.　利用修正的用户偏好对物品候选集进行降序排序
18.　从$\mathcal{V}_{\text{cand}}$中选出 top-$k$ 物品$\mathcal{V}_k$
19.　IF u accepts $\mathcal{V}_k$ THEN
20.　推荐成功并退出.
21.　ELSE
22.　更新：$\mathcal{V}_{\text{cand}}=\mathcal{V}_{\text{cand}}\backslash\mathcal{V}_k$；$\mathcal{V}_u^-=\mathcal{V}_u^-\cup\mathcal{V}_k$
23.　END IF
24.　END IF
25. END FOR
26. 推荐失败并退出.

3.2　模型框架

图 1 给出了 NCPR 的工作流程，展示了不同模块间的关系和消息传递的过程. NCPR 基于 Lei 等人[15]所构建的知识图谱进行对话推荐，该知识图谱是包含用户 u、物品 v 和属性 p 三种节点的无向异构图，包含了用户-物品、用户-用户、用户-属性和物品-属性等四种类型的边. 具体地说，用户-物品之间的边(u,v)代表用户 u 和物品 v 之间至少发生过一次交互，用户-用户之间的边(u_0,u_1)代表用户 u_0 和用户 u_1 之间存在好友关系，用户-属性之间的边(u,p)代表用户 u 在历史交互中表现过对属性 p 的偏好，物品-属性之间的边(v,p)代表物品 v 带有 p 属性. NCPR 遵循 SCPR 中一些设定：(1) 和用户的交互相当于模型在图上游走的过程. 即模型会维护一条路径 P，代表交互过程中按时间顺序被用户确认的偏好属性，并在图上推理出下一个要到达的顶点，即向用户询问的属性或推荐的物品；(2) 不考虑不同关系的语义信息，即只关心两个顶点之间是否有一条边；(3) 将知识图谱当作一个无向图，模型只能到达图上的属性顶点和物品顶点，且不能重复到达同一顶点. 假设模型当前的游走路径为 $P=p_0,p_1,p_2,\cdots,p_t$，停留在 p_t，要转移到与 p_t 相邻的属性顶点或物品顶点. 这个过程可以被细分为以下的三个步骤：推理、决策和转移. 相较于 SCPR，NCPR 在这三个模块中均有改进：

(1) 推理模块. SCPR 使用二阶因子分解机对物品候选集进行打分，仅使用属性粒度的正反馈来建模用户的实时偏好，这种做法忽略了用户属性粒度的负反馈和物品粒度的负反馈对建模用户实时偏好的影响. NCPR 修改了物品打分模型，对物品进行打分时会利用交互过程中用户给出的所有反馈来建模用户的实时偏好. 我们期望随着交互轮次的增加，用户偏好中实时偏好的占比越来越大；

(2) 决策模块. SCPR 收到用户的负反馈后，会给决策智能体一个固定的奖励值，我们认为在真实场景下随着交互轮次的增加，用户会逐渐失去耐心，因此负反馈奖励应该不断减小. NCPR 使用奖励塑形，当收到用户的负反馈后，给决策智能体的负反馈奖励会随着交互轮次的增加不断减小；

(3) 转移模块. SCPR 在图上游走的过程中，只会游走到当前节点的相邻属性节点(两个属性节点的最短路径中不包含其它属性节点)，比如在图 2 中属性 p_0 的相邻节点为 p_1 和 p_4. 这种图结构中相邻属性的限制可以有效地减少属性候选集大小，即$\mathcal{P}_{\text{cand}}=\mathcal{AA}_t\backslash(\mathcal{P}_u^+\cup\mathcal{P}_u^-)$，其中$\mathcal{AA}_t$为属性节点 p_t 的所有相邻属性集合. NCPR 在此基础上使用协同过滤算法基于属性粒度的负反馈进一步缩小属性候选集大小，当询问的属性被用户拒绝(如 p_4)，NCPR 从属性候选集中删去与该属性非常相似的部分属性，如图 2 中属性 p_0 的相邻属性节点 p_2，可以缩短游走路径，减少交互轮次.

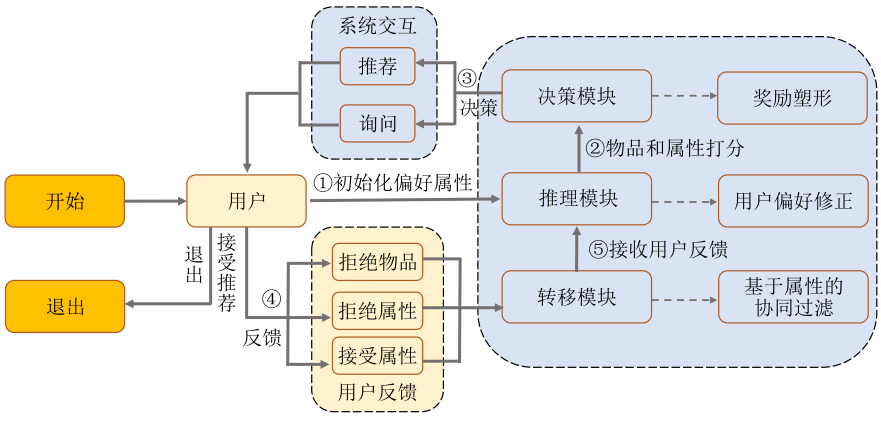

图 1　NCPR 工作流程

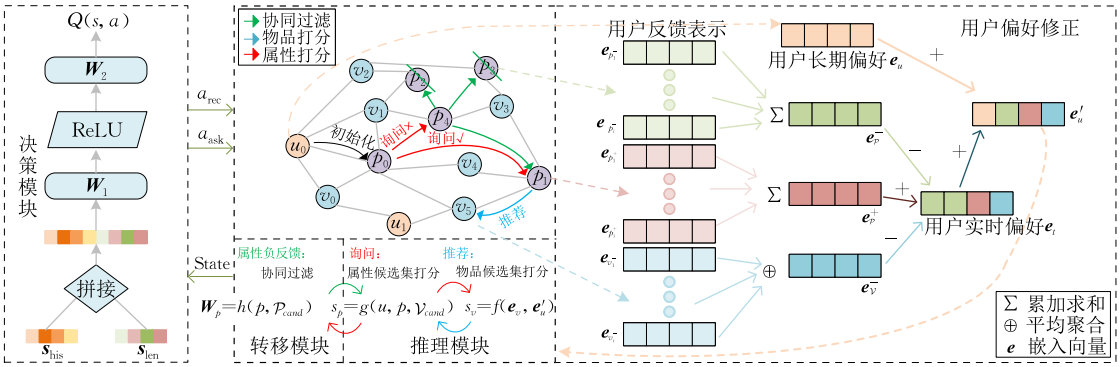

图 2 NCPR 模型结构图

图 2 展示了 NCPR 的整体框架，我们将在第 3.3 节具体介绍相关模块的工作细节.

3.3 NCPR 模型

3.3.1 推理模块

在该模块中，模型会对物品候选集和属性候选集进行打分，帮助决策模块解决询问什么属性或推荐什么物品的问题.

