

分类号：

单位代码：10140

密 级：公 开

学 号：4032032343

遼寧大學

硕 士 学 位 论 文

中文题目：融合多模态知识图谱与注意力机制的个性化信息推荐研究

Research on Personalized Information Recommendation
Integrated Multi-modal

英文题目：Knowledge Graph and Attention Mechanism

论文作者：陈 稳

指导教师：冯 勇 教 授

专 业：计算机软件与理论

完成时间：二〇二三年五月

申请辽宁大学硕士学位论文

融合多模态知识图谱与注意力机制的个性
化信息推荐研究

Research on Personalized Information Recommendation
Integrated Multi-modal Knowledge Graph and Attention
Mechanism

作 者： 陈 稳

指导教师： 冯 勇 教授

专 业： 计 算 机 软 件 与 理 论

答辩日期： 2023 年 5 月 16 日

二〇二三年三月·中国辽宁

摘要

随着数字技术和互联网的迅速发展，我们生活在一个信息爆炸的时代，每天都会被大量的信息和选择所包围。推荐系统作为一种智能化的技术，能够利用用户的历史数据和个人偏好，为用户推荐与其兴趣相关的商品、服务或内容，以便帮助用户更快、更有效地找到他们想要的东西。然而，推荐系统目前也面临着一些问题，其中最主要的是数据稀疏问题。数据稀疏指的是用户的历史行为数据过少，难以对用户进行准确的兴趣分析。在这种情况下，推荐系统可能会出现推荐错误或者无法为用户提供满意的推荐服务。针对此问题，目前的研究趋势是将外部信息引入到推荐模型中，其中知识图谱被广泛应用。然而，现有基于知识图谱的推荐方法主要侧重于利用知识图谱中的路径结构信息或知识图谱中的语义信息，忽略了知识图谱内部丰富的高阶信息。

本文针对推荐系统存在的数据稀疏问题以及知识图谱内部高阶信息没有得到充分利用的问题，提出了一种融合多模态知识图谱与注意力机制的个性化信息推荐方法，该方法主要包含了多模态数据的构建与预处理方法和基于关系空间的注意力机制传播模型两个部分，具体研究内容如下：

（1）针对推荐系统在实际应用中面临的数据稀疏性问题，本文提出了一种多模态数据的构建与预处理方法，该方法将多模态信息分为内部和外部两个部分。其中，将与物品相对应的视觉图像（例如电影海报）以基于实体的形式被用作内部多模态信息来扩展知识图谱推荐。在知识图谱嵌入过程中往往将物品的文本属性直接嵌入到一个 ID 中，从而忽略了物品所包含的一些语义信息，比如电影名信息。因此，本文引入物品的标题文本语义作为外部多模态特征来弥补知识图谱嵌入带来的语义缺失问题。同时将用户特征属性加入外部多模态信息中可以更准确地分析出用户的兴趣爱好。通过这种方法，推荐系统可以更好地应对数据稀疏性问题。

（2）针对知识图谱推荐中高阶信息挖掘不充分的问题，本文提出了一种基于关系空间的注意力机制传播模型。具体而言，该模型首先从内部多模态知识图谱层中得到多模态特征向量作为输入，并通过基于关系的感知注意力机制来学

习传播过程中的每个邻居权重，这样可以在关系空间中找到关系更近的实体。在特征聚合阶段，本文采用相加、连接和双端混合三种方式进行聚合效果测试，最后，将外部多模态用户和物品特征向量加入，融合得到最终的用户和物品特征向量以用来推荐预测。

为了检验本文所提的融合多模态知识图谱与注意力机制的个性化信息推荐方法的有效性，本文基于 Multi-MovieLens1M 数据集进行了对比实验，在准确率、召回率等指标上与几种当前主流推荐方法进行了比较。经实验验证，本文所提方法比 MKR、RippleNet 等知识图谱推荐方法在准确率、召回率以及 ACC 等指标上均有所提升。同时，针对数据稀疏性做了对比实验，实验表明在保证推荐准确性的前提下，所提方法可以有效缓解推荐系统的数据稀疏性问题。

关键词：个性化信息推荐，多模态，知识图谱，卷积神经网络，注意力机制

ABSTRACT

With the rapid development of digital technology and the Internet, we live in an era of information explosion, surrounded by a large amount of information and choices every day. As an intelligent technology, the recommendation system can use users' historical data and personal preferences to recommend goods, services or content related to their interests to users, so as to help users find what they want faster and more effectively. However, the recommendation system is also facing some problems, the most important of which is data sparseness. Data sparsity refers to the fact that the user's historical behavior data is too small to accurately analyze and recommend the user's interest. In this case, the recommendation system may have recommendation errors or cannot provide satisfactory recommendations for users. Aiming at this problem, the current research trend is to introduce external information into the recommendation model, in which the knowledge graph is widely used. However, the existing knowledge graph recommendation methods mainly focus on using the path structure information in the knowledge graph or the semantic information in the knowledge graph, ignoring the rich high-order information inside the knowledge graph.

Aiming at the problem of data sparseness in the recommendation system and the problem that the high-order information inside the knowledge graph has not been fully utilized, this thesis proposed a personalized information recommendation method that combined multi-modal knowledge graph and attention mechanism. The method mainly included two parts : the knowledge graph representation method that integrated multi-modal information and the attention mechanism propagation model based on relational space. The specific research contents were as follows :

(1) Addressing the problem of data sparsity faced by recommendation systems in practical applications, this thesis proposed a new knowledge graph representation method that integrated multi-modal information. Specifically, visual images (such as movie posters) corresponding to items were used as internal multimodal information

in an entity-based form to expand knowledge graph recommendation. However, in the process of knowledge graph embedding, the text attributes of items were directly embedded into an ID, thus ignoring some semantic information contained in items, such as movie name information. Therefore, this thesis introduced the title text semantics of items as external multimodal features to make up for the semantic loss caused by knowledge graph embedding. At the same time, adding user feature attributes to external multimodal information can more accurately analyze users' interests. Through this method, the recommendation system can better deal with the problem of data sparsity.

(2) Tackling the problem of insufficient mining of high-order information in knowledge graph recommendation, this thesis proposed an attention mechanism propagation fusion model based on relational space. Specifically, the model first obtained multimodal feature vectors from the internal multimodal knowledge graph layer as input, and learned the weight of each neighbor in the propagation process through the relationship-based perceptual attention mechanism, so that entities with closer relationships could be found in the relationship space. In the feature aggregation stage, this thesis used addition, connection and double-ended mixing to test the aggregation effect. Finally, the external multi-modal user and item feature vectors were added, and the final user and item feature vectors were fused to predict the recommendation.

In order to test the effectiveness of the personalized information recommendation method combining multi-modal knowledge graph and attention mechanism proposed in this thesis, this thesis conducted a comparative experiment based on the Multi-MovieLens1M dataset, and compared it with several current mainstream recommendation methods in terms of accuracy and recall rate. Experimental results showed that the proposed method was better than MKR, RippleNet and other knowledge graph recommendation methods in terms of accuracy, recall and ACC. A comparative experiment was conducted on data sparsity. The experiment showed that the data sparsity problem of the recommendation system was further effectively alleviated while ensuring the accuracy of the recommendation.

Key Words: Personalized Information Recommendation, Multimodal, Knowledge Graph, Convolutional Neural Network, Attention Mechanism

目 录

第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	3
1.2 国内外研究现状.....	4
1.2.1 推荐系统.....	4
1.2.2 多模态表示学习.....	6
1.3 主要研究工作.....	7
1.4 论文结构与安排.....	8
第 2 章 相关理论	9
2.1 知识图谱.....	9
2.1.1 知识图谱嵌入方法.....	10
2.1.2 多模态知识图谱表示方法.....	13
2.2 基于知识图谱的推荐方法	15
2.2.1 基于嵌入的方法.....	15
2.2.2 基于路径的方法.....	16
2.3 深度学习.....	17
2.3.1 多层感知机.....	17
2.3.2 卷积神经网络.....	19
2.3.3 注意力机制.....	20
2.4 本章小结.....	21
第 3 章 多模态数据的构建与预处理方法	23
3.1 方法思想.....	23
3.2 内部多模态.....	25
3.2.1 多模态数据集构建.....	25
3.2.2 协作知识图谱构建.....	26
3.2.3 预处理方法.....	28
3.3 外部多模态.....	31
3.4 方法实现.....	33
3.5 本章小结.....	34
第 4 章 基于关系空间的注意力机制传播模型	35
4.1 模型框架.....	35
4.2 模型描述.....	36
4.2.1 基于关系空间的注意力传播.....	37

4.2.2 信息聚合	38
4.2.3 高阶传播	40
4.2.4 模型预测	42
4.2.5 模型实现	42
4.3 模型优化与训练	43
4.4 本章小结	44
第 5 章 实验分析	45
5.1 实验环境	45
5.2 对比方法介绍	45
5.3 实验设置与评价指标	46
5.4 实验结果与分析	47
5.4.1 点击率 CTR 预测	48
5.4.2 Top-K 推荐	49
5.4.3 稀疏性验证	50
5.5 消融实验	51
5.5.1 超参数深度 L 的选取	51
5.5.2 多模态与注意力层影响	53
5.5.3 聚合方式的影响	54
5.6 本章小结	55
第 6 章 总结与展望	56
6.1 总结	56
6.2 展望	57
参考文献	58

图 表 目 录

图目录

图 1-1 基于知识图谱的推荐示意图	2
图 1-2 电影海报示例图	3
图 2-1 电影领域相关知识图谱示例图	10
图 2-2 TransR的几何解释.....	11
图 2-3 DistMult模型结构	12
图 2-4 MKBE模型结构.....	14
图 2-5 多层感知机结构	18
图 2-6 卷积神经网络架构	19
图 2-7 注意力机制图	21
图 3-1 含有视觉（图片）信息的内部多模态知识图谱结构	24
图 3-2 用户物品特征图	25
图 3-3 用户-物品交互图	27
图 3-4 内部多模态协作知识图谱	28
图 3-5 TransE 模型结构.....	28
图 3-6 多种关系情况下TransE面临的局限性.....	29
图 3-7 TransR 模型结构.....	29
图 3-8 VGG16框架图	31
图 3-9 用户特征MLP处理模块结构	32
图 3-10 物品语义特征TextCNN处理模块结构.....	32
图 4-1 MKGA模型结构	35
图 4-2 内部多模态注意力传播层	37
图 4-3 基于关系的注意力传播聚合方法	39
图 4-4 K=1时，GCN高阶传播图.....	40
图 4-5 K=2时，GCN高阶传播图.....	41
图 5-1 CTR预测中的AUC，ACC	48

图表目录

图 5-2 Top-K推荐召回率	49
图 5-3 Top-K推荐准确率	50
图 5-4 嵌入传播层数(L)的影响	52
图 5-5 不同模态以及注意力机制的影响	53
图 5-6 不同聚合器的影响	54

表目录

表 3-1 实验数据表	26
表 5-1 不同训练比率下的AUC值	51

第 1 章 绪论

本章整体概述了论文的主要内容和组织结构。首先，详细介绍了本文选题的背景，之后从理论和实际意义两个层面出发，总结归纳了所做工作的研究意义。其次，对当前国内外的多模态表示学习和知识图谱推荐技术进行了概述，分析了现有研究成果存在的问题，并针对这些问题提出本文的解决方法。接下来详细阐述了本文主要研究内容。最后展示了本文的结构安排。本章内容对下文的后续研究工作起到了概括和指导作用。

1.1 研究背景及意义

1.1.1 研究背景

随着数字时代的快速发展，大量数据不断涌现，这些数据蕴含了海量的信息和知识，成为了研究者们挖掘有用信息和提高决策效率的重要资源。然而，数据的规模和复杂性给数据分析和决策工作带来了巨大挑战，因此如何从大数据中高效地提取有用的信息和知识，成为当前研究的重点任务之一。在这样的背景下，推荐系统作为一种重要的信息过滤和推送工具，已经在电商、社交网络、在线影音娱乐等领域得到广泛应用。推荐系统通过分析用户的历史行为和偏好，自动推荐用户可能感兴趣的物品或内容，从而提高了用户体验和满意度，促进了用户和平台之间的互动。Resnick 在 1997 年提出了第一个推荐系统^[1]，根据用户的行为数据向用户推荐相应的项目或网页。之后，推荐系统不断发展，历经了基于内容推荐、基于协同过滤推荐和混合推荐等多个阶段。就目前的应用情况来看，基于协同过滤的推荐算法^[2]是应用最为广泛的一种，基本流程是根据用户的历史行为数据，搜索到与之相似的用户或物品，再将其推荐给用户。

然而，协同过滤算法也存在着一些局限性，最大的问题是数据稀疏性问题。当用户和物品数量非常庞大，而用户只对小部分物品进行过操作或者评分时，推荐系统往往会面临数据稀疏性的问题。为了提高推荐算法的准确性和覆盖率，研

究者们开始探索使用其他类型的辅助信息。这些信息包括用户的社交网络关系、用户的地理位置、用户对物品的评论和评分等等。通过利用这些辅助信息，推荐系统可以更好地理解用户的兴趣和需求，从而提高推荐的准确性和覆盖率。

近年来，基于知识图谱的个性化推荐，是当前推荐系统研究的热点之一。知识图谱^[3]是一种能够将实体、关系、属性等知识组织起来，形成结构化的、可计算的知识表示方式的技术。如图 1-1 所示，知识图谱可以为个性化推荐提供更丰富的辅助信息。例如，知识图谱中的实体和属性可以为推荐系统提供更精细的描述和分类，从而更好地帮助用户理解物品的特征和差异。同时，知识图谱中的关系信息也可以为推荐系统提供更全面的背景信息，从而更好地反映不同实体之间的相互影响和依赖关系。此外，知识图谱作为一种表示知识的图形结构，还能够捕捉实体之间的语义关系，更好地反映实体间的复杂关联。通过将知识图谱与个性化推荐相结合，可以利用图形结构中的高阶信息进行更深入的挖掘。例如，对于某个用户，知识图谱可以根据用户的历史兴趣和行为，挖掘出与用户相关的实体和属性，并将它们表示为一个子图。然后，利用该子图可以推断用户的潜在兴趣和需求，从而提高推荐的准确性和覆盖率。因此，知识图谱作为一种高效的信息表示和挖掘工具，对于个性化信息推荐研究具有重要意义。

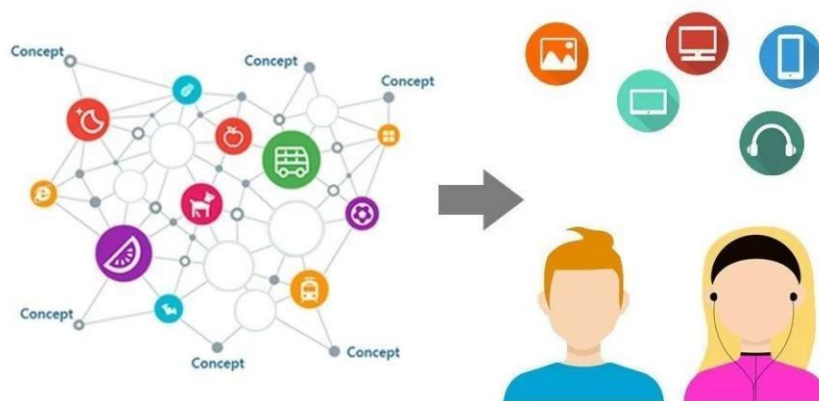


图1-1 基于知识图谱的推荐示意图

随着计算机技术的飞速发展，信息化时代已然来临，数据的形式和数量都发生了巨大的变化。移动设备和社交媒体的普及使得用户产生的数据形式和内容更加丰富，同时这也为推荐系统提供了更多的信息。多模态数据是一种同时包含

多种类型信息的数据，包括但不限于图像、文本、语音、视频等，它能够提供更加丰富的信息，帮助推荐系统更好地理解用户的兴趣和需求，从而提高推荐准确率。多模态信息正是传统的基于知识图谱推荐系统容易忽略的地方，比如语义文本描述，而这些文本特征在推荐系统中具有重要的作用。例如，在看电影之前，用户首先需要看到电影标题中的信息。传统知识图谱推荐模型会将一个物品的文本语义属性直接嵌入到一个ID中，这种方法会导致文本语义属性的深层信息被忽略。这样的电影标题所包含的一些语义信息就浪费了。另一方面，多模态信息也会影响用户的决策过程。例如，在电影院里，用户接触到最多的是电影的海报信息，通过海报里传达的人物、内容或主题类型第一时间判断用户自己是否感兴趣。同样地，在浏览腾讯视频、优酷、Youtube等网站时，如图1-2所示，用户所接触到的也是电影或电视剧的宣传图片。

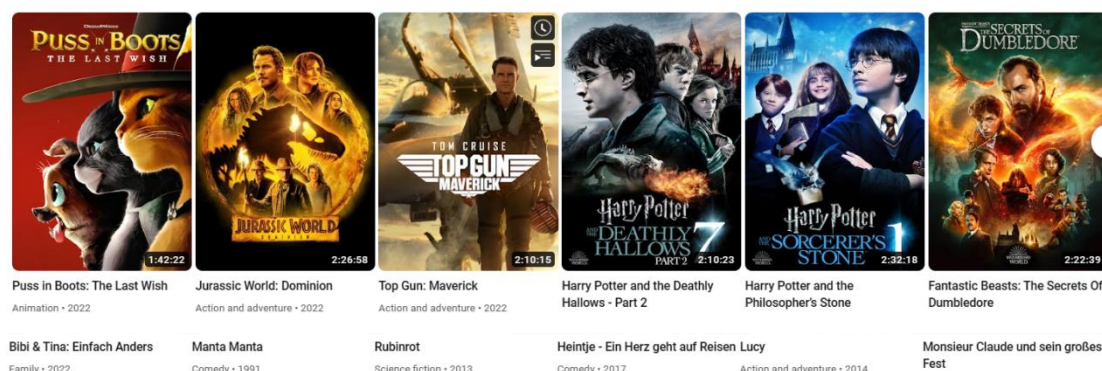


图1-2 电影海报示例图

1.1.2 研究意义

本文的理论意义在于能够进一步的提升推荐方法的效率与准确率。在推荐系统中，数据稀疏性问题是一个非常重要的问题，为了解决推荐系统中存在的数据稀疏性问题，目前的研究趋势是将外部信息引入到推荐模型中，其中知识图谱被广泛应用。然而，现有的知识图谱推荐方法主要侧重于利用知识图谱中的路径结构探索或知识图谱中的语义信息两大方面，缺乏对多模态信息的利用，同时对于知识图谱内部丰富的高阶信息没有充分提取挖掘。因此，本文提出了一种融合多模态知识图谱与注意力机制的个性化信息推荐方法。该方法主要包含了多模态数据的构建与预处理方法和基于关系空间的注意力机制传播模型两个部分。

论文将多模态信息分为外部多模态信息和内部多模态知识图谱信息两个方面。其中，外部多模态信息包括物品的文本语义属性和用户属性，这些信息被纳入到多模态信息中，从而丰富了数据特征信息。而内部多模态知识图谱则在传统知识图谱实体类型的基础上，加入了图像实体，构建了新型的内部多模态知识图谱。这一举措有助于缓解数据稀疏性问题。同时，为改进节点的嵌入方法和更好地挖掘高阶信息，本文还提出了基于关系空间的注意力机制传播模型，通过采用注意力机制来区分邻居的重要程度。本文所提方法不仅可以深度挖掘知识图谱中蕴藏的高阶信息，还可以利用多模态信息提高推荐系统的准确性和效果。

