文章编号:2096-1472(2024)01-0058-06
DOI:10.19644/j.cnki.issn2096-1472.2024.001.014

基于用户反馈和对话历史的对话式推荐技术研究

杨 畅,姚 越,方霖枫,周仁杰

(杭州电子科技大学计算机学院,浙江 杭州 310012)

✉ yangchang@hdu.edu.cn; yaoyue@hdu.edu.cn; terrywu59@gmail.com; rjzhou@hdu.edu.cn

摘 要:针对对话式推荐系统中忽视用户负反馈和对话历史信息的问题,创新地提出了一个对话式推荐模型。首先,该模型通过构建负反馈图捕获和编码用户拒绝的属性和项目信息,结合动态奖励函数,使系统能更精准地理解用户的实时偏好。其次,将序列模型融入智能体实现了对每步对话状态的编码,从而能基于全局对话状态做出更准确的推荐决策。为验证模型的有效性,在 LastFM 和 LastFM * 数据集上开展实验,相较于最优的基线模型,本文方法的推荐成功率分别提升了 21%和 13.7%,平均推荐轮次数也分别降低了 1.56 轮和 2.35 轮。实验结果表明,用户负反馈和对话历史的深度整合,为对话式推荐系统带来更高的准确性。

关键词:推荐系统;对话式推荐;强化学习;图表示学习

中图分类号:TP315 **文献标志码**:A

Research on Conversational Recommender Technology Based on User Feedback and Conversation History

YANG Chang, YAO Yue, FANG Linfeng, ZHOU Renjie

(*School of Computer Science, Hangzhou Dianzi University, Hangzhou* 310012, *China*)

✉ yangchang@hdu.edu.cn; yaoyue@hdu.edu.cn; terrywu59@gmail.com; rjzhou@hdu.edu.cn

Abstract: This paper proposes an innovative conversational recommender model to address the issue of overlooking user negative feedback and dialogue history information in conversational recommender systems. Firstly, this model captures and encodes the attributes and item information that users refuse by constructing a negative feedback graph, and combines dynamic reward function to make the system understand the user' real-time preferences more accurately. Secondly, the sequence model is integrated into the intelligent agent to realize encoding of each dialogue state, so that more accurate recommendation decisions can be made based on the global dialogue state. To verify the effectiveness of the model, experiments are conducted on the LastFM and LastFM* datasets. Compared to the optimal baseline model, the recommendation success rate of the proposed method has increased by 21% and 13.7%, respectively, and the average number of recommendation rounds has decreased by 1.56 and 2.35 rounds, respectively. The experimental results indicate that the deep integration of user negative feedback and dialogue history brings higher accuracy to conversational recommendation systems.

Key words: recommender system; conversational recommender; reinforcement learning; graph representation learning

0 引言(Introduction)

随着互联网技术的迅猛发展,人们逐渐从信息匮乏的时代走入了信息过载的时代,因此如何应对信息过载问题成为一个迫切需要解决的问题。推荐系统的出现有效地缓解了这一问题。然而,传统的推荐方法在处理新用户冷启动和充分利用用户反馈问题时仍面临挑战。相较之下,对话式推荐系统

收稿日期:2023-05-11
基金项目:国家重点研发计划项目(2022YFB3105401)

(Conversational Recommender System, CRS)通过与用户直接交互收集"显式反馈",从而实现更精确的推荐。鉴于对话式推荐系统在电商和社交媒体中具有巨大的应用潜力,其已成为业界和学术界关注的焦点。本文提出一种新方法,可以解决对话推荐中的准确状态建模和高效策略制定问题。该方法整合了用户负反馈图以优化状态表示,并采用动态奖励函数更有效地指导策略学习。此外,通过在智能体中加入历史对话状态的序列模型编码,进一步增强了模型的推荐性能。

1　相关工作(Related works)

1.1　传统推荐系统

传统的推荐系统通常依赖用户与项目的交互矩阵,并用机器学习方法挖掘其中的隐含关系。例如,协同过滤通过矩阵分解等手段将用户与项目之间的交互映射到低维空间,进而提取特征、找寻相似用户,并推荐用户喜好的项目[1]。SALAKHUTDINOV等[2]引入深度学习架构,使深度学习在推荐系统领域得到广泛应用。HIDASI等[3]用循环神经网络[4](Recurrent Neural Network, RNN)对用户的交互序列进行建模,预测下一个点击的项目。HE等[5]则使用生成对抗网络,分别在生成和判别模型中进行数据捕获和预测。WANG等[6]融合了协同过滤和图表示学习,开发了一个神经图协同过滤框架,通过传播用户偏好的项目特征提升推荐结果的准确性。

然而,以上方法各有一定的局限性,尤其体现在解决新用户冷启动和充分利用用户在线反馈问题方面。相较之下,对话推荐系统通过与用户直接对话并询问问题以了解用户需求[7]。这种对话式推荐在多个环境下已受到研究者的广泛关注。

