

中图分类号：TP391.3

论文编号：10006SY1921110

北京航空航天大学
硕士学位论文

基于异质信息网络传播机制的
推荐算法研究

作者姓名 陈清泓

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 Recommendation Algorithm on Heterogeneous Information Propagation Network

A Dissertation Submitted for the Degree of Master

Candidate: Chen Qinghong

Supervisor: Assoc. Prof. Lin Guangyan

School of Software

Beihang University, Beijing, China

中图分类号：TP391.3

论文编号：10006SY1921110

硕 士 学 位 论 文

基于异质信息网络传播机制的 推荐算法研究

作者姓名	陈清泓	申请学位级别	学术硕士
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	2019 年 9 月 1 日	起至	2021 年 12 月 6 日止
论文提交日期	2021 年 11 月 22 日	论文答辩日期	2021 年 12 月 6 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

推荐系统已经成为当今解决信息过载的主要工具，目前先进的推荐算法聚焦在如何利用外部的辅助信息（如知识图谱、社交网络等），来解决传统协同过滤的数据稀疏性和冷启动问题。这些辅助信息用图的形式展现，由于图中节点和关系的异质性，把它们统称为异质信息网络。本文研究基于异质信息网络传播机制的推荐算法，主要研究传播机制的两阶段：信息传播和信息聚合，并形成完整的推荐算法。本文主要工作内容如下：

（1）在信息传播阶段，针对现有方法中存在的建模颗粒度过粗、信息传播路径过于发散的不足，本文提出基于分离式表征的信息传播方法，细粒度刻画信息的不同层面，并采用基于关系连接的自适应采样方法对传播到的信息进行甄别和去噪。

（2）在信息聚合阶段，针对现有方法中存在的无法有效地区分不同信息的重要性、忽视信息在传播过程中的衰减现象的不足，本文提出基于分层注意力机制的信息聚合方法。第一层注意力机制用于区分同一信息层内各信息的重要程度，第二层注意力机制用于聚合各个不同传播深度的信息层，建模信息衰减现象。

（3）在信息传播和信息聚合阶段的研究基础上，本文提出一种基于分离式表征与分层注意力机制的异质信息推荐模型（DHAN）。该模型能够充分发挥基于分离式表征的信息传播方法和基于分层注意力机制的信息聚合方法的优点，形成端到端的推荐算法。

本文在论文、书籍、电影和音乐四个不同推荐场景的公开数据集上验证算法的性能，结果表明本文提出的 DHAN 算法在 AUC 指标上分别相对提升 1.4%、2.2%、0.4% 和 1.9%。本文还进行了消融实验，并对超参数进行实验分析，实验结果验证了算法的有效性。

关键词：推荐算法，异质信息网络，分离式表征，注意力机制

Abstract

Recommender system has become the main tool to solve the problem of information overload. To overcome the sparsity issue and cold start problem of traditional recommender system, the advanced recommendation algorithm focuses on how to put external auxiliary information, such as knowledge graph and social network, into use. The auxiliary information is presented in the form of graph. Due to the heterogeneity of nodes and relationships in the graph, it is referred to as heterogeneous information network. This paper conducts research of recommendation algorithm on heterogeneous information propagation network, which focuses on two stages of the propagation mechanism: information propagation and information aggregation. Based on the propagation mechanism, an integral recommendation algorithm is presented. The main contents of this work are as follows:

(1) In the stage of information propagation, as the existing studies have the problems of coarse modelling and divergent information dissemination, this paper proposes an information propagation method that is based on disentangled representation to describe different levels of information in fine granularity. Moreover, in this paper, an adaptive sampling algorithm based on relationship is also proposed to denoise the transmitted information.

(2) In the stage of information aggregation, as the existing methods cannot distinguish the importance of different information and ignore the attenuation of information in the process of propagation, this paper proposes an information aggregation method based on hierarchical attention mechanism. The first hierarchy of attention mechanism is used to distinguish the importance of information in the same level. The second hierarchy of attention mechanism is used to aggregate information layers of different propagation depths, and to model the phenomenon of information attenuation.

(3) Based on the proposed methods of information propagation and information aggregation, this paper presents a disentangled and hierarchical attentive network (DHAN) for heterogenous-information based recommendation, which takes full advantages of the method of information propagation that is based on disentangled representation and the method of

information aggregation that is based on hierarchical attention mechanism to form an end-to-end recommendation algorithm.

This paper verifies the performance of the algorithm on the public data sets under four different recommending scenarios about papers, books, movies and music recommendation. The experimental results show that the proposed DHAN improves the AUC by 1.4%, 2.2%, 0.4% and 1.9% respectively. The ablation experiment is carried out, and the hyper-parameters are also analyzed. The experimental results verify DHAN's effectiveness.

Keywords: Recommendation Algorithm, Heterogeneous Information Network, Disentangled Representation, Attention Mechanism

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 推荐系统研究现状.....	2
1.2.2 基于异质信息网络的推荐算法研究现状.....	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.....	5
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.....	8
2.1 异质信息网络	8
2.2 图表示学习	8
2.2.1 图表示学习的介绍.....	8
2.2.2 图表示学习在推荐中的应用.....	12
2.3 分离式表征	13
2.3.1 分离式表征的介绍.....	13
2.3.2 分离式表征在推荐中的应用.....	14
2.4 注意力机制	14
2.4.1 注意力机制的介绍.....	14
2.4.2 注意力机制在推荐中的应用.....	16
2.5 本章小结	17
第三章 基于分离式表征的信息传播方法.....	18
3.1 信息传播概念与现有问题分析	18
3.1.1 信息传播的概念.....	18
3.1.2 信息传播的一般流程.....	18

3.1.3 目前的方法存在的问题.....	19
3.2 基于分离式表征的信息传播方法的提出	20
3.2.1 问题描述.....	20
3.2.2 信息传播方法.....	21
3.3 本章小结	23
第四章 基于分层注意力机制的信息聚合方法.....	24
4.1 信息聚合概念与现有问题分析	24
4.1.1 信息聚合的概念.....	24
4.1.2 信息聚合的一般方法.....	24
4.1.3 目前的方法存在的问题.....	25
4.2 基于分层注意力机制的信息聚合方法的提出	26
4.2.1 问题描述.....	26
4.2.2 信息聚合方法.....	27
4.3 本章小结	29
第五章 基于分离式表征与分层注意力机制的异质信息推荐网络.....	31
5.1 问题描述与形式化定义	31
5.2 DHAN 模型结构	33
5.2.1 异质信息表示层.....	33
5.2.2 信息传播层.....	34
5.2.3 信息聚合层.....	35
5.2.4 融合预测层.....	36
5.3 模型训练	37
5.4 本章小结	37
第六章 实验与验证.....	39
6.1 实验方案设计	39
6.1.1 实验数据集.....	39
6.1.2 实验环境.....	41

6.1.3 对比算法.....	41
6.1.4 评估指标.....	42
6.2 实验结果与分析	44
6.2.1 对比实验的结果与分析	44
6.2.2 DHAN 模型消融实验结果与分析	47
6.2.3 DHAN 模型超参数探索分析	47
6.3 本章小结	50
总结与展望	51
参考文献	53
攻读硕士期间取得的学术成果	57
致谢	58

图 清 单

图 1 异质信息网络示例图	4
图 2 研究路线示意图	6
图 3 异质信息网络中的关联路径示例图	9
图 4 NTN 方法结构示意图	11
图 5 点乘注意力机制示意图	15
图 6 多头注意力机制示意图	16
图 7 层内注意力机制结构示意图	27
图 8 DHAN 网络结构示意图	33
图 9(a) CiteULike 上 Recall@K 的结果	45
图 9(b) Book-Crossing 上 Recall@K 的结果	45
图 9(c) MovieLens 上 Recall@K 的结果	46
图 9(d) Last.FM 上 Recall@K 的结果	46
图 10(a) CiteULike 上 DHAN 的 AUC 随向量维度变化	49
图 10(b) Book-Crossing 上 DHAN 的 AUC 随向量维度变化	49
图 10(c) MovieLens 上 DHAN 的 AUC 随向量维度变化	50
图 10(d) Last.FM 上 DHAN 的 AUC 随向量维度变化	50

表 清 单

表 1	基于分离式表征的信息传播方法中定义的数学符号	20
表 2	基于分层注意力机制的信息聚合方法中定义的数学符号	26
表 3	DHAN 中定义的数学符号	32
表 4	数据集描述	40
表 5	服务器硬件与软件配置	41
表 6	各算法在点击率预估场景下的 AUC 和 F1 指标对比	44
表 7	DHAN 变体模型在不同数据集下的 AUC 结果	47
表 8	DHAN 不同因素个数对 AUC 结果的影响	48
表 9	DHAN 不同信息传播深度对 AUC 结果的影响	48

第一章 绪论

1.1 论文背景和意义

随着科学技术的飞速发展，当下的科技资源（主要包括科研机构、科研成果、科技文献等），在时时刻刻产生，并以多种多样的形式存在。这类数据由于符合了海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征，所以又被称作科技大数据^[1]。

科技大数据给科研工作者提供了丰富的资源和便利的获取途径，但也带来了信息检索的挑战。通常情况下，科研人员想要获取感兴趣的科技文献，往往是在大型文献检索库（例如 IEEE Xplore、ACM 等数字图书馆）中通过关键词进行查询，然后再从结果中选择需要的文献。这样的方式只适用于科技工作者有明确的文献查找需求的情况，如果科研人员并不能精准了解自己想阅读的科技文献所使用的关键词时，搜索的结果可能并不会符合需求。除了搜索，推荐系统也是一种信息检索的方式。它的主动性相比于关键词搜索更强，它不需要科研人员给出关键词，而是根据其历史浏览文献等行为，分析研究偏好，进而推荐其可能感兴趣的科技成果。推荐系统大大节省了科技工作者寻找感兴趣的科技资源的时间，缓解了信息爆炸带来的检索问题。

科技大数据（例如科技文献、专家学者、课题项目等）彼此相连，共同组成了大数据网络。在这样的信息网络中，由于包含多种类型的节点和边，所以统称为异质信息网络。传统的推荐方法（如协同过滤系列方法）只考虑了用户的行为，因此在推荐时丢失了外部丰富的语义信息，导致推荐效率不高。利用异质信息网络辅助推荐，可以从科研人员与文献的历史交互行为（例如浏览、下载、收藏等）中，挖掘更深层次的用户偏好，得到更加鲜明的用户画像，真正实现个性化的推荐服务。

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。

本文研究基于异质信息网络传播机制的推荐算法，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 推荐系统研究现状

推荐系统是当今流行的一种信息检索工具，它的服务对象统称为用户。在推荐系统中，用户一般包含 ID、年龄和性别等静态信息，此外还有近期浏览记录和社交关系等动态信息。推荐给用户的对象统称为物品。物品可以是实物商品，也可以是虚拟服务，例如论文、歌曲和影片等。

上世纪 90 年代，推荐系统开始受到研究者们的关注。2006 年，美国流媒体播放平台 Netflix 放出用户对电影评分的数据集，奖金 100 万美元，鼓励各支参赛队伍研究推荐算法，要求在给定数据集上将推荐准确度提升 10%及以上。这一开创性的竞赛，大大促进了推荐算法研究。此后，美国计算机协会 ACM 从 2007 年开始一直举办推荐系统的顶级会议 ResSys，备受学术界和工业界关注。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，SIGIR、WWW、ICDM 等知识发现、数据挖掘领域的重要国际会议每年有大量推荐系统相关研究成果发表。

推荐系统发展到今天，已经有很多算法被提出。按照方法的不同，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中的“内容”指的是物品属性。这类方法的核心思想是挖掘用户曾经交互过的物品间的共有属性，从而提取用户属性，然后根据用户属性为他推荐相关物品。它包含三个过程：第一个过程是给物品打标签；第二个过程一般是利用统计的方法，分析用户交互过的物品，得出用户标签；最后一个过程是对于待推荐物品的标签和用户标签，计算两者的相似度，并以此为依据，向用户推荐特定物品集合。基于内容的推荐系统主要依赖于物品的标签体系，它可以有效解决物品冷启动问题，并且对数据稀疏问题不敏感。但是给物品打标签是一项需要人工高度干预的工作，并且费时费力。此外，物品标签质量直接决定最后的推荐效果。

(2) 协同过滤的含义指的是推荐系统中存在一些集体智慧：一般品味相似的人，在物品选择上会有相似的行为。基于协同过滤的推荐系统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。

基于物品的协同过滤方法的核心思想与基于内容的推荐算法类似，但是基于物品的协同过滤方法不需要依赖物品标签体系，而是根据物品被交互的历史数据来计算物品间的相似度。这种方法的优点在于计算简单，适合离线计算，但是它对于没有历史交互记录的新物品不友好。

与基于物品的协同过滤方法不同，基于用户的协同过滤方法是从用户角度出发，为用户寻找其相似用户，然后根据相似用户的交互记录，为用户推荐他也可能感兴趣的物品。这类方法不需要对商品特征进行挖掘，与“协同过滤”的本质思想很接近，但是它存在用户冷启动的问题。

基于模型的协同过滤方法是利用统计或机器学习方法，从用户-物品交互数据中学习一个概率模型，预测用户与物品产生交互的概率。这种方法简单有效，但是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将多种推荐技术混合而成的推荐算法，其中混合的方法包括多模型的融合（多为基于协同过滤的方法和其他方法的融合）、推荐结果的加权融合、不同数据源的特征组合等。它可以对不同的推荐场景进行个性化建模，充分利用场景中的各类信息，例如时间信息、上下文信息和知识信息等。

1.2.2 基于异质信息网络的推荐算法研究现状

目前主流的推荐系统研究方法通常会引入外部信息网络辅助推荐决策。外部信息网络常见的表现形式有知识图谱、事件图谱和社交网络等，它们与用户-物品交互网络共同构成异质信息网络。异质信息网络中的每一个节点都不是孤立的，在经过特定关系的连接后，一个节点会顺着关系走到其他节点。例如当推荐系统中被推荐的物品与外部信息网络进行节点匹配后，一个物品会对应外部信息网络中的一个或多个节点，因此可以说外部信息网络中的节点包含了被推荐的物品的外部信息。目前，基于异质信息网络的推荐系统受到越来越多研究者的关注，在该领域研究发展中形成了两类方法：基于路径的方法和基于传播的方法。

(1) 基于路径的方法

从图 1 可以看到，用户小明到文献《协同知识感知注意力网络》存在多条可能的路径。基于路径的方法就是要挖掘节点间存在的推荐路径。示例图中可以挖掘出其中一条

推荐路径：用户-文献-关键词-文献。这种定义在网络模式上，链接两个节点的复合关系就称为元路径。

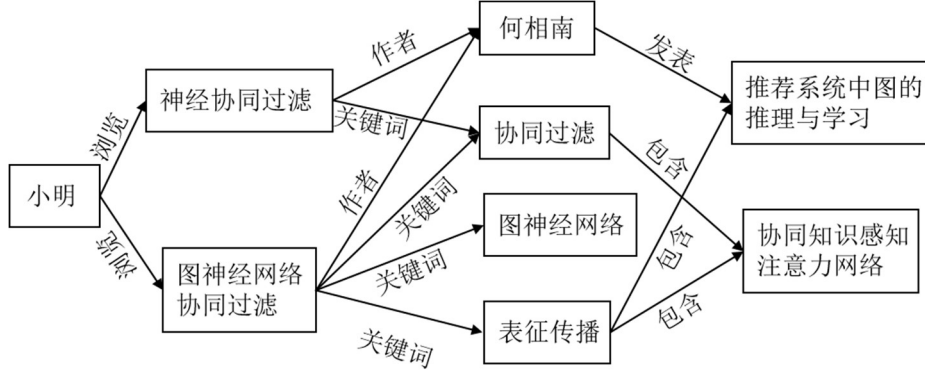


图 1 异质信息网络示例图

Yu 等人^[3]提出一种在异质信息网络中基于元路径的推荐方法（Personalized Entity Recommendation, PER）。PER 通过预先手动设计用户和物品在异质信息网络中的关联路径，然后利用 PathSim^[4]，一种在异构信息图中基于元路径的节点相似度衡量算法，得到用户 u_i 到物品 e_j 之间的用户偏好特征值：