本文研究的实际意义表现在：一方面，该方法通过多个角度来丰富数据特征，缓解数据稀疏和冷启动问题，同时挖掘知识图谱内部高阶信息以提高准确性，从而满足用户的不同需求和喜好。该方法结合了多种数据源，如海报图像、用户和物品属性等信息，对用户和物品进行全方位的分析，实现更精准的推荐。该方法不仅能够提高用户的体验度和满意度，还能增强用户对推荐系统的黏度，从而提升商家的线上访问量和销售额。商家可以通过该方法更好地了解用户的需求和偏好，扩大用户群体和提高盈利能力。另一方面，该推荐方法对于推动经济发展也有积极的影响。推荐方法是线上商家平台的核心竞争力，良好的使用感受可以提高用户对平台的忠诚度，提升用户粘性。在当前国内国外经济双循环形势下，促使线上消费额不断提升，这也为线上经营者带来了收益。因此，本文所提融合多模态知识图谱与注意力机制的个性化信息推荐方法具有良好的应用价值和前景。

1.2 国内外研究现状

1.2.1 推荐系统

个性化推荐系统^[1]是一种为用户提供个性化推荐服务的应用系统，其目的是依据用户的个性化需求，从海量的信息中筛选出与之相符的信息，以提高信息利用率和用户满意度。早期的个性化推荐研究主要基于内容过滤，即根据用户的历史行为和偏好，向其推荐感兴趣的相关内容。内容过滤方法通过分析文本、图片、音频等内容信息来推荐相关物品。Wang 等人^[4]提出了一种基于内容过滤的学术

论文推荐系统,该系统利用了论文的元数据和全文内容来计算相似性,并提供了一个用户界面以获取用户反馈。作者使用了 **TF-IDF** 方法计算论文间的相似度,并使用 **Jaccard** 相似度来计算用户的兴趣相似度。实验结果表明,该系统在精度和覆盖率方面比其他方法表现更好。但是基于内容的推荐算法只考虑物品的属性,而没有考虑用户的交互行为,因此可能会出现无法推荐用户感兴趣但是没有被描述到的物品。协同过滤推荐方法可以根据用户或物品之间的相似性来进行推荐,具有良好的推荐效果。协同过滤的优点在于可以发现用户间的相似度和共同点,提供个性化推荐服务。如 **Liu** 等人^[2]提出了一种利用多通道用户反馈的深度协同过滤模型。该模型考虑到用户的多种反馈信息,包括评分、浏览和购买行为,并通过多通道嵌入的方式将这些信息进行融合。但是,该方法也存在冷启动问题和数据稀疏问题等缺点。针对上述问题,混合推荐方法被提出,将基于内容过滤和协同过滤的推荐算法结合起来。如 **Channarong** 等人^[5]将基于内容过滤和协同过滤推荐结果进行加权平均,可以得到更好的推荐效果。同时,研究者也在这一领域中探索了多种混合推荐算法,例如基于规则的推荐算法^[6]、基于图的推荐算法^[7]和基于矩阵分解的推荐算法^[8]等。这些混合推荐算法的优点在于可以充分利用内容信息和用户行为信息,提供更为准确的个性化推荐服务,但从根本上并没有解决数据稀疏性问题,并且对于大规模的数据和复杂模型无法有效处理。

近年来,随着深度学习技术的不断发展,一些深度学习模型^[9]也被应用于个性化推荐系统,如基于神经网络的推荐模型和深度协同过滤模型。这些模型不仅具有更强的推荐效果,还可以处理大规模的数据和复杂模型。如 **Li** 等人^[10]提出的基于神经网络的推荐模型可以通过自编码器 (**AutoEncoder**) 来学习用户和物品的潜在特征,从而提高推荐效果。在深度协同过滤方面,**Zhang** 等人^[11]通过引入多层神经网络来捕捉用户和物品之间的非线性关系,进一步提高推荐效果。然而,这些深度学习模型也存在一些缺点,如模型的可解释性较差、需要大量的计算资源等。知识图谱^[3]是近年来被广泛应用于个性化推荐领域的一种技术,它是一种用于描述实体之间关系的图形化知识表示方法,可以利用知识图谱中的实体、关系和属性等信息为用户提供更准确和个性化的推荐服务。在知识图谱推荐方面,近年来涌现出很多研究成果。如王利娥等人^[12]提出了一种基于知识图谱的新闻推荐算法,该模型使用知识图谱来表示新闻之间的语义关系,并结合用户历

史行为数据进行推荐,取得了较好的实验效果。Guo 等人^[13]提出了一种基于知识图谱的个性化音乐推荐系统。该系统利用知识图谱表示音乐数据之间的语义关系,结合用户的历史听歌行为和音乐特征,采用协同过滤算法进行推荐,为用户提供更准确和个性化的音乐推荐服务。然而目前的知识图谱推荐通常只关注实体之间的一阶关系,即实体之间的直接关系。但在实际应用场景中,很多实体之间存在更高阶的关系,例如通过共同拥有属性或者参与共同事件等方式产生的间接关系。而目前知识图谱推荐对更高阶的关系和复杂的图结构没有进行充分的挖掘和利用。

1.2.2 多模态表示学习

多模态表示学习^[14]是指通过将多种数据以数据序列的形式呈现,学习并得到最终的嵌入向量表示。这种方法可以将多种数据类型(如图像、音频、文本等)进行有效地结合,以提高预测的准确性和发现数据内在的关系特征。多模态学习在其他领域也有着广泛应用。

在语音识别领域^[15],多模态学习可以结合图像和文本等信息,以提高语音识别的准确性和稳定性。在图像识别领域^[16],多模态学习可以结合文本和语音等信息,以提高图像分类、目标检测和图像生成等任务的性能。变量解码和多模自动编码也是多模态学习的常见应用。变量解码^[17]是指在多模态学习中,将不同形式的数据分别表示为不同的潜在变量,并且通过这些潜在变量来进行跨模态信息的交互。具体来说,变量解码的过程可以被视为将每种数据形式(例如图像和文本)映射到其各自的潜在变量,然后使用这些潜在变量作为输入,通过某种方式对它们进行组合以生成多模态数据。多模自动编码^[18]则是一种学习多模态数据表示的方法,它将多种形式的数据编码为低维表示,然后使用解码器将它们恢复为原始数据形式。与传统的自动编码器不同的是,多模自动编码器同时处理多个数据形式,因此它可以通过学习数据的交互来提高模型的性能。在多模态推荐中,多模态学习可以结合多种数据类型,如知识图谱、用户画像、内容理解、海报和声音等信息,以生成最终的向量表示。如 Yang^[19]等人提出的基于注意力的多模态神经网络模型(AMNN)可以学习多模态微博的表示并推荐相关的主题标签。李琳^[20]提出利用用户的多种数据类型(如音乐播放记录、用户标签、歌手信息等)

进行音乐推荐,以提高推荐的准确性和个性化程度。Tao 等人^[21]提出了一种称为 MGAT 的方法,它通过利用微视频项目的视觉、听觉和文本特征来执行电影特征嵌入。

1.3 主要研究工作

本文针对推荐系统存在的数据稀疏性问题以及知识图谱内部高阶信息没有得到充分利用的现状,提出了一种融合多模态知识图谱与注意力机制的个性化信息推荐方法,该方法主要包含了融合多模态数据的构建与预处理方法和基于关系空间的注意力机制传播模型两个部分。本文的主要研究内容如下:

(1) 现有的推荐系统存在着数据稀疏性问题,如果一个用户对物品的评价或行为数据非常少或者没有,将难以进行训练和推荐。针对这一问题,本文提出一种多模态数据的构建与预处理方法,将多模态信息分为内部多模态知识图谱信息和外部多模态信息。首先,利用视觉图像作为内部多模态信息对知识图谱推荐进行扩展。然后将物品的语义文本和用户属性作为外部多模态特征弥补知识图谱嵌入带来的语义损失问题,从而更好地提高推荐系统的有效性和准确性。

(2) 针对在知识图谱推荐中的高阶信息挖掘不充分现状,本文提出了一种基于关系空间的注意力机制传播模型。首先将从内部多模态知识图谱中得到的多模态特征向量作为输入,然后采用基于关系空间的感知注意力机制来学习传播过程中的每个邻居权重,同时挖掘聚合用户和物品的高阶信息,在关系空间中找到关系更近的实体。在特征聚合阶段,本文使用相加、连接和双端混合三种方式进行聚合效果测试,实验表明采用双端混合的聚合方式效果最好,可以更好的聚合特征以提高推荐性能。最后,将外部多模态用户和物品特征向量加入与其融合得到最终的用户和物品特征向量,使用预测函数进行推荐预测。

(3) 本文开展了充分的实验分析,主要工作是在构建的多模态数据集 Multi-MovieLens1M 上进行对比实验,以准确率、召回率等作为评价指标,比较本文所提方法与当前主流推荐方法在推荐效果上的差异,经过实验验证,本文所提方法对比目前主流推荐方法在准确率、召回率以及 ACC 等指标上均有所提升。同时,通过稀疏实验验证了在稀疏场景中本文所提方法的可行性。最后,通过消融实验详细说明了本文所提方法的各模块效用,并研究它们之间的相互作用,以进一步

验证所提方法的有效性。

1.4 论文结构与安排

本文共分为六个章节，分别介绍如下：

第1章 绪论。本章首先对推荐系统相关的研究背景进行详细介绍，提出了对多模态知识图谱推荐开展研究的目的和意义。然后针对推荐系统和多模态学习表示的国内外研究现状进行描述，并指出了已有研究成果存在的不足，最后概述了本文的主要研究内容和结构安排。

第2章 相关理论。本章主要对于研究所用的相关理论和技术进行概述。按照本文研究的具体内容，依次介绍了知识图谱、基于知识图谱的推荐以及所涉及到的深度学习相关理论技术，并总结了现有研究成果的贡献和不足。

第3章 多模态数据的构建与预处理方法。本章首先介绍了多模态数据的构建与预处理的方法思想，提出对知识图谱进行多模态扩展；然后将物品的相关图像作为多模态扩展实体连接物品本身形成内部多模态知识图谱，同时引入物品和用户本身特征作为外部多模态辅助信息，为后文的多模态信息融合做准备。

第4章 基于关系空间的注意力机制传播模型。本章首先给出了模型的框架，然后将上一章得到的内外部多模态用户物品特征信息通过基于关系空间的注意力机制模型进行融合处理，最后使用预测函数进行推荐预测。

第5章 实验分析。本章首先说明了实验环境和实验用到的数据集；然后介绍了对比方法和实验用到的评价指标；最后通过与主流知识图谱推荐方法MKR,RippleNet等进行对比实验，验证了本文所提方法的有效性。

第6章 总结与展望。在总结本文所做研究工作的基础上，归纳本文研究内容和待完善之处，并展望了今后的工作重点。

第2章 相关理论

本章主要阐述了融合多模态知识图谱与注意力机制的个性化信息推荐研究方法中所涉及到的相关理论知识。首先对于本文所使用的知识图谱进行了介绍，描述了关于知识图谱方面的相关表示技术，如知识图谱的表示方法以及多模态知识图谱的表示方法；然后进一步介绍了基于知识图谱的推荐相关技术，如基于嵌入的方法和基于路径的方法等；最后介绍了深度学习方面常用的几种模型架构，包括多层感知机、卷积神经网络以及注意力机制等，本章内容为论文研究提供理论基础和技术支撑。

2.1 知识图谱

知识图谱^[22]是一个描述实体之间关系的图形化知识表示方法，是机器学习、自然语言处理、人工智能等领域中重要的研究方向之一。早在上世纪 60 年代提出的语义网是一种将网络数据转换为可读性更强的格式方法，这一技术为知识图谱的出现奠定了基础。知识图谱^[22]就是在此基础上发展起来的，它通过引入更多的语义概念和数据挖掘技术，使知识图谱能够更好地理解和表示实体之间的关系。

知识图谱的核心是实体和关系。实体是指现实世界中的一个具体或抽象的事物，如人、公司、产品等。关系是实体之间的连接或联系，如“父亲”和“出生于”等。如图 2-1 所示的电影领域相关知识图谱就是不同的实体和关系相连而形成。实体和关系可以采用三元组的形式来表示，即 (h, r, t) ，其中 h 是头实体， r 是关系， t 是尾实体。

知识图谱的应用范围非常广泛，包括搜索引擎、智能问答、推荐系统、智能客服、自动化决策等领域。例如，搜索引擎^[23]可以使用知识图谱来理解用户的查询意图和相关实体之间的关系，从而能够提供更准确的搜索结果。智能问答^[24]可以使用知识图谱来回答用户的问题，通过查询知识图谱中的实体和关系，找到最相关的答案。推荐系统^[25]可以使用知识图谱来发现不同实体之间的关系，从而向用户推荐更符合他们兴趣和需求的内容。知识图谱可以由人工或自动构建，通常

采用语义网技术和图数据库技术进行存储和查询。

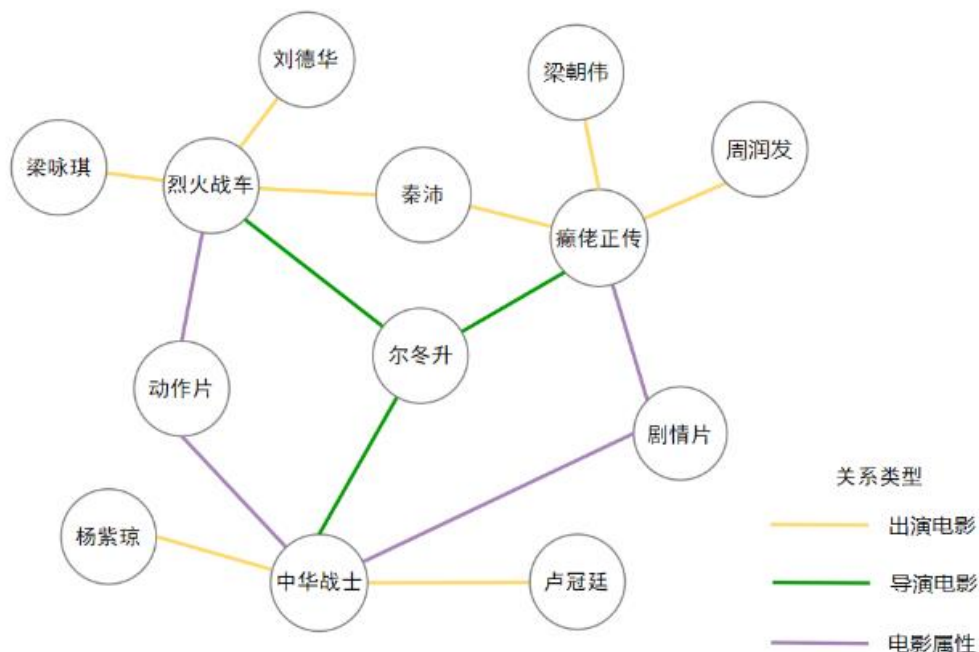


图2-1 电影领域相关知识图谱示例图

2.1.1 知识图谱嵌入方法

知识图谱嵌入（Knowledge Graph Embedding）^[26]是将知识图谱中的实体和关系映射到低维向量空间的一种技术。嵌入方法可以将实体和关系表示为低维向量，使得实体和关系之间的语义关系可以在向量空间中进行计算。在知识图谱应用中，嵌入方法常用于实体对齐、关系预测、推荐系统等任务。嵌入方法可被分为两种，平移距离的翻译模型以及语义匹配模型。

（1）翻译模型

TransR^[27]是在 TransE^[28]的基础上进一步改进的，如图 2-2 所示，它的思想是将实体和关系映射到不同的向量空间中，即将实体和关系表示为不同的向量，并定义一个矩阵 M_r 来描述实体和关系之间的映射关系。假设实体 e 和关系 r 分别表示为向量 $e \in \mathbb{R}^{k_e}$ 和 $r \in \mathbb{R}^{k_r}$ ，则实体和关系可以表示为如公式(2.1)和(2.2)所示：

$$e_M = M_r e \quad (2.1)$$

$$r_M = M_r r \quad (2.2)$$

其中, e_M 和 r_M 分别表示在矩阵 M_r 的作用下, 实体 e 和关系 r 映射到新的向量空间中的向量。

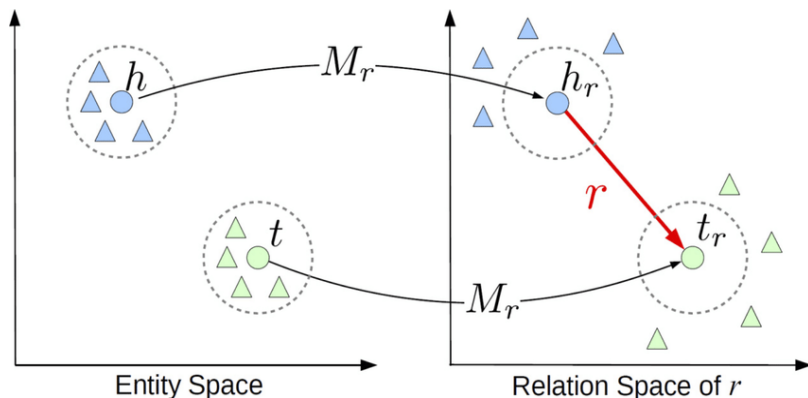


图2-2 TransR的几何解释

然后, 可以定义一个函数 f 来计算投影后的向量, 实体和关系可以采用三元组的形式来表示, 即 (h, r, t) , 其中 h 是头实体, r 是关系, t 是尾实体。如公式 (2.3) 所示:

$$f_r(h) = e_M + r_M \quad (2.3)$$

TransR 的损失函数分为两部分: 边界损失函数和关系损失函数。

边界损失函数用于约束实体嵌入向量在不同关系空间中的变换, 防止实体嵌入向量在变换后跨越不同关系空间的边界。边界损失函数如公式 (2.4) 所示:

$$L_b = \sum_{(h,r,t) \in S} [\|M_r h - m_r\|^2 + \|M_r t - m_r\|^2 - 2\alpha] \quad (2.4)$$

其中, S 表示训练集中的三元组集合, h 和 t 分别表示头实体和尾实体对应的嵌入向量, r 表示关系对应的嵌入矩阵, M_r 表示将头实体嵌入向量变换到关系空间中的变换矩阵, m_r 表示关系空间中的原点, α 表示边界的宽度。

关系损失函数用于度量不同关系之间的语义距离, 使得相似关系之间的距离更小。关系损失函数如公式 (2.5) 所示:

$$L_r = \sum_{(h,r,t) \in S} [\|r_h + m_r - r_t\|^2 - \|r_h + m_r - r_{t'}\|^2 + \gamma] \quad (2.5)$$

其中, r_h 和 r_t 分别表示头实体和尾实体在关系空间中的嵌入向量, t' 表示与 h 在训练集中出现在同一三元组中但不是尾实体的实体, γ 表示损失函数的边界

宽度。

总的损失函数如公式 (2.6) 所示:

$$L = L_b + \beta L_r \quad (2.6)$$

其中, L_b 为边界损失函数, L_r 为关系损失函数, β 为权衡两部分损失函数的超参数。

(2) 语义匹配模型

DistMult^[29] 是一种基于矩阵分解的知识图谱嵌入模型, 如图 2-3 所示, 它通过对三元组中的实体和关系进行低维度向量表示来学习知识图谱的嵌入。DistMult 模型的核心思想是使用标量积来计算三元组中头实体、关系和尾实体之间的相似度, 它的目标函数可以定义为公式 (2.7):

$$f(h, r, t) = \langle h, r, t \rangle = e_h^T R_r e_t \quad (2.7)$$

其中, e_h 和 e_t 分别是头实体和尾实体在向量空间中的表示, R_r 是关系 r 的表示矩阵, " \langle, \rangle " 表示两个向量的标量积。

