

密级： 公开

中图分类号：TP391



浙江工商大学

硕士学位论文

论文题目： 基于图神经网络与对比学习
的标签感知推荐算法研究

作者姓名： 章寅

学科专业： 信息与通信工程

研究方向： 推荐系统

指导教师： 董黎刚

提交日期： 2023 年 6 月

**Dissertation Submitted to Zhejiang Gongshang University for
Master's Degree of Engineering**

Tag-aware recommender systems based on graph neural networks and
contrastive learning

Author: Yin Zhang

Major: Information and Communication Engineering

Supervisor: Prof. Ligang Dong



Jun. 2023

School of Information and Electronic Engineering

Zhejiang Gongshang University

Hangzhou, 310018, P.R. China

摘 要

随着信息技术和互联网大数据的飞速发展，现代社会已经进入了信息爆炸的时代。然而，数据资源在应用的数据平台中大量积累，信息超载已经成为互联网用户面临的重要挑战之一。推荐系统作为大数据时代缓解信息超载的有效工具之一，可以为用户提供个性化推荐服务。在 Web 3.0 时代，用户和数据资源的关系更为丰富。在一些网站上用户可以创造并分享各种物品，同时为物品标注各种标签。这些标签不仅可以反映用户自身的兴趣偏好和看待物品的态度，而且物品本身蕴含的丰富内容也可以为推荐系统提供更多信息。因此，标签感知推荐系统将这类社会标注数据作为协同过滤信息，为用户提供更为精准的个性化推荐服务。

当前，标签感知推荐系统已成为解决推荐系统中的稀疏性、公平性等问题的有效途径。然而，标签感知推荐系统也有其局限。其中，标签数据自有的一词多义与多词同义问题限制了标签感知推荐系统的推荐性能。虽然针对这些问题的标签感知推荐模型被一些研究人员提出，但依旧存在社会标签数据如何系统性组织、现有的深度学习模型如何有效地迁移以及如何为算法设定合适的优化方向等问题。为了更好地利用社会标签数据，提升推荐性能，本文针对以上研究问题，提出了社会标注图的定义，并在此基础上提出了两个新型的标签感知推荐模型。本文提出的模型分别通过轻量化的图神经网络与对比学习方法提高了标签感知推荐算法的表现，并降低训练难度，缓解数据中的流行度偏差。主要研究贡献如下：

(1) 本文提出了一种新的社会标注图 (Folksonomy Graph, FG)，由 <用户-标签> 图和 <物品-标签> 图组成，降低了社会标注图的复杂度，为后续模型的设计和优化提供了便利。

(2) 基于图神经网络，本文提出了一种轻量化社会标注图协同过滤 (Light Folksonomy Graph Collaborative Filtering, LFGCF)。为了适应推荐系统的特性，该模型去除了图卷积神经网络的特征变换和非线性激活组件，并采用加权和聚合函数进行消息传播。这种方式提高了模型的精准度，并减轻了模型的训练难度。

(3) 本文探讨了推荐系统中存在的数据偏差，并提出了一种标签图对比学习框架 (Tag-aware Graph Contrastive Learning, TAGCL)。该模型使用对比学习和知识图谱联合优化模型，在训练过程中对标签进行同时采样，从而有效地提高了模型的推荐精准度和

公平性。

(4) 为了评估 LFGCF 和 TAGCL 的性能, 本文设计了一系列实验, 并与当前主流的推荐算法模型在召回率、准确率、归一化累计折损增益 (normalized Discounted cumulative Gain, NDCG)、平均逆排名 (Mean Peciprocal Rank, MRR) 指标上进行对比。在三个公开的学术数据集 MovieLens、Last.FM 和 Delicious 上, 本文提出的模型 TAGCL 的召回率对比通用推荐模型有 4.6% 的提升, 准确率有 5.0% 的提升, NDCG 有 4.21% 的提升, MRR 有 1.62% 的提升。对比标签感知推荐模型, TAGCL 在召回率有 5.18% 的提升, 准确率有 8.0% 的提升, NDCG 有 7.22% 的提升, MRR 有 5.64% 的提升。对于数据偏差较大的 MovieLens 与 Last.FM 降低了 17% 的平均推荐流行度。最后, 本文还在一个真实运行的推荐系统 BibSonomy 上测试 TAGCL 性能, 实验结果证明 TAGCL 对比与基线模型的召回率提高了 1%。

关键字: 标签感知推荐算法; 图神经网络; 对比学习; 协同过滤; 个性化推荐

Abstract

With the rapid development of information technology and big data on the Internet, modern society has entered the era of information explosion. However, data resources are accumulated in large quantities in the data platforms of applications, and information overload has become one of the important challenges faced by Internet users. As one of the effective tools to relieve information overload in the era of big data, recommendation system can provide personalized recommendation service for users. In the Web 3.0 era, the relationship between users and data resources is richer. On some websites, users can create and share various items, and tag them with various labels. These tags not only reflect the user's own interests and preferences and attitudes towards the items, but also provide more information to the recommendation system with the rich content of the items themselves. Therefore, label-aware recommendation systems use such socially labeled data as collaborative filtering information to provide users with more accurate personalized recommendation services.

