doi: 10.19920/j.cnki.jmsc.2026.06.004

面向高卷入度产品的对话推荐系统①
——基于模块化多阶段方法

柴一栋[1,2]，周永行[1,2]，姜元春[1,2*]，刘春丽[1,2]，袁　昆[1,2]，刘业政[1,2]
（1. 合肥工业大学管理学院，合肥 230009；2. 网络空间行为与管理安徽省哲学社会科学重点实验室，合肥 230009）

摘要： 随着汽车、家电等高卷入度产品的日益丰富，如何设计推荐系统辅助消费者选择高卷入度产品成为重要的研究问题. 本研究围绕高卷入度产品功能属性复杂、价值高昂和消费者咨询具有多阶段性的特点，提出一种面向高卷入度产品的模块化多阶段对话推荐方法. 本研究采用“系统提问－用户回答”的人机交互范式，通过提问获取用户偏好并生成推荐结果. 针对多阶段咨询问题，本研究引入阶段状态建立多阶段强化学习模型；针对功能属性复杂加剧偏好获取和产品推荐任务的矛盾性问题，本研究构建了基于强化学习的模块化对话推荐系统，包括基于强化学习的对话策略和属性选择策略，以及基于知识图谱和理想点的产品推荐方法. 基于国内知名汽车论坛真实购买数据和仿真用户交互数据的实验表明，所提方法能够在降低交互次数的情况下，取得更高的推荐准确率.

关键词： 高卷入度产品；模块化对话推荐；多阶段强化学习；知识图谱；理想点法
中图分类号： C93　　**文献标识码：** A　　**文章编号：** 1007－9807(2026)06－0046－17

0　引　言

高卷入度产品是指消费者在购买过程中需要投入大量时间和精力，反复权衡产品各方面性能，才能做出购买决策的产品，如汽车、冰箱、洗衣机等[1,2]. 该类产品通常具有价值高昂、功能属性复杂、使用周期长等特点，因此消费者在购买该类商品时，经常陷入选择困境[3]. 高卷入度产品的日益丰富进一步增加了消费者的选择难度[4,5]. 因此，如何辅助消费者进行高卷入度产品的购买决策成为重要的研究问题[6].

作为辅助消费者购买决策的重要手段，个性化推荐近年来得到了广泛的研究与应用. 研究者提出了协同过滤推荐、基于内容的推荐以及混合推荐等方法[7-9]. 由于具有更好的非线性变换能力、表征学习能力以及序列建模能力，深度学习近年来成为构建推荐系统的主流方法[10,11]. 现有基于深度学习的个性化推荐方法利用交易记录、评分数据等信息，对用户偏好和产品特征进行建模，取得了良好的推荐效果[12]. 但是，对于高卷入度产品而言，消费者的购买频率极低，导致消费者的高卷入度产品历史购买信息几乎为零. 因此，以丰富的历史购买数据为基础的深度学习推荐方法难以实现高卷入度产品的推荐任务.

对话推荐系统(conversational recommendation system)通过机器提问与消费者回答的交互过程获取消费者偏好，克服历史购买数据缺失的难题. 对话推荐策略通常采用“系统提问－用户回答”的范式[13]，利用系统与消费者的实时交互数据捕捉消费者细粒度偏好特征，进而给出精准的推荐结果[14]，在面向电影、音乐等低卷入度产品推荐

① 收稿日期：2023－04－19；修订日期：2024－07－30.
基金项目：国家自然科学基金资助项目(72342011；72322019；72171071；72101079；72101076；72271083).
通讯作者：姜元春(1980—)，男，山东莱西人，博士，教授，博士生导师. Email：ycjiang@hfut.edu.cn

场景中取得了优异表现[15,16].

然而,已有对话推荐方法在高卷入度产品推荐情境下依然存在两点不足. 首先,高卷入度产品功能属性复杂、价值高昂的特点加剧了对话推荐矛盾性问题. 具体而言,由于高卷入度产品功能属性复杂,为了提高推荐的准确性,对话推荐系统需要进行更多的提问以获取消费者偏好. 同时,价值高昂也对推荐准确性提出更高要求. 但是,过多的提问次数会增加用户的交互负担,让用户失去耐心,即加剧提问与推荐的矛盾性. 其次,与低卷入度产品相比,高卷入度产品价值高昂的特点使高卷入度产品的消费者呈现出不同的消费模式. 消费者通常不会在第一次到店咨询商家后便直接购买高卷入度产品,而是反复多次到店咨询商家以充分了解情况. 在不同咨询阶段中,消费者的心态是不同的,如第一次咨询更多是为了了解信息,而越往后其购买意愿通常越强,这就需要对话推荐方法考虑不同对话轮次,即消费者咨询的多阶段性问题. 本研究使用“阶段”代指用户在购买过程中多次到店与商家咨询,使用“轮次”代指用户在每一阶段中的多次交互. 因此,如何解决高卷入度产品对话推荐中的矛盾性问题以及如何建模消费者咨询的多阶段性,进而设计面向高卷入度产品的对话推荐系统有待进一步探索.

针对以上问题,本研究提出了一种新颖的面向高卷入度产品的对话推荐方法. 首先,本研究设计模块化对话推荐模型,包括三个决策函数: 1) 基于强化学习的动作函数,决策每次交互的动作形式,即向用户提问还是向用户推荐; 2) 基于强化学习的属性提问函数,决策向用户提问什么属性; 3) 基于知识图谱和理想点的产品推荐函数,决策应该推荐什么产品. 通过分别优化三个决策函数,解决提问与推荐任务的矛盾性问题. 其次,本研究针对咨询多阶段性的问题,提出多阶段对话推荐方法,通过为强化学习引入阶段变量和调整损失函数,对不同咨询阶段的消费者设计不同的决策模型. 本研究基于国内知名汽车论坛的真实数据,设计用户交互仿真实验,发现所提模型可以协同优化用户偏好获取和产品推荐,在较少交互次数中获得更好的推荐结果.

本研究的创新之处在于: 1) 首次设计了针对高卷入度产品的模块化多阶段对话推荐系统,包括基于强化学习的对话动作函数、基于强化学习的提问函数以及基于知识图谱和理想点的推荐函数; 2) 首次提出基于知识图谱和理想点的产品推荐方法,综合消费者正向偏好与负向偏好更有效地进行产品推荐; 3) 首次设计了针对高卷入度产品的多阶段对话推荐系统,根据阶段状态对交互策略进行动态优化调整,更好地适用高卷入度产品用户的购买行为.

1　文献综述

1.1　高卷入度产品推荐方法

产品的卷入度是消费者对产品的感知重要性与兴趣[1]. 相对于低卷入度产品,高卷入度产品通常具有功能属性复杂、价值高昂、使用周期长等特点. 已有研究表明,消费者的高卷入度产品购买决策过程受到功能、价格[17]、可靠性[18]、在线口碑[19]、情感价值[20]和社会价值[21]等多种因素影响. 同时,消费者在高卷入度产品购买决策过程具有多阶段性的特点[22]. 消费者的购买决策更加谨慎,他们通常会反复多次到店咨询商家以不断了解新情况. 随着消费者掌握的信息不断增多,其内在心态也会发生变化,如相比于首次到店咨询,通常越往后阶段的消费者获取信息的意愿越弱而决定购买产品的意愿越强.

个性化推荐是辅助消费者产品购买决策的有效手段. 然而,功能属性复杂,价值高昂以及购买决策过程多阶段等特点使得设计高效的高卷入度产品推荐方法面临挑战. 首先,高卷入度产品功能属性复杂以及消费者决策多阶段的特征要求推荐系统能够在不同阶段准确地分析出消费者在不同功能属性上的偏好. 针对此,已有研究提出了基于多属性效用理论[23]的推荐方法,该类方法对不同功能属性的效用加权求得产品的总效用,并将总效用最高的产品推荐给消费者. 例如 Huang 等设计了一种基于径向基网络的效用函数来获取用户在不同属性上的偏好为消费者推荐笔记本电脑[24]. 然而,以上方法依赖于消费者效用函数的设计,但是设计适用于所有消费者的效用函数非常困难. 其次,高卷入度产品价值高、使用生命周期长,消费者对高卷入度产品的购买频率极低,导

致产品的历史购买数据极其稀疏[25]. 为了缓解该问题,目前主流方法是采用基于批判的(critiquing-based)交互式方法获取消费者的偏好信息,让消费者在交互界面上添加对产品属性的限制. 例如,Burke 等基于单元批判设计了 FindMe 系统,为用户推荐数码照相机[26];Pear 等提出基于案例的批判方法,辅助用户复杂产品的购买决策[27];James 等提出了一种动态增量批判的方法[28]. 然而,基于批判的交互方法缺乏个性化,且只能实现对产品的检索. 事实上,关于推荐领域数据稀疏性问题的研究历来已久,引入外部信息以减轻数据稀疏性是常用策略. 例如,Dong 等提出基于用户和基于产品的混合协同过滤方法[8];Chen 等引入产品的评论信息并将评论进行聚类,以增强产品的特征表示[29];考虑高卷入度产品在线口碑的动态性,Jiang 等提出了基于评论的混合协同过滤算法[9];Capdevila 等引入用户关系、产品知识等信息来缓解数据稀疏性问题[30]. 引入外部信息的方法需要借助于大量的交互数据,然而对高卷入度产品而言,由于缺乏历史交互记录数据,上述方法难以奏效.

在推荐过程中引入主动对话,建立对话推荐系统,通过对话而非历史购买记录获取消费者偏好,不仅可以应对数据稀疏性的挑战,也可以获取消费者对不同功能属性的偏好,为辅助消费者进行功能属性复杂的高卷入度产品多阶段购买决策提供了可行思路. 对话推荐系统综述如下.