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

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，是取值为 1，否则为 0。

利用用户和物品之间的用户偏好特征矩阵进行矩阵分解^[5]，可以得到元路径的特征向量，最后根据元路径对应的用户和物品向量，计算得到推荐结果。

由于是依赖于元路径，该方法有一定的可解释性。但是元路径的抽取需要领域知识或经验，这是一项耗时耗力的工作，并且元路径的选择直接决定了最终的推荐效果。当推荐场景或异质信息网络发生变化时，元路径也需要被重新被定义。

（2）基于传播的方法

异质信息网络节点间的关系错综复杂，不可能穷举出其中存在的所有推荐元路径。因此，出现了一些方法，考虑了用户和物品在异质信息网络中对应的邻接节点，并将它们继续传播到高阶邻接节点，以更充分地利用异质信息网络中的附加信息。这样的方法称为基于传播的推荐方法，也称为基于图神经网络的推荐方法。

基于传播的方法以向量形式对异质信息网络中的节点和关系进行数值建模，并充分

利用异质信息网络中的结构信息和节点、关系的语义信息。在这类方法中，信息传播机制的设计是研究的核心。Wang 等人^[6]提出涟漪网络 (RippleNet)，这是一个利用异质信息网络实现推荐的端到端框架，它通过将用户交互过的物品在异质信息网络中不断传播，探索用户潜在兴趣。具体来说，RippleNet 借鉴水波的传播，以用户感兴趣的物品为种子，并以种子集合为中心，在异质信息网络中一圈一圈地向外扩散到异质信息网络中的其他节点上，这个过程称之为偏好传播。Wang 等人^[7]提出协同知识感知的注意力机制网络 (Collaborative Knowledge-Aware Attentive Network, CKAN)，CKAN 和 RippleNet 创建初始种子集合的思想一致，在用户侧以用户交互过的物品为种子，在物品侧通过两层协同过滤，以物品-用户-物品为路径得到代表物品的种子，然后通过注意力机制聚合种子邻居的方式得到最终的用户和物品表示，进而预估用户点击物品的概率。

目前的研究中，基于传播的方法是主流，但是目前大部分的研究工作中，关于传播机制的研究还不完善。例如在 RippleNet 中，传播只在用户侧进行；CKAN 算法在传播的过程中，信息的建模是粗粒度的；在信息聚合时，CKAN 没有考虑信息衰减的因素。

1.3 论文研究目标及内容

1.3.1 研究目标

本文针对细粒度的信息传播和高效的信息聚合问题，研究异质信息网络中的信息传播机制，提出一种基于异质信息网络传播机制的推荐算法。通过该算法，扩展推荐系统中用户的潜在兴趣，挖掘物品的潜在特征，提高用户对特定物品点击率预估的准确率。

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 所示中。首先调研、获取推荐相关数据集以及对应的外部数据，构成异质信息网络数据。其次，调研基于异质信息网络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，深入研究分离式表征，注意力机制一系列先进的深度学习技术，设计基于分离式表征的信息传播方法，利用分离式表征技术和图神经网络技术对异质信息进行细粒度传播建模。然后，对于获取到的多阶信息嵌入集合，采用基于分层注意力机制的信息聚合方法，使得信息聚合阶段能够感知同一阶内和不同阶间信息重要性的差异。接下来，整合信息传播和信息聚合两阶段的关键方案，设计一个端到端的基于异质信息网络的推荐算法。最

后在论文、数据、电影和音乐四个不同领域的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。

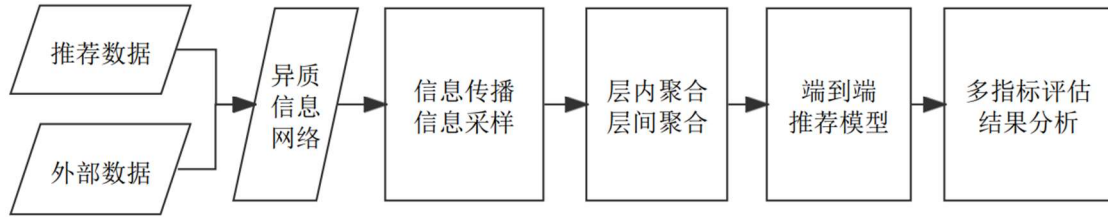


图2 研究路线示意图

1.3.3 研究内容

为了达到研究目标，本文重点研究信息传播与信息聚合两个重要阶段，并形成端到端的推荐算法。本文的主要研究内容包括以下几个方面：

（1）信息传播阶段的设计与建模

随着图神经网络的发展，目前主流的基于异质信息网络的推荐方法都聚焦在图神经网络的应用上。应用图神经网络建模信息传播，主要依靠从图中节点出发传播邻居信息。而信息不止单一的维度，就如用户购买物品，不止处于一个意图；物品除了本身的物理属性外，还有协同属性。本文的主要研究之一是如何将信息解耦，即细粒度刻画信息，并融入当前的图神经网络建模方案中。由于信息传播过程中，总会伴有与原始节点信息关联度很低的噪声信息的影响。因此，信息传播阶段还要设计一种基于信息间关系连接的自适应采样方法。

（2）信息聚合阶段的设计与建模

在空域图神经网络方法中，邻居有远近之分，可以分为低阶邻居和高阶邻居，因此信息也可以分为低阶邻域信息和高阶邻域信息。一般情况下，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。在信息聚合阶段，不仅要考虑每个节点周围邻居节点不同的重要性，也要考虑不同距离的邻居聚合的不同影响力，即在信息聚合阶段建模信息传播过程中的衰减现象。

（3）端到端推荐算法的提出与验证

在上述两个关键方法设计的基础上，本文最重要的研究内容是提出一种基于异质信息网络传播机制的推荐算法。该算法以信息传播的形式对用户和物品进行建模，充分利用它们在异质信息网络中的语义和结构信息，并进行端到端的模型训练，得到精确的用

户和物品的向量数值表示，从而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不同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

本文对基于异质信息网络传播机制的推荐算法进行研究，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异质信息网络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。首先介绍图表示学习的相关技术与方法，然后介绍分离式表征学习的基础概念及其在推荐系统领域的应用，最后介绍注意力机制及其在推荐系统领域的应用。

第三章首先介绍信息传播的相关概念和一般流程，并指出现有方法存在的问题，然后详细介绍针对现有问题而改进提出的基于分离式表征的信息传播方法。

第四章首先明确信息聚合的相关概念，接着介绍信息聚合的一般方法和它们存在的问题，最后对提出的基于分层注意力机制的信息聚合方法进行详细介绍。

第五章对本文提出的基于分离式表征与分层注意力机制的推荐网络 DHAN 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。

第六章是实验部分，先介绍实验的基本设置、数据集、选取的对比方法进行介绍，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，验证所提出的算法的有效性，最后深入分析 DHAN 模型各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，首先详细介绍与本文核心方法相关的图表示学习算法，并分类描述其在推荐系统中的主流应用，接着介绍本文使用的分离式表征和注意力机制相关的基础概念，并介绍它们在推荐系统中的应用情况。

2.1 异质信息网络

异质信息网络是一种特殊形式的图，它包含多种类型的节点和边，它可以形式化地表示为 $G = \{V, \mathcal{E}\}$ ，其中 V 表示节点集合， \mathcal{E} 表示边集合。异质信息网络中还存在一个节点映射函数 $\phi: V \rightarrow A$ 和边映射函数 $\varphi: \mathcal{E} \rightarrow R$ 。其中 A 是预定义的节点类型集合， R 是边类型集合，他们之间满足关系式 $|A| + |R| > 2$ 。

在异质信息网络中，任意两个节点之间可以通过不同的语义关联路径相连，这样的路径称为元路径，它可以形式化地定义为 $A_1 (R_1) \rightarrow A_2 (R_2) \rightarrow \dots A_l (R_l) \rightarrow A_{l+1}$ ， $R = R_1 \circ R_2 \dots \circ R_l$ 称为 A_1 到 A_{l+1} 之间的复合关系。 A_2 由于距离 A_1 只有一个单位关系的间隔，因此 A_2 称为 A_1 的一阶邻居， A_{l+1} 为第 l 阶邻居。

图 3 给出了一个异质信息网络中关联路径的示例。 $\{i_1, i_2, i_3, i_4\}$ 是 u_2 的一阶邻居聚合。用户 u_2 到物品 i_6 可以通过路径 $u_2 (r_1) \rightarrow i_1 (r_3) \rightarrow i_6$ ，路径 $u_2 (r_3) \rightarrow i_3 (r_5) \rightarrow i_6$ 和路径 $u_2 (r_4) \rightarrow i_4 (r_6) \rightarrow i_6$ 相连接。

2.2 图表示学习

2.2.1 图表示学习的介绍

现实世界中，很多数据都是以图的结构形式展现，例如学术引文网络、社交网络、知识图谱和事件图谱等。在图相关的研究中，有很多下游任务（例如图中节点的分类，节点之间连边的预测等等）都需要依赖图表示学习这一基础任务来完成。

图表示学习是将图中各个节点表示成一个低维向量，经过训练后，图中相似的节点在低维向量空间中相接近，其中对于“相似”和“接近”如何定义，在各个方法中都不一样。目前，图表示学习技术主要分为图嵌入、知识图谱嵌入和图神经网络三大类。

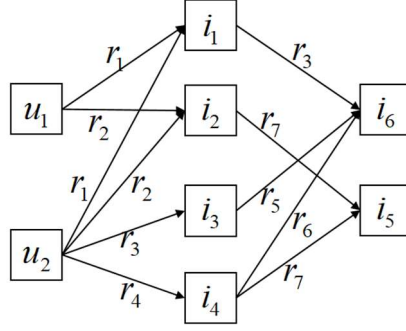


图 3 异质信息网络中的关联路径示例图

(1) 图嵌入

图嵌入技术最早是从词嵌入技术 Word2Vec^[8]发展而来。Word2Vec 工具技术在 2013 年由谷歌开源，它通过在语料库中，学习句子序列中词与词的先后关系和共现关系，训练得到词的低维向量表示。在 Word2Vec 开源以后，DeepWalk^[9]、Node2Vec^[10]等图嵌入方法被提出，它们都从 Word2Vec 中获得启发。DeepWalk 和 Node2Vec 算法都需要先在图中进行随机游走，采样得到特定长度的节点序列，这个序列就类比 Word2Vec 中的句子，节点类比于词。在随机游走生成若干条序列样本后，利用 Word2Vec 的 skip-gram 模式就可以训练得到图中各个节点的向量表示。

这两个图嵌入算法的不同之处主要体现在随机游走的方式上。DeepWalk 中的随机游走是一个深度优先遍历（Deep First Search, DFS）的过程，它可以重复访问已经游走过的节点。Node2Vec 改进了 DeepWalk 的随机游走过程，它同时考虑 DFS 和广度优先遍历（Breath First Search, BFS）。具体来说，Node2Vec 在随机游走时，不再是随机选择，而是根据概率选择。对于给定当前节点 v ，访问下一个节点 x 的概率为：

$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{if } (v, x) \in E \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

其中， π_{vx} 是从节点 v 到 x 的访问概率， Z 为归一化系数， E 是边集合。

为了控制游走时的 DFS 和 BFS 策略，Node2Vec 引入两个超参数 p 和 q 。假设当前随机游走经过边 (t, v) 到达节点 v ，公式(2.1)中的 π_{vx} 就可以写成：

$$\pi_{vx} = \alpha_{pq}(t, x) \cdot \omega_{vx} \quad (2.2)$$

其中， ω_{vx} 是节点 v 和 x 之间边的权重， $\alpha_{pq}(t, x)$ 为节点之间的转移概率，用公式表示为：

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases} \quad (2.3)$$

其中, d_{tx} 为节点 t 到节点 x 之间的最短路径。

(2) 知识图谱嵌入

大部分的图嵌入方法只考虑节点之间的相似性, 而弱化了对图中关系的建模, 甚至忽视了关系。在知识图谱中关系是描述节点间相似性和相关度的重要信息, 它可以为嵌入学习提供丰富的语义、结构信息。因此, 知识图谱嵌入脱胎于图嵌入技术, 独立发展起来。

在知识图谱嵌入方法中, 基于翻译的知识图谱嵌入方法是最早被提出的一类方法。该系列方法的核心思想是: 头实体向量 \mathbf{h} 和关系向量 \mathbf{r} , 二者之和应当接近于尾实体向量 \mathbf{t} , 即 $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ 。一般用 L1 或者 L2 范数来衡量“接近”。TransE^[11]是最经典的基于翻译的方法, 它在 2013 年被提出。TransE 通过替换实际存在的三元组中的头或尾实体, 用随机负采样的一个实体构造成负样本, 损失函数的形式为:

$$L(\mathbf{h}, \mathbf{r}, \mathbf{t}) = \max(0, d_{pos} - d_{neg} + margin) \quad (2.4)$$

其中 d_{pos} 和 d_{neg} 分别表示正、负样本三元组, $d = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$ 为 L1 或 L2 范数, $margin$ 是一个常数, 表示正负样本的最大距离。在此之后, 衍生出了一系列基于 TransE 的 Trans 系列方法, 例如 TransH^[12], TransR^[13] 和 TransD^[14] 等。

除了基于翻译的知识图谱嵌入方法, 还有基于张量分解和基于神经网络的知识图谱嵌入方法。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, 这类方法直接使用多层的神经网络拟合知识图谱的结构, 输入是实体-关系三元组 $(\mathbf{h}, \mathbf{r}, \mathbf{t})$, 输出是该三元组存在的概率。这一类型的方法有 SME^[15]、NTN^[16] 和 MLP^[17] 等。

NTN 整合和拓展了各种早期的神经网络模型, 提出了神经张量网络。NTN 替换了神经网络模型中的一个线性层为一个双线性张量层, 从多个维度关联两个实体向量, 结构示意图如图 4 所示。

NTN 中计算两个实体间存在特定关系概率的得分函数为:

$$f_r(\mathbf{h}, \mathbf{t}) = \mathbf{r}^T \tanh(\mathbf{h}^T \mathbf{M}_r \mathbf{t} + \mathbf{M}_r^1 \mathbf{h} + \mathbf{M}_r^2 \mathbf{t} + b_r) \quad (2.5)$$

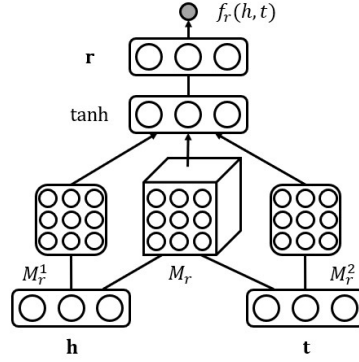


图4 NTN方法结构示意图

其中 M_r 是一个三维张量，对应于知识图谱中的一条关系。比起用二维矩阵去表示一条关系，该三维张量能进一步地从不同语义层面上进行表示，并且从形式上来看，它能够捕获实体和关系之间存在的双线性关系。 $M_r^1 h + M_r^2 t + b_r$ 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。由于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，在 NTN 方法中对于实体和关系表示向量的初始化，还使用预先训练好的实体词向量的平均值作为实体与关系的初始表示，对模型的效果有了进一步的提升。

（3）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。

早期的图神经网络通过循环神经网络不断迭代传递邻域信息，学习得到节点表示。然而循环神经网络的计算量很大，且计算机不易并行处理，因此训练需要花费大量时间。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。复杂的图神经网络和图像的最大不同点在于：图网络在欧几里得空间中是不规则的，每一个节点周围的邻居节点数量并不相同，而在图像中实现卷积操作是很容易的事情，于是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，它从信号处理的角度引入滤波器来定义图卷积。具体地，先将欧式空间下的图信号投影变换到谱空间中，然后利用空间平移不变性和傅里叶变换实现谱空间下的卷积操作。然而，基于谱的图卷积神经网络需要将整个图加载到内存中才能执行图卷积，当图过大时，这个方法就遇到了瓶颈。于是研究推进到了基于空间的图卷积神

