

中图分类号：TP391.3

论文编号：10006SY2021112

北京航空航天大学
硕士学位论文

基于知识图谱的多任务推荐方法
研究

作者姓名	柳啸峰
学科专业	软件工程
指导教师	林广艳 副教授
培养学院	软件学院

Research of Multi-Task Recommendation Algorithm based on Knowledge Graph

A Dissertation Submitted for the Degree of Master

Candidate: Liu Xiaofeng

Supervisor: Prof. Lin Guangyan

School of Software

Beihang University, Beijing, China

中图分类号：TP391.3

论文编号：10006SY2021112

硕 士 学 位 论 文

基于知识图谱的多任务推荐方法研究

作者姓名	柳啸峰	申请学位级别	工学硕士
指导教师姓名	林广艳	职 称	副教授
学科专业	软件工程	研究方向	软件工程
学习时间自	2020 年 09 月 01 日	起至	2023 年 05 月 18 日止
论文提交日期	2023 年 04 月 23 日	论文答辩日期	2023 年 05 月 18 日
学位授予单位	北京航空航天大学	学位授予日期	年 月 日

关于学位论文的独创性声明

本人郑重声明：所呈交的论文是本人在指导教师指导下独立进行研究工作所取得的成果，论文中有关资料和数据是实事求是的。尽所知，除文中已经加以标注和致谢外，本论文不包含其他人已经发表或撰写过的研究成果，也不包含本人或他人为获得北京航空航天大学或其它教育机构的学位或学历证书而使用过的材料。与我一同工作的同志对研究所做的任何贡献均已在论文中做出了明确的说明。

若有不实之处，本人愿意承担相关法律责任。

学位论文作者签名：_____

日期：____年____月____日

学位论文使用授权书

本人完全同意北京航空航天大学有权使用本学位论文（包括但不限于其印刷版和电子版），使用方式包括但不限于：保留学位论文，按规定向国家有关部门（机构）送交学位论文，以学术交流为目的赠送和交换学位论文，允许学位论文被查阅、借阅和复印，将学位论文的全部或部分内容编入有关数据库进行检索，采用影印、缩印或其他复制手段保存学位论文。

保密学位论文在解密后的使用授权同上。

学位论文作者签名：_____

日期：____年____月____日

指导教师签名：_____

日期：____年____月____日

摘 要

近年来,随着图谱构建技术的发展,知识图谱的应用日益广泛。推荐系统常引入知识图谱作为辅助信息来丰富物品的特征,挖掘用户的兴趣,解决传统推荐算法的冷启动和数据稀疏性问题。这类推荐算法的效果与图谱的完备性息息相关,为此,本文研究基于知识图谱的多任务推荐方法,提出一种物品推荐与图谱补全联合网络(MRKC),主要工作内容如下:

(1)针对现有方法中物品与节点建模同质化、任务间参数交叉学习能力有限和图谱完备性欠缺的问题,本文提出基于高阶共享单元的多任务学习框架,异质化建模物品与节点,并设计基于异阶特征交叉的参数共享单元,联合训练推荐任务和图谱补全任务。

(2)针对现有方法中推荐任务存在的忽视用户物品交互中的协同信息对图谱补全的正向作用、忽略多阶交互中的协同信息差异的问题,本文提出基于图卷积神经网络的协同信息感知方法,利用图卷积神经网络挖掘用户物品交互过程中的协同信息,同时引入门控单元自适应建模异阶信息的通入量。

(3)针对现有方法中图谱补全任务存在的弱化关系建模、忽略长距离信息传播衰减现象的问题,本文采用多层注意力机制分别建模同阶邻域不同关系下尾节点的重要性差异和异阶邻域信息长距离传播的衰减。

本文在书籍、电影和音乐三个通用数据集上验证了方法的性能,结果表明MRKC在AUC指标上分别相对提升1.1%、0.7%和1.0%;同时本文在自建的论文数据集上验证了方法对论文推荐效果的提升。本文还进行了图谱补全实验和消融实验,并对超参数进行实验分析,实验结果验证了方法的有效性。

关键词: 推荐算法, 知识图谱, 多任务学习, 图卷积神经网络, 图谱补全

Abstract

In recent years, with the development of graph construction technology, the application of knowledge graphs has become increasingly widespread. Recommendation systems often introduce knowledge graphs as auxiliary information to enrich the features of items, explore users' interests, and effectively solve the cold start and data sparsity problems of traditional recommendation algorithms. The recommendation effect is closely related to the completeness of the graph. For this reason, this paper studies multi-task recommendation methods based on knowledge graphs, and proposes a joint network of recommendation and graph completion (MRKC) based on knowledge graphs, with the following main work:

(1) In response to the problems of homogenization of item and node modeling, limited cross-learning ability between tasks, and lack of graph completeness in existing methods, this paper proposes a multi-task learning framework based on higher-order shared units. The framework heterogeneously models items and nodes and designs shared units based on heterogeneous feature cross. It jointly trains recommendation tasks and knowledge graph completion tasks.

(2) In response to the shortcomings of existing methods in the inability to effectively perceive collaborative information in user-item interactions and the neglect of collaborative information differences in multi-order interactions, this paper proposes a collaborative information-aware method based on graph convolutional neural networks. It fully mines the collaborative information in the user-item interaction process using graph convolutional neural networks and introduces gate units to adaptively model the input volume of different-order information.

(3) In response to the shortcomings of existing methods in weakened relationship modeling and neglect of the decay phenomenon of long-distance information propagation in graph completion tasks, this paper uses a multi-layer attention mechanism to respectively model the importance differences of tail nodes under different relationships in the same-order neighborhood and the decay of information propagation in the different-order neighborhood.

This paper validates the performance of the proposed MRKC algorithm on three common datasets of books, movies, and music. The results show that the MRKC algorithm improves the AUC metric by 1.1%, 0.7%, and 1.0%, respectively. The algorithm's effectiveness on paper recommendation is also verified on two self-built paper datasets. The paper also conducts experiments on graph representation learning and ablation experiments, and experimental analysis of hyperparameters. The experimental results confirm the effectiveness of the algorithm.

Keywords: Recommendation Algorithm, Knowledge Graph, Multi-task Learning, Graph Convolutional Neural Network, Graph Completion

目 录

第一章 绪论	1
1.1 论文背景和意义	1
1.2 国内外研究现状	2
1.2.1 传统推荐算法	2
1.2.2 基于知识图谱的推荐算法	3
1.3 论文研究目标及内容	5
1.3.1 研究目标	5
1.3.2 研究路线	6
1.3.3 研究内容	6
1.4 论文组织与安排	7
第二章 相关理论基础	9
2.1 多任务学习	9
2.1.1 参数共享机制	9
2.1.2 多任务推荐方法	10
2.2 图神经网络	12
2.3 图表示学习	13
2.3.1 图表示学习方法	13
2.3.2 基于表示学习的知识图谱补全	15
2.4 注意力机制	16
2.4.1 注意力建模	16
2.4.2 基于注意力的推荐方法	17
2.5 本章小结	18
第三章 基于高阶共享单元的多任务学习	19
3.1 参数共享侧问题描述与分析	19
3.2 基于高阶共享单元的多任务学习框架	21
3.2.1 高阶共享单元	21

3.2.2 多任务交替学习框架	24
3.3 本章小结	25
第四章 基于图卷积神经网络的协同信息感知	26
4.1 推荐侧问题描述与分析	26
4.2 基于图卷积神经网络的协同推荐方法的提出	27
4.2.1 协同节点传播	28
4.2.2 协同消息构造	29
4.2.3 协同消息聚合	30
4.3 本章小结	31
第五章 基于多层注意力机制的图谱补全	33
5.1 图谱侧问题描述与分析	33
5.2 基于多层注意力机制的图谱补全方法的提出	34
5.2.1 同阶邻域信息的传播与聚合	34
5.2.2 异阶邻域信息的传播与聚合	36
5.3 本章小结	36
第六章 基于知识图谱的推荐与补全联合网络	37
6.1 形式化定义	37
6.2 基于知识图谱的推荐与补全联合网络的提出	37
6.2.1 模型架构	37
6.2.2 图表示层	39
6.2.3 信息传播与聚合层	39
6.2.4 信息共享层	42
6.2.5 训练层	42
6.3 模型训练	43
6.4 本章小结	43
第七章 实验与验证	44
7.1 实验方案设计	44

7.1.1 实验数据集	44
7.1.2 实验环境	45
7.1.3 对比算法	46
7.1.4 评估指标	47
7.2 实验结果与分析	49
7.2.1 公开数据集对比实验	49
7.2.2 论文推荐场景对比实验	51
7.2.3 图谱补全实验	52
7.2.4 消融实验	52
7.2.5 模型超参数探索	53
7.3 本章小结	56
总结与展望	57
参考文献	59
攻读硕士期间取得的学术成果	63
致谢	64

图 清 单

图 1	科技文献知识图谱示例图	4
图 2	研究路线示意图	6
图 3	参数硬共享架构图	10
图 4	参数软共享架构图	10
图 5	DUPN 多任务架构图	11
图 6	NTN 模型架构图	15
图 7	点乘注意力机制示意图	16
图 8	多头注意力机制示意图	17
图 9	基于高阶共享单元的多任务学习框架图	21
图 10	高阶共享单元架构图	22
图 11	用户物品交互二分图	28
图 12	用户物品连通图	28
图 13	层内注意力机制结构示意图	34
图 14	MRKC 网络结构示意图	38
图 15	book 数据集上 Recall@K 的结果	49
图 16	movie 数据集上 Recall@K 的结果	50
图 17	music 数据集上 Recall@K 的结果	50
图 18	book 数据集上 AUC 指标随物品向量维度变化	53
图 19	movie 数据集上 AUC 指标随物品向量维度变化	54
图 20	music 数据集上 AUC 指标随物品向量维度变化	54
图 21	article 数据集上 AUC 指标随物品向量维度变化	54
图 22	paper 数据集上 AUC 指标随物品向量维度变化	55

表 清 单

表 1	相关参数列表	20
表 2	基于图卷积神经网络的协同信息感知方法中定义的数学符号	27
表 3	基于分层注意力机制的图谱补全方法中定义的数学符号	33
表 4	基于知识图谱的推荐与补全联合网络中定义的数学符号	37
表 5	数据集描述	44
表 6	服务器硬件与软件配置	45
表 7	各算法在点击率预估场景下的 AUC、ACC 和 F1 指标对比	48
表 8	各算法在论文数据集上的 AUC 和 F1 指标对比	51
表 9	各算法在图谱学习场景下的 RMSE 指标对比	51
表 10	MRKC 变体模型在不同数据集下的 AUC 结果	51
表 11	MRKC 不同训练间隔对 AUC 结果的影响	52
表 12	MRKC 不同信息传播深度对 AUC 结果的影响	53

第一章 绪论

1.1 论文背景和意义

近年来,科学技术迅猛发展,以科研机构、科研成果、科技文献为主的科技数据不断涌现。由于这类数据符合海量、多样、流转快速和价值密度低等特征,因此被称为科技大数据^[1]。

科技大数据提供了海量的资源和便捷的获取途径,但也给科研工作者带来了信息检索上的挑战。以科技文献为例,科研人员常需要通过关键词检索来获得感兴趣的科技文献,但这种被动搜索的方式只适用于科研人员有明确的领域文献查找需求,如果科研人员无法精准定位查找的关键词,搜索结果往往偏差较大。推荐系统则是一种主动式的信息检索方式。与搜索相比,推荐系统不需要提供关键词,它会对科研人员的阅读偏好、研究领域以及历史浏览行为进行分析,然后推荐其可能感兴趣的科技文献。推荐系统不仅能够大大节省科研人员寻找相关文献的时间,也可能给科研人员带来新的兴趣和灵感,从而有效缓解需求模糊、信息爆炸带来的检索问题。

传统的推荐方法仅考虑科研人员的历史行为,忽略了科技大数据丰富的语义信息,因此推荐效率不高。近年来,知识图谱相关技术蓬勃发展,科技文献、专家学者和课题项目等相互关联的数据构成的知识图谱为科技大数据的推荐提供了新的灵感。利用知识图谱辅助推荐,可以有效利用图谱的结构优势和语义信息,丰富科技资源的表征,挖掘用户历史交互行为中更深层次的偏好,刻画用户的个性化画像,从而实现精准的科技资源推荐服务。

本文的课题来源于实验室参与的国家重点研发计划“科技大数据理论与技术研究”项目中的“面向服务的知识服务构件开发”子课题。该课题面向科技服务的开放协同与精准服务需求,研究、建立多领域科技资源大数据模型,开发分布式结构的科技资源表示、个性化推荐等知识服务构件。在科技资源图谱建构与表示研究的基础上,本文研究基于知识图谱的多任务推荐方法,为不同场景提供个性化推荐服务,为科技大数据平台项目的推荐服务提供坚实的理论支撑。

1.2 国内外研究现状

1.2.1 传统推荐算法

推荐系统是目前主流的主动式信息检索工具，其服务对象统称为用户，推荐给用户的对象统称为物品。除了性别、年龄与姓名等静态信息，对用户的建模一般还包括社交关系和历史浏览等动态信息。物品一般为实物商品，也可以是电影、音乐与论文等虚拟商品。

上世纪 90 年代，推荐系统开始进入研究人员的视野。随着美国计算机协会 ACM 推动开展推荐系统顶级会议 ResSys，二十多年来推荐技术不断发展和沉淀，现在推荐系统已经成为信息检索领域热门的研究方向。按照方法的不同，传统推荐算法可以分为：基于内容的推荐算法、基于协同过滤的推荐算法和混合推荐算法^[2]。

(1) 基于内容的推荐算法核心思想是对物品多种属性的建模。这类方法首先根据物品属性构建物品特征，利用统计学方法分析用户交互过的相关物品，提取物品间的共同属性，作为用户的兴趣特征；再依据用户的兴趣特征和待推荐物品的特征，计算两者的相似度，从而为用户推荐相似度高的物品。基于内容的推荐算法主要依赖于对物品属性的建模，它对稀疏的用户物品交互并不敏感，同时对系统中的新物品非常友好，能够有效解决物品冷启动问题。但物品的特征建模是一件费时费力的人工工作，并且特征质量直接决定最后的推荐效果，使得建模方法在不同推荐场景下难以泛化。

(2) 基于协同过滤的推荐算法主要考虑推荐场景中存在的群体效应：一般兴趣相似的用户在物品选择上会有相似的行为。基于协同过滤的推荐算法就是利用这一特点向用户推荐他可能感兴趣的物品。这类算法还可以细分为：基于物品、基于用户和基于模型的协同过滤方法^[3]。

类似基于内容的推荐算法，基于物品的协同过滤方法首先计算物品间的相似度，但其并不依赖于对物品属性的建模，而是考虑物品的交互历史——交互过的用户重合度越高，则物品相似度越高。这种方法特征建模简单，但是它并不适用于没有历史交互记录的新物品，会引起物品的冷启动问题。

区别于基于物品的协同过滤方法，基于用户的协同过滤方法是从用户角度出发——交互过的物品重合度越高，用户间越相似——为用户寻找其相似用户，然后为用户推荐其相似用户交互过的物品。这类方法仅考虑不同用户间的历史交互记录，与“协同过滤”

的本质思想很接近，不需要过多关注物品的建模，因此它存在一定程度的用户冷启动问题。

基于模型的协同过滤方法的核心思想是利用统计或机器学习方法（矩阵分解、聚类等），从交互记录中学习一个用于预测用户物品交互概率的模型。一般情况下，这种方法学习得到的模型结构简单有效，同时大小远小于原始的交互矩阵，空间需求较低。但这类方法的推荐结果不具有很好的可解释性，同时也存在数据稀疏和用户、物品的冷启动问题。

（3）混合推荐算法指的是利用两种或者两种以上推荐算法来配合，克服单个算法存在的问题，期望更好地提升推荐的效果，主要包括不同数据的特征组合、数据增强混合、多源推荐结果的加权组合和多模型的融合等。该类方法可以对不同的推荐场景进行个性化组合和建模，能够充分利用场景中的上下文、特征值和时间戳等各类信息。

1.2.2 基于知识图谱的推荐算法

近年来，基于大数据的知识图谱技术发展迅速。知识图谱本质上是语义网络的知识库^[4]，内部节点与关系为现实世界的客观知识，通常可以形式化地由三元组表示。基于知识图谱的推荐算法能够利用知识图谱的图结构优势和丰富的语义信息来提升推荐算法的性能。目前，在该领域研究发展中主要形成了两类方法：基于路径的方法和基于传播的方法。

（1）基于路径的方法

在知识图谱中，节点间存在多种关系，元路径则是表示节点与其关系的一种方式。元路径一般指连接两个节点的路径，它可以被看作是一种长距离的关系。例如在科技文献图谱中，如图 1 所示，可以构建用户-论文-作者-论文和用户-论文-关键字-论文两条元路径。基于元路径的推荐算法通过利用不同类型节点之间的元路径来挖掘用户和物品之间的关系，从而实现推荐。

Yu 等人^[5]提出一种在知识图谱中基于元路径的推荐方法（Personalized Entity Recommendation, PER）。PER 模型首先人工设计用户和物品在知识图谱中的关联元路径，利用 PathSim^[6]，由公式(1.1)计算不同节点间的相似度，得到用户 u_i 到物品 e_j 之间的用户偏好特征值。

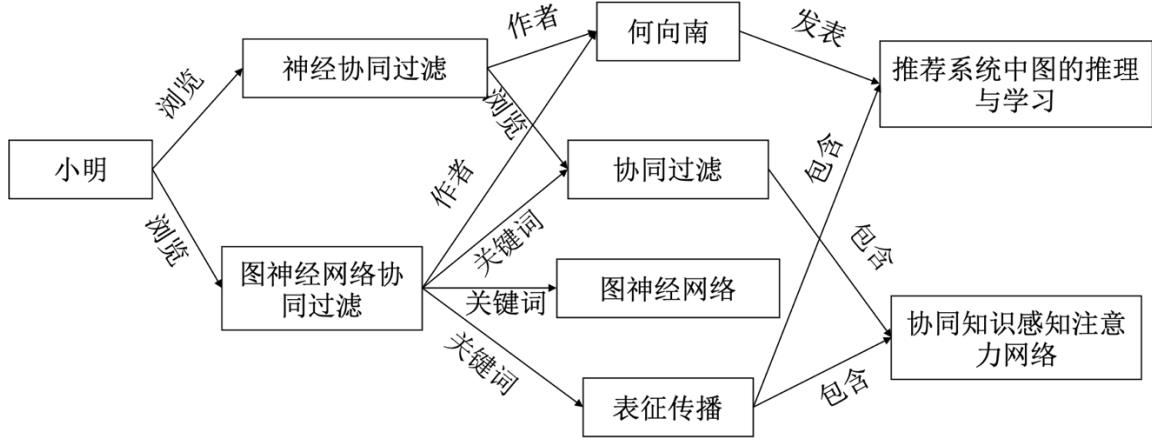


图1 科技文献知识图谱示例图

$$s(u_i, e_j | P) = \sum_{e \in I} \frac{2 \times R_{u_i, e} \times |P_{e \sim e_j} : P_{e \sim e_j} \in P'|}{|P_{e \sim e} : P_{e \sim e} \in P'| + |P_{e_j \sim e_j} : P_{e_j \sim e_j} \in P'|} \quad (1.1)$$

其中 $P_{e \sim e_j}$ 、 $P_{e \sim e}$ 和 $P_{e_j \sim e_j}$ 分别表示 e 与 e_j 、 e 与 e 、 e_j 与 e_j 之间的路径， P 和 P' 表示元路径的复合关系， $R_{u_i, e}$ 表示 u_i 与 e 是否直接相连，是则取值为1，否则为0。用户偏好特征值构建用户和物品之间的用户偏好特征矩阵，通过矩阵分解^[7]可以得到元路径的特征向量，进而计算用户和物品向量，从而预测用户与待推荐物品的交互概率。

基于元路径的推荐算法依赖于元路径的抽取，虽然该方法有一定的可解释性，但元路径的构建耗时耗力，并且对于种类繁杂的科技资源，该方法对相关领域知识和经验要求较高。当推荐场景或知识图谱发生变化时，元路径需要重新定义，而科技数据往往日新月异，因此该类算法无法很好地支持科技数据的推荐。

(2) 基于传播的方法

知识图谱中关系种类繁多、节点连接复杂，无法有效抽取其中存在的所有推荐元路径。因此，一些学者提出基于传播的推荐方法，在物品对应知识图谱中节点的基础上，考虑图谱的中心-邻域结构，传播和聚合来自邻接节点的信息，将信息传播到高阶邻接节点。

基于传播的方法充分利用知识图谱中的图结构信息和语义信息，向量建模图谱中的节点和关系。Wang 等人^[8]提出涟漪网络 (RippleNet)，这是一个经典的基于知识图谱传播机制的推荐模型，该模型通过在知识图谱中不断传播用户交互过的物品，探索用户潜在兴趣。RippleNet 借鉴水波传播机制，以用户交互过的物品为种子，并在知识图谱上以