1.2　基于属性的对话式推荐系统

近年来,学者们在基于属性的对话式推荐技术方面有多项探索。最初,CRM(Conversational Recommender Model)[8]提出了单轮对话推荐方法,不论推荐结果是否被接受,对话都将在一轮推荐后结束。该方法用基于长短期记忆网络[9](Long Short-term Memory,LSTM)的编码模块编码用户输入的属性-值对,并用二路因子分解机[10]预测用户对每个项目的评分。后续研究将CRM扩展到多轮交互,根据先前反馈优化后续推荐,实现了更精确的推荐。

LEI等[11]提出了一个多轮对话推荐框架,包含评估(Estimation)、动作(Action)和反应(Reflection)三个模块。评估模块采用因子分解机和贝叶斯个性化排序算法,动作模块基于强化学习选取最优策略,反应模块则根据用户反馈更新损失函数[12]。CPR(Conversational Path Reasoning)[13]使用图模型捕捉候选属性和项目间的关系,通过多轮对话和用户互动,不断修剪候选产品和属性,从而提高推荐性能。DENG等[14]综合先前研究,将推荐、询问和决策任务统一在一个强化学习框架中,用基于图的马尔可夫决策过程(Markov Decision Process, MDP)和图卷积神经网络[15](Graph Convolution Network, GCN)、Transformer[16]学习对话状态,再用深度Q网络[17](Deep Q-Network, DQN)进行价值学习。

2　模型构建(Model building)

2.1　融合负反馈的对话状态建模

本文通过将用户、商品、商品信息等嵌入基于图的MDP环境中对用户当前的对话状态进行建模。在用户与智能体交互的过程中,不断裁剪环境中的结点,每一步交互都将得到表示用户正反馈和负反馈信息的动态图。图卷积神经网络是一种聚合节点周围信息的有效模型,能捕获到节点的高阶连通性[18]。因此,本文模型在每一步交互结束后,使用图卷积神经网络分别对用户的正反馈图和反馈图进行图卷积,提取图中结点的结构特征,得到每个结点的表示向量。通过注意力机制对上一步中的正、负反馈结点的表示向量进行增强,以此将网络的关注点聚焦于关键的结点上。通过一个全连接层将经过聚合的结点表示编码成最终的对话状态表示向量。融合负反馈的图表示学习如图1所示。

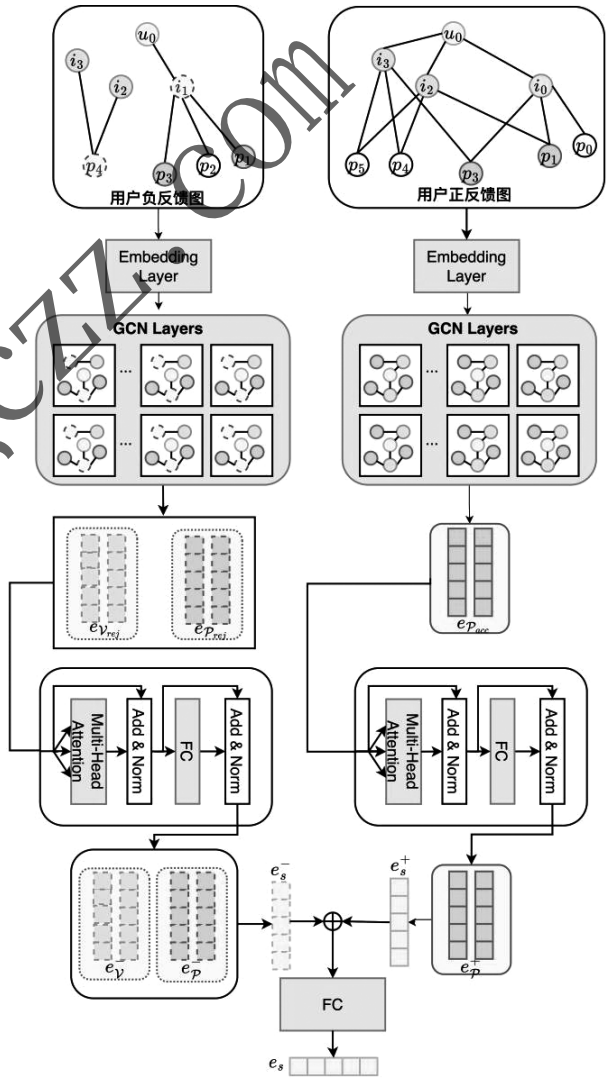

图1　融合负反馈的图表示学习

Fig. 1 Graph representation learning incorporating negative feedback

2.1.1　基于图的MDP环境

本文中将基于图的MDP环境定义为由状态空间S、动作空间A、状态转移函数$T:S\times A\to S$和奖励函数$R:S\times A\to\mathbb{R}$构成的四元组(S,A,T,R)。其中,状态包含了给定用户u的