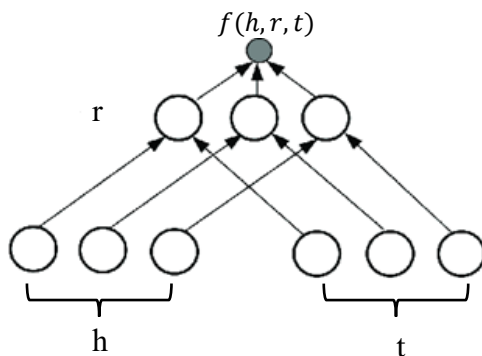


图2-3 DistMult模型结构

为了使得学习到的嵌入更具有区分度, 需要通过最大化正例的得分和负例的得分之间的差异来优化模型, 可以使用以下的损失函数, 如公式 (2.8) 所示:

$$\mathcal{L} = - \sum_{(h,r,t) \in \mathcal{K}} \sum_{(h',r,t') \in \mathcal{K}'} \log \sigma(f(h,r,t) - f(h',r,t')) + \lambda(|e_h|^2 + |e_t|^2 + |R_r|^2) \quad (2.8)$$

其中, \mathcal{K} 表示训练集, \mathcal{K}' 表示从 \mathcal{K} 中随机采样得到的负例集合, $\sigma(\cdot)$ 是 sigmoid 函数, λ 是正则化参数, $e_h, e_t \in \mathbb{R}^k$, $R_r \in \mathbb{R}^{k \times k}$ 。

相比于其他的矩阵分解模型, DistMult 模型具有计算效率高和空间复杂度低的优点, 因此在许多知识图谱应用中得到了广泛的应用。但它也有一些局限性,

例如不能很好地处理具有对称性质的关系和多重关系等问题。

2.1.2 多模态知识图谱表示方法

多模态知识图谱表示学习是将多种模态的数据（例如文本、图像、音频等）以及知识图谱中的实体和关系进行统一的表示，以便更好地理解 and 利用这些数据。多模态知识图谱表示学习方法主要包括两种：基于特征的方法和基于实体的方法。基于特征的方法^[30]通常将每种模态的数据都转化为向量表示，然后将这些向量进行拼接或者加权求和，得到一个综合的向量表示。基于实体的方法^[31]则是将每个实体表示为一个向量，而这个向量同时包含了实体本身的信息和与其相关的多种模态数据的信息。这种方法可以同时考虑实体本身的语义信息和其所关联的模态数据的语义信息，从而得到更加准确的表示。

（1）基于特征的方法

基于特征的方法是一种用于多模态知识图谱表示学习的技术，它将知识图谱中的实体和关系表示为多种不同类型的特征向量，包括结构特征、文本特征、图像特征等等，然后将这些特征向量融合在一起，得到一个综合的多模态表示。这种方法的优点是可以充分利用多种不同类型的信息，提高知识图谱的表示能力和应用效果。

具体来说，基于特征的方法通常包括以下步骤：

① 提取实体和关系的多种特征：例如，可以使用自然语言处理技术从文本中提取实体的词向量表示，使用图像处理技术从图像中提取实体的视觉特征，使用知识图谱本身的结构信息提取实体和关系的拓扑特征等等。

② 将不同类型的特征向量融合在一起：将不同类型的特征向量通过一些融合技术进行合并，例如将其串联^[32]在一起，加权求和，或者通过一些深度神经网络模型^[33]进行融合。这个过程的目的是得到一个综合的、包含多种信息的实体和关系表示向量。

③ 用表示向量进行下游任务：将得到的多模态表示向量用于下游任务，例如实体分类、关系预测、推荐等等。在这些任务中，可以使用不同的机器学习算法，例如支持向量机、逻辑回归^[34]、神经网络等等，来对实体和关系进行分类、预测、推荐等操作。

基于特征的方法是一种有效的多模态知识图谱表示学习技术，它能够利用不同类型的信息提高知识图谱的表示能力和应用效果。然而，由于特征提取和融合过程的复杂性，这种方法往往需要大量的人工设计和优化，因此在实际应用中还需要进一步研究和改进。

(2) 基于实体的方法

基于实体的方法是其中的重要研究方向之一。相对于基于特征的方法，基于实体的方法不仅考虑了多模态信息对实体属性的影响，还将多模态信息作为结构化知识加入到知识图谱中。这种方法不仅能够有效地保留多模态信息的丰富性，还能够通过建立新的关系将多模态实体与其他实体进行关联，从而更好地挖掘知识图谱中的信息。

MKBE^[13]是基于实体方法的代表性工作之一，如图2-4所示，其核心思想是将不同模态的信息视为结构化知识，并将其直接作为实体加入到知识图谱中。在MKBE中，视觉模态数据使用预训练的VGG^[35]模型进行编码，数值模态数据使用全连接层进行编码，而结构化的知识则使用嵌入层进行编码。通过这种方法，MKBE能够将不同模态的信息有效地结合起来，提高知识图谱的表示能力。

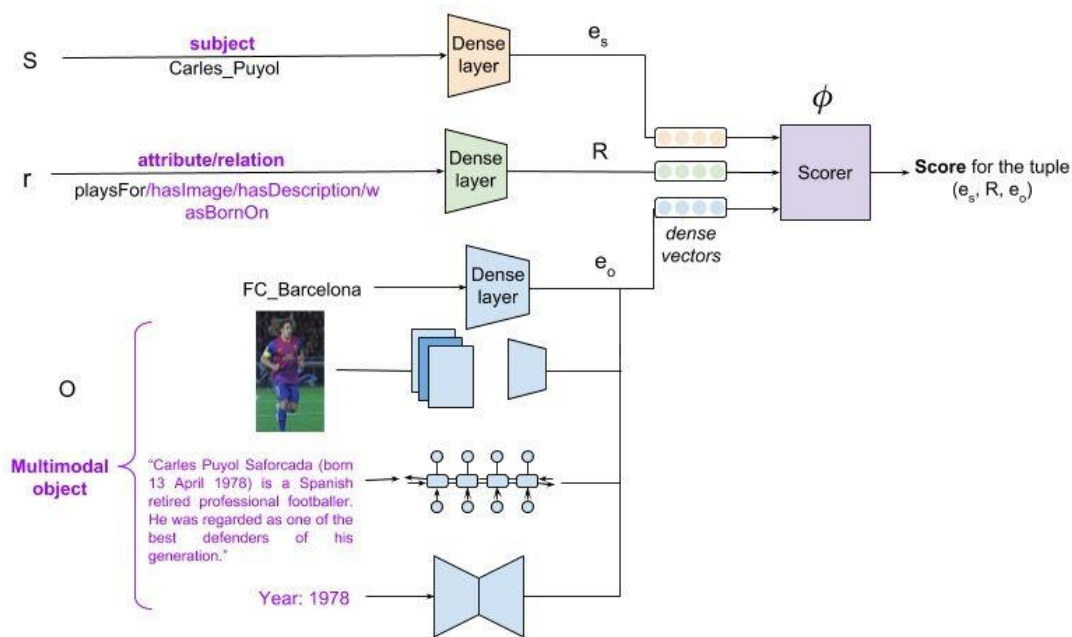


图2-4 MKBE模型结构

总的来说,基于实体的方法是多模态知识图谱表示学习的重要研究方向之一。这种方法能够有效地将多模态信息结合起来,提高知识图谱的表示能力,从而更好地挖掘知识图谱中的信息。

2.2 基于知识图谱的推荐方法

传统的推荐系统往往仅仅基于用户历史行为进行推荐,无法准确反映用户的兴趣和偏好。而知识图谱作为一种结构化的知识表达方式,可以将丰富的语义信息组织起来,从而为推荐系统提供更加全面、准确的数据支持。基于知识图谱的推荐方法不仅可以利用知识图谱中的实体和关系信息解决推荐系统中的冷启动和数据稀疏等问题,还可以将用户的历史行为路径与知识图谱中的信息相结合,生成更加个性化的推荐结果。因此,基于知识图谱的推荐方法已成为推荐系统研究领域的热门方向之一,受到学术界和工业界的广泛关注和应用。利用知识图谱的推荐系统一般分为两大类:基于嵌入的方法和基于路径的方法。

2.2.1 基于嵌入的方法

基于嵌入的方法^[36]是将知识图谱中的实体和关系映射为低维嵌入向量,将这些嵌入向量作为输入,通过一个推荐模型来预测用户与项目之间的关系,从而实现推荐。基于嵌入的方法是一种常见的知识图谱推荐方法,其重要优点是可以捕捉到知识图谱中的丰富信息,包括实体的语义、属性以及关系等信息。这些信息可以提供有助于产生更准确推荐的上下文信息,从而提高推荐的质量和效果。在实际应用中,基于嵌入的方法可以与不同的推荐模型结合使用。例如,Zhang等人^[37]提出的CKE就是一种将知识图谱嵌入方法与基于协同过滤的推荐模型相结合的方法。该算法将知识图谱中的项目和用户映射为低维嵌入向量,将用户和项目的交互行为作为用户偏好和项目属性之间的关系,与知识图谱中的属性关系共同学习得到用户和项目的表示向量,然后将这些向量输入到基于协同过滤的推荐模型中,计算用户和项目之间的相似度,预测用户对项目的评分。徐等人^[38]提出了一种KGE-CF模型,它结合了多层感知机以捕捉高阶非线性特征,可以更深层次地学习用户与项目之间的交互信息,从而提高推荐的质量。该模型将

用户历史交互的项目与知识图谱中的实体进行映射，并使用知识图谱的翻译模型进行训练，从而得到实体嵌入向量和关系向量。Song 等人^[39]提出了一种基于深度强化学习的推荐模型，该模型分为嵌入部分和排序部分。嵌入部分使用 TransE 算法将用户和项目表示为低维向量，排序部分使用 DQN 算法^[40]将推荐问题转化为强化学习问题，并根据用户反馈调整推荐列表，可以对推荐列表进行排序以提高整体推荐效果。

2.2.2 基于路径的方法

基于路径的方法是一种常见的知识图谱推荐方法，它利用知识图谱中实体之间的关系路径来预测用户与项目之间的关系，从而实现推荐。这种方法的一个重要优点是它可以捕捉到实体之间的关系和上下文信息，从而提高推荐的质量和效果。具体地，基于路径的方法一般包含以下几个步骤：

（1）路径提取：从知识图谱中提取与用户历史行为相关的路径。这些路径可以是用户历史行为节点到其他实体节点的任意长度路径，例如，用户-交互行为-项目或用户-兴趣标签-项目等路径。

（2）路径嵌入：将路径映射为低维嵌入向量。这一步通常可以使用诸如深度学习等技术，将路径中的实体和关系表示为向量形式，并将这些向量通过聚合、加权等方式组合成路径的嵌入向量。

（3）路径评分：使用一些方法对路径进行评分。例如，基于路径的相似度计算方法，例如余弦相似度等，来衡量路径与用户历史行为的相似程度。

（4）推荐排序：将得到的路径评分与其他推荐算法（例如基于嵌入的推荐算法或基于协同过滤的推荐算法）中得到的推荐列表进行排序，得到最终的推荐结果。

有一些基于路径的方法将知识图谱看作一个异质的信息网络，利用不同元路径中实体的语义相似性细化用户和项目的表示。例如刘丹阳^[41]提出了一种基于异构图注意力网络的新闻跨平台推荐模型。该模型通过一种基于路径的注意力机制，结合了知识图谱信息，能够在不同平台的新闻推荐中取得更好的性能。该模型在知识图谱中学习实体和关系的表示，并使用注意力机制来自适应地选择路径以捕捉实体和关系之间的上下文信息，最终将得到的表示用于推荐任务。

中。此外还有一些方法将实体嵌入和路径嵌入进行联合学习，例如 Li 等人^[42]提出了一种统一的基于路径的方法来进行知识图谱嵌入和推荐，通过基于路径的注意机制联合学习实体嵌入和路径嵌入，以更好地捕捉实体之间的复杂关系。同时，该方法还考虑到了实体和路径之间的交互关系，从而进一步提高了嵌入的质量和推荐的效果。

2.3 深度学习

深度学习应用于推荐系统的基本思想是通过学习用户和物品之间的映射关系来实现个性化推荐。在这个过程中，深度学习模型将用户和物品表示为向量，并利用这些向量进行推荐。具体来说，深度学习模型使用神经网络来学习用户和物品之间的非线性映射关系，从而更好地捕捉它们之间的相似性和差异性。这些模型可以从用户的历史行为、物品的特征和其他上下文信息中进行学习，以生成个性化的推荐结果。深度学习模型^[43]在推荐系统中的优点之一是能够处理大规模、高维度的数据集，并能够自动学习高质量的特征表示。此外，深度学习模型还可以捕捉到用户和物品之间的复杂关系，比如长期和短期的兴趣漂移以及用户和物品的交互效应等。这使得深度学习模型比传统的推荐算法更加适用于实际场景中的个性化信息推荐。下面将对本文涉及到的深度学习技术进行简单介绍。

2.3.1 多层感知机

多层感知机（Multilayer Perceptron, MLP）^[44]是一种基于前馈神经网络的监督学习算法，如图 2-5 所示，它由多个全连接层组成，可以用于解决分类和回归问题。在每个全连接层中，输入数据通过权重和偏置进行线性变换，然后通过一个激活函数进行非线性映射。多层感知机的输出层通常是一个 softmax 层或一个线性层，用于进行分类或回归预测。下面是多层感知机的一般形式：

给定输入数据 $X \in R^{n \times d}$ ，其中 n 表示样本数， d 表示输入特征的维度。多层感知机通过将输入数据传递给多个全连接层，其中第 i 层的输出 $H^{(i)}$ 可以表示为如公式（2.9）所示：

$$H^{(i)} = f(W^{(i)}H^{(i-1)} + b^{(i)}) \quad (2.9)$$

其中 $W^{(i)} \in R^{h_{i-1} \times h_i}$ 和 $b^{(i)} \in R^{1 \times h_i}$ 分别表示第 i 层的权重和偏置, h_{i-1} 和 h_i 表示第 $i-1$ 层和第 i 层的神经元数目, $H^{(0)} = X$ 表示输入数据。 f 是非线性激活函数, 如 sigmoid 函数、ReLU 函数、tanh 函数等。

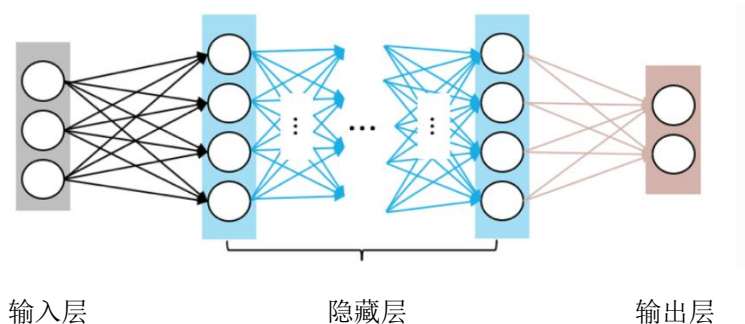


图2-5 多层感知机结构

多层感知机的输出层通常是一个 softmax 层或一个线性层。对于分类问题, 输出层通常使用 softmax 函数将每个类别的预测概率归一化到 0 到 1 之间, 如公式 (2.10) 所示:

$$\hat{y} = \text{softmax}(W^{(L)}H^{(L-1)} + b^{(L)}) \quad (2.10)$$

其中 L 表示输出层的编号, \hat{y} 表示每个类别的预测概率, $W^{(L)} \in R^{h_{L-1} \times c}$ 和 $b^{(L)} \in R^{1 \times c}$ 分别表示输出层的权重和偏置, c 表示类别数目。

对于回归问题, 输出层通常是一个线性层, 直接输出预测值, 如公式 (2.11) 所示:

$$\hat{y} = W^{(L)}H^{(L-1)} + b^{(L)} \quad (2.11)$$

在训练多层感知机时, 通常使用反向传播算法进行优化, 目标是最小化损失函数。常见的损失函数包括平均方差损失 (Mean Squared Error, MSE) 和交叉熵损失函数 (Cross-entropy loss)。以 MSE 为例, 其定义如公式 (2.12) 所示:

$$L(\theta) = \frac{1}{2n} \sum_{i=1}^n |\hat{y}_i - y_i|^2 \quad (2.12)$$

其中, θ 表示模型参数, n 表示样本数, \hat{y}_i 表示第 i 个样本的预测值, y_i 表示第 i 个样本的真实值。通过反向传播算法计算损失函数对模型参数的梯度, 并使用优化算法 (如随机梯度下降) 更新模型参数。训练多层感知机通常需要大量的数据和计算资源, 但它在许多实际问题中都取得了非常好的效果。

总之，多层感知机是一种常用的神经网络结构，可以用于解决分类和回归问题。它通过多个全连接层和非线性激活函数进行非线性映射，同时使用反向传播算法进行优化，可以实现较为准确的预测。

2.3.2 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）^[45]是一类广泛应用于计算机视觉和自然语言处理等领域的神经网络模型。它的核心思想是通过卷积和池化操作实现特征提取，从而有效地解决图像和文本等复杂数据的分类和识别问题。如图 2-6 所示的卷积神经网络架构，主要分为卷积层，池化层和全连接层，以下是卷积神经网络的详细介绍：

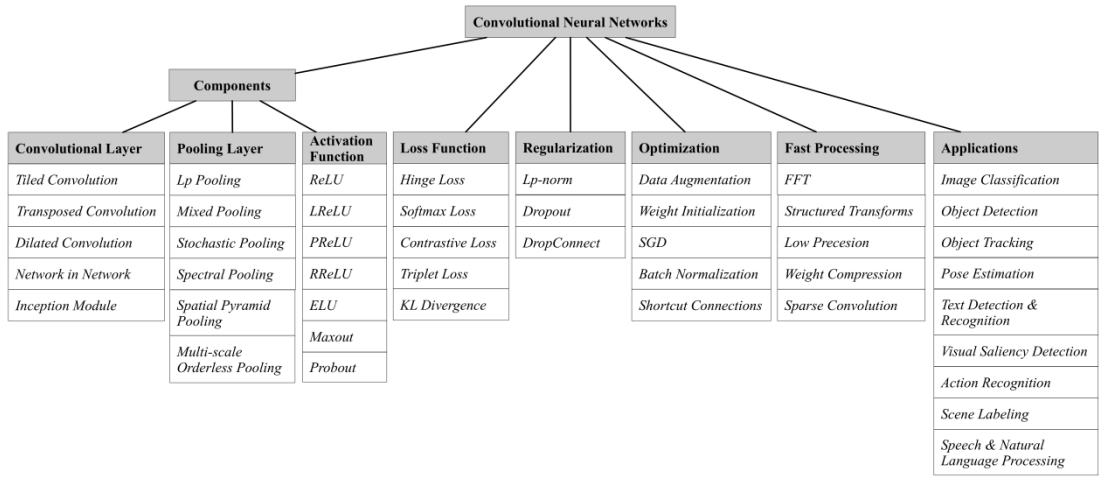


图2-6 卷积神经网络架构

（1）卷积层（Convolutional Layer）

卷积层是卷积神经网络的核心组件之一，它主要用于从原始数据中提取特征。具体来说，卷积层使用一些固定大小的滤波器（也称为卷积核或权重），将它们与输入数据进行卷积运算，得到一组新的特征映射。卷积运算可以看作是将滤波器在输入数据上进行滑动，计算出每个位置的卷积结果。滤波器的大小通常是正方形或矩形，比如 3x3、5x5 等，它的深度与输入数据的深度相同。

（2）池化层（Pooling Layer）

池化层是卷积神经网络中另一个重要的组件，它主要用于对特征图进行下采样，从而减少计算量和内存消耗。池化层通常会对特征图的每个子区域进行池化

操作，输出一个新的特征图。池化操作通常有两种方式，分别为最大池化（Max Pooling）和平均池化（Average Pooling）。最大池化会选取特征图中每个子区域中的最大值作为池化结果，而平均池化则是取子区域中的平均值。

（3）全连接层（Fully Connected Layer）

全连接层又可称为密集连接层（Dense Layer），包含多个神经元且每个神经元与上一层的所有神经元都有连接。它主要用于将卷积层和池化层提取的特征进行分类或回归等任务，是卷积神经网络中的常见组件之一。

除了以上三种层之外，卷积神经网络中还包括其他的组件，如批量归一化（Batch Normalization）、残差网络（Residual Network）等。这些组件的使用可以进一步提高卷积神经网络的性能。