对物品候选集进行打分时，NCPR 首先利用交互过程中用户给出的所有反馈来建模用户的实时偏好，其中属性粒度的反馈可以直接反映用户细粒度的偏好，因此将交互过程中属性粒度的正反馈、负反馈进行累加：

$$\boldsymbol{e}_{\mathcal{P}}^{+}=Sum(\{\boldsymbol{e}_{p^+}|p^+\in\mathcal{P}_u^+\})=\sum_{p^+\in\mathcal{P}_u^+}\boldsymbol{e}_{p^+} \tag{1}$$

$$\boldsymbol{e}_{\mathcal{P}}^{-}=Sum(\{\boldsymbol{e}_{p^-}|p^-\in\mathcal{P}_u^-\})=\sum_{p^-\in\mathcal{P}_u^-}\boldsymbol{e}_{p^-} \tag{2}$$

一个用户拒绝一个物品的理由是多样的，因此物品粒度的反馈不能直接反映用户的实时偏好，我们采用对所有物品粒度的反馈取平均的方式来“投票”选出用户不喜欢的特征：

$$\boldsymbol{e}_{\mathcal{V}}^{-}=Avg(\{\boldsymbol{e}_{v^-}|v^-\in\mathcal{V}_u^-\})=\frac{1}{|\mathcal{V}_u^-|}\sum_{v^-\in\mathcal{V}_u^-}\boldsymbol{e}_{v^-} \tag{3}$$

其中 Sum 代表累加函数，Avg 代表平均聚合函数，$\mathcal{P}_u^+$ 代表属性的正反馈集合，$\mathcal{P}_u^-$ 代表属性的负反馈集合，$\mathcal{V}_u^-$ 代表物品的负反馈集合，$\boldsymbol{e}_{p^+}$ 为属性的正反馈集合中 p^+ 的嵌入向量，$\boldsymbol{e}_{p^-}$ 为属性的负反馈集合中 p^- 的嵌入向量，$\boldsymbol{e}_{\mathcal{V}}^-$ 为物品的负反馈集合中 v^- 的嵌入向量.

然后通过聚合不同类型的反馈信息来修正 NCPR 中用户的偏好表示，e_u 建模用户的长期偏好，$e_{\mathcal{P}}^+-e_{\mathcal{P}}^--e_{\mathcal{V}}^-$ 建模用户的实时偏好：

$$e_t=e_{\mathcal{P}}^+-e_{\mathcal{P}}^--e_{\mathcal{V}}^- \tag{4}$$

$$e'_u=e_u+e_t \tag{5}$$

其中 e_u 为 FM 预训练得到的用户偏好表示，e_t 为用户的实时偏好，e'_u 为修正后的用户偏好表示.

接下来利用修正后的用户偏好表示对物品候选集进行打分，即建模用户对物品的兴趣：

$$s_v=f(\boldsymbol{e}_v,\boldsymbol{e}'_u) \tag{6}$$

$$f(\boldsymbol{e}_v,\boldsymbol{e}'_u)=\boldsymbol{e}'_u\boldsymbol{e}_v \tag{7}$$

$$f(\boldsymbol{e}_v,\boldsymbol{e}'_u)=e_ue_v+\boldsymbol{e}_{\mathcal{P}}^+e_v-(\boldsymbol{e}_{\mathcal{P}}^-e_v+\boldsymbol{e}_{\mathcal{V}}^-e_v) \tag{8}$$

其中 $\boldsymbol{e}'_u$ 表示用户 u 修正后的嵌入向量，e_v 表示物品 v 的嵌入向量，其中 e_ue_v 建模用户 u 对物品 v 的长期偏好，即与当前对话状态无关，$e_{\mathcal{P}}^+e_v$ 建模当前对话状态下物品 v 与用户偏好属性的相关程度，$e_{\mathcal{P}}^-e_v+e_{\mathcal{V}}^-e_v$ 使用用户的负反馈来建模当前对话状态下用户 u 对物品 v 的不满意程度.

对话推荐系统包含了很多的组件，在当前研究领域下，学者们[15,18]会对前人工作中可用的组件进行复用. 为了证明 NCPR 中关键设计的有效性，我们不希望对某些组件进行技巧性的改进来提升实验效果，例如将 DQN 替换成 DDQN 等. NCPR 的两个关键设计分别是用户偏好修正和基于属性的协同过滤，推理模块的属性打分模型和决策模块的强化模型分别复用了 SCPR 中的信息熵模型[45]和 DQN 模型，详细过程可以在原工作[15]中看到，这里做一个简单的介绍.

属性打分模型的作用主要是帮助对话推荐系统根据当前的物品候选集的状态对属性候选集中每个属性带来的预期收益进行一个打分排序，即帮助模型决定询问用户哪个属性可以更好地消除物品的不确定性，在较短的交互轮次内做出符合用户个性化的推荐. 对于属性候选集 $\mathcal{P}_{\text{cand}}$ 中任意一个属性 p，其打分模型可以被定义为

$$s_p=g(u,p,\mathcal{V}_{\text{cand}}) \tag{9}$$

信息熵已经被证明是一种有效的用于不确定性估计的方法[45]，因此属性打分模型采用了信息熵[15]：

$$g(u,p,\mathcal{V}_{\text{cand}})=-\text{prob}(p)\cdot\log_2(\text{prob}(p)),$$

$$\text{prob}(p)=\frac{|\mathcal{V}_{\text{cand}}\cap\mathcal{V}_p|}{|\mathcal{V}_{\text{cand}}|} \tag{10}$$

其中$\mathcal{V}_{\text{cand}}$为物品候选集集合，$\mathcal{V}_p$代表所有包含属性$p$的物品构成的集合.

3.3.2　决策模块

当推理模块完成物品打分和属性打分后，决策模块开始工作，该模块的主要作用是帮助 NCPR 从动作集合$\{a_{\text{ask}},a_{\text{rec}}\}$中做出选择.

该模块采用 DQN(Deep Q-Network)作为策略网络，该策略网络的状态动作价值函数由一个两层的前馈神经网络近似：

$$\varphi(\boldsymbol{s},a|\boldsymbol{w})\approx Q_\pi(\boldsymbol{s},a) \tag{11}$$

$$Q_\pi(\boldsymbol{s},a)=E_\pi[R_t|S_t=\boldsymbol{s},A_t=a] \tag{12}$$

其中$Q_\pi(\boldsymbol{s},a)$即为动作价值函数，表示系统处于状态$\boldsymbol{s}$时执行动作a，并一直延续策略π直到当前会话结束所能获得的累积收益的期望值，$\varphi(\boldsymbol{s},a|\boldsymbol{w})$为状态动作价值函数的近似表示，$\boldsymbol{w}$即为图 2 所示的$\boldsymbol{W}_1$和$\boldsymbol{W}_2$，即不同层的权重矩阵，$\varphi(\boldsymbol{s},a|\boldsymbol{w})$中使用 ReLU 作为激活函数.

策略网络的输入状态向量$\boldsymbol{s}$由$\boldsymbol{s}_{his}$和$\boldsymbol{s}_{len}$两个状态向量拼接而来：

$$\boldsymbol{s}=\boldsymbol{s}_{\text{his}}\oplus\boldsymbol{s}_{\text{len}} \tag{13}$$

其中$\boldsymbol{s}_{\text{his}}$为用户反馈编码而来的一个长度为$T$的状态向量，该向量的每一维记录对应交互轮次用户给出的反馈，该状态向量可以帮助模型学习到更好的交互策略，比如用户已经给出了多个属性粒度的正反馈，那么当前做出推荐动作就有很大概率获得较高收益.$\boldsymbol{s}_{\text{len}}$为物品候选集长度编码而来的一个状态向量，该状态向量用于提醒模型在物品候选集足够小时应倾向于向用户做出推荐.例如，目前物品候选集的大小为 10，而设定的物品推荐数目k为 15，此时模型继续向用户询问属性显然是不合理的.

$r_{\text{rec\_suc}}$，$r_{\text{rec\_fail}}$，$r_{\text{ask\_suc}}$，$r_{\text{ask_fail}}$和r_{quit}分别代表模型收到物品粒度的正/负反馈、属性粒度的正/负反馈和交互轮次达到预设的最大交互轮次T时给智能体的 5 种不同奖励.我们认为在真实场景中随着交互轮次的增加，用户会逐渐失去耐心，因此我们认为负反馈奖励应该随着交互轮次的增加不断减小，因此我们使用奖励塑形使得负反馈奖励随着交互轮次的增加不断减小：

$$r=r_0*\left(1-\frac{(T-t-1)}{T}\right) \tag{14}$$

其中T为预设的最大交互轮次，当交互轮次大于T时用户就会失去耐心选择离开，t为当前交互轮次，取值范围为$[1,T]$，r_0为奖励预设的初始值，r为奖励塑形后的奖励值.