Currently, label-aware recommendation systems have become an effective way to solve the problems of sparsity and fairness in recommendation systems. However, label-aware recommendation systems also have their limitations. Among them, the label data's own problems of multi-word meaning and multi-word synonymy limit the recommendation performance of label-aware recommendation systems. Although label-aware recommendation models for these problems have been proposed by some researchers, there are still problems such as how to systematically organize social label data, how to effectively migrate existing deep learning models, and how to set appropriate optimization directions for algorithms. In order to make better use of social labeling data and improve recommendation performance, this thesis proposes the definition of social labeling graph for the above research problems, and based on this, two novel label-aware recommendation models are proposed. The models proposed in this thesis improve the performance of label-aware recommendation algorithms through lightweight graph neural networks and contrast learning methods, respectively, and reduce the training difficulty and mitigate the prevalence bias in the data. The main research contributions are as follows:

- (1) This thesis proposes a new social labeling graph Folksonomy Graph consisting of

<user-label> graph and <item-label> graph, which reduces the complexity of the social labeling graph and facilitates the design and optimization of subsequent models.

(2) Based on graph neural network, this thesis proposes a Light Folksonomy Graph Collaborative Filtering (LFGCF). In order to adapt to the characteristics of recommender systems, the model removes the feature transformation and nonlinear activation components of graph convolutional neural networks, and uses weighting and aggregation functions for message propagation. This approach improves the accuracy of the model and reduces the training difficulty of the model.

(3) This thesis explores the data bias present in recommender systems and proposes a Tag-aware Graph Contrastive Learning framework (TAGCL). The model uses contrast learning and knowledge graph to jointly optimize the model and sample tags simultaneously during the training process, thus effectively improving the recommendation accuracy and fairness of the model.

(4) In order to evaluate the performance of LFGCF and TAGCL, a series of experiments are designed and compared with the current mainstream recommendation algorithm models in terms of recall, accuracy, Normalized Discounted cumulative Gain, Mean Reciprocal Rank metrics are compared. On three publicly available academic datasets MovieLens, Last.FM, and Delicious, the proposed model TAGCL has a 4.6% improvement in recall, 5.0% improvement in accuracy, 4.21% improvement in NDCG, and 1.62% improvement in MRR compared to the generic recommendation model. Compared with the tag-aware recommendation model, TAGCL has a 5.18% improvement in recall, 8.0% improvement in accuracy, 7.22% improvement in NDCG, and 5.64% improvement in MRR. For MovieLens and Last.FM, which have large data bias, the average recommendation popularity is reduced by 17%. Finally, this thesis also tests TAGCL performance on a real-running recommendation system BibSonomy, and the experimental results demonstrate that TAGCL improves the recall rate by 1% compared to the baseline model.

Keywords: Tag-aware recommender systems; Graph neural networks; Contrastive learning; Contrastive Learning; Personalized recommender

目 录

中文摘要	I
英文摘要	III
插图	VIII
表格	IX
1 绪论	1
1.1 研究背景及意义	1
1.2 研究现状及存在问题	3
1.2.1 国内外研究现状	4
1.2.2 现有研究存在的问题	6
1.3 研究内容与贡献	6
1.3.1 本文研究内容	7
1.3.2 研究贡献	7
1.4 本文创新点	8
1.5 本文组织架构	9
2 标签感知推荐系统	11
2.1 社会标签数据	11
2.1.1 社会标签的研究	11
2.1.2 社会标注图	13
2.2 标签感知推荐算法	13
2.3 本章小结	14
3 基于图神经网络的标签感知推荐算法	15
3.1 图神经网络	15
3.1.1 针对图表征学习的图神经网络	15
3.1.2 基于图神经网络的推荐算法	18
3.2 基于轻量化图卷积的标签感知推荐模型 LFGCF	21