2.3.3 注意力机制

注意力机制^[46]（Attention Mechanism）能够为输入数据的不同部分分配不同的注意力权重，它一直被广泛应用深度学习模型中以提高模型的表现。最初注意力机制被用于机器翻译中，随后被扩展应用于语音识别、图像处理、自然语言处理等领域。在深度学习模型中，通常会有多个输入信号，比如图像中的不同像素点、文本中的不同词语等。这些输入信号可能存在一定的相关性，也可能存在差异性。注意力机制的数学表示通常包含以下三个步骤：首先计算输入信号的权重，然后根据权重对输入信号进行加权，最后将加权后的信号作为输出。以文本分类任务为例，假设输入是一个长度为 n 的文本序列 $x = (x_1, x_2, \dots, x_n)$ ，每个 x_i 是一个 d 维的词向量。注意力机制的数学表示可以如下：

（1）计算权重

首先需要计算每个输入 x_i 的权重 w_i ，权重表示了 x_i 在整个输入序列中的重要性。常用的计算权重的方法如点积注意力（Dot-Product Attention），如具体公式（2.13）所示：

$$w_i = \frac{\exp(\text{score}(q, k_i))}{\sum_{j=1}^n \exp(\text{score}(q, k_j))} \quad (2.13)$$

其中 q 是查询向量， k_i 是与 x_i 相对应的键向量， $\text{score}(q, k_i)$ 是一个评分函数，用于评估 q 和 k_i 之间的关联程度。点积注意力计算的是 q 和 k_i 的点积，即

$\text{score}(q, k_i) = q^T k_i$ 。

(2) 计算加权和

根据权重 w_i 对每个输入 x_i 进行加权求和，从而得到加权和 z ，如公式(2.14)所示：

$$z = \sum_{i=1}^n w_i x_i \quad (2.14)$$

(3) 输出

将加权和 z 作为输出，可以用于后续的分类或其他任务，总体过程如图2-7所示。

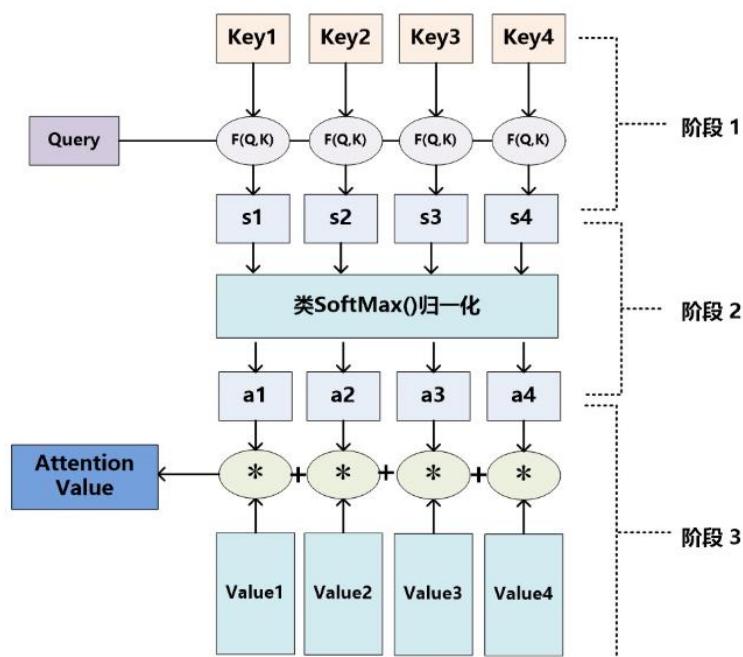


图2-7 注意力机制图

2.4 本章小结

本章首先对知识图谱的发展历程以及相关概念进行了简要的介绍；其次列举了常见的知识图谱嵌入方法，并介绍了多模态知识图谱的两种表示方法，即基于特征的方法和基于实体的方法。同时对于部分方法的工作原理和不足之处进行了介绍以便更为系统地阐述本文地研究内容和创新点；然后将本文所用的基于知识图谱的推荐方法进行了介绍；最后列举了多层感知机、卷积神经网络和注

注意力机制等深度学习模型架构。本章综述的内容为论文研究工作的开展提供了坚实的理论和技术支撑。

第3章 多模态数据的构建与预处理方法

通过第2章中对现有知识图谱推荐方法的贡献与不足的分析,本文提出了一种融合多模态知识图谱与注意力机制的个性化信息推荐方法。本章对其中的多模态数据的构建与预处理方法进行介绍,首先对知识图谱进行多模态扩展,采用基于实体的方法将物品的相关图像作为多模态特征连接物品实体本身形成内部多模态知识图谱,从而更加有效地丰富了知识图谱。同时引入物品和用户自身特征作为外部多模态辅助信息,为多模态信息融合做好准备。本章为后续章节基于关系空间的注意力机制传播模型提供了重要的数据基础。

3.1 方法思想

目前推荐系统的主流技术主要采用协同过滤推荐、基于内容的推荐以及混合推荐等多种技术。通过协同过滤推荐,可以有效地利用用户的历史行为信息,但是,随着用户产生的交互数据越来越少时,这种方法的精确度就会受到影响;同时,对于新加入的用户或物品会存在冷启动等问题。因此,为了提高推荐的效率和性能,本章通过使用知识图谱作为扩充关联信息,从而有效减轻数据稀疏和冷启动问题带来的影响。

随着时间的推移,知识图谱已经成为一种重要的工具,它能够帮助人们更好地理解日常生活中的相关性。通过将知识图谱与推荐相结合,不仅能更好地了解用户的需求,而且还能更准确地提供有价值的建议,从而在解释性方面提高用户对推荐系统的信任度。但知识图谱推荐更多的侧重于结构联系方面,对于更加丰富的多模态信息没有利用。因此,本章提出多模态数据的构建与预处理方法,在引入知识图谱的基础上,加入多模态信息,更多维度的描述物品用户特征,从而对于解决数据稀疏问题提供了有力支撑,可以更好的提高推荐系统的性能和准确性。

在方法的预处理阶段,本文把多模态信息分为了内部多模态知识图谱和外部多模态信息。其中内部多模态知识图谱是对传统知识图谱的扩充,将视觉信息,例如图片以基于实体的方法对物品进行补充修饰,形成一个范围更大的知识图

谱。做到这一点，需要对传统数据集进行修改，因为目前传统知识图谱的数据集只包含结构信息，例如导演-指导-影片这种结构知识，本文将图片插入之后，变为电影-关系（hasImage）-宣传海报，如图 3-1 所示。在此基础上，本章使用改进的 TransR 方法将知识图谱进行嵌入向量表示，加入对应编码器用于提取相关图片特征。

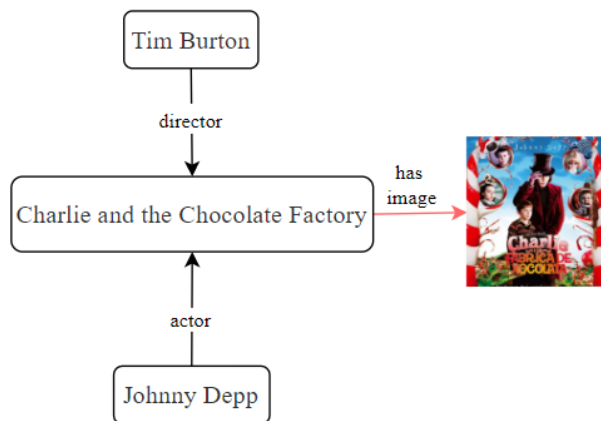


图3-1 含有视觉（图片）信息的内部多模态知识图谱结构

对于外部多模态信息，将用户物品特征作为外部多模态信息纳入进来。在实际物品推荐过程中，需要收集来自不同来源的数据，包括物品信息、用户信息和用户与物品之间的交互信息。每个用户有 n 个特征值，这些值表示用户的上下文，从而表示用户。如图 3-2 所示，每个物品都有其文本语义属性值。但知识图谱嵌入会将文本属性（如电影片名）视为通用属性，变为一个 id 值，忽略了属性值本身的内在含义，也忽略了用户的相关属性。所以，本章加入外部多模态信息，将用户属性和物品属性作为外部多模态特征来进行深度多模态推荐，从而补全了知识图谱嵌入忽略的文本语义属性。

本章充分利用视觉图像作为内部多模态信息对知识图谱推荐进行扩展，利用物品用户文本语义作为外部多模态特征弥补知识图谱嵌入带来的语义损失的不足。最后将多模态特征作为下一章基于关系空间的注意力机制传播模型的输入，可以更好地提高推荐系统的有效性和准确性。

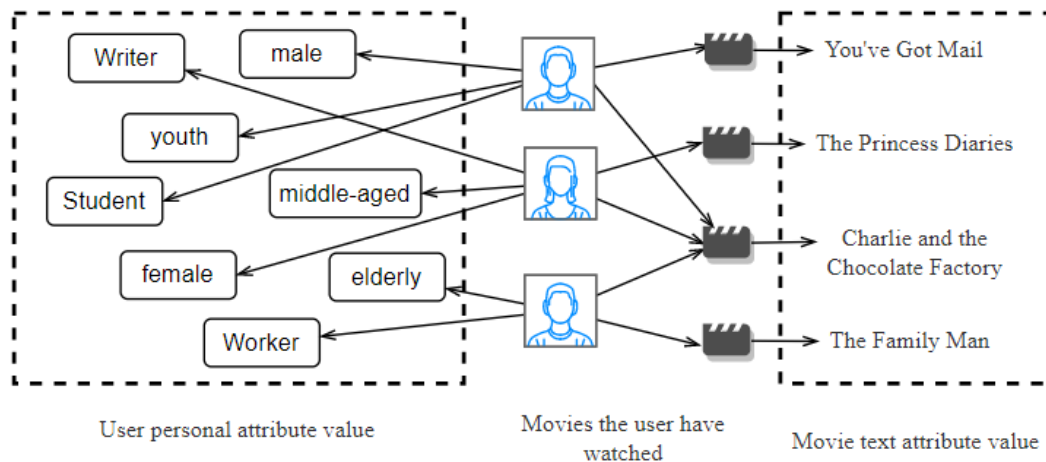


图3-2 用户物品特征图

3.2 内部多模态

3.2.1 多模态数据集构建

MovieLens1M 是一个广泛使用的推荐数据集，它可以帮助研究人员对推荐算法进行有效的评估和比较。每一条数据都反映了用户对已观看的电影的评分，评分范围从 1 分到 5 分不等，其中 5 分代表最高分，1 分则为最低分，其余分数按照从高到低的顺序递减。MovieLens1M 数据集可以提供用户对影片的评分，但在大多数情况下，建模的难点在于如何准确预测点击率，也就是二分类问题。因此，本文把用户的评价作为一种隐性的信息，同时将用户评分过的影片视为已浏览的行为。因此，根据用户的评分，可以将已获得打分信息的电影标记为 1，而未获得打分的电影则被标记为 0。

多模态 MovieLens1M 知识图谱的构建参考了 KGNN-LS^[47]方法，数据源通过利用微软的 Satori^[48]知识图谱进行扩展，可以更有效地构建知识图谱。Satori 知识图谱是一个包含丰富实体和关系信息的大型知识图谱，其中包含了从多个领域获取的数据，如常识知识、网络文本、实体属性、实体关系等。通过使用此知识图谱，可以更全面地了解电影的信息，同时也能在知识图谱中找到与电影相关的实体和关系。首先通过 Satori 知识图谱找到一组置信度超过 0.9 的集合，

称之为集合 S 。然后将 MovieLens 数据集中的电影名称与集合 S 中的尾实体进行匹配，得到对应的 Satori ID。接着，将这些电影 ID 与集合 S 中的头实体进行匹配，筛选出可以匹配的三元组，形成最终的知识图谱。

对于 Multi-MovieLens1M 知识图谱多模态图像实体部分的构建，本文采用了从 MovieLens1M 中的统一资源定位符（URL）映射到因特网电影数据库（Internet Movie DataBase, IMDB）中的图像数据。然后使用预训练的 VGG16 模型^[30]来提取图片相关特征。映射后的数据基本信息见表 3-1。这种方法能够有效地为知识图谱提供丰富的多媒体信息，增加知识图谱的表现力和可用性。通过在 IMDB 中获取电影的图像数据，可以为电影实体添加更加直观、形象的信息，同时也能够为用户提供更好的电影推荐体验。在本文中，采用了 IMDB 作为图像数据源，因为 IMDB 是一个包含大量电影信息的数据库，其中包括了电影的海报、剧照、场景等多种类型的图像数据。通过将 MovieLens1M 中的 URL 映射到 IMDB 中的对应图像，能够从 IMDB 中获取到更加全面和准确的图像数据，为知识图谱提供更加丰富的视觉信息。

表3-1 实验数据表

数据集	指标类型	指标量	数据集	指标类型	指标量
电影评分数据	用户数量	610	知识图谱数据	用户实体数量	610
	项目数量	9742		项目实体数量	9742
	评分数量	100836		演职人员关系	52904
	评分级别	1,2,3,4,5		电影类别关系	24226
	电影海报数量	9730		项目图片数量	9730

3.2.2 协作知识图谱构建

知识图谱是一种非常丰富的图结构，其汇集了客观世界中各种实体之间的复杂关联，并将其组织成结构化的信息。在知识图谱中， E 代表一个实体集合，而 R 则是一个关系集合，它们可以被抽象地表示为一个三元组集合 G ，其中 $G = \{(h, r, t) \mid h, t \in E \mid r \in R\}$ 。本文将用户对物品的交互矩阵变换为用户-物品交互图的形式，如图 3-3 所示，即如果用户对物品进行过打分，矩阵值为 1，否则为 0。然后将其变成图连接的形式，通过绿色用户节点与蓝色物品节点连接起来。

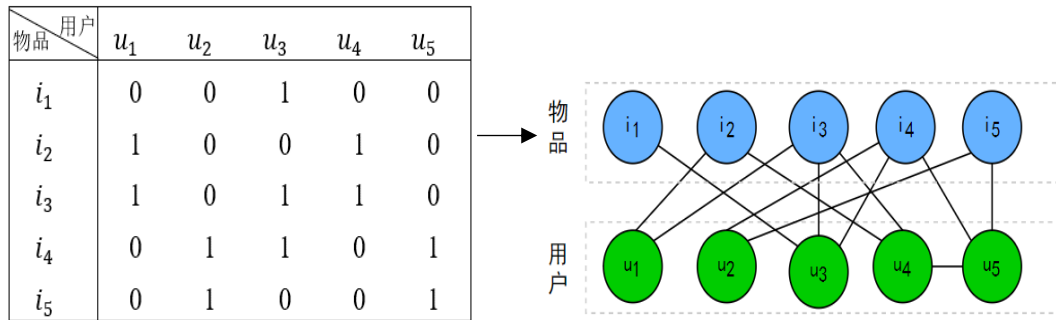


图3-3 用户-物品交互图

协作多模态知识图谱是一种新型的异质信息结构图，其是由用户物品交互图和多模态知识图谱相结合而形成的。在该知识图谱中，用户与物品被看作是一个完全独立的实体，它们之间的关系可以通过 $\{(e_u, y_{ui}, e_i) \mid e_u \in \mathcal{U}, e_i \in \mathcal{I}\}$ 来表示。

其中 e_u 代表的是用户的相关实体， e_i 则指物品的相关实体， y_{ui} 则描述了用户与物品之间的相连关系，如果用户与物品之间进行互动时， y_{ui} 的值应该设置为1，否则设置为0。 \mathcal{U} 是用户集合， \mathcal{I} 是物品相关集合。通过将多模态的知识图谱整合到用户物品交互图中，协作多模态知识图谱可以保留所有 y_{ui} 等于1的边，并将用户和物品的交互行为表示为三元组形式 $(e_u, Interact, e_i)$ ，其中 $Interact$ 代表用户物品之间的联系。此外，物品也具备一些特定的知识，比如关联实体，这些实体可以用来组成完整的知识图谱。由于用户物品交互图和知识图谱均包括了物品节点，从图3-4可以看出，紫色节点 i 不仅存在于用户物品交互图中，而且还存在于知识图谱中，因此，可以利用这些节点通过实体对齐操作将物品和实体连接起来组成一个完整的图，即协作多模态知识图谱。

通过构建协作多模态知识图谱，可以将用户物品交互图中的物品和用户节点视为实体，并将其有机地结合在一起，从而形成一个完整的知识图谱网络。如图3-4所示，多种颜色节点代表着不同类型的实体，蓝色代表用户实体，紫色代表物品实体，绿色代表知识图谱中相关实体，黄色代表图像实体，而线段则描绘出了不同的关系特征。这样构建的协作多模态知识图谱能够为模型提供丰富的异质信息，以支持更高效的知识推理和用户行为预测。

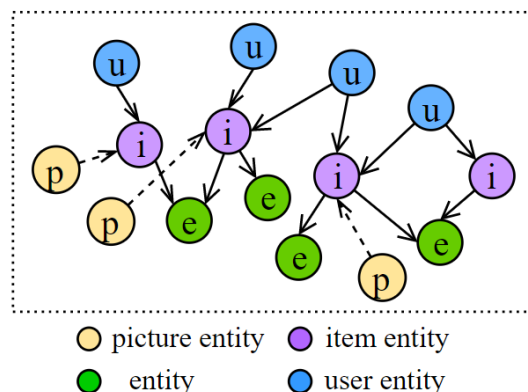


图3-4 内部多模态协作知识图谱

3.2.3 预处理方法

（1）改进的知识图谱嵌入 TransR 方法

多模态协作知识图谱构建完成后，需要对其进行向量化表示。通过将零散的知识要素嵌入到知识图谱中，可以得到构建知识表示的结果，即知识向量。然后对于向量进行处理提取特征。如前文所述，可以将知识图谱形式化为一个三元组集合 G ，它包含了一个头实体 h 、尾实体 t 以及一个两者之间的关系向量 r 。采用 TransE 翻译技术，能够将头部实体 h 映射到 r 关系中，从而获取尾部实体 t 。如图 3-5 所示。当三者之间的关系被确认为真时，即 $h + r - t \approx 0$ 。如果其绝对值很大，则说明三者不构成关联。

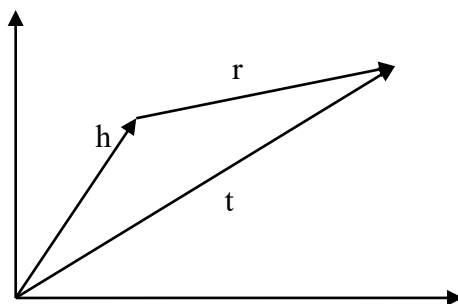


图 3-5 TransE 模型结构

TransE 是一种非常有效的知识嵌入表示方法，其可以有效地处理单个实体之间的关系，但是，如果实体之间存在多个关系，由于它们被嵌入到同一个平面

中,就无法准确地反映出多个实体之间的关系类型,从而导致相似度计算结果的不准确。如图 3-6 所示,由于张艺谋导演过的电影有多个,而 TransE 只能处理单个关系,这种情况下经过 TransE 处理就只能把《英雄》、《十面埋伏》、《金陵十三钗》等电影训练成同一部电影向量,导致知识损失。而通过使用 TransR 这种改进的翻译方法,就能够更好地展现实体之间的多元关系。



图3-6 多种关系情况下TransE面临的局限性

在实际场景中,两个实体之间可能有多种关系,如果一个电影与其他电影具有多种关系,那么其之间的相似性就应该更高,因此在推荐时更有可能出现在用户列表中。TransR 模型的原理如图 3-7 所示。