种子集合为中心，一圈一圈地向外传播信息，通过偏好传播得到多阶邻居，从而实现用户兴趣的延伸和扩展。Wang 等人^[9]提出协同知识感知的注意力网络(Collaborative Knowledge-Aware Attentive Network, CKAN)，在用户侧，CKAN 模型同样以用户交互过的物品为种子，创建初始种子集合；在物品侧，利用两层协同过滤，以物品-用户-物品为路径构建物品的初始种子集合。通过注意力机制聚合图谱中多阶传播的种子信息，得到最终的用户和物品向量表示，进而预估用户交互物品的概率。Chen 等人^[10]提出的具有个性化推荐协作指导的注意力知识感知图卷积网络(CG-KGR, attentive Knowledge-aware Graph convolutional networks with Collaborative Guidance for personalized Recommendation)则认为知识图谱中的信息可能并不总是有助于向所有用户推荐，因此CG-KGR 将历史交互封装为交互信息摘要，然后以此为指导，从知识图谱中提取信息，最终提供更精确的个性化推荐。

除了这两大类主流方法，近年来也有部分学者对基于知识图谱的推荐算法作出了新的尝试。Geng 等人^[11]提出了一个新颖的路径语言建模推荐框架(Path Language Modeling over Knowledge Graphs for Explainable Recommendation, PLM-Rec)。通过路径序列解码，PLM-Rec 在由节点和边组成的路径上学习一个语言模型，在捕捉用户行为的同时，消除了图谱多跳的限制，从而有效提高了推荐的效果和可解释性。Park 等人^[12]则引入情感分析，首先分析用户对物品的评论和评价，构建了一个情感感知知识图谱，然后提出基于强化学习的语义感知策略，对图谱上的物品进行情感感知和推荐。

目前主流基于知识图谱的推荐算法往往基于一个假设——认为相关数据构建的知识图谱是完备的——通过在知识图谱中传播与聚合用户兴趣来提升推荐效果，这类方法并不关注图谱的学习，忽略了推荐任务对图谱完备性提升的正向作用。同时，这类算法同质化推荐系统中的物品与图谱的节点，在同一向量空间下表示二者，忽略了对其自身背景信息的建模，也引入了不同背景的噪声。

1.3 论文研究目标及内容

1.3.1 研究目标

本文针对物品与节点的异质化建模和图谱完备性提升的问题，提出一种基于知识图谱的多任务推荐方法，在补全知识图谱的基础上，进一步挖掘物品的潜在特征，扩展推

荐系统中用户的潜在兴趣，提高用户对特定物品点击率预估的准确率。此方法在传统推荐场景，例如电影、音乐推荐等，相比于传统的推荐方法，能取得更优异的推荐效果。此外，该方法能匹配适用于科技大数据中面向科研人员的文献推荐服务。

1.3.2 研究路线

为了达成研究目标，本文的研究路线如图 2 中所示。首先调研、获取用户物品交互数据以及物品关联的外部数据，分别构建用户物品连通图和知识图谱。其次，调研基于知识图谱推荐算法的相关理论和现有方法，并从现有方法中归纳出研究方向，找到研究突破口。接着，分别建模物品与节点，通过深度学习特征交叉相关先进方法，设计基于高阶共享单元的多任务学习框架，利用多任务学习机制来共享任务间的相关信息。然后，在推荐侧采用图卷积神经网络捕捉用户与物品交互中的深层协同信息，并用门控单元控制信息的通入量；在图谱学习侧采用多层注意力机制来分别建模同阶信息差异性和异阶信息衰减。接下来，将推荐侧模型和图谱学习侧模型整合到基于高阶共享单元的多任务学习框架中，设计一个端到端的基于知识图谱的推荐模型。最后在书籍、电影和音乐三个不同领域的公开数据集和自建的论文数据集上进行实验比较，经过多个指标的评估，验证提出算法的有效性，并对实验结果进行深入分析。

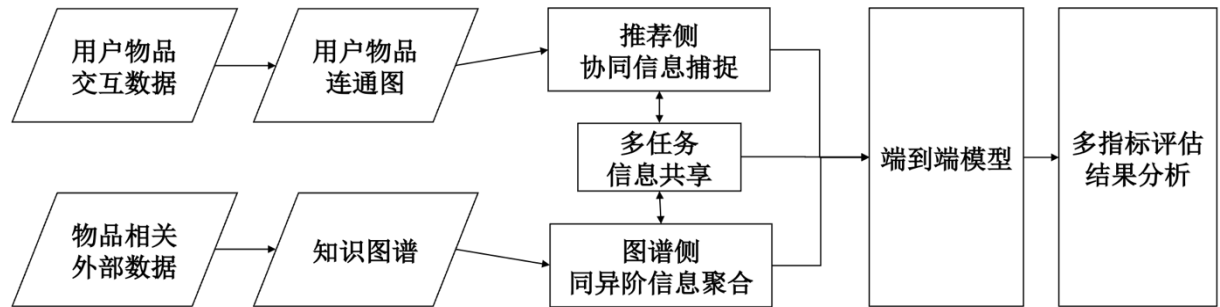


图 2 研究路线示意图

1.3.3 研究内容

为了达到研究目标，本文的主要研究内容包括以下几个方面：

(1) 多任务学习与共享

现有基于知识图谱的推荐算法大多使用相同嵌入向量来表示物品和对应节点，借助图谱的图结构，传播与聚合信息来丰富用户和物品的表示。但是物品和节点是同一个对象在不同场景下的两种描述，其向量表示应该包含自身环境特征，因此这类算法的表示

是次优的。因此本文重要研究之一是对物品和节点在不同向量空间下进行的编码，构建推荐与图谱补全双侧任务，并针对现有特征交叉方法学习能力有限问题进行改进，加强物品与节点的交叉学习，提高推荐任务和图谱补全任务的双向推动作用。

（2）用户物品深度协同信息的捕捉

用户物品原始交互中存在丰富的协同信息，这类信息既能有效表征用户的兴趣，又能体现物品间的某种关联性。然而目前基于知识图谱的推荐算法大多将用户物品交互数据作为输入，构造用户的历史交互物品记录，再融入知识图谱中进行信息的传播与聚合，用户和物品交互中的协同信息常常被忽略。因此在多任务下的推荐侧，本文重点研究如何高效地捕捉用户和物品交互记录中的多层协同信息，同时改进高阶信息聚合的方式，控制不同距离下的协同信息对用户和物品的影响，深度刻画用户在交互过程中的兴趣，并加强推荐任务对知识图谱补全任务的正向作用。

（3）图谱多阶邻域信息的建模

知识图谱常以中心-领域的形式建构，中心节点拥有一阶到多阶的邻居集合。一般来说，图谱中的关系可以很好地表示多样的用户兴趣，不同的兴趣对交互的影响程度各不相同。因此在图谱侧，本文重点研究同阶邻域下，如何基于关系有效建模尾节点的重要性差异，区分同阶邻居对中心节点的不同影响，进而提升图谱补全任务对推荐任务的正向推动作用，提高推荐效果的可解释性。另一方面，邻居有远近之分，低阶邻域信息往往与原始节点信息更加相关，而高阶邻域信息的相关度则会低一些，因此建模异阶邻域信息在高阶传播时的衰减现象也是本文的研究要点之一。

（4）端到端推荐算法的提出与验证

在上述关键方法设计的基础上，本文研究提出一种基于知识图谱的多任务推荐算法。该算法异质化建模对推荐侧的用户和物品与图谱侧的节点和关系，通过信息的交叉与共享完善物品与节点的表示，同时构建推荐任务和图谱学习任务，充分学习两侧的语义和结构信息，然后进行端到端的模型训练，在提升图谱完备性的基础上，得到更加优异的推荐结果。最后，将提出的算法在论文推荐和其他领域的公开推荐数据集上进行对比验证，并对实验结果加以分析。

1.4 论文组织与安排

本文对基于知识图谱的多任务推荐方法进行研究，下面是各个章节的内容安排：

第一章先介绍论文的研究背景与意义、推荐系统整个大领域与基于知识图谱的推荐系统这一小领域的研究现状。然后，明确论文的研究目标与研究路线，概括论文的主要研究内容。最后是论文的组织与安排。

第二章是对论文所涉及的背景知识、相关理论基础和应用情况的介绍。首先介绍多任务学习的基础概念及其在推荐系统领域的应用，然后介绍图神经网络和图表示学习的相关技术与方法，最后介绍注意力机制及其在推荐系统领域的应用。

第三章首先指出现有基于知识图谱的多任务推荐方法存在的问题，然后详细介绍为改进现有问题而提出的基于高阶共享单元的多任务学习框架。

第四章首先介绍目前基于知识图谱的多任务推荐方法中推荐侧任务存在的问题，然后对提出的基于图卷积神经网络的协同信息感知方法进行详细介绍。

第五章首先对目前基于知识图谱的多任务推荐方法中的图谱学习进行介绍，分析现存问题，然后针对现有问题提出了基于多层注意力机制的图谱补全方法，并作详细介绍。

第六章对本文提出的基于知识图谱的推荐与补全联合网络(MRKC)进行详细介绍。首先对该网络要解决的问题进行形式化定义与描述，并列出相关数学符号的定义，然后具体介绍网络的各个模块，最后对模型训练所使用的损失函数和训练方式进行介绍。

第七章是实验部分，首先介绍实验的基本设置、数据集、选取的对比方法，接着说明了用于对比的推荐系统评估指标，然后对实验的结果进行详细论述，验证所提算法的有效性，最后深入分析 MRKC 模型多个超参数对结果的影响。

总结与展望部分，首先对本文的研究工作进行总结，接着对后续研究和未来发展进行展望。

第二章 相关理论基础

本章介绍本文涉及到的相关理论基础，首先介绍与本文基础框架相关的多任务学习方法，并概述目前多任务学习在推荐系统中的应用；其次介绍目前主流的图神经网络，并介绍图卷积神经网络对推荐系统的影响；然后介绍图表示学习算法，并分类描述知识嵌入相关研究，简要介绍基于图谱表示学习的图谱补全研究；最后介绍注意力机制的基础概念，并概括其在推荐系统中的应用情况。

2.1 多任务学习

多任务学习是迁移学习方法中的一种^[13]，它将多个相关的任务协同训练，不同任务通过共享参数来交叉学习，利用梯度的反向传播来分享和补充不同任务下的背景相关信息，提高多个任务的泛化效果。

2.1.1 参数共享机制

多任务学习的前提是不同任务间具有相关或可共享的参数。一方面，浅层的参数共享能够促进不同任务互相补充领域信息，促进任务间信息的获取与穿透；另一方面，不同任务的环境噪声各不相同，共享参数在模型反向传播时能够兼顾到多个任务的反馈，从而通过平均环境噪声来约束不同任务的表征学习，得到泛化能力更强的向量表示，有效降低过拟合的风险。参数共享主要分为硬共享机制和软共享机制^[14]。

（1）参数硬共享机制

硬共享机制的核心思想是多个任务共享同一底层神经网络，如图 3 所示，是目前多任务学习中最常见的一种方式。底层神经网络可以根据任务间相似的特征来构建，上层则可以根据不同任务特异化搭建输出层。共享的任务越多，模型就越能捕捉到不同任务间的相似特征，提高底层网络输出向量的泛化能力，进而降低了模型的过拟合风险。但由于所有任务共享底层网络，硬共享机制对任务间的相似度要求较高，差异度较大的任务间优化时会产生一定冲突。

（2）参数软共享机制

软共享机制中并不设置共享的参数，每个任务都有自己的参数和模型架构，但不同任务间通过限制正则化距离来保障参数的相似，进而约束不同任务的训练，如图 4 所示。

软共享机制很好地解决了硬共享机制的共享参数优化冲突问题，但需要训练的参数极大地增加，其约束机制也很大程度上受传统多任务学习中正则化技术的影响，需要更多的训练数据学习任务间的相似^[15]。

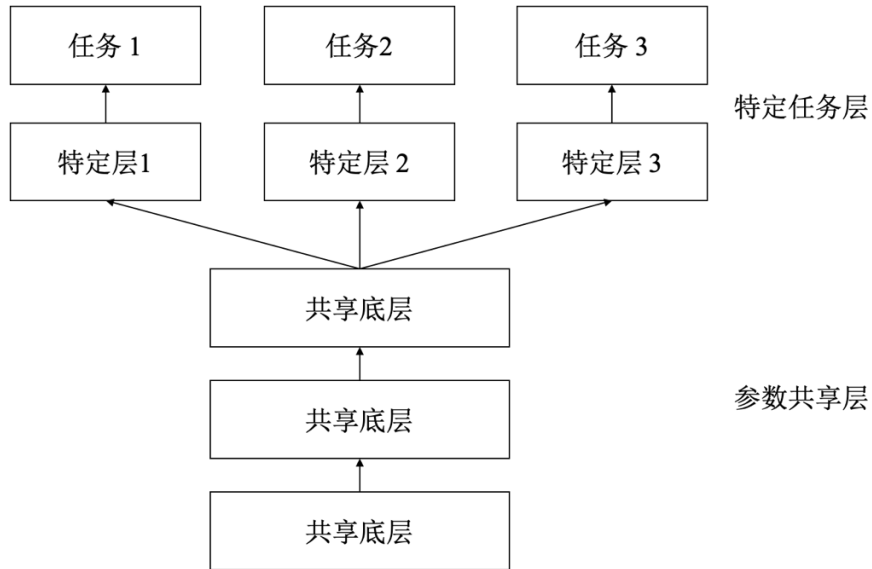


图3 参数硬共享架构图

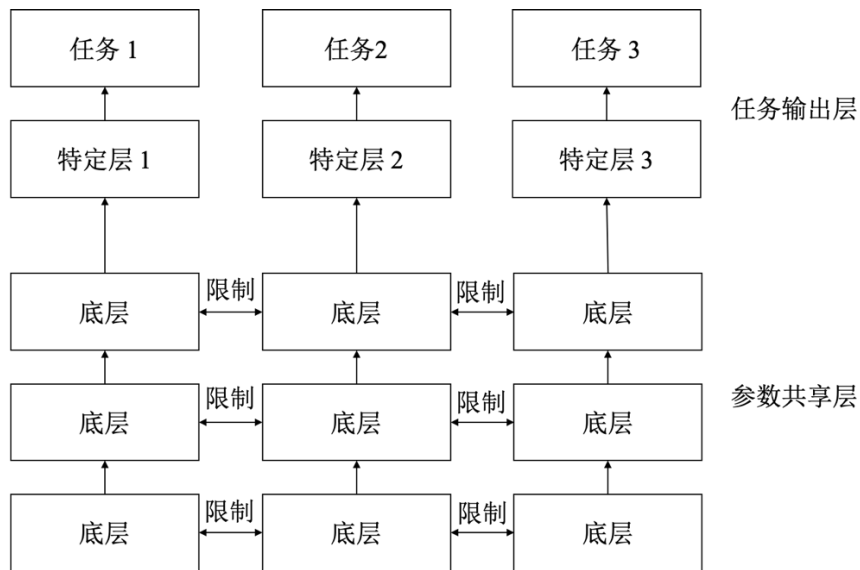


图4 参数软共享架构图

2.1.2 多任务推荐方法

推荐场景下，常常需要同时考虑点击率(CTR)、转化率(CVR)、点击转化率(CTCVR)、视频播放时长和用户停留时长这些指标。如果对每个任务单独优化，需要花费很多人力，并且无法学习到其他任务的背景信息。利用不同任务的关联性——例如 CTR 和 CVR——进行联合训练往往能够同时优化多个任务。

阿里团队的 ESMM (Entire Space Multi-Task Model) 模型^[16]借鉴多任务学习的思路, 通过底层参数软共享同时拟合 CTR 和 CTCVR。由于 CVR 可由 CTR 和 CTCVR 推导出来, 同时 CTR 可将有点击行为的用户物品作为正样本, 没有点击行为的用户物品作为负样本, 而 CTCVR 可将同时有点击行为和购买行为的物品作为正样本, 其他作为负样本, 因此 CTR 和 CTCVR 都可从整个样本空间进行训练, 从而有效解决了传统 CVR 预估中存在的样本选择偏差问题。另外由于 CTR 任务的训练样本量要远超过 CVR 任务的训练样本量, 不同任务间共享的嵌入层能够使得 CVR 任务也能够从只有曝光没有点击的样本中学习, 从而缓解训练数据稀疏的问题。

DUPN (Perceive Your Users in Depth: Learning Universal User Representations from Multiple E-commerce Tasks) 模型^[17]则通过硬共享机制共享不同任务的底层参数。如图 5 所示, 模型分为行为序列输入层、Embedding 嵌入层、LSTM 层、Attention 注意力层、下游多任务层。模型输入时除了用户行为序列外, 还引入了一些外部辅助信息, 丰富物品的表示, 一定程度上缓解了商品的冷启动问题。Embedding 层主要对序列中物品的特征做处理。得到序列行为的嵌入表示之后, 经过一个 LSTM 层, 把序列信息考虑进来, 并通过 attention 层去区分和聚合用户不同行为的重要性, 最终拼接上用户表征作为用户的表示。

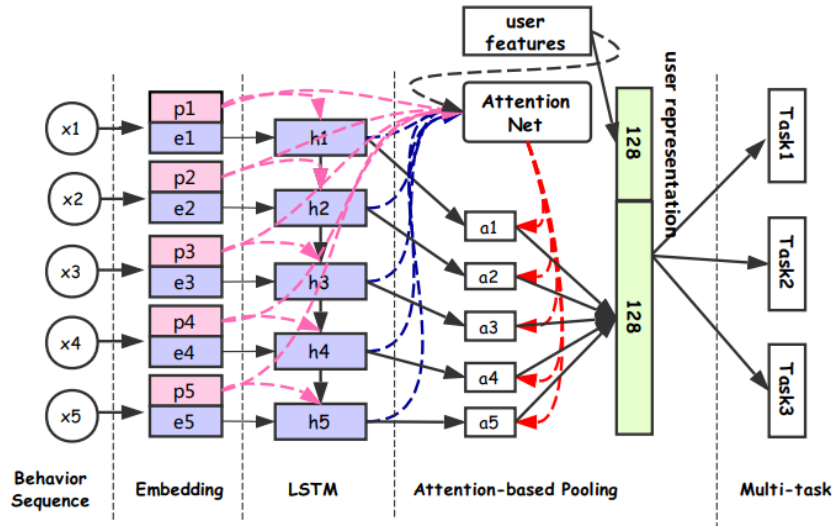


图 5 DUPN 多任务架构图

DUPN 模型输出的用户表示可以用于下游多种任务的训练: CTR、L2R(Learning to Rank)、用户购买力预测 PPP (Price Preference Prediction) 和用户潮流偏好 FIFP (Fashion Icon Following Prediction) 等。DUPN 模型将学习到的用户、商品嵌入向量迁移到其它

下游任务中，对于每个任务而言，其他的任务都视作正则化限制，可以使学习得到的用户表示泛化性更强，更可靠。

2.2 图神经网络

图神经网络（Graph Neural Network, GNN）的概念首先由 Gori 等人^[18]提出，并由 Scarselli 等人^[19]对其理论进一步补完善。早期的图神经网络的循环神经网络架构无法并行处理，训练需要花费大量时间，其发展受到限制。随着卷积神经网络在计算机视觉领域大获成功，图神经网络从中受到了启发。但是图像在欧几里得空间中是规则的，其有固定数量的邻居，使得图像上的卷积操作是一件直观且容易的事情。为了引入卷积思想，图神经网络的研究开始往非欧式空间中发展。

基于谱的图卷积神经网络（Graph Convolution Networks, GCN）^[20]是图神经网络领域非常经典的工作，类似信号处理，它引入图滤波器来定义图上的卷积操作。首先将欧式空间下的图信号投影变换到谱空间中，然后利用傅里叶变换和空间平移不变性实现谱空间下的卷积操作。然而，基于谱的图卷积神经网络训练时需要加载整个图，内存的限制使得该方法无法进一步发展。于是基于空间的图卷积神经网络^[21]被提出来，它基于节点的邻域节点集合来定义图卷积。假设有一组图数据，其中有 N 个节点，节点邻接矩阵表示为 A 。在基于空间的图卷积神经网络下，图卷积操作可以定义为：

$$H^t = \sigma \left(D^{-\frac{1}{2}} A' D^{-\frac{1}{2}} H^{t-1} W^{t-1} \right) \quad (2.1)$$

其中， $A' = A + I$ ， I 表示单位矩阵， D 是 A 的度矩阵， H^t 表示 t 层的特征矩阵， W 为每一层可学习的参数矩阵， σ 是一个非线性激活函数。

受到注意力机制影响，图神经网络进一步发展成图注意力网络（Graph Attention Networks, GAT）^[22]，在邻域信息聚合时建模尾节点的重要性差异。GAT 中，图卷积操作表示为：

$$h_i^t = \sigma \left(\sum_{j \in N_i} \alpha(h_i^{t-1}, h_j^{t-1}) W^{t-1} h_j^{t-1} \right) \quad (2.2)$$