1.2 对话推荐系统

对话推荐系统是通过与消费者对话询问相关问题动态地更新用户偏好,解决历史购买记录稀疏问题,从而达到为用户提供推荐服务目的的智能系统[31]. 对话推荐系统目前已经在各类智能客服平台得到了广泛应用[32].

现有对话推荐系统的研究主要分为两类:1)基于对话回复生成的方法;2)基于对话推荐策略的方法. 第一类研究将对话推荐系统视为一种任务导向的对话系统,其主要任务是根据用户的表述生成相应的回复. 此类研究通常采用端到端的方法建立统一的模型,将产品的名称视作字词,并通过复制机制[33]、指针机制[34]等方法加入到生成的对话中. 由于回复的生成需要根据用户的表述,通过条件概率的方式逐字生成语句,因此,该类方法会倾向生成训练语料中出现频率高却无意义的回复[33]. 第二类研究建立模块化的对话推荐系统,将对话内容理解生成与推荐策略分离,并且只关注交互推荐策略. 该类研究通过提问获取用户关于产品属性的偏好,并根据用户对属性的偏好产生推荐[15]. 具体而言,这类研究使用“系统提问 - 用户回答”的对话交互范式[13],将对话推荐分解成三个问题[15],每一轮次采取什么动作(继续提问,还是进行推荐)? 提问什么(询问产品的哪个属性)? 推荐什么产品? 针对“继续提问还是进行推荐”的问题,系统需要根据当前的状态,选择向消费者提问或者推荐. Zhang 等使用基于规则的方法来选择当前交互动作. 例如,每提问 m 次就进行一次推荐[35]. 考虑到每次交互动作的选择是动态序列决策问题. Deng 等通过信息熵的方法选择一个属性集合,使用基于用户偏好的方法选择一个产品集合,构成由产品和属性构成的动作集合,再利用深度策略网络的强化学习方法从动作集合中决策出当前应该选择的动作[36];Lei 等人将推荐动作与所有属性映射在统一的动作空间中,并利用强化学习决策出当前交互应该选择进行推荐还是提问[15];Greco 等将对话推荐系统划分为易于管理的任务,并使用层次强化学习(hierarchical reinforcement learning)来优化每次交互应该进行决策的任务[37]. 针对“提问什么”的问题,系统需要选择一个产品属性展开提问,以获取消费者关于该属性的偏好. 属性问题的选择也是动态序列决策问题,当前轮次属性选择的决策将会影响后续交互的决策[38]. Sun 等利用用户在历史交互中表达的偏好属性通过深度策略网络的强化学习方法优化属性选择过程[39];Li 等将产品和属性映射在统一的动作空间中,并利用基于汤普森采样的强化学习选择当前应该提问的属性[16];Dhingra 等基于贪心思想,考虑在每次提问时能够达到当前轮次最优决策提出了选择当前状态下信息熵最大的属性进行提问[40];Xia 等利用学习者的行为数据对在线教育平台中的问题池中的学习路径进行建模,并利用马尔科夫链方法为新用户选择每次交互学习的问题,从而提高长期收益[41];Shi 等利用知识图谱来组织学习对象. 并利用启发式的方法为在线教育平台用户选择个性化学习问题[42]. 针对“推荐什么产品”的问题,系统需要根据当前获取的用户偏好信息,为用户产

生推荐列表. Lei 等利用用户与产品的历史交互数据预训练了基于属性感知的贝叶斯个性化排序算法,并在对话交互过程中结合用户的属性偏好信息进行个性化推荐[15]; Li 等利用用户与产品历史交互数据训练基于自编码器(autoencoder)的方法为用户进行推荐产品[31]. 考虑用户产品交互数据的稀疏性问题,Lei 等人利用产品的领域知识构建了知识图谱,通过对话交互获取的属性偏好信息在知识图谱上进行交互路径推理,使用随机游走的方式为用户推荐产品[43].

虽然研究者提出各类对话推荐方法,并解决了电影、音乐等低卷入度产品数据稀疏性问题[44]. 然而,高卷入度产品推荐不仅存在数据稀疏性的问题,而且面临着产品功能属性复杂,消费者需要进行多阶段决策等特征的挑战. 如何针对以上挑战,构建面向高卷入度产品的对话推荐方法,现有研究尚存理论空白.

2　面向高卷入度产品的模块化多阶段对话推荐系统

2.1　问题描述

2.1.1　面向高卷入度产品的对话推荐系统

给定用户 $u \in U$,高卷入度产品 $v \in V$,v_u 表示用户 u 购买过的产品,P_{v_u} 表示产品 v_u 的属性集合. 推荐系统为每个消费者 u 从 V 中识别出该消费者可能感兴趣的产品,其核心在于分析 u 对 v 的偏好 $s_{v,u}$,如式(1)所示

$$s_{v,u} = f_S(G_v, A_u) \tag{1}$$

其中 f_S 表示偏好计算函数,G_v 表示高卷入度产品 v 的表征,A_u 表示用户 u 的表征. G_v 和 A_u 由表征模型 f_D 根据已知交互数据 D 得到

$$[G_v, A_u] = f_D(D, v, u) \tag{2}$$

结合式(1)和式(2),记 $f_R = f_S \circ f_D$,则推荐任务可以表示为

$$s_{v,u} = f_R(D, v, u) \tag{3}$$

由于高卷入度产品的购买稀疏性,系统已知交互数据为空,即 $D = \varnothing$. 因此,推荐系统一方面需要通过提问函数主动提问用户以获取高质量交互数据 D,记提问函数为 f_I; 另一方面需要基于交互数据 D 分析产品和用户的表征,基于 f_R 进行产品推荐. 由于每次交互只能提问或推荐,因此需要选择执行哪一个动作,相应的动作函数记为 f_C. 因此,面向高卷入度产品的对话推荐系统包括三个模块,即动作函数 f_C、提问函数 f_I 和推荐函数 f_R. 面向高卷入度产品的对话推荐系统的流程图如图 1 所示.

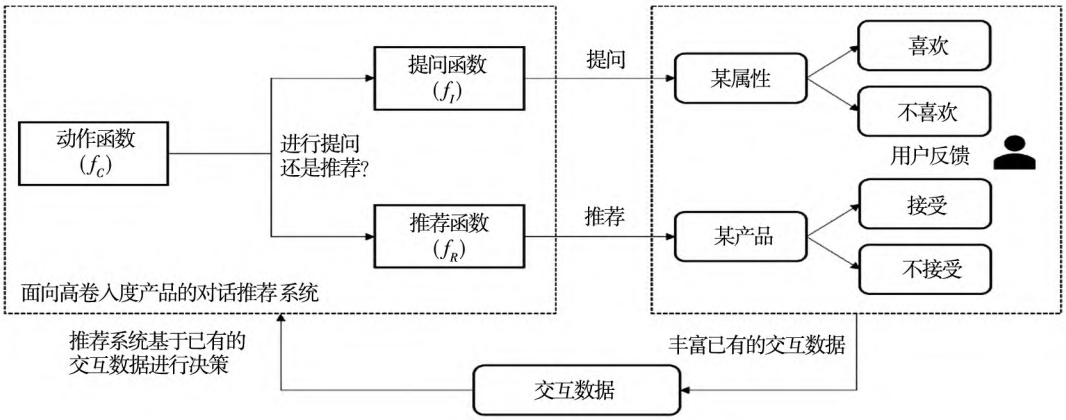

图 1　面向高卷入度产品的对话推荐系统的流程图

Fig. 1 Flowchart of the conversational recommendation system for high-involvement products

2.1.2　面向高卷入度产品对话推荐方法的矛盾性和多阶段性问题

数据 D 直接影响用户和产品表征学习的效果,进而影响推荐精度. 为了提升推荐函数 f_R 的准确性,提问函数 f_I 需要通过多次提问以获得丰富的交互数据. 然而获取交互数据会引入额外成本(如反复提问让用户失去耐心导致用户流失),因此提问函数 f_I 需要减少交互次数[38]. 因此,数据获取与产品推荐之间存在着矛盾性. 虽然许多交互式产品推荐均面临矛盾性的问题,但是高卷

入度产品具有功能属性复杂的特点，使得高卷入度产品的矛盾性问题更加突出. 功能属性复杂使得推荐系统需要获取更多的交互数据 D 才能精确推测消费者需求，这增加了数据获取任务的挑战. 由于高卷入度产品价值高昂，消费者错误的购买决策会造成巨大经济损失，因此消费者对推荐错误的容忍性较低，这增加了产品推荐任务的挑战. 因此，高卷入度产品的以上特点加剧了数据获取与产品推荐之间的矛盾性.

与电影、音乐等低卷入度产品相比，高卷入度产品价值高昂，消费者在购买过程中会更加谨慎，因此在购买高卷入度产品前通常会多次反复咨询商家，如许多用户会隔一段时间后再次咨询商家，以充分了解产品的情况. 同时由于消费者在不同咨询阶段时的信息掌握程度以及购买意向存在显著差异[22]，因此需要对处于不同咨询阶段的消费者设计不同的策略，从而更有利于理解和精准刻画用户购买决策过程，更高效地为用户提供精准推荐.

2.2 解决方案：基于多阶段强化学习的模块化对话推荐系统

假设消费者最多需要进行 K 阶段咨询，在每一阶段 k 的推荐系统的损失记为 L_{total}^{k}，多阶段推荐系统的总损失为不同阶段损失之和，即 $L_{\text{total}} = \sum_{k=1}^{K} L_{\text{total}}^{k}$. 本研究建立基于多阶段强化学习的模块化对话推荐系统降低 L_{total}. 在每一阶段 k 中，损失函数 L_{total}^{k} 由动作函数、提问函数和推荐函数共同决定，因此本研究通过优化每一阶段中的动作函数、提问函数和推荐函数以降低 L_{total}.