3.3.3　转移模块

当 NCPR 收到用户给出的属性粒度的正/负反馈或物品粒度的负反馈后，该模块开始工作，作用是适应用户的在线反馈.

当 NCPR 收到关于物品集合$\mathcal{V}_k$的负反馈，更新物品候选集$\mathcal{V}_{\text{cand}}$，并将其加入用户的拒绝物品集$\mathcal{V}_u^-$中(算法 1 第 22 行).

当 NCPR 收到关于属性p_{t+1}的正反馈，更新属性候选集$\mathcal{P}_{\text{cand}}$和物品候选集$\mathcal{V}_{\text{cand}}$，并将其加入用户的偏好属性集$\mathcal{P}_u^+$中(算法 1 第 9 行，第 14 行)，此时游走路径更新为$P=p_0,p_1,p_2,\cdots,p_t,p_{t+1}$.

当 NCPR 收到关于属性p_{t+1}的负反馈，会对知识图谱进行剪枝操作.该模块利用基于属性的协同过滤模型对经过信息熵模型排序过的属性候选集进行重排序(算法 1 第 11 行)，即从属性候选集$\mathcal{P}_{\text{cand}}$中移除那些与$p_{t+1}$相关性较高的属性，并将这些属性与$p_{t+1}$一起加入$\mathcal{P}_u^-$中，然后更新属性候选集$\mathcal{P}_{\text{cand}}$(算法 1 第 11 行).其模型定义如下：

$$\boldsymbol{W}_p=h(p,\mathcal{P}_{\text{cand}}) \tag{15}$$

$$h(p,\mathcal{P}_{\text{cand}})=\{\text{sim}(p,p_i)|p_i\in\mathcal{P}_{\text{cand}}\} \tag{16}$$

其中$\boldsymbol{W}_p$表示属性p与属性候选集中属性的相似系数矩阵.本文中，NCPR 设置每轮向用户询问单个属性，因此$\boldsymbol{W}_p\in\mathbb{R}^{1\times d}$.计算属性$p_{t+1}$的相似度矩阵时，我们测试了三种不同的相似度度量函数来验证该设计的有效性：

(1)余弦相似系数：

$$w_{p_{t+1}p_i}=\text{sim}(p_{t+1},p_i)=\frac{\boldsymbol{e}_{p_{t+1}}\cdot\boldsymbol{e}_{p_i}}{\|\boldsymbol{e}_{p_{t+1}}\|\|\boldsymbol{e}_{p_i}\|}=\frac{\boldsymbol{e}_{p_{t+1}}\cdot\boldsymbol{e}_{p_i}}{\sqrt{e_{p_{t+1}}^2}\sqrt{e_{p_i}^2}} \tag{17}$$

(2)皮尔逊相似系数：

$$w_{p_{t+1}p_i}=\text{sim}(p_{t+1},p_i)=\frac{(\boldsymbol{e}_{p_{t+1}}-\bar{e}_{p_{t+1}})\cdot(\boldsymbol{e}_{p_i}-\bar{e}_{p_i})}{\|\boldsymbol{e}_{p_{t+1}}-\bar{e}_{p_{t+1}}\|\|\boldsymbol{e}_{p_i}-\bar{e}_{p_i}\|}=\frac{(\boldsymbol{e}_{p_{t+1}}-\bar{e}_{p_{t+1}})\cdot(\boldsymbol{e}_{p_i}-\bar{e}_{p_i})}{\sqrt{(\boldsymbol{e}_{p_{t+1}}-\bar{e}_{p_{t+1}})^2}\sqrt{(\boldsymbol{e}_{p_i}-\bar{e}_{p_i})^2}} \tag{18}$$

(3)杰卡德相似系数

$$w_{p_{t+1}p_i}=\text{sim}(p_{t+1},p_i)=\frac{|\mathcal{V}_{p_{t+1}}\cap\mathcal{V}_{p_i}|}{|\mathcal{V}_{p_{t+1}}\cup\mathcal{V}_{p_i}|} \tag{19}$$

其中$\boldsymbol{e}_{p_{t+1}}$和$\boldsymbol{e}_{p_i}$为属性p_{t+1}和p_i的嵌入向量表示，$\bar{e}_{p_{t+1}}$和$\bar{e}_{p_i}$为$\boldsymbol{e}_{p_{t+1}}$和$\boldsymbol{e}_{p_i}$的平均值，$(\boldsymbol{e}_{p_{t+1}}-\bar{e}_{p_{t+1}})$称为$\boldsymbol{e}_{p_{t+1}}$的中心化.

考虑到协同过滤模型处理稀疏矩阵的能力较

弱，该模型具有以下优势：(1) NCPR 引入了知识图谱，传统协同过滤方法无法捕捉辅助信息之间的相关性，从而降低了推荐的准确性，基于属性的协同过滤模型可以利用知识图谱中多源结构性数据来缓解数据稀疏的问题；(2) 在一般场景中，物品-属性矩阵不会像用户-物品矩阵那样稀疏. 如果物品-属性矩阵很稀疏，则可以利用式(17)、(18)两种基于隐向量计算属性间相似度的方法，来弥补协同过滤模型处理稀疏数据能力不足的问题. 值得一提的是，我们并没有构建基于物品的协同过滤模型，即 NCPR 收到用户关于物品粒度的负反馈后，仅将该反馈用于用户偏好表示的修正. 原因有以下三点：

(1) 毫无疑问，协同过滤模型通过大量物品的训练可以学习到物品之间的共同属性，但这个前提首先就是样本要足够多. 在对话推荐中，系统推荐物品列表长度为预设值 k，考虑到实际情况这个值不能过大，例如本实验设置 $k=10$. 因此，当前会话中用户拒绝过的物品数量累积起来也达不到可以学习到物品之间共同属性的程度.

(2) 上面提到的通过大量物品的训练学习物品之间的共同属性与我们构建的基于属性的协同过滤模型中使用的杰卡德相似系数相似，即衡量两个物品集之间的相似度作为两个属性之间的相似度.

(3) 假设不考虑学习物品之间的关系，利用基于物品的协同过滤模型对用户拒绝的每一个物品进行协同过滤也是不合适的. Bi 等人[46]提到物品粒度的反馈很难被充分利用，因为用户拒绝一个物品的原因是多样的，物品粒度的负反馈不能明确地表示用户偏好. 例如，用户给出一个物品粒度的负反馈——“不喜欢黑色的风衣”，可能用户仅仅是不喜欢风衣这个属性，但是模型同样会把黑色作为一个特征，以此过滤掉物品候选集中那些带有黑色属性的物品. 然而，如果用户在与 CRS 交互的过程中给出过“喜欢黑色”这个属性粒度的正反馈，那么模型就可以推断出“风衣”是用户不喜欢的属性. 由此可见，充分利用属性粒度的反馈与物品粒度的反馈之间的关系可以帮助 CRS 更好地学习用户偏好表示，提升用户的满意度.

4 实验结果及分析

本文进行了多组实验来验证 NCPR 的有效性. 本章将主要回答以下问题：

问题 1. NCPR 与对话推荐系统领域的其它工作相比，效果如何？

问题 2. NCPR 中用户偏好表示修正和基于属性的协同过滤这两个关键设计是否都有效？

问题 3. NCPR 中参数设置是否合理？

4.1 数据集

本文使用对话推荐系统领域四个常用的基准数据集对 NCPR 进行评估，数据处理遵循 SCPR[15]中的设置：不考虑评论数少于 10 条的用户，以此减小数据稀疏性，并将数据按照 7∶1.5∶1.5 的比例划分为训练集、验证集和测试集. 数据集的统计如表 1 所示.