其中 $\alpha(\cdot)$ 是注意力函数，它用来计算邻居节点 j 对目标节点 i 的贡献值。

推荐系统中存在多种图——用户-物品交互二分图，外部信息网络（例如社交网络、异质信息网络和知识图谱等）。随着图神经网络研究进展迅速，利用图神经网络进行推荐的研究工作也越来越多。

基于图神经网络的推荐算法主要利用图神经网络技术聚合推荐系统中的用户与物品的邻域信息，丰富用户偏好和物品表征。神经图协同过滤（Neural Graph Collaborative Filtering, NGCF）^[23]最早在用户-物品交互二分图上应用图神经网络，通过协同信号的传播和聚合，丰富用户和物品的表示，进而取得了更优异的推荐效果。此后，有学者简化 NGCF，提出基于 GCN 改进模型的 LightGCN^[24]算法。它简化了 NGCF 的邻域传播操作，舍弃了邻域信息聚合时的特征变换和非线性变换部分，使得模型训练变得更加容易，学习得到更加稳定、鲁棒性更高的用户和物品表示，进一步提高了推荐的实验效率。

基于知识图谱的推荐算法中也引入图神经网络加以改进，基于知识图谱的图注意力网络（Knowledge Graph Attention Network, KGAT）^[25]是目前推荐效率最好的方法之一。具体地，它首先将知识图谱和用户-物品交互二分图融合为协同知识图（Collaborative Knowledge Graph, CKG），该图将用户和物品统一融合到节点中，将用户-物品交互融合到关系中；其次利用传统知识图谱嵌入方法 TransR^[26]学习图中所有节点和关系的低维向量表示；然后采用信息传播层、基于知识的注意力层和信息聚合层三层结构，利用图卷积神经网络分别聚合节点的多阶邻域信息，得到最终用户、物品和节点的向量表示。最后对用户和物品向量进行点积操作，计算用户与物品交互的概率值。

2.3 图表示学习

2.3.1 图表示学习方法

图表示学习将图中节点嵌入到同一低维向量空间中，训练后相似节点的嵌入向量相似——“相似性”由具体方法定义——进而应用于下游任务（例如图节点分类、链路预测等）。目前，图表示学习方法主要分为图嵌入、知识图谱嵌入和图神经网络三大类^[27]。

（1）图嵌入方法

图嵌入方法最早是从词嵌入技术 Word2Vec^[28]发展而来。Word2Vec 主要学习语句序列中不同词间的共现关系和前后关系，通过语料库的训练得到每个词的低维向量表示。在 2013 年谷歌开源 Word2Vec 以后，部分学者从中获得启发，相继提出 DeepWalk^[29]、

Node2Vec^[30]等图嵌入方法。类比 Word2Vec 中的句子序列，DeepWalk 和 Node2Vec 算法将图上节点类比于词，通过图上的随机游走采样得到若干特定长度的节点序列，再利用 Word2Vec 的 skip-gram 模式进行训练得到图中各个节点的低维向量表示。

(2) 知识图谱嵌入方法

普通无权图中边关系单一，因此图嵌入方法大都仅考虑节点间的相似性，弱化甚至忽视对关系的建模。知识图谱中关系是连接节点的重要信息，其丰富的结构和语义信息能够有效描述节点间的相关度，因此基于知识图谱的嵌入方法开始关注关系的建模。在知识图谱嵌入方法中，基于翻译的方法是最早被提出的一类方法。该系列方法的核心思想是：头节点向量 h 和关系向量 r 的和应当接近于尾节点向量 t ，即 $h + r \approx t$ ——这一点符合直观理解——向量一般用 L1 或者 L2 范数来衡量“接近”。2013 年提出的 TransE^[31]是最经典的基于翻译的方法，其通过随机负采样的节点替换实际存在的头或尾节点构造负样本，其损失函数定义为：

$$L(h, r, t) = \max(0, d_{pos} - d_{neg} + \text{margin}) \quad (2.3)$$

其中 d_{pos} 和 d_{neg} 分别表示正、负样本三元组， $d = ||h + r - t||$ 为 L1 或 L2 范数， margin 是一个常数，表示正负样本的最大距离。在 TransE 基础上衍生出了一系列 Trans 类方法，例如 TransH^[32]，TransR^[26]和 TransD^[33]等。

(3) 基于神经网络的方法

随着深度学习的发展，基于神经网络的方法是目前的主流研究方向，基于张量分解和基于图神经网络的图谱表示学习方法相继被提出。这类方法直接输入三元组 (h, r, t) ，使用多层的神经网络拟合知识图谱的结构，输出该三元组存在的概率。这类方法主要有 SME^[34]、NTN^[35]和 MLP^[36]等。NTN 整合和拓展了各种早期的神经网络模型，将神经网络模型中的一个线性层替换为一个双线性张量层，从多个维度关联两个节点向量，结构如图 6 所示。

NTN 中计算两个节点间存在特定关系概率的得分函数为：

$$f_r(h, t) = r^T \tanh(h^T M_r t + M_r^1 h + M_r^2 t + b_r) \quad (2.4)$$

其中张量层采用三维张量 M_r ，比起二维矩阵，该三维张量能进一步地从不同语义层面上表示每一条关系。 $M_r^1 h + M_r^2 t + b_r$ 则是常规的神经网络线性层，包含权重矩阵和偏置， \tanh 是非线性激活函数， r^T 是网络中的线性层。该模型同时包含了张量层以及常规

的线性层，因此拥有很强的表达能力。此外，NTN 使用预先训练好的节点词向量的平均值初始化节点与关系的表示，对模型的效果有了进一步的提升。

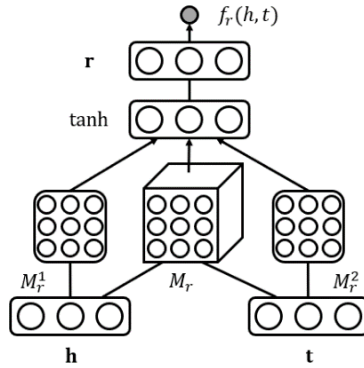


图 6 NTN 模型架构图

2.3.2 基于表示学习的知识图谱补全

知识图谱的构建基本是通过人工或算法，利用知识抽取技术从文本中识别节点、抽取关系。然而受限于图谱构建技术和客观因素，人工构建的知识库无法充分、完整，存在于世界中的很多实体、关系并没有体现在知识图谱中，存在很多内容缺失，因此图谱是不完备的。当训练下游任务的数据在知识图谱中不存在时，任务的效果会大打折扣。因此对于现有知识图谱的完善，即知识图谱补全任务，是当前知识图谱领域的主要研究方向之一。

对知识图谱进行补全，需要去挖掘知识图谱中各类客观事实之间的潜在规律，为了有效表征图谱中的节点和关系，基于路径的 PRA 方法^[37]、基于强化学习的 DeepPath 方法^[38]、基于元学习的 One-Shot Relational Learning for Knowledge Graphs 方法^[39]等相继被提出。随着图神经网络的发展，知识图谱的三元组大多抽象转化为标识符形式，为了更加有效地表征图谱中的结构和语义信息，部分学者将图表示学习引入知识图谱补全任务中。基于图表示学习的图谱补全首先通过对知识图谱已有结构和信息的感知，学习规律，将三元组中的节点和关系嵌入到连续低维向量空间中，即对知识图谱进行初始编码；再设计有效的节点间、节点与关系间的信息交互，丰富图上的特征交叉与融合；最后根据特定需求设计得分函数，将知识图谱中事实存在的预测概率同原始事实数据进行比对，即通过损失函数来优化。虽然知识图谱补全任务是图表示学习嵌入后的上层应用，但由

于当前的知识图谱补全任务都是依托于知识图谱嵌入来完成的，并且学术上一般不对两者进行算法研究上的区分，因此本文也不将这两者区别看待。

2.4 注意力机制

2.4.1 注意力建模

对于文字，图片等序列形式的数据，通常会采用循环神经网络（Recurrent Neural Network, RNN）^[40]架构，利用 RNN 单向地从后往前或者从前往后进行信息传播，然而 RNN 只能顺序计算，限制了其并行训练的能力，无法训练大规模数据；同时其结构容易出现梯度消失的现象，很难建模长序列信息。为了建模长距离序列数据的重要性差异，注意力机制火速发展起来，注意力机制的核心思想是模型收集信息时，应当给予重点信息更多的关注，主动弱化甚至忽略无关的部分。这一思想可以用人的视觉形象地描述：当人们看到一张图片时，首先会注意到画面的重要部分，然后才会查看其他部分。因此，在注意力机制的研究中，如何捕捉信息的重点部分，区分重点与非重点是一个难题。Transformer^[41]第一次提出用注意力机制并行提取信息的重点部分，大大提升了 RNN 类模型的训练效率；同时自适应地学习重点与非重点信息的不同权重，并尽可能地保留序列中的全部原有信息，有效解决了 RNN 无法建模长序列的问题。

Transformer 中的注意力机制分为：点乘注意力机制和多头注意力机制。点乘注意力机制是最常见的一种结构，如图 7 所示。

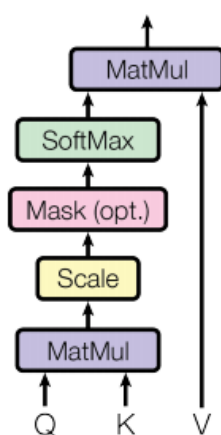


图 7 点乘注意力机制示意图^[41]

图中的 Q 、 K 和 V 分别代表查询向量 Query，键向量 Key 和值向量 Value。点乘注意力机制首先根据 Q 和 K 点乘计算二者的相似性权重，再进行归一化处理，最后根据归一化权重对数值向量 V 进行加权求和。这个过程用公式表示为：

$$a_i = \frac{q_i k_i}{\sum_{j=1}^l q_j k_j} v_i \quad (2.5)$$

其中， a_i 代表序列中第 i 个元素经过注意力机制后的向量表示， l 为序列长度。

为了考虑不同因素下 K 对 V 的重要性，多头注意力机制设置多个点乘注意力机制，通过若干组不同的线性矩阵将 Q 、 K 和 V 映射到不同向量空间下，再分别送入各自对应的点乘注意力机制中，拼接多因素下的重要性结果，最后通过线性映射得到最终的输出。图 8 展示了多头注意力机制的结构。

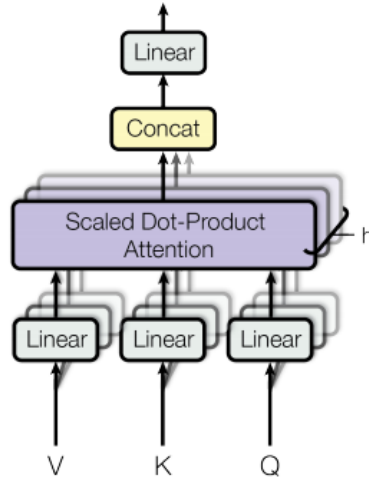


图 8 多头注意力机制示意图^[41]

在 Transformer 后，还出现了其他变种的注意力机制，例如自注意力机制^[42]，共同注意力机制^[43]等。

2.4.2 基于注意力的推荐方法

注意力机制最先在自然语言处理领域用于处理序列化语句，随后在多种领域研究中被广泛应用。推荐系统中，用户的浏览记录序列中也存在某种规律，因此目前注意力机制在推荐中的应用主要集中在序列建模上。

阿里巴巴的研究团队提出深度兴趣网络（Deep Interest Network, DIN）^[44]，DIN 应用注意力机制，为不同待推荐物品从用户的历史浏览序列中提取不同的用户兴趣向量表示。具体来说，DIN 以候选物品 c 的向量表示 V_c 作为查询向量，用户历史序列中的物品 i

的向量表示 V_i 作为键向量，学习序列中的每个物品对于候选物品的重要性权重，再以序列中物品的向量表示 V_i 为值向量，加权求和得到用户的兴趣表示，用公式表示为：

$$V_u = \sum_{i=1}^n g(V_i, V_c) V_i \quad (2.6)$$

其中 V_u 表示用户的兴趣向量， $g(V_i, V_c)$ 代表注意力计算方法。

微软提出的基于个性化注意力机制的新闻推荐算法（Neural News Recommendation with Personalized Attention, NPA）^[45]则是利用两个编码器分别编码新闻和用户来实现高效推荐。在新闻编码器中，NPA 把新闻的信息的多样性——出版社、标题和正文等——视为不同的视角，提取新闻的多维度表征。另外，不同的用户可能关注同一篇新闻的不同方面，为此，NPA 提出新闻级注意力机制，考虑不同用户对同一篇新闻的不同关注面，提取用户新闻偏好的个性化表征。

在基于知识图谱的推荐方法中，基于知识图谱的图卷积网络（Knowledge Graph Convolutional Network, KGCN）^[46]也应用了注意力机制。KGCN 采用点式注意力机制，在图中进行信息传播时以用户、关系和目标节点作为查询向量，邻居节点同时作为键向量和值向量，计算不同邻居节点对中心节点的重要性。

2.5 本章小结

本章介绍本文涉及到的相关研究概念，为后续章节奠定理论基础。首先详细介绍与本文基础框架相关的多任务学习方法，并概述目前多任务学习在推荐系统中的应用；其次介绍目前主流的图神经网络，并介绍图卷积神经网络对推荐系统的影响；然后介绍图表示学习算法，并分类描述知识嵌入相关研究；最后介绍本文使用注意力机制的基础概念，并介绍其在推荐系统中的应用情况。

第三章 基于高阶共享单元的多任务学习

在多任务推荐方法的研究中，参数共享单元是不同任务间相互学习和传播信息的通用方法之一，也是研究的核心之一。参数共享单元设计的主要目标是充分利用不同任务间的相关性，辅助推荐任务。设计有效的参数共享单元能够使得两个任务间相关的参数相互学习和共享，聚合来自不同任务的高阶信息。本章分析现有基于多任务学习的推荐方法中的问题，然后详细介绍基于高阶共享单元的联合任务学习框架。

3.1 参数共享侧问题描述与分析

在基于知识图谱的多任务推荐算法研究中，如何设计合理的参数共享单元是研究的核心之一。现有的方法中，大多存在以下几个问题：

(1) 忽略推荐中的物品和图谱中对应节点的区别。推荐系统能利用知识图谱的前提是推荐系统中的物品与知识图谱中的节点存在着一对一，或者一对多的关系。现有算法大多使用相同嵌入向量来表示物品和对应节点，并以此作为推荐任务和图谱表示学习任务的桥梁进行训练。但是，物品和节点是同一个对象在不同场景下的两种描述，同质化物品与节点，在同一向量空间下表示二者，会引入不同背景下的环境噪声，同时忽略了不同任务对其自身领域信息的建模。因此应该在不同向量空间下分别建模物品与节点，自适应学习不同任务下的自身环境特征，进而学习到二者更适合自身任务的表示。

(2) 任务间参数交叉学习能力有限。目前基于知识图谱的推荐算法大多基于参数软共享机制，设置参数约束用于不同参数间的特征交互学习。例如 DCN 模型^[47]中参数交叉公式为：

$$X_{l+1} = X_0 X_l^T W_l + X_l + b_l \quad (3.1)$$

这样的交叉单元特征交叉简单，两个参数只能学习到彼此 $O(l)$ 阶的特征，而 KTUP 模型^[4]中则直接将物品表示与对应节点表示相加作为最终物品表示；而 MKR 模型^[48]中采用交叉压缩单元来学习物品和节点：

$$V_{l+1} = C_l W_l^{VV} + C_l^T W_l^{EV} + b_l^V = v_l e_l^T W_l^{VV} + e_l v_l^T W_l^{EV} + b_l^V \quad (3.2)$$

该单元虽然能够学习到 $O(2^l)$ 阶的特征，但其中 W 和 C 均为 $R^{d \times d}$ 的方阵，要求交叉的特征都是同一向量空间下，导致该单元无法交叉学习不同维度的向量，对物品和节点的表示学习是次优的。

(3) 弱化了推荐任务与图谱补全任务的相互正向作用。目前基于知识图谱的多任务推荐方法常将用户或物品放入知识图谱中，借助图谱的语义信息和结构信息，通过传播与聚合信息来丰富用户和物品的表示，然而受限于图谱构建技术和客观因素，存在于世界中的很多节点、关系并没有体现在知识图谱中，存在很多内容缺失，图谱的不完备性制约了这类算法的发展。用户和物品的交互中蕴含着一定的物品关联信息，例如同一个用户交互过的物品往往具有一定相关性，捕捉这些关联信息能够有效辅助知识图谱的表示学习，并进一步运用于下游的图谱链路预测等相关任务，提高图谱完备性的同时辅助提升推荐效果。

基于以上问题，本文提出一种基于高阶共享单元的联合任务学习框架，将推荐中的物品和知识图谱中对应节点的表示区分开来，使用不同维度的向量进行编码，方便学习对应场景下的环境特征。同时设计基于异阶特征交叉的共享单元，适应不同维度特征间的高阶信息交互，互相学习物品表示与节点表示。最后依次训练推荐任务和知识图谱补全任务，在提升知识图谱完备性的基础上，提高推荐预测的准确率。

本节使用数学符号对本文要解决的问题进行了定义，表1给出了本节所使用到的数学符号。

表1 相关参数列表

符号	描述
$U = \{u_1, u_2, \dots, u_M\}$	用户集合， M 为用户数量
$V = \{v_1, v_2, \dots, v_N\}$	物品集合， N 为物品数量
h, r, t	头节点、关系和尾节点
E, R	知识图谱中的节点、关系集合
W_l^0, W_l^1	转置矩阵
$W_l^{VV}, W_l^{VE}, W_l^{EV}, W_l^{EE}$	压缩矩阵
v_l, h_l	物品 v 、节点 h 的 l 阶表示
S	高阶共享单元
d_v	物品表示的维度
d_h	节点表示的维度

3.2 基于高阶共享单元的多任务学习框架

本文提出基于高阶共享单元的多任务学习框架如图 9 所示。下面对它进行详细介绍。

框架输入层输入原始的用户物品交互数据和物品相关知识图谱。考虑物品和节点是同一个对象在不同场景下的两种描述，框架嵌入层针对物品和节点进行异质化编码使其向量表示包含自身环境特征，初始化物品向量表示 $v \in R^{d^v}$ ，节点向量表示 $h \in R^{d^h}$ ，其中 d^v 、 d^h 表示物品和节点的向量维度，为预先定义的超参数。框架在推荐侧和图谱侧分别构建对应任务，在局部视角下分别学习用户与物品、节点与关系的表示。经过推荐侧和图谱侧任务学习后，得到物品和节点的初阶表示 v_0 、 h_0 。为了让推荐任务和图谱任务学习来自对方任务的领域信息，框架定义高阶共享单元 S ，多阶交叉学习物品 v 与对应节点 h ，得到二者的高阶交互向量 v_l 、 h_l 。最后，框架分别定义点击预估层和图谱补全层，分别训练推荐任务和图谱任务。

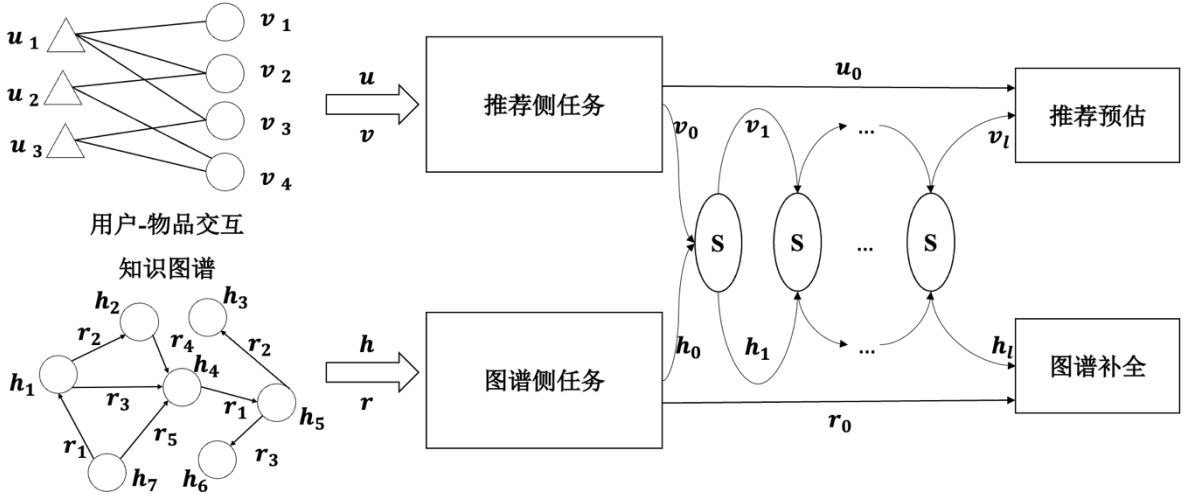


图 9 基于高阶共享单元的多任务学习框架图

3.2.1 高阶共享单元

对于物品 $v_0 = [v_0^{(1)} \dots v_0^{(d^v)}]^T$ 和节点 $h_0 = [h_0^{(1)} \dots h_0^{(d^h)}]^T$ ，推荐系统中常用 $v_0^{(i)} h_0^{(j)}$ 表示物品和节点的交叉特征，因为其可以衡量多个原始特征的相互作用^[45]，其中 $i \in \{1, 2 \dots d^v\}$ ， $j \in \{1, 2 \dots d^h\}$ 。根据 Weierstrass 近似定理^[49]，在某些平滑性假设下，任何函数都可以被多项式近似到一个任意的精度，因此推荐中常用高阶多项式去拟合可能的