情况下，时刻 t 的对话历史 $H_u^{(t)}=[P_{\text{acc}}^{(t)},P_{\text{rej}}^{(t)},V_{\text{rej}}^{(t)}]$ 和该时刻下与用户相关的正反馈图、负反馈图 $G_u^{(t)}=[G_u^{+(t)},G_u^{-(t)}]$。例如，在基于图的 MDP 环境下(图 2)，用户 u_0 的目标项目是 i_0 且在时刻 t 已经接受了属性 p_1 和 p_3，拒绝了项目 i_1 和属性 p_4，则对话历史可以表示为 $H_{u_0}^{(t)}=[\{p_1,p_3\},\{p_4\},\{i_1\}]$。

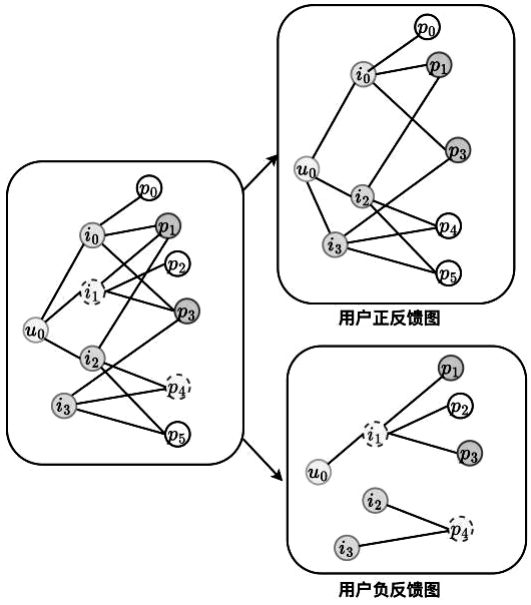

图 2 基于图的 MDP 环境

Fig. 2 Graph-based MDP environment

动作空间 $A=\{V_{\text{cand}}^{(t)},P_{\text{cand}}^{(t)}\}$，分别代表候选项目和候选属性。其中，候选项目定义为 $V_{\text{cand}}^{(t)}=V_{P_{\text{acc}}^{(t)}}\setminus V_{\text{rej}}^{(t)},V_{P_{\text{acc}}^{(t)}}$ 代表所有与 $P_{\text{acc}}^{(t)}$ 中属性直接相连的项目。候选属性定义为 $P_{\text{cand}}^{(t)}=P_{V_{\text{cand}}^{(t)}}\setminus(P_{\text{acc}}^{(t)}\cup P_{\text{rej}}^{(t)})$，即候选项目包含的所有属性减去已经交互过的属性。

用户反馈图表示为由结点和邻接矩阵构成的二元组 $G=(N,A)$。对于正反馈图 $G_u^{+(t)}$，结点集合 $N^{+(t)}=\{u\}\cup P_{\text{acc}}^{(t)}\cup P_{\text{cand}}^{(t)}\cup V_{\text{cand}}^{(t)}$，邻接矩阵则定义为公式(1)，用户和候选项目之间的边权重定义如公式(2)所示。

$$A_{i,j}^{(t)}=\begin{cases}w_v^{(t)}, & n_i=u,n_j\in V_{\text{cand}}\\ \sigma(\boldsymbol{e}_{n_i}^{\mathrm{T}}\boldsymbol{e}_{n_j}), & n_i\in V,n_j\in P_{\text{cand}}\\ 0, & \text{其他}\end{cases}\tag{1}$$

$$w_v^{(t)}=\sigma\Big(\boldsymbol{e}_u^{\mathrm{T}}\boldsymbol{e}_v+\sum_{p\in P_{\text{acc}}^{(t)}}\boldsymbol{e}_v^{\mathrm{T}}\boldsymbol{e}_p-\sum_{p\in P_{\text{rej}}^{(t)}\cap P_v^{(t)}}\boldsymbol{e}_v^{\mathrm{T}}\boldsymbol{e}_p\Big)\tag{2}$$

对于负反馈图 $G_u^{-(t)}$，结点集合 $N^{-(t)}=\{u\}\cup P_{\text{rej}}^{(t)}\cup V_{\text{rej}}^{(t)}\cup P_{\text{cand}}^{(t)}\cup V_{\text{cand}}^{(t)}$，邻接矩阵则定义如公式(3)所示。

$$A_{i,j}^{(t)}=\begin{cases}\sigma(\boldsymbol{e}_u^{\mathrm{T}}\boldsymbol{e}_{n_j}), & n_j\in V_{\text{rej}}\\ \sigma(\boldsymbol{e}_{n_i}^{\mathrm{T}}\boldsymbol{e}_{n_j}), & n_i\in V_{\text{rej}},n_j\in P_{\text{cand}}\\ \sigma(\boldsymbol{e}_{n_i}^{\mathrm{T}}\boldsymbol{e}_{n_j}), & n_i\in P_{\text{rej}},n_j\in V_{\text{cand}}\\ 0, & \text{其他}\end{cases}\tag{3}$$