首先，通过基于强化学习的动作函数，预测不同动作的期望损失，从而决定每次交互中选择提问动作还是推荐动作（即模块一：基于强化学习的动作选择）. 具体而言，如果系统当前掌握的用户偏好信息不足，那么动作函数继续采取询问动作 a_{ask} 以获取更多偏好信息，减少推荐损失. 如果系统充分掌握了用户的偏好信息，那么动作函数采取推荐动作 a_{rec}，以减少交互损失. 本研究将与模块一直接相关的损失记为 L_{dialog}^{k}，表示由于决策数据获取动作和推荐产品动作而引发的损失. 如果用户在本轮交互中途退出，而且再隔一段时间后（如几日）后又到店咨询，则交互进入新的阶段，模块一中的状态变量进行更新，从而优化不同阶段的动作函数.

其次，由于用户提问是直接询问用户关于某属性的喜好，因此优化提问函数等价于优化用户偏好属性的选择方法. 属性选择是一个序列决策的过程，每一次对用户偏好的提问都会影响用户的状态表示及最终推荐的准确性. 鉴于强化学习建模序列决策过程的优势[45, 46]，本研究依然采用强化学习来确定用户偏好（即模块二，基于强化学习的提问函数）. 该模块从属性候选集中选择出一个属性向用户 u 进行提问，以获取该用户关于该属性的偏好. 模块二直接相关的损失记为 $L_{\text{preference}}^{k}$，表示由于频繁提问引发的数据获取的成本.

最后，本研究利用产品与属性间的领域知识，建立知识图谱来刻画产品和属性的表征，基于理想点法[47]对候选产品进行排序. 正理想点由每个属性的最优取值构成的虚拟最优产品，负理想点是由每个属性的最劣取值构成的虚拟最劣产品. 与正理想点距离近且与负理想点距离远的即为现实可行的最优产品（用户最感兴趣的产品）（即模块三：基于知识图谱和理想点的推荐函数）. 该模块生成推荐列表 rec_{list}，如果当推荐列表中有用户感兴趣的产品时，用户接受则交互结束. 如果用户拒绝推荐列表并引发相应损失记为 L_{rec}^{k}，同时更新模块一和模块二中的状态信息，进入下一次交互.

与端到端的模型直接针对总体损失 L_{total} 进行优化求解不同，基于多阶段强化学习的模块化对话推荐系统将总体损失 L_{total} 进行分解，首先将 L_{total} 拆解为不同阶段的损失 L_{total}^{k}，再将 L_{total}^{k} 进一步拆解为 L_{dialog}^{k}、$L_{\text{preference}}^{k}$ 和 L_{rec}^{k}. 由于每个模块的参数众多且存在相互影响，导致直接对总体损失求解较为困难. 特别地，高卷入度产品功能属性复杂、价值高昂和消费者咨询多阶段的特点进一步加剧了直接求解的难度. 因此通过模块化的思路，本研究通过将复杂问题简化为一系列相对容易处理的子问题，实现对目标的优化.

2.3 基于强化学习的动作函数

基于强化学习动作函数 f_C 中不同阶段 k 轮次 t 应当采取的对话动作（a_{ask} 或 a_{rec}）由状态 $\boldsymbol{state}_t^{k,u}$

决定. 由于消费者在开始阶段对高卷入度产品的购买可能性较低,推荐成功率也较低,因此阶段状态 $stage$ 会影响强化学习的对话策略. 根据已有研究[15],产品候选集大小 $|V_t^{k,u}|$、属性候选集信息熵均值 $AH(s)$、当前轮次 t、用户偏好属性的数量 $|P_{k,u,t}^{+}|$、最近两轮的候选集变化率 $\frac{|V_t^{k,u}|-|V_{t-1}^{k,u}|}{|V_{t-1}^{k,u}|}$ 均会影响对话动作的选择. 候选集越大、属性信息熵均值越大、已交互轮次越小、用户偏好的属性数越少、候选集大小变化率越大时,系统对用户偏好的不确定性越大,采取 a_{ask} 的收益可能更大; 反之亦然. 因此,本研究构造状态 $\boldsymbol{state}_t^{k,u}$ 函数如下

$$\boldsymbol{state}_t^{k,u}=\left[stage,|V_t^{k,u}|,AH(s),t,|P_{k,u,t}^{+}|,\frac{|V_t^{k,u}|-|V_{t-1}^{k,u}|}{|V_{t-1}^{k,u}|}\right] \tag{4}$$

当系统采取动作 a_t^k 后,模型得到用户环境相应的损失反馈 r_t^k. 当系统采取推荐动作 a_{rec} 并且推荐成功时获得的损失反馈为 $r_t^k=r_{rec_{succ}}$,推荐失败时获得损失反馈 $r_t^k=r_{rec_{fail}}$; 当系统采取提问动作 a_{ask} 并且选择了用户偏好的属性时获得的损失反馈为 $r_t^k=r_{ask_{succ}}$,否则获得损失反馈 $r_t^k=r_{ask_{fail}}$; 当交互达到最大次数 $\max Turn$ 时,得到的损失反馈为 $r_t^k=r_{quit}$. 假设 γ 为折扣率,累计损失反馈 L_{dialog}^k 如式(5)所示

$$L_{\text{dialog}}^k=\sum_{t=1}^{T}\gamma^t r_t^k \tag{5}$$

系统优化的目标是 k 阶段及以后阶段的累计损失反馈 L_{dialog}^k 之和(记为 $L_{\text{dialog}}^{k:K}$)最小. 优化 $L_{\text{dialog}}^{k:K}$ 损失函数需要知道第 t 次采取不同动作时,产生的期望累计损失(即 Q 值). 由于本研究中用户的状态空间为连续的且状态数量是无限的,本研究采用 deep Q-network(DQN)[48],通过构造两层的神经网络拟合状态、动作与 Q 值的关系,如式(6)所示

$$Q(\boldsymbol{state}_t^{k,u},a_t^k;\theta^k)=\mathrm{ReLU}(\boldsymbol{W}_2^k(\mathrm{ReLU}(\boldsymbol{W}_1^k\boldsymbol{state}_t^{k,u}+\boldsymbol{b}_1^k))+b_2^k) \tag{6}$$

其中 $\boldsymbol{\theta}^k=\{\boldsymbol{W}_1^k,\boldsymbol{W}_2^k,\boldsymbol{b}_1^k,b_2^k\}$ 为可训练的参数.

由于 ε-贪心策略可以对不同状态下所有动作进行充分的探索,因此本研究根据 ε-贪心策略,选择动作 a_t^k,如式(7)所示

$$a_t^k=\begin{cases}\arg\max\limits_{a_t^k}Q(\boldsymbol{state}_t^{k,u},a_t^k;\boldsymbol{\theta}^k), & c>1-\varepsilon\\ \mathrm{randomchoice}(a_{rec},a_{ask}), & \text{否则}\end{cases}$$
$$c\sim U(0,1) \tag{7}$$

利用时序差分法优化神经网络参数 $\boldsymbol{\theta}^k$[49],其中损失函数为

$$\mathrm{loss}_{\text{dialog}}^k=\left[Q(\boldsymbol{state}_t^{k,u},a_t^k;\boldsymbol{\theta}^k)-\left(r_t^k+\gamma\max_{a_{t+1}}Q(\boldsymbol{state}_t^{k,u},a_{t+1}^k;\boldsymbol{\theta}^k)\right)\right]^2 \tag{8}$$

通过 $\mathrm{loss}_{\text{dialog}}^k$ 优化动作函数,降低累计损失反馈 $L_{\text{dialog}}^{k:K}$,得到高质量的动作函数 f_C.

2.4　基于强化学习的提问函数

提问函数 f_I 的目标在于优化属性选择策略,从而能够通过较少的交互轮次准确识别用户的属性偏好. 用户属性偏好状态使用用户偏好的属性集 $P_{k,u,t}^{+}$ 以及所有候选属性的信息熵 $H_t^{k,u}$ 表示,即 $\overline{\boldsymbol{state}}_t^{k,u}=(P_{k,u,t}^{+},H_t^{k,u})$,其中属性 p_i 的信息熵 $H_{t,i}^{k,u}$ 的计算如式(9)所示

$$H_{t,i}^{k,u}=\begin{cases}-(c(p_i)\log_2(c(p_i)))-(1-c(p_i))\times\log_2(1-c(p_i)), & p_i\in P_t^{k,u}\\ 0, & \text{否则}\end{cases} \tag{9}$$

其中 $P_t^{k,u}$ 表示当前交互系统可以进行提问的属性候选集,$c(p_i)$ 表示属性 p_i 的概率. 提问函数使用参数为 $\overline{\boldsymbol{\theta}}^k=\{\boldsymbol{W}_3,\boldsymbol{b}_3\}$ 的神经网络作为策略函数,输出 $\overline{Q}(\overline{\boldsymbol{state}}_t^{k,u},p_{t,i}^{k,u};\overline{\boldsymbol{\theta}}^k)$ 表示选择属性 $p_{t,i}^{k,u}$ 对用户 u 进行提问能够带来的价值. 如果提问的是用户偏好的属性,可以用该属性构建用户的偏好表征,因此可以更快速地获取用户偏好,从而提高用户交互的效率; 如果提问的是信息熵大的属性,可以更大程度地降低产品候选集的大小,更快地确定用户感兴趣的产品. 本研究中 $\overline{Q}(\overline{\boldsymbol{state}}_t^{k,u},p_{t,i}^{k,u};\overline{\boldsymbol{\theta}}^k)$ 的计算如式(10)所示

$$\overline{Q}(\overline{\boldsymbol{state}}_t^{k,u},p_{t,i}^{k,u};\overline{\boldsymbol{\theta}}^k)=\mathrm{ReLU}(\boldsymbol{e}_{p_{t,i}^{k,u}}\boldsymbol{v}_t^{k,u}+H_{t,i}^{k,u}) \tag{10}$$