神经网络^[21]，它基于节点的空间关系定义图卷积。假设有一组图数据，其中有 N 个节点，每个节点都有所属的特征，它们组成一个 $N \times D$ 维的矩阵 X ，各个节点之间的关系可以组成一个邻接矩阵 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：

$$H^t = \sigma \left(D^{-\frac{1}{2}} A' D^{-\frac{1}{2}} H^{t-1} W^{t-1} \right) \quad (2.6)$$

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。

图神经网络受到注意力机制影响，将注意力机制应用在了图结构数据中，由此有了图注意力网络（Graph Attention Networks, GAT）^[22]。关于注意力机制的详细介绍在 2.4 节，这里可以将注意力机制理解为：在聚合信息的过程中确定邻居节点的重要性，它是一个权重值。GAT 中，图卷积操作表示为：

$$h_i^t = \sigma \left(\sum_{j \in N_i} \alpha(h_i^{t-1}, h_j^{t-1}) W^{t-1} h_j^{t-1} \right) \quad (2.7)$$

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。

除了 GCN 和 GAT 外，图神经网络在发展中还出现了图自编码器、图生成网络和图时空网络等方法。目前，图神经网络的概念变得越来越宽泛，应用的范围也越来越广，图嵌入和知识图谱嵌入中也有图神经网络的影子。

2.2.2 图表示学习在推荐中的应用

随着图表示学习在近年来取得了一系列突破，在其下游的推荐任务中，利用图表示学习进行推荐的研究工作也越来越多。推荐系统中的图有用户-物品交互二分图，也有包含外部信息（例如知识图谱，社交网络等）的辅助信息网络。

基于图表示学习的推荐算法主要利用图表示学习技术更好得学习推荐系统中的用户偏好和物品表征。

神经图协同过滤（Neural Graph Collaborative Filtering, NGCF）^[23]是最早将图神经网络应用在用户-物品二分图上的推荐算法。NGCF 通过在二分图上利用 GCN，传播用户和物品的嵌入表征，以此优化它们随机初始化的向量表示。由于将深层协同信息编码进向量表示中，NGCF 相比于传统的推荐算法取得了更优异的推荐效果。此后，基于 GCN 改进模型的 LightGCN^[24]算法被提出。LightGCN 是 NGCF 的简化版，它舍弃了 GCN 中的特征变换和非线性变换部分，只包含最基本的邻域传播和聚合部分，这样的简化操作，

使得模型更容易被训练，学习得到的用户和物品表征也更稳定，在相同的实验设置下，它比 NGCF 的推荐效果又有了较大的改进。

知识图谱，社交网络都属于外部信息网络。知识图谱经常作为外部知识引入到推荐系统中，目前主流的方法之一是基于知识图谱的图注意力网络（Knowledge Graph Attention Network, KGAT）^[25]，它利用了知识图谱嵌入和图注意力网络两种图表示学习方法。具体地，它将知识图谱和用户-物品交互图融合为一张异质图，取名为协同知识图（Collaborative Knowledge Graph, CKG），此图融合了用户-物品的交互信息以及知识图谱中的节点-关系信息。KGAT 首先利用知识图谱嵌入方法 TransR 学习图中节点和关系的低维向量。然后采用图神经网络框架，以三层结构：信息传播层、基于知识的注意力层和信息聚合层，实现了高效的信息传播和聚合，得到最终用户、物品和实体的向量表示。最后利用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。

在基于社交网络的推荐系统中，具有代表性的是基于图神经网络的图模型推荐算法 GraphRec^[26]，它主要解决的是用户对物品的评分预测问题。GraphRec 的重点在于用户和物品建模。在用户建模部分，因为用户既处于评分空间，又属于社交空间中，所以用户的表示应该在这两个空间中进行融合表征。物品建模部分的思想 and 用户建模部分相似，物品在评分空间中有关联的用户，用户又处于社交空间中，因此同样要进行融合表征。融合表征的方法通过图神经网络框架来完成，通过向量的深层传播与聚合得到最终的用户和物品表征。最后通过一个多层感知机拟合用户对物品的打分，进行有监督的学习。

在独立发展出来的基于异质信息网络的推荐算法中，图表示学习一直都是方法的核心。基于异质信息网络的推荐模型（Heterogeneous Information Network, HINRec）^[27]使用随机游走策略生成节点序列，对不同的元路径都会学习其向量表示，通过融和多个元路径的向量表示，最终得到目标节点的嵌入表征。HINRec 将得到的融合向量表示和传统的矩阵分解结合做评分预测任务，最终生成推荐列表。

2.3 分离式表征

2.3.1 分离式表征的介绍

分离式表征的最早概念是在 2013 年由 Bengio 等人^[28]提出，它是表示学习中的一种方法。数据的分布不是稳定不变的，但是变化的数据中总是包含一些不变的隐因子，分

离式表征的核心思想是分离出变化数据中的不变隐因子，并把这些隐因子用向量形式表示出来。经过分离式表征后的向量表示是彼此独立的，各个隐因子之间互不干扰，各向量表示的取值只与对应的隐因子有关。分离式表征作用于下游任务，需要根据下游任务进行监督或半监督的学习。同时，分离式表征对数据也有一定要求，它需要数据场景满足一定可解释的假设，它在小样本任务中表现不俗。

2.3.2 分离式表征在推荐中的应用

在推荐系统中，传统的推荐算法假定了所有用户和物品之间的交互关系是无差别的，而用户对物品的交互意图本是多种多样的，这就导致学习到的用户嵌入表征只蕴含着粗粒度的用户和物品交互意图，即意图是纠缠在一起、未解开的，模型并没有显式地去建模和挖掘这种细粒度的意图，因此学习到的用户嵌入表示是次优的。

基于分离式表征的图协同过滤方法（Disentangled Graph Collaborative Filtering, DGCF）^[29]提出的目的是希望去挖掘和解开用户的意图，使得学习到的表征能够表达出用户多种多样的意图。DGCF 建模用户-物品交互关系中用户的隐意图分布，并将隐意图映射到向量表示上，从而形成分离式表征。此外，DGCF 在图神经网络的基础上应用分离式表征。图神经网络的方法考虑了节点与其相连接的邻居节点间的相互影响，DGCF 在此基础上认为节点与节点相连是受一些潜在因素影响的，因此可以通过分离式表征建模用户多种多样的交互意图。

2.4 注意力机制

2.4.1 注意力机制的介绍

注意力机制可以用人的视觉系统形象地描述：当人们看到一张图片时，注意力通常会聚焦于画面的主体部分，然后才会关注画面中的边缘部分。这意味着人们在关注信息时，会优先关注信息的重点，主动忽视无关的部分。深度学习中的注意力机制，就是用来抽象这一现象的具体方法。注意力机制目前广泛应用于自然语言处理、计算机视觉和推荐系统中。

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。在注意力机制还没有出现之前，通常会采用循环神经网络（Recurrent Neural Network, RNN）^[30]架构，将文字，图片等用序列的形式处理，利用 RNN 从前往后或者从后往前

进行信息传播，信息在传播过程中，会自动忽视长远的不重要的信息。然而 RNN 的主要问题在于它很难建模长序列信息，容易出现梯度消失的现象。此外，它只能顺序计算，限制了其并行训练的能力。Transformer^[31]突破了 RNN 无法建模长序列和无法并行计算的问题，它第一次提出用注意力机制并行提取信息重点部分，学习重点与非重点信息不同权重，并尽可能地保留序列中的全部原有信息。

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。点乘注意力机制的结构如图 5 所示，它也是最常见的一种结构。

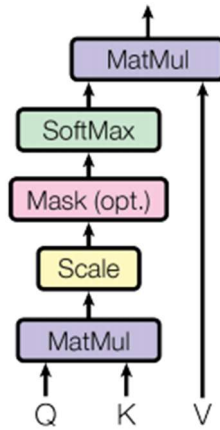


图 5 点乘注意力机制示意图^[31]

图中的 Q、K 和 V 分别代表 Query, Key 和 Value，它们在不同场景下有不同的含义。例如在搜索场景下，Q 代表搜索词向量，K 代表搜索结果向量，V 是数值向量。点乘注意力机制分为两个过程，第一个过程是根据 Q 和 K 计算权重，第二个过程是根据权重对数值向量进行加权求和。在第一个过程中，需要先根据 Q 和 K 计算它们的相似性或者相关性，点乘注意力机制的“点乘”，就代表的是利用点乘的方式计算。接着，就是对原始相似度或相关性数值进行归一化处理。这个过程可以用公式表示为：

$$a_i = \frac{q_i k_i}{\sum_{j=1}^l q_j k_j} v_i \quad (2.8)$$

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。

多头注意力机制指的是堆叠多个点乘注意力机制，图 6 展示了多头注意力机制的示意图。它先通过若干组不同的线性映射矩阵，得到若干个不同 Q、K 和 V 矩阵；然后分别送入各自的点乘注意力机制中；拼接结果后，最后通过线性映射得到最终的输出。

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^[32]，共同注意力机制^[33]等。

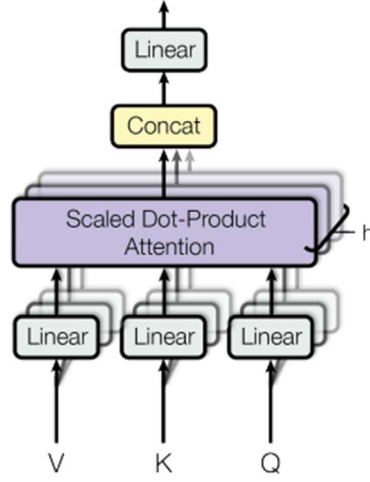


图 6 多头注意力机制示意图^[31]

2.4.2 注意力机制在推荐中的应用

注意力机制最先应用在自然语言处理领域，随后又在推荐系统领域的研究中被广泛应用。主要原因在于推荐系统和自然语言一样，也存在一种规律性的序列，即用户的浏览记录序列。目前，注意力机制在推荐中的应用也集中在序列建模上。

阿里巴巴的研究团队提出深度兴趣网络（Deep Interest Network, DIN）^[34]，DIN 应用注意力机制，从用户的历史浏览序列中自适应提取用户的兴趣向量表示。对于不同的候选物品，得到的用户兴趣的向量不同。具体来说，DIN 以候选物品 c 的向量表示 V_c 作为 Query，用户历史序列中的物品 i 的向量表示 V_i 作为 Key 和 Value，学习得到各物品的权重，并通过加权求和池化的方式得到用户的兴趣表示，用公式表示为：

$$V_u = \sum_{i=1}^n g(V_i, V_c) V_i \quad (2.9)$$

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。和传统点式注意力机制不同的是，DIN 中没有对权重进行归一化的过程。

微软在 2019 年提出基于个性化注意力机制的新闻推荐算法（Neural News Recommendation with Personalized Attention, NPA）^[35]。NPA 中利用两个编码器实现高效推荐：新闻编码器和用户编码器。在新闻编码器中，由于新闻的信息具有多样性，比如标题、正文和标签等，因此 NPA 把它们视为不同的视角，提出了视角级注意力机制，用

来提取新闻的表征。用户编码器中，不同的用户会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，提取用户个性化表征。最后计算内积，得到用户对新闻的个性化偏好得分。

在基于异质信息网络的推荐方法中，也有注意力机制的应用。基于知识图谱的图卷积网络（Knowledge Graph Convolutional Network, KGCN）^[36]从物品侧沿着图中的关系进行信息传播，在考虑邻居节点时，KGCN 以用户特征、关系特征和目标节点特征融合作为 Query，邻居节点特征作为 Key 和 Value，采用了点式注意力机制。物品侧的个性化建模大大提升了推荐效果。

目前，注意力机制渗透了推荐系统研究领域的各个环节，已经成为了一项重要技术。

2.5 本章小结

本章介绍了本文涉及的相关理论基础，为后续章节奠定基础。首先详细介绍了与本文核心方法相关的各种图表示学习算法以及其在推荐系统中的应用。接着，对本文所使用的分离式表征学习相关概念和在推荐系统中的应用发展进行了介绍。最后，介绍了本文用到的注意力机制相关理论和应用发展情况。

第三章 基于分离式表征的信息传播方法

在基于异质信息网络传播机制的推荐算法研究中，信息传播方法是利用图神经网络实现邻域传播的通用方法之一，也是研究的核心之一。信息传播方法设计的主要目的是充分利用异质信息网络中的结构信息和语义信息，辅助推荐任务。设计有效的信息传播方法，能够使得获取的信息表征刻画得更精准，表征的含义更丰富。本章首先介绍信息传播的概念，并分析现有方法的问题，然后详细介绍基于分离式表征的信息传播方法。

3.1 信息传播概念与现有问题分析

3.1.1 信息传播的概念

信息传播指的是信息在图神经网络中的传递以及特征的变换。在本文基于异质信息网络的推荐系统背景下，信息传播指的是目标节点在异质信息网络中，借助节点间的关联关系，获取其对应的一阶或多阶邻域节点集合，然后利用图神经网络，对邻域节点集合进行特征变换，并用向量的形式表示。

3.1.2 信息传播的一般流程

信息传播主要分为：节点匹配、邻域采样、高阶传播和信息转换四个步骤。

（1）节点匹配

推荐系统中存在用户-物品交互二分图，想要利用外部信息网络辅助推荐，首先需要在外部的信息网络中找到对应的节点，这个过程称为节点匹配。以物品举例，通常物品与外部的知识图谱相连，中间需要经过实体链接的过程，节点匹配和这个过程类似。例如论文《基于异质信息网络的论文推荐系统》，首先进行分词，然后利用实体链接技术匹配出知识图谱中的实体，即节点“异质信息网络”、“论文”和“推荐系统”。目前节点匹配的方法有基于图方法^[37]、基于概率生成模型^[38]、基于主题模型^[39]等。

（2）邻域采样

在匹配节点后，用户-物品交互二分图就和外部信息网络产生连接，组成异质信息网络。图神经网络的思想是将目标节点的邻域考虑进来，以此扩充目标节点的向量表示。在图神经网络框架下，需要获取邻域节点集合和边集合。一般集合是由三元组组成，三元组包含节点-边-节点，为了方便描述，分别称它们为头节点、关系和尾节点。然而对

于不同的节点，其邻域三元组集合中的三元组数量并不一致，因此常会对邻域集合做采样。这么做一方面可以解决上述问题，方便模型分批次训练以及更新参数。另一方面，通过采样可以有效缓解邻域节点中的噪声问题。一般采样的方式^[40]包括：随机采样、基于元路径的采样和基于相似度计算的采样等。

（3）高阶传播

高阶传播指的是从目标节点或每一阶邻域节点集合开始，继续向深处传播的操作。具体来说，它以目标节点或者邻域三元组集合中的尾节点为中心节点出发，寻找下一阶邻域节点和关系，得到高阶的邻域节点集合和三元组集合。在高阶传播过程中，也需要对邻域进行采样操作。

（4）信息转换

信息转换指的是将传播的低阶到高阶邻域节点集合转换成可计算的向量形式。其中，节点本身是不可以计算的，当把节点用低维向量的形式表示时，可以转换为可计算的信息向量。常见的转换方法有 One-Hot 独热编码、嵌入化和矩阵乘积变换等。

3.1.3 目前的方法存在的问题

在基于异质信息网络传播机制的推荐系统研究中，信息传播方法是研究的核心之一，但是目前的方法中，大多存在以下几个问题：

（1）没有从用户和物品双侧进行传播。在信息传播过程中，从用户、物品侧单侧进行传播是目前大多数基于异质信息网络的推荐系统研究方法选择的策略。这么做，一方面是为了防止模型的复杂度过高，加快模型的收敛速度。另一方面，由于用户或者物品只有其中一种直接与外部信息网络相连，例如知识图谱只连接物品，社交网络只连接用户，因此单侧传播更直接地利用异质信息网络特点。例如 Wang 等人^[36]提出基于图卷积神经网络的端到端的框架 KG-CNN，它只以物品为中心，挖掘物品在异质信息网络中的相关属性。但是，双侧传播的计算复杂度可以降低，设计有效的方法可以解决这个问题。此外，从双侧传播可以使得用户和物品都直接融合外部信息网络的辅助信息，更精准刻画用户和物品表征，提升推荐系统的效果。