表 1 数据集统计

数据集	用户数量	物品数量	属性数量	交互次数
LastFM	1801	7432	33	76693
Yelp	27675	70311	29	1368606
LastFM*	1801	7432	8438	76693
Yelp*	27675	70311	590	1368606

LastFM① 数据集用于歌手/乐队推荐，Yelp② 数据集用于商户推荐. Lei 等人[14]对数据集中原始属性进行手动处理，保留了 LastFM 数据集中 33 个高频属性及 Yelp 数据集中 590 个高频属性，同时考虑到 Yelp 数据集的属性空间较大会导致交互轮次过多，因此为两个数据集设置了不同的问题场景. LastFM 数据集被用于在二值问题场景下对模型进行评估，即用户通过接受或拒绝被询问的属性来表达自己当前的偏好. Lei 等人[14]对 Yelp 数据集的原始属性进行了二级分类，将 590 个高频属性分为 29 个类，比如一星级饭店、二星级饭店、三星级饭店的父类为星级饭店. Yelp 数据集被用于在枚举问题场景下对模型进行评估. 在枚举问题场景下，推荐系统每次选择一个父类属性向用户询问，要求用户在该父类下的所有孩子属性中进行选择(用户可以选择任意个孩子属性). Lei 等人[15]认为实际场景中手动合并原始属性并不现实，因此使用原始属性对两个数据集进行了重构，分别称为 LastFM* 和 Yelp*. 公平起见，我们采用两个版本的数据集进行实验.

4.2 实验设置

4.2.1 用户模拟器

对话推荐系统需要在与用户交互的过程中对模型进行训练和评估. 然而使用真实场景下的对话数据来训练模型代价太大，因此我们通过构建用户模拟器[13]来完成这一工作. 构建用户模拟器[47]是一种常见的做法，也是当前研究领域大家通用的方法，当

① https://grouplens.org/datasets/hetrec-2011/
② https://www.yelp.com/dataset/

前研究工作[14-15,18]都遵循用户非常明确自己的偏好这一相同的假设来构建用户模拟器，在相同的用户模拟器设置下对比不同的方法是公平的. 用户模拟器使用离线数据来模拟真实场景下的对话，用户模拟器为交互历史中每个用户-物品对(u,v)构建一个会话，将物品v作为用户u的心仪物品. 会话开始时，模拟器从物品v的属性集合$\mathcal{P}_v$中随机选择一个属性p_0作为用户u的偏好属性，接下来如算法1所示，进入“模型选择动作-模拟器给出反馈”的循环，直到推荐成功或交互轮次达到预设的最大值T. 在交互过程中模型选择动作为a_{ask}时，模拟器只接受物品v的属性集合$\mathcal{P}_v$中的属性. 当模型选择动作为a_{rec}并且给出的推荐物品列表中包含物品v时，模拟器接受这次推荐. 值得一提的是，模型在枚举问题场景下训练和评估时，当模型选择动作为a_{ask}时，会选出得分最高的一级属性，将一级属性下的所有二级属性作为本轮询问的属性列表，如果目标物品v的属性集中包含二级属性列表中的一个或多个属性，则本轮询问的一级属性作为用户的正反馈，否则本轮询问的一级属性作为用户的负反馈.

4.2.2　模型参数设置

用户模拟器中，设定会话的最大交互轮次为15，推荐物品列表长度为10（当物品候选集大小小于10时，推荐物品列表长度为$len(\mathcal{P}_{\text{cand}})$）.

推理模块的因子分解机中，嵌入向量的维度为64，使用带有正则化项的SGD作为优化器，正则化项为0.001，模型的学习率设置为0.01.

决策模块的DQN模型中，使用RMSprop作为优化器并将目标网络的参数更新频率设置为20，智能体的奖励设置为：$r_{\text{rec\_suc}}=1$，$r_{\text{rec\_fail}}=-0.1$，$r_{\text{ask\_suc}}=0.01$，$r_{\text{ask_fail}}=-0.1$和$r_{\text{quit}}=-0.3$，衰减因子γ设置为0.999. 在数据存储上，记忆池大小为50 000，每轮批量梯度下降时从经验池中均匀随机采样128个转移样本.

转移模块的基于属性的协同过滤模型中，我们通过参数分析实验选用杰卡德相似系数作为相似度量方法，并在不同的数据集上设置了不同的超参数，这里使用δ_{jaccard}表示杰卡德相似系数阈值. 在LastFM数据集上，超参数设置为$\delta_{\text{jaccard}}=0.2$；在Yelp数据集上，超参数设置为$\delta_{\text{jaccard}}=0.5$；在LastFM*数据集上，超参数设置为$\delta_{\text{jaccard}}=0.3$；在Yelp*数据集上，超参数设置为$\delta_{\text{jaccard}}=0.4$.

4.2.3　基线模型

在现有的工作中，我们选择以下几个与本文方法最相关的基线模型作为对比：

(1) 贪心模型(Abs Greedy)[48]. 每轮交互都向用户推荐物品，收到物品粒度的负反馈时，将这些物品作为负样本对推荐模型进行在线更新，直到做出成功推荐或交互轮次达到预设的最大值.

(2) 最大熵模型(Max Entropy). 每轮交互按照设定的概率从动作集合$\{a_{\text{ask}},a_{\text{rec}}\}$中做出选择，如果动作为$a_{\text{ask}}$，拿出属性候选集中熵最大的属性向用户询问；如果动作为a_{rec}，从物品候选集中选择top-k物品进行推荐.

(3) CRM[13]. 由标准对话系统(包含NLU、DM、NLG)演变而来，包含状态追踪器、策略网络和推荐器三个部分. 其中状态追踪器记录用户当前偏好，策略网络用于学习与用户交互的合适策略，推荐器使用因子分解机(FM)预测符合用户当前偏好的物品.

(4) EAR[14]. 通过评估、行动和反射三阶段完成对话模块和推荐模块的交互. 评估阶段利用FM对物品候选集和属性候选集进行排序，行动阶段学习一个合适的对话策略，反射阶段利用物品粒度的负反馈更新FM.

(5) SCPR[15]. 基于CPR框架实现，证明了CPR框架的有效性. 该模型启发了NCPR的实现，是最具可比性的基线模型.

(6) UNICORN[18]. 当前领域最先进的方法，使用一种统一的对话策略来解决对话推荐系统中三个决策问题，即询问什么属性、推荐什么物品和什么时候询问或推荐.

4.2.4　评价指标

本文使用第t轮的推荐准确率(*success rate*@t，$SR@t$)来衡量对话推荐模型在t轮交互中做出成功推荐的比例，使用平均交互轮次AT(Average Turn)来衡量一次会话中，用户和模拟器的平均交互次数. $SR@t$越高代表模型在指定t轮交互内推荐效果越好，AT越低代表模型的交互策略越好，即模型的效率越高.

4.3　对比实验及分析(问题1)

我们将NCPR与六种基线模型进行了对比，证明了NCPR相较于对话推荐系统领域的先进工作，能够在两种评价指标上取得更好的表现，实验结果如表2所示. 图3展示了NCPR与SCPR在不同交互轮次($t=1,2,\cdots,15$)下的准确率(SR)与相对准确率(SR^*)，其中SR^*表示模型相较于SCPR在准确率上的差值. 从图3和表2中分析得出以下结论.