状态转移函数 $T:S\times A\to S$ 根据 t 时刻的状态 $\boldsymbol{s}_t$ 和智能体在动作空间 A 中选择的动作 a_t 给出下一个状态 $\boldsymbol{s}_{t+1}$。动作空间中的动作有两类，分别为询问属性 p 和推荐项目 i。如果用户在 t 时刻接受了询问的属性 p_t，则 $P_{\text{acc}}^{(t+1)}=P_{\text{acc}}^{(t)}\cup p_t$，否则 $P_{\text{rej}}^{(t+1)}=P_{\text{rej}}^{(t)}\cup p_t$；如果用户拒绝了推荐的项目 i_t，则 $V_{\text{rej}}^{(t+1)}=V_{\text{rej}}^{(t)}\cup i_t$，否则推荐成功，本轮推荐结束。

2.1.2 动态奖励函数

奖励函数 $R:S\times A\to\mathbb{R}$ 给出在当前状态 $\boldsymbol{s}_t$ 下执行动作 a_t 的奖励 r_t。本文采用根据环境变化而变化的动态奖励函数，首先按照表 1 根据当前对话的轮次缩放基础奖励，然后按照公式(4)根据候选项目空间中的项目数量的缩减比例计算最终的奖励。

表 1 基础奖励

Tab.1 Basic reward

情况	符号	计算方式
询问成功	$r_{\text{ask\_suc}}$	$0.02\times\left(1-\frac{turn}{max\_turn}\right)$
询问失败	$r_{\text{ask\_fail}}$	$-0.05\times\left(1+\frac{turn}{max\_turn}\right)$
推荐失败	$r_{\text{rec\_fail}}$	$-0.2\times\left(1-\frac{turn}{max\_turn}\right)$
推荐成功	$r_{\text{rec\_suc}}$	1
到达最大轮数	$r_{\text{until\_T}}$	-0.3

$$r^{(t)}(\boldsymbol{s}_t,a_t)=\begin{cases}r_{\text{ask\_suc}}+r_{\text{ask\_suc}}\times\left(1-\frac{|V_{\text{cand}}^{(t+1)}|}{V_{\text{cand}}^{(t)}}\right), & \text{询问成功}\\ r_{\text{ask\_fail}}-r_{\text{ask\_fail}}\times\left(1-\frac{|V_{\text{cand}}^{(t+1)}|}{V_{\text{cand}}^{(t)}}\right), & \text{询问失败}\\ r_{\text{rec\_fail}}-r_{\text{rec\_fail}}\times\left(1-\frac{|V_{\text{cand}}^{(t+1)}|}{V_{\text{cand}}^{(t)}}\right), & \text{推荐失败}\\ r_{\text{rec\_suc}}, & \text{推荐成功}\\ r_{\text{util\_T}}, & \text{到达最大轮数}\end{cases}\tag{4}$$

2.1.3 融合负反馈的图表示学习

由于本文将对话式推荐建模为基于图的 MDP 环境下的统一策略学习问题，因此需要将对话和图的结构信息编码为状态向量表示。本文采用 Trans-E[19] 进行预训练，确保异构图中的节点表示具备结构化特征。然后使用 GCN 对环境中的正、负反馈图进行图卷积以捕捉表示当前对话状态的图中的高阶结构信息。对于第 l 层图卷积，计算方法如公式(5)至公式(7)所示，其中 N_i 是第 i 个节点 n_i 的邻接节点，W_l 和 B_l 是训练的参数。

$$\boldsymbol{e}_i^{(l+1)}=\text{ReLU}\Big(\sum_{j\in N_i}\Lambda_{i,j}W_l\boldsymbol{e}_j^{(l)}+B_l\boldsymbol{e}_i^{(l)}\Big)\tag{5}$$

$$\boldsymbol{\Lambda}=\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\tag{6}$$

$$D_{ii}=\sum_j A_{i,j}\tag{7}$$

经过 GCN 编码后，可以得到结合了高阶结构信息的结点表示，接着，将图上所有结点的表示向量视作序列，利用 Transformer Encoder 来进一步增强结点的表示，从而让网络更加聚焦在重要的结点上，如公式(8)至公式(9)所示；最后，对编码的结果做平均聚合，如公式(10)所示。

$$\boldsymbol{h}^*=\text{MultiHead}(W^Q\boldsymbol{h}^{(l)},W^K\boldsymbol{h}^{(l)},W^V\boldsymbol{h}^{(l)})\tag{8}$$

$$\boldsymbol{h}^{(l+1)}=\text{LayerNorm}(\text{FNN}(\boldsymbol{h}^{*}+\boldsymbol{h}^{(l)})) \tag{9}$$

$$\boldsymbol{g}_t=\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{h}_j^{L} \tag{10}$$