3.2.1	模型架构	21
3.2.2	消息传播操作.....	23
3.2.3	轻量化聚合操作	23
3.2.4	卷积层合并与模型预测	24
3.2.5	基于知识图谱的嵌入层	25
3.2.6	模型预测输出与联合优化	25
3.3	本章小结	26
4	基于对比学习的标签感知推荐算法	27
4.1	对比学习	27
4.1.1	对比学习回顾.....	27
4.1.2	噪声对比估计.....	28
4.1.3	推荐算法的优化方向.....	32
4.2	推荐系统的偏差	32
4.2.1	推荐系统中存在的偏差	32
4.2.2	推荐系统的流行性偏差	34
4.3	基于图对比学习的标签感知推荐模型 TAGCL	35
4.3.1	模型架构	35
4.3.2	嵌入表征的数据增强.....	35
4.3.3	多任务学习框架	36
4.3.4	优化目标函数.....	38
4.4	本章小结	39
5	标签感知推荐的算法实验评估	40
5.1	数据集	40
5.2	实验方法	41
5.2.1	实验设计	41
5.2.2	对比模型	44
5.3	实验评估	45
5.3.1	实验设置	45

5.3.2 性能对比实验分析	45
5.3.3 LFGCF 模型评估	48
5.3.4 TAGCL 模型评估	49
5.3.5 模型复杂度分析	53
5.3.6 工业数据集上的应用	53
5.4 本章小结	54
6 总结与展望	56
6.1 工作总结	56
6.2 未来研究展望	57
参考文献	58
攻读硕士学位期间取得的研究成果	66
致谢	67
独创性声明和论文使用授权说明	68

插图

1.1 标签感知推荐系统在电影推荐网站中的应用	2
1.2 本文组织架构	9
2.1 用户、标签与物品	12
3.1 空域视角下的图神经网络	17
3.2 推荐系统中的图数据	19
3.3 LFGCF 的模型结构	22
4.1 一个经典的对比学习例子	28
4.2 推荐系统中的反馈回路与偏差	33
4.3 TAGCL 的模型架构	36
5.1 图神经网络层数对 TAGCL 性能的影响	51
5.2 嵌入表征维度大小对 TAGCL 性能的影响	52
5.3 不同个性化序列长度下的 TAGCL 推荐性能	53

表 格

5.1 数据集统计	41
5.2 二分类混淆矩阵	43
5.3 数据集 MovieLens 下的模型对比实验	47
5.4 数据集 Last.FM 下的模型对比实验	47
5.5 数据集 Delicious 下的模型对比实验	48
5.6 LFGCF 消融实验	48
5.7 TAGCL 消融实验	50
5.8 模型在 Delicious 上的复杂度	54
5.9 TAGCL 在 BibSonomy 上的应用	54

1 绪论

推荐系统 (Recommender Systems) 是一种信息过滤系统, 其研究的目的是让系统预测用户 (user) 对物品 (item) 的“评分”或“偏好”。目前推荐系统以其不同的推荐目标, 可以划分为序列推荐 (Sequential Recommender)、知识图谱推荐 (Knowledge-based Recommender)、对话推荐 (Conversational Recommender) 以及标签感知推荐系统 (Tag-aware Recommender) 等。本文主要研究范围是标签感知推荐。

1.1 研究背景及意义

随着信息技术和互联网大数据的飞速发展, 现代社会已经进入了信息爆炸的时代。在这个信息量急剧上升的时代, 各大内容服务平台的数据资源规模呈指数级增长。然而, 数据资源在应用的数据平台中大量积累, 导致了信息超载 (information overload) 的问题, 这已经成为互联网用户面临的重要挑战之一^[1]。

传统的搜索引擎已经难以为用户提供高效的信息检索服务, 用户无法从丰富但相似、庞杂且无序的信息中有效地获取所需的信息。在这种信息超载的大环境下, 如何为用户过滤冗余信息, 提供独特而精准的个性化推荐服务已经成为了各大内容服务平台的核心业务之一。为此, 这些平台相继构建起符合平台业务需求的工业规模推荐系统, 基于用户与平台的交互信息, 预测用户对数据资源的评分和偏好, 从而提供以用户为中心的个性化推荐服务^[2]。

例如, 淘宝、京东等电子商务平台^[3]致力于为用户提供感兴趣的商品; 抖音、快手等短视频平台^[4]为用户推送符合偏好的视频; 小红书、美团、大众点评、爱彼迎等生活服务类信息平台^[5]为用户的衣食住行等领域推荐可能需要的服务。上述公司均通过构建大型的推荐系统, 提高了平台“长尾”数据资源利用率, 为平台创造了更多价值^[6]。同时, 推荐系统帮助发现新商品, 提供决策服务, 极大程度上缓解了用户面对信息超载时的窘境。