交叉特征。对于第 l 阶的物品表示 v_l ，其第一范式 $\|v_l\|_1$ 中最大度的项表示为 $k \prod_i^{d^v} v_0^{(i)\alpha_i} \prod_j^{d^h} h_0^{(j)\beta_j}$ ， k 为常数，其中 $\{\sum_i^{d^v} \alpha_i + \sum_j^{d^h} \beta_j\}$ 常用来衡量特征交叉的能力。

为了有效共享学习物品与对应节点，本框架设计了一种应用于不同维度特征空间下的交叉共享单元，交叉学习物品和对应节点的高阶交互特征，结构如图 10 所示。

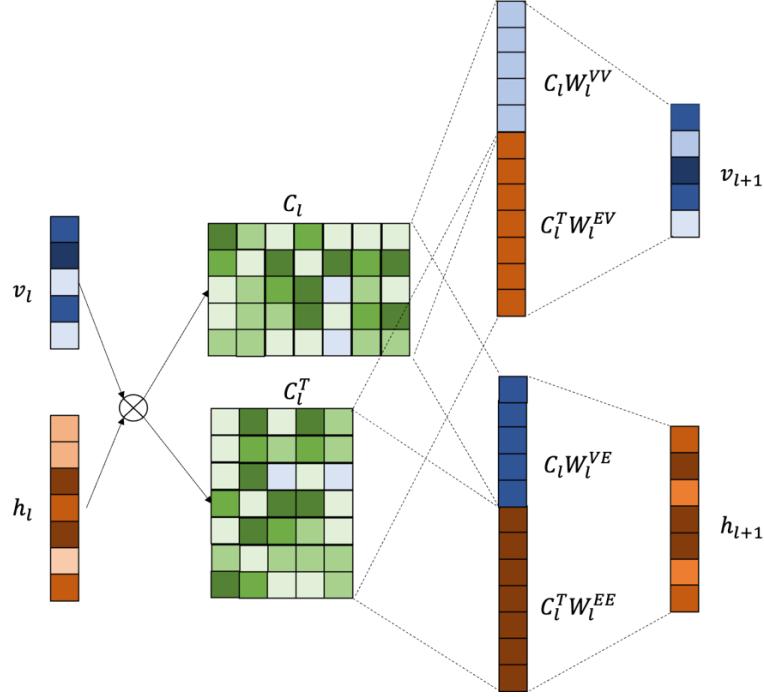


图 10 高阶共享单元架构图

以第 l 阶为例，为了显式建模了物品与节点所有可能的特征交叉 $v_l^{(i)}h_l^{(j)}$ ，高阶共享单元首先构造特征交互矩阵 $C_l \in R^{d^v \times d^h}$ ，定义为：

$$C_l = v_l h_l^T = \begin{bmatrix} v_l^{(1)} h_l^{(1)} & \cdots & v_l^{(1)} h_l^{(d^h)} \\ \vdots & \ddots & \vdots \\ v_l^{(d^v)} h_l^{(1)} & \cdots & v_l^{(d^v)} h_l^{(d^h)} \end{bmatrix} \quad (3.3)$$

C_l 中蕴含了物品与节点特征交叉的丰富语义信息。为了下一阶的特征交叉，需要将学习到的高阶交叉特征矩阵分别进行横向压缩和纵向压缩，表示为：

$$\begin{aligned} X_l &= C_l W_l^- \\ Y_l &= C_l^T W_l^- \end{aligned} \quad (3.4)$$

其中 X, Y 分别表示从垂直方向和水平方向压缩后得到的特征向量, W_l^V 、 W_l^E 分别表示从垂直方向和水平方向对特征交叉矩阵进行压缩的矩阵。再将压缩后的特征向量映射到物品与节点对应的向量空间下, 定义下一阶物品与节点表示为:

$$\begin{aligned} v_{l+1} &= W_0(X_l^{VV} || Y_l^{EV}) + b_l^V = W_l^0(v_l h_l^T W_l^{VV} || h_l v_l^T W_l^{EV}) + b_l^V \\ h_{l+1} &= W_1(X_l^{VE} || Y_l^{EE}) + b_l^E = W_l^1(v_l h_l^T W_l^{VE} || h_l v_l^T W_l^{EE}) + b_l^E \end{aligned} \quad (3.5)$$

其中 $W_l^0 \in R^{d^v \times (d^v + d^h)}$ 、 $W_l^1 \in R^{d^h \times (d^v + d^h)}$ 、 $W_l^{VV} \in R^{d^h}$ 、 $W_l^{EV} \in R^{d^v}$ 、 $W_l^{VE} \in R^{d^h}$ 、 $W_l^{EE} \in R^{d^v}$ 为可训练的权重矩阵, b_l^V 、 b_l^E 为可训练的偏置矩阵。

对于物品 $v_0 = [v_0^{(1)} \dots v_0^{(d^v)}]^T$, 节点 $h_0 = [h_0^{(1)} \dots h_0^{(d^h)}]^T$, 经过多阶共享单元的交叉共享, 第 l 阶的物品表示 v_l 和节点表示 h_l 的第一范式中最大度为 $O(2^l)$ 阶, 即假设 $\|v_l\|_1$ 中最大度项为 $k \prod_i^{d^v} v_0^{(i)\alpha_i} \prod_j^{d^h} h_0^{(j)\beta_j}$, 有 $\sum_i^{d^v} \alpha_i = 2^{l-1}$, $\sum_j^{d^h} \beta_j = 2^{l-1}$, 使得物品和节点都能学习到来自对方的 $O(2^l)$ 阶交叉特征, 下面简要证明该结论。

以第 l 阶的物品表示 v_l 为例, 其中:

$$\begin{aligned} W_l^0 &= \begin{bmatrix} W_l^{0(1,1)} & \dots & W_l^{0(1,d^v+d^h)} \\ \vdots & \ddots & \vdots \\ W_l^{0(d^v,1)} & \dots & W_l^{0(d^v,d^v+d^h)} \end{bmatrix} \\ W_l^{VV} &= [W_l^{VV(1)} \dots W_l^{EV(d^h)}]^T \\ W_l^{EV} &= [W_l^{EV(1)} \dots W_l^{EV(d^v)}]^T \\ b_l^V &= [b_l^{V(1)} \dots b_l^{V(d^v)}]^T \end{aligned} \quad (3.6)$$

当 $l = 1$ 时, 有:

$$\begin{aligned} v_1 &= W_0(v_0 h_0^T W_0^{VV} || h_0 v_0^T W_0^{EV}) + b_0^V \\ &= \begin{bmatrix} \sum_i^{d^v} W_0^{0(1,i)} v_0^{(i)} \sum_j^{d^h} h_0^{(j)} W_0^{VV(j)} + \sum_i^{d^h} W_0^{0(1,d^v+i)} h_0^{(i)} \sum_j^{d^v} v_0^{(j)} W_0^{EV(j)} + b_0^{V(1)} \\ \vdots \\ \sum_i^{d^v} W_0^{0(d^v,i)} v_0^{(i)} \sum_j^{d^h} h_0^{(j)} W_0^{VV(j)} + \sum_i^{d^h} W_0^{0(d^v,d^v+i)} h_0^{(i)} \sum_j^{d^v} v_0^{(j)} W_0^{EV(j)} + b_0^{V(d^v)} \end{bmatrix} \end{aligned} \quad (3.7)$$

则 v_1 的第一范式为:

$$\begin{aligned} \|v_1\|_1 = & \left| \sum_i^{d^v} \sum_j^{d^v} W_0^{0(j,i)} v_0^{(i)} \sum_k^{d^h} h_0^{(k)} W_0^{VV(k)} \right. \\ & \left. + \sum_i^{d^h} \sum_j^{d^v} W_0^{0(j,d^v+i)} h_0^{(i)} \sum_k^{d^v} v_0^{(k)} W_0^{EV(k)} + \sum_i^{d^v} b_0^{V(i)} \right| \end{aligned} \quad (3.8)$$

显然, $\|v_1\|_1$ 中最大度的项为 $k_{\alpha,\beta}^1 v_0^{(i)} h_0^{(j)}$, 则有 $\sum_i^{d^v} \alpha_i = 1 = 2^{l-1}$, $\sum_i^{d^v} \beta_i = 1 = 2^{l-1}$ 。

假设 $l = n$ 时, $v_n^{(x)}$ 、 $h_n^{(y)} = k_{\alpha,\beta}^n \prod_i^{d^v} v_0^{(i)\alpha_i} \prod_j^{d^h} h_0^{(j)\beta_j}$ 分别表示 v_n 、 h_n 中的最大度项, 且有 $\sum_i^{d^v} \alpha_i = 2^{n-1}$, $\sum_i^{d^v} \beta_i = 2^{n-1}$ 。则当 $l = n + 1$ 时, 有:

$$\begin{aligned} \|v_{n+1}\|_1 = & \left| \sum_i^{d^v} \sum_j^{d^v} W_n^{0(j,i)} v_n^{(i)} \sum_k^{d^h} h_n^{(k)} W_n^{VV(k)} \right. \\ & \left. + \sum_i^{d^h} \sum_j^{d^v} W_n^{0(j,d^v+i)} h_n^{(i)} \sum_k^{d^v} v_n^{(k)} W_n^{EV(k)} + \sum_i^{d^v} b_n^{V(i)} \right| \end{aligned} \quad (3.9)$$

不难看出, $\|v_{n+1}\|_1$ 中的最大度项为 $k_{\alpha,\beta}^{n+1} v_n^{(x)} h_n^{(y)}$, 则对 v_{n+1} 有 $\sum_i^{d^v} \alpha_i = 2^{(n+1)-1}$, $\sum_i^{d^v} \beta_i = 2^{(n+1)-1}$ 。根据以上数学归纳法得证, 经过 l 阶的高阶共享单元, 物品和节点都能学习到来自对方的 $O(2^l)$ 阶交叉特征。

高阶共享单元的设计一方面有效缓解了 DCN、KTUP 等模型的特征交叉能力较低的问题, 将特征交叉从 $O(l)$ 阶提升到 $O(2^l)$ 阶, 使得物品和节点学到了更多来自对方的信息; 另一方面不同维度下的特征交叉适配了对物品和节点的差异化建模, 使得二者能在包含自身环境信息编码的基础上进行信息交互与融合。

3.2.2 多任务交替学习框架

目前大多基于知识图谱的多任务推荐方法通过依次训练或联合训练进行多任务间协同。依次训练指的是知识图谱首先通过 **embedding** 得到节点与关系向量, 然后引入推荐系统学习用户向量与物品向量, 进行模型训练, 例如 wang 等人提出 DKN 模型^[50]先将新闻图谱进行训练, 得到节点的向量表示, 再引入推荐模块进行点击率预估。但依次训练出来的节点向量表示目的是拟合图谱结构, 导致其更加适用于知识图谱的相关任务, 如链路预测等, 并不适用于推荐。

联合训练指的是同时进行知识图谱的 **embedding** 与模型的训练。这类方法的损失函数一般为：

$$Loss_{total} = Loss_{Rec} + \gamma Loss_{KGE} + \lambda ||\theta||_2^2 \quad (3.10)$$

其中 γ 控制图谱表示学习损失的权重， $||\theta||_2^2$ 为参数正则化， λ 为参数正则化权重。图谱学习任务同推荐任务联合训练主要基于物品和节点的同一表示下，任务间的依赖性导致物品和节点无法很好地针对自身环境建模。因此，本文采用交替训练的方法训练知识图谱补全任务和推荐任务：先对推荐任务进行训练，再经过固定轮次后，训练图谱补全任务。其中轮次数为一个超参数。以此迭代，直到模型收敛。

3.3 本章小结

本章对基于高阶交互单元的多任务学习框架进行了详细介绍。首先介绍了多任务学习参数共享的概念，分析目前基于知识图谱的多任务推荐方法中参数共享所存在的问题。接着，基于目前方法存在的不足，介绍了本文提出的基于高阶交互单元的多任务学习框架，并用形式化的方法详细介绍了高阶共享单元的设计，证明了其高阶特征交叉学习能力。

第四章 基于图卷积神经网络的协同信息感知

在基于知识图谱的多任务推荐方法中，推荐侧任务的设计是提高推荐效率的基础。设计高效的信息传播与聚合的结构，能够有效传播用户的深层兴趣，丰富物品的自身表示，在提升推荐预估准确率的同时，也能增强对图谱补全任务训练。本章首先介绍目前基于知识图谱的多任务推荐方法中推荐侧任务的问题，然后详细介绍基于图卷积神经网络的协同信息感知方法。

4.1 推荐侧问题描述与分析

(1) 忽略了用户物品交互中的协同信息对图谱补全的正向作用。目前基于知识图谱的多任务推荐方法大多将用户物品交互数据作为输入，构造用户的历史交互物品记录和物品的交互用户记录，再结合知识图谱进行信息传播与聚合。例如 wang 等人提出的 CKAN 模型^[9]将用户的多阶交互物品和物品的多阶相似物品作为初始种子放入知识图谱中进行传播，进而聚合用户兴趣和丰富物品表示；Chen 等人提出的 DHAN 模型^[51]在用户物品交互图同知识图谱的融合图上进行信息的传播与聚合，通过分离式表征细粒度刻画用户和物品。这些方法能够有效利用图谱结构在用户和物品双侧进行信息的传播，却割裂了用户与物品的交互关系，忽略了隐藏在用户物品交互过程中丰富的协同信息，导致模型无法有效利用这类信息去挖掘用户的深层兴趣，交互物品间的关联性也无法有效推动图谱补全，推荐效果无法进一步提升。

(2) 忽视了用户物品多阶交互中的协同信息差异。通过交互记录的高阶连接，用户可能有着多阶的交互物品，这些物品中蕴含的协同信息对用户兴趣的表征至关重要。目前的方法大多采用简单的求和或拼接方式聚合多阶交互的协同信息，例如在 NGCF^[23]构建用户-物品交互记录转置为高阶连通图，将不同阶信息直接拼接起来；LightGCN^[24]进一步简化 NGCF，放弃了特征变换和非线性激活的使用，将不同阶信息取均值。然而，随着交互的传播，传播的物品可能与初始用户完全不相关，这类方法没有考虑不同阶交互中的协同信息对于初始用户或物品不同的信息量，忽略了对高阶协同信息噪声的建模，导致模型引入过多的环境噪声，限制了对推荐物品的预测。

基于以上问题，本文提出一种基于图卷积神经网络的协同信息感知方法。它首先根

据用户-物品交互记录构建用户-物品连通图，将三元组形式转化为图结构。进一步，构建用户与物品的多阶交互邻居集合，利用图卷积神经网络挖掘用户物品连通图的结构信息，充分感知用户物品交互过程中的协同信息。同时引入门控单元，在协同信息聚合阶段自适应建模不同阶信息的通入量，降低环境噪声对表征的影响。

为了方便介绍本节的问题背景与提出的协同信息感知方法，表 2 总结了本节所有用到的数学符号及其对应描述。

表 2 基于图卷积神经网络的协同信息感知方法中定义的数学符号

符号	描述
$U = \{u_1, u_2, \dots, u_M\}$	用户集合， M 为用户数量
$V = \{v_1, v_2, \dots, v_N\}$	物品集合， N 为物品数量
$G = \{(h, t) h, t \in U \cup V, I_{(h,t)} = 1\}$	用户物品的交互集合
S_u^l, S_v^l	用户 u 、物品 v 第 l 阶的邻域节点集合
$M_{u \leftarrow v}^{(l)}$	交互物品 v 对用户 u 第 l 阶的信息贡献
$M_{u \leftarrow u}^{(l)}$	用户 u 对自身第 l 阶的信息贡献
$M_{v \leftarrow u}^{(l)}$	交互用户 u 对物品 v 第 l 阶的信息贡献
$M_{v \leftarrow v}^{(l)}$	物品 v 对自身第 l 阶的信息贡献
$e_u^{(l)}, e_v^{(l)}$	用户 u 、物品 v 第 l 阶的信息聚合向量
$LeakRelu$	$LeakRelu$ 激活函数
W_1, W_2, W_3	转置权重矩阵
$E^{(l)}$	用户与物品表示向量拼接矩阵
A	用户物品邻接矩阵

4.2 基于图卷积神经网络的协同推荐方法的提出

用户物品的交互数据可以形象地表示为一个二分图，如图 11 所示。其中左侧为用户集合 $U = \{u_1, u_2, \dots, u_M\}$ ，右侧为物品集合 $V = \{v_1, v_2, \dots, v_N\}$ ：

传统基于协同过滤的推荐算法常从用户-物品交互矩阵中提取相关特征，虽然这类方法简单直观，但实际应用的交互规模往往达到数百万甚至更大，难以构建交互矩阵，导致很难提炼出想要的协同信息。

因此，本文将用户物品交互二分图转置成用户物品交互连通图，如图 12 所示，利用来自用户物品交互的高阶连通性作为图卷积神经网络传播与聚合协同信息的基础。以用户 u_1 为例，图 11 表示用户物品的交互二分图，图 12 为转置的用户物品连通图。高阶连通性表示从路径长度大于 1 的任意节点到达 u_1 的路径。这种高阶连通性承载了协同信息的丰富语义。例如，路径 $u_1 \leftarrow v_2 \leftarrow u_2$ 表示 u_1 和 u_2 之间的行为相似性，因

为两个用户都与 v_2 进行了交互；较长的路径 $u_1 \leftarrow v_2 \leftarrow u_2 \leftarrow v_4$ 表明 u_1 很可能采用 v_4 ，因为它的相似用户 u_2 之前已经消费过 v_4 。此外，从整体来看，相比物品 v_5 ， u_1 更可能对物品 v_4 感兴趣，因为有两条高阶通路链接 u_1 和 v_4 。

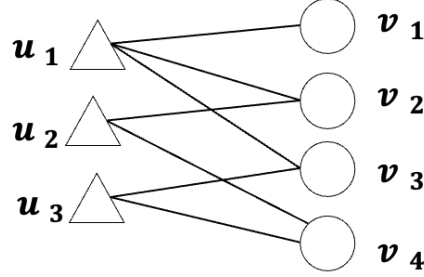


图 11 用户物品交互二分图

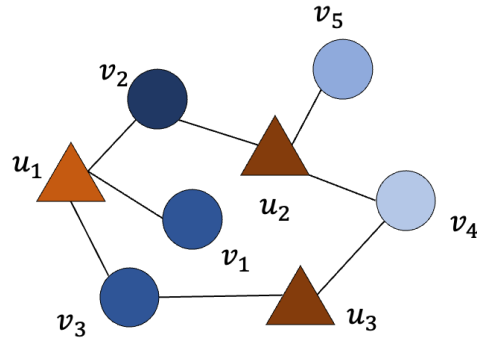


图 12 用户物品连通图

4.2.1 协同节点传播

用户物品连通图由用户物品以及表示交互的连接组成，它可以表示为 $G = \{(h, t) | h, t \in U \cup V, I_{(h, t)} = 1\}$ ，其中节点 h 和 t 分别属于用户集合和物品集合， I 为交互函数， $I = 1$ 表示当前物品和用户交互过。协同节点传播是从特定节点出发，沿着用户物品连通图中的连接传播，得到不同距离的邻域节点集合。

从用户物品连通图中节点出发，沿着连接可以传播得到的第 l 阶的邻域节点集合，它表示为：

$$\begin{aligned} S_u^l &= \{t | (h, t) \in G \text{ and } h \in S_u^{l-1}\} \\ S_v^l &= \{t | (h, t) \in G \text{ and } h \in S_v^{l-1}\} \end{aligned} \quad (4.1)$$

其中， l 表示传播的距离。 S_u^{l-1} 、 S_v^{l-1} 分别表示用户和物品第 $l-1$ 阶的邻域节点集合， S_u^0 、 S_v^0 表示用户、物品节点本身。

4.2.2 协同消息构造

直观来说，交互物品能够很好的表示用户偏好；类似地，消费过某物品的用户也可表现该物品的某些特征。因此，本文基于用户物品连通图的高阶连通性，利用图卷积神经网络构造交互节点间的信息贡献来丰富彼此的表示。对于用户 u ，定义其交互物品对自身的信息贡献为：

$$M_{u \leftarrow v} = \frac{1}{\sqrt{|S_u^1| |S_v^1|}} (W_1 e_v + W_2 (e_v \odot e_u)) \quad (4.2)$$