表 2 NCPR 与基线模型结果对比

模型	LastFM		Yelp		LastFM*		Yelp*	
	SR@15	*AT*	*SR*@15	*AT*	*SR*@15	*AT*	*SR*@15	*AT*
Abs Greedy	0.222	13.48	0.264	12.57	0.635	8.66	0.189	13.43
Max Entropy	0.283	13.91	0.921	6.59	0.669	9.33	0.398	13.42
CRM	0.325	13.75	0.923	6.25	0.580	10.79	0.177	13.69
EAR	0.429	12.88	0.967	5.74	0.595	10.51	0.182	13.63
SCPR	0.465	12.86	0.973	5.67	0.709	8.43	0.489	12.62
UNICORN	0.535	**11.82**	0.985	5.33	0.788	7.58	0.520	**11.31**
NCPR	**0.562***	11.87	**0.989***	**4.53***	**0.825***	**7.17***	**0.534***	12.07

注：下划线"_"代表基线模型最优结果，带"*"表示方法相较于所有基线模型显著提升（t 检验，$p<0.01$）

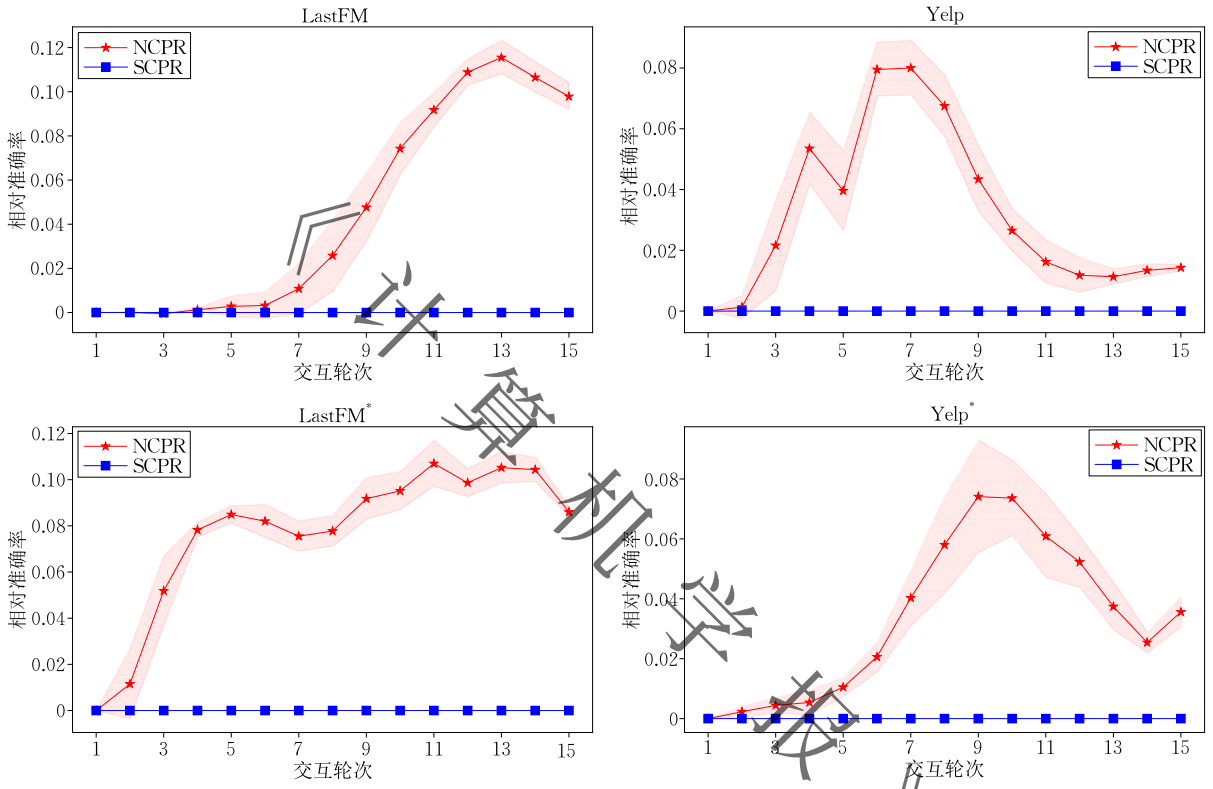

图 3 NCPR 和 SCPR 在不同轮次下的准确率和相对准确率的对比（阴影部分表示标准差）

（1）对话刚开始时（$t<3$）模型的准确率都接近于 0，原因是模型在交互的前期并不了解用户的实时偏好导致推荐准确率较低，而对话策略的目标是最大化累积收益，因此更倾向于向用户询问属性以更好地建模用户的实时偏好.

（2）贪心模型在四个数据集上的表现都较差，原因是属性粒度的反馈能够直接反映用户的实时偏好，而用户拒绝一个物品的原因却是多样的. 这一结果体现了在对话推荐任务中获取用户属性粒度反馈的重要性.

（3）在 Yelp 数据集下，贪心模型表现特别差，而其它模型准确率则很高. 原因在于 Yelp 数据集被用于在枚举问题场景（具体工作方式我们在数据集部分进行了详细介绍）下对模型进行评估，即每轮询问中模型都可以获得多个用户对于二级属性的反馈，因此模型（贪心模型除外，贪心模型无法获取用户属性粒度的反馈）可以在少数轮次的交互中快速理解用户实时偏好，达到非常高的准确率，这也是 NCPR 在 Yelp 数据集上提升较小的原因.

（4）CRM 和 EAR 在属性空间较大的数据集（LastFM* 和 Yelp*）上的表现还不如贪心模型. 主要原因是 CRM 和 EAR 中强化模型的决策空间太大导致无法学习到有效策略，而 SCPR、UNICORN 和 NCPR 限制了强化模型的决策空间，因此表现要优于其它基线模型.

（5）NCPR 的表现在大多数情况下优于六个基线模型，而且几乎在会话的每一轮，NCPR 相较于 SCPR 在准确率上均有明显提升，证明了基于负反馈的用户偏好表示修正和利用协同过滤算法对属性候选集进行限制的有效性. 在 LastFM 和 Yelp* 这两

个数据集上，NCPR 在平均交互轮次上的表现稍逊于最先进的基线模型 UNICORN，原因在于 NCPR 和 UNICORN 都是基于 SCPR 的改进，但是出发点不同. NCPR 通过用户偏好修正提高模型选择推荐时的成功率并对属性候选集重排序帮助模型选择预期价值最大的属性进行询问，UNICORN 使用一种更智能的交互策略来帮助模型进行决策，可以理解为 NCPR 主要改进对话推荐系统的推荐模块，而 UNICORN 主要改进对话推荐系统的对话模块. 因此 NCPR 在推荐准确率方面提升显著，在所有数据集上相较于六个基线模型，均取得了最优表现，在平均交互轮次方面，NCPR 和 UNICORN 的表现均优于 SCPR，但是在不同场景下提升程度不同.

(6) NCPR 相较于所有的基线模型是显著有效的. 我们进行了显著性检验，检验返回的 p 值小于 0.01，证明 NCPR 相较于最优基线模型的提升是统计显著的.

4.4　消融实验及分析(问题 2)

我们设计消融实验来证明 NCPR 中两个关键设计的有效性. NCPR-v 表示模型仅对用户偏好表示进行动态的修正，NCPR-x 表示模型仅保留基于属性的协同过滤模型，实验结果如表 3 所示. 这里我们仅报告 LastFM 和 Yelp 数据集(两种不同场景设置)下的实验结果，在另外两个数据集上也可以得出类似结论.

表 3　消融实验结果对比

数据集	模型	*SR*@15	*AT*
LastFM	NCPR-v	0.541	11.99
	NCPR-x	0.494	12.31
	NCPR	**0.562***	**11.87***
Yelp	NCPR-v	0.987	4.62
	NCPR-x	0.984	4.55
	NCPR	**0.989***	**4.53***

注：带“*”代表方法相较于所有对比方法有显著提升($p<0.01$).

图 4 展示了 NCPR-v、NCPR-x、NCPR 和 SCPR 四种模型在不同交互轮次下的准确率和相对准确率的对比. 从图 4 和表 3 中分析得出以下结论：

(1) 在两个数据集上，随着交互轮次的增加，NCPR-v 的变化趋势与 NCPR 相近，原因是用户偏好修正模型负责建模用户长期偏好和实时偏好之间关系，直接影响最终的推荐效果. 基于属性的协同过滤模型用于限制属性候选集的大小，并且为用户偏好修正模型提供更多的负样本，间接影响最终的推荐结果.