工业界的推荐系统通常由前端推荐展示页面、数据流模块以及推荐算法三个部分组成。其中, 推荐算法为推荐系统最核心的组件, 其功能为数据流模块中提取用户与数据资源之间的交互关系, 并根据用户的交互行为预测用户未来的交互行为, 最终为用户提供个性化的推荐结果。基于内容的推荐系统通常以用户的显式反馈为数据对象, 例如用户对推荐资源的评价, 通常以 1-5 进行评分, 并预测用户对未评价物品的评分。基于协

同过滤的推荐系统通常以用户的隐式反馈为数据对象，例如用户的点击、购买、收藏等行为，推荐算法通过这些隐式反馈来预测用户未来的交互行为^[7]。由于用户隐式反馈的数据容易收集，且更为密集，并且隐式反馈更能反映用户的真实需求，因此基于隐式反馈数据的推荐系统在工业界得到了广泛的应用，并成为了学术研究的主流方向。因此，本文的推荐算法研究将围绕推荐系统，并且在基于用户隐式反馈数据的推荐系统数据中进行研究，利用图神经网络方法为用户提供个性化的 Top-K 推荐序列。



图 1.1 标签感知推荐系统在电影推荐网站中的应用

在 Web 3.0 时代，用户和数据资源的关系更为丰富。用户不仅可以广泛地参与到网络互动中，并且他们所创造的数据和内容可以被下一代智能互联网归纳整理成更有价值的内容^[8]。许多网站提供了大众分类法（folksonomy）的功能来实现这一目标。用户可以创造并分享各种物品，同时为物品标注各种标签（tag），因此可以获得形式为 < 用户，标签，物品 > 的社会标签数据（folksonomy records）。如图 1.1 所示，不同的用户为电影“肖申克的救赎”标注了“越狱”或“发人深省”这样标签。内容服务平台通过积累大众标注行为（tagging behavior）为用户和物品之间构建起桥梁。这种标注行为不仅可以

表达用户自身的偏好，还可以对物品内容赋予更丰富、动态的理解。因此，大众标注行为作为一种用户行为，不仅可以反映用户自身的兴趣偏好和看待物品的态度，而且物品本身蕴含的丰富内容也可以为推荐系统提供更多信息。

现有推荐系统对标签信息的应用有两类，一类是标签推荐系统（Tag Recommender Systems），其目的是在用户为物品标注标签时推荐更准确表达用户观点和需求的标签，从而构建更具规范性的标签系统，缓解不同用户知识背景带来的标签多词同义性（redundancy），或标签一词多义性（ambiguity），如图 1.1 红框中推荐的标签“IMDB top 250”。另一类是标签感知推荐系统（Tag-aware Recommender Systems），其目的是将标签信息作为一种内容信息引入到推荐系统中，从而为用户提供个性化的物品资源推荐服务。这类应用可以运用系统内大量的大社会标注数据，有助于解决推荐系统的稀疏性和数据偏差问题。如图 1.1 蓝框中的标签，这些标签被用于推荐模型的训练，从而为用户带来更好的个性化推荐体验。

此外，由于大多数推荐方法都依赖于平台直接收集到的数据来训练模型，这将导致推荐结果继承先前模型中存在的偏见^[9]。推荐模型倾向于将系统内流行的物品推荐给用户，因此引发了显著的流行性偏差，这使得处在长尾（long-tail）位置的物品难以被系统推荐，导致推荐系统性能下降^[6]。如果这样的模型被直接部署到在线服务当中，整个推荐系统将会加强数据中存在的流行性偏见问题。例如，对于电影推荐系统为例，在数据中贡献巨大的热门电影可能会被系统更为频繁的推荐，甚至对已经观看过这些电影的用户也会推荐它们。因此，推荐系统会对系统中的所有用户产生偏见，对不热门的电影产生影响，加强了系统的长尾特性。

当前，图神经网络、对比学习等新一代技术的涌现为解决推荐系统中存在的问题带来新的机遇。如何充分利用内涵丰富而又简洁的标签来提高推荐系统的准确性，缓解系统中的稀疏性和公平性等问题，克服上述所提到的多词同义性和一词多义性问题，已成为探索标签感知推荐系统的当务之急。