其中 $W_1, W_2 \in R^{d \times d}$ 是可训练的权重矩阵，用于提取有用的传播信息， d 为特征维度。 $|S_u^1|$ 、 $|S_v^1|$ 分别表示与用户 u 和物品 i 直接有交互的物品或者用户数量，即为用户物品连通图中用户 u 和物品 i 的度。 $\frac{1}{\sqrt{|S_u^1| |S_v^1|}}$ 为图拉普拉斯范数（graph Laplacian norm）^[49]，它表示物品传递给用户信息的权重。直观地，如果某个物品的度越小，表示这个物品越“独特”，则该权重越大，越能够体现用户的个性偏好；用户的度越小，表示该用户的兴趣越“集中”，那么他的历史数据中的每个物品的权重都应该增大，表示每个物品都能够较大的反映该用户偏好。同时，与仅考虑 e_v 贡献的传统图卷积网络不同，本文不仅考虑物品自身的特征，还考虑了物品对于当前用户的依赖度，将 e_v 和 e_u 之间的交互编码为通过 $e_v \odot e_u$ 传递的消息，其中 \odot 表示元素积乘法，使得相似的物品能够向中心用户传递更多信息。

同时，为了保留节点自身的特征，本文还考虑节点自身信息的自连接，定义为：

$$M_{u \leftarrow u} = W_1 e_u \quad (4.3)$$

其中 W_1 为(4.2)式中相同的可训练权重矩阵。同理，对于物品节点 v ，其交互用户对自身的的信息贡献为：

$$M_{v \leftarrow u} = \frac{1}{\sqrt{|S_u^1| |S_v^1|}} (W_1 e_u + W_2 (e_v \odot e_u)) \quad (4.4)$$

物品节点 v 自身信息的自连接为：

$$M_{v \leftarrow v} = W_1 e_v \quad (4.5)$$

4.2.3 协同消息聚合

对于一个用户节点 u ，经过消息构造得到了来自多个交互物品的协同信息贡献和自连接信息贡献，为了综合考虑，本文定义一阶协同信息的聚合方式为：

$$e_u^{(1)} = LeakRelu \left(M_{u \leftarrow u} + \sum_{v \in S_u^1} M_{u \leftarrow v} \right) \quad (4.6)$$

$e_u^{(1)}$ 是用户 u 经过第一次聚合得到最终表示向量。为了有效学习用户物品连通图中可能的负样本偏差，采用 $LeakRelu$ 作为激活函数。

同样的，对于物品节点 v ，其一阶信息的聚合方式为：

$$e_v^{(1)} = LeakRelu \left(M_{v \leftarrow v} + \sum_{u \in S_v^1} M_{v \leftarrow u} \right) \quad (4.7)$$

在用户物品连通图中，一阶协同信息聚合可以有效增强用户和物品的表示。根据连通图的高阶连通性，本文将协同信息的传播与聚合拓展到多阶，使得用户和物品能够接收来自高阶交互的协同消息。对于用户节点 u ，定义第 l 阶信息聚合为：

$$e_u^{(l)} = LeakRelu \left(M_{u \leftarrow u}^{(l)} + \sum_{i \in S_u^1} M_{u \leftarrow v}^{(l)} \right) \quad (4.8)$$

$M_{u \leftarrow u}^{(l)}$ 、 $M_{u \leftarrow v}^{(l)}$ 分别表示用户 u 第 l 阶自连接信息贡献和交互物品信息贡献，分别定义为：

$$M_{u \leftarrow v}^{(l)} = \frac{1}{\sqrt{|S_u^1| |S_v^1|}} \left(W_1^{(l)} e_v^{(l-1)} + W_2^{(l)} (e_v^{(l-1)} \odot e_u^{(l-1)}) \right) \quad (4.9)$$

$$M_{u \leftarrow u}^{(l)} = W_1^{(l)} e_u^{(l-1)}$$

其中 $W_1^{(l)}$ 、 $W_2^{(l)} \in R^{d \times d}$ 为可训练的权重矩阵。 $e_u^{(l-1)}$ 、 $e_i^{(l-1)}$ 分别为用户和物品节点第 $l-1$ 阶聚合信息后的表示。

同理，对于物品节点 v ，其第 l 阶信息聚合定义为：

$$e_v^{(l)} = LeakRelu \left(M_{v \leftarrow v}^{(l)} + \sum_{u \in S_v^1} M_{v \leftarrow u}^{(l)} \right) \quad (4.10)$$

$$M_{v \leftarrow u}^{(l)} = \frac{1}{\sqrt{|S_u^1| |S_v^1|}} \left(W_1^{(l)} e_u^{(l-1)} + W_2^{(l)} (e_v^{(l-1)} \odot e_u^{(l-1)}) \right)$$

$$M_{v \leftarrow v}^{(l)} = W_1^{(l)} e_v^{(l-1)}$$

为了方便张量运算，多阶的信息聚合可以改写为矩阵形式：

$$E^{(l)} = \text{LeakyRelu} \left((L + T)E^{(l-1)}W_1^{(l)} + LE^{(l-1)} \odot E^{(l-1)}W_2^{(l)} \right) \quad (4.11)$$

其中 $E^{(l)} \in R^{(N+M) \times d}$ 表示用户和物品经过 l 阶信息聚合后的表示拼接张量。 E^0 表示用户和物品初始表示拼接张量。 T 表示用户物品的 ID 映射矩阵。 L 表示用户物品连通图上的对称归一化拉普拉斯矩阵，定义为：

$$L = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \quad (4.12)$$

其中 A 为用户物品邻接矩阵，定义为：

$$A = \begin{bmatrix} \mathbf{0} & I \\ I^T & \mathbf{0} \end{bmatrix} \quad (4.13)$$

其中 $I \in R^{N \times M}$ 表示用户物品的交互矩阵， $\mathbf{0}$ 表示全零矩阵， D 表示对角度矩阵。

经过高阶传播与聚合后，分别得到用户和物品的多阶协同信息表示 $\{e^{(1)}, \dots, e^{(l)}\}$ 。随着交互的传播，用户或物品的交互邻居越来越多，传统方法采用求和或拼接的操作会引入过多的环境噪声，因此聚合后的嵌入表示是次优的。本文从 LSTM^[50]、GRU^[51] 等序列模型中的门控单元得到启发，在多阶协同信息聚合时引入门控单元，自适应控制不同阶协同信息的通入量，降低环境噪声。用户或物品的最终聚合表示为：

$$e^* = e^{(0)} + \sum_{i=1}^l \alpha_i e^{(i)} \quad (4.14)$$

其中 α_i 是门控单元的输出值，表示为：

$$\alpha_i = \sigma(W_3(e^{(0)} || e^{(i)}) + b) \quad (4.15)$$

其中 $W_3 \in R^{d \times d}$ 为可训练的权重矩阵， $e^{(0)}$ 用户或物品的初始表示。

4.3 本章小结

本章详细介绍了本文提出的基于图卷积神经网络的协同信息感知方法。首先指出了目前基于知识图谱的多任务推荐方法中推荐侧任务存在的问题。然后详细介绍了利用图卷积神经网络对多阶协同信息的构造与聚合，并阐述了基于门控单元对高阶协同信息通入量的建模，自适应地控制噪声。

第五章 基于多层注意力机制的图谱补全

在基于知识图谱的推荐方法中，图谱的补全是提升图谱完备性，辅助推荐效果的关键研究之一。本章首先分析基于知识图谱的多任务推荐方法中图谱补全的问题，然后详细介绍基于多层注意力机制的图谱补全方法。

5.1 图谱侧问题描述与分析

基于知识图谱的推荐方法往往将用户交互过的物品作为初始种子，根据图谱的中心-邻域结构，拓展其多阶邻居，构造节点间的关联信息，并设计有效的信息传播与聚合来充分利用图谱的邻域信息，丰富节点的自身表示，进行图谱的表示学习。目前基于知识图谱的推荐算法，在图谱学习时普遍存在以下问题：

(1) 同阶邻域信息聚合时忽视了对节点间关系的建模。图谱中的关系可以很好地表示多样的用户兴趣，不同的兴趣对交互的影响程度各不相同。但目前的方法在处理同一阶邻域嵌入集合中各节点的信息时，往往会弱化甚至忽视对节点间不同关系的建模，一方面会导致图谱补全时无法区分同一阶邻域内，不同关系下的尾节点对中心节点传递信息的差异；另一方面，有效建模关系能够提升图谱补全任务对推荐任务的正向推动作用，提高推荐效果的可解释性。

(2) 异阶邻域信息聚合时忽略了长距离信息传播的衰减现象。经过高阶传播，节点的邻居集合极大地拓展，远距离的邻居节点和关系与中心节点可能千差万别，对推荐任务来说甚至会噪声信息。目前的方法在异阶信息的聚合阶段，没有将高阶的邻域信息衰减考虑到算法建模中。例如在 CKAN^[9]算法中，在多阶信息聚合时直接使用求和操作，没有考虑邻域信息在高阶传播后的衰减现象，使得不同阶邻域信息平均地影响中心节点的信息通入量，导致模型对领域信息传播的阶数有很大限制，因此 CKAN 无法处理远距离的噪声信息。

基于上述问题，本文利用分层注意力机制改进了传统方法中同阶邻域信息和异阶邻域信息的聚合方式。在同阶邻域信息聚合阶段，该方法首先根据中心节点、尾节点和关联关系计算尾节点在当前关系下对中心节点的影响力因素，进而聚合多个尾节点得到当前阶邻域的聚合信息；在不同阶邻域信息聚合阶段，该方法利用注意力机制控制多阶邻

域的聚合信息在传播时的衰减，聚合得到多阶邻域传播到中心节点的信息。相比于传统方法，多层注意力机制的设计能够充分发挥图谱中关系对节点信息聚合的作用，高效地实现邻域信息聚合，同时有效抑制信息在长距离传播中引入的噪声。

本节使用数学符号对本文要解决的问题进行了定义，表 3 给出了本节所使用到的数学符号。

表 3 基于分层注意力机制的图谱补全方法中定义的数学符号

符号	描述
$S_{h_0}^l$	节点 h_0 第 l 阶的邻接节点集合
$T_{h_0}^l$	节点 h_0 第 l 阶的三元组集合
$G = \{(h, r, t) h, t \in E, r \in R\}$	知识图谱中的三元组集合
$e_i^h, e_i^t \in E$	第 i 个头/尾节点的向量表示
$e_i^r \in R$	第 i 个关系的向量表示
$\pi(e_i^h, e_i^t)$	第 i 个三元组的注意力权值
W_0, W_1, W_2	权重参数矩阵
b_0, b_1, b_2	偏置参数矩阵
$ReLU$	ReLU 激活函数
σ	Sigmoid 激活函数
$e_{h_0}^{(l)}$	节点 h_0 第 l 阶的聚合向量表示
L_{h_0}	节点 h_0 第 l 阶的聚合向量表示集合

5.2 基于多层注意力机制的图谱补全方法的提出

5.2.1 同阶邻域信息的传播与聚合

知识图谱可以形式化的表示为三元组集合 $G = \{(h, r, t) | h, t \in E, r \in R\}$ ，其中节点 h 和 t 分别称为头节点和尾节点，它们共同组成节点集合 E ， r 表示关系，所有关系组成关系集合 R 。

在知识图谱中从节点 h_0 出发，沿着已有关系可以得到的第 l 阶的邻域节点集合 $S_{h_0}^l$ ，表示为：

$$S_{h_0}^l = \{t | (h, r, t) \in G \text{ and } h \in S_{h_0}^{l-1}\} \quad (5.1)$$

其中， l 表示传播的阶数， $S_{h_0}^{l-1}$ 是节点 h_0 第 $l-1$ 阶的邻域节点集合， $S_{h_0}^0$ 表示节点 h_0 本身。

同理可以得到节点 h_0 的第 l 阶邻域三元组集合 $T_{h_0}^l$ ，表示为：

$$T_{h_0}^l = \{(h, r, t) | (h, r, t) \in G \text{ and } h \in S_{h_0}^{l-1}\} \quad (5.2)$$

邻域信息在同阶进行聚合时，本文借用注意力机制，如图 13 所示，首先将头节点向量和关系向量输入多层感知机，计算出注意力权重，从而考虑不同关系下的尾节点对中心节点信息传播的差异；再结合注意力权重和对应尾节点向量，得到当前尾节点对中心节点传递的信息，最后通过聚合多个尾节点的信息向量得到当前阶的信息聚合向量。

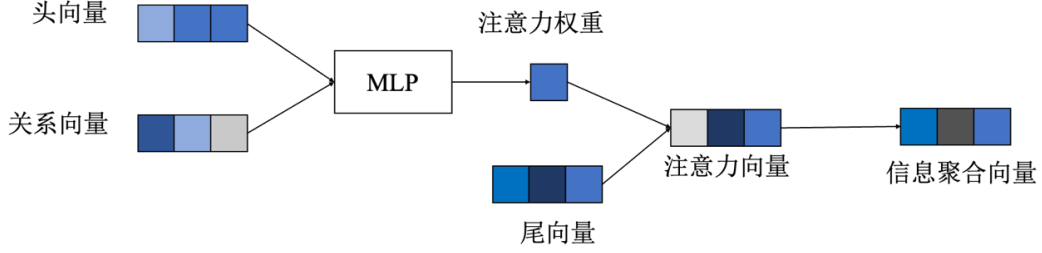


图 13 层内注意力机制结构示意图^[51]

具体地，对于节点 h_0 的第 l 阶邻域三元组集合 $T_{h_0}^l$ ，其中第 i 个三元组的头节点向量表示为 e_i^h ，尾节点向量表示为 e_i^t ，关系向量表示为 e_i^r ，经过一个两层的感知机，得到注意力权重。此过程用公式表示为：

$$\begin{aligned} z_0 &= \text{ReLU}(W_0(e_i^h || e_i^r) + b_0) \\ z_1 &= \text{ReLU}(W_1 z_0 + b_1) \\ \pi'(e_i^h, e_i^r) &= \sigma(W_2 z_1 + b_2) \end{aligned} \quad (5.3)$$

为了防止某个尾节点计算的注意力权重过大或者过小，在训练过程中造成梯度爆炸或梯度消失的问题，本文对得到的所有注意力权重通过 SOFTMAX 函数进行归一化。具体地，对于节点 h_0 的第 l 阶邻域而言，第 i 个三元组得到的注意力权重表示为：

$$\pi(e_i^h, e_i^r) = \frac{\exp(\pi'(e_i^h, e_i^r))}{\sum_{(h', r', t') \in S^l} \exp(\pi'(e_i^{h'}, e_i^{r'}))} \quad (5.4)$$

尾节点向量与对应的注意力权重进行相乘，就得到了该尾节点对中心节点传递的信息。将同阶所有三元组中的信息进行聚合，得到节点 h_0 的第 l 阶邻域聚合信息 $e^{(l)}$ ，表示为：

$$e^{(l)}_{h_0} = \sum_{i=1}^{|S_{h_0}^l|} \pi(e_i^h, e_i^r) e_i^t \quad (5.5)$$

利用注意力机制对不同阶邻域分别进行聚合，得到了节点 h 多阶邻域的信息聚合向量集合 L_{h_0} ，表示为：

$$L_{h_0} = \{e^{(1)}_{h_0}, e^{(2)}_{h_0}, \dots, e^{(l)}_{h_0}\} \quad (5.6)$$

5.2.2 异阶邻域信息的传播与聚合

得到节点 h_0 多阶邻域的信息聚合向量集合后，需要为其有效聚合来自不同阶的信息。传统的均值或者求和聚合，没有考虑长距离信息传播的衰减现象，本文利用注意力机制改进异阶邻域信息的聚合，控制多阶邻域的聚合信息在传播时的衰减。具体地，首先拼接节点 h_0 的初始表示和第 l 阶聚合后的信息聚合向量作为输入，输出注意力 $\tau^{(l)}$ ，表示为：

$$\tau^{(l)} = \sigma(W^{(l)}(e^{(0)}_{h_0} || e^{(l)}_{h_0}) + b_3) \quad (5.7)$$

其中 $e^{(0)}_{h_0}$ 为节点的初始向量表示。同样，为了防止训练过程中出现梯度爆炸或者梯度消失问题，采用 SOFTMAX 函数对注意力权重进行归一化。对节点 h_0 来说，所能得到的归一化注意力权重输出为：

$$\tau^{(l)} = \frac{\exp(\tau^{(l)})}{\sum_{\{\tau^{(n)} | n \in \{1, 2, \dots, N\}\}} \exp(\tau^{(n)})} \quad (5.8)$$

其中 N 为传播的最远距离。计算出各阶的注意力权重后，需要将不同阶的信息聚合向量进行异阶聚合，建模信息在长距离下的衰减。因此，节点 h_0 的的异阶邻域信息聚合向量最终表示为：

$$e = \sum_{(e^{(i)}_{h_0} \in L_{h_0})} e^{(i)}_{h_0} \tau^{(i)} \quad (5.9)$$

5.3 本章小结

本章对基于多层注意力机制的图谱补全方法进行了详细介绍。首先分析了现有基于知识图谱的推荐方法中同阶和异阶邻域信息传播与聚合所存在的问题。接着，基于目前方法存在的不足，介绍了本文提出的基于多层注意力机制的图谱补全方法，并用形式化的方法详细介绍了对同阶邻域信息重要性的建模和异阶邻域信息长距离衰减的建模。

第六章 基于知识图谱的推荐与补全联合网络

在基于知识图谱的多任务推荐方法研究中，推荐侧协同信息的捕捉与图谱补全是关键核心，并通过共享单元进行高效的信息交互。有了以上设计之后，需要提出一个完整的推荐算法，形成端到端的模型，同时需要定义模型的训练、更新的方式，以充分发挥多任务机制下，推荐任务与图谱补全任务联合训练的优势。本章首先用形式化的方式对问题进行描述，然后详细介绍基于知识图谱的推荐与补全联合网络，并介绍模型训练的方式。

6.1 形式化定义

推荐系统的主要对象是用户和物品。假设共有 M 位用户， N 个物品，那么用户集合可以表示为 $U = \{u_1, u_2, \dots, u_M\}$ ，物品集合表示为 $V = \{v_1, v_2, \dots, v_N\}$ 。用户与物品的历史交互表示 $G_1 = \{(u, y_{uv}, v) | u \in U, v \in V\}$ ，其中 y_{uv} 表示交互函数，这种交互不是如评分之类的显式反馈数据，而是如查看、购买等隐式反馈数据^[51]，如果用户 u 与物品 v 之间存在交互，则 $y_{uv} = 1$ ，否则为 $y_{uv} = 0$ 。

知识图谱可以形式化地表示为三元组集合 $G_2 = \{(h, r, t) | h, t \in E, r \in R\}$ ，将节点 h 和 t 分别称为头节点和尾节点，它们共同组成节点集合 E ， r 表示关系，所有关系组成关系集合 R 。推荐系统中的物品往往与知识图谱中的一个或多个节点具有匹配关系，它们组成集合 $A = \{(v, h) | v \in V, h \in E\}$ 。

为了更好地描述基于知识图谱的推荐与补全联合网络模型，详细介绍其内部方法，表4列出了本章用到的数学符号及其对应描述。

6.2 基于知识图谱的推荐与补全联合网络的提出

6.2.1 模型架构

针对基于知识图谱的推荐方法现存问题，本文提出名为基于知识图谱的推荐与补全联合网络 MRKC (Multi-task Recommendation based on Knowledge Graph Completion)。MRKC 的结构如图 14 所示，分为图表示层、信息传播与聚合层、信息共享层和训练层。

表 4 基于知识图谱的推荐与补全联合网络中定义的数学符号

符号	描述
$U = \{u_1, u_2, \dots, u_M\}$	用户集合, M 为用户个数
$V = \{v_1, v_2, \dots, v_N\}$	物品集合, N 为物品个数
$G_1 = \{(u, y_{uv}, v) u \in U, v \in V\}$	用户-物品交互二分图
$G_2 = \{(h, r, t) h, t \in E, r \in R\}$	知识图谱三元组集合
$A = \{(v, e) v \in V, e \in E\}$	交互二分图与外部信息网络中节点的匹配集合
h, r, t	头节点、关系和尾节点向量表示
$W_{0 \sim 5}, W_l^0, W_l^1$	转置矩阵
$W_l^{VV}, W_l^{VE}, W_l^{EV}, W_l^{EE}$	压缩矩阵
v_l, h_l	物品 v 、节点 h 的 l 阶表示
d_v	物品表示的维度
d_h	节点表示的维度
S_u^l, S_v^l	用户 u 、物品 v 第 l 阶的节点集合
$M_{u \leftarrow v}^{(l)}$	交互物品 v 对用户 u 第 l 阶的信息贡献
$M_{u \leftarrow u}^{(l)}$	用户 u 对自身第 l 阶的信息贡献
$M_{v \leftarrow u}^{(l)}$	交互用户 u 对物品 v 第 l 阶的信息贡献
$M_{v \leftarrow v}^{(l)}$	物品 v 对自身第 l 阶的信息贡献
$e_u^{(l)}, e_v^{(l)}$	用户 u 、物品 v 第 l 阶的信息聚合向量
$E^{(l)}$	用户与物品第 l 阶表示向量矩阵
A	用户物品邻接矩阵
$S_{h_0}^l$	节点 h_0 第 l 阶的邻接节点集合
$T_{h_0}^l$	节点 h_0 第 l 阶的三元组集合
$e_{h_0}^{(l)}$	节点 h_0 第 l 阶的聚合向量表示
L_{h_0}	节点 h_0 聚合向量表示集合
$e_i^h, e_i^t \in E$	第 i 个头/尾节点的向量表示
$e_i^r \in R$	第 i 个关系的向量表示
$\pi(e_i^h, e_i^r)$	第 i 个三元组的注意力权值
b	偏置参数矩阵
$ReLU$	ReLU 激活函数
σ	Sigmoid 激活函数
$LeakRelu$	LeakRelu激活函数