本文分别选取正、负反馈图上经过聚合的用户表示向量 $\boldsymbol{e}_s^{+}$ 和 $\boldsymbol{e}_s^{-}$ 进行拼接,使用全连接神经网络进行最后的编码,如公式(11)所示。

$$\boldsymbol{e}_s=\text{ReLU}(\text{FNN}([\boldsymbol{e}_s^{+},\boldsymbol{e}_s^{-}])) \tag{11}$$

2.2 基于状态序列建模的对话策略

对话策略的学习是使用时间差分(Temporal Difference, TD)算法[20]优化 DQN 进行的,随着环境给出当前状态和动作空间,基于 DQN 的智能体给出动作空间中各个动作的价值,并选出最大价值的动作执行,基于状态序列建模的对话策略如图 3 所示。

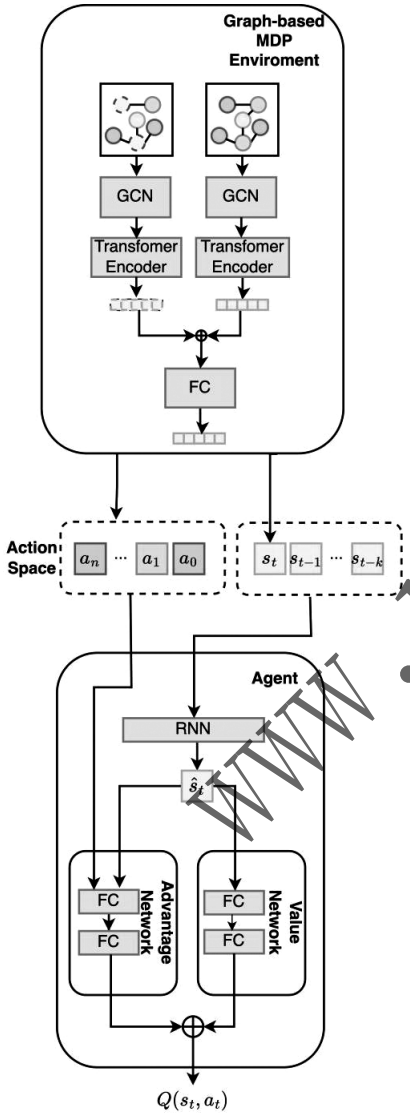

图 3　基于状态序列建模的对话策略

Fig. 3 Conversation policy based on state sequence modeling

2.2.1 状态序列建模

记本次对话的状态向量序列为 $S_t=\{\boldsymbol{s}_{t-k},\boldsymbol{s}_{t-k+1},\cdots,\boldsymbol{s}_t\}$,其中每一个状态向量都携带了特定轮次对话的细节信息。本文利用 RNN 将状态向量序列进行编码,以提取每一次对话轮次间的状态变化特征。这样的设计旨在捕获对话动态变化的本质特征,以改善推荐准确性和效率。本文采用门控循环单元[21](Gated Recurrent Unit, GRU)的原因是它对长序列数据有优异的记忆和学习能力,尤其在处理具有复杂时序依赖的任务时的表现更加优异。门控循环单元结构简单、计算高效且能解决长时序依赖问题,这些优点使它成为此类任务的理想选择。在实现细节上,状态向量通过 GRU 层进行处理,每一个状态都被更新并与前一个状态相关联。公式(12)表示状态向量序列通过 GRU 进行编码后的输出,记为 $\hat{\boldsymbol{s}}_t$。

$$\hat{\boldsymbol{s}}_t=\text{GRU}(\{\boldsymbol{s}_{t-k},\boldsymbol{s}_{t-k+1},\cdots,\boldsymbol{s}_t\}) \tag{12}$$

2.2.2 价值学习

最优动作价值函数 $Q^{*}(\boldsymbol{s}_t,a_t)$ 为采取最优策略 π^{*} 能得到的最大折扣期望汇报,根据 Bellman 方程,最优动作价值函数定义为公式(13),其中 γ 为折扣因子。

$$Q^{*}(\boldsymbol{s}_t,a_t)=\mathbb{E}_{s_{t+1}}\left[r_t+\gamma\max_{a_{t+1}\in A_{t+1}}Q^{*}(\boldsymbol{s}_{t+1},a_{t+1})\,|\,\boldsymbol{s}_t,a_t\right] \tag{13}$$