其中 $\boldsymbol{e}_{p_{t,i}^{k,u}}$ 表示属性 p_i 的嵌入式表示(由 Translating on Hyperplanes(TransH)[50] 方法得到,见下节),$v_t^{k,u}$ 表示当前交互用户 u 对产品属性的个性化偏好,由式(11)得到

$$\boldsymbol{v}_t^{k,u}=\mathrm{ReLU}\left(\frac{1}{|P_{k,u,t}^{+}|}\boldsymbol{W}_3\sum_{p\in P_{k,u,t}^{+}}\boldsymbol{e}_p+\boldsymbol{b}_3\right) \tag{11}$$

提问函数使用 ε-贪心策略决策选择的属性

$p_t^{k,u}$，当用户喜欢属性 $p_t^{k,u}$ 时，得到即时反馈 $r_t^k = r_{ask_{succ}}$，否则得到即时反馈 $r_t^k = r_{ask_{fail}}$. 对 $\overline{Q}(\overline{\boldsymbol{state}}_t^{k,u}, p_t^{k,u}; \overline{\boldsymbol{\theta}}^k)$ 的准确拟合可以更准确地得到用户对于属性的偏好信息以及更加快速地降低产品候选集大小. 类似地，本研究使用时序差分方法进行优化，其中损失函数为

$$L_{\text{preference}}^k = \left[\overline{Q}\left(\overline{\boldsymbol{state}}_t^{k,u}, p_t^{k,u}; \overline{\boldsymbol{\theta}}^k \right) - \left(r_t^k + \gamma \max_{p_{t,i}^{k,u}} \overline{Q}(\overline{\boldsymbol{state}}_t^{k,u}, p_t^{k,u}; \overline{\boldsymbol{\theta}}^k) \right) \right]^2 \tag{12}$$

2.5 基于知识图谱和理想点的推荐函数

推荐函数 f_R 的目标在于利用产品领域知识和已获取的交互数据 D 学习高卷入度产品表征和用户偏好表征，做出推荐决策以降低每一阶段的推荐损失 L_{rec}^k. 由于交互数据稀疏，传统方法难以利用交互数据直接对产品、属性以及用户的偏好进行直接表征. 知识图谱在个性化推荐方法研究中作为外部知识数据，在解决交互数据稀疏性问题中发挥了重要作用[36,51]. 因此，本研究利用高卷入度产品丰富的的领域知识(如每款产品都有详细的功能配置信息)构建知识图谱 G，并利用知识图谱对产品和属性进行表征学习得到产品和属性的嵌入式向量表示. 知识图谱由三元组构成，其中的节点表示产品和属性，边表示产品与属性之间的关系，因此三元组(产品，关系，属性)表示节点产品与属性之间的关系. 以汽车产品知识图谱为例，如图 2 所示，(“宝马 3 系”，“百公里油耗”，“6L ~ 10L”)表示宝马 3 系车型的每百公里油耗是 6L ~ 10L. 在知识图谱中具有共同连接的节点相似性较高，例如“宝马 3 系”车型与“奥迪 A4L 车型”都连接了“中型车”，“轿车”等属性节点，因此，这两款汽车产品具有较高的相似度. 由于知识图谱节点之间具有多种关系(如“一对多”和“多对一”等)，而 TransH 模型能够让同一个节点在不同的关系下拥有不同的表示，因此，本研究采用基于距离的“翻译”模型 TransH[50]. 将关系 r 的表征记为 $\boldsymbol{e}_r$，将产品 v、属性 p 在关系 r 空间中的表征分别记为 $\boldsymbol{e}_v$，$\boldsymbol{e}_p$，将产品 v 和属性 p 在关系 r 空间中的距离表示为 $f_r(v,p)$，则 $f_r(v,p)$ 计算如式(13)所示

$$f_r(v,p) = \| \boldsymbol{e}_v + \boldsymbol{e}_r - \boldsymbol{e}_p \|_2^2 \tag{13}$$

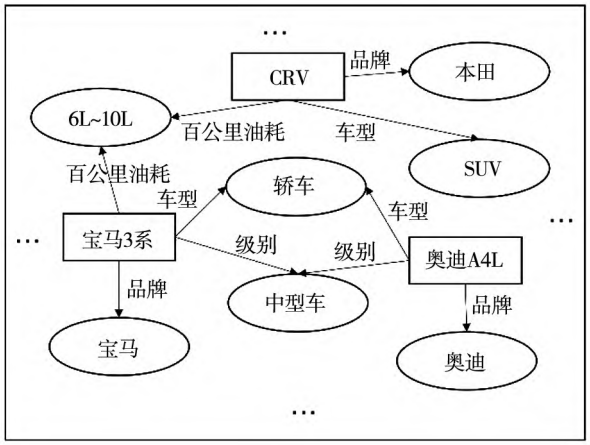

图 2 汽车产品属性知识图谱示例图

Fig. 2 Example of automotive product attribute knowledge graph

下面利用交互数据 D 中用户偏好的属性集 P_u^+ 与用户不偏好的属性集 P_u^- 来代表用户的正向偏好与负向偏好. 基于理想点法[47]在知识图谱表征空间中分别计算候选产品 v 与用户正向偏好(正理想点)的距离 $score_v^+$ 和负向偏好(负理想点)的距离 $score_v^-$，如式(14)和式(15)所示

$$score_v^+ = \frac{1}{|p_u^+|} \sum_{p \in P_u^+} f_r(v,p) \tag{14}$$

$$score_v^- = \frac{1}{|p_u^-|} \sum_{p \in P_u^-} f_r(v,p) \tag{15}$$

由于离用户 u 正向偏好越近，负向偏好越远的产品越能满足用户的偏好，即让推荐的损失函数 L_{rec}^k 更小. 因此，本研究通过式(16)对候选集进行排序，为用户生成推荐列表 rec_{list}.

$$rec_{\text{list}} = \text{rank}\left(\frac{score_v^-}{score_v^+} \right) v \in V_t^{k,u} \tag{16}$$

根据式(14) ~ 式(16)可知，rec_{list} 取决于产品与属性的表征学习. 即一旦给定产品与属性的表征，推荐列表 rec_{list} 就是确定的. 因此，如果产品和属性的表征学习得较好，则推荐列表的质量更高. 因此，本研究将推荐函数的损失 L_{rec}^k 定义为知识图谱节点与关系的表征损失，即

$$L_{\text{rec}}^k = \sum_{\substack{(p,r,v) \in G, \\ (p^-,r^-,v^-) \notin G}} [f_r(v,p) - f_{r^-}(v^-,p^-)] \tag{17}$$

其中 p^-，r^-，v^- 为知识图谱的中三元组的负样本.

本研究提出的面向高卷入度产品的对话推荐算法如算法 1 所示.

算法 1 高卷入度产品对话推荐算法

1. For a user u:
2. For 每一阶段 $k = 1,2,\cdots K$:
3. For 每一轮次 $t = 1,2,\cdots T$:
4. If $t = 1$ and $k = 1$:
5. 用户向表达初始化偏好
6. 通过式(4)计算 $state_1^{u,k}$
7. 更新阶段 $stage = k$
8. 通过式(6)计算 $Q(state_t^{k,u}, a_t^k; \boldsymbol{\theta}^k)$
9. 根据 ε - 贪心策略选择系统动作 a_t^k
10. If $a_t^k = a_{ask}$:
11. 计算式(10)计算选择属性 $p_{t,t}^{k,u}$ 价值
12. 根据 ε - 贪心策略确定提问属性 $p_t^{k,u}$
13. If $p_t^{k,u} \in P_{v_u}$:
14. 用户确认
15. Else:
16. 用户否认
17. Else If $a_t^k = a_{\text{rec}}$:
18. 通过式(16)计算 rec_{list}
19. If 用户喜欢的产品在 rec_{list} 中:
20. 推荐成功,交互结束
21. Else:
22. 推荐失败,进入下一轮交互

3　实验结果

3.1　数据集描述及初始化数据生成

对话推荐系统需要在与用户的交互过程中进行训练和评估. 然而,基于真实的在线交互进行模型训练与评估会带来代价昂贵、可操作性弱且可重复性低等问题[16]. 因此,与现有主流研究一致[36],本研究采用了真实数据与仿真交互相结合的方式进行模型训练与评估.

鉴于汽车产品的市场规模,本研究利用汽车产品数据构建高卷入度产品推荐场景. 首先从国内知名汽车网站上采集市场上现有的 1 100 种车型的配置信息和 19 194 个用户的购买记录信息,其中汽车平均价格为 20.79 万元,属性共有 667 个. 基于汽车的配置信息,构建真实汽车产品的知识图谱.

基于获取汽车产品的信息,仿真生成用户偏好. 与已有研究一致[15,36],本研究假设用户对自己购买产品的属性都是偏好的. 因此,对于用户的真实购买记录 (u,v),本研究将产品 v 作为目标产品,目标产品的属性集合 P_{v_u} 作为在对话中用户偏好的属性. 对于初始化偏好,本研究针对汽车市场中较为流行的车型,基于信息熵的方法,选择了从区分度比较大的属性集合中进行属性采样,从而作为用户的初始偏好. 具体而言,本研究发现"价格"与"车型"的区分度最高,因此从用户目标车型的"价格"和"车型"属性中随机进行随机采样这两个属性的取值(如"价格" = "15 万 ~ 20 万","车型" = "轿车")作为用户初始偏好.