(2) 在两个数据集上，NCPR-v 的相对准确率都是先上升后下降. 先上升的原因是相较于 SCPR 只利用属性粒度的正反馈来建模用户的实时偏好，NCPR-v 会利用交互过程中用户给出的所有反馈，因此在交互前期 NCPR-v 能够更好地建模用户的实时偏好. 相对准确率下降的原因是，随着交互轮次

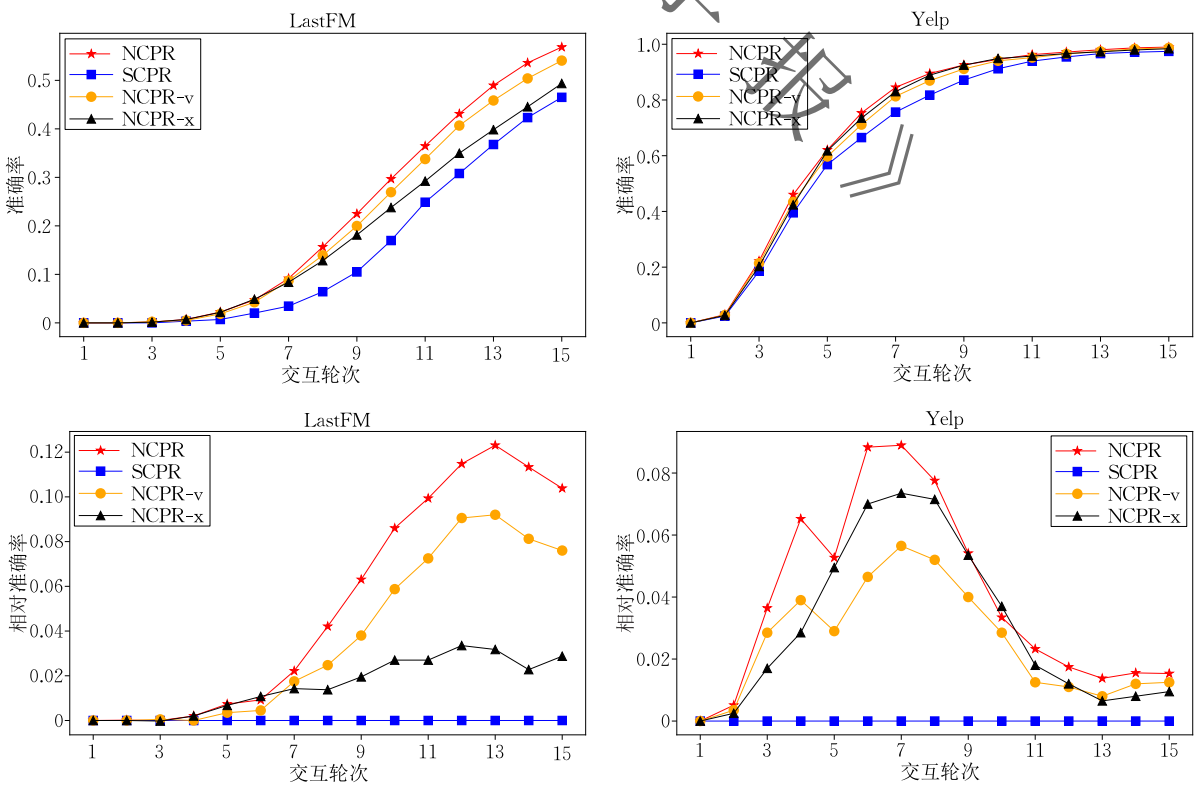

图 4　不同轮次下消融实验结果对比

的不断增多，模型对用户的实时偏好更加了解，此时模型询问用户心仪属性的概率就越来越大，因此正反馈的比例会逐渐增大，此时 NCPR-v 的优势就会稍微减弱. 这种现象在 Yelp 数据集上尤为明显的原因是枚举问题场景下每轮交互用户都会给出多个二级属性粒度的反馈.

(3) 在 Yelp 数据集上，NCPR-x 的相对准确率随着交互轮次的增加先上升后下降. 这是因为 Yelp 数据集中一级属性个数较少且相对独立，协同过滤模型经过前几轮的交互已经过滤掉仅有的相似属性对，之后模型再收到一级属性粒度的负反馈也无法进一步减少第一级中候选属性的个数.

(4) NCPR 保留任意一个关键设计，相较于 SCPR 在两种评价指标上的表现都有一定程度的提升. 同时 NCPR-v 和 NCPR-x 的综合表现比不上 NCPR，这是因为基于属性的协同过滤模型可以为用户偏好修正模型提供更多的属性负样本，帮助其更好的对用户偏好表示进行修正. 综上，NCPR 中的两个关键设计都是有效的.

4.5 参数分析实验(问题 3)

我们设计参数分析实验来证明 NCPR 中相似度量方法以及相似系数阈值选择的合理性. 首先将基于不同相似度量方法的 NCPR-x 在不同相似系数阈值下进行对比，实验结果如图 5 所示. 这里我们仅报告 LastFM 和 Yelp 数据集下的实验结果，因为在另外两个数据集上也可以得出相似结论.

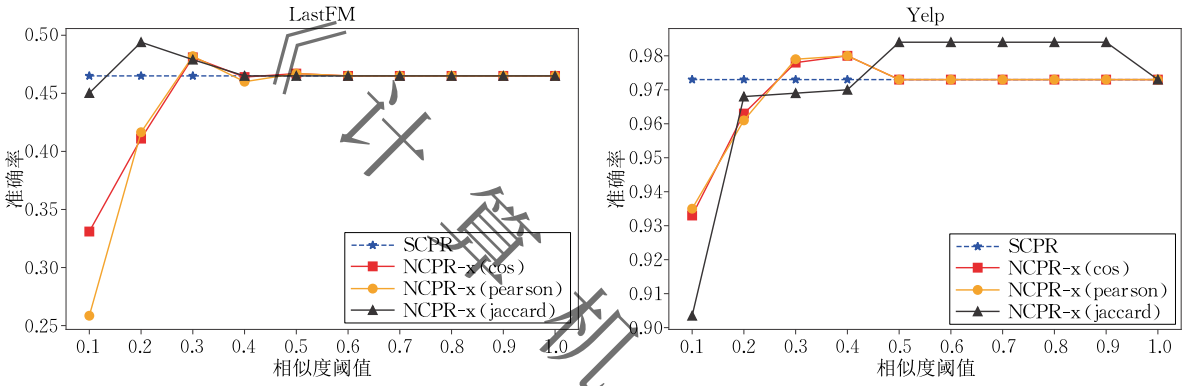

图 5 不同相似度阈值下模型的准确率对比

我们为基于不同相似度度量方法的 NCPR-x 选择各自最优的相似系数阈值进行对比，实验结果如表 4 所示.

表 4 基于不同相似度度量方法的 NCPR-x 结果对比

数据集	模型(NCPR-x)	*SR*@15	*AT*
LastFM	$\delta_{cos}=0.3$	0.481	12.46
	$\delta_{pearson}=0.3$	0.482	12.46
	$\delta_{jaccard}=0.2$	**0.494***	**12.31***
Yelp	$\delta_{cos}=0.4$	0.980	4.78
	$\delta_{pearson}=0.4$	0.980	4.78
	$\delta_{jaccard}=0.5$	**0.984***	**4.55***

注：带“*”代表方法相较于所有对比方法有显著提升 ($p<0.01$).

图 6 展示了三种相似度度量方法各自选择最优的相似系数阈值时，NCPR-x 在不同交互轮次下的准确率和相对准确率. 从表 4 和图 6 中可以分析得出以下结论：

(1) 基于不同相似度度量方法的 NCPR-x 在两种评价指标上的表现均优于 SCPR，证明我们基于属性的协同过滤模型是设计合理且有效的.

(2) 阈值设置较小时，基于不同相似度度量方法的 NCPR-x 表现都较差. 这是因为阈值设置较小会导致协同过滤模型筛掉很多本不相关的属性，导致 NCPR-x 无法学习到正确的用户偏好表示.