图表示层中, 首先将用户物品二分图转置成用户物品连通图, 同知识图谱一起, 将所有节点用低维向量进行表示, 便于后续计算, 其中物品与对应节点分开编码。信息传播与聚合层中, 在高阶连通图上通过图卷积神经网络捕捉用户与物品的协同信息, 聚合用户与物品表示; 在知识图谱中通过分层注意力机制聚合经过高阶传播得到的不同阶邻居节点-关系集合, 更新目标节点的表示。信息共享层中交叉学习物品和对应节点的高阶交互特征, 丰富二者的向量表示。训练层一方面将训练得到的用户物品表示经过多层感

知机输出用户与物品交互的概率值；另一方面将训练得到的节点和关系表示进行打分操作。

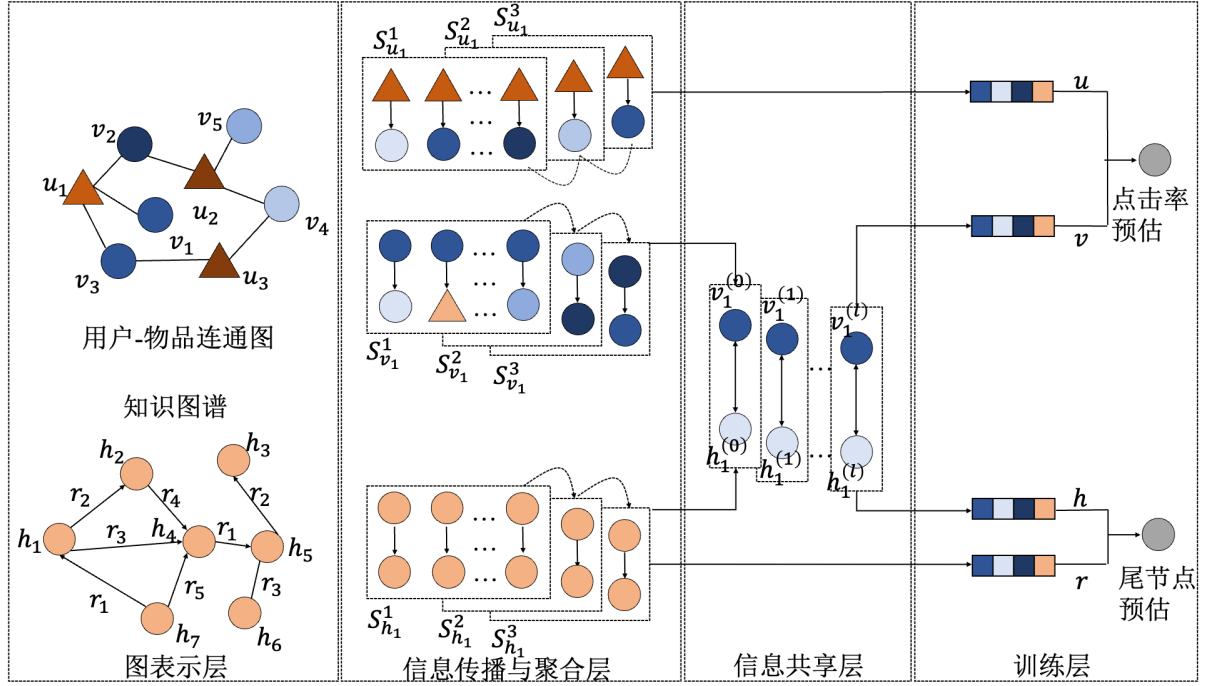


图 14 MRKC 网络结构示意图

6.2.2 图表示层

图表示层将用户物品交互二分图转置成用户物品连通图，利用来自用户项交互的高阶连通性作为图卷积神经网络传播与聚合信息的基础。用户物品连通图可以表示为 $G_1 = \{(h, t) | h, t \in U \cup V, I_{(h, t)} = 1\}$ ，其中节点 h 和 t 分别属于用户集合和物品集合， I 为交互函数， $I = 1$ 表示当前物品和用户交互过。知识图谱则用通用的三元组形式表示，即 $G_2 = \{(h, r, t) | h, t \in E, r \in R\}$ 。考虑物品和节点是同一个对象在不同场景下的两种描述，为了使其向量表示包含自身环境特征，对物品和节点分别编码，初始化物品向量表示 $v \in R^{d^v}$ ，节点向量表示 $h \in R^{d^h}$ ，其中 d^v 、 d^h 表示物品和节点的向量维度，为预先定义的超参数。

6.2.3 信息传播与聚合层

信息传播与聚合层包括推荐侧用户和物品的协作信号捕捉与知识图谱侧的图谱补全。

(1) 推荐侧

从用户物品连通图中节点出发，沿着连接可以传播得到的第 l 阶的邻域节点集合，它表示为：

$$\begin{aligned} S_u^l &= \{t | (h, t) \in G \text{ and } h \in S_u^{l-1}\} \\ S_v^l &= \{t | (h, t) \in G \text{ and } h \in S_v^{l-1}\} \end{aligned} \quad (6.1)$$

对于用户 u ，定义其交互物品对自身的信息贡献为：

$$M_{u \leftarrow v} = \frac{1}{\sqrt{|S_u^1| |S_v^1|}} (W_1 e_v + W_2 (e_v \odot e_u)) \quad (6.2)$$

同时，为了保留节点自身的特征，还考虑节点自身信息的自连接：

$$M_{u \leftarrow u} = W_1 e_u \quad (6.3)$$

其中 W_1 为式中相同的可训练权重矩阵。同理，对于物品节点 v ，其交互用户对自身的信息贡献为：

$$M_{v \leftarrow u} = \frac{1}{\sqrt{|S_u^1| |S_v^1|}} (W_1 e_u + W_2 (e_v \odot e_u)) \quad (6.4)$$

物品节点 v 自身信息的自连接为：

$$M_{v \leftarrow v} = W_1 e_v \quad (6.5)$$

对于用户节点 u ，定义第 l 阶信息聚合为：

$$e_u^{(l)} = LeakRelu \left(M_{u \leftarrow u}^{(l)} + \sum_{v \in S_u^1} M_{u \leftarrow v}^{(l)} \right) \quad (6.6)$$

$M_{u \leftarrow u}^{(l)}$ 、 $M_{u \leftarrow v}^{(l)}$ 分别表示用户 u 第 l 阶自连接信息贡献和交互物品信息贡献，分别定义为：

$$M_{u \leftarrow v}^{(l)} = \frac{1}{\sqrt{|S_u^1| |S_v^1|}} \left(W_1^{(l)} e_v^{(l-1)} + W_2^{(l)} (e_v^{(l-1)} \odot e_u^{(l-1)}) \right) \quad (6.7)$$

$$M_{u \leftarrow u}^{(l)} = W_1^{(l)} e_u^{(l-1)}$$

同理，对于物品节点 v ，其第 l 阶信息聚合定义为：

$$e_v^{(l)} = LeakRelu \left(M_{v \leftarrow v}^{(l)} + \sum_{u \in S_v^1} M_{v \leftarrow u}^{(l)} \right) \quad (6.8)$$

$$M_{v \leftarrow u}^{(l)} = \frac{1}{\sqrt{|S_u^1| |S_v^1|}} \left(W_1^{(l)} e_u^{(l-1)} + W_2^{(l)} (e_v^{(l-1)} \odot e_u^{(l-1)}) \right)$$

$$M_{v \leftarrow v}^{(l)} = W_1^{(l)} e_v^{(l-1)}$$

用户或物品的最终聚合表示为:

$$e^* = e^{(0)} + \sum_{i=1}^l \alpha_i e^{(i)} \quad (6.9)$$

其中 α_j 是门控单元的输出值, 控制信息的通入量, 表示为:

$$\alpha_i = \sigma(W_3(e^{(0)} || e^{(i)}) + b_0) \quad (6.10)$$

其中 $W_3 \in R^{d \times d}$ 为可训练的权重矩阵, $e^{(0)}$ 为用户或物品的初始表示。

(2) 图谱侧

在知识图谱中, 从节点 h_0 出发, 沿着关系可以传播得到第 l 阶的邻域节点集合, 它表示为:

$$S_{h_0}^l = \{t | (h, r, t) \in G \text{ and } h \in S_{h_0}^{l-1}\} \quad (6.11)$$

同理, 传播得到节点 h_0 第 l 阶的邻域三元组集合表示为:

$$T_{h_0}^l = \{(h, r, t) | (h, r, t) \in G \text{ and } h \in S_{h_0}^{l-1}\} \quad (6.12)$$

对于节点 h_0 的第 l 阶邻域而言, 第 i 个三元组得到的注意力权重表示为:

$$\begin{aligned} z_0 &= ReLU(W_0(e_i^h || e_i^r) + b_1) \\ z_1 &= ReLU(W_4 z_0 + b_2) \\ \pi'(e_i^h, e_i^r) &= \sigma(W_5 z_1 + b_3) \\ \pi(e_i^h, e_i^r) &= \frac{\exp(\pi'(e_i^h, e_i^r))}{\sum_{(h', r', t') \in S^l} \exp(\pi'(e_i^h, e_i^r))} \end{aligned} \quad (6.13)$$

将同阶所有三元组中的信息进行聚合, 得到节点 h 的第 l 阶邻域聚合信息 $e_{h_0}^{(l)}$, 表示为:

$$e_{h_0}^{(l)} = \sum_{i=1}^{|S^l|} \pi(e_i^h, e_i^r) e_i^t \quad (6.14)$$

利用注意力机制对不同阶邻域分别进行聚合, 得到了节点 h 多阶邻域的信息聚合向量集合 L_{h_0} , 表示为:

$$L_{h_0} = \{e_{h_0}^{(1)}, e_{h_0}^{(2)}, \dots, e_{h_0}^{(l)}\} \quad (6.15)$$

对节点 h 来说, 所能得到的归一化注意力权重输出为:

$$\begin{aligned}\tau'^{(l)} &= \sigma \left(W^{(l)} \left(e_{h_0}^{(0)} || e_{h_0}^{(l)} \right) + b_4 \right) \\ \tau^{(l)} &= \frac{\exp \left(\tau'^{(l)} \right)}{\sum_{\{\tau'^{(n)} | n \in \{1, 2, \dots, N\}\}} \exp \left(\tau'^{(n)} \right)}\end{aligned}\quad (6.16)$$

节点 h 的异阶邻域信息聚合向量最终表示为:

$$e = \sum_{(e_{h_0}^{(i)} \in L_{h_0})} e_{h_0}^{(i)} \tau^{(i)} \quad (6.17)$$

6.2.4 信息共享层

经过信息传播与聚合层, 物品 v 和对应节点 h 的嵌入表示都得到了更新。由于它们融合了各自邻域的辅助信息, 向量表示能力得到了增强。为了提高知识图谱对推荐任务的辅助作用, 同时将用户与物品间的协作信号传入知识图谱, 物品与节点一般需要进行多阶的特征交叉学习。以第 l 阶为例首先构造特征交互矩阵 $C_l \in R^{d^v \times d^h}$, 定义为:

$$C_l = v_l h_l^T = \begin{bmatrix} v_l^{(1)} h_l^{(1)} & \dots & v_l^{(1)} h_l^{(d^h)} \\ \vdots & \ddots & \vdots \\ v_l^{(d^v)} h_l^{(1)} & \dots & v_l^{(d^v)} h_l^{(d^h)} \end{bmatrix} \quad (6.18)$$

为了下一阶的特征交叉, 再将高阶交叉特征映射到物品与节点对应的向量空间下, 定义下一阶物品与节点表示为:

$$\begin{aligned}v_{l+1} &= W_l^0 (v_l h_l^T W_l^{VV} || h_l v_l^T W_l^{EV}) + b_l^V \\ h_{l+1} &= W_l^1 (v_l h_l^T W_l^{VE} || h_l v_l^T W_l^{EE}) + b_l^E\end{aligned}\quad (6.19)$$

经过多阶交叉学习后可以得到最终的物品与对应节点表示。

6.2.5 训练层

经过信息共享层, 物品 v 和节点 h 的嵌入表示都得到了更新, 二者都携带来自对方的高阶特征信息, 向量表示能力进一步增强。

在推荐侧, 经过信息传播与聚合层后, 用户的嵌入表示分别记为 e_u ; 经过信息共享层后, 物品 v 的嵌入表示记为 e_v , 组合得到多层感知机的输入 x :

$$x = e_u || e_v \quad (6.20)$$

经过多层感知机拟合, 并以 Sigmoid 函数作为激活函数, 最后得到预测交互概率值 \hat{y}_{uv} :

$$\hat{y}_{uv} = \sigma(MLP(x)) \quad (6.21)$$

在知识图谱补全侧, 作为辅助任务, 考虑到模型复杂度要求, 本文选用较为简单

得分函数:

$$score(h, r, t) = \sigma(t^T t') \quad (6.22)$$

其中 t' 是尾节点的预测向量表示, t 为尾节点的真实向量表示。

6.3 模型训练

本文是对推荐系统下的隐式反馈进行建模, 隐式反馈的标签为 0 或 1。交叉熵损失函数是二分类模型训练一般采用的损失函数, 在推荐点击率预估中, 也采用了该函数, 其公式表示为:

$$L_{Rec} = \sum_{u \in U} \left(\sum_{v \in V} (-y_{uv} \log \hat{y}_{uv} - (1 - y_{uv}) \log(1 - \hat{y}_{uv})) \right) + \lambda \|\Theta\|_2^2 \quad (6.23)$$

其中, Θ 为 MRKC 模型中所有参数, $\|\cdot\|_2^2$ 为参数的 L2 正则项表示, λ 为正则化系数, 为一个超参数。在知识图谱补全任务中, 使用贝叶斯个性排序 (Bayesian Personalized Ranking, BPR) 损失函数^[52], 用公式表示为:

$$L_{KGC} = \sum_{(h, r, t, t') \in G} -\ln \sigma(g(h, r, t') - g(h, r, t)) \quad (6.24)$$

因此模型最终的损失函数表示如下:

$$\begin{aligned} L_{total} = & \sum_{u \in U} \left(\sum_{v \in V} (-y_{uv} \log \hat{y}_{uv} - (1 - y_{uv}) \log(1 - \hat{y}_{uv})) \right) \\ & + \sum_{(h, r, t, t') \in G} -\ln \sigma(g(h, r, t') - g(h, r, t)) + \lambda \|\Theta\|_2^2 \end{aligned} \quad (6.25)$$

MRKC 在模型训练时, 采用迭代训练的方式: 先对推荐任务进行训练, 再经过固定轮次后, 训练图谱补全任务。其中轮次数为一个超参数。以此迭代, 直到模型收敛。

6.4 本章小结

本章详细介绍了本文所提出的基于知识图谱的推荐与补全联合网络 MRKC。首先, 形式化地描述了该算法所要解决的问题场景, 即基于知识图谱的推荐场景。然后, 对 MRKC 模型结构进行分层地详细介绍, 推导了图表示层、信息传播与聚合层、信息共享层和训练层。最后, 介绍了模型训练使用的损失函数以及训练方式。

第七章 实验与验证

为了验证 MRKC 算法的有效性,本章进行实验验证,首先详细介绍实验的设计方法,包括用于实验的数据集、实验环境、对比算法和评估指标,然后对实验结果进行分析,并对算法的超参数进行探索分析。

7.1 实验方案设计

7.1.1 实验数据集

由于本文的背景是科技大数据推荐场景,因此特别选择了论文推荐场景的相关数据集。通用场景下选择了书籍、电影和音乐三个不同的推荐场景。下面是对各个数据集的具体介绍。

(1) article 数据集

article 为自建数据集,主要基于 CiteULike-a^[56]数据。CiteULike 是一个帮助学者管理文献的标签工具,它可以自动抽取文献的作者、出版社、会议名、页码等详细信息。article 数据集包含了约 5500 位学术工作者对 11700 余条文献的收藏行为数据。

(2) paper 数据集

paper 为自建数据集,主要基于 CiteULike-t^[56]数据。Paper 数据集包含了约 8000 位学术工作者对 13700 余条文献的收藏行为数据,数据集中还包含一个引文网络和标签子网络,它们共同构成知识图谱。

(3) book 数据集

book 数据集来源于 Book-Crossing^[57]数据。Book-Crossing 是一个图书分享平台,平台中以图书作为社交载体,用户根据喜好对图书进行打分。Book-Crossing 数据由平台上 17860 名用户对 14967 本图书的评分组成,数据中还提供了图书的详细信息,包括书的作者、出版社和出版日期等。

(4) movie 数据集

movie 数据集来源于 MovieLens-1M^[58]数据。MovieLens 是用于推荐研究的电影数据网站。MovieLens-1M 从相关电影资料库上收集了 6036 名用户根据喜好程度对 2347

部电影的评分，分数范围 1 至 5 分。除了用户对影片的评分信息外，MovieLens-1M 数据还提供了用户的身份信息和电影的相关概要信息。

(5) music 数据集

music 来源于 Last.FM^[59]数据。Last.FM 数据从音乐社区平台收集了 1872 名用户对 3846 首歌曲的历史收听记录的曲名和播放次数。除此之外，Last.FM 数据集中还提供歌曲的歌手和曲名等基本信息以及用户的相关基础信息。

除了 paper 和 article 数据集中的交互为收藏形式的隐式交互，book、movie 和 music 数据集中都是显式交互。本模型用于预测用户是否可能与物品产生交互行为，因此需要转化 book、movie 和 music 数据集中的交互形式。由于 movie 数据集相对稠密，因此本文设置评分阈值，将所有评分大于等于阈值的用户-物品对作为正样本，评分小于阈值的用户-物品对作为负样本；考虑到 book 和 music 数据集的稀疏性，实验直接将用户-交互物品对记为正样本，并从未评分的物品中以正负样本 1:1 的比例随机抽取负样本。

除了 article 和 paper 数据集本身自带知识图谱外，其他数据集需要从微软开源的知识图谱 Satori^[60]中抽取与物品有关的信息形成知识图谱，以弥补自身数据量的不足。如果物品无法对应知识图谱中的任一节点，为了简化实验，会在数据预处理阶段删除该物品。最终五个数据集的详细信息如表 5 所示。实验中，将各数据集以 6:2:2 的比例随机划分训练集、验证集和测试集。

表 5 数据集描述

数据来源	数据类别	article	paper	book	movie	music
用户-物品 交互数据	用户数	5547	7884	17860	6036	1872
	物品数	11714	13782	14967	2347	3846
	交互数	295146	190936	139746	753772	42346
知识图谱 数据	节点数	11714	13782	24039	6729	9366
	关系数	1	4	10	7	60
	三元组数	89418	65130	19793	20195	15518

7.1.2 实验环境

实验中用到的服务器硬件和软件配置如表 6 所示。硬件部分，实验使用的服务器的 CPU 处理器为 Intel(R) Core(TM) i9-10920X CPU @ 3.50GHz，内存为 128G，GPU 为

GeForce RTX 3090，显存为 24G。软件部分，实验在 Ubuntu 20.04.1 版本的操作系统中进行；编程语言及版本为 Python 3.7；深度学习框架及版本为 TensorFlow 1.15。

表 6 服务器硬件与软件配置

项目	版本/内容
CPU	Intel(R) Core(TM) i9-10920X CPU @ 3.50GHz
内存	128G
GPU	GeForce GTX 3090
显存	24G
操作系统	Ubuntu 20.04.1
Python	3.7

7.1.3 对比算法

为了验证 MRKC 算法的有效性，本文对比了多个基准推荐算法进行实验。下面是各个基准算法的详细介绍。

(1) 知识协同嵌入算法 (Collaborative Knowledge base Embedding, CKE) [61]

CKE 是利用知识嵌入改进传统协同过滤方法的代表性算法。它首先通过分解用户物品交互矩阵得到用户和物品特征，再利用嵌入方法 TransE 对物品相关知识图谱的辅助知识信息进行编码，丰富物品的特征表示，进而有效提升推荐算法预测的准确性。

(2) 个性化节点推荐算法 (Personalized Entity Recommendation, PER) [5]

PER 是一种基于元路径的知识图谱推荐算法，它首先人工预定义用户和物品在知识图谱中的关联元路径，利用相似度搜索方法计算不同节点间的相似度，寻找用户可能感兴趣的物品，实现物品的个性化推荐。

(3) 涟漪网络 (RippleNet) [8]

RippleNet 是一个经典的基于知识图谱传播机制的推荐算法，该算法借鉴水波传播机制，以用户交互过的物品为种子，并以种子集合为中心，在知识图谱中一圈一圈地向外扩散，通过偏好传播得到多阶邻居，从而实现用户兴趣的延伸和扩展。

(4) 知识图谱卷积网络 (Knowledge Graph Convolutional Network, KGCN) [46]