3.2　评价指标

与已有对话推荐系统研究一致[15,16,43],本研究采用的评估指标如下.

1) t 轮次成功率: 在对话推荐系统与用户交互 t 次时的推荐成功率.

2) k 阶段成功率: 在对话推荐系统与用户第 k 阶段交互的推荐成功率.

3) 对话策略损失值: 在对话推荐系统与用户的交互过程中,动作函数的累计损失.

4) 平均交互次数: 对话推荐系统为用户推荐出满意产品而需要与用户交互的平均次数.

其中 t 轮次成功率和 k 阶段成功率越高、累计损失值越低、平均交互次数越小,表明系统性能越好.

3.3　参数设置

根据已有文献[43],在提问决策中,如果提问成功,环境应该反馈较小的损失值; 如果提问失败,环境应该给予较大的损失值; 在推荐决策中,如果推荐失败,环境应该给予较小的损失值; 如果推荐成功,环境应该给予非常小的损失值; 当交互次数达到最大时,环境应该给予非常大的损失值. 因此,本研究 DQN 损失值设置如表 1 所示. 同时,贪心策略 $\varepsilon = 0.9$,折扣率 $\gamma = 0.95$,推荐函数中,产品、属性和关系的嵌入表示维度为 64,推荐列表的长度为 5. 阶段数和每个阶段的交互次数的设置与应用场景有很大关系. 以本研究汽车推荐

为例，用户购买汽车主要关注购车用车成本、车辆性能以及乘坐舒适性三类主题属性，因此，当系统需要向用户探究其更多类型的偏好属性才能获得精准推荐时，系统会自动切换交互主题进入下一阶段，此时需要通过改变损失反馈调整对话策略，即基于交互的主题发生变化来划分会话阶段，因此本研究将最大阶段数 K 设置为三阶段. 在模型预训练阶段，反复试验表明，对于每类主题属性下的用户兴趣偏好，系统最高通过 5 次交互就能获取. 例如，在购车用车成本主题中，系统通过探究价格、油耗、能源类型等属性的偏好水平就能获取用户的成本偏好. 因此，本研究将每个阶段系统与用户交互的最大次数 T 设置为 5 次，最大交互次数 $\max Turn = 3 \times 5 = 15$. 本研究的实验平台硬件环境为 Intel 4215R CPU@ 3.2GHz、128GB RAM 以及 RTX 3090 GPU，软件环境为 Python 3.7 和 Pytorch 1.13. 本研究所提模型在测试时，每次对话决策、属性选择和产品推荐决策的整体用时小于 10 ms.

表 1 DQN 损失值设置

Table 1 DQN loss setting

损失类别	阶段 1	阶段 2	阶段 3
询问成功	-0.01	-0.01	-0.01
询问失败	0.10	0.10	0.10
推荐成功	-1.00	-0.75	-0.50
推荐失败	0.10	0.10	0.10
退出会话	0.30	0.30	0.30

3.4 对比算法

实验 1 为验证模块化设计的优越性，本研究选取了端到端的对比方法：Conversational Recommendation Model (CRM) [39]、Estimation-Action-Reflection (EAR) [15] 和 Conversational Thompson Sampling (CONTS) [16]. CRM 利用神经网络同时对对话策略与产品推荐方法进行优化. EAR 方法使用强化学习同时对对话策略与属性选择策略进行决策. CONTS 方法通过贝叶斯多臂老虎机同时决策对话策略、属性选择策略以及产品推荐方法.

实验 2 为了验证所设计的动作函数（模块一）的优越性，本研究对比算法包括随机选择的方法、基于人工规则的方法 Rule-based [35] 和基于策略梯度的强化学习算法 Policy Gradient (PG) [52] 等目前对话策略的主流方法. 在本研究中，基于人工规则的方法规则_n 表示每提问 $n-1$ 次推荐 1 次. PG 方法需要在系统与用户完成交互后才能根据交互数据更新参数.

实验 3 为了验证所设计的提问函数（模块二）的优越性，本研究对比算法有基于策略梯度的强化学习算法 PG [52]、基于值学习的强化学习算法 DQN [48]、基于最大熵的方法 MaxEnt [40] 以及基于流行度的方法 Popular based (POP) [53] 等主流属性选择方法.

实验 4 为了验证所设计的推荐函数（模块三）的优越性，本研究对比算法包括基于流行度的方法 [53]、因子分解机算法 Factorization Machine (FM) [43]、TransR [54]、TransD [55]、TransH [50] 以及 Knowledge Graph Attention Network (KGAT) [51] 等主流产品推荐方法.

实验 5 为了验证所提多阶段方法的优越性，本研究将其与单阶段方法进行了对比. 为了验证差异是由多阶段设计引起的，单阶段方法也采用所提的模块化方法. 单阶段方法重复用于每一阶段，每一阶段只利用本阶段信息和本阶段的损失函数.

实验 6 敏感性分析，由于推荐失败损失值越高，表明数据获取损失相对产品推荐损失的重要性越低，因此本实验测试了推荐失败损失的不同取值对实验结果的影响，反映数据获取损失相对推荐失败损失的重要性对实验结果的影响. 同时，由于贪心策略参数 ε 会影响对话动作的选择，因此本实验分析 ε 的不同取值对实验结果的影响. 此外，本研究对强化学习三个阶段损失反馈和累计损失中的折扣率均进行了敏感性分析.

实验 7 案例分析，通过具体案例展示所提面向高卷入度产品的模块化多阶段对话推荐过程②.

3.5 实验 1 与端到端方法的对比实验

如表 2 所示，相比于端到端的方法，本研究提出的模块化方法在阶段 1 成功率、阶段 2 成功率、

② 由于版面限制，案例分析结果可联系作者备索.

阶段 3 成功率以及 10 轮次成功率,15 轮次成功率上至少分别提升了 7.64%、14.09%、11.13%、10.16%、10.33%. 同时,在对话策略损失值上至少降低 11.68%,在平均交互次数上至少降低 5.01%. 因此,所提方法的各项评估指标均显著优于对比方法. 由于端到端的方法需要同时对对话策略、属性选择和产品推荐进行优化,各模块间的矛盾性使得优化难度显著增高,在目前可行技术背景下,难以同时保证各模块的优化效果. 本研究通过模块化的设计有效降低了整体任务的难度,能够利用较少的对话交互实现较高的推荐准确率,有效缓解了偏好信息获取任务和产品推荐任务之间的矛盾性问题,最终取得更好的整体效果.

3.6　实验 2　对话策略方法的对比实验

实验结果如表 3 所示. 基于值学习的 DQN 方法在阶段 1 成功率、阶段 2 成功率和阶段 3 成功率分别为 0.295 8、0.249 4 和 0.161 8,而第 10 次交互和第 15 次交互的推荐成功率分别为 0.495 4 和 0.599 0,对话策略损失为 0.616 2,平均交互次数为 9.774 次,均显著优于对比方法. 这说明虽然其他对话策略也可以通过模块化设计缓解矛盾性问题,但是本研究所设计基于 DQN 的方法能够更有效地处理偏好信息获取与用户、产品表征学习之间的关系,降低数据获取与产品推荐之间的矛盾性,因此取得了更佳的效果.

3.7　实验 3　属性选择方法的对比实验

如表 4 所示,基于强化学习的属性选择方法在各项评估指标上均显著优于对比方法. 在推荐准确率指标阶段 2 成功率和阶段 3 成功率上分别提升 1.92% 和 1.38%,在 10 轮次成功率和 15 轮次成功率上分别提升了 1.23% 和 1.27%. 同时本研究发现,基于流行度的属性选择方法表现最差,并且,与通过选择信息熵最大的属性方法相比,本研究提出的融合用户个性化属性偏好在交互成功率指标阶段 1 成功率、阶段 2 成功率、阶段 3 成功率和 10 轮次成功率、15 轮次成功率以及平均交互次数指标优于最大熵方法,这进一步表明了针对不同用户进行个性化属性选择的重要性. 以上实验结果表明本研究所采取的强化学习方法在融合用户个性化属性偏好和候选属性信息进行序列决策的有效性.

表 2　与端到端方法实验结果对比

Table 2 Comparison of experimental results of end-to-end methods

模型	阶段 1 成功率	阶段 2 成功率	阶段 3 成功率	10 轮次成功率	15 轮次成功率	对话策略损失	平均交互次数
EAR	0.219 0 (35.07%)	0.175 5 (42.11%)	0.109 9 (47.22%)	0.359 4 (37.84%)	0.429 7 (39.40%)	0.857 0 (-28.10%)	11.257 8 (-13.18%)
CRM	0.230 4 (28.39%)	0.166 9 (49.43%)	0.122 0 (32.62%)	0.364 0 (36.10%)	0.442 0 (35.52%)	0.845 6 (-27.13%)	11.160 4 (-12.42%)
CONTS	0.274 8 (7.64%)	0.218 6 (14.09%)	0.145 6 (11.13%)	0.449 7 (10.16%)	0.542 9 (10.33%)	0.697 7 (-11.68%)	10.289 3 (-5.01%)
OURS	0.295 8	0.249 4	0.161 8	0.495 4	0.599 0	0.616 2	9.774 0