本文采用 Dueling Q-Network[22]的设定,将价值网络实现为状态价值网络和动作优势网络两个独立的神经网络,如公式(14)所示。其中:$f_{\theta_V}(\boldsymbol{s}_t)$ 是状态价值网络,θ_V 是网络的参数,$f_{\theta_A}(\boldsymbol{s}_t,a_t)$ 是动作优势网络,θ_A 是网络的参数,θ_S 是图表示学习的参数。

$$Q(\boldsymbol{s}_t,a_t)=f_{\theta_V}(\boldsymbol{s}_t)+f_{\theta_A}(f_{\theta_S}(\boldsymbol{s}_t),a_t) \tag{14}$$

在每轮对话中,智能体都能从基于图的 *MDP* 环境中得到当前时刻 t 的状态的表示向量 $\boldsymbol{s}_t=\boldsymbol{e}_s$ 和动作空间 A_t,接着通过状态序列建模将本次对话中的每轮对话状态向量进行编码,并得到包含状态变化特征的状态表示向量 $\hat{\boldsymbol{s}}_t$。智能体利用动作价值函数 $Q(\boldsymbol{s}_t,a_t)$ 估计动作空间 A 中的各个动作的价值,并使用 ε-greedy 策略从动作空间中选出要执行的动作 a_t。环境则根据智能体选择的动作给出动作 a_t 的奖励 r_t,然后转移到下一个状态 $\boldsymbol{s}_{t+1}$ 并更新动作空间 A_{t+1}。

记六元组 $(\boldsymbol{s}_t,a_t,r_t,\boldsymbol{s}_{t+1},A_{t+1},S_{t-1})$ 为一条经验,为了提高价值网络的学习效率,将每次对话得到的经验存放在重放记忆(Replay Memory)D 中,每次采样一个小批量用于更新网络参数。如公式(15)至公式(16)所示,本文通过最小化均方差损失的方式优化价值网络,其中 y_t 是 TD 误差。

$$L(\theta_Q,\theta_S)=\mathbb{E}_{\boldsymbol{s}_t,a_t,r_t,\boldsymbol{s}_{t+1},A_{t+1},S_{t-1}\sim D}\left[(y_t-Q(\boldsymbol{s}_t,a_t;\theta_Q,\theta_S))^2\right] \tag{15}$$

$$y_t=r_t+\gamma\max_{a_{t+1}\in A_{t+1}}Q(\boldsymbol{s}_{t+1},a_{t+1};\theta_Q,\theta_S) \tag{16}$$

3 实验(Experiment)

3.1 数据集

实验数据集采用 LastFM 数据集和 LastFM * 数据集。其中,LastFM 数据集是由 Last. fm 发布的音乐推荐数据集,包含来自 1 801 位用户的听歌记录。为了便于建模,LEI 等[11]将 LastFM 的原始属性人工手动合并成 33 个粗粒度属性组。同时,LEI 等[13]认为手动合并属性在实际应用中并非最佳实践,因此他们在 LastFM 数据集的基础上使用原始的属性重构出 LastFM * 数据集。LastFM 数据集和 LastFM * 数据集的数据量见表 2。

表2 实验数据集的数据量

Tab.2 The amount of data in the experimental dataset

类型/个	LastFM	LastFM *
用户数	1 801	1 801
项目数	7 432	7 432
交互数	76 693	76 693
属性数	33	8 438

3.2 实现细节

本文将最大对话轮次数 T 设为 15 轮，将推荐列表的大小 K 设为 10，并将每个数据集分割为 7∶1.5∶1.5 用于训练、验证和测试。本文使用 OpenKE[23] 中实现的 Trans-E 算法，在基于训练集构建的图上进行节点表示的预训练。对于实验中所有的基线方法，本文都使用用户模拟器[11]进行了 10 000 轮对话的在线训练。超参数的设置如下：嵌入大小和图表示学习的输出层大小分别设置为 64 和 100；GCN 层数 L_g，Transformer 层数 L_T 和 GRU 的层数 L_s 分别设置为 2、1 和 2；选中的候选属性 K_p 和项目 K_v 的数量都设置为 10；在 DQN 的训练过程中，经验回放缓冲区的大小为 50 000，从经验回放缓冲区中每次采样的小批量的大小为 128；学习率和 $L2$ 范数正则化系数分别设置为 10^{-4} 和 10^{-6}，使用 Adam 优化器；折扣因子 γ 和更新频率 τ 分别设为 0.999 和 0.01。

3.3 实验结果与分析

3.3.1 对比实验

为了验证本文所提出模型的有效性，比较了本文提出的模型与四个基线模型在 15 轮推荐准确率（$SR@15$）和平均推荐成功轮数（AT）指标上的性能表现，实验结果如表 3 所示。本文分别实现了三个不同版本的推荐模型，如表 3 中的(a)将用户负反馈信息分别建模为正反馈图和负反馈图；(b)在(a)的基础上实现了动态奖励函数；(c)又在(b)的基础上在智能体中使用 GRU 对历史对话状态序列进行编码。表 3 中的结果表明，本文提出的模型(c)在两个数据集上均取得了最佳结果，相较于次优模型分别在推荐成功率指标上提升了 21%和 13.7%，在平均推荐轮次数指标上降低了 1.56 轮和 2.35 轮。