1.2 研究现状及存在问题

本节将介绍标签感知推荐系统当前的研究现状，以及深度学习与对比学习在推荐系统中的应用。最后提出当前研究存在的问题。

1.2.1 国内外研究现状

当前, 标签感知推荐系统已成为解决推荐系统中的稀疏性、公平性等问题的有效途径。该系统通过利用社会标签提供的信息, 提高推荐准确性, 从而增强用户对推荐结果的满意度。因此, 如何充分利用标签信息来改善推荐效果成为了标签感知推荐系统的首要任务。在这一领域, 研究者们提出了一系列的模型和算法。例如, Zhen 等人^[10]提出了 TagiCofi 模型, 通过利用标签信息更好地描述用户之间的相似性, 从而提高协同过滤效果。为了解决标签的多词同义性问题, Shepisten 等人^[11]提出了一种基于分层聚类的标签过滤方法, 从而提升推荐性能。此外, 还有一些研究在用户、物品和标签之间建立关系, 以提高推荐效果。例如, Peng 等人^[12]提出了一种基于物品-标签对的推荐方法, 而 Zhang 等人^[13-14]则通过构建用户-标签-物品三部图来建模推荐系统, 从而提高准确率、用户满意度和物品新颖性。除此之外, 一些研究也利用标签挖掘用户和物品之间的相似性。例如, Zhao 等人^[15]提出了 FolkRank++ 模型, 通过利用标签信息挖掘用户和物品之间的相似性来改善推荐效果。同时, Randle 等人^[16]提出了一种基于矩阵分解的推荐模型 RTF, 并引入了基于个性化排名的评估方式, 从而为标签感知推荐系统带来了新的发展机遇。Li 等人^[17]则在 Randle 等人工作的基础上提出了 BPR-T 模型, 进一步完善了标签感知推荐系统的性能。

近年来, 随着深度学习技术的飞速发展, 越来越多基于深度学习的推荐模型应运而生。现有研究表明, 采用深度学习模型能够有效提升推荐系统的性能。在深度学习模型中, 主流的模型包括多层感知机 (Multi-layer Perceptron, MLP)^[18-21], 自编码器 (Autoencoder, AE)^[22-23], 注意力网络 (Neural Attention Network, NAN)^[24-25]以及图神经网络 (Graph Neural Network, GNN)^[26-27]等。其中, 多层感知机作为一种广泛使用的深度学习模型, 其强大的非线性学习能力是通过网络结构的宽度 (即神经元数量) 和深度 (即神经网络层数) 来实现的。自编码器则在多层感知机的基础上增加了额外的监督约束, 引入噪声干扰来学习更为稳健的嵌入表征 (embedding)。注意力网络则广泛应用于序列信息的建模场景, 能够识别数据中更为重要的特征。同时, 由于深度学习在图片、文章等结构性数据中的应用更为广泛, 推荐系统中的用户-物品交互关系则更适合采用图建模的方式。因此, 图神经网络成为了当前推荐系统领域的热门研究方向。这些深度学习模型的应用对于提高推荐系统的性能和准确性具有重要意义。

近年来,随着深度学习和图神经网络的不断发展,越来越多的推荐算法开始采用图神经网络来学习用户和物品的表示。然而,这些基于图神经网络的推荐算法通常使用传统推荐系统的优化目标函数,如基于排名的贝叶斯个性化排序(Bayesian Personalized Ranking, BPR)或均方误差(Mean Squared Error, MSE)等。具体来说,这些目标函数的训练目标和真实推荐的目标并不一致,导致训练效率低下,同时数据中的偏差问题也会被放大^[28-29]。因此,研究人员开始将对比学习引入到推荐系统中。为了解决传统推荐系统目标函数的问题,Mao 等人^[28]提出了一种基于余弦对比损失函数(Cosine Contrastive Loss, CCL)的推荐系统损失函数。他们将协同过滤分为交互编码器、损失函数以及负采样机制,并将余弦对比损失函数作为推荐系统损失函数进行优化,取得了出色的效果。另一方面,Hao 等人^[29]将对比学习应用于解决推荐系统中存在的样本不均衡问题,通过实验表明引入对比学习可以缓解流行性偏差等问题。除了引入对比学习外,还有其他一些方法也被用来解决推荐系统目标函数的问题。例如,一些研究者在现有的逐点型损失函数、配对型损失函数以外,引入软最大归一化损失^[30],以更准确地反映推荐系统的目标。针对传统推荐系统目标函数存在的问题,引入对比学习可以提高推荐算法的效果。未来的研究可以继续探索更有效的目标函数和学习方法来提高推荐系统的性能。