表 3　对话策略实验结果对比

Table 3 Comparison of experimental results of dialog strategies

模型	阶段 1 成功率	阶段 2 成功率	阶段 3 成功率	10 轮次成功率	15 轮次成功率	对话策略损失	平均交互次数
随机动作	0.129 4 (128.59%)	0.109 6 (127.55%)	0.087 1 (85.76%)	0.217 1 (128.19%)	0.272 9 (119.49%)	1.083 0 (-43.10%)	12.755 5 (-23.37%)
规则_3	0.212 4 (39.27%)	0.165 0 (51.15%)	0.110 1 (46.96%)	0.344 4 (43.84%)	0.414 8 (44.41%)	0.852 9 (-27.75%)	11.456 9 (-14.69%)
规则_5	0.188 5 (56.92%)	0.167 2 (49.16%)	0.116 7 (38.64%)	0.322 3 (53.70%)	0.396 9 (50.92%)	0.893 4 (-31.03%)	11.865 8 (-17.63%)
策略梯度	0.281 7 (5.05%)	0.240 6 (3.66%)	0.130 0 (24.46%)	0.448 4 (10.48%)	0.531 7 (12.66%)	0.680 5 (-9.45%)	10.041 9 (-2.67%)
OURS	0.295 8	0.249 4	0.161 8	0.495 4	0.599 0	0.616 2	9.774 0

3.8 实验 4 产品推荐方法的对比实验

如表 5 所示，相较对比方法，本研究提出的基于知识图谱和理想点的推荐方法在阶段 1 成功率、阶段 2 成功率和阶段 3 成功率相较对比方法至少提升 6.94%、2.05% 和 10.52%. 同时，在对话策略损失值上降低了 6.81%，在平均交互次数上降低了 3.53%. 同时，本研究发现只考虑候选产品与用户正向偏好距离的方法 TransH 的各评估指标均较低，这说明本研究引入理想点在产品推荐中具有有效性. 本研究也发现 TransD、TranR 相比于 TransH 效果稍差，这也验证了本研究选择 TransH 的科学性. FM、KGAT 等方法由于需要使用大量的交互数据进行训练，因此在有限交互数据条件下也难以取得较高的推荐准确率. 综上所述，在面向高卷入度产品的对话推荐系统中，基于知识图谱和理想点的推荐方法可以有效地利用产品自身的属性信息和交互数据为用户推荐满意的产品.

表 4 属性选择方法实验结果对比

Table 4 Comparison of experimental results of attribute selection methods

模型	阶段 1 成功率	阶段 2 成功率	阶段 3 成功率	10 轮次成功率	15 轮次成功率	对话策略损失	平均交互次数
最大熵	0.246 9 (19.81%)	0.189 1 (31.89%)	0.143 8 (12.52%)	0.398 2 (24.41%)	0.490 2 (22.20%)	0.778 8 (−20.88%)	10.709 8 (−8.74%)
流行度	0.210 3 (40.66%)	0.149 8 (66.49%)	0.112 3 (44.08%)	0.330 2 (50.03%)	0.402 0 (49.00%)	0.883 1 (−30.22%)	11.395 1 (−14.23%)
策略梯度	0.293 6 (0.75%)	0.244 7 (1.92%)	0.159 6 (1.38%)	0.489 4 (1.23%)	0.591 5 (1.27%)	0.621 1 (−0.79%)	9.816 4 (−0.43%)
OURS	0.295 8	0.249 4	0.161 8	0.495 4	0.599 0	0.616 2	9.774 0

表 5 产品推荐方法实验结果对比

Table 5 Comparison of experimental results of recommendation methods

模型	阶段 1 成功率	阶段 2 成功率	阶段 3 成功率	10 轮次成功率	15 轮次成功率	对话策略损失	平均交互次数
流行度	0.214 9 (37.65%)	0.182 1 (36.96%)	0.126 4 (28.01%)	0.360 6 (37.38%)	0.441 5 (35.67%)	0.845 1 (−27.09%)	11.252 6 (−13.14%)
FM	0.258 6 (14.39%)	0.232 1 (7.45%)	0.146 4 (10.52%)	0.444 2 (11.53%)	0.537 9 (11.36%)	0.717 1 (−14.07%)	10.405 4 (−6.07%)
KGAT	0.276 0 (7.17%)	0.244 4 (2.05%)	0.141 7 (14.18%)	0.471 5 (5.07%)	0.562 2 (6.57%)	0.661 2 (−6.81%)	10.131 2 (−3.53%)
TransD	0.273 3 (8.23%)	0.238 2 (4.70%)	0.139 6 (15.90%)	0.463 9 (6.78%)	0.553 3 (8.26%)	0.708 1 (−12.98%)	10.198 5 (−4.16%)
TransR	0.271 6 (8.91%)	0.238 3 (4.66%)	0.139 5 (15.99%)	0.462 3 (7.16%)	0.551 6 (8.59%)	0.704 9 (−12.58%)	10.217 5 (−4.34%)
TransH	0.276 6 (6.94%)	0.240 0 (3.92%)	0.140 8 (14.91%)	0.468 6 (5.72%)	0.558 7 (7.21%)	0.687 2 (−10.33%)	10.155 0 (−3.75%)
OURS	0.295 8	0.249 4	0.161 8	0.495 4	0.599 0	0.616 2	9.774 0

3.9 实验 5 与单阶段对话推荐对比实验

多阶段对话推荐与单阶段对话推荐的比较结果如图 3 所示，所提多阶段模型在阶段 1 成功率、阶段 2 成功率、阶段 3 成功率、对话策略损失值和平均交互次数上均优于单阶段模型. 同时，本研究发现，阶段 3 成功率小于阶段 2 成功率，而阶段 2 成功率小于阶段 1 成功率，这是由于在用户与对话推荐系统的不同交互阶段中，进入下一阶段用户的均未被上一阶段成功推荐，对这些用户的推荐难度更大，因此命中率呈现递减趋势. 本研究所提多阶段模型通过引入阶段状态变量，调整不同阶段损失反馈函数，减缓了递减的趋势. 同时相比单阶段模型，所提模型在不同阶段的推荐效果上均有显著的提升，这也验证了多阶段建模的有效性.

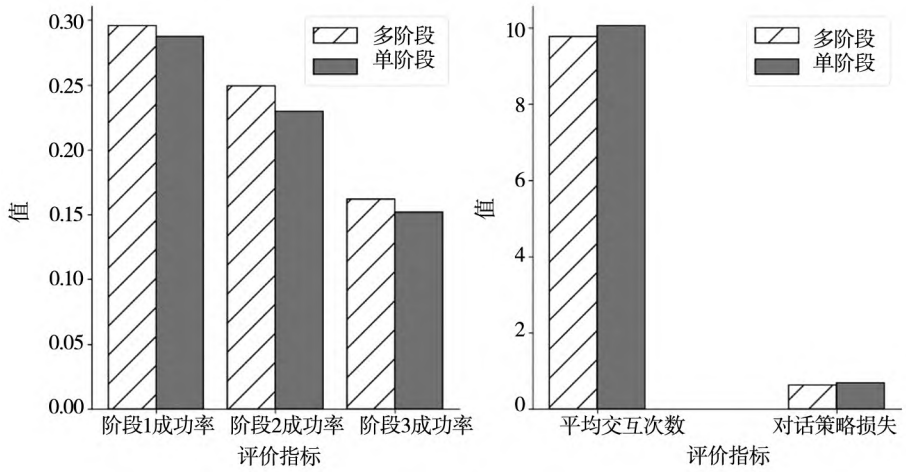

图 3　与单阶段模型实验结果对比

Fig. 3 Comparison of experimental results of single stage model

3.10　实验 6　敏感性分析

推荐失败损失值的敏感性分析实验结果如图 4 所示. 由于交互过程中最多只会出现一次推荐成功的交互,故推荐次数的增加表明推荐精度的损失增加. 同时,提问次数增加说明数据获取的损失增加. 根据图 4 可以发现,随着推荐失败损失值增加,系统倾向进行提问,数据获取损失增加; 随着推荐失败损失值降低,系统倾向进行推荐,推荐损失增加. 当取值为 0.1 时,推荐损失和数据获取损失的矛盾性取得了较好的均衡,达到最佳实验效果.

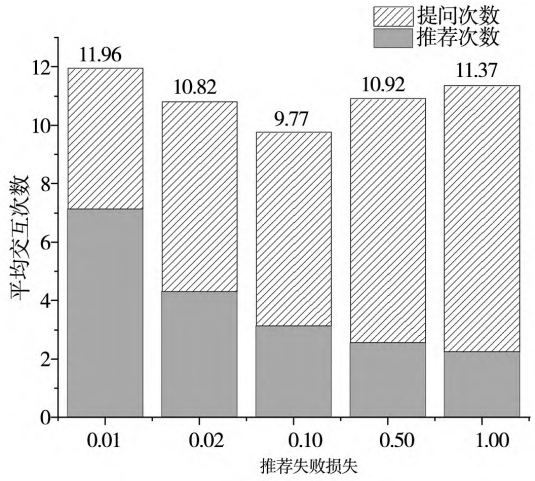

图 4　推荐失败损失值的敏感性分析

Fig. 4 Sensitivity analysis of the loss on failure of recommendation

贪心策略参数 ε 的敏感性分析实验结果如图 5 所示. 当 ε 较小时,动作函数偏向于进行探索,随机选择动作 a_t^k ,导致推荐准确率较低; 当 ε = 0.9 时,动作函数取得了较好的探索与利用的平衡,从而取得最优效果; 当 ε 较大时,动作函数偏向于利用已学到的信息而缺乏探索动力,导致推荐准确率下降.