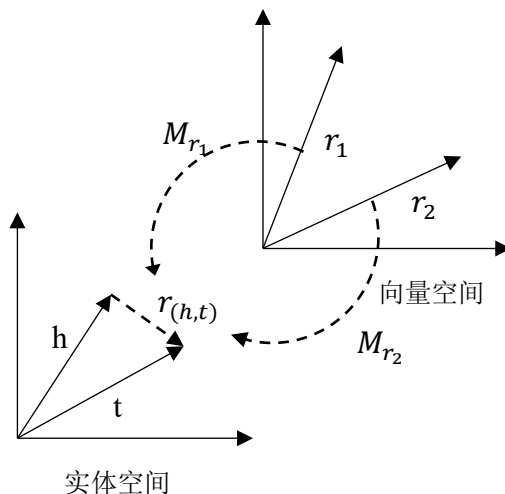


图3-7 TransR 模型结构

TransR 采用与 TransE 不同的方式来表示知识图谱中的关系。具体而言, TransR 将头实体和尾实体嵌入到一个平面上,并使用一个独立的关系空间来学

习表示关系向量。当两个实体之间存在 n 个关系时，第 i 个关系可以用向量 r_i 来表示，而其他关系则存储在一个独立的矩阵空间 \mathbf{M}_r 中。

当一个实体与另一个实体之间建立 n 个相互依赖的关联关系时，第 i 个关联关系可以通过一个向量 r_i 来描述，之后将全部关联关系保留到一个独立的矩阵空间 \mathbf{M}_{r_i} 中。因此，TransR 的关系向量可以表示为如公式 (3.1) 所示：

$$r'_{(k,t)} = r_i M_{r_i}, i = 1, 2, \dots, n \quad (3.1)$$

TransR 可以将多元的实体间关系转换成一个矩阵 \mathbf{M}_{r_i} 。采用这一映射转换方式，可以将头部实体与尾部实体的关联关系转换为一个向量 $r_{(h,t)}$ 。因此，TransR 方法中的头实体与尾实体之间的关系映射可以用如下公式 (3.2) 来描述：

$$h + r_{(h,t)} \approx t \quad (3.2)$$

TransR 方法不仅可以表示每个实体，还能保留实体间的多元关系，这比 TransE 方法更加准确。然而，因为 TransR 将实体与实体之间的多元关系都集中存储在了独立的关系矩阵 \mathbf{M}_{r_i} 中，导致了计算复杂度大大增加。因此，为了减少关系矩阵的复杂性，本文选择使用对角矩阵 \mathbf{A}_{r_i} 来替换 \mathbf{M}_{r_i} 。这时，向量映射的形式如公式 (3.3) 所示：

$$r_{h,t} = r_i \mathbf{A}_{r_i}, i = 1, 2, \dots, v \quad (3.3)$$

(2) 图像编码器

上节提到的改进的 TransR 方法适用于处理常规的知识图谱三元组实体元素进行嵌入向量表示，但对于复杂的图像实体特征提取能力有限，需要专门的图像编码器对其进行编码向量化表示，因此，本文采用了 VGG16 卷积神经网络^[49]对图像进行特征提取，以将图像实体转化为向量表示。如图 3-8，VGG16 是一个经典的卷积神经网络，其预训练的权重能够很好地提取图像中的高级特征。

然而，高维度的数据会导致计算复杂度很高，因此本文使用 PCA 算法^[50]降低数据的维度，同时保留了图片数据的主要特征。PCA 算法被广泛应用于降维领域，它能够有效地将高维信息转换成低维信息，大大减少了计算的难度和复杂性。PCA 能够将数据压缩到一个较低的维度，同时保留数据的主要特征。当物品不存在图片时，使用 0 向量作为物品的图片特征。这样可以确保当前物品不参与图片特征的融合，同时避免了在融合时出现异常情况。

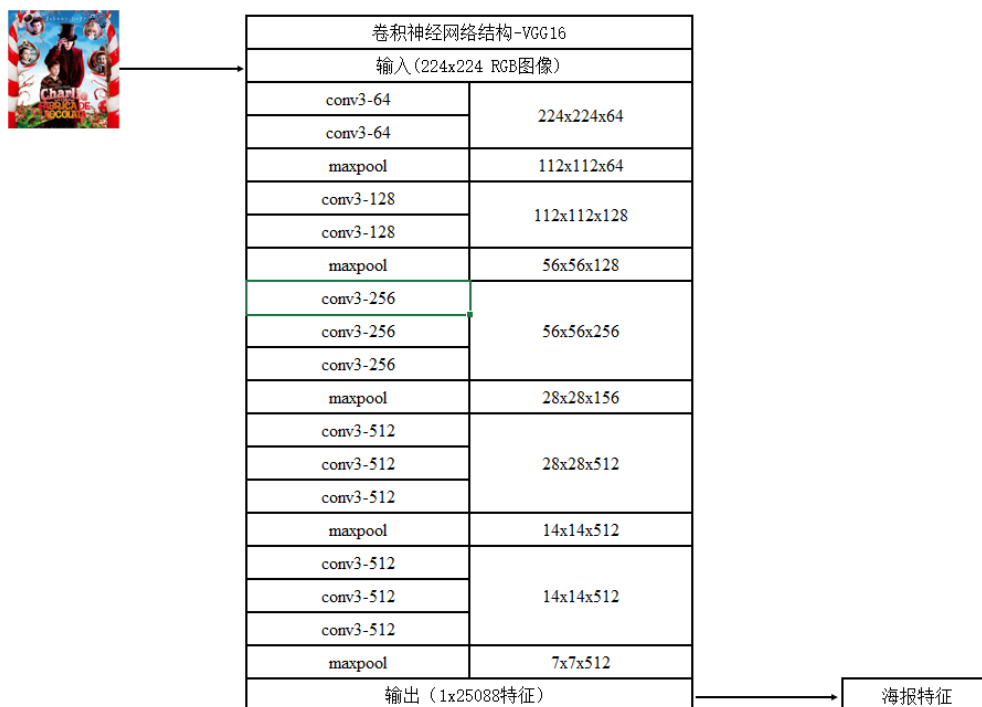


图3-8 VGG16框架图

3.3 外部多模态

本文研究将外部多模态信息引入知识图谱中，主要包括用户属性信息和物品名称信息。在传统知识图谱模型中，比如 MKR^[51]，KGCN^[52]都没有对用户属性进行相应的处理，像 MKR 只是简单的利用了物品属性特征，忽略物品属性之间的内在关系。而本文提出的 MKGA 模型不仅充分挖掘了用户属性和物品名称信息，同时也分别构建了统一的向量化表征，实现用户属性之间和属性与物品名称的充分信息融合。

用户属性信息（如用户 ID、性别、职业等）表示为 $y: \{y_1, y_2 \dots y_n\}$ ，其中 y_i 表示用户属性之一。通过对逐个属性采用全连接网络，可以获得用户属性的特征向量组 \bar{y} ，具体如公式(3.4)所示：

$$\bar{y} = f(w_1 y + b_1) \quad (3.4)$$

其中 w_1 表示权重， b_1 表示偏差， $f(\cdot)$ 表示激活函数（本文采用 Relu 函数）。处理过程如图 3-9 所示。

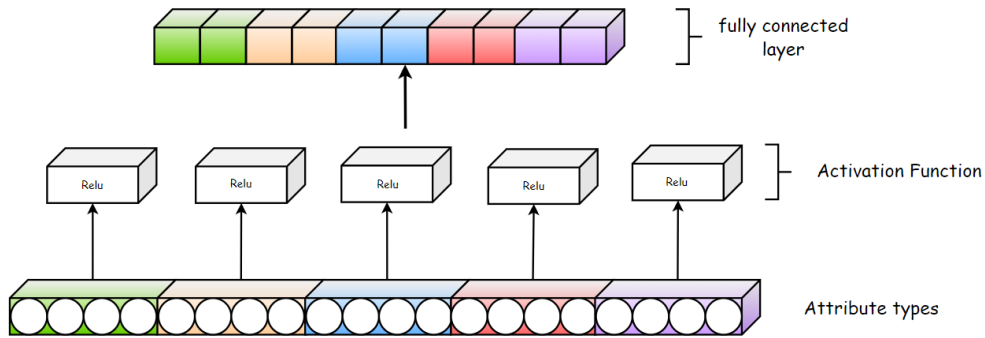


图3-9 用户特征MLP处理模块结构

为获得用户属性的向量化表征，通过首位相连操作（*concatenate*），计算出用户属性向量，如公式（3.5）所示：

$$u = concatenate(\bar{y}) \quad (3.5)$$

其中 $u \in R^{d_1}$ ， d_1 为用户属性特征维度。为了充分融合不同属性信息，本文进一步将 u 通过一层全连接层（Fully-connected layer, FC），其公式表达如（3.6）所示：

$$u' = FC(u) \quad (3.6)$$

物品名称描述通常为文本语义句子，为了充分建模文本语义信息，本文采用TextCNN网络实现细粒度文本语句信息提取。TextCNN模型结构如图3-10所示。

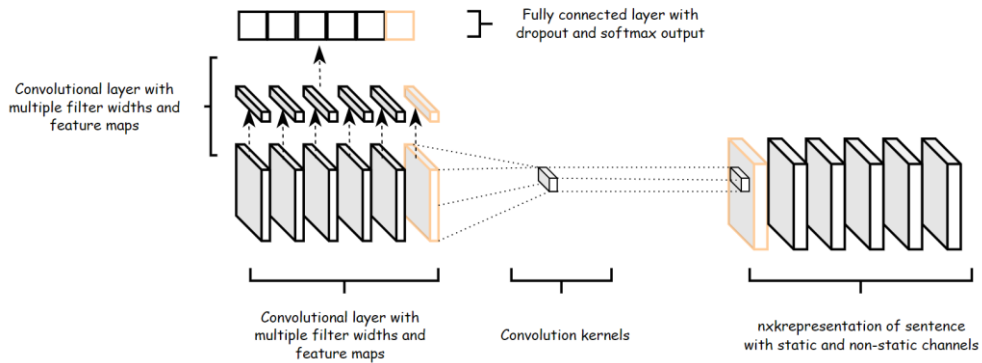


图3-10 物品语义特征TextCNN处理模块结构

具体来说，假设物品文本包含最多 n 个单词，每个单词表征为一 K 维特征向量，因此，对于任意物品文本名称，其原始特征可表示为一 $n \times K$ 的实数矩阵 X 。

TextCNN 采用两类不同类型的卷积核在特征平面上进行处理,包括滑动和固定。在此结构中,卷积核的尺寸被分别定义为 $2 \times k$ 、 $3 \times k$ 、 $4 \times k$ 、 $5 \times k$,其中 k 代表着嵌入的长度。采用最大池化算法,可以有效地提取出各个特征图的最高值,并将其组合在一起形成一个完整的特征向量。数学上,经过 TextCNN 处理后,获得的物品文本特征向量如公式 (3.7) 所示:

$$v' = \text{TextCNN}(X) \quad (3.7)$$

其中 $v \in \mathbb{R}^{d_2}$, d_2 为特征维度。

3.4 方法实现

该方法通过算法 3-1 进行实现。首先通过实体对齐算法得到物品的相关实体集合;然后利用用户项目交互矩阵和知识图谱融合形成多模态协作知识图谱。嵌入过程中,针对不同类型实体,分别使用改进的 TransR 方法和 VGG16 算法进行相应预处理;最后针对外部多模态分别使用多层感知机网络和 TextCNN 网络进行特征提取得到用户属性特征和物品的文本特征,以下是伪代码描述:

算法 3-1: 多模态数据的构建与预处理算法

Input: interaction matrix \mathbf{Y} , knowledge graph \mathbf{G} , user attribute set \mathbf{U} , item visual **image_Images**, item title text **item_text**

Output: item text features, user attribute **feature{u,i}**

```

1  n=0;
2  Initialize model parameters;
3  for each item and e do
4      if item != e then
5          n = n+1;                                // entity alignment operation
6      else
7           $A = \{(v, e) | v \in V, e \in \mathcal{E}\}$ ;
8  end for
9   $\text{MKG} = \{(e_u, y_{ui}, e_i) | e_u \in \mathcal{U}, e_i \in I\}$ ; // Building a multimodal collaborative
                                                knowledge graph

```

```

10  $r_{h,t} = r_i A_{r_i}, i = 1, 2, \dots, v;$  //Embedding using improved TranR method
11  $model \leftarrow VGG16(item\_images);$ 
12 for each U and item_text do
13      $u' \leftarrow FC(u);$  // extract user attribute features
14      $v' \leftarrow TextCNN(X);$  // Extracting item text attributes using textCNN
15 end for
16 return feature{u,i}

```

3.5 本章小结

本章主要提出了多模态数据的构建与预处理方法。一方面，引入图像多模态信息构建内部多模态知识图谱，在此基础上将用户物品点击交互记录转换为用户物品交互图的形式，然后与多模态知识图谱结合，组合成多模态协作知识图谱。之后将多模态协作知识图谱进行向量化表示，分别针对不同数据类型提出改进的 TransR 和图像编码方法，从而实现内部多模态知识图谱向量化表示；另一方面，加入外部多模态用户物品特征，对应使用 MLP 多层感知机对其进行特征提取，得到更为丰富的语义信息，从而对物品数据进行了有效扩展，增强了推荐效果。

第 4 章 基于关系空间的注意力机制传播模型

本章以第 3 章中经过预处理的多模态数据信息为基础, 提出基于关系空间的注意力机制传播模型以实现用户物品推荐。首先, 提出模型框架, 以内部多模态数据特征作为输入对象; 然后, 通过基于关系空间的注意力机制逐层传播, 聚合形成用户物品特征向量, 对于注意力传播过程以单层进行阐述说明; 之后, 将外部多模态数据与其融合进行推荐预测; 最后, 对本章的主要工作进行总结。

4.1 模型框架

物品和用户属性值作为辅助信息添加到推荐系统中, 需要以知识图谱三元组的形式构建 (物品、属性类型、属性值)。但是, 传统知识图谱推荐模型将文本属性 (如电影片名) 视为通用属性, 忽略了属性值本身的内在含义, 也忽略了用户相关属性。而且知识图谱大多只关注于内部结构语义信息, 没有考虑将其他类型的数据信息作为辅助特征。同时传统推荐模型无法有效挖掘用户物品实体内部隐藏关系。为了解决以上问题, 本章提出融合多模态知识图谱与注意力机制的推荐方法 (Multimodal Knowledge Graph Attention, MKGA) 将多模态知识图谱和注意力机制进行融合, 使用加入视觉信息的知识图、用户属性和物品属性作为多模态信息来进行深度多模态推荐, 同时融合注意力机制自动学习相关权重从而可以更好地聚合高阶邻域信息。方法框架如图 4-1 所示。该框架主要由三个部分组成: 数据预处理层、注意力层以及预测层。

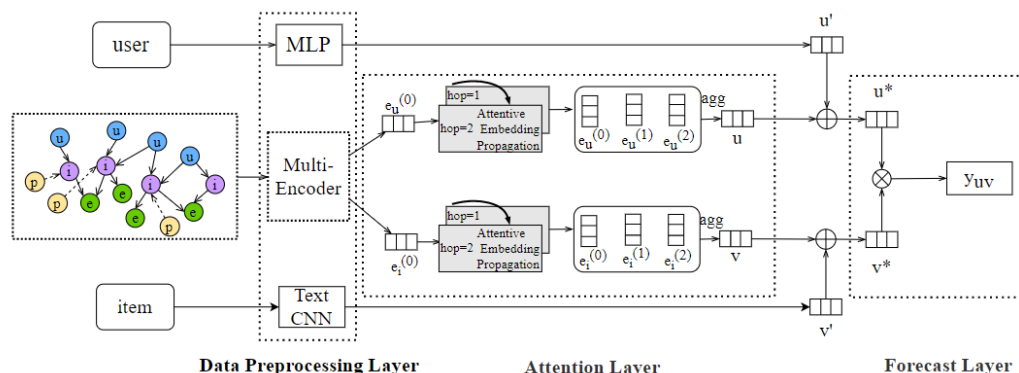


图4-1 MKGA模型结构

(1) 数据预处理层：如第 3 章所述，针对外部多模态信息，应在用户和物品之间进行分类讨论。对于 MKGA 模型，所有的用户属性均需要在多层感知机（Multilayer Perceptrons, MLP）处理，比如所有的用户 ID、性别、职业等属性都需要经过训练、研究和整合来输出用户特征向量。所有的物品属性也需要在 TextCNN 模型中提取特征，比如物品的文本语义属性通过在 TextCNN 模型中训练得到物品特征向量。对于内部多模态知识图谱，通过构建多模态数据集，图像编码器，从而得到内部多模态特征向量。

(2) 注意力层：通过分析多模态知识图谱模块中的多种特征，结合基于关系空间的知识感知注意力机制，对传播过程中的每一个邻居进行权重分析，并从中提取出有价值的信息，形成知识图谱注意力层中的用户和物品的特征向量。在特征聚合阶段，本章使用相加，连接和双端混合三种方式进行聚合效果测试，之后将外部多模态数据与其融合形成最终的用户物品特征向量。

(3) 预测层：通过注意力层与数据预处理层外部多模态数据聚合得到最终的用户和物品特征向量，最后使用预测函数进行预测。

4.2 模型描述

当前，许多基于实体的多模态知识图谱表示学习方法只关注单一的三元组，而忽视了多模态场景中信息聚合的重要性。尽管之前也有研究工作，如 MFB(Multimodal Factorized Bilinear pooling)^[53]和 CMIN(Common and Modal-specific Interaction Network)^[54]，但是它们仍然局限于三元组的个体评分，对其进行训练提升而忽视了多模态信息的交互和整合。在知识图谱中，实体和实体之间以图形式存在，因此从图结构入手进行处理是一种符合逻辑的思路。本文使用图卷积网络（Graph Convolutional Network, GCN）对图结构加以处理。

GCN^[55]是专门用于处理图结构的一种新型网络结构。本文将主要从消息的传输方式和信息的整合方法这两个方面对 GCN 进行阐述解释。在 GCN 中，每个节点都有一个特征向量。通过传递节点之间的信息和聚合相邻节点的信息，GCN 可以学习到每个节点的表示。具体来说，GCN 使用一个邻接矩阵来描述节点之间的关系，该矩阵可以将节点之间的关系表示为一种图形式。GCN 使用邻接矩阵来计算节点之间的信息传递，将相邻节点的特征向量聚合成一个新的特

征向量。这个过程可以重复进行，直到 GCN 学习到每个节点的表示。

通过对 GCN 技术的深入研究，可以将其应用于本文提出的模型中，以达到更好的效果。因此，本文提出了基于关系空间的注意力机制传播模型，通过基于关系空间的迭代传播路径将高阶邻居信息聚合到目标节点上，从而实现了高效的信息传递。通过观察图 4-2，可以得知，在内部多模态注意力传播层中，多模态图像信息已经被纳入到了注意力机制传播的路径中。

具体地，本文模型将图像知识作为一种实体，通过多模态知识图谱注意力层的传播聚合，得到相关用户物品的实体向量。最后将其与外部多模态信息如用户物品特征融合进行推荐。注意力传播层由信息传播、知识感知注意力和信息聚合三个部分组成。本章首先对其单层进行介绍，之后将其泛化到多个层传播，以达到更好的推荐作用。

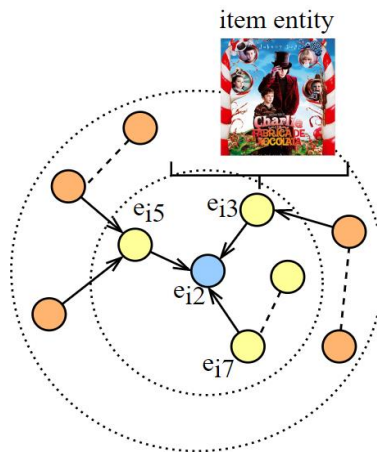


图4-2 内部多模态注意力传播层

4.2.1 基于关系空间的注意力传播

对于每个实体都能够参与到多个三元组当中，成为两个三元组之间的纽带，从而将它们之间的信息传递出去。以 $e_1 \xrightarrow{r_2} v_2 \xrightarrow{-r_1} u_2$ 和 $e_2 \xrightarrow{r_3} v_2 \xrightarrow{-r_1} u_2$ 为例，物品 v_2 将其相关实体属性 e_1 和 e_2 作为前驱输入属性来强化其本体特征，然后贡献用户 u_2 的喜好，可以通过将信息从 e_1 传播到 u_2 来进行。利用这一机制来实现实体与其周围邻接实体之间的信息共享和传递。