(3) 随着阈值的增大，基于不同相似度度量方法的 NCPR-x 的表现会达到最优后开始逐渐下降. 这是因为随着设定的阈值不断增大，NCPR-x 会从属性候选集中过滤掉那些真正与用户给出的属性负反馈相关的属性，此时模型会取得较好的表现. 当阈值大到一定程度时(比如 LastFM 中取 0.8)，属性相似度矩阵中除对角线外的元素都小于该阈值，协同过滤模型就失去了作用.

(4) NCPR-x(cos)和 NCPR-x(pearson)在两种评价指标上的表现相近，NCPR-x(jaccard)表现最好. 这是因为 NCPR-x(cos)和 NCPR-x(pearson)都是基于属性的嵌入向量来计算属性之间的相似度，皮尔逊相似度的计算相当于在计算余弦相似度的基础上对属性的嵌入向量进行中心化，我们在实验过程中发现属性的嵌入向量的平均值很小(1e-4 级别)，因此 NCPR-x(cos)和 NCPR-x(pearson)表现相近. 而 NCPR-x(jaccard)相当于利用了知识图谱引入的外部信息，不受因子分解机训练的影响，因此表现要优于 NCPR-x(pearson)和 NCPR-x(cos).

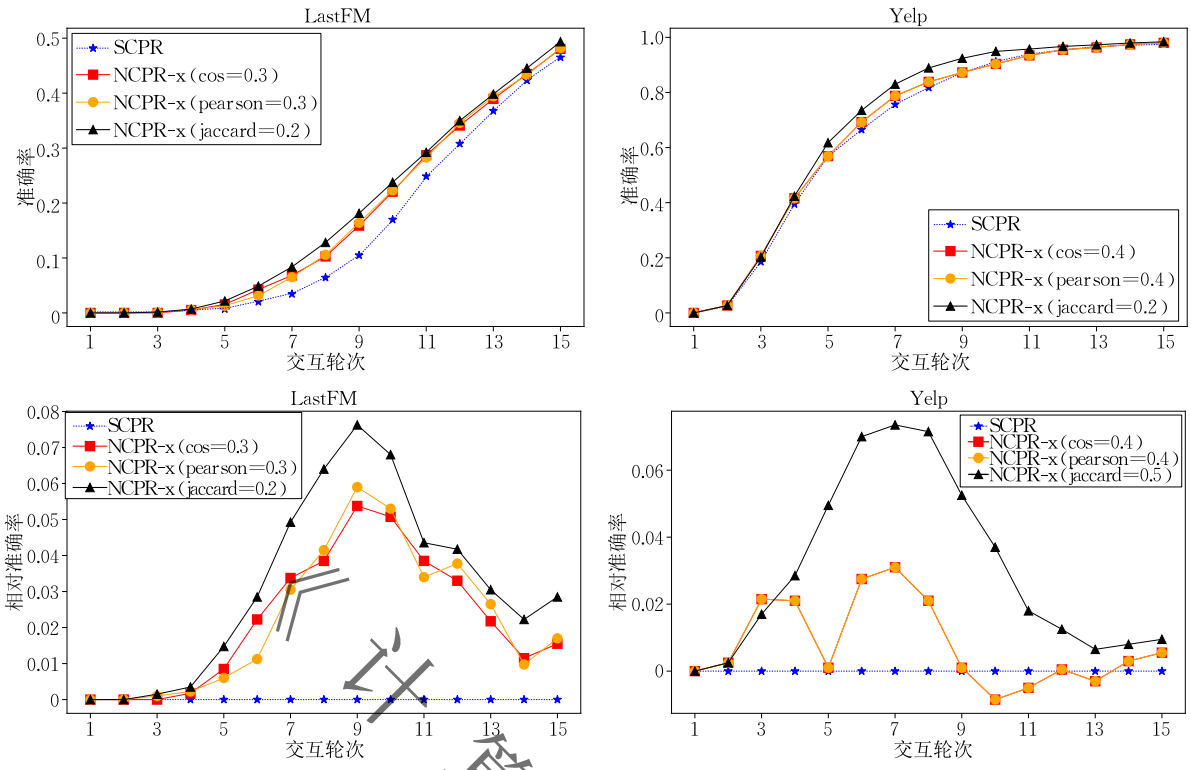

图 6　不同轮次下参数分析实验结果对比

5　总结与展望

在本文中，我们提出了一种新的对话推荐模型 NCPR，用于解决当前的研究工作无法充分利用用户交互过程中的负反馈的问题. 相较于那些将用户的在线反馈作为独立的特征或者训练实例的对话推荐模型，NCPR 能够有效地利用用户给出的负反馈信息，帮助属性候选集重排序的同时获得更多属性粒度的负样本，用于用户偏好表示的在线更新. 本文通过一系列的实验分析并验证了 NCPR 在推荐准确率和平均交互轮次两种评价指标上的表现均优于当前先进的基线模型. 最后，我们设计并实现了一个网页端的对话推荐系统，证明了 NCPR 在真实的对话推荐场景下的有效性.

对话推荐系统领域还有很多方向可以进行优化，在下一阶段的工作中，我们希望聚焦于以下方向：

(1) 设计更贴近真实场景的用户模拟器. 当前的用户模拟器的设计基于用户完全了解自己的偏好这种假设，但这种假设存在很大的局限性，因为真实场景下用户很多时候并不了解自己到底喜欢哪个物品.

(2) 融入多模态信息. 当前的研究工作还都是基于文本信息，为了让对话推荐系统更加智能，可以尝试引入多模态信息，如语音、图像等. 这样就可以利用其它领域的先进方法来造福对话推荐系统领域.

(3) 对推理模块和决策模块进行协同优化[5]. 这两个模块在任务上存在交叉，然而目前很多对话推荐系统的研究工作是将这两个模块分开进行优化的，所以对这两个模块进行协同优化是一个值得尝试的方向.

致　谢　衷心感谢老师们和学姐对本工作的指导！

参　考　文　献

[1] Chen J, Zhang H, He X, et al. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo, Japan, 2017: 335-344

[2] He X, Liao L, Zhang H, et al. Neural collaborative filtering//Proceedings of the 26th International Conference on World Wide Web. Perth, Australia, 2017: 173-182

[3] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 30-37

[4] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:1205.2618, 2012

[5] Gao C, Lei W, He X, et al. Advances and challenges in conversational recommender systems: A survey. arXiv preprint arXiv:2101.09459, 2021

[6] Jin X, Lei W, Ren Z, et al. Explicit state tracking with semi-supervision for neural dialogue generation//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino, Italy, 2018: 1403-1412

[7] Lei W, Jin X, Kan M Y, et al. Sequicity: Simplifying task-oriented dialogue systems with single sequence-to-sequence architectures//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia, 2018: 1437-1447

[8] Liao L, Ma Y, He X, et al. Knowledge-aware multimodal dialogue systems//Proceedings of the 26th ACM International Conference on Multimedia. Seoul Republic of Korea, 2018: 801-809

[9] Wang H, Wu Q, Wang H. Factorization bandits for interactive recommendation//Proceedings of the 31st AAAI Conference on Artificial Intelligence. California, USA, 2017: 2695-2702

[10] Chen H, Dai X, Cai H, et al. Large-scale interactive recommendation with tree-structured policy gradient//Proceedings of the AAAI Conference on Artificial Intelligence: Volume 33. Hawaii, USA, 2019: 3312-3320

[11] Zhou S, Dai X, Chen H, et al. Interactive recommender system via knowledge graph-enhanced reinforcement learning//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA, 2020: 179-188

[12] Loepp B, Hussein T, Ziegler J. Choice-based preference elicitation for collaborative filtering recommender systems//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Toronto, Canada, 2014: 3085-3094

[13] Sun Y, Zhang Y. Conversational recommender system//Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. Ann Arbor, USA, 2018: 235-244

[14] Lei W, He X, Miao Y, et al. Estimation-action-reflection: Towards deep interaction between conversational and recommender systems//Proceedings of the 13th International Conference on Web Search and Data Mining. Houston TX, USA, 2020: 304-312