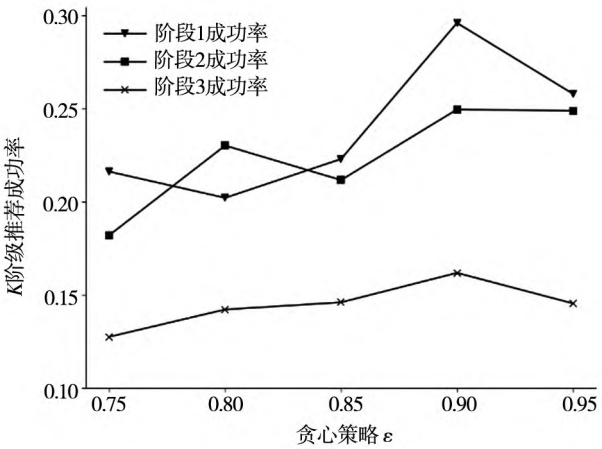

图 5　贪心策略 ε 值的敏感性分析

Fig. 5 Sensitivity analysis of the ε of the greedy

动作函数三个阶段损失反馈的敏感性分析实验结果如图 6 所示,对于所有阶段,推荐失败损失与推荐成功损失的取值对实验结果的影响较大. 并且每个阶段的损失反馈设置对该阶段的命中率影响较大,验证了本研究通过多阶段建模实现对不同阶段设置不同损失反馈的有效性.

累计损失中的折扣率的敏感性分析实验结果如图 7 所示. 由于折扣率表示未来损失的重要性. 未来损失重要性较小时,模型未能充分考虑未来

损失的影响，因此在阶段 1 成功率，阶段 2 成功率和阶段 3 成功率指标上表现较差；当折扣率 $\gamma = 1$ 时，模型认为未来损失与当前损失同等重要，导致模型过度考虑了未来损失. 当折扣率 $\gamma = 0.95$ 时，模型在当前损失和未来损失的重要性上取得较好的权衡.

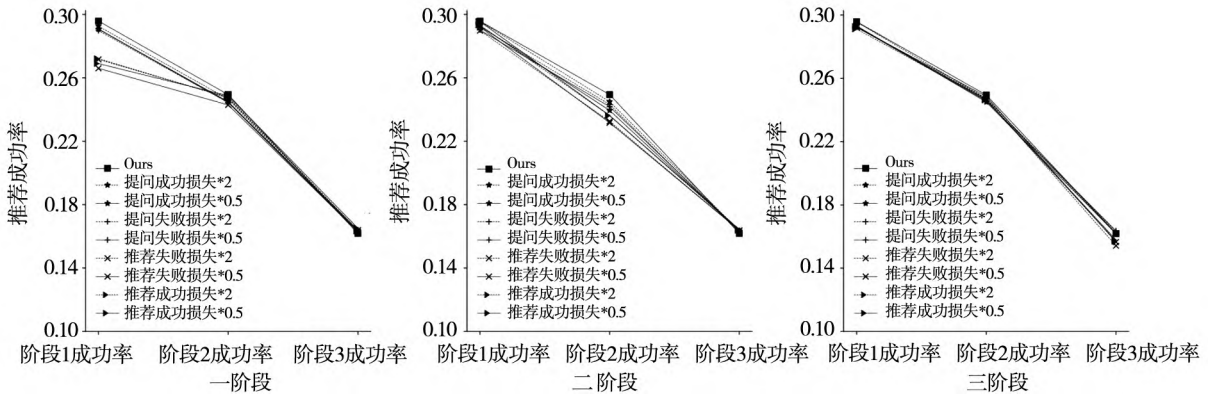

图 6　强化学习损失的敏感性分析

Fig. 6 Sensitivity analysis of reinforcement learning loss

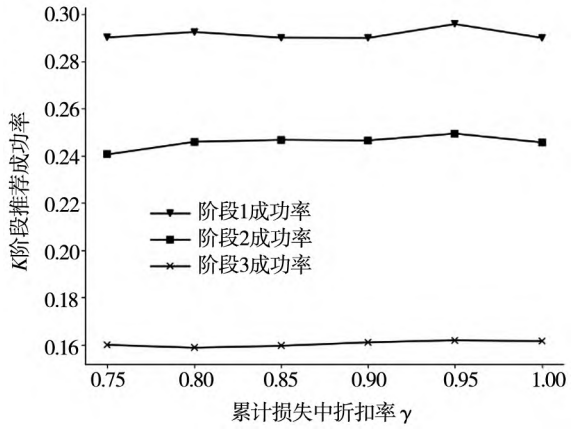

图 7　累计损失中 γ 的敏感性分析

Fig. 7 Sensitivity analysis of γ of total loss

4　结束语

本研究研究了面向高卷入度产品的对话推荐问题，围绕高卷入度产品功能属性复杂、价值高昂和消费者咨询具有多阶段性的特点，分析了对话推荐系统任务的矛盾性难题以及咨询多阶段性的挑战，进而设计了面向高卷入度产品的模块化多阶段对话推荐方法，包括通过引入阶段状态建立多阶段强化学习模型，以及通过将端到端问题进行分解并建立基于强化学习的模块化对话推荐系统. 在模块化推荐系统中，本研究设计了基于强化学习的对话策略，决策提问或推荐动作的选择；构建了基于强化学习的属性选择方法，对属性选择进行优化；构建了基于知识图谱和理想点的产品推荐方法，对产品推荐进行优化. 实验表明所提方法在推荐准确率、交互次数以及对话策略损失等指标上均显著优于对比方法.

本研究的管理意义包括以下三点：首先，对于高卷入度产品的消费者，本研究所提的对话推荐系统能够更准确地识别消费者偏好，同时尽量避免因无效对话而引发的消费者厌倦问题，提高消费者购物体验和满意度；其次，对于高卷入度产品的销售者，一方面本研究所提的对话推荐系统能够通过提高推荐准确率，帮助销售者提升利润；另一方面多阶段式的设计也能够辅助销售者针对不同阶段的消费者采取不同的营销策略；最后，对于对话推荐的研究者，本研究提出的面向高卷入度产品的基于多阶段强化学习的模块化对话推荐系统，为对话推荐提供了新的思路，丰富了对话推荐的理论与方法体系.

本研究不足以及未来研究方向包括以下四点：1）本研究根据已有对话推荐方法文献仿真对话交互的过程，并假设消费者对所购买的车型的所有功能属性都是偏好的，在一定程度上与现实情况存在差异，例如消费者可能会需要在不同的功能属性上进行取舍. 未来可以基于更加现实的数据进行模型构建和评估；2）本研究通过提问用户关于高卷入度产品的功能属性来确定用户的偏

好,在实际情况中,消费者可能关注高卷入度产品的情感价值和社会价值,例如通过购买豪华汽车来彰显社会身份. 因此未来研究可以将高卷入度产品的情感价值和社会价值考虑在内,构建更加全面的推荐系统; 3) 由于缺乏真实的用户交互数据,本研究所提出的方法设置每一个阶段的交互次数为固定值. 在实际情况中,不同消费者在不同阶段的交互次数可能是不同的. 未来可以针对不同的消费者设立不同的交互次数,从而更好地反映不同消费者的行为差异; 4) 本研究只针对汽车推荐情景进行验证,未来研究可以进一步验证本研究所提方法在冰箱、洗衣机、空调等其他高卷入度产品推荐上的推荐效果.

参 考 文 献:

[1] Gu B, Park J, Konana P. Research note: The impact of external word-of-mouth sources on retailer sales of high-involvement products [J]. Information Systems Research, 2012, 23(1): 182 - 196.

[2] 李　创, 叶露露, 王丽萍. 新能源汽车消费促进政策对潜在消费者购买意愿的影响 [J]. 中国管理科学, 2021, 29(10): 151 - 164.
Li Chuang, Ye Lulu, Wang Liping. The influence of new energy vehicle consumption promotion policy on the purchase intention of potential consumers [J]. Chinese Journal of Management Science, 2021, 29(10): 151 - 164. (in Chinese)

[3] Dong X, Yu L, Wu Z, et al. A hybrid collaborative filtering model with deep structure for recommender systems [C]. Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, 2017: 1309 - 1315.

[4] Qian Y, Jiang Y, Shang J, et al. Why some products compete and others don't: A competitive attribution model from customer perspective [J]. Decision Support Systems, 2023, 169: 113956.

[5] 方晓丹. 全国居民收入比 2010 年增加一倍 居民消费支出稳步恢复 [EB/OL]. https://www.ndrc.gov.cn/fggz/jyysr/jysrsbxf/202101/t20210126_1265751.html, 2021.
Fang Xiaodan. Personal income doubled over 2010 Consumer spending recovered steadily [EB/OL]. https://www.ndrc.gov.cn/fggz/jyysr/jysrsbxf/202101/t20210126_1265751.html, 2021. (in Chinese)

[6] Jiang C, Duan R, Jain H K, et al. Hybrid collaborative filtering for high-involvement products: A solution to opinion sparsity and dynamics [J]. Decision Support Systems, 2015, 79: 195 - 208.

[7] 刘业政, 吴　锋, 孙见山, 等. 基于群偏好与用户偏好协同演化的群推荐方法 [J]. 系统工程理论与实践, 2021, 41(3): 537 - 553.
Liu Yezheng, Wu Feng, Sun Jianshan, et al. Group recommendation method based on co-evolution of group preference and user preference [J]. System Engineering: Theory & Practice, 2021, 41(3): 537 - 553. (in Chinese)

[8] 朱国玮, 周　利. 基于遗忘函数和领域最近邻的混合推荐研究 [J]. 管理科学学报, 2012, 15(5): 55 - 64.
Zhu Guowei, Zhou Li. Hybrid recommendation based on forgetting curve and domain nearest neighbor [J]. Journal of Management Sciences in China, 2012, 15(5): 55 - 64. (in Chinese)

[9] Zhou B, Zou T. Competing for recommendations: The strategic impact of personalized product recommendations in online marketplaces [J]. Marketing Science, 2023, 42(2): 360 - 376.

[10] Zhang S, Yao L, Sun A, et al. Deep learning based recommender system: A survey and new perspectives [J]. ACM Computing Surveys (CSUR), 2019, 52(1): 1 - 38.