考虑实体 h ，可以使用 $S_h = \{(h, r, t) | (h, r, t) \in \mathcal{G}\}$ 来表示，其中 S_h 是头部实体

的三元组集合。用 e_{s_h} 来衡量 h 的 S_h 嵌入的实体表达，如公式（4.3）所示：

$$e_{s_h} = \sum_{(h,r,t) \in S_h} \theta(h,r,t) e_t \quad (4.3)$$

其中 $\theta(h,r,t)$ 是控制边缘 (h,r,t) 上每次传播的衰减因子，表示从尾实体 t 传播到头实体 h 的信息量取决于关系 r 。

本文通过关系注意力机制实现 $\theta(h,r,t)$ ，其公式如（4.4）所示：

$$\theta(h,r,t) = (W_r e_t)^T \text{LeakyReLU}(W_r e_h + e_r) \quad (4.4)$$

其中， $W_r \in R^{k \times d}$ 是关系 r 的变换矩阵，它将实体从 d 维实体空间投射到 k 维关系空间。激活函数选择 LeakyReLU，LeakyReLU 作为一种修正线性单元（ReLU）的变体，它在处理负数输入时不会将输出直接归零，而是采用一个小的斜率来避免信息的丢失，从而使得模型更加稳健，鲁棒性更高。具体的计算方法如公式（4.5）所示：

$$\text{LeakyReLU}(x) = \begin{cases} x & x \geq 0 \\ \text{negativeslope} \times x & \text{otherwise} \end{cases} \quad (4.5)$$

接下来，使用 softmax 将和 h 相关联的全部三元组的系数进行归一化操作，如公式（4.6）所示：

$$\theta(h,r,t) = \frac{\exp(\theta(h,r,t))}{\sum_{(h,r',t') \in N_h} \exp(\theta(h,r',t'))} \quad (4.6)$$

通过评估注意力分数，可以确定哪些邻居节点需要被特别关注，以便及时发现并传递协作信息。本文提出的 MKGA 模型与 GCN^[55] 和 GraphSage^[56] 中的信息传播有所不同，它们将两个节点之间的衰减因子分别设置为 $1/\sqrt{|N_h||N_t|}$ 或者 $1/|N_t|$ ，而本文 MKGA 模型则采用了更加多元的方法，不仅考虑了图的邻近结构，还考虑了邻居的不同重要性。另外，与以节点表示的方式进行构建的 GAT^[57] 不同，本文还利用 e_h 和 e_t 的相互影响，对其关系空间进行建模，以便能够有效地捕捉到传播的更多细节信息。

4.2.2 信息聚合

根据图 4-3 所示，针对特定目标实体，传播层需要将其邻居信息传递到其目标本体上面，但是从避免目标实体的信息被邻居信息所覆盖的角度出发，本文加入聚合层，以实现邻居实体和目标实体之间的有效聚合。

考虑聚合头节点本身的嵌入向量（也就是 e_h ）的信息和头节点对应的 e_{s_h} 的

嵌入，来重新得到 h 节点的嵌入表示。此时定义 h 头节点的新的嵌入表示如公式 (4.7) 所示：

$$e_h^{(1)} = f(e_h, e_{s_h}) \quad (4.7)$$

本文将研究三种不同的特征交互聚合方法，这些方法可以用于提取不同聚合方法的特征信息，并在模型中进行融合，以实现更好的性能。下面将从设计原理的角度分别对这些聚合方式进行描述：

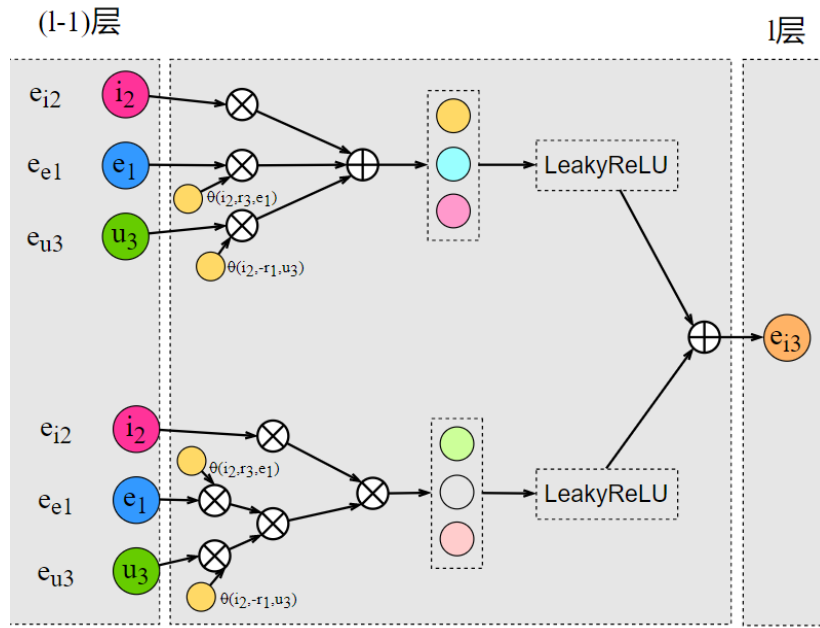


图4-3 基于关系的注意力传播聚合方法

(1) 相加聚合方法 (Add Aggregation Method)

相加聚合方法考虑 e_h 和 e_{s_h} 之间的逐元素相加的特征交互，可以通过公式 (4.8) 获得。

$$f_{add} = \tanh(W(e_h + e_{s_h})) \quad (4.8)$$

其中，本文使用 \tanh 作为激活函数， $W \in \mathbb{R}^{d' \times d}$ 为迭代优化的权重矩阵， d' 是变换大小，这些权重矩阵能够帮助从传播的信息中提取出有价值的内容。

(2) 拼接聚合方法 (Concatenation aggregation method)

向量拼接是一种常用的特征融合方式，在深度学习模型中广泛应用。它将不同的特征向量拼接成一个更长的向量，从而提供更多的信息给后续的层。这种方式可以避免信息的损失，保留多种特征之间的关联性，提高模型的性能和鲁棒性。

在本文中，使用线性变换将 e_h 和 e_{s_h} 拼接起来，如公式（4.9）所示：

$$f_{Concatenation} = \tanh(W(e_h \parallel e_{s_h})) \quad (4.9)$$

其中， \parallel 是拼接操作。

（3）双向混合聚合方法（Bi-Mix Aggregation Method）

本文针对模型中的特征交互问题，如图 4-3 所示，考虑了 e_h 和 e_{s_h} 之间的两种不同交互方式。具体来说，将特征的逐元素相加和逐元素点乘进行结合，构建了一种新的特征表示方式，具体如公式（4.10）所示：

$$f_{Bi-Mix} = \tanh(W_1(e_h + e_{s_h})) + \tanh(W_2(e_h \odot e_{s_h})) \quad (4.10)$$

其中， \odot 表示元素——点乘，也叫做哈达玛积。 $W_1, W_2 \in R^{d' \times d}$ 是权重矩阵。

Bi-Mix Aggregation Method 方法对 e_h 和 e_{s_h} 之间的交互特征进行了编码。该方法可以使正在传播的信息对 e_h 和 e_{s_h} 之间的密切相关性更加灵敏，以此传递来自相似实体的更多消息。

4.2.3 高阶传播

4.2 章讨论了 GCN 的基本概念，但仅仅关注了它的单一层面，而忽略了其他层面的信息传输。GCN 技术允许采用多层级的神经网络结构，从而实现跨越多个跳点的信息传输。根据图 4-4，当 $K=1$ 时，一个节点的信息可以通过一个直接相邻的邻居节点来进行传输。通过一阶传播，一个蓝色的节点被用来接收它周围四个黄色节点的数据。同样地，以黄色节点为起点，一阶传播可以将橙色节点的信息有效地聚合到黄色节点上，从而实现更加精准的信息传播。为了更好地理解，本示例中，传播过程会将信息以不同的颜色进行染色，比如蓝色的目标节点会被染成黄色，而黄色节点则会被染成橙色。

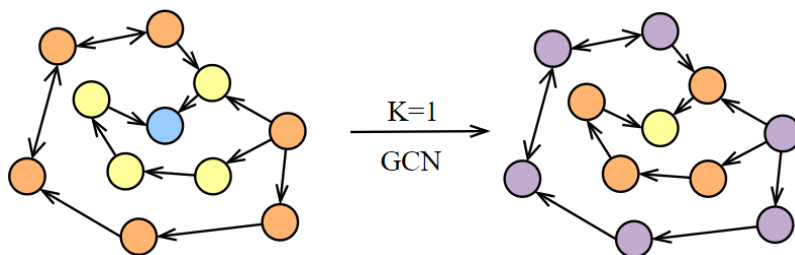


图4-4 $K=1$ 时，GCN高阶传播图

如图 4-5 所示，当 $K=2$ 时，由于两层 GCN 的叠加，此时的节点已经包含

了它们的邻居节点的信息，因此，在第二次传输的时候，就能够将更多的信息传输至目标节点，比如，在两次传输之后，最中央的节点就会由原本的蓝色变成橙色，并且已经得到两次传播的信息。从图 4-4 和图 4-5 中可以清楚地观察到信息传播的过程。在这里，可以将与目标节点直接相连的节点称为一阶邻居，而借助另一个节点才能相连的节点称为二阶邻居。 n 阶邻居可以同理以此类推。

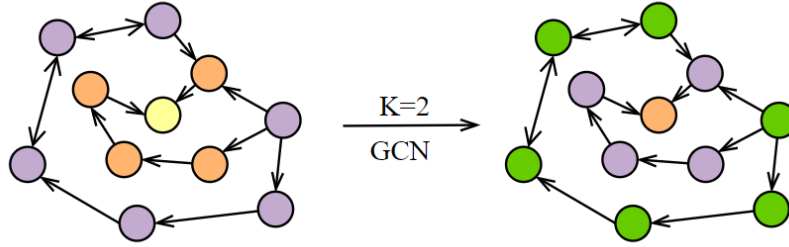


图4-5 K=2时，GCN高阶传播图

为了提高效率，可以在 GCN 层中添加更多跳的邻居，这样就能够把更多的信息集成在目标节点上。但是，这样做也有缺陷，因为跳跃次数越多相关性就越低，而 GCN 层的数量也会随之增加，这样就容易造成过拟合和过渡平滑的问题。通常情况下，将 GCN 层数增加至 2-3 层就足够了。为了可以聚合更多跳的邻居信息传播，获取到更多的高阶层次信息，第 1 个步骤中，可以通过递归的方式来描述实体如公式(4.11)所示。

$$e_h^{(l)} = f(e_h^{(l-1)}, e_{S_h}^{(l-1)}) \quad (4.11)$$

其中实体 h 在 $e_{S_h}^{(l-1)}$ 邻域网络实体中传播的信息定义如公式（4.12）所示：

$$e_{S_h}^{(l-1)} = \sum_{(h,r,t) \in S_h} \theta(h,r,t) e_t^{(l-1)} \quad (4.12)$$

在信息传播过程中，实体 h 的表示是通过 $e_h^{(l-1)}$ 得到的，该表示存储了它(1-1)跳邻居的相关信息。在初始迭代中， e_h^0 被设置为 e_h ，有助于在第 1 层更好地表示实体 h 。

通过不断的传递和迭代，此时将节点 h 的多跳的信息都传入了 h 的嵌入编码中。且通过注意力机制保证了不同信息的相关权重，从而实现了高阶信息的传递。

4.2.4 模型预测

经过 L 层的注意力传播处理后,可以得到用户节点 u 和物品节点 v 的多个表示,分别为 $\{e_u^{(1)}, \dots, e_u^{(L)}\}$ 和 $\{e_v^{(1)}, \dots, e_v^{(L)}\}$ 。每一层的输出都是基于以 u (或 v) 为根的树形结构的信息聚合,因此各个层的输出重点在于不同级别的连接性信息,同时,与外部多模态的用户物品特征向量进行融合。为了将每一步的表示连接成一个向量,本文采用了层聚合机制^[58],具体如公式 (4.13) 和 (4.14) 所示:

$$e_u^* = e_u^{(0)} \parallel \dots \parallel e_u^{(L)} \parallel u' \quad (4.13)$$

$$e_v^* = e_v^{(0)} \parallel \dots \parallel e_v^{(L)} \parallel v' \quad (4.14)$$

在这里,使用符号 \parallel 表示连接运算。其作用不仅在于扩展了执行嵌入传播操作所需的初始嵌入表示,并且可以根据需要调节 L 值,以达到更好的传输效果。

最后,通过计算用户与物品之间的内积,可以估算出它们之间的相似度,如公式 (4.15) 所示:

$$\hat{y}(u, v) = e_u^{*T} e_v^* \quad (4.15)$$

4.2.5 模型实现

本章从图卷积神经网络出发,引入注意力机制,对于实体高阶的信息进行挖掘和归纳,在知识图谱传播过程中通过衰减因子 θ 不断迭代得到邻域集合;之后再通过聚合方法对实体邻域及其本体进行特征融合,该模型通过算法 4-1 加以实现,以下是伪代码描述:

算法 4-1: 基于关系空间的注意力传播算法

Input: interaction matrix Y , knowledge graph G ,

neighborhood sampling map $\mathcal{S}: e \rightarrow 2^{\mathcal{E}}$;

Output: Prediction function $\mathcal{F}(u, v \mid \Theta, Y, \mathcal{G})$

```

1  while MKGA not converge do;
2  for  $(u, v)$  in  $Y$  do;
3     $\{\mathcal{M}[i]\}_{i=0}^H \leftarrow \text{Get-Receptive-Field}(v)$ ;
4     $e^u[0] \leftarrow e, \forall e \in \mathcal{M}[0]$ ;
5    for  $h = 1, \dots, H$  do
```

```

6      for  $e \in \mathcal{M}[h]$  do
7           $\mathbf{e}_{\mathcal{S}(e)}^u[h-1] \leftarrow \sum_{e' \in \mathcal{S}(e)} \tilde{\theta}_{r_{e,e'}}^u \mathbf{e}'^u[h-1];$ 
8           $\mathbf{e}^u[h] \leftarrow \text{agg}(\mathbf{e}_{\mathcal{S}(e)}^u[h-1], \mathbf{e}^u[h-1]);$ 
9       $v \leftarrow \mathbf{e}^u[H] || v';$ 
10     Calculate predicted probability  $\hat{y}_{uv} = f(u, v);$ 
11     Update parameters by gradient descent;
12 return  $\mathcal{F}$ 

```

4.3 模型优化与训练

使用第 3 章提出的改进的TransR方法, 通过优化翻译原理 $\mathbf{e}_h^r + \mathbf{e}_r \approx \mathbf{e}_t^r$, 可以将给定的三元组 (h, r, t) 中的实体及其关系有效地联系起来, 从而实现对实体之间的关系进行更好地表达。其中, $\mathbf{e}_h, \mathbf{e}_t \in R^d$ 和 $\mathbf{e}_r \in R^k$ 是 h, t, r 的嵌入向量; $\mathbf{e}_h^r, \mathbf{e}_t^r$ 可以通过关系 r 的空间投影的形式来表达出来。根据给定的三元组 (h, r, t) , 本文可以推导出一个具有可靠性的评估标准, 其合理性得分可以如公式 (4.16) 所示:

$$g(h, t, r) = \|\mathbf{W}_r \mathbf{e}_h + \mathbf{e}_r - \mathbf{W}_r \mathbf{e}_t\|_2^2 \quad (4.16)$$

关系矩阵 $\mathbf{W}_r \in R^{m \times n}$ 是关系 r 的转换矩阵, 其中实体空间维度为 m , 关系空间维度为 n , \mathbf{W}_r 将实体映射到了关系空间中。对于三元组 (h, r, t) , 其值越小意味着该三元组越可能是真实存在的, 同样其值越高则说明关系越远。

改进的TransR训练方法通过比较有效组和无效组之间的相对排序, 并利用公式 (4.17) 来估算损失函数, 以提高模型的准确性和可靠性。

$$\mathcal{L}_{KG} = \sum_{(h,r,t,t')} -\ln \sigma(g(h,r,t') - g(h,r,t)) \quad (4.17)$$

对于 $\mathcal{T} = \{(h, r, t, t') | (h, r, t) \in \mathcal{G}, (h, r, t') \notin \mathcal{G}\}$, (h, r, t') 则利用随机抽取的 t' 来取代原来真实的三元组中的 t , 从而构造出无效三元组进行迭代优化。 $\sigma(\cdot)$ 是 sigmoid 函数。通过三元组粒度的建模, 这一层可以实现正则化和注入表示, 从而大大提升模型的表达能力, 使得实体和关系更加清晰、准确。

MKGA 损失函数为公式 (4.18) 所示:

$$\mathcal{L}_{CF} = \sum_{(u,i,j) \in O} -\ln \sigma(\hat{y}(u,i) - \hat{y}(u,j)) \quad (4.18)$$

其中 $O = \{(u,i,j) | (u,i) \in R^+, (u,j) \in R^-\}$ 表示训练集, R^+ 表示用户 u 和物品 j 之间观察到的 (正) 交互, 而 R^- 是采样的未观察 (负) 交互集; $\sigma(\cdot)$ 是 sigmoid 函数。

最后, 使用目标函数来联合学习如公式 (4.19) 所示:

$$\mathcal{L}_{MKGA} = \mathcal{L}_{KG} + \mathcal{L}_{CF} + \lambda \|\theta\|_2^2 \quad (4.19)$$

其中, $\theta = \{E, W_r, \forall l \in \mathcal{R}, W_1^{(l)}, W_2^{(l)}, \forall l \in \{1, \dots, L\}\}$, 记作模型参数集, E 是嵌入所有实体和关系的表。为了避免过拟合, 使用 L2 正则化。

4.4 本章小结

为了解决现有推荐系统存在的数据稀疏性问题, 本文提出了一种融合多模态知识图谱和注意力机制的个性化信息推荐研究方法。本章从图卷积神经网络出发, 将第 3 章的多模态数据的构建与预处理方法信息作为本章 MKGA 模型的输入, 引入注意力机制, 对于实体高阶的信息进行挖掘和归纳, 其中在融合阶段, 将外部多模态信息中所包含的用户特征和物品语义特征通过实体对齐的方式融入到最终的实体表示特征, 最后进行优化训练, 使用预测函数进行预测。目的是通过使用多元化多角度的辅助信息, 对高阶信息进行充分挖掘以保证推荐的准确性和精确度, 缓解数据稀疏性问题。

第 5 章 实验分析

本文在第 4 章中深入探讨了所使用的模型、损失函数和训练方法,以期获得更好的推荐效果。本章将通过具体实验以验证本文所提方法的可行性和有效性。首先,将详细阐述实验所需的软硬件环境,然后介绍实验的各个组成部分,包括评估指标和对比方法。接下来,将本文提出的方法与当前主流基线方法进行比较。同时,通过稀疏实验验证了在稀疏场景中本文所提方法的可行性。最后,通过消融实验详细说明了本文所提方法的各模块效用,并研究它们之间的相互作用,以进一步验证所提方法的有效性。

5.1 实验环境

本文使用了 PyCharm 2021 作为运行开发环境,并在 Windows 10 操作系统平台下进行实验,搭配了 AMD 锐龙 4 系标压处理器(2.9GHz)、16G 运行内存,以及 GTX 2060 6G 显卡。

5.2 对比方法介绍

为了验证方法的有效性,本文选用五种主流基于知识图谱的推荐方法作为比较对象:

(1) MKR^[51]: MKR 是 Wang 等人提出了一种新的多任务学习模型,该模型结合了知识图谱和深度学习,具体来说, MKR 使用了知识图谱中的实体和关系,将其融入到深度学习模型中,以提高模型的表示能力和泛化能力。此外, MKR 框架还考虑了多任务学习的问题,即如何在不同的任务之间共享知识,来提高模型的效率和精度。