KGCN 是基于图卷积神经网络的一种推荐模型，常被选作基于知识图谱推荐研究领域的对比基准算法。KGCN 从节点的邻域中进行采样，层内聚合来自样本尾节点的信息。随着邻域的扩展，用户潜在兴趣从低阶邻居扩展到高阶邻居。

(5) 图谱增强的推荐算法 (Multi-task Learning for KG enhanced Recommendation, MKR) [48]

MKR 模型是借助知识图谱增强的推荐算法, 它将推荐任务和图谱表示学习任务分开, 在两侧分别使用多层感知机来学习用户和关系的表示, 并设计交叉压缩单元来沟通学习物品与节点间的特征, 通过多任务机制使得两侧任务均学习到来自对方的信息, 并在高层分别计算推荐点击率和图谱表示得分。

7.1.4 评估指标

为了简化实验, 推荐系统的评估主要采用离线指标评估的方法。本文在选取评估指标时主要考虑了两种推荐场景: 点击率预估以及最佳 K 个物品推荐 (TOP-K)。

(1) 点击率预估

点击率预估是一个二分类场景, 模型输入用户和待预测物品的相关信息, 输出用户是否会点击该物品。在该场景下常使用 AUC^[62]、ACC 和 F1 这三个评价指标。

二分类场景下, 主要有四种样本类型: 实际为真, 预测为真是真阳性(True Positive, TP); 实际为假, 预测为真的是假阳性(False Positive, FP); 实际为假, 预测为假的是真阴性(True Negative, TN); 实际为真, 预测为假的是假阴性(False Negative, FN)。AUC 表示 ROC (Receiver Operator Curve) 曲线下的面积, ROC 曲线的横坐标是假阳率 (假阳性所占比例), 纵坐标是真阳率 (真阳性所占比例)。对于给定的一对正样本和负样本, 本实验的 AUC 取值正样本得分大于负样本得分的概率。假定在数据集中, 正样本的个数为 M , 负样本的个数为 N , 所有样本按照预测点击率得分从高到底排列, $rank_i$ 表示第 i 条正样本的排序序号。AUC 的计算公式表示为:

$$AUC = \frac{\sum_{i=1}^N rank_i - \frac{M \times (M + 1)}{2}}{M \times N} \quad (7.1)$$

ACC, 即准确率 (Accuracy), 可以通俗的认为是预测正确的概率, 计算公式如下:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (7.2)$$

在上式中, 分子表示所有预测正确的样例, 分母表示所有样例。

在点击率预估场景下, 本实验中选择 0.5 作为预测阈值, 当预测点击率大于等于 0.5 时计为正样本, 小于计为负样本。再根据用户与物品产生交互的真实标签, 计算出精确率 (Precision) 和召回率 (Recall)。精确率表达的是所有预测为真的样本中有多少

是真正的正样本，其计算公式为：

$$Precision = \frac{TP}{TP + FP} \quad (7.3)$$

召回率描述对所有正样本的预测中有多少为真，其计算公式为：

$$Recall = \frac{TP}{TP + FN} \quad (7.4)$$

F1 是模型精确率和召回率的调和平均，用于评估二分类模型训练效果的指标。F1 越高，说明模型的预测效果越好。其计算公式为：

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (7.5)$$

(2) TOP-K 推荐

在 TOP-K 推荐场景下，需要按照预测点击率对待推荐物品进行排序，向用户推荐前 K 个物品。与点击率预估场景不同，TOP-K 推荐关注推荐列表中是否含有用户真正感兴趣的，而不关注列表内物品的相关排序。本实验选择召回率 (Recall@K) 作为 TOP-K 推荐场景下的评估指标。

Recall@K 指的是模型向用户推荐的 K 个物品中，用户真正交互过的物品个数所占的比例。比例越高，表示模型推荐得越符合用户兴趣。定义 $T(u)$ 为用户在测试集中真实交互的物品记录， $R(u)$ 表示给用户推荐的 K 个物品，则召回率的计算公式为：

$$Recall@K = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |T(u)|} \quad (7.6)$$

(3) 图谱补全场景

在图谱补全场景，本实验采用均方根误差 (RMSE) 指标来衡量知识图谱模块所预测出的尾部节点和真实三元组关系中的尾部节点的差距。RMSE 的计算公式表示为：

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (p_i - a_i)^2}{n}} \quad (7.7)$$

其中， p_i 代表的是预测的尾部节点向量， a_i 代表的是真实的尾部节点向量。因此，RMSE 的结果越小代表着预测尾部节点和真实尾部节点之间的差异小，说明知识图谱的学习更加充分。

7.2 实验结果与分析

7.2.1 公开数据集对比实验

点击率预估场景下，本文提出的 MRKC 算法与各个对比方法在三个公开数据集上的 AUC、ACC 和 F1 的实验结果如表 7 所示，下划线数字为次优指标，加粗数字为最优指标。从表中结果可知，MRKC 算法在 book、movie 数据集中所有指标都取得了最优的表现。相对于次优结果，book 数据集上 MRKC 的 AUC、ACC 指标相对提升 1.1%、0.2%，在 F1 指标上同次优结果持平；movie 数据集上的 AUC、ACC、F1 指标相对提升 0.7%、0.9%和 1.1%；music 数据集上的 AUC、F1 指标分别提升 1.0%和 0.3%。这表明了 MRKC 算法在通用推荐场景下有着很好的表现。

表 7 各算法在点击率预估场景下的 AUC、ACC 和 F1 指标对比

算法	book			movie			music		
	AUC	ACC	F1	AUC	ACC	F1	AUC	ACC	F1
CKE	0.671	0.633	0.602	0.801	0.742	0.785	0.744	0.673	0.689
PER	0.623	0.588	0.568	0.710	0.664	0.688	0.633	0.596	0.566
RippleNet	0.729	0.662	0.658	<u>0.918</u>	<u>0.842</u>	<u>0.845</u>	0.768	0.691	0.715
KGCN	0.725	0.689	0.654	0.907	0.835	0.834	0.793	0.744	0.722
MKR	<u>0.734</u>	<u>0.702</u>	0.661	0.914	0.840	0.84	<u>0.793</u>	0.753	<u>0.730</u>
MRKC	0.742	0.704	0.661	0.925	0.850	0.854	0.801	<u>0.749</u>	0.732

值得注意的是，在 music 数据集中，MRKC 算法的 AUC 指标结果高于 MKR，但 ACC 指标结果略有下降，这表明模型对于正负样本的排序能力有所提升，但打分能力略有下降，可能的原因是 music 数据集中的用户与物品交互稠密度仅为 0.059%，其中对应的知识图谱的节点平均度为 3.3，训练数据较为稀疏，导致对用户和物品的表征学习有限，虽然模型能够更好地将正样本排在负样本前面，但无法更为精准地预测用户点击物品的概率。

横向比较三个数据集上的表现可以看出，MRKC 在 book、music 和 movie 数据集上的提升效果逐渐上升。通过计算，book、music 和 movie 数据集上用户物品交互稠密度分别为 0.052%、0.059%和 5.32%，对应知识图谱上的节点平均度分别为 1.6、3.3 和 6.0。由此可以分析得出，MRKC 的推荐表现很大程度上依赖于训练数据集的稀疏程度：

MRKC 在稠密数据集上学习能力更强，推荐的表现更加优异，也表明 MRKC 在完备度更高的图谱上更具优势。

图 15-图 17 表示 MRKC 模型在 TOP-K 推荐场景下与各个对比方法分别于 book、movie 和 music 数据集中的 Recall 指标实验结果。分析各个数据集下的召回率结果，可以得到一个统一的结论：随着 K 值越来越大，MRKC 相比于其他推荐算法在召回率指标上提升越明显。这说明随着推荐列表的扩大，MRKC 倾向于把所有正样本全部找出来。然而，观察三张折线图，会发现当 K 值取小于等于 10 时，MRKC 算法的优势没有特别明显，甚至在 book 数据集上低于 MKR 模型。主要原因是因为 book 数据集中的负样本采样自用户未交互的物品中，这部分物品也有可能是用户所喜爱的，具有较大的潜力与用户进行交互，所以 MRKC 算法给这些负样本也打出了较高的预测得分。

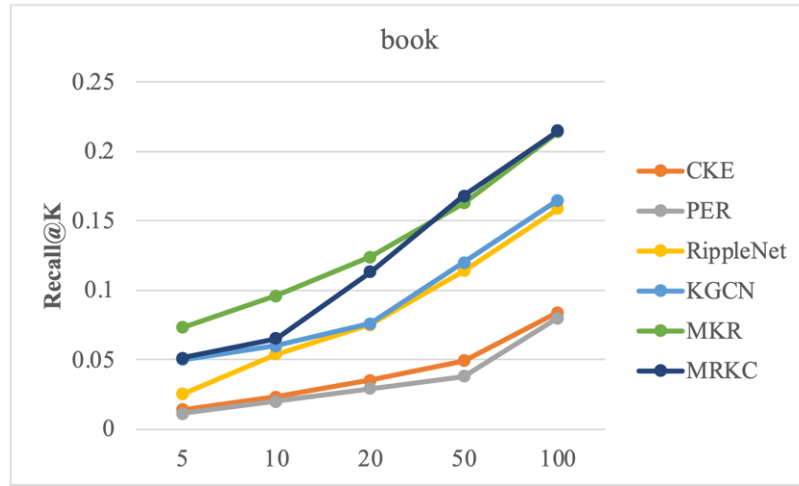


图 15 book 数据集上 Recall@K 的结果

在 book 和 music 数据集上对比各个基准算法，可以发现基于路径的方法 PER 表现效果最差，可能的原因是这两个数据集中图谱比较稀疏，预先定义的元路径无法完整，难以充分发挥图谱的辅助作用，导致推荐效果不佳；知识协同嵌入方法 CKE 的推荐效果仅高于 PER，与其他基于传播的推荐算法有较大差距，说明基于图谱嵌入的算法更加适用于图谱相关任务，无法充分发挥知识图谱对推荐的优势。RippleNet、KGCN 都是基于传播的推荐算法，但是性能都被 MRKC 超过，这验证了物品与节点异质化建模的有效性，体现了联合图谱补全对推荐的正向作用。同时，多任务推荐方法 MKR 在大多数指标下也低于 MRKC，这一方面表明了捕捉用户物品交互中协同信息的重要性，另一方面也说明了 MRKC 中多层注意力机制在信息同异阶信息传播与聚合时的有效性。

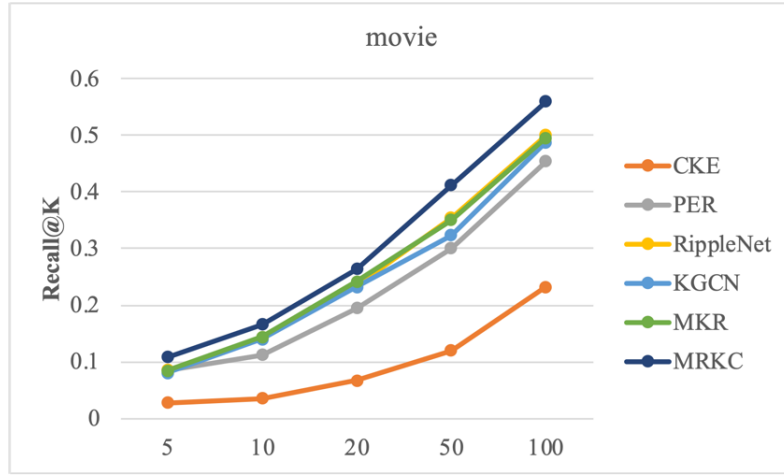


图 16 movie 数据集上 Recall@K 的结果

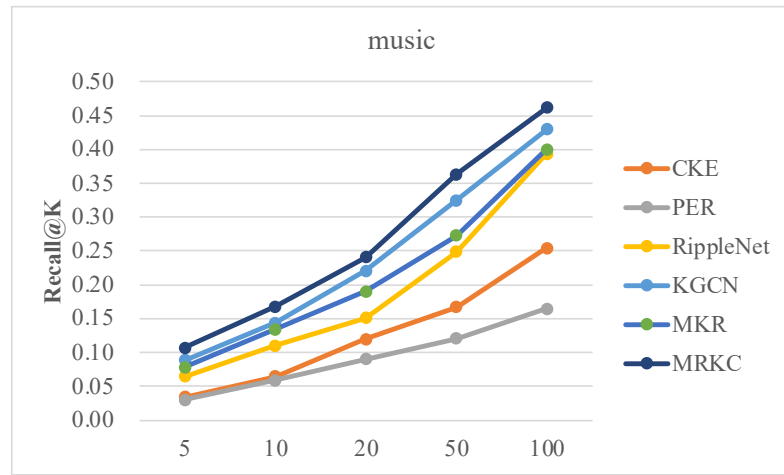


图 17 music 数据集上 Recall@K 的结果

7.2.2 论文推荐场景对比实验

表 8 展示了在论文推荐领域, MRKC 模型与对比算法在 paper 和 article 数据集上的 AUC 与 F1 指标的实验结果。从表中结果可知, MRKC 算法在 paper 和 article 数据集中两个指标都取得了最优异的效果。相对于次优结果, article 数据集上 MRKC 的 AUC、F1 指标相对提升 1.6% 和 1.6%; paper 数据集上的 AUC、F1 指标相对提升 2.1% 和 2.3%。这表明了 MRKC 算法在论文场景下的推荐效果有着较高的提升。

与通用数据集的结果不同, 论文数据集上 PER 的推荐效果明显高于 CKE, 可能的原因是自建的 article 和 paper 数据集中关系数量较少, 元路径的设定相对简单和完整, 使得 PER 模型能够更好地学习到用户的收藏兴趣, 从而带来推荐效果的提升。RippleNet 是表现最好的基准算法, 推荐效果相较于 MKR 有着较大提升, 验证了挖掘协同信息能够更加充分地利用关系简单的论文知识图谱, MRKC 效果的提升则进一步

说明了传播和聚合协同信息能够进一步丰富用户和物品的表征，从而有效提高论文的推荐效率。

表 8 各算法在论文数据集上的 AUC 和 F1 指标对比

算法	article		paper	
	AUC	F1	AUC	F1
CKE	0.715	0.655	0.722	0.668
PER	0.720	0.701	0.745	0.672
RippleNet	<u>0.824</u>	<u>0.763</u>	<u>0.767</u>	<u>0.700</u>
MKR	0.791	0.745	0.759	0.699
MRKC	0.837	0.775	0.783	0.716

7.2.3 图谱补全实验

表 9 各算法在图谱学习场景下的 RMSE 指标对比

算法	book	movie	music	article
TransE	0.319	0.596	0.480	0.451
MKR	0.302	0.558	0.471	0.432
MRKC/T	0.287	0.554	0.451	0.430
MRKC	0.184	0.450	0.388	0.374

表 9 展示了 MRKC 算法与基准算法在图谱补全场景下 RMSE 指标的实验结果。MRKC/T 消融了 MRKC 中图谱的同阶与异阶邻域信息聚合。从实验结果上看，MRKC 对比基础的 TransE 和 MKR 效果提升明显，说明推荐任务对图谱补全任务的效果有着显著的正向作用，充分挖掘用户与物品的协同信息有助于图谱的表示学习；同时 MRKC 较于 MRKC/T 也有较大效果提升，说明注意力机制下的同阶与异阶邻域信息聚合能够丰富节点表示，对图谱的补全起着至关重要的作用。

7.2.4 消融实验

表 10 MRKC 变体模型在不同数据集下的 AUC 结果

算法	book	movie	music	article
MRKC/S	0.735	0.900	0.780	0.750
MRKC/R	0.738	0.910	0.783	0.755
MRKC/C	0.740	0.911	0.785	0.773
MRKC	0.742	0.925	0.801	0.837

为了进一步验证 MRKC 算法中各个模块的有效性，消融实验对比了 MRKC 的三种变体。MRKC/S 将高阶共享单元退化为普通的交叉压缩单元，采用相同纬度建模物品与节点；MRKC/R 算法消融了模型中应用于用户-物品交互连通图上的图卷积神经网络，

为了保证和图谱中的节点与关系迭代次数相同，仅用多层全连接神经网络替代；MRKC/C 算法直接对邻域信息进行平均池化，同时在不同阶邻居信息聚合时直接求和池化。表 10 展示了各个变体在四个数据集上的 AUC 指标的结果。从表中可知，MRKC/S 在四个数据集上的 AUC 指标有明显的效果下降，这表明异质化建模并通过高阶共享单元学习物品与节点的表征对推荐性能的提升有着巨大的优势；MRKC/R 的效果也低于 MRKC，说明用户与物品交互中协同信息的捕获能够有效提升推荐效果；MRKC/C 相比于 MRKC 在 AUC 上均有略微下降，这反映了区分同阶信息重要性和建模异阶信息衰减对图谱补全有着正向作用。

7.2.5 模型超参数探索

MRKC 模型中有不同的超参数，本节继续深入实验，对超参数继续探究，加深对 MRKC 模型的分析。

(1) 训练间隔

在基于高阶交互单元的多任务训练框架下，推荐任务与图谱任务的训练轮次差异对 MRKC 的实验结果有很大影响。表 11 详细展示了三个不同的训练间隔对 MRKC 模型的 AUC 结果的影响——训练间隔表示训练多次推荐任务后训练图谱任务——其中加粗数字表示最佳效果。从结果来看，book、movie、music、article 和 paper 数据集上的训练间隔分别为 2、3、2、2 和 1 时，AUC 结果达到最优。实验结果表明，如果训练间隔过小，图谱任务的环境噪声会过多引入推荐任务，导致推荐效果的下降；在一定程度上增加推荐任务的训练轮次，能够充分利用知识图谱作为辅助信息扩充用户和物品向量的表征能力，但是当训练轮次过大时，会导致物品与对应节点的交叉学习不足，导致图谱学习无法进一步推升推荐任务的效果。

表 11 MRKC 不同训练间隔对 AUC 结果的影响

训练间隔	book	movie	music	article	paper
1	0.735	0.918	0.788	0.835	0.783
2	0.740	0.919	0.801	0.837	0.780
3	0.730	0.925	0.791	0.832	0.777

(2) 信息传播阶数

在协同信息的传播和图谱邻域信息的传播中，信息传播阶数将会直接影响到推荐和图谱学习效果。表 12 展示了不同的信息传播阶数对多个数据集下 AUC 结果的影响，其

中加粗数字表示最优效果。从结果来看，当 book、movie、music、article 和 paper 数据集上的信息传播阶数分别为 3、3、1、3 和 2 时，AUC 指标值达到最佳。实验结果表明，在一定程度上增加信息传播的深度，能够更加有效地结合邻域信息，提升用户、物品和图谱中各个节点向量表征能力；但是当传播深度达到一定限度后，传播到的节点与中心节点差异较大，可能会带来一定的噪声干扰。对于不同数据集，需要试验确定最佳的信息传播阶数。

表 12 MRKC 不同信息传播深度对 AUC 结果的影响

信息传播深度	book	movie	music	article	paper
1	0.735	0.918	0.801	0.835	0.776
2	0.735	0.920	0.793	0.836	0.783
3	0.740	0.925	0.790	0.837	0.777

(3) 物品与节点表征维度

为了验证高阶共享单元的有效性，探究物品与节点不同建模维度对实验的影响，本文在五个数据集上实验了不同向量维度对点击率预估场景下 AUC 指标的影响，为了简化实验，防止引入过多超参数，本文在实验时固定图谱中节点和关系的向量维度，所取值来自最佳参数。实验结果如图 18-图 22 所示。从实验结果来看，当用户与物品的向量维度分别取 64、64、8、32 和 16 时，book、movie、music、paper 和 article 数据集下的 AUC 指标达到最大值。这表明在一定范围内，增加用户物品向量维度能有效编码更多推荐领域信息，提升共享单元对物品和节点的表征学习，从而带来 AUC 指标的提高；但是维度超过一定阈值后，会使得用户物品向量引入了一定的环境噪声信息，随着共享单元传入图谱学习中，进而影响了两个任务的学习与泛化，导致其 AUC 指标出现下降。