[11] Chen S, Qiu H, Zhao S, et al. When more is less: The other side of artificial intelligence recommendation [J]. Journal of Management Science and Engineering, 2022, 7(2): 213 - 232.

[12] He X, He Z, Song J, et al. Nais: Neural attentive item similarity model for recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(12): 2354 - 2366.

[13] Zhang Y, Chen X, Ai Q, et al. Towards conversational search and recommendation: System ask, user respond [C]. Pro-

ceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, 2018: 177 – 186.

[14] Jannach D, Manzoor A, Cai W, et al. A survey on conversational recommender systems [J]. ACM Computing Surveys (CSUR), 2021, 54(5): 1 – 36.

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

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

[17] 鞠晴江, 鞠 鹏, 代文强, 等. 新能源汽车补贴政策与保有量影响研究: 单位补贴、销售奖励与产品差异化 [J]. 管理科学学报, 2021, 24(6): 101 – 116.
Ju Qingjiang, Ju Peng, Dai Wenqiang, et al. Adoption of new energy vehicles under subsidy policies: Unit subsidies, sales incentives and product differentiation [J]. Journal of Management Sciences in China, 2021, 24(6): 101 – 116. (in Chinese)

[18] Santandreu J, Shurden M C. Purchase decisions for high involvement products: The new generation of buyers [J]. Journal of Marketing Development and Competitiveness, 2017, 11(2): 88 – 92.

[19] Lovett M J, Peres R, Shachar R. On brands and word of mouth [J]. Journal of Marketing Research, 2013, 50(4): 427 – 444.

[20] Nayeem T, Casidy R. The role of external influences in high involvement purchase behaviour [J]. Marketing Intelligence & Planning, 2013, 31(7): 732 – 745.

[21] Peng L, Zhang W, Wang X, et al. Moderating effects of time pressure on the relationship between perceived value and purchase intention in social e-commerce sales promotion: Considering the impact of product involvement [J]. Information & Management, 2019, 56(2): 317 – 328.

[22] Andrews R L, Currim I S. Multi-stage purchase decision models: Accommodating response heterogeneity, common demand shocks, and endogeneity using disaggregate data [J]. International Journal of Research in Marketing, 2009, 26(3): 197 – 206.

[23] Huber G P. Multi-attribute utility models: A review of field and field-like studies [J]. Management Science, 1974, 20(10): 1393 – 1402.

[24] Huang S l. Designing utility-based recommender systems for e-commerce: Evaluation of preference-elicitation methods [J]. Electronic Commerce Research and Applications, 2011, 10(4): 398 – 407.

[25] Dinsdale A, Williamann P. The future of auto retailing [EB/OL]. https://www2.deloitte.com/us/en/insights/focus/future-of-mobility/automotive-retail-industry-mobility-ecosystems.html, 2016.

[26] Burke R D, Hammond K J, Yound B. The FindMe approach to assisted browsing [J]. IEEE Expert, 1997, 12(4): 32 – 40.

[27] Pu P, Chen L. Integrating tradeoff support in product search tools for e-commerce sites [C]. Proceedings of the 6th ACM Conference on Electronic Commerce, Vancouver, 2005: 269 – 278.

[28] Reilly J, Mccarthy K, Mcginty L, et al. Incremental critiquing [J]. Knowledge-Based Systems, 2005, 4(18): 143 – 151.

[29] Chen L, Wang F. Preference-based clustering reviews for augmenting e-commerce recommendation [J]. Knowledge-Based Systems, 2013, 50: 44 – 59.

[30] Capdevila J, Arias M, Arratia A. Geosrs: A hybrid social recommender system for geolocated data [J]. Information Systems, 2016, 57: 111 – 128.

[31] Li R, Kahou S, Schulz H, et al. Towards deep conversational recommendations [C]. Proceedings of the 32nd International Conference on Neural Information Processing Systems, Red Hook, 2018: 9748 – 9758.

[32] Qiu M, Li F L, Wang S, et al. Alime chat: A sequence to sequence and rerank based chatbot engine [C]. Proceedings of

the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, 2017: 498 – 503.

[33] Lu Y, Bao J, Song Y, et al. RevCore: Review-augmented conversational recommendation [C]. Meeting of the Association for Computational Linguistics, 2021: 1161 – 1173.

[34] Liao L, Takanobu R, Ma Y, et al. Topic-guided relational conversational recommender in multi-domain [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(5): 2485 – 2496.

[35] Zhang X, Xie H, Li H, et al. Conversational contextual bandit: Algorithm and application [C]. Proceedings of the Web Conference 2020, Taipei, 2020: 662 – 672.

[36] Deng Y, Li Y, Sun F, et al. Unified conversational recommendation policy learning via graph-based reinforcement learning [C]. The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Houston, 2021: 304 – 213.

[37] Greco C, Suglia A, Basile P, et al. Converse-et-impera: Exploiting deep learning and hierarchical reinforcement learning for conversational recommender systems [C]. Conference of the Italian Association for Artificial Intelligence, Bari, 2017: 372 – 386.

[38] **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，**2021, 33(12): 4616 – 4643.
Zhao Mengyuan, Huang Xiaowen, Sang Jitao, et al. A survey on conversational recommender system [J]. Journal of Software, 2021, 33(12): 4616 – 4643. (in Chinese)

[39] Sun Y, Zhang Y. Conversational recommender system [C]. The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, 2018: 235 – 244.

[40] Dhingra B, Li L, Li X, et al. Towards end-to-end reinforcement learning of dialogue agents for information access [C]. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, 2017: 484 – 495.

[41] Xia M, Sun M, Wei H, et al. Peerlens: Peer-inspired interactive learning path planning in online question pool [C]. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow Scotland, 2019: 1 – 12.

[42] Shi D, Wang T, Xing H, et al. A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning [J]. Knowledge-Based Systems, 2020, 195: 105618.

[43] Lei W, Zhang G, He X, et al. Interactive path reasoning on graph for conversational recommendation [C]. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, 2020: 2073 – 2083.

[44] Wei Y, Wang X, Li Q, et al. Contrastive learning for cold-start recommendation [C]. Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event, 2021: 5382 – 5390.

[45] Deng Y, Chen H, Shao S, et al. Multi-objective vehicle rebalancing for ridehailing system using a reinforcement learning approach [J]. Journal of Management Science and Engineering, 2022, 7(2): 346 – 364.

[46] **刘冠男，曲金铭，李小琳，等. 基于深度强化学习的救护车动态重定位调度研究 [J]. 管理科学学报，**2020, 23(2): 39 – 53.
Liu Guannan, Qu Jinming, Li Xiaolin, et al. Dynamic ambulance redeployment based on deep reinforcement learning [J]. Journal of Management Sciences in China, 2020, 23(2): 39 – 53. (in Chinese)

[47] Tzeng G H, Huang J J. Multiple Attribute Decision Making: Methods and Applications [M]. Boca Raton: CRC Press, 2011.

[48] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518(7540): 529 – 533.

[49] Hasselt H. Double Q-learning [C]. Advances in Neural Information Processing Systems, Vancouver, 2010: 2613 – 2621.

[50] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes [C]. Proceedings of the AAAI Conference on Artificial Intelligence, Québec, 2014: 1112 – 1119.

[51] Wang X, He X, Cao Y, et al. KGAT: Knowledge graph attention network for recommendation [C]. The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, 2019: 950 – 958.

[52] Sutton R S, McAllester D A, Singh S P, et al. Policy gradient methods for reinforcement learning with function approximation [C]. Advances in Neural Information Processing Systems, Denver, 2000: 1057 – 1063.

[53] Prawesh S, Padmanabhan B. The "most popular news" recommender: Count amplification and manipulation resistance [J]. Information Systems Research, 2014, 25(3): 569 – 589.

[54] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion [C]. Twenty-ninth AAAI Conference on Artificial Intelligence, Austin, 2015: 2181 – 2187.

[55] Ji G, He S, Xu L, et al. Knowledge graph embedding via dynamic mapping matrix [C]. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, 2015: 687 – 696.

Conversational recommendation system for high-involvement products: A modular multi-stage approach

CHAI Yi-dong[1,2], *ZHOU Yong-hang*[1,2], *JIANG Yuan-chun*[1,2*], *LIU Chun-li*[1,2], *YUAN Kun*[1,2], *LIU Ye-zheng*[1,2]

1. School of Management, Hefei University of Technology, Hefei 230009, China;
2. Key Laboratory of Philosophy and Social Sciences for Cyberspace Behavior and Management, Hefei 230009, China

Abstract: With the increasing enrichment of high-involvement products, such as automobiles and home appliances, designing a recommendation system to assist consumers in choosing high-involvement products has become an important research issue. Focusing on the attributes of high-volume and high value products, as well as multistage of consumer consulting, this paper proposes a modular multistage conversational recommendation method for high-involvement products. The proposed method adopts the paradigm of "system query-user answer" to obtain user preferences through questions and generate recommendation results based on user answers. For the issue of multistage consulting, the proposed method introduces the state variable of the stage to the reinforcement learning algorithm. To conquer the contradiction problem between the tasks of preference acquisition and product recommendation, this paper constructs a modular conversational recommendation system. The system includes three components: A dialogue strategy based on reinforcement learning, an attribute selection method based on reinforcement learning, and a product selection method based on knowledge graph and ideal point method. Experiments based on a real purchase dataset on a well-known Chinese auto forum and on simulated user interaction data indicate that, compared with the benchmark method, the proposed method can achieve higher recommendation accuracy with fewer interactions.

Key words: high-involvement products; modular conversational recommendation; multistage reinforcement learning; knowledge graph; ideal point method