[15] Lei W, Zhang G, He X, et al. Interactive path reasoning on graph for conversational recommendation//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. CA, USA, 2020: 2073-2083

[16] Zhao X, Zhang L, Ding Z, et al. Recommendations with negative feedback via pairwise deep reinforcement learning//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018: 1040-1048

[17] Xie R, Ling C, Wang Y, et al. Deep feedback network for recommendation//Proceedings of the IJCAI-PRICAI. Yokohama, Japan, 2020: 2519-2525

[18] Deng Y, Li Y, Sun F, et al. Unified conversational recommendation policy learning via graph-based reinforcement learning//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. Canada, 2021: 1431-1441

[19] Ma C, Kang P, Liu X. Hierarchical gating networks for sequential recommendation//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, USA, 2019: 825- 833

[20] Ji S, Pan S, Cambria E, et al. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(2): 494-514

[21] Wang H, Zhang F, Xie X, et al. DKN: Deep knowledge-aware network for news recommendation//Proceedings of the 2018 World Wide Web Conference. Lyon, France, 2018: 1835-1844

[22] Huang J, Zhao W X, Dou H, et al. Improving sequential recommendation with knowledge-enhanced memory networks//Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. Ann Arbor, USA, 2018: 505-514

[23] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion//Proceedings of the Advances in Neural Information Processing Systems. Lake Tahoe, USA, 2013: 926-934

[24] Guo Q, Zhuang F, Qin C, et al. A survey on knowledge graph-based recommender systems. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(8): 3549-3568

[25] Huang X, Qian S, Fang Q, et al. Meta-path augmented sequential recommendation with contextual co-attention network. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2020, 16(2): 1-24

[26] Huang X, Fang Q, Qian S, et al. Explainable interaction-driven user modeling over knowledge graph for sequential recommendation//Proceedings of the 27th ACM International Conference on Multimedia. New York, USA, 2019: 548-556

[27] Yu X, Ren X, Gu Q, et al. Collaborative filtering with entity similarity regularization in heterogeneous information networks//Proceedings of the 2nd IJCAI Workshop on Heterogeneous Information Network Analysis (HINA 2013). Beijing, China, 2013, 27

[28] Yu X, Ren X, Sun Y, et al. Recommendation in heterogeneous information networks with implicit user feedback//Proceedings of the 7th ACM Conference on Recommender Systems. Hong Kong, China, 2013: 347-350

[29] Yu X, Ren X, Sun Y, et al. Personalized entity recommendation: A heterogeneous information network approach//Proceedings of the 7th ACM International Conference on Web Search and Data Mining. New York, USA, 2014: 283-292

[30] Shi C, Zhang Z, Luo P, et al. Semantic path based personalized recommendation on weighted heterogeneous information networks//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York, USA, 2015: 453-462

[31] Hu B, Shi C, Zhao W X, et al. Leveraging meta-path based context for top-n recommendation with a neural co-attention model//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018: 1531-1540

[32] Zhao H, Yao Q, Li J, et al. Meta-graph based recommendation fusion over heterogeneous information networks//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, Canada, 2017: 635-644

[33] Zhang X, Xie H, Li H, et al. Conversational contextual bandit: Algorithm and application//Proceedings of the Web Conference 2020. New York, USA, 2020: 662-672

[34] Li L, Chu W, Langford J, et al. A contextual-bandit approach to personalized news article recommendation//Proceedings of the 19th International Conference on World Wide Web. Raleigh North Carolina, USA, 2010: 661-670

[35] Li S, Lei W, Wu Q, et al. Seamlessly unifying attributes and items: Conversational recommendation for cold-start users. ACM Transactions on Information Systems (TOIS), 2021, 39(4): 1-29

[36] Xie Z, Yu T, Zhao C, et al. Comparison-based conversational recommender system with relative bandit feedback//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA, 2021: 1400-1409

[37] Zhang Y, Chen X, Ai Q, et al. Towards conversational search and recommendation: System ask, user respond//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino, Italy, 2018: 177-186

[38] Li R, Kahou S, Schulz H, et al. Towards deep conversational recommendations. arXiv preprint arXiv:1812.07617, 2018

[39] Zhou K, Zhao W X, Bian S, et al. Improving conversational recommender systems via knowledge graph based semantic fusion//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. CA, USA, 2020: 1006-1014

[40] Lu Y, Bao J, Song Y, et al. RevCore: Review-augmented conversational recommendation. arXiv preprint arXiv:2106.00957, 2021

[41] Liang Z, Hu H, Xu C, et al. Learning neural templates for recommender dialogue system. arXiv preprint arXiv:2109.12302, 2021

[42] Zhou Y, Zhou K, Zhao W X, et al. C2-CRS: Coarse-to-fine contrastive learning for conversational recommender system. arXiv preprint arXiv:2201.02732, 2022

[43] Christakopoulou K, Beutel A, Li R, et al. Q&R: A two-stage approach toward interactive recommendation//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018: 139-148

[44] Ren X, Yin H, Chen T, et al. Learning to ask appropriate questions in conversational recommendation//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. Canada, 2021: 808-817

[45] Wu J, Li M, Lee C H. A probabilistic framework for representing dialog systems and entropy-based dialog management through dynamic stochastic state evolution. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(11): 2026-2035

[46] Bi K, Ai Q, Zhang Y, et al. Conversational product search based on negative feedback//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing, China, 2019: 359-368

[47] Chandramohan S, Geist M, Lefevre F, et al. User simulation in dialogue systems using inverse reinforcement learning//Proceedings of the 20th Annual Conference of the International Speech Communication Association. Florence, Italy, 2011: 1025-1028

[48] Christakopoulou K, Radlinski F, Hofmann K. Towards conversational recommender systems//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 815-824

ZHU Li-Xi, M. S. candidate. His research interests include user modeling and recommendation systems.

HUANG Xiao-Wen, Ph. D., lecturer. Her research interests include multimedia computing, data mining, user modeling, recommender systems.

ZHAO Meng-Yuan, M. S. candidate. Her research interests include user modeling and recommendation systems.

SANG Ji-Tao, Ph. D., professor, Ph. D. supervisor. His research interests include machine learning and cognitive computing, artificial intelligence and applications and other fields.

Background

Traditional recommendation systems work in a static way, only mining user interests and modeling user preferences from interaction history. This static way of working leads to a certain flaw in traditional recommender systems, i. e. they cannot answer two questions: what do users like at the moment, and what are the reasons why they like an item. Conversational recommendation systems can dynamically interact with users by introducing conversational technology and get real-time feedback from them during the interaction, so that when the system does not know what a user likes and why they like an item, it can simply ask the user directly. However, current CRS studies more often uses positive feedback on the granularity of attributes given by the user to model the user's real-time preferences, ignoring the impact of negative user feedback on modeling the user's real-time preferences, but negative feedback is an important part of user feedback and can also indicate real-time user preferences, making it difficult to make fine-grained corrections to the user preference representation, which means it is difficult to effectively balance the relationship between users' long-term preference and real-time preference. At the same time, current state-of-the-art work in the field only takes advantage of the natural advantages of graph structure to limit the size of the attribute candidate set, which suffers from the problem of too many interaction rounds due to the large attribute candidate set.

In this paper, we propose a new conversational recommendation model, NCPR, to address the problem that current research work does not fully exploit negative feedback during user interaction. Compared to those conversational recommendation models that use users' online feedback as independent features or training instances, NCPR can effectively exploit the negative feedback given by users to help reorder attribute candidate sets while obtaining more negative samples of attribute granularity for online updating of user preference representations. In this paper, we analyse and validate through a series of experiments on four benchmark datasets that NCPR outperforms current state-of-the-art baseline models in terms of both recommendation accuracy and average interaction rounds. Finally, we design and implement a web-based conversational recommendation system to demonstrate the effectiveness of NCPR in a real-world conversational recommendation scenario.

This research is partly supported by the Fundamental Research Funds for the Central Universities under Grant No. 2021RC217, the National Natural Science Foundation of China under Grant No. 62202041.