随着社交媒体和电子商务的迅猛发展,标签感知推荐系统已经成为个性化推荐研究领域的热门话题。尽管标签能够提供一些有用的信息,但由于标签数据的稀疏性,使得标签感知推荐系统在实践中面临着一些挑战。为了解决这些问题,近年来,越来越多的研究者开始探索利用深度学习技术来构建标签感知推荐系统。其中,Xu 等人^[31]提出了一种名为 DSPR 的模型,该模型使用多层感知器将用户和物品的嵌入表征映射到标签的特征空间。通过这种方式,模型可以利用神经网络的强大表征学习能力来解决标签数据的稀疏性问题。Zuo 等人^[32]则提出了一种名为 CFA 的模型,该模型使用稀疏自编码器(Sparse Autoencoder, SAE)将用户在标签空间中的嵌入表征用于基于用户的协同过滤中。这种方法不仅可以解决标签数据的稀疏性问题,而且还可以提高推荐准确性。Chen 等人^[33]提出了一种名为 AIRec 的模型,该模型使用多层注意力网络来捕捉不同用户对于标签的偏好,从而提高推荐性能。此外,一些研究者还尝试利用图神经网络来解决标签数据中存在的多词同义性和一词多义性等问题。Huang 等人^[34]利用图神经网络对于多跳邻居语义信息的捕获能力,提出了一种高质量的个性化标签推荐模型。Chen 等

人^[35]在存在用户、标签、节点的异构图（heterogeneous graph）上，使用区分节点类别的消息传递机制来增强个性化推荐性能。这些基于深度学习的标签感知推荐模型为解决标签数据稀疏性问题和多词同义性、一词多义性等问题提供了有力的工具和方法，将为未来的个性化推荐系统的研究和实践带来新的思路 and 方向。

1.2.2 现有研究存在的问题

目前，标签感知推荐系统中的基于深度学习的算法解决了标签空间稀疏性、一词多义和多词同义等问题，从而在推荐准确性和性能等方面展现出了优越性。然而，现有的研究方法也存在着一些问题，这些问题限制了模型性能的提升。

（1）现有的标签感知推荐系统涉及到的交互类型繁多，如何高效组织数据，并进一步建模数据中的多种交互类型极具挑战。在标签感知推荐系统中，有一些模型将标签作为用户的特征进行编码^[32-33]，而另一些模型将标签作为与用户、物品同级的实体进行建模^[13,35]。然而，这些建模方法在直接应用于图神经网络的模型时，会进一步放大数据的稀疏性，因此需要进行进一步的调整；

（2）现有应用图神经网络的模型仅仅使用了原始神经网络的设定，而没有根据标签感知推荐系统进行进一步调整。有些模型使用了复杂的消息传播机制^[26,35]，这使得模型的训练难度大幅增加，需要更多的训练轮数才能使得模型收敛。相比之下，简单的消息传播机制不仅有效降低了训练难度，还通过多跳邻居提供了丰富的上下文语义，从而有效地缓解了一词多义和多词同义问题；

（3）现有模型所采用的优化机制往往忽略了图神经网络自身的特点，盲目地沿用传统推荐算法的优化目标进行模型优化。大多数模型仅仅使用 BPR 单一损失作为优化方向，忽略了现代推荐系统中广泛存在的流行性偏差问题。在多任务学习的设定下，联合使用知识图谱、对比学习等方法可以有效地克服数据中存在的偏见，使得推荐系统得到的结果更为公平。

1.3 研究内容与贡献

本节将介绍本文的主要研究内容，介绍研究该课题的必要性。同时总结本文成果，概括本文对于该领域的学术贡献。

1.3.1 本文研究内容

本文旨在将标签信息融入推荐算法，以构建用户和物品的桥梁。针对标签感知推荐系统目前面临的问题，本文将深入探讨社会标签信息中存在的图数据结构，为用户-标签-物品的交互过程做出合理的建模。同时，将利用图神经网络和对比学习，提出两种标签感知推荐模型，最终为用户带来优秀的 Top-K 个性化推荐能力。本文的研究内容主要包括以下四个方面：

(1) 深入探索标签感知推荐系统

本文将标签感知推荐系统的研究重点划分为数据与算法两个部分。对于社会标注数据部分，本文分析过去文献对数据的建模方式存在的问题，为用户-标签-物品的交互过程做出合理的建模。同时，对于标签感知推荐算法，本文将标签感知推荐系统抽象为数学形式，为后续研究打下基础。

(2) 提出基于图神经网络的标签感知推荐模型

为避免图神经网络中冗余的架构，本文去除了不适用于标签感知推荐系统的部分，提出了轻量化设计的图神经网络算法。该模型主要由轻量化设计的图神经网络结构组成。此外，本文利用基于知识图谱的标签关系映射的正则化函数为该模型学习到更稳健的特征。