（2）没有对邻居节点采用有效的采样策略。目前的方法大多以随机采样的方式，在异质信息网络传播过程中控制邻居节点个数。这样的处理方法带来两个方面的弊端：一方面，随机采样的方式使得邻居节点完全不受控制，给模型带来很大的随机性，进而影

响模型的训练。例如 RippleNet^[6]在以初始种子为起点向异质信息网络扩散时,采用随机的方式采样固定数量节点。因为引入太多随机性,导致推荐结果收敛很慢。另一方面,由于节点在传播过程中,沿着一些关系继续传播,对推荐结果并没有收益,因此传播到的关系和节点算作噪声信息,而随机采样必然会导致噪声信息增多,影响最终推荐效果。

(3) 没有细粒度建模信息。在目前主流的信息传播方案中,信息构造的方式是比较简单的,多为直接嵌入化处理,这样的信息表示是随机且耦合在一起的。而信息是多颗粒度的,在用户层面,用户与物品的交互意图多种多样,例如由于兴趣、生活必需等。在物品层面,一个物品有多个层面的信息维度,例如物品的价格因素、实用性因素等。

基于以上问题,本文提出一种基于分离式表征的信息传播方法。从传播方向看,它能为用户和物品两侧同时进行邻居节点扩充,提升辅助信息的利用效率。在高阶传播过程中,该方法提出一种基于关系连接的自适应采样方案,有效抑制信息在传播过程中遇到的噪声,提升采样的可靠性。在信息转换方面,该方案能够细粒度建模用户多意图和物品多属性。

3.2 基于分离式表征的信息传播方法的提出

3.2.1 问题描述

为了方便介绍本节的问题背景与提出的信息传播方法,表 1 总结了本节所有用到的数学符号及其对应描述。

表 1 基于分离式表征的信息传播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
h, r, t	头节点、关系和尾节点
E, R	异质信息网络中的节点、关系集合
$G = \{(h, r, t) h, t \in E, r \in R\}$	异质信息网络中的三元组集合
E_u^l	用户 u 第 l 阶的节点集合
E_v^l	物品 v 第 l 阶的节点集合
S_u^l	用户 u 第 l 阶的三元组集合
S_v^l	物品 v 第 l 阶的三元组集合
W	转置矩阵
$M_{u,k}^l$	用户 u 在 k 因素下第 l 阶的信息嵌入集合
$M_{v,k}^l$	物品 v 在 k 因素下第 l 阶的信息嵌入集合

本节提到的异质信息网络由用户-物品交互二分图和与用户或物品相关的外部异质信息网络共同组成。信息传播指的是从用户或物品节点开始传播，在异质信息网络中寻找它们的邻域节点集合。随着传播向高阶进行，会得到从低阶到高阶的邻域节点集合，最终经过信息转换后，各集合中的节点都会得到其低维向量表征。分离式表征是对向量表征的细粒度刻画，每个节点都有其建模因素下的向量表示。以用户 u 举例，最终会获得其在 k 因素下第 l 阶的信息嵌入集合 $M_{u,k}^l$ 。

3.2.2 信息传播方法

本文提出基于分离式表征的信息传播方法，它分为两部分：邻域节点传播和细粒度信息建模。下面对它们进行详细介绍。

(1) 邻域节点传播

异质信息网络由三元组构成，它可以表示为 $G = \{(h, r, t) | h, t \in E, r \in R\}$ ，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 。邻域节点传播是从特定节点出发（一般为用户或物品节点），沿着异质信息网络中的关系传播，得到不同距离的邻域节点集合和邻域三元组集合。

以用户为例，从异质信息网络中用户 u 节点出发，沿着关系可以传播得到的第 l 阶的邻域节点集合，它表示为：

$$E_u^l = \{t | (h, r, t) \in G \text{ and } h \in E_u^{l-1}\} \quad (3.1)$$

其中， l 表示传播的距离。 E_u^{l-1} 是第 $l-1$ 阶的邻域节点集合， E_u^0 表示用户节点本身。

同理，传播得到用户 u 第 l 阶的邻域三元组集合表示为：

$$S_u^l = \{(h, r, t) | (h, r, t) \in G \text{ and } h \in E_u^{l-1}\} \quad (3.2)$$

类似地，从异质信息网络中物品 v 节点出发，获取的第 l 阶的邻域节点集合以及邻域三元组集合表示为：

$$E_v^l = \{t | (h, r, t) \in G \text{ and } h \in E_v^{l-1}\} \quad (3.3)$$

$$S_v^l = \{(h, r, t) | (h, r, t) \in G \text{ and } h \in E_v^{l-1}\} \quad (3.4)$$

由于每一阶邻域节点集合中包含的节点数并不一致，这会导致模型后续难以处理计算，因此需要对邻域节点集合进行采样。本文从基于图神经网络的推荐模型 GraphSAGE^[41]和 PinSAGE^[42]中获得启发，提出一种基于关系连接的自适应采样方法，它的采样原理是优先采样彼此间有关系连接的节点。具体过程如算法 1 所示。

算法 1 基于关系连接的自适应采样方法

输入：用户 u ，物品 v ，未采样的三元组集合 S_u^l 和 S_v^l ；

输出：采样后的目标集合 k_u 和 k_v ，大小为 n ；

for set s in $\{S_u^1, S_u^2, \dots, S_u^L, S_v^1, S_v^2, \dots, S_v^L\}$ do:

 for tail node $e_1 \in s$ do:

 for tail node $e_2 \in s$ do:

 if not e_1, e_2 is connected then:

 add s to $k \in \{k_u, k_v\}$;

 end if

 end for

 end for

end for

for set s in $\{S_u^1, S_u^2, \dots, S_u^L\}$ do:

 if $|s| \neq n$ then:

 random sampling to make $|s| = n$;

 end if

end for

for set s in $\{S_v^1, S_v^2, \dots, S_v^L\}$ do:

 if $|s| \neq n$ then:

 random sampling to make $|s| = n$;

 end if

end for

(2) 细粒度信息建模

以用户 u 为例，在得到第 l 阶的节点集合 E_u^l 后，可以对集合中的节点进行细粒度信息建模。细粒度建模主要采用基于分离式表征的方法，将节点包含的信息分为 k 个因素分别建模。具体地，先将节点集合 E_u^l 中每个节点 $\{e_{1,u}^l, e_{2,u}^l, \dots, e_{i,u}^l, \dots, e_{n,u}^l\}$ 进行向量化处理，然后用转置矩阵相乘，将它们置于同一信息空间中。以 $e_{i,u}^l$ 为例，常规的信息转换方法表示为：

$$m_{i,u}^l = W^T E(e_{i,u}^l) \quad (3.5)$$

其中 $m_{i,u}^l$ 为节点 $e_{i,u}^l$ 的信息表示， W 为转置矩阵， $E(\cdot)$ 为向量嵌入化操作。

若采用分离式表征方法，需要得到 K 个因素下的耦合信息表示，这里巧妙地设置 K 个转置矩阵 $\{W_1, W_2, \dots, W_K\}$ ，对应地可以得到 $e_{i,u}^l$ 在第 k 因素下的信息表示：

$$m_{i,u,k}^l = W_k^T E(e_{i,u}^l) \quad (3.6)$$

因此，用户 u 第 l 阶的信息嵌入集合表示为：

$$M_{u,k}^l = \{m_{1,u,k}^l, m_{2,u,k}^l, \dots, m_{n,u,k}^l\} \quad (3.7)$$

用同样的方法可以得到物品 v 第 l 阶的信息嵌入集合：

$$M_{v,k}^l = \{m_{1,v,k}^l, m_{2,v,k}^l, \dots, m_{n,v,k}^l\} \quad (3.8)$$

3.3 本章小结

本章详细介绍了本文提出的基于分离式表征的信息传播方法。首先明确了信息传播的概念，介绍了信息传播的一般流程。随后，指出了现有的信息传播方法存在的问题。最后，详细介绍了本文提出的基于分离式表征的信息传播方法以及基于关系连接的自适应采样方法。

第四章 基于分层注意力机制的信息聚合方法

和信息传播方法一样，信息聚合方法也是基于异质信息网络传播机制的推荐算法研究的核心之一。信息经过传播后，需要聚合更新异质信息网络中的目标节点。信息聚合方法与信息传播方法相辅相成，需要定制化设计，以发挥信息传播方法设计的优势。本章首先介绍信息聚合的概念，并对现有方法的问题进行分析，然后详细介绍基于分层注意力机制的信息聚合方法。

4.1 信息聚合概念与现有问题分析

4.1.1 信息聚合的概念

信息聚合是指将信息传播到的低阶或高阶的信息嵌入集合融合，并累积到目标节点上，得到目标节点最终的向量表示。从信息聚合的概念可知，信息聚合包含两个阶段：层内聚合和层间聚合。层内聚合作用在信息嵌入集合内，指的是将同一阶邻居节点的信息表示，融合为一个定长的向量表示。层间聚合是指将层内聚合得到的不同阶的向量表示，聚合到目标节点上。不管是层内聚合还是层间聚合，聚合器的选择直接影响聚合的效果。

4.1.2 信息聚合的一般方法

在信息聚合时，一般方法采用的聚合器主要有三大类：基于池化操作的聚合器、基于注意力机制的聚合器以及混合聚合器。

（1）基于池化操作的聚合器

池化操作的概念来源于卷积神经网络（Graph Convolutional Network, CNN）^[43]，在CNN中，原始图片经过卷积层后，需要将卷积层输出的特征整合，池化操作就是对获得的特征进行降维，以方便计算。常用的池化方法有最大池化，平均池化和求和池化。最大池化可以理解为将图像局部放大，取视野内最大像素值作为该局部的特征，从而实现降维的目的。平均池化是将局部特征取平均作为特征，求和池化以求和作为特征。池化操作也可以作用于图神经网络聚合邻居信息阶段，它可以将同一阶信息嵌入集合压缩成一个定长的向量特征。在KGCN^[36]、KGNN-LS^[44]信息聚合时，都用到了平均池化操作。

（2）基于注意力机制的聚合器

注意力机制从自然语言处理领域发展而来，现在已经成为了深度学习在众多其他领域应用的核心方法之一。在信息聚合的应用方面，注意力机制一般用于区分同一信息嵌入集合内不同信息的重要度，以决定在聚合时，优先考虑哪些邻域节点信息。例如 KGAT^[25]专门设计了一个基于知识的注意力层，将注意力机制应用到基于异质信息网络的推荐系统中。此外，注意力机制还天然带有可解释的潜力，可以通过权重可视化，评估模型是否正确学习到异质信息网络中的关系特点。

（3）混合聚合器

混合聚合器指的是利用两种或两种以上不同的聚合器实现聚合的方法。混合聚合器的优势是能根据不同的应用场景定制化选择不同的聚合器，实现高效聚合，提升模型的效果。

4.1.3 目前的方法存在的问题

目前基于异质信息网络的推荐算法，在信息聚合时使用的方法普遍存在以下问题：

（1）在层内聚合阶段，忽视异质信息网络中的节点间关系的建模。目前的方法在处理同一阶信息嵌入集合中各节点的信息时，会弱化甚至忽视集合中的节点间关系的建模，导致无法区分同一阶不同信息的重要性。

（2）在层间聚合阶段，忽视了不同距离信息嵌入集合间存在的信息衰减现象。随着信息传播越向深处，远距离的邻居节点和关系的信息与原始节点的信息可能千差万别，甚至会噪声信息。目前的方法在信息聚合阶段，没有将这一因素考虑到算法建模中。例如在 CKAN^[7]算法中，在层间聚合时直接使用求和池化操作，没有考虑不同阶信息嵌入集合包含与原始节点不同的信息量，导致算法对信息传播的层数有很大的限制，因此 CKAN 无法处理远距离的噪声信息。

（3）没有对细粒度建模设计专门的信息聚合方法。在信息传播阶段细粒度建模信息的同时，也需要在聚合阶段，专门设计细粒度信息聚合的方法。目前基于异质信息网络的推荐算法在细粒度建模信息聚合时，大多沿用粗粒度下的聚合方法，导致无法充分发挥细粒度建模用户多意图和物品多属性的重要价值。

基于上述问题，本文对层内和层间聚合这两个信息聚合阶段，设计了一种基于分层注意力机制的方法。在层内聚合阶段，该方法能根据目标节点、信息嵌入集合中的邻居节点信息和关联关系三者得到定长的信息聚合向量。在层间聚合阶段，该方法能建模信

息传播的衰减现象，得到不同传播距离的信息嵌入集合对目标节点的聚合权重。分层注意力机制的设计，能高效地实现信息聚合。相比于传统聚合方法，它具有抑制噪声，充分利用异质信息网络中辅助信息的作用。

4.2 基于分层注意力机制的信息聚合方法的提出

4.2.1 问题描述

本节使用数学符号对本文要解决的问题进行了定义，表 2 给出了本节所使用到的数学符号。

表 2 基于分层注意力机制的信息聚合方法中定义的数学符号

符号	描述
S_u^l	用户 u 第 l 阶的三元组集合
S_v^l	物品 v 第 l 阶的三元组集合
E_u^l	用户 u 第 l 阶的节点集合
E_v^l	物品 v 第 l 阶的节点集合
$G = \{(h, r, t) h, t \in E, r \in R\}$	异质信息网络中的三元组集合
$e_i^h, e_i^t \in E$	第 i 个头/尾节点的向量表示
$r_i \in R$	第 i 个关系的向量表示
$\pi(e_i^h, r_i)$	第 i 个三元组的注意力权值
$W_0, W_1, W_2, W_{u,k}^{(l)}, W_{v,k}^{(l)}$	权重参数矩阵
$b_0, b_1, b_2, b_{u,k}^{(l)}, b_{v,k}^{(l)}$	偏置参数矩阵
$ReLU$	ReLU 激活函数
σ	Sigmoid 激活函数
$e_{u,k}^{(l)}$	用户 u 在 k 因素下第 l 阶的聚合向量表示
$e_{v,k}^{(l)}$	物品 v 在 k 因素下第 l 阶的聚合向量表示
$T_{u,k}$	用户 u 在 k 因素下 L 阶的向量表示集合
$T_{v,k}$	物品 v 在 k 因素下 L 阶的向量表示集合
$g_{u,k}^{(l)}$	用户 u 在 k 因素下第 l 阶向量表示的权重
$g_{v,k}^{(l)}$	物品 v 在 k 因素下第 l 阶向量表示的权重
e_u	用户 u 聚合后的向量表示
e_v	物品 v 聚合后的向量表示

4.2.2 信息聚合方法

基于分层注意力机制的信息聚合方法主要由两部分组成：层内信息聚合和层间信息聚合。层内聚合中的注意力机制用于区分同一阶邻域节点信息的不同重要度，经过层内聚合，同一阶的信息嵌入集合会聚合成一个定长的向量。层间聚合中的注意力机制用于区分不同阶邻域信息对中心节点表征的不同贡献度。理想状况下，它能建模出信息越向深处传播，信息量越发衰减，从而贡献度降低的现象。下面对层内聚合和层间聚合方法做具体介绍。

(1) 基于注意力机制的层内向量聚合

在层内聚合时，本文提出层内注意力机制聚合方法，它的具体结构如图 7 所示。层内注意力机制的输入是头节点向量和关系向量，经过多层感知机输出注意力权重，再把它和尾节点向量相乘得到注意力向量，最后注意力与关系向量进行元素积操作得到层内信息聚合向量。