表3 对比实验结果

Tab.3 Comparative experiment result

算法	LastFM		LastFM *	
	SR@15	*AT*	*SR@15*	*AT*
CRM[8]	0.325	13.75	0.580	10.79
EAR[11]	0.429	12.88	0.595	10.51
SCPR[13]	0.465	12.86	0.709	8.43
UNICORN[14]	0.535	11.82	0.788	7.58
(a)融合负反馈图	0.642	10.93	0.894	5.31
(b)(a)+动态奖励函数	0.695	10.87	0.900	5.30
(c)(b)+序列模型	**0.745**	**10.26**	**0.925**	**5.23**

3.3.2 不同轮次内的推荐成功率

图 4 和图 5 展示了在 LastFM 和 LastFM * 上，本文模型在每个对话轮次 t 下的推荐成功率（$SR@t$）。虽然动态奖励函数在推荐初期可能降低成功率，但是它能提升模型的长期推荐效果。这是因为，在初期，询问属性有助于了解用户偏好，进而缩小推荐范围。短期内这可能会降低成功率，但从长远看，能更精确地满足用户需求，提高推荐成功率。模型在不同数据集上使用了不同推荐策略。例如，在属性丰富的 LastFM * 上，模型更倾向于早期推荐，而在属性较少的 LastFM 数据集上则较为保守，前几轮几乎没有推荐。引入 GRU 对历史对话进行编码进一步提高了推荐效果，说明历史对话对于了解用户需求和偏好很重要，GRU 能有效捕获这些信息，为模型提供准确的上、下文。

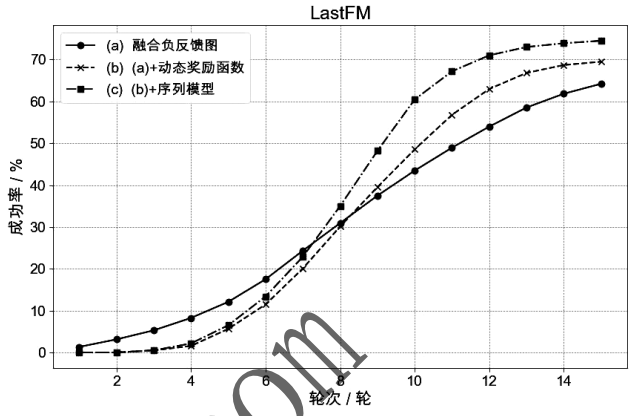

图 4 本文模型在 LastFM 数据集上每个不同轮次的推荐成功率

Fig. 4 Recommendation success rates of the text model on LastFM dataset for different rounds

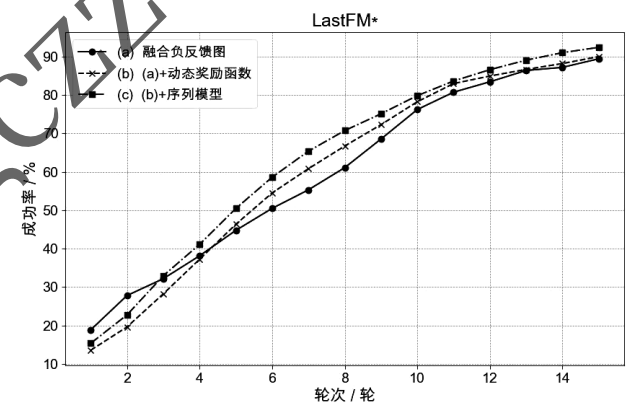

图 5 本文模型在 LastFM * 数据集上每个不同轮次的推荐成功率

Fig. 5 Recommendation success rates of the text model on LastFM * dataset for different rounds

4 结论(Conclusion)

本文通过深入探究对话式推荐系统中用户负反馈及对话历史信息的有效性，成功构建了一个全新的对话式推荐模型。负反馈图不仅为系统提供了更全面的用户偏好画像，动态奖励函数也成功地缓解了奖励稀疏的问题，促使系统能更好地从用户反馈中学习。此外，序列模型的引入进一步优化了推荐效果，证明了对话历史信息在揭示用户深层次需求和偏好中的关键作用。在实验中，该模型被应用于 LastFM 和 LastFM * 两个数据集，与最优的基线模型相比推荐成功率分别提升 21%和 13.7%，平均推荐轮次数也分别降低了 1.56 轮和 2.35 轮。尽管本文提出的模型考虑了用户负反馈和对话历史信息，但如何有效地处理用户历史反馈的漂移和变化，并在不断学习的过程中保持推荐系统的稳定性，是一个值得未来深入研究的方向。

参考文献(References)

[1] KOREN Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]//LI Y, LIU B, SARAWAGI S. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2008: 426-434.