(3) 提出基于对比学习的标签感知推荐模型

为避免在数据中广泛存在的推荐系统不公平现象，本文提出了基于图对比学习的标签感知推荐系统。在轻量化图神经网络算法的基础上，使用带有知识图谱与对比学习的多任务优化任务进行模型优化。

(4) 实验验证和结果分析

本文提出的两个模型在三个公开数据集上与标签感知推荐系统领域主流模型进行对比。实验结果表明，本文提出的模型具有优越的 Top-K 推荐性能。同时，本文还在一个真实运行的系统数据中验证了模型的有效性。

1.3.2 研究贡献

本文主要探讨了推荐系统中的标签感知问题以及不公平现象，通过重新设计的社会标注图 (Folksonomy Graph, FG) 的基础上，提出了轻量化社会标注图协同过滤 (Light

Folksonomy Graph Collaborative Filtering, LFGCF) 和标签图对比学习框架 (Tag-aware Graph Contrastive Learning, TAGCL)

(1) 本文首先提出了一种新的社会标注图, 该图由用户-标签图和物品-标签图组成, 降低了异构图的复杂度, 方便后续模型的设计和优化;

(2) 本文基于图神经网络, 提出了一种轻量化的标签感知推荐模型 LFGCF。为了适应推荐系统的特性, 该模型去除了图卷积神经网络的特征变换和非线性激活组件, 并采用加权和聚合函数进行消息传播, 从而提高了模型推荐的精准度, 并降低了模型的训练难度;

(3) 本文探讨了推荐系统中存在的不公平现象, 提出了一种新的模型 TAGCL。该模型使用对比学习和知识图谱联合优化模型, 在训练过程中对标签进行同时采样, 从而有效地提高了模型推荐的精准度和公平性;

(4) 本文设计了一系列实验以评估 LFGCF 和 TAGCL 的性能, 并与当前主流的推荐算法模型进行对比。本文还在一个真实运行的系统数据中验证了提出模型的有效性。

综合来看, 本文的研究贡献为标签感知推荐系统和数据中的不公平问题提供了新的思路和解决方案。

1.4 本文创新点

本文的主要创新点有以下三点:

(1) 本文提出一种新的社会标注图结构。该图分别由用户-标签图和物品-标签图组成, 创新性地降低了异构图的复杂度;

(2) 本文提出了基于轻量化图神经网络的推荐模型 LFGCF。本文去除图卷积神经网络的特征变换与非线性激活组件, 并使用加权和聚合函数进行消息传播。创新性的轻量化设计降低了模型的训练难度, 并提高了模型推荐的精准度;

(3) 本文提出了基于对比学习的推荐模型 TAGCL。在模型优化时, 使用对比学习与知识图谱联合优化模型, 在训练过程中对标签进行同时采样。创新性地应用了多任务学习框架, 提高模型推荐精准度的同时优化了推荐系统中不公平的现象。

1.5 本文组织架构

本论文分为 6 章进行组织，第 1 章引言介绍标签感知推荐的研究背景和研究意义，并阐述本文的主要研究内容与创新点。第 2 章介绍社会标签数据和标签感知推荐系统的定义。第 3 章介绍基于轻量化图神经网络的推荐算法研究工作。第 4 章介绍基于对比学习的推荐算法研究工作。第 5 章介绍推荐算法实验结果与数据分析。最后在第 6 章总结全文，具体安排如下：