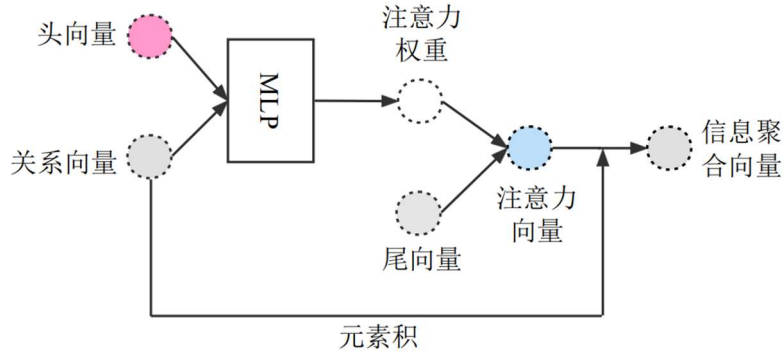


图 7 层内注意力机制结构示意图

具体地，同一阶邻域三元组集合 S_u^l 中的三元组形式化地表示为 (h, r, t) 。因为细粒度建模，本节所有节点和关系都在 k 因素下进行表征。令其中第 i 个三元组的头节点向量表示为 $e_{i,k}^h$ ，尾节点向量表示为 $e_{i,k}^t$ ，经过第三章信息传播中的消息变换操作，尾节点对应的消息向量为 $m_{i,k}^t$ ，关系向量表示为 $r_{i,k}$ ，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：

$$z_0 = \text{ReLU}(W_0(e_{i,k}^h || r_{i,k}) + b_0) \quad (4.1)$$

$$z_1 = \text{ReLU}(W_1 z_0 + b_1) \quad (4.2)$$

$$\pi'(e_{i,k}^h, r_{i,k}) = \sigma(W_2 z_1 + b_2) \quad (4.3)$$

为了防止某个节点被分配的注意力权重过大,而使得其他节点的权重无限趋近于 0,在训练过程中造成梯度消失的问题,本文对得到的所有注意力权重进行归一化处理。具体地,采用点式注意力机制的方法,通过 SOFTMAX 函数^[45]进行归一化。

对于用户 u 第 l 阶的三元组集合而言,第 i 个三元组得到的注意力权重用公式表示为:

$$\pi_u(e_{i,k}^h, r_{i,k}) = \frac{\exp(\pi'(e_{i,k}^h, r_{i,k}))}{\sum_{(h',r',t') \in S_u^l} \exp(\pi'(e', r'))} \quad (4.4)$$

同理,可以得到物品 v 第 l 阶的第 i 个三元组在集合中的注意力权重:

$$\pi_v(e_{i,k}^h, r_{i,k}) = \frac{\exp(\pi'(e_{i,k}^h, r_{i,k}))}{\sum_{(h',r',t') \in S_v^l} \exp(\pi'(e', r'))} \quad (4.5)$$

将所有三元组中的尾节点向量表示与对应的注意力权重进行相乘求和,就得到了注意力向量。因为关系在聚合时起到至关重要的作用,为了显式强调关系的重要性,将注意力向量与关系向量进行点积操作,就得到用户 u 和物品 v 第 l 阶三元组层内聚合后的向量表示。这个过程用公式可以表示为:

$$e_{u,k}^{(l)} = \sum_{i=1}^{|S_u^l|} \pi_u(e_{i,k}^h, r_{i,k}) m_{i,k}^t e_{i,k}^r \quad (4.6)$$

$$e_{v,k}^{(l)} = \sum_{i=1}^{|S_v^l|} \pi_v(e_{i,k}^h, r_{i,k}) m_{i,k}^t e_{i,k}^r \quad (4.7)$$

最终,经过基于注意力机制的层内向量聚合方法后,就得到用户或物品特定因素下每一阶信息嵌入集合的定长向量表示。因此,用户 u 和物品 v 在 k 因素下所得到的所有 L 阶的向量表示集合就表示为:

$$T_{u,k} = \{e_{u,k}^{(1)}, e_{u,k}^{(2)}, \dots, e_{u,k}^{(l)}, \dots, e_{u,k}^{(L)}\} \quad (4.8)$$

$$T_{v,k} = \{e_{v,k}^{(1)}, e_{v,k}^{(2)}, \dots, e_{v,k}^{(l)}, \dots, e_{v,k}^{(L)}\} \quad (4.9)$$

(2) 基于注意力机制的层间向量聚合

在得到用户和物品 k 因素下 L 阶的信息向量表示集合后,需要将各阶的信息向量表示聚合成一个定长向量。传统的求和或均值聚合,没有考虑信息传播越深信息信息量越少的现象,本文提出基于注意力机制的层间向量聚合方法建模信息衰减现象,对不同因素下不同阶的信息向量自适应分配权重,最终经过加权求和,实现层间向量聚合。

层间注意力机制是一个典型的点式注意力机制，它以目标节点的初始表示和第 l 层内聚合后的向量表示作为输入，最终输出注意力权值。具体地，用户 u 在 k 因素下第 l 阶向量表示的权重用公式表示为：

$$g'_{u,k}^{(l)} = \sigma \left(W_{u,k}^{(l)} \left(e_{u,k}^{(0)} || e_{u,k}^{(l)} \right) + b_{u,k}^{(l)} \right) \quad (4.10)$$

其中 $e_{u,k}^{(0)}$ 为用户 u 在 k 因素下的初始向量表征。

为了防止训练过程中出现梯度消失问题，采用 SOFTMAX 函数进行归一化。对用户 u 在 k 因素下的第 l 阶向量表示来说，所能得到的归一化注意力权重输出为：

$$g_{u,k}^{(l)} = \frac{\exp(g'_{u,k}^{(l)})}{\sum_{\{g_{u,k}^{(n)} | n \in \{1, 2, \dots, L\}\}} \exp(g'_{u,k}^{(n)})} \quad (4.11)$$

其中 L 为传播的最远距离。

同样地，物品 v 在 k 因素下的第 l 阶向量表示的注意力权重计算公式为：

$$g'_{v,k}^{(l)} = \sigma \left(W_{v,k}^{(l)} \left(e_{v,k}^{(0)} || e_{v,k}^{(l)} \right) + b_{v,k}^{(l)} \right) \quad (4.12)$$

$$g_{v,k}^{(l)} = \frac{\exp(g'_{v,k}^{(l)})}{\sum_{\{g_{v,k}^{(n)} | n \in \{1, 2, \dots, L\}\}} \exp(g'_{v,k}^{(n)})} \quad (4.13)$$

在经过层间注意力机制后，可以对不同阶的向量表示进行加权聚合，模拟信息衰减。最后融合 K 个因素下的聚合向量就能得到最终的节点表示。因此，用户 u 和物品 v 的聚合向量最终表示为：

$$e_u = \sum_{k=1}^K \sum_{(e_{u,k}^{(l)} \in T_{u,k})} e_{u,k}^{(l)} g_{u,k}^{(l)} \quad (4.14)$$

$$e_v = \sum_{k=1}^K \sum_{(e_{v,k}^{(l)} \in T_{v,k})} e_{v,k}^{(l)} g_{v,k}^{(l)} \quad (4.15)$$

4.3 本章小结

本章对基于分层注意力机制的信息聚合方法进行了详细介绍。首先明确了信息聚合的相关概念，介绍了信息聚合的一般方法，并分析了现有方法中所存在的问题。接着，

基于目前方法存在的不足，介绍了本文提出的基于分层注意力机制的信息聚合方法，并用形式化的方法详细介绍了层内注意力机制和层间注意力机制。

第五章 基于分离式表征与分层注意力机制的异质信息推荐网络

在基于异质信息网络传播机制的推荐算法研究中，信息传播方法和信息聚合方法是关键核心。有了上述两个核心方法的设计之后，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，以充分发挥所设计的信息传播和信息聚合方法的优势。本章首先用形式化的方式对问题进行描述，然后详细介绍基于分离式表征与分层注意力机制的异质信息推荐网络，并介绍模型训练的方式。

5.1 问题描述与形式化定义

在基于异质信息网络的推荐系统中，用户和物品是两个重要的对象。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 $V = \{v_1, v_2, \dots, v_N\}$ 。用户与物品的交互二分图表示 $G_1 = \{(u, y_{uv}, v) | u \in U, v \in V\}$ ，如果用户 u 与物品 v 之间存在交互，则 $y_{uv} = 1$ ，否则为 $y_{uv} = 0$ 。这种交互行为不是如打分之类的显式反馈，而是如点击、点赞等隐式反馈^[46]。

外部信息网络为推荐系统提供辅助信息。外部信息网络中包含大量的节点-关系-节点的三元组，可表示为 $G_2 = \{(h, r, t) | h, t \in E, r \in R\}$ ，其中 (h, r, t) 表示三元组，为了方便描述，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 。推荐系统中的用户或者物品往往与外部信息网络中的一个或多个节点具有匹配关系，它们组成集合 $A = \{(v, e) | v \in V, e \in E\}$ 。

用户-物品交互二分图中蕴含着整体的用户偏好，外部信息网络中包含与用户或者物品相关的辅助信息（用户社交信息或物品属性信息等）。通过用户或者物品与节点的匹配关系，可以融合成异质信息网络，本文将融合后的图取名为协同异质图，表示为 $G = \{(h, r, t) | h, t \in E', r \in R'\}$ ，其中 $E' = E \cup U$ ， $R' = R \cup \{\text{Interact}\}$ ，Interact表示用户和物品的交互行为。在已知协同异质图 G 的情况下，本文的任务目标为学习一个点击率预估算法 f ，预测给定用户 u 对特定候选物品 v 的交互概率 \hat{y}_{uv} ，用数学符号可以表示为：

$$\hat{y}_{uv} = f(u, v | \theta, G) \quad (5.1)$$

其中， f 为点击率预估算法，即推荐模型， θ 为模型的所有参数。

为了更好地描述 DHAN 模型，详细介绍其内部方法，表 3 列出了本章用到的数学符号及其对应描述。

表 3 DHAN 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
$G_2 = \{(h, r, t) h, t \in E, r \in R\}$	外部信息网络, 三元组集合
$G = \{(h, r, t) h, t \in E', r \in R'\}$	协同异质图, 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
$f(\cdot)$	DHAN 模型
θ	DHAN 模型的所有参数
$g(\cdot)$	TransR 得分函数
h, r, t	头节点、关系和尾节点向量表示
M_r	关系矩阵
h_r, t_r	关系空间中头和尾节点向量表示
S_u^l, S_v^l	用户 u 、物品 v 第 l 阶的三元组集合
E_u^l, E_v^l	用户 u 、物品 v 第 l 阶的节点集合
$M_{u,k}^l, M_{v,k}^l$	用户 u 、物品 v 在 k 因素下第 l 阶的信息嵌入集合
$e_i^h, e_i^t \in E$	第 i 个头/尾节点的向量表示
$e_{i,u}^l, e_{i,v}^l$	用户 u 、物品 v 第 l 阶邻域三元组中第 i 个节点
$r_i \in R$	第 i 个关系的向量表示
$\pi(e_i^h, r_i)$	第 i 个三元组的注意力权值
$W_0, W_1, W_2, W_{u,k}^{(l)}, W_{v,k}^{(l)}$	权重参数矩阵
$b_0, b_1, b_2, b_{u,k}^{(l)}, b_{v,k}^{(l)}$	偏置参数矩阵
$ReLU$	ReLU 激活函数
σ	Sigmoid 激活函数
$e_{u,k}^{(l)}, e_{v,k}^{(l)}$	用户 u 、物品 v 在 k 因素下第 l 阶的聚合向量
$T_{u,k}, T_{v,k}$	用户 u 、物品 v 在 k 因素下 L 阶向量表示集合
$g_{u,k}^{(l)}, g_{v,k}^{(l)}$	用户 u 、物品 v 第 l 阶向量表示的权重
e_u, e_v	用户 u 、物品 v 聚合后的向量表示
$\ \cdot \ $	向量间的拼接操作
x_u, x_v	用户 u 、物品 v 的基础特征向量
$MLP(\cdot)$	多层感知机
\hat{y}_{uv}	用户 u 与物品 v 的交互概率

5.2 DHAN 模型结构

针对基于异质信息网络的推荐问题，本文提出名为基于分离式表征和分层注意力机制的推荐网络（Disentangled and Hierarchical Attentive Network, DHAN）。DHAN 的结构如图 8 所示，分为异质信息表示层、信息传播层、信息聚合层和融合预测层。

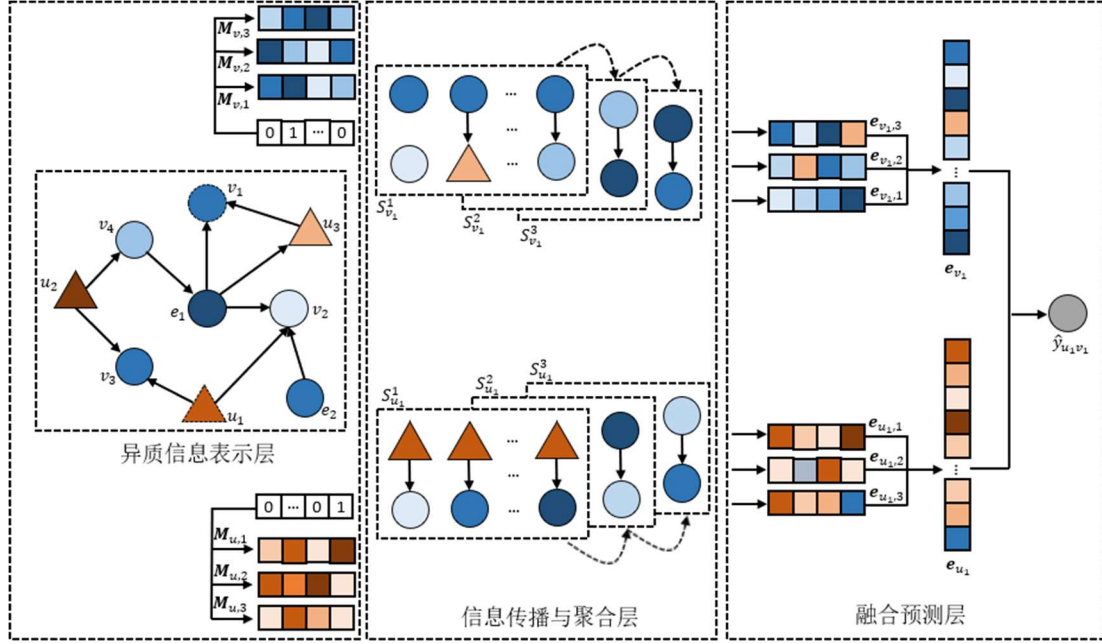


图 8 DHAN 网络结构示意图

异质信息表示层将协同异质图中的所有节点用低维向量进行表示，便于后续计算；信息传播层利用图神经网络架构方法，采样目标节点的邻域，并经过高阶传播得到不同阶的邻居节点-关系集合，同时利用分离式表征方法，对节点和关系进行细粒度建模；信息聚合层通过分层注意力机制，对信息传播获取到的节点-关系集合进行聚合，更新目标节点的表示；融合预测层将用户与物品的基本特征与 DHAN 模型获取的特征表示进行拼接融合，最后经过多层感知机输出用户与物品交互的概率值。

5.2.1 异质信息表示层

协同异质图中包含了用户与物品交互的协同信息和外部信息网络的辅助信息。异质信息表示是为了将协同异质图的结构和语义信息融合在节点和关系的低维嵌入中，常用的方法有图嵌入，知识图谱表示学习，图神经网络表示学习方法等。由于协同异质图和知识图谱有相同的三元组形式的表示，同时都强调了关系建模的重要性，最终综合实验效果和模型训练的复杂度，在本文中，采取的是知识图谱表示方法 TransR^[13]。

TransR 从 TransE^[11]发展而来, TransE 假设节点和关系是在同一个语义空间的向量, 这样相似的节点会在空间中相近的位置, 然而每一个节点都可以有很多方面, 而不同的关系关注的是节点不同的方面。因此, TransR 模型对不同的关系建立各自的关系空间, 通过一个关系矩阵 M_r 将节点映射到对应关系空间中, 映射后的头尾节点向量分别可以表示为 $h_r = hM_r$, $t_r = tM_r$, 在关系空间中关系向量 r 满足 $h_r + r \approx t_r$, 因此对于头节点向量 h 、关系向量 r 和尾节点向量 t , 得到得分函数:

$$g(h, r, t) = -||h_r + r - t_r|| \quad (5.2)$$

异质信息表示的训练使用贝叶斯个性排序 (Bayesian Personalized Ranking, BPR) 损失函数, 用公式表示为:

$$L_{IG} = \sum_{(h,r,t,t') \in G} -\ln \sigma(g(h, r, t') - g(h, r, t)) \quad (5.3)$$