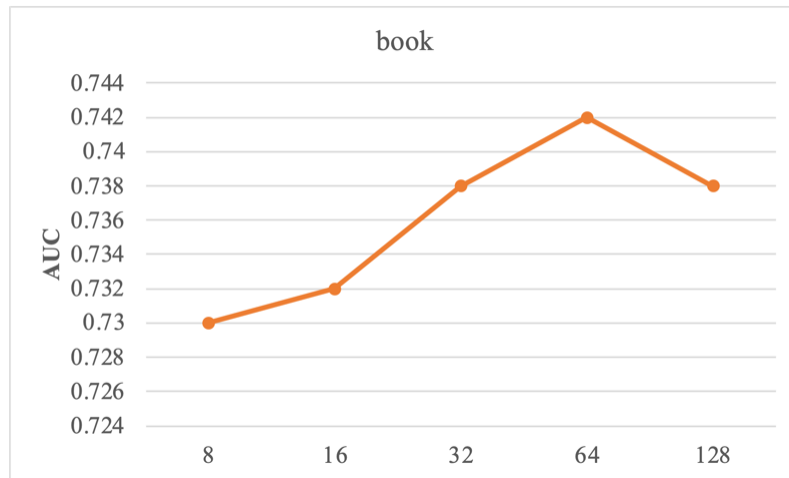


图 18 book 数据集上 AUC 指标随物品向量维度变化

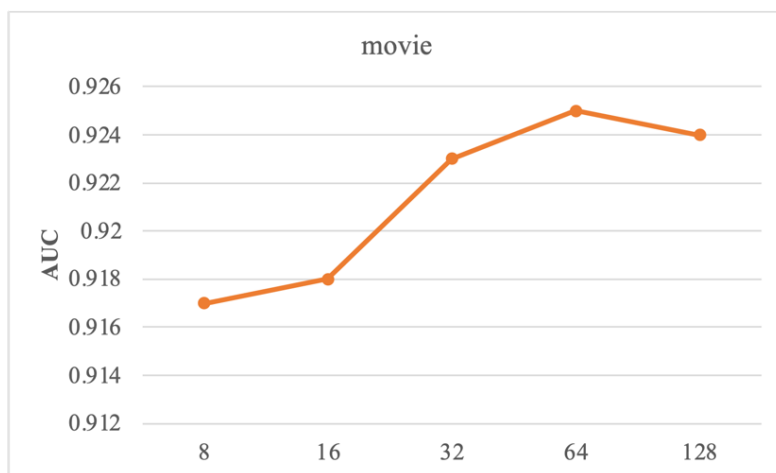


图 19 movie 数据集上 AUC 指标随物品向量维度变化

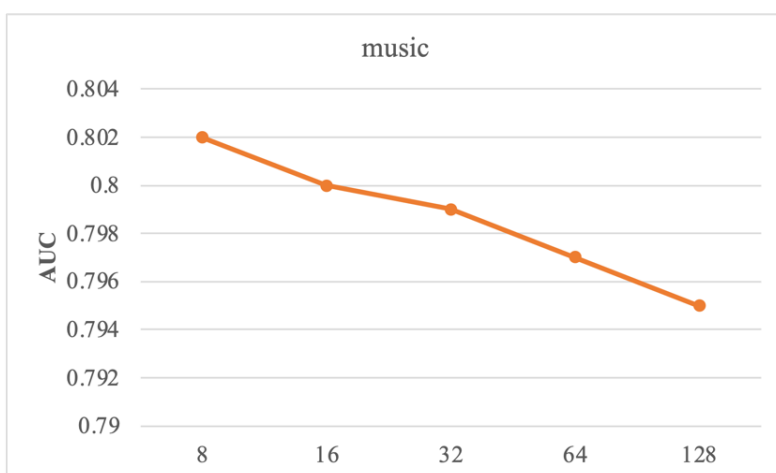


图 20 music 数据集上 AUC 指标随物品向量维度变化

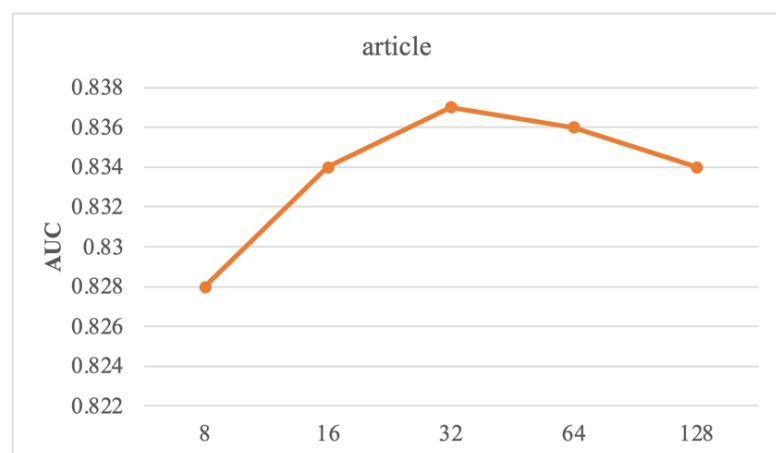


图 21 article 数据集上 AUC 指标随物品向量维度变化

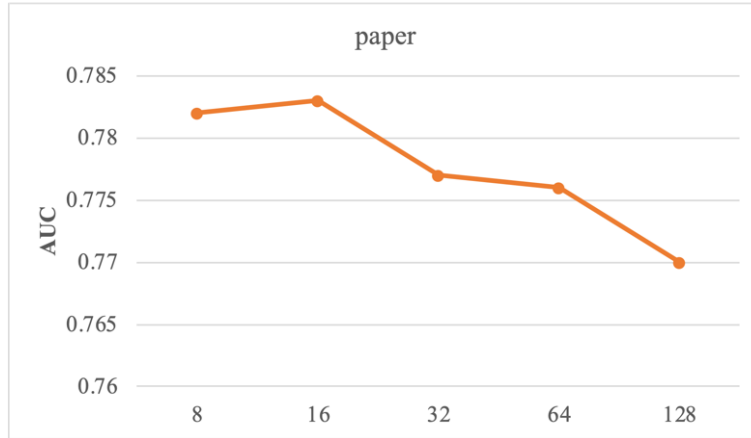


图 22 paper 数据集上 AUC 指标随物品向量维度变化

7.3 本章小结

本章对本文提出的 MRKC 算法进行实验和验证。首先介绍实验数据集、软硬件设备环境以及用于对比实验的基准算法和评估指标。接着是对实验结果的分析，这里分为四部分：第一部分是展示三个通用数据集上，MRKC 算法与对比方法在各个评估指标下的实验结果，并从中归纳实验结论；第二部分为 MRKC 在论文推荐场景下与相关对比算法的实验结果；第三部分是 MRKC 模型的消融实验，主要是对高阶共享单元、协同信息的捕捉以及多阶注意力机制的进行消融，并对消融结果加以分析；第四部分是对 MRKC 算法的超参数进行深入分析，探究模型受超参数的影响并分析可能的原因。

总结与展望

本文聚焦基于知识图谱的多任务推荐算法研究。针对现有工作存在的不足，本文重点研究物品节点的异质化建模、协同信息的感知以及图谱上同异阶邻域信息的传播，主要内容包含以下方面：

(1) 针对现有方法中物品与节点建模同质化、任务间参数交叉学习能力有限和图谱完备性欠缺的问题，本文提出基于高阶共享单元的多任务学习框架，异质化建模物品与节点，并设计基于异阶特征交叉的共享单元，联合训练推荐任务和知识图谱补全任务。

(2) 针对现有方法中推荐任务存在的忽略用户物品交互中的协同信息对图谱补全的正向作用、忽视多阶交互中的协同信息差异的不足，本文提出基于图卷积神经网络的协同信息感知方法，利用图卷积神经网络充分挖掘用户物品交互过程中的协同信息，同时引入门控单元自适应建模不同阶信息的通入量。

(3) 针对现有方法中图谱补全任务存在的弱化关系建模、忽略长距离信息传播衰减现象的不足，本文采用多层注意力机制分别建模同阶邻域不同关系下尾节点的重要性差异和异阶邻域信息长距离传播的衰减。

本文基于上述内容，提出了基于知识图谱的推荐与补全网络 **MRKC**。在基于高阶共享单元的多任务学习框架基础上，**MRKC** 算法集成了基于图卷积神经网络的协同信息感知方法和基于多层注意力机制的图谱补全方法，形成了一个端到端的可训练的模型。

本文在书籍、电影和音乐三个通用数据集上验证了算法的性能，结果表明本文提出的 **MRKC** 算法在 AUC 指标上分别相对提升 1.1%、0.7% 和 1.0%；两个自建的论文数据集上在 AUC 指标上相对提升 1.6% 和 2.1%。本文还进行图谱补全实验和消融实验论证，并对超参数进行实验分析，并对相关实验现象进行了解释。

然而因为作者的能力有限，研究内容还存在很多待完善的地方，更多改进点有待发掘。根据作者对该研究方向的认识和了解，未来可能的发展方向包括：

(1) 本文提出的 **MRKC** 算法仅考虑了点击率预估场景下推荐任务与图谱补全任务的联合训练。然而现实推荐场景下，常常需要同时考虑多个指标，如 CTR 点击率、CVR 转化率、视频播放时长和用户停留时长等。因此未来计划引入新的推荐子任务，

进一步增强推荐任务对图谱补全的正向作用，将图谱中的语义信息引入多个推荐子任务中，增强向量的泛化能力，提升推荐效果。

（2）目前推荐侧和图谱侧的信息传播过程都依赖于邻域的多阶扩展，但是随着阶数增加，邻域节点集合呈现爆炸式增长，本文使用的随机采样即会带来信息损失，有可能引入环境噪声。因此，未来可以考虑采用更加适合的采样策略，降低噪声的引入和减小语义信息的损失，提高邻域信息表达能力。

参考文献

- [1] 曾文, 车尧. 科技大数据的情报分析技术研究[J]. 情报科学, 2019, 37(3): 93–96.
- [2] 孟祥武, 胡勋, 王立才, 等. 移动推荐系统及其应用[J]. 软件学报, 2013, 24(1): 91-108.
- [3] ZHANG Y, SONG W. A Collaborative Filtering Recommendation Algorithm Based on Item Genre and Rating Similarity[M]. 2009. DOI 10.1109/CINC.2009.219
- [4] CAO Y X, WANG X, HE X N, et al. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences[C]//The world wide web conference. 2019: 151-161.DOI 10.1145/3308558.3313705
- [5] Yu X, Ren X, Sun Y, et al. Personalized Entity Recommendation: A Heterogeneous Information Network Approach[C]//Proceedings of the 7th ACM International Conference on Web Search and Data Mining. 2014: 283-292.
- [6] Sun Y, Han J, Yan X, et al. Pathsim: Meta Path-based Top-k Similarity Search in Heterogeneous Information Networks[J]. Proceedings of the VLDB Endowment. 2011, 4(11): 992-1003.
- [7] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian Personalized Ranking from Implicit Feedback[C]//Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. 2009: 452-461.
- [8] Wang H, Zhang F, Wang J, et al. Ripplenet: Propagating User Preferences on the Knowledge Graph for Recommender Systems[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018: 417-426.
- [9] Wang Z, Lin G, Tan H, et al. CKAN: Collaborative Knowledge-aware Attentive Network for Recommender Systems[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 219-228.
- [10] Chen Y, Yang Y, Wang Y, et al. Attentive knowledge-aware graph convolutional networks with collaborative guidance for personalized recommendation[C]//2022 IEEE 38th International Conference on Data Engineering (ICDE). IEEE, 2022: 299-311.
- [11] Geng S, Fu Z, Tan J, et al. Path language modeling over knowledge graphsfor explainable recommendation[C]//Proceedings of the ACM Web Conference 2022. 2022: 946-955.
- [12] Park S J, Chae D K, Bae H K, et al. Reinforcement learning over sentiment-augmented knowledge graphs towards accurate and explainable recommendation[C]//Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. 2022: 784-793.
- [13] Zhang Y, Yang Q. An overview of multi-task learning[J]. National Science Review, 2018, 5(1): 30-43.
- [14] Ruder S. An overview of multi-task learning in deep neural networks[J]. arXiv preprint arXiv:1706.05098, 2017.
- [15] Zhang Y, Yang Q. A survey on multi-task learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 34(12): 5586-5609.

- [16] Ma X, Zhao L, Huang G, et al. Entire space multi-task model: An effective approach for estimating post-click conversion rate[C]//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018: 1137-1140.
- [17] Ni Y, Ou D, Liu S, et al. Perceive your users in depth: Learning universal user representations from multiple e-commerce tasks[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 596-605.
- [18] Scarselli F, Tsoi A C, Gori M, et al. Graphical-based Learning Environments for Pattern Recognition[C]//Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR). 2004: 42-56.
- [19] Scarselli F, Gori M, Tsoi A C, et al. The Graph Neural Network Model[J]. IEEE Transactions on Neural Networks. 2008, 20(1): 61-80.
- [20] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs[J]. arXiv preprint arXiv:1312.6203, 2013.
- [21] Henaff M, Bruna J, LeCun Y. Deep Convolutional Networks on Graph-structured Data[J]. arXiv preprint arXiv:1506.05163, 2015.
- [22] Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks[J]. arXiv preprint arXiv:1710.10903, 2017.
- [23] Wang X, He X, Wang M, et al. Neural Graph Collaborative Filtering[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2019: 165-174.
- [24] He X, Deng K, Wang X, et al. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 639-648.
- [25] Wang X, He X, Cao Y, et al. KGAT: Knowledge Graph Attention Network for Recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2019: 950-958.
- [26] Lin H, Liu Y, Wang W, et al. Learning Entity and Relation Embeddings for Knowledge Resolution[J]. Procedia Computer Science, 2017, 108: 345-354.
- [27] Wang Q, Mao Z, Wang B, et al. Knowledge Graph Embedding: A Survey of Approaches and Applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [28] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [29] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online Learning of Social Representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014: 701-710.
- [30] Grover A, Leskovec J. Node2vec: Scalable Feature Learning for Networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference. 2016: 855-864.

-
- [31] Bordes A, Usunier N, Garcia-Duran A, et al. Translating Embeddings for Modeling Multi-relational Data[C]//Advances in Neural Information Processing Systems. 2013: 2787-2795.
 - [32] Wang Z, Zhang J, Feng J, et al. Knowledge Graph Embedding by Translating on Hyperplanes[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2014: 1112-1119.
 - [33] Ji G, He S, Xu L, et al. Knowledge Graph Embedding via Dynamic Mapping Matrix[C]//Meeting of the Association for Computational Linguistics & the International Joint Conference on Natural Language Processing. 2015: 687-696.
 - [34] Yoshua, Bengio, Jason, et al. A Semantic Matching Energy Function for Learning with Multi-relational Data Application to Word-sense Disambiguation[J]. Machine learning, 2014, 94(2):233-259.
 - [35] Socher R, Chen D, Manning C D, et al. Reasoning with Neural Tensor Networks for Knowledge Base Completion[C]//Advances in Neural Information Processing Systems. 2013: 926-934.
 - [36] Wang Q, Mao Z, Wang B, et al. Knowledge Graph Embedding: A Survey of Approaches and Applications[J]. IEEE Transactions on Knowledge and Data Engineering. 2017, 29(12): 2724-2743.
 - [37] Mazumder S, Liu B. Context-aware path ranking for knowledge base completion[J]. arXiv preprint arXiv:1712.07745, 2017.
 - [38] Xiong W, Hoang T, Wang W Y. Deeppath: A reinforcement learning method for knowledge graph reasoning[J]. arXiv preprint arXiv:1707.06690, 2017.
 - [39] Xiong W, Yu M, Chang S, et al. One-shot relational learning for knowledge graphs[J]. arXiv preprint arXiv:1808.09040, 2018.
 - [40] Schuster M, Paliwal K K. Bidirectional Recurrent Neural Networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681.
 - [41] Vaswani A, Shazeer N, Parmar N, et al. Attention is All You Need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 6000–6010
 - [42] Shaw P, Uszkoreit J, Vaswani A. Self-attention with Relative Position Representations[J]. arXiv preprint arXiv:1803.02155, 2018.
 - [43] Tay Y, Luu A T, Hui S C. Multi-pointer Co-attention Networks for Recommendation[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2018: 2309-2318.
 - [44] Zhou G, Zhu X, Song C, et al. Deep Interest Network for Click-through Rate Prediction[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2018: 1059-1068.
 - [45] Wu C, Wu F, An M, et al. NPA: Neural News Recommendation with Personalized Attention[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019: 2576–2584.
 - [46] Wang H, Zhao M, Xie X, et al. Knowledge Graph Convolutional Networks for Recommender Systems[C]//The World Wide Web Conference. 2019: 3307-3313
 - [47] Rawat W, Wang Z. Deep convolutional neural networks for image classification: A comprehensive review[J]. Neural computation, 2017, 29(9): 2352-2449.

- [48] Wang H, Zhang F, Zhao M, et al. Multi-task feature learning for knowledge graph enhanced recommendation[C]//The world wide web conference. 2019: 2000-2010.
- [49] Stone M H. The generalized Weierstrass approximation theorem[J]. Mathematics Magazine, 1948, 21(5): 237-254.
- [50] Wang H, Zhang F, Xie X, et al. DKN: Deep knowledge-aware network for news recommendation[C]//Proceedings of the 2018 world wide web conference. 2018: 1835-1844.
- [51] 陈清泓, 林广艳, 柳啸峰, 等. 基于分离式表征的知识图谱推荐算法[J]. 武汉大学学报(理学版), 2021, 67(6): 517-524.
- [52] Feng C M, Gao Y L, Liu J X, et al. Joint-norm constraint and graph-Laplacian PCA method for feature extraction[J]. BioMed Research International, 2017, 2017.
- [53] SHI X J, CHEN Z R, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]//Advances in neural information processing systems. 2015: 802-810.
- [54] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014. DOI 10.3115/v1/D14-1179
- [55] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback[J]. arXiv preprint arXiv:1205.2618, 2012.
- [56] Wang H, Chen B, Li W J. Collaborative Topic Regression with Social Regularization for Tag Recommendation[C]//Proceedings of the 23rd International Joint Conference on Artificial Intelligence. 2013: 2719-2725.
- [57] Shahrokhi F, Székely L A, Sýkora O, et al. The Book Crossing Number of a Graph[J]. Journal of Graph Theory, 1996, 21(4): 413-424.
- [58] Verbert K, Drachsler H, Manouselis N, et al. Dataset-driven Research for Improving Recommender Systems for Learning[C]//Proceedings of the 1st International Conference on Learning Analytics and Knowledge. 2011: 44-53.
- [59] Schedl M. The lfm-1b Dataset for Music Retrieval and Recommendation[C]//Proceedings of the 2016 ACM International Conference on Multimedia Retrieval. 2016: 103-110.
- [60] Zhao W X, He G, Dou H, et al. A Dataset for Linking Knowledge Bases with Recommender Systems[J]. arXiv preprint arXiv:1807.11141, 2018.
- [61] Zhang F, Yuan N J, Lian D, et al. Collaborative Knowledge Base Embedding for Recommender Systems[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 353-362
- [62] Lobo J M, Jiménez - Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models[J]. Global ecology and Biogeography, 2008, 17(2): 145-151.

攻读硕士期间取得的学术成果

- [1] 陈清泓, 林广艳, **柳啸峰**, 谭火彬. 基于分离式表征的知识图谱推荐算法[J]. 武汉大学学报(理学版), 2021, 67(6): 517-524.

致谢

转眼间，研究生三年就要结束了，近二十年的学生生涯也即将画上句号。想到写这篇致谢，竟忽然无从下笔，脑海中浮现了太多熟悉的面孔，与他们在一起的瞬间都是我人生中熠熠生辉的闪光点。接下来让我用一些心里话来表达对他们的感激之情。

首先要感谢林广艳老师和谭火彬老师，感谢他们对我的学业和生活的支持和帮助。林老师更像一位母亲，不仅指导我的学术研究，帮助我规划未来的职业方向，还会时刻关心我们的身体健康，以身作则地告诫我们要多运动，也会时常与我谈心，缓解我的学习压力和生活烦恼。谭老师则更像一位朋友，不仅能够专业地指导我们的实验，修改我们的论文，平常相处也非常的随和，给我们提供了一个轻松愉快的实验室环境。遇到两位老师是我读研期间最幸福的事情。

其次要感谢我的父母，感谢他们一直默默地付出，做我最坚强的后盾。曾经年少反叛，拒绝和他们交流，甚至把这种生疏当成一种骄傲，以为自己足够成熟。可是没有他们辛苦的工作，也不会有我现在的的生活；感谢他们在深夜烦躁委屈时安慰我，在迷茫无措时鼓励我大胆去做我想做的事情，让我有了继续前进的勇气。没有他们，就没有我今天的一切。也要感谢我的姐姐，感谢她陪我度过无数个挣扎彷徨的日夜，无论我做出什么选择，都始终坚定地站在我身边。也感谢她以亲身经历告诫我的那些人生教条——“人生为什么要努力？因为最痛苦的事情不是失败，而是我本可以”——指引着我前进。同时也要感谢我的女朋友黄可心，感谢她一直以来真心的陪伴，让我的人生充满爱与浪漫。

最后还要感谢一直陪伴我的小伙伴们。感谢王泽学长和陈清泓学长对我研究方向提供的建议和帮助，你们对待学术认真的态度是我毕生追赶的榜样。感谢于九阳学弟对我小论文工作的帮助。感谢陈国涛同学对我的包容和理解，我们在七年的舍友生涯里一起经历了各种酸甜苦辣，让我学到了你身上很多美好的品质。感谢强巴桑布、李哲明、贺俞凯和费云帆几位兄弟一直以来陪伴在我的身边。也感谢王哲同学对我的支持和理解，我们留下了许多共同追逐篮球梦想的美好瞬间。

篇幅有限，还有很多帮助过我的人无法一一罗列，请允许我在此向你们道一声谢谢。寥寥数语，难以表达我对大家的感激之情，就先祝各位身体健康，万事如意。