[2] SALAKHUTDINOV R, MNIH A, HINTON G. Restricted Boltzmann machines for collaborative filtering[C]// GHAHRAMANI Z. Proceedings of the 24th International Conference on Machine Learning. New York: ACM, 2007: 791-798.

[3] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session-based recommendations with recurrent neural networks[DB/OL]. (2016-03-29)[2023-02-21]. http://arxiv.org/abs/1511.06939.

[4] 刘建伟,宋志妍. 循环神经网络研究综述[J]. 控制与决策,2022,37(11):2753-2768.

[5] HE X, HE Z, DU X, et al. Adversarial Personalized Ranking for Recommendation[C]//COLLINS-THOMPSON K, MEI Q Z, DAVISON B, et al. Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. New York: ACM, 2018: 355-364.

[6] WANG X, HE X N, WANG M, et al. Neural graph collaborative filtering[C]//PIWOWARSKI B, CHEVALIER M, GAUSSIER E, et al. Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2019: 165-174.

[7] 赵梦媛,黄晓雯,桑基韬,等. 对话推荐算法研究综述[J]. 软件学报,2022,33(12):4616-4643.

[8] SUN Y M, ZHANG Y. Conversational recommender system[C]//COLLINS-THOMPSON K, MEI Q Z, DAVISON B, et al. Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. New York: ACM, 2018: 235-244.

[9] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[10] RENDLE S. Factorization machines[C]//IEEE. Proceedings of the 2010 IEEE International Conference on Data Mining. Piscataway: IEEE, 2010: 995-1000.

[11] LEI W Q, HE X N, MIAO Y S, et al. Estimation-action-reflection: towards deep interaction between conversational and recommender systems[C]//CAVERLEE J, HU X B, LALMAS M, et al. Proceedings of the 13th International Conference on Web Search and Data Mining. New York: ACM, 2020: 304-312.

[12] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: bayesian personalized ranking from implicit feedback [C]//MCALLESTER D. Proceedings of the Twenty-fifth Conference on Uncertainty in Artificial Intelligence. Arlington: AUAI Press, 2009: 452-461.

[13] LEI W Q, ZHANG G Y, HE X N, et al. Interactive path reasoning on graph for conversational recommendation[C]// GUPTA R, LIU Y, SHAH M, et al. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 2073-2083.

[14] DENG Y, LI Y L, SUN F, et al. Unified conversational recommendation policy learning via graph-based reinforcement learning[C]//DIAZ F, SHAH C, SUEL T, et al. Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2021: 1431-1441.

[15] YING R, HE R N, CHEN K F, et al. Graph convolutional neural networks for web-scale recommender systems[C]// GUO Y, FAROOQ F. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 974-983.

[16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[DB/OL]. (2017-12-06)[2023-03-10]. https://arxiv.org/abs/1706.03762v5.

[17] FAN J Q, WANG Z R, XIE Y C, et al. A theoretical analysis of deep Q-learning[DB/OL]. (2020-02-24)[2023-03-14]. https://arxiv.org/abs/1901.00137v3.

[18] 潘润超,虞启山,熊松霏,等. 基于深度图神经网络的协同推荐算法[J]. 计算机应用,2023,43(9):2741-2746.

[19] BORDES A, USUNIER N, GARCÍA-DURÁN A, et al. Translating embeddings for modeling multi-relational data [C]//BURGES C J C, BOTTOU L, WELLING M, et al. Proceedings of the 26th International Conference on Neural Information Processing Systems: Volume 2. New York: Curran Associates Inc., 2013: 2787-2795.

[20] 余力,杜启翰,岳博妍,等. 基于强化学习的推荐研究综述[J]. 计算机科学,2021,48(10):1-18.

[21] DEY R, SALEM F M. Gate-variants of Gated Recurrent Unit (GRU) neural networks[C]//IEEE. Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems. Piscataway: IEEE, 2017: 1597-1600.

[22] WANG Z Y, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[C]// BALCAN M F, WEINBERGER K Q. Proceedings of the 33rd International Conference on International Conference on Machine Learning : Volume 48. New York: ACM, 2016: 1995-2003.

[23] HAN X, CAO S L, LV X, et al. OpenKE: an open toolkit for knowledge embedding[C]//BLANCO E, LU W. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Brussels: Association for Computational Linguistics, 2018: 139-144.

作者简介:

杨　畅(1998-),男,硕士生。研究领域:推荐系统。

姚　越(1998-),女,硕士生。研究领域:增量学习,自然语言处理。

方霖枫(1999-),男,硕士生。研究领域:推荐系统。

周仁杰(1983-),男,博士,副教授。研究领域:数据智能,自然语言处理,推荐系统。本文通信作者。