5.2.2 信息传播层

本文在第三章介绍的基于分离式表征的传播方法应用在基于异质信息网络的推荐场景中就是 DHAN 的信息传播层。在 DHAN 中, 用户或者物品, 通过信息传播层, 可以同时进行协同信息和辅助信息传播。用户-物品的交互数据中存在着群体智慧, 这属于协同过滤的思想。外部信息网络又包含了用户或物品的辅助信息。因此可以说, 协同异质图中的节点邻居既有协同信息也有丰富的外部辅助信息。此外, 充分考虑一个节点的邻居节点信息也正符合了空域图神经网络的思想。DHAN 还引入分离式表征, 细粒度刻画消息的不同方面, 多层次建模消息。

具体地, 在信息传播层中, 对用户 u 和物品 v 进行 l 阶信息传播, 可以获取对应传播后的节点集合和三元组集合:

$$E_u^l = \{t | (h, r, t) \in G \text{ and } h \in E_u^{l-1}\} \quad (5.4)$$

$$E_v^l = \{t | (h, r, t) \in G \text{ and } h \in E_v^{l-1}\} \quad (5.5)$$

$$S_u^l = \{(h, r, t) | (h, r, t) \in G \text{ and } h \in E_u^{l-1}\} \quad (5.6)$$

$$S_v^l = \{(h, r, t) | (h, r, t) \in G \text{ and } h \in E_v^{l-1}\} \quad (5.7)$$

所得到的不同阶的节点集合和三元组集合会采用第三章的基于关系连接的自适应采样方法进行采样, 该方法能有效抑制噪声, 减小信息传播造成的偏差。

再经过第三章提出的基于分离式表征的信息转换方法后，可以得到信息传播 l 阶后，用户 u 和物品 v 在 K 个因素下信息嵌入集合：

$$M_{u,k}^l = \{m_{1,u,k}^l, m_{2,u,k}^l, \dots, m_{i,u,k}^l, \dots, m_{n,u,k}^l\} \quad (5.8)$$

$$M_{v,k}^l = \{m_{1,v,k}^l, m_{2,v,k}^l, \dots, m_{i,v,k}^l, \dots, m_{n,v,k}^l\} \quad (5.9)$$

其中：

$$m_{i,u,k}^l = W_k^T E(e_{i,u}^l) \quad (5.10)$$

$$m_{i,v,k}^l = W_k^T E(e_{i,v}^l) \quad (5.11)$$

5.2.3 信息聚合层

本文在第四章提出的基于分层注意力机制的信息聚合方法应用在推荐场景中就是 DHAN 的信息聚合层。分层注意力机制分别作用在层内聚合和层间聚合两阶段。

在层内聚合阶段，利用第四章中的基于注意力机制的层内向量聚合方法，可以得到层内信息聚合向量。具体地，在用户侧，对于 k 因素下同一阶三元组集合中第 i 个三元组的头节点向量 $e_{i,k}^h$ ，尾节点向量 $e_{i,k}^t$ ，与其对应的消息向量 $m_{i,k}^t$ 和关系向量 $r_{i,k}$ ，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：

$$\pi(e_{i,k}^h, r_{i,k}) = \sigma \left(W_2 \left(\text{ReLU} \left(W_1 \left(\text{ReLU}(W_0(e_{i,k}^h || r_{i,k}) + b_0) \right) + b_1 \right) \right) + b_2 \right) \quad (5.12)$$

$$\pi_u(e_{i,k}^h, r_{i,k}) = \frac{\exp(\pi(e_{i,k}^h, r_{i,k}))}{\sum_{(h', r', t' \in S_u^l)} \exp(\pi(e', r'))} \quad (5.13)$$

将所有三元组中的尾节点向量表示与对应的注意力权重进行相乘并求和，再将注意力向量与关系向量进行点积操作，就得到用户 u 在 k 因素下第 l 阶三元组层内聚合后的向量表示：

$$e_{u,k}^{(l)} = \sum_{i=1}^{|S_u^l|} \pi_u(e_{i,k}^h, r_{i,k}) m_{i,k}^t e_{i,k}^r \quad (5.14)$$

在物品侧用同样的方法可以得到物品 v 在 k 因素下第 l 阶三元组层内聚合后的向量表示：

$$e_{v,k}^{(l)} = \sum_{i=1}^{|S_v^l|} \pi_v(e_{i,k}^h, r_{i,k}) m_{i,k}^t e_{i,k}^r \quad (5.15)$$

最终用户 u 和物品 v 在 k 因素下得到的所有 L 阶的向量表示集合就表示为:

$$T_{u,k} = \{e_{u,k}^{(1)}, e_{u,k}^{(2)}, \dots, e_{u,k}^{(l)}, \dots, e_{u,k}^{(L)}\} \quad (5.16)$$

$$T_{v,k} = \{e_{v,k}^{(1)}, e_{v,k}^{(2)}, \dots, e_{v,k}^{(l)}, \dots, e_{v,k}^{(L)}\} \quad (5.17)$$

在层间知识聚合阶段, DHAN 使用第四章所提出的基于注意力机制的层间向量聚合方法, 对用户 u , 不同建模因素下, 不同传播距离的信息向量进行聚合, 计算公式为:

$$g'_{u,k}^{(l)} = \sigma \left(W_{u,k}^{(l)} (e_{u,k}^{(0)} || e_{u,k}^{(l)}) + b_{u,k}^{(l)} \right) \quad (5.18)$$

$$g_{u,k}^{(l)} = \frac{\exp(g'_{u,k}^{(l)})}{\sum_{\{g_{u,k}^{(n)} | n \in \{1, 2, \dots, L\}\}} \exp(g'_{u,k}^{(n)})} \quad (5.19)$$

对于物品 v , 利用同样的方法可以得到:

$$g'_{v,k}^{(l)} = \sigma \left(W_{v,k}^{(l)} (e_{v,k}^{(0)} || e_{v,k}^{(l)}) + b_{v,k}^{(l)} \right) \quad (5.20)$$

$$g_{v,k}^{(1)} = \frac{\exp(g'_{v,k}^{(1)})}{\sum_{\{g_{v,k}^{(n)} | n \in \{1, 2, \dots, L\}\}} \exp(g'_{v,k}^{(n)})} \quad (5.21)$$

经过层间注意力机制, 特定因素 k 下, 不同阶的向量表示可以进行加权聚合, 最后融合所有 K 个考虑因素, 得到用户 u 和物品 v 的嵌入表示, 用公式表示为:

$$e_u = \sum_{k=1}^K \sum_{(e_{u,k}^{(l)} \in T_{u,k})} e_{u,k}^{(l)} g_{u,k}^{(l)} \quad (5.22)$$

$$e_v = \sum_{k=1}^K \sum_{(e_{v,k}^{(l)} \in T_{v,k})} e_{v,k}^{(l)} g_{v,k}^{(l)} \quad (5.23)$$

5.2.4 融合预测层

经过信息传播层和信息聚合层, 用户 u 和物品 v 的嵌入表示都得到了更新。由于它们融合了邻域的辅助信息, 向量表示能力得到了增强。但是用户和物品本身也有属性特征, 因此可以将两部分特征表示进行融合, 就成为了用户 u 和物品 v 的最终向量表示。把它们进行拼接融合, 作为输入, 再经过多层感知拟合用户与物品交互的标签, 就得到最终用户与物品交互的概率值。

经过信息传播和聚合层后，用户和物品的嵌入表示分别记为 e_u 和 e_v ，用户基础信息特征的向量表示记为 x_u ，物品属性等信息经过编码后的向量表示记为 x_v ，将它们与用户和融合，就得到多层感知机的输入 x ：

$$x = e_u || e_v || x_u || x_v \quad (5.24)$$

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} ：

$$\hat{y}_{uv} = \sigma(MLP(x)) \quad (5.25)$$

5.3 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。在二分类问题下，模型训练一般使用交叉熵损失函数。DHAN 在训练时，也采用了该函数，其公式表示为：

$$L_{cf} = \sum_{u \in U} \left(\sum_{v \in V} (-y_{uv} \log \hat{y}_{uv} - (1 - y_{uv}) \log(1 - \hat{y}_{uv})) \right) + \lambda ||\Theta||_2^2 \quad (5.26)$$

其中， Θ 为 DHAN 模型中所有参数， $||\cdot||_2^2$ 为参数的 L2 正则项表示，它可以有效防止模型在训练时出现过拟合， λ 为正则化系数，为一个超参数。

由于异质信息层在异质信息表示时也存在一个损失函数，因此 DHAN，模型最终的损失函数表示如下：

$$L = \sum_{u \in U} \left(\sum_{v \in V} (-y_{uv} \log \hat{y}_{uv} - (1 - y_{uv}) \log(1 - \hat{y}_{uv})) \right) + \sum_{(h,r,t,t') \in G} -\ln \sigma(g(h,r,t') - g(h,r,t)) + \lambda ||\Theta||_2^2 \quad (5.27)$$

DHAN 在模型训练时，采用迭代训练的方式：先对异质信息表示层训练，再经过固定轮次后，训练推荐任务。其中轮次数为一个超参数。以此迭代，直到模型收敛。

5.4 本章小结

本章详细介绍了本文所提出的基于分离式表征和分层注意力机制的异质信息推荐模型 DHAN。首先，形式化地描述了该算法所要解决的问题场景，即基于异质信息网络

的推荐场景。然后，对 DHAN 模型结构进行分层地详细介绍，推导了异质信息表示层、信息传播层、信息聚合层和融合预测层四层结构。最后，介绍了模型的训练使用的损失函数以及训练方式。

第六章 实验与验证

为了验证 DHAN 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。

6.1 实验方案设计

6.1.1 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) CiteULike-t 数据集

CiteULike^[47]是一个用于文献收藏的标签工具。它可以帮助学者管理文献,除此之外,它还可以自动抽取文献的作者、期刊名、文章卷期、页码等详细信息。CiteULike-t 数据集包含了约 8000 位学术工作者对 25000 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,它们共同构成异质信息网络。

(2) Book-Crossing 数据集

Book-Crossing^[48]是一个以图书为载体的社交分享平台,平台中的用户互相传阅图书,并对图书进行 0 至 10 分的打分,分数越高代表对该图书的喜好程度和评价越高。Book-Crossing 数据集是由社区内 17860 名用户对 14967 本图书的评分组成,数据集中还提供了各本图书的详细信息,包括书的作者、出版社和出版日期等。

(3) MovieLens-20M 数据集

MovieLens^[49]是一个以研究为目的的电影推荐网站。MovieLens-20M 数据集从 IMDB 等互联网电影资料库上收集了 138159 名用户对 16954 部电影的评分,分数范围 1 至 5 分,代表用户对影片的喜好程度。除了用户对影片的评分信息外,数据集还提供了全部电影的概要信息和所有用户的身份信息。

(4) Last.FM 数据集

Last.FM^[50]是一个音乐社区平台。Last.FM 数据集收集了近 2000 名用户在一段时间内的所有听歌记录，包含歌曲曲名以及播放次数。除此之外，Last.FM 数据集中还提供用户基础信息以及歌曲的基本信息。

除了 CiteULike-t 数据集中的交互为收藏形式的隐式交互，Book-Crossing、MovieLens-20M 和 Last.FM 数据集中的用户-物品交互都是评分和收听次数形式的显式交互。由于实验的目的是预测用户对某物品是否会产生行为交互，而不是对物品的评分或交互次数预测，因此需要对 Book-Crossing、MovieLens-20M 和 Last.FM 数据集中的交互形式做转化。对于 MovieLens-20M 数据集，实验通过设置评分阈值的方式，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；对于 Book-Crossing 和 Last.FM 数据集，实验时考虑到它们数据的稀疏性，并没有对歌曲的播放次数或图书评分设置阈值，而是将用户与其所有交互过的物品构成的用户-物品对记为正样本，并随机从每位用户未评分过的物品中以正负样本 1:1 的比例抽取负样本，最后，所有正样本构成用户-物品交互二分图。

除了 CiteULike-t 数据集本身自带异质信息网络外，其他数据集虽然都有关于用户或物品的元数据，但是数据量还不够丰富，因此需要引入外部信息网络合并成异质信息网络，常见的外部信息网络，包括社交网络，知识图谱，事件图谱等。本文在实验中选择了知识图谱作为外部辅助信息网络。具体构造异质信息网络的方法是：从微软开源的知识图谱 Satori^[51]中抽取与原始数据集中的物品有关的信息构成外部信息网络，再与用户-物品交互二分图合并组成实验使用的异质信息网络。需要特别说明的是：如果交互二分图中的物品无法匹配 Satori 中的任一节点，会在数据预处理阶段，将该物品进行剔除，以简化实验。最终四个数据集的详细信息如表 4 所示。实验中，将各数据集以 6:2:2 的比例随机划分训练、验证和测试集。

表 4 数据集描述

数据来源	数据类别	CiteULike-t	Book-Crossing	MovieLens	Last.FM
用户-物品交互数据	用户数	7947	19676	138159	1872
	物品数	25975	20003	16954	3846
	交互数	134860	172576	13501622	42346
外部信息网络数据	节点数	78921	25787	102569	9366
	关系数	4	18	32	60
	三元组数	132565	60787	499474	15518

6.1.2 实验环境

实验中用到的服务器硬件和软件配置如表 5 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 版本号的操作系统中进行；编程语言及版本为 Python 3.6.10；深度学习框架及版本为 PyTorch 1.5.0。

表 5 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	3.6.10

6.1.3 对比算法

为了验证 DHAN 算法对比其他现有算法的优越性，本文选取了各个不同的基准推荐算法进行对比实验。按照类别来分，它们可以分为传统推荐算法和基于异质信息网络的推荐算法。对于基于异质信息网络的推荐算法，本文选取了基于路径和基于传播的方法中典型的算法。由于 DHAN 是属于基于传播的方法，因此本文着重对比此类方法下提出的主流算法。下面是各个基准算法的详细介绍。

(1) 贝叶斯个性排序算法 (Bayesian Personalized Ranking, BPR)

BPR^[5]是一种传统的基于协同过滤的推荐算法，该算法没有利用任何外部辅助信息，而是基于贝叶斯理论和矩阵分解方法，将用户-物品交互矩阵分解为用户矩阵和物品矩阵，以此得到各用户和物品的向量表示，通过向量的相似度衡量方法就可以得到一个用户对多个物品的个性化兴趣排序。

(2) 协同知识表征嵌入算法 (Collaborative Knowledge base Embedding, CKE)

CKE^[52]是一种基于协同过滤和知识嵌入的代表性算法，它是从协同过滤算法演化而来。在协同过滤方法分解得到用户和物品特征后，利用 CKE 对物品相关的辅助知识信息进行编码，从而丰富物品的特征表示。编码采用的是知识图谱嵌入方法 TransE。物品特征丰富后，可以有效提升推荐算法的准确性。

（3）个性化实体推荐算法（Personalized Entity Recommendation, PER）

PER^[3]是一种基于元路径的异质信息网络推荐算法，它通过预先定义用户-物品在异质信息网络中的元路径，然后利用基于元路径的相似搜索算法，寻找用户最有可能感兴趣的物品，实现个性化实体推荐。

（4）涟漪网络（RippleNet）

RippleNet^[6]是一种经典的基于异质信息网络传播机制的推荐算法。RippleNet 的中文名叫涟漪网络或者叫水波纹网络，它把信息传播的过程用水波泛起涟漪形象化地表现出来。在 RippleNet 算法中，用户交互过的物品称为种子，各个种子在异质信息网络中进行传播，扩散到其他节点，从而实现用户兴趣的延伸和扩展。

（5）基于知识图谱的图卷积网络（Knowledge Graph Convolutional Network, KGCN）

KGCN^[36]也是一种基于异质信息网络传播机制的推荐算法，它在推荐性能上表现优异，常被选取为对比基准算法。KGCN 是基于图卷积神经网络的一种推荐模型，它以异质信息网络中每个节点的邻居作为感受域，并从中进行采样。随着感受域的扩展，节点信息从低阶邻居扩展到高阶邻居，从而挖掘用户潜在的远程兴趣。