(2) PER^[59]: 提出了一种基于异构信息网络的个性化实体推荐方法(PER),将知识图谱视为异构信息网络,并提取基于元路径的特征来表示用户和物品之间的连接关系。其中,元路径是由多个节点和关系构成的路径,可以表示不同实体之间的语义关系。PER 方法首先通过元路径挖掘算法来找到用户和物品之间

的重要关联性,并提取出这些关联性对应的元路径特征。然后,将这些特征用于构建一个实体推荐模型,该模型通过学习用户和物品之间的关系来进行个性化推荐。

(3) DKN^[60]: DKN 的主要贡献在于结合了知识图谱和深度神经网络的优势,以提高新闻推荐的准确性和个性化程度。DKN 模型的基本思想是将新闻和用户都表示为向量,然后使用知识图谱中的实体和关系来扩展新闻和用户的表示。模型使用卷积神经网络(CNN)和长短期记忆网络(LSTM)来学习新闻和用户的表示,并使用注意力机制来加强表示的相关性。此外,模型还使用了一个多任务框架,同时优化点击率和多样性,以提高推荐系统的效果。

(4) Wide&Deep^[61]: Wide&Deep 主要介绍了一种结合广度和深度学习的推荐系统模型。该模型由两部分组成:宽部分(Wide)和深部分(Deep)。宽部分是一个线性模型,主要用于处理高维稀疏特征。该部分利用了广告领域中的经验:相似的广告倾向于具有相似的特征。因此,宽部分使用了一组简单的交叉特征(Cross-product feature transformations)来捕获这种相似性。这些特征是由原始输入特征的笛卡尔积组成的,然后通过一个线性模型进行预测。深部分是一个神经网络模型,主要用于捕捉输入特征之间的非线性关系。该部分的输入特征通过一个多层神经网络进行处理,并最终将其映射到一个低维空间中。这个低维向量可以用于推荐系统的目标预测。

(5) RippleNet^[62]: 该方法基于异构图模型,将用户、商品和交互信息表示为节点,通过学习用户与商品之间的交互,来推荐用户可能感兴趣的物品。具体来说,RippleNet 采用了一种新颖的训练方式,将推荐过程视为基于随机游走的图嵌入任务。首先,利用随机游走生成的序列学习节点嵌入,然后通过注意力机制将不同嵌入的信息融合起来,形成用户和商品之间的关联。此外,该方法还使用了自适应的负采样策略,有效地解决了传统负采样中样本分布不均匀的问题。

5.3 实验设置与评价指标

在 MKGA 模型实验中,将第 3 章创建的多模态数据集 Multi-MovieLens1M 按照训练集、验证集和测试集的比例设置为 6:2:2。设置 epochs 数 20, $\lambda_2 = 10^{-8}$ 。知识图谱模块的学习率为 0.001。对于 TextCNN 模块,filter_num=2,dropout=0.5。

在 CTR 预测实验和 Top-K 推荐中,本文采用正确率 (accuracy, ACC)、ROC 曲线下的面积 (area under curve, AUC)、精确率 (Precision@K) 和召回率 (Recall@K) 等多种指标来评估模型性能。其中, ACC 指标衡量模型正确分类的比例, AUC 指标则用于度量 ROC 曲线的准确性, ROC 曲线的横坐标为假阳率 (False Positive Rate, FPR), 纵坐标为真阳率 (True Positive Rate, TPR), 得到 ROC 曲线, 然后对 ROC 曲线的横轴做积分便可计算得到 AUC 的值。精确率和召回率则分别衡量推荐结果中被正确推荐的样本占总样本数的比例和被正确推荐的样本占有所有应该被推荐的样本的比例。这些指标的综合考虑可以更全面地评估推荐方法的效果。

TP(True Positive)、TN(True Negative)、FP(False Positive)和 FN(False Negative) 是分类问题中常用的四种评价指标, 用来衡量分类方法的性能。其中, TP 表示实际的正样本数量与模型预测为正样本的数量的交集, 即正确地识别出的正样本数量; TN 表示实际的负样本数量与模型预测为负样本的数量的交集, 即正确地识别出的负样本数量; FP 表示实际的负样本数量与模型预测为正样本的数量的交集, 即错误地将负样本识别为正样本的数量; FN 表示实际的正样本数量与神经网络认为是负样本的数量的交集, 即错误地将正样本识别为负样本的数量。这些指标可以用来评价推荐方法在不同场景下的性能表现。具体说明及计算方式如公式 (5.1) - (5.5) 所示:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (5.1)$$

$$PRECISION = \frac{TP}{TP+FP} \quad (5.2)$$

$$RECALL = \frac{TP}{TP+FN} \quad (5.3)$$

$$FPR = \frac{FP}{N} \quad (5.4)$$

$$TPR = \frac{TP}{P} \quad (5.5)$$

5.4 实验结果与分析

5.4.1 点击率 CTR 预测

图 5-1 展示了 MKGA 模型在 Multi-MovieLens1M 数据集上的整体性能，观察实验结果可以看出 MKGA 模型优于其他模型。

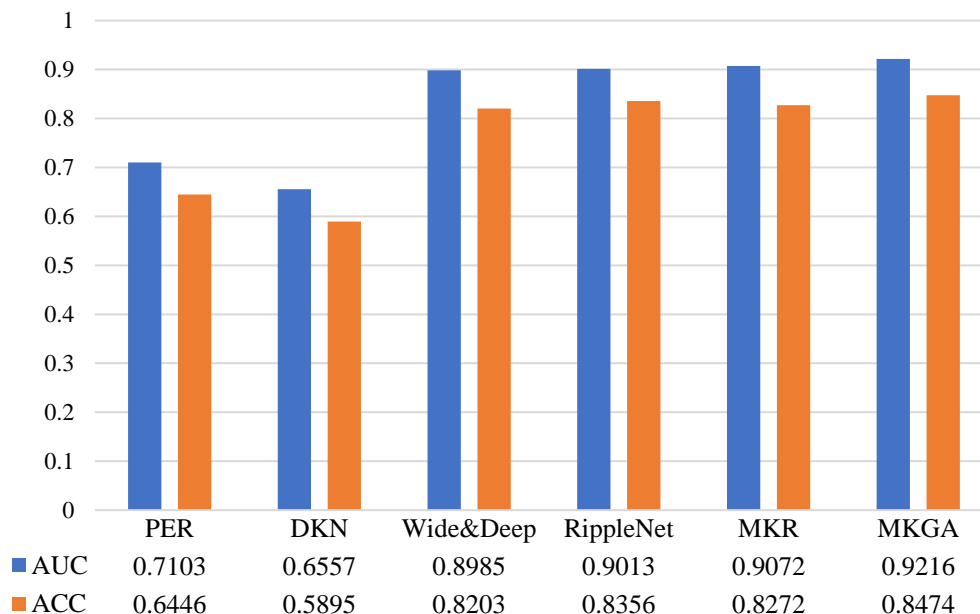


图5-1 CTR预测中的AUC，ACC

在 Multi-MovieLens1M 数据集中，MKGA 相比 PER，DKN，Wide&Deep，RippleNet，MKR，在 AUC 指标上分别提高了 21.1%，26.5%，2.3%，2%，1.44%；在 ACC 指标上分别提高了 20.3%，25.7%，2.7%，1.2%，2.0%。

•PER: PER 在电影推荐上表现不佳，因为用户定义的元路径在现实应用中几乎不可能是最优的。

•DKN: 鉴于 DKN 属于新闻领域推荐，其数据集中的文本内容长度比较短，因此 DKN 方法的结果在数据集中也不尽如人意。

•Wide&Deep: Wide&Deep 模型的性能不如 MKGA 模型，因为该模型只拼接属性，并没有像 MKGA 模型那样将文本语义分析以及视觉图像作为辅助信息集成到训练当中。

•RippleNet: RippleNet 利用注意力机制合并两跳相邻物品从而丰富用户表

明了注意力机制对高阶连通性或邻居建模的积极影响。但是，与 MKGA 模型相比，因为忽略实体本身属性且没有处理，文本属性、用户属性的一些固有信息会丢失。

•MKR: 对于 MKR 模型，因为没有考虑注意力机制，容易受噪音干扰，所以性能不如 MKGA 好。

因此，通过使用基于关系空间的注意力机制和多模态视觉图像特征提取技术是有效的，同时可以通过引入用户属性以及物品名称，可以进一步提高方法的性能。具体来说，本文所提 MKGA 模型表现最好。

5.4.2 Top-K 推荐

MKGA 模型在 Top-K 推荐中也取得了出色的表现，具体如图 5-2 和图 5-3 所示。

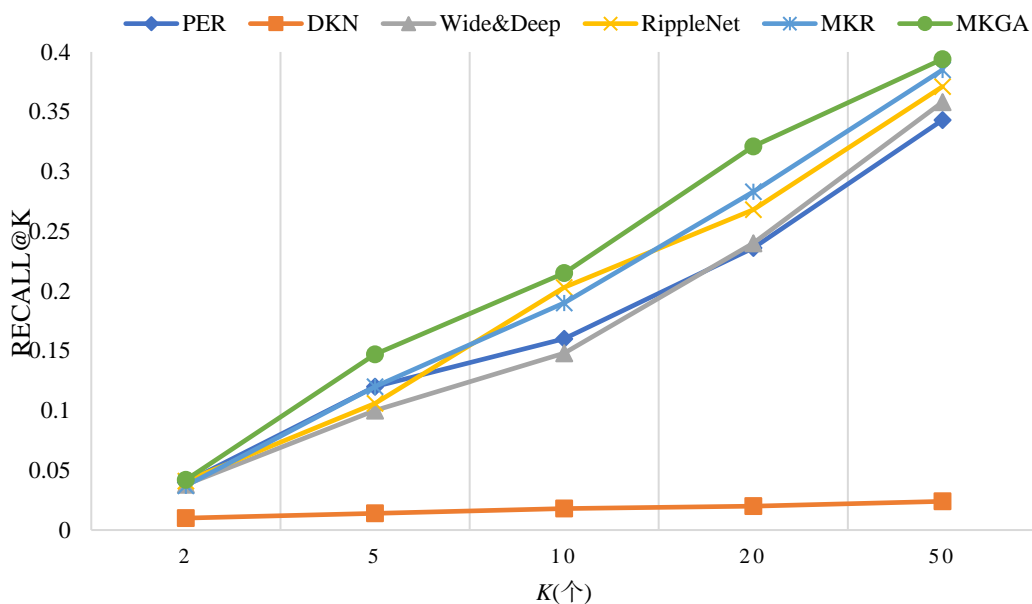


图5-2 Top-K推荐召回率

从图 5-2 和图 5-3 中可以看出，本文 MKGA 模型的准确率、召回率在 Multi-MovieLens1M 数据集上表现优异，表明添加注意力机制和多模态特征可以有效

提升模型性能。实验数据对比表明,引入知识图谱后,基线模型 **RippleNet** 和 **MKR** 模型均优于 **Wide&Deep** 和 **PER** 模型。这说明,知识图谱中的实体和关系信息有助于提高推荐性能。由于 **DKN** 应用于新闻推荐领域,适合处理长文本,本数据集中的文本长度比较短,所以尽管 **DKN** 引入了知识图谱,但方法结果在数据集表现也不尽如人意。本文所提 **MKGA** 模型对比基线模型 **RippleNet** 表现优异,是因为本文模型在基于注意力机制的基础上还添加了视觉图像和用户物品语义属性等信息特征,进一步丰富了特征向量表示,因而产生性能上的明显提升。

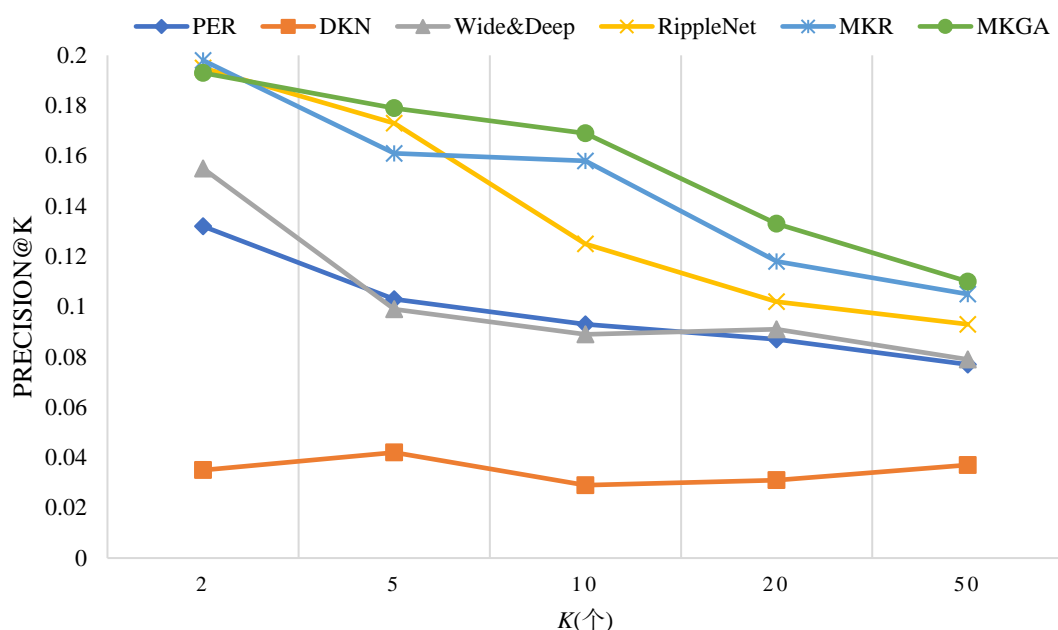


图5-3 Top-K推荐准确率

5.4.3 稀疏性验证

将知识图谱应用于推荐系统的主要目标之一是缓解系统的稀疏性问题。本文提出的 **MKGA** 模型可以进一步缓解推荐系统的稀疏性。为了探究在稀疏场景中知识图谱数据预处理层的有效性,将 **Multi-MovieLens1M** 数据集的训练集比例从 100%降至 20% (同时保持验证和测试集不变),并报告所有 CTR 预测的 AUC 结果,具体结果见表 5-1。

表5-1 不同训练比率下的AUC值

模型	20%	40%	60%	80%	100%
PER	0.6073	0.6388	0.6634	0.6886	0.7103
DKN	0.5824	0.6012	0.6203	0.6385	0.6554
Wide&Deep	0.8029	0.8156	0.8405	0.8760	0.8982
RippleNet	0.8639	0.8825	0.8858	0.8916	0.8962
MKR	0.8743	0.8804	0.8971	0.9056	0.9172
MKGA	0.8776	0.8897	0.8996	0.9089	0.9319

根据实验结果显示，MKGA 模型在数据较为稀疏的情况下仍然具有良好的推荐效果，相对于其他推荐模型来说，MKGA 模型有效地缓解了数据稀疏问题。

5.5 消融实验

在上一节中，本文提出的方法与当前主流的多种推荐算法在数据集上进行了测试对比。在本小节中，通过执行消融实验的方式来验证所提出的模型在不同场景下的性能。消融实验是一种有效的实验方法，可以通过对模型的不同部分进行剔除或更改来评估模型的贡献和效果。在本文中，将利用消融实验来深入研究所提出的模型，并从中发现其优点和不足之处，以便更好地了解该模型的工作原理。具体来说，本文将研究模型深度、不同模态和注意力机制对模型性能的影响，以及聚合层的作用。为了得到科学的结果，本研究将在每个消融实验中严格控制其他变量。首先，将深入探究模型深度如何影响其性能表现。通过对多种深度的模型进行评估，以找到最适合的深度 L 参数。接下来，将探讨不同的模态和注意力机制如何影响模型的表现。最后，将研究聚合层对模型性能的影响。通过将不同的聚合层进行比较，以确定哪些聚合层对模型性能的贡献最大。通过进行这些消融实验，能够更深入地了解所提出的模型在不同情况下的性能表现，确定模型的最佳配置，从而进一步提高推荐系统的性能。

5.5.1 超参数深度 L 的选取

可以通过调整 MKGA 的深度（例如 L ）来研究使用多层信息传递的性能效

果,使用 MKGA1 来表示具有一层的模型,并使用类似的符号表示其他层。如图 5-4 所示可以得出以下观察结果:

(1)增加 MKGA 的深度能够显著提高性能。可以明显看出,相比于 MKGA1, MKGA2 和 MKGA3 在各个指标方面均获得了持续的提高。这种改进可以归因于其对用户、物品以及实体之间高阶关系的有效建模,而建模的实现则是通过二阶和三阶的连接,这些连接有效地捕捉了数据中的复杂关联关系,从而提高了模型的性能。

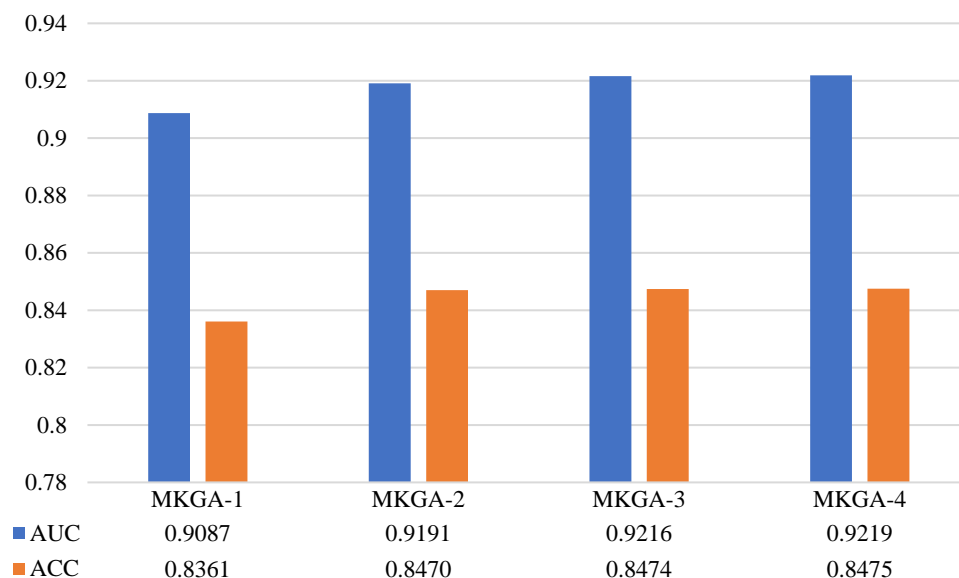


图5-4 嵌入传播层数(L)的影响

(2)在将 MKGA3 进一步改进成 MKGA4 的过程中,发现该模型只取得了边际的改进。这表明,在考虑实体之间的三阶关系时, MKGA3 已经能够有效地捕获协作信号,而叠加额外的一层并没有显著提高模型的性能。这可能是因为在 MKGA3 中,通过三阶连接建立的高阶关系已经足够丰富和复杂,能够有效地描述用户、物品和实体之间的协作行为。因此,进一步增加模型的深度并不能带来更大的性能提升。

(3)联合分析表 5-1 和图 5-4, MKGA1 在大多数情况下始终优于其他基线。它再次验证了这种基于注意力的嵌入传播的有效性,实验表明它可以更好地模拟了一阶关系。

5.5.2 多模态与注意力层影响

本小节主要研究了不同模态对于推荐结果的影响。为此，本文通过实验，对比了在考虑内部和外部模态的知识图谱以及引入注意力机制的情况下推荐系统的性能指标。具体实验结果如图 5-5 所示。其中，“baseline”表示知识图谱中只有结构化知识的实体和关系，“base+In_picture”表示知识图谱中有结构化和视觉的实体以及关系，“base+In_picture+Out_feature”表示知识图谱中有结构化、视觉的实体和关系以及外部用户物品特征，“base+In_picture+Out_feature+attention”表示知识图谱中有结构化、视觉的实体和关系、外部用户物品特征以及注意力模块。