（6）基于知识图谱的图注意力网络（Knowledge Graph Attention Network, KGAT）

KGAT^[25]是当前流行的基于异质信息网络传播机制的推荐算法，它将知识图谱作为外部辅助信息，与用户-物品交互二分图合并到异质信息网络中，该信息网络称为协同知识图。在协同知识图中，KGAT 基于图注意力神经网络，结构分为三层：信息传播层、基于知识表征的注意力机制层和信息聚合层，实现用户兴趣发现和物品特性挖掘。

6.1.4 评估指标

对推荐系统的评估一直以来都是一项重要的工作，虽然推荐效果可以通过用户调研得到，但是这种方法需要耗费大量的人力物力且过于主观，因此现在主要采用离线评估的方法。根据推荐场景的不同，评估的方法也不相同。本文在选取评估指标时考虑了两种推荐场景：点击率预估场景以及前 K 个物品（TOP-K）推荐场景。

（1）点击率预估场景

点击率预估场景是指输入用户和物品特征，模型输出用户对该物品的点击概率。在该场景下常使用 AUC 和 F1 这两个评价指标。

AUC 表示 ROC (Receiver Operator Curve) 曲线下的面积, ROC 曲线的横坐标是伪阳率 (负样本预测为真的比率), 纵坐标是正阳率 (正样本预测为真的比率)。这些比率和分类阈值有关, 但在推荐系统中并不关心阈值的取值。因此, 这里的 AUC 取它的物理含义, 即给定一对正样本和负样本, 正样本概率得分大于负样本概率得分的概率计为 AUC 的取值。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:

$$AUC = \frac{\sum_{i=1}^N rank_i - \frac{M \times (M + 1)}{2}}{M \times N} \quad (6.1)$$

F1 是一个用于评估二分类模型精确度的指标。在二分类问题场景下, 首先需要定义一个打分阈值, 用于区分预测为正样本还是负样本。本实验中选择的阈值为 0.5, 当点击概率大于等于 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, 计算出准确率和召回率。定义预测为真, 实际也为真的是正阳(True Positive, TP)样本; 预测为真, 实际为假的是伪阳(False Positive, FP)样本; 预测为假, 实际也为假的是正阴(True Negative, TN)样本; 预测为假, 实际为真的是伪阴(False Negative, FN)样本。

准确率 (Precision) 表达的是预测为真的样本中有多少是真正的正样本。准确率越高, 说明模型效果越好。其计算公式为:

$$Precision = \frac{TP}{TP + FP} \quad (6.2)$$

召回率 (Recall) 描述所有样本中的正例有多少被预测正确了。召回率越高, 代表表现越好。其计算公式为:

$$Recall = \frac{TP}{TP + FN} \quad (6.3)$$

F1 同时考虑模型的精确率和召回率, 因此它可以视为模型精确率和召回率的调和平均。F1 越高, 说明模型的表现越好。其计算公式为:

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (6.4)$$

(2) TOP-K 推荐场景

TOP-K 推荐指的是在给用户的待推荐的物品集合中, 按照点击概率从高到低排序, 选择前 K 个物品形成列表推荐给用户。与点击率预估的评估方法关注推荐的顺序不同,

TOP-K 推荐关注的是给用户推荐的前 K 个商品中是否有用户真正感兴趣的，而不关注用户感兴趣的物品位置。在本文中，选择 Recall@K 作为 TOP-K 场景下的评估指标。

Recall@K 召回率指的是模型给用户推荐的前 K 个物品中，用户真实交互的物品个数所占的比率。比率越高，代表模型推荐得越精准。定义 $T(u)$ 为用户在测试集中真实的交互记录， $R(u)$ 为给用户推荐的物品列表，物品按照模型给出的预测得分由高到低排序，则召回率的计算公式为：

$$Recall@K = \frac{\sum_u |R(u)[1:K] \cap T(u)|}{\sum_u |T(u)|} \quad (6.5)$$

6.2 实验结果与分析

6.2.1 对比实验的结果与分析

本文提出的 DHAN 算法与各个对比方法在点击率预估场景下的 AUC 和 F1 的实验结果如表 6 所示，加粗数字为最优指标。从表中结果可知，DHAN 算法在论文、书籍、电影和音乐四个不同的推荐场景的数据集中都取得了最优异的表现。相对于次优结果，DHAN 在 AUC 指标上相对提升 1.4%、2.2%、0.4% 和 1.9%，在 F1 指标上相对提升 1.7%、1.5%、0.3% 和 3.0%。这表明了 DHAN 算法不仅在论文推荐场景中有很好的表现，在通用推荐场景中也具有优势。

表 6 各算法在点击率预估场景下的 AUC 和 F1 指标对比

算法	CiteULike		Book-Crossing		MovieLens		Last.FM	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
BPR	0.832	0.765	0.675	0.636	0.963	0.920	0.766	0.692
CKE	0.812	0.741	0.678	0.620	0.929	0.906	0.743	0.675
PER	0.766	0.697	0.605	0.572	0.838	0.792	0.641	0.603
RippleNet	0.862	0.783	0.720	0.645	0.976	0.920	0.778	0.702
KGCN	0.845	0.774	0.686	0.631	0.974	0.924	0.802	0.722
KGAT	0.848	0.782	0.725	0.654	0.972	0.922	0.825	0.739
DHAN	0.876	0.796	0.747	0.664	0.980	0.929	0.841	0.761

TOP-K 推荐场景下 DHAN 与各个对比方法分别在 CiteULike-t、Book-Crossing、MovieLens-20M 和 Last.FM 数据集中的召回率指标实验结果如图 9(a)-图 9(d)所示。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，DHAN 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，DHAN 倾

向于把所有正样本全部找出来。然而，观察四张折线图，会发现当 K 值取小于等于 10 时，DHAN 算法的优势没有特别明显。可能的解释是 DHAN 算法给一些负样本也打出了较高的预测得分，这部分物品同样具有被用户交互的潜质。

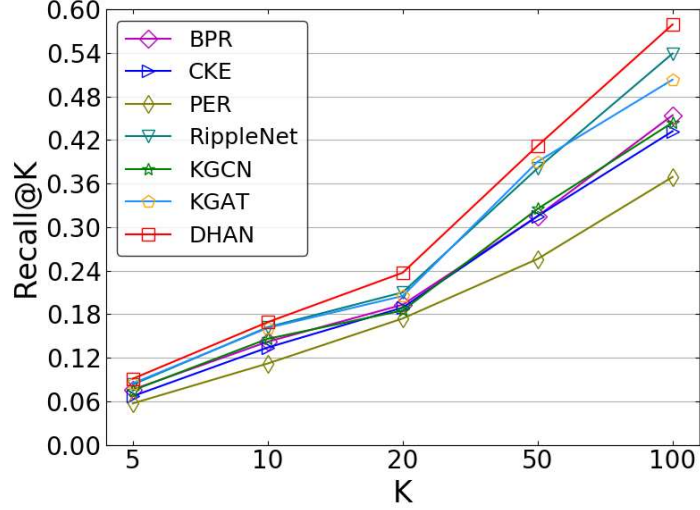


图 9(a) CiteULike 上 Recall@K 的结果

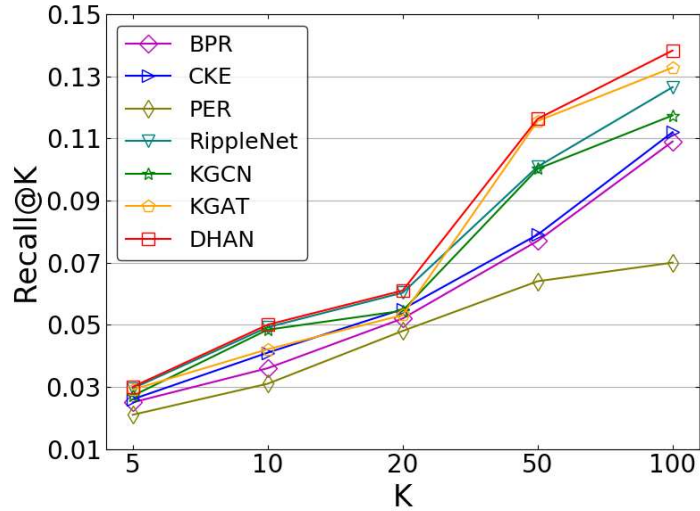


图 10(b) Book-Crossing 上 Recall@K 的结果

横向比较四个数据集上各推荐算法的表现，可以看到各算法在 MovieLens-20M、CiteULike、Last.FM 和 Book-Crossing 的表现依次呈现递减趋势。再观察这四个数据集中每位用户与物品交互的平均个数，由此可以分析出推荐算法在各个数据集上的整体表现很大程度上依赖数据的稀疏程度。在数据稠密的情况下，推荐算法的表现越优异。这一实验结论也和工业届推荐系统的实际情况相符合。

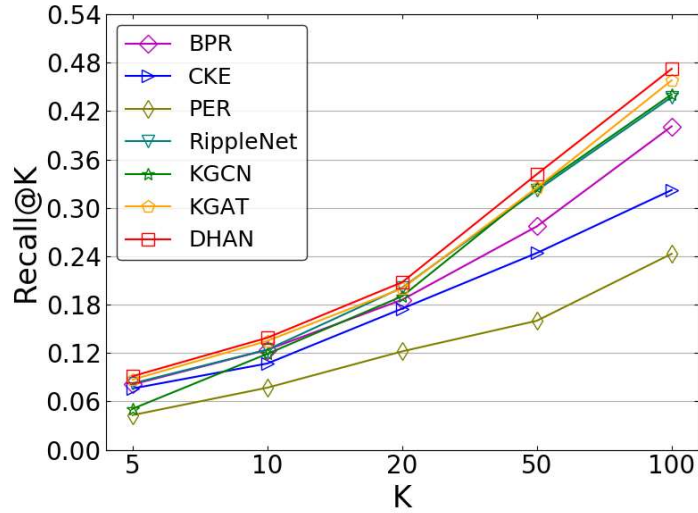


图 9(c) MovieLens 上 Recall@K 的结果

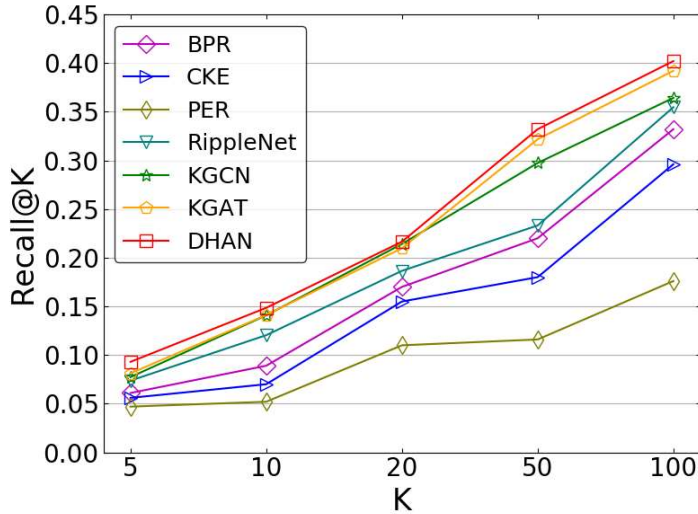


图 9(d) Last.FM 上 Recall@K 的结果

对比各个推荐算法,可以发现大部分基于异质信息网络的推荐算法的效果超过了经典的协同过滤算法,这说明了外部辅助信息对推荐效果的正向影响。基于知识嵌入的协同过滤方法 CKE 在大部分指标上比 BPR 略低,一方面说明了基于知识嵌入的协同过滤算法在推荐效果中并不能完全发挥知识图谱嵌入的优势,另一方面说明只考虑异质信息网络中自身节点的信息,忽视了其邻域信息的影响,就无法有效挖掘用户潜在的兴趣。在基于异质信息网络传播机制的一些算法中,基于路径的方法 PER 表现效果是所有基准算法最差的,可能的原因是没有预先定义完整的元路径,导致最终的推荐效果不佳。RippleNet、KGCN 和 KGAT 都是基于传播的方法,但是性能都被 DHAN 算法超过,这

一方面验证了在信息传播阶段对信息细粒度建模的优势，另一方面说明 DHAN 在信息聚合时使用分层注意力机制的有效性。

6.2.2 DHAN 模型消融实验结果与分析

为了进一步验证 DHAN 算法中信息传播机制设计的有效性，本文进行了消融实验，主要实验对比了 DHAN 三种变体算法。DHAN/D 算法消融了信息传播阶段的分离式表征部分，将异质信息网络中所有的节点和关系嵌入到一个定长维度的向量中；DHAN/A 算法不考虑信息聚合阶段，层内聚合时同一阶邻域信息的重要性因素，即去掉了 DHAN 算法的第一层注意力机制部分，而是简单地对邻域信息进行求和池化；DHAN/G 算法在层间聚合时消融了第二层注意力机制部分，不对信息衰减进行建模，而是直接求和池化。

表 7 展示了各个变体模型在不同数据集下的 AUC 结果，从表中结果可以发现，DHAN/D 在各个数据集上的 AUC 都有明显的性能损失。这表明在信息传播阶段，粗粒度建模得到的用户和物品表征是次优的，推荐效果不能发挥到最佳，证明了分离式表征学习具有无可比拟的优势。DHAN/A 和 DHAN/G 相比于 DHAN 在 AUC 指标性能略有下降，这反映了在信息聚合阶段，聚合邻域信息时区分邻域信息的重要性，也验证了对信息传播衰减进行建模的必要性。

表 7 DHAN 变体模型在不同数据集下的 AUC 结果

模型	CiteULike	Book-Crossing	MovieLens	Last.FM
DHAN/D	0.839	0.724	0.974	0.830
DHAN/A	0.849	0.741	0.976	0.835
DHAN/G	0.854	0.743	0.977	0.840
DHAN	0.876	0.747	0.980	0.841

6.2.3 DHAN 模型超参数探索分析

DHAN 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 DHAN 模型的分析。

(1) 建模因素个数

在信息传播阶段，利用分离式表征细粒度建模信息是 DHAN 中的一个重要模块，建模时考虑的因素个数对于 DHAN 的实验结果有很大影响。表 8 详细地展示了四个不同的因素个数对 DHAN 模型 AUC 结果的影响，其中粗体为最优值。从实验结果来看，不同的数据集最优结果对应的建模因素个数不同，当论文、书籍、电影和音乐四个数据

集的因素个数分别取 4、3、2 和 2 时，实验指标达到最高。实验结果表明，并不是建模时考虑的因素越多越好，因素越多，参数量也就越大，如果数据集的数据量过小，很容易造成过拟合，导致测试结果不佳，因此选择合适的建模因素个数至关重要。

表 8 DHAN 不同因素个数对 AUC 结果的影响

因素个数	CiteULike	Book-Crossing	MovieLens	Last.FM
1	0.859	0.724	0.974	0.835
2	0.864	0.739	0.980	0.841
3	0.873	0.747	0.971	0.838
4	0.876	0.743	0.966	0.829

(2) 信息传播深度

在信息传播阶段，信息传播深度是一个十分重要的超参数，它直接影响最终的推荐效果。表 9 展示了不同信息传播深度对点击率预估场景下的 AUC 结果的影响，其中加粗数字表示最优值。从结果来看，当论文、书籍、电影和音乐四个场景下的数据传播深度分别为 3、2、1 和 3 时，AUC 指标值达到最优。实验结果表明，在一定程度上增加信息传播的深度，能够有效地扩充用户、物品和异质信息网络中各个节点的辅助额外信息，提升它们的嵌入表征能力，但是当传播深度达到一定限度后，由于传播到达的信息与原始节点信息相隔较远，可能成为噪声干扰，破坏了向量表征的鲁棒性。因此，信息传播深度这个超参数，要根据不同的数据集而定。从实验结果来看，数据集的数据稀疏程度影响较大。当数据稠密的情况下，异质信息网络能提供丰富的邻域信息扩充原始节点表征，随着传播深度加深，信息产生爆炸，会引入越来越多的噪声。在数据稀疏的情况下，信息需要传播得更远一些，获取足够的邻居节点集合，以扩充原始节点表征。这也能解释为什么在 CiteULike 和 Last.FM 数据集下，DHAN 算法取得最优 AUC 表现时，传播深度比另外两个数据集更深的现象。

表 9 DHAN 不同信息传播深度对 AUC 结果的影响

传播深度	CiteULike	Book-Crossing	MovieLens	Last.FM
1	0.861	0.739	0.980	0.828
2	0.870	0.747	0.970	0.837
3	0.876	0.745	0.964	0.841
4	0.865	0.738	0.956	0.832

(3) 向量表征维度

为了防止引入过多超参数，本文在实验时对各节点（包括用户和物品）和关系都使用了相同维度的向量表征。为了探究不同向量表征维度对实验结果的影响，本文在四个数据集上实验了不同的向量嵌入维度对点击率预估场景下 AUC 的影响，实验结果如图 10(a)-图 10(d)所示。从实验结果来看，当向量表征维度分别取 64、128、64 和 128 时，论文、书籍、电影和音乐数据集下的 AUC 指标达到最大值。这表明在一定范围内，增加向量表征维度可以提升 DHAN 在 AUC 性能上的表现，但是并不是维度越大越好，在超过一定阈值后，其 AUC 性能会出现下降。一个可能的解释是随着向量的表征维度增大，该向量能有效编码更多的信息，但是超过阈值后，会使得模型出现过拟合的现象，造成效果的下降。

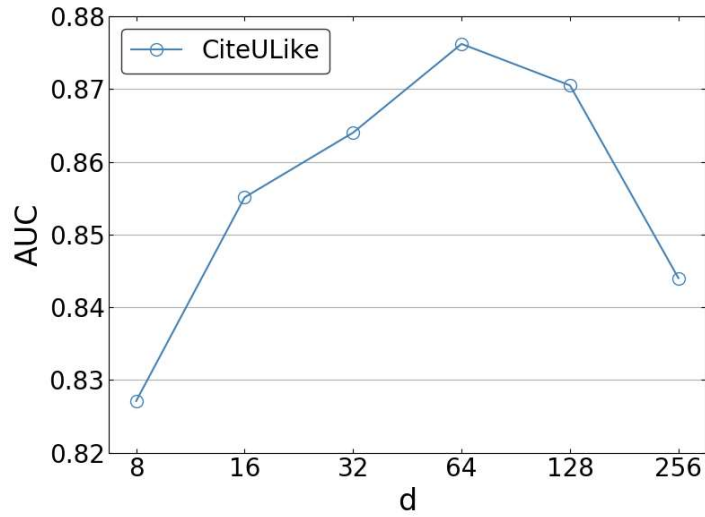


图 10(a) CiteULike 上 DHAN 的 AUC 随向量维度变化

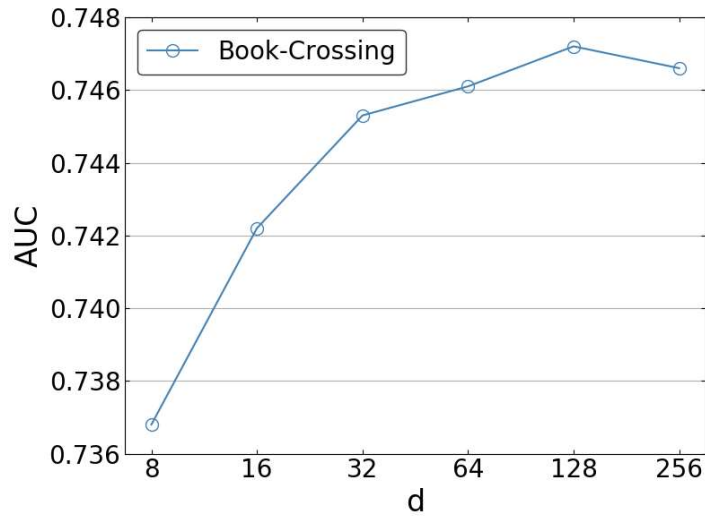


图 10(b) Book-Crossing 上 DHAN 的 AUC 随向量维度变化

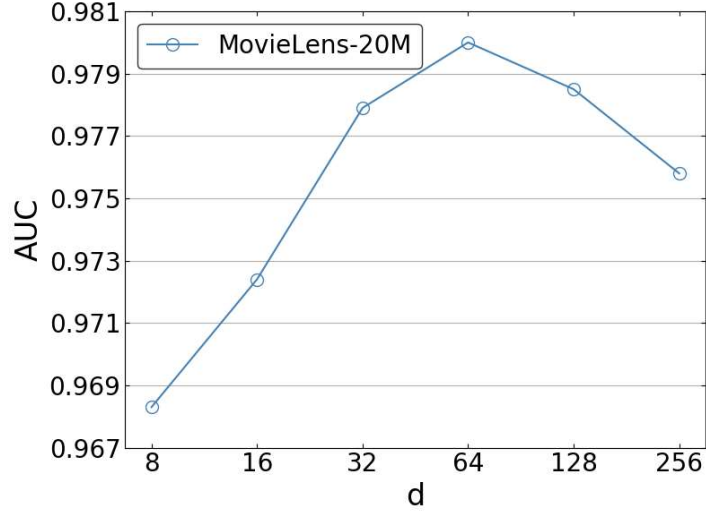


图 10(c) MovieLens 上 DHAN 的 AUC 随向量维度变化

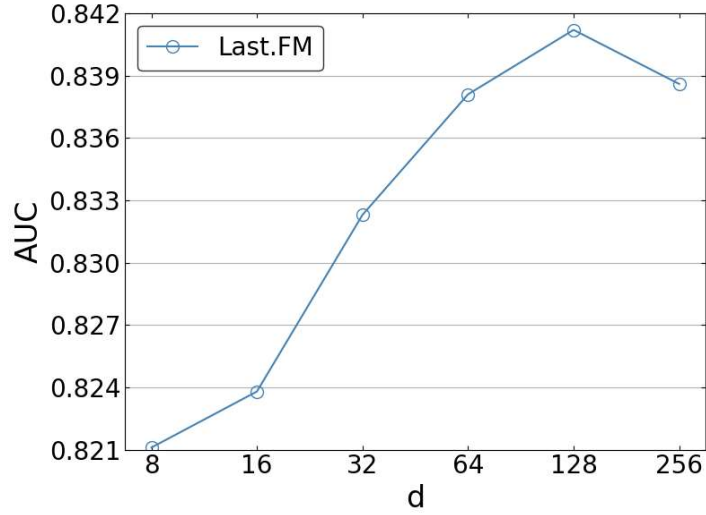


图 10(d) Last.FM 上 DHAN 的 AUC 随向量维度变化

6.3 本章小结

本章对本文提出的 DHAN 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，用于对比实验的基准算法以及评估指标。接着是对实验结果的分析，这里分为三部分：第一部分是展示 DHAN 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 DHAN 的消融实验，主要是对分离式表征与分层注意力机制进行消融，并对消融结果加以分析；第三部分是对 DHAN 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

本文聚焦基于异质信息网络传播机制的推荐算法研究，重点研究基于传播的推荐方法中的信息传播机制。信息传播机制包括两个阶段：信息传播和信息聚合。本文分别分析两阶段现有工作存在的不足，分别提出基于分离式表征的信息传播方法和基于分层注意力机制的信息聚合方法，主要内容包含以下方面：

(1) 在信息传播阶段，针对现有方法普遍存在无法细粒度建模信息，信息传播过程中噪声较大的问题。本文提出基于分离式表征的信息传播方法，细粒度建模用户、物品，从而更好地分析用户意图，建模物品的多属性。为了抑制信息在传播过程中的噪声，本文采用基于关系连接的自适应采样方法，优先考虑有关联关系的邻居信息，提升了信息传播的有效性。

(2) 在信息聚合阶段，针对现有方法中存在的不足：无法有效区分信息间相关度和重要性，无法建模信息衰减现象。本文提出基于分层注意力机制的信息聚合方法。第一层注意力机制用于层内聚合，目的是区分层内各个节点与目标节点的相关性，从而确定其重要性。第二层注意力机制作用于层间聚合，建模信息衰减现象，使得信息聚合处理更加精细化，大大提升信息聚合的能力。

本文在信息传播和信息聚合两阶段的研究基础上，提出一种基于分离式表征与分层注意力机制的异质信息推荐网络 DHAN。DHAN 算法集成了基于分离式表征的信息传播和基于分层注意力机制的信息聚合两阶段的所有优点，形成了一个端到端的可训练的模型。

本文将 DHAN 算法与主流的推荐算法在多个评估指标下进行了实验对比，实验结果表明 DHAN 算法在论文、书籍、电影和音乐推荐数据集上的 AUC 指标分别相对提升 1.4%、2.2%、0.4% 和 1.9%。本文还对 DHAN 算法的超参数进行了分析，并对相关实验现象进行了解释。

本文初步验证了细粒度建模信息传播，分层建模信息聚合方法的有效性，然而因为作者的研究时间有限，研究内容还存在很多待完善的地方，未来更多研究点还有待发掘。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) 在信息传播过程中,信息是逐步发散的,随着信息向深处传播,传播到的节点数量和信息量会越来越大,由此会在信息传播过程中引入非常多噪声。现有的方法都是通过采样解决这类问题,然而采样必然带来信息的损失,未来的发展方向之一在于利用非采样策略进行信息传播,非采样可以利用异质信息网络中更多语义和结构信息,提升信息的表达能力。

(2) 现在用于研究的异质信息网络多是静态的,网络中的节点和关系都是固定的,然而现实世界中,信息是无时无刻更新的,辅助于推荐系统的异质信息网络也是处于动态更新的过程中。因此未来的研究方向在于对动态图进行研究,可能的研究点包括:如何不更新全图的情况,增量更新动态信息;如何解决新节点、新关系的冷启动问题。

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] 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.
- [4] Sun Y, Han J, Yan X, et al. Pathsirn: Meta Path-based Top-k Similarity Search in Heterogeneous Information Networks[J]. Proceedings of the VLDB Endowment. 2011, 4(11): 992-1003.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [9] 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.
- [10] Grover A, Leskovec J. Node2vec: Scalable Feature Learning for Networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference. 2016: 855-864.
- [11] 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.
- [12] 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.
- [13] Lin H, Liu Y, Wang W, et al. Learning Entity and Relation Embeddings for Knowledge Resolution[J]. Procedia Computer Science, 2017, 108: 345-354.
- [14] 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.
- [15] 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.
- [16] 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.

-
- [17] 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.
- [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] Fan W, Ma Y, Li Q, et al. Graph Neural Networks for Social Recommendation[C]//The World Wide Web Conference. 2019: 417-426.
- [27] Shi C, Hu B, Zhao X, et al. Heterogeneous Information Network Embedding for Recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 31(2): 357-370.
- [28] Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828.
- [29] Wang X, Jin H, Zhang A, et al. Disentangled Graph Collaborative Filtering[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 1001-1010.
- [30] Schuster M, Paliwal K K. Bidirectional Recurrent Neural Networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681.
- [31] 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
- [32] Shaw P, Uszkoreit J, Vaswani A. Self-attention with Relative Position Representations[J]. arXiv preprint arXiv:1803.02155, 2018.

- [33] 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.
- [34] 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.
- [35] 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.
- [36] Wang H, Zhao M, Xie X, et al. Knowledge Graph Convolutional Networks for Recommender Systems[C]//The World Wide Web Conference. 2019: 3307-3313
- [37] Melnik S, Garcia-Molina H, Rahm E. Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching[C]//Proceedings of the 18th International Conference on Data Engineering. 2002: 117-128.
- [38] Song J, Yang Y, Li X, et al. Robust Hashing with Local Models for Approximate Similarity Search[J]. IEEE Transactions on Cybernetics, 2014, 44(7): 1225-1236.
- [39] Santos C T, Euzenat J, Meilicke C, et al. Evaluation Design and Collection of Test Data for Matching Tools[J]. 2009.
- [40] Chen J, Shtykh R Y, Jin Q. A Web Recommender System Based on Dynamic Sampling of User Information Access Behaviors[C]//Proceedings of the 9th IEEE International Conference on Computer and Information Technology. IEEE, 2009, 2: 172-177.
- [41] Oh J, Cho K, Bruna J. Advancing GraphSAGE with A Data-Driven Node Sampling[J]. arXiv preprint arXiv:1904.12935, 2019.
- [42] Ying R, He R, Chen K, et al. Graph Convolutional Neural Networks for Web-Scale Recommender Systems[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2018: 974–983.
- [43] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [44] Wang H, Zhang F, Zhang M, et al. Knowledge-aware Graph Neural Networks with Label Smoothness Regularization for Recommender Systems[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 968-977.
- [45] Memisevic R, Zach C, Pollefeys M, et al. Gated Softmax Classification[C]//Advances in Neural Information Processing Systems. 2010: 1603-1611.
- [46] 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.
- [47] 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.

- [48] 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.
- [49] 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.
- [50] 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.
- [51] 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.
- [52] 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.

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.
- [2] **Chen Qinghong**, Tan Huobin, Lin Guangyan and Wang Ze. A Hierarchical Knowledge and Interest Propagation Network for Recommender Systems[C]//Proceedings of 20th IEEE International Conference on Data Mining Workshops. 2020: 119-126.
- [3] Wang Ze, Lin Guangyan, Tan Huobin, **Chen Qinghong** and Liu Xiyang. 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.
- [4] Liu Xiyang, Tan Huobin, **Chen Qinghong** and Lin Guangyan. RAGAT: Relation Aware Graph Attention Network for Knowledge Graph Completion[J]. IEEE Access, 2021, 9: 20840-20849.

致谢

2019年9月我开始了在北航的研究生生活，全新的校园环境，未知的实验室氛围都让我忐忑，但这种忐忑的心理在我进入实验室和林广艳导师第一次谈话后就消解了。仍记得谈话中，林老师问我是想做工程还是做学术，我表达了想做学术。谈话的最后，林老师告诉我一定不会后悔选择北航，选择现在的实验室。两年多过去了，我可以很坚定地说，我不仅不后悔，反而感到很幸运。

这篇毕业论文是我两年多来的科研成果。学术这条路并不好走，这一路上多亏有林老师和谭火彬老师的支持和帮助。每周一次的组会上，两位老师几乎从不缺席地听大家分享领域内的前沿工作，并鼓励大家从这些工作中提取创新点，思考可以改进的方面，大胆实验，实现自己的想法。正是有了两位老师的鼓励，我才有动力以第一作者的身份发表了两篇小论文。在此特别感谢谭老师帮助我修改小论文，完成投稿、定稿等工作，感谢林老师对大论文撰写和修改提供的宝贵建议。两位老师对待科研严谨认真的态度，待人友善温和的性格将会一直影响我。

我还要感谢实验室的学长学姐、同级同学和学弟学妹们，在实验室大家庭里我感受到了大家的和谐温暖。特别感谢王泽学长为实验室开辟新的研究方向，带领我入门推荐系统，并在研究工作中指导和锻炼我。虽然难以望其项背，但学长一直是我学习的榜样。感谢柳啸峰学弟帮助我处理数据，进行对比实验以及对小论文的修改和校稿工作。感谢刘希阳同学，在知识图谱领域研究方面的指点。

此外，还要感谢杨弘毅同学在小论文英语语法方面的指导。感谢李丹阳、何明智同学在推荐系统研究方面的探讨和灵感启发。

最后特别感谢我的父母，他们在求学途中一直默默鼓励和支持我，无论我走到哪里，他们都是我最坚实的依靠。感谢北航的培养，“德才兼备，知行合一”的校训将一直激励我做更好的北航人。

即将离开学校和实验室，走上工作岗位，心中满是不舍。不舍得和18年的校园生活说再见，不舍得和北京说再见，不舍得和实验室的大家说再见……然而生活总是要向前走的，终身学习的场所不只在校园，研究生毕业是全新生活的开始，我们也会在更高处再相见。祝福实验室发展越来越好。