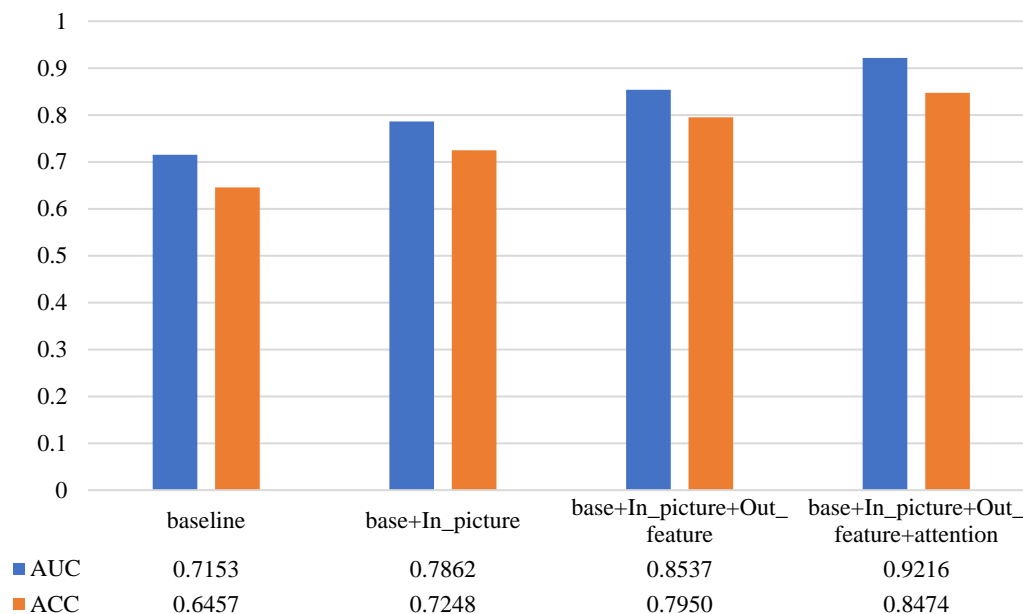


图5-5 不同模态以及注意力机制的影响

通过对比结果发现：

在推荐效果上，内部视觉模态的引入相比于基线标准模型有较好的提升。可能是由于视觉信息（例如图片）更容易引起用户的关注，因此在提高推荐效果方面，视觉模态具有更为显著的作用。再次同时引入外部用户物品特征对比只引入视觉信息的模型有着更大的提升。这表明加入用户类型特征以及物品语义特征能够明显地提升推荐效果。最后，加入注意力机制的综合模型明显优于只考虑多模态特征模型。这是因为通过注意力机制分配聚合可以得到更大范围的综合特

征,从而提高推荐的准确性。因此,可以得出本文提出的融合多模态知识图谱和注意力机制的推荐模型在推荐方面表现出色。

在实验过程中,为了得到科学的结果,本文对各组消融实验中的其他变量进行了严格控制。总之,本小节的实验结果表明,使用多模态知识图谱和注意力机制的推荐模型在推荐方面具有较好的表现,这为推荐系统的应用提供了一定的参考和启示。

5.5.3 聚合方式的影响

本组实验旨在研究不同聚合方法对于知识图谱嵌入向量的学习效果的影响。研究采用三种不同的聚合方式,分别是相加组合方式(Add)、拼接组合方式(Concatenation)和双向混合组合方式(Bi-Mix)。这三种组合方式的具体操作已在章节 4.2.2 中介绍。

实验结果如图 5-6 所示,Add 方法在所有情况下均优于 Concatenation 方法。这可能是由于采用 Concatenation 方法导致了实体表示 e_h 与其邻居网络表示 e_{s_h} 之间的交互被忽略。这一点表明在信息聚合和传播过程中,特征之间的相互作用是至关重要的,不能被忽略。与 Add 方法相比,Bi-Mix 的性能表现证明,通过结合额外的特征交互,可以改善表示学习效果。这再次验证了 Bi-Mix 方法的合理性和有效性。

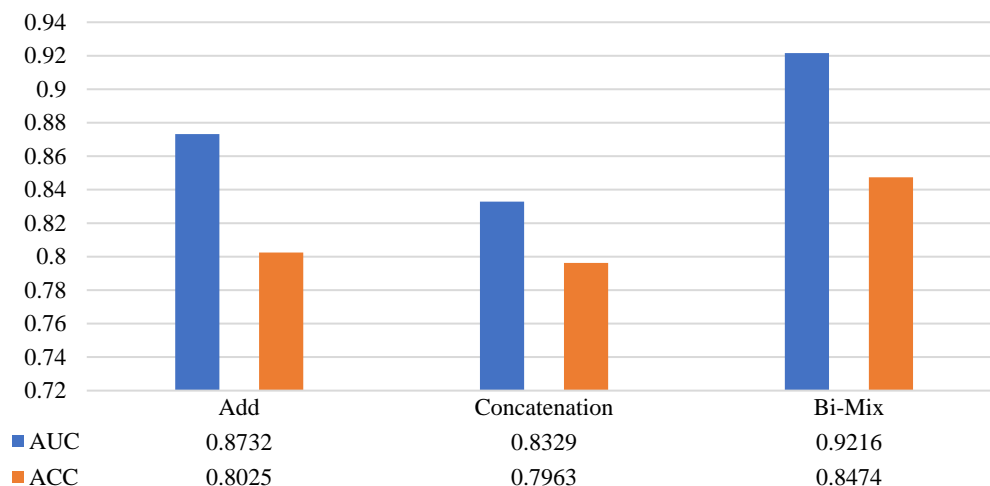


图5-6 不同聚合器的影响

综上所述,不同聚合方法对知识图谱嵌入向量的学习效果具有显著影响。在聚合过程中,特征交互应得到充分考虑,而 Bi-Mix 方法可以结合额外的特征交互,提高表示学习效果。

5.6 本章小结

本章首先阐述了本文 MKGA 模型所需要的实验环境和实验数据,并介绍了本文所采用的对比方法;接着描述了实验所需的评价指标;然后针对推荐方法的准确性和缓解数据稀疏性不足方面进行了实验分析。最后,在本章的消融实验部分,对本文所提模型的各个模块进行了详细对比分析。实验结果表明,本文所提出的方法在准确率、召回率等指标上均有明显提升,这表明本文所提出的方法是可行有效的。

第6章 总结与展望

本章对全文研究工作进行了总结，并对本文研究中需要完善之处进行了深入分析。此外，本章还对下一步研究工作提出了展望。通过对全文进行归纳概括，更加明确了本文的主要研究内容及其意义，有助于体现本文的研究价值。为了进一步提高推荐效果，本章认为未来可以从研究方法、数据来源等方面进行研究工作，并提出了相应的改进措施。

6.1 总结

在当今信息爆炸的时代，推荐技术起着至关重要的作用。在互联网上，有大量的信息、产品和服务可供选择，但是用户往往难以确定他们真正需要的是什么。此时，推荐技术可以帮助用户快速找到符合其兴趣和需求的产品和服务，提高用户的满意度和使用体验，同时也能够帮助企业提高销售额和用户留存率。但目前推荐系统存在数据稀疏性问题，针对此问题现有研究成果给出了一些解决方法，例如矩阵分解、基于邻域等方法，但这些方法没有从根本上解决数据稀疏性问题。为此，本文提出了一种融合多模态知识图谱与注意力机制的个性化信息推荐方法，该方法主要包含了多模态数据的构建与预处理方法和基于关系空间的注意力机制传播模型两个部分，本文的研究内容如下：

(1) 推荐系统在实际应用中面临着数据稀疏性问题，当用户对物品的评价或行为数据非常少甚至没有时，推荐算法难以进行训练和产生推荐。为了解决这个问题，本文提出了一种多模态数据的构建与预处理方法。该方法将多模态信息分为内部和外部两类。其中，视觉图像被用作内部多模态信息来扩展知识图谱推荐。同时，将物品的文本语义用户属性作为外部多模态特征来弥补知识图谱嵌入带来的语义损失问题。通过这种方法，推荐系统可以更好地应对数据稀疏性问题，从而提高推荐的效果。

(2) 针对知识图谱推荐中高阶信息挖掘不充分的不足，本文提出了一种基于关系空间的注意力机制传播模型。首先，从内部多模态知识图谱模块得到的多模态特征向量作为输入，并采用基于关系的感知注意力机制来学习传播过程中

每个邻居的权重,同时挖掘聚合用户和物品的高阶信息。这样可以在关系空间中找到关系更近的实体,并得到知识图谱注意力模块中的用户和物品的特征向量。在特征聚合阶段,本文采用相加、连接和双端混合三种方式进行聚合效果测试。最后,通过注意力推荐模块将外部多模态用户和物品特征向量加入融合得到最终的用户和物品特征向量,使用预测函数进行预测。这种方法可以更充分地挖掘知识图谱中的高阶信息,提高了推荐系统的准确性。

6.2 展望

本文提出了一种融合多模态知识图谱与注意力机制的个性化信息推荐研究方法,旨在解决推荐系统中数据稀疏性和推荐准确性问题。该方法融合了多模态知识图谱和注意力机制,有效地缓解了数据稀疏性的同时也提高了推荐准确性。然而,在实际应用中所提方法还存在有待完善之处,计划在以下两方面加以提升:

(1) 随着时代的发展,用户交流沟通的信息种类也越来越丰富,其中一个提升方向是进一步扩充多模态信息的应用。例如用户与物品交互的用户历史行为记录可以在某种程度上反应用户对物品的偏好行为。因此,未来的工作可以考虑将用户历史行为作为相关属性加入到特征向量中,从而更加丰富全面地表示用户的特征向量。

(2) 当进行知识图谱中的实体相关邻域传播时,如果直接邻域实体数量不足,则会导致单个行为对整体的影响过大;如果邻域实体数量过多,新增的邻居的行为对整个聚合的影响将会变得微乎其微。因此,需要更深入地探索如何在不同密度的邻域之间取得平衡。未来,可以通过设定合适的阈值,如限定最多只取K值的直接邻居进行特征融合,以更加准确地反映用户的兴趣爱好。

参考文献

- [1] Resnick P, Varian HR. Recommender systems[J]. Communications of the ACM, 1997, 40(3): 56-58.
- [2] Xiao X, Wen J, Zhou W, et al. Multi-interaction fusion collaborative filtering for social recommendation[J]. Expert Systems with Applications, 2022, 205: 117-131.
- [3] 赵晔辉, 柳林, 王海龙, 等. 知识图谱推荐系统研究综述[J]. 计算机科学与探索, 2023, 17(04): 771-791.
- [4] Wang D, Liang Y, Xu D, et al. A content-based recommender system for computer science publications[J]. Knowledge-Based Systems, 2018, 157: 1-9.
- [5] Channarong C, Paosirikul C, Maneeroj S, et al. HybridBERT4Rec: a hybrid (content-based filtering and collaborative filtering) recommender system based on BERT[J]. IEEE Access, 2022, 10: 56193-56206.
- [6] 纪文璐, 王海龙, 苏贵斌, 等. 基于关联规则算法的推荐方法研究综述[J]. 计算机工程与应用, 2020, 56(22): 33-41.
- [7] Akbar A, Agarwal P, Obaid A J. Recommendation engines-neural embedding to graph-based: Techniques and evaluations[J]. International Journal of Nonlinear Analysis and Applications, 2022, 13(1): 2411-2423.
- [8] Ran X, Wang Y, Zhang L Y, et al. A differentially private nonnegative matrix factorization for recommender system[J]. Information Sciences, 2022, 592: 21-35.
- [9] 黄立威, 江碧涛, 吕守业, 等. 基于深度学习的推荐系统研究综述[J]. 计算机学报, 2018, 41(07): 1619-1647.
- [10] Li Y, Ren J, Liu J, et al. Deep sparse autoencoder prediction model based on adversarial learning for cross-domain recommendations[J]. Knowledge-Based Systems, 2021, 220: 106-117.
- [11] 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.
- [12] 王利娥, 李小聪, 刘红翼. 融合知识图谱和差分隐私的新闻推荐方法[J]. 计算机应用, 2022, 42(05): 1339-1346.
- [13] Guo Q, Zhuang F, Qin C, et al. A survey on knowledge graph-based recommender systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(8): 3549-3568.
- [14] 潘梦竹, 李千目, 邱天. 深度多模态表示学习的研究综述[J]. 计算机工程与应用, 2023, 59(02): 48-64.
- [15] Lee K Y, Sheehan L, Lee K, et al. The continuation and recommendation intention of artificial

- intelligence-based voice assistant systems (AIVAS): the influence of personal traits[J]. Internet Research, 2021, 31(5): 1899-1939.
- [16] Jian M, Guo J, Zhang C, et al. Semantic manifold modularization-based ranking for image recommendation[J]. Pattern Recognition, 2021, 120: 108-122.
- [17] 方毅. 基于多变量模式分析的 fMRI 解码与应用研究[D]. 华南理工大学, 2017.
- [18] Li Y Y, Ramadas P, Gibson J. Multimode Tree-Coding of Speech with Pre-/Post-Weighting[J]. Applied Sciences, 2022, 12(4): 2026.
- [19] Yang Q, Wu G, Li Y, et al. Amnn: Attention-based multimodal neural network model for hashtag recommendation[J]. IEEE Transactions on Computational Social Systems, 2020, 7(3): 768-779.
- [20] 李琳, 唐守廉. 基于多层注意力表示的音乐推荐模型[J]. 电子学报, 2020, 48(09): 1672-1679.
- [21] Tao Z, Wei Y, Wang X, et al. Mgat: Multimodal graph attention network for recommendation[J]. Information Processing & Management, 2020, 57(5): 102-118.
- [22] Tiddi I, Schlobach S. Knowledge graphs as tools for explainable machine learning: A survey[J]. Artificial Intelligence, 2022, 302: 26-31.
- [23] Wu X, Tang Y, Zhou C, et al. An Intelligent Search Engine Based on Knowledge Graph for Power Equipment Management[C]//2022 5th International Conference on Energy, Electrical and Power Engineering (CEEPE). IEEE, 2022: 370-374.
- [24] Zhao F, Li Y, Hou J, et al. Improving question answering over incomplete knowledge graphs with relation prediction[J]. Neural Computing and Applications, 2022: 1-18.
- [25] 曹宗胜, 许倩倩, 李朝鹏, 等. 基于对偶四元数的协同知识图谱推荐模型[J]. 计算机学报, 2022, 45(10): 2221-2242.
- [26] Zhang Z, Cai J, Zhang Y, et al. Learning hierarchy-aware knowledge graph embeddings for link prediction[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(03): 3065-3072.
- [27] Chen X, Jia S, Xiang Y. A review: Knowledge reasoning over knowledge graph[J]. Expert Systems with Applications, 2020, 141: 112-120.
- [28] Yang H, Liu J. Knowledge graph representation learning as groupoid: unifying TransE, RotatE, QuatE, ComplEx[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021: 2311-2320.
- [29] Alam M M, Rony M R A H, Nayyeri M, et al. Language Model Guided Knowledge Graph Embeddings[J]. IEEE Access, 2022, 10: 76008-76020.
- [30] R.Xie, Z. Liu, H. Luan, et al. Imageembodied knowledge representation learning[C]. Proceedings of the TwentySixth International Joint Conference on Artificial Intelligence, 2017, 31403146.
- [31] H. MoussellySergie, T. Botschen, I. Gurevych, et al. A multimodal translationbased approach for knowledge graph representation learning[C]. Proceedings of the Seventh Joint Conference

- on Lexical and Computational Semantics, New Orleans, Louisiana, 2018, 225234.
- [32] 宋其杰. 基于深度卷积特征融合的监控场景车辆身份分层鉴别系统[D]. 南京邮电大学, 2021.
- [33] Xu L, Du Z, Mao R, et al. GSAM: A deep neural network model for extracting computational representations of Chinese addresses fused with geospatial feature[J]. Computers, environment and urban systems, 2020, 81: 101-108.
- [34] Bailly A, Blanc C, Francis É, et al. Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models[J]. Computer Methods and Programs in Biomedicine, 2022, 213: 31-39.
- [35] Wang S H, Khan M A, Zhang Y D. VISPNN: VGG-inspired stochastic pooling neural network[J]. Computers, materials & continua, 2022, 70(2): 81-88.
- [36] Wang Q, Hao Y. ALSTM: An attention-based long short-term memory framework for knowledge base reasoning[J]. Neurocomputing, 2020, 399: 342-351.
- [37] 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.
- [38] 徐兵, 弋沛玉, 王金策, 等. 知识图谱嵌入的高阶协同过滤推荐系统[J]. 计算机科学, 2021, 48(S2): 244-250.
- [39] Song W, Duan Z, Yang Z, et al. Explainable knowledge graph-based recommendation via deep reinforcement learning[J]. arXiv preprint arXiv:1906.09506, 2019, 13-20.
- [40] Guo J, Wang Y, An H, et al. IIDQN: an incentive improved DQN algorithm in EBSN recommender system[J]. Security and Communication Networks, 2022, 156-164.
- [41] 刘丹阳. 结合知识图谱的个性化新闻推荐系统[D]. 中国科学技术大学, 2022.
- [42] Li J, Xu Z, Tang Y, et al. Deep hybrid knowledge graph embedding for top-n recommendation[C]//Web Information Systems and Applications: 17th International Conference, WISA 2020, Guangzhou, China, September 23–25, 2020, Proceedings 17. Springer International Publishing, 2020: 59-70.
- [43] Da'u A, Salim N. Recommendation system based on deep learning methods: a systematic review and new directions[J]. Artificial Intelligence Review, 2020, 53(4): 2709-2748.
- [44] Westby I, Yang X, Liu T, et al. FPGA acceleration on a multi-layer perceptron neural network for digit recognition[J]. The Journal of Supercomputing, 2021, 77(12): 14356-14373.
- [45] Li Zewen, Liu Fan, Yang Wenjie, Peng Shouheng, Zhou Jun. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. [J]. IEEE transactions on neural networks and learning systems, 2021, 254-265.
- [46] 任欢, 王旭光. 注意力机制综述[J]. 计算机应用, 2021, 41(S1): 1-6.
- [47] H.Wang, F.Zhang, M.Zhang, et al. Knowledgeaware 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, 968977.

- [48] Aliyu F M, Yahaya Y I. An Investigation of the Accuracy of Knowledge Graph-base Search Engines: Google knowledge Graph, Bing Satori and Wolfram Alpha[J]. International Journal of Scientific & Engineering Research, 2021, 12(1): 11-15.
- [49] Mascarenhas S, Agarwal M. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification[C]//2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON). IEEE, 2021, 1: 96-99.
- [50] 徐岩, 徐竟泽, 曾建行, 等. 融合 HOG 与 PCA 算法的人脸识别[J]. 计算机与数字工程, 2022, 50(11): 2544-2547.
- [51] 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.
- [52] Wang H, Zhao M, Xie X, et al. Knowledge graph convolutional networks for recommender systems[C]//The world wide web conference. 2019: 3307-3313.
- [53] Kumari R, Ekbal A. Amfb: Attention based multimodal factorized bilinear pooling for multimodal fake news detection[J]. Expert Systems with Applications, 2021, 184: 115-131.
- [54] He L, Chen H, Wang D, et al. Click-through rate prediction with multi-modal hypergraphs[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021: 690-699.
- [55] 张家想. 基于图卷积神经网络的人体骨架行为识别方法研究[D]. 江南大学, 2022.
- [56] Lo W W, Layeghy S, Sarhan M, et al. E-graphsage: A graph neural network based intrusion detection system for iot[C]//NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium. IEEE, 2022: 1-9.
- [57] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903, 2017.
- [58] Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. 2018. Representation Learning on Graphs with Jumping Knowledge Networks. In ICML, Vol. 80. 5449–5458.
- [59] Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network[C]//The world wide web conference. 2019: 2022-2032.
- [60] 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..
- [61] Cheng H T, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems[C]//Proceedings of the 1st workshop on deep learning for recommender systems. 2016: 7-10.
- [62] 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.