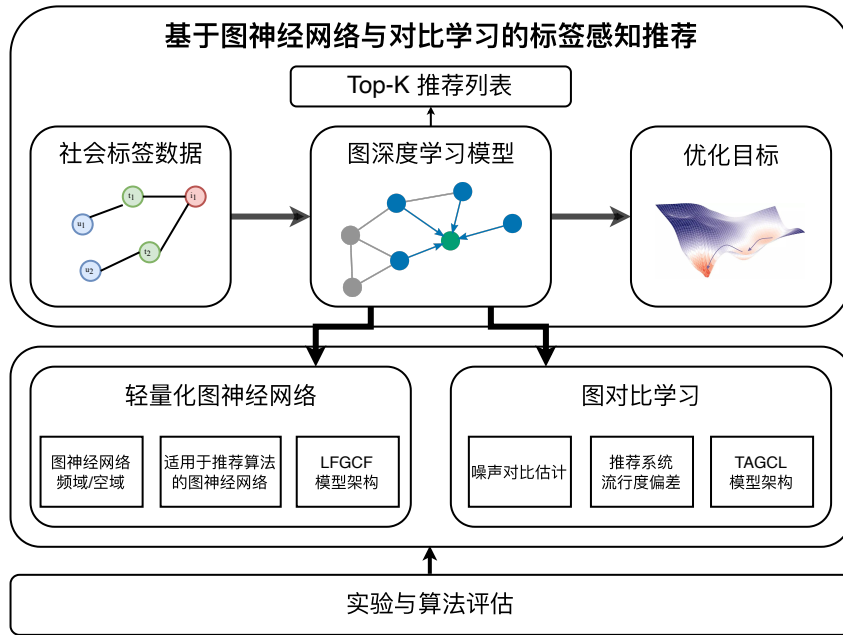


图 1.2 本文组织架构

第 1 章介绍了标签感知推荐系统的背景和研究意义，并总结了国内外研究现状及其存在的问题。同时概括了本文的研究内容、贡献和创新点，并提出了本文的架构。

第 2 章首先综述了基于深度学习、图神经网络和对比学习的推荐算法的相关研究。接着深入探讨了现有研究存在的问题，并阐述了本文的技术路线。对研究中所涉及的图神经网络、对比学习和推荐系统的公平性等理论和技术进行了介绍。最后，构建了本文的核心建模——社会标注图，并定义了标签感知推荐系统的数学形式。

第 3 章首先详细地介绍了图神经网络，并且介绍了图神经网络应用与推荐算法的两个模型。之后详细介绍了本文提出的基于轻量化图卷积网络的 LFGCF。特别是如何去除原始图神经网络中不必要的组件，从而构建轻量化的图神经网络模块，从消息传播层、卷积层合并和模型输出角度介绍模型，并介绍了模型的训练细节。

第4章首先介绍了对比学习并推导了对比学习中最重要损失函数 InfoNCE。之后详细介绍本文提出的基于轻量化图卷积网络的 TAGCL，着重描述如何利用对比学习以及副标签采样和 TransT 解决数据中的流行度偏差，并介绍了 TAGCL 的训练细节。

第5章主要对本文提出的两个模型 LFGCF 和 TAGCL 进行实验验证和性能分析。首先简要介绍了实验使用的数据集、数据预处理方法、实验流程、评估指标和对比模型，然后阐述了实验模型相关参数配置。评估结果表明，与现有标签感知推荐模型相比，本文提出的解决方案具有显著的性能提升。此外，设计了多组实验对模型的结构和相关参数进行评估。最后，本文在一个真实运行的推荐系统数据中进行离线实验，结果表明模型可以有效的提高推荐性能。

第6章对本文的研究工作进行简要总结，并对下一阶段的工作进行展望，为后续标签感知推荐模型的研究指明方向。

2 标签感知推荐系统

标签感知推荐系统是当前推荐系统研究的一个研究热点，其研究内容主要分为数据和算法两个方面。在数据部分，最为关键的是如何准确地定义社会标签数据，并将其有效地融入到标签感知推荐算法中。社会标签数据的定义对于整个推荐系统至关重要。算法部分则负责从社会标签数据中学习用户、标签、物品的表征，并将其运用到最终的推荐环节中，为用户提供个性化的推荐服务。

2.1 社会标签数据

社会标签数据是指用户在社交网络和互联网等平台上由用户自主添加的标签信息。社会标签数据以其丰富的语义信息和用户生成的特点，成为推荐系统中重要的数据来源之一。本节将重点介绍标签感知推荐系统中常用的社会标签数据建模方式，并对社会标签数据的建模定义进行详细讲解。

2.1.1 社会标签的研究

社会标签数据是标签感知推荐系统实现个性化推荐的关键基础。社会标签数据被广泛应用于解决用户和物品之间信息缺失的问题，从而更加准确地描述用户的偏好和物品的特征^[8,36]。这些社会标签的数据是通过标签感知推荐系统中的大众分类工具（即 Folksonomy）进行积累的。该工具允许用户对他们所交互的物品（例如电影、音乐和书签）自由地进行标注。由此组成的社会标签数据是由简明扼要的单词或短语组成，可以反映单个用户的观点，同时也可以获得大众集体智慧的评价。从这个角度来看，社会标签数据中的标签可以作为用户和物品之间交互的桥梁。通过对社会标签数据的探索，标签感知推荐系统可以为用户提供精确的个性化 Top-K 推荐结果^[17,32,35]。因此，将社会标签数据引入推荐系统可以提高推荐的质量，同时在一定程度上提高推荐系统的可解释性。

在标签感知推荐系统中，如何准确地定义和处理社会标签数据是实现个性化推荐的基础。常用的社会标签数据建模方式包括基于用户的标签建模、基于物品的标签建模和基于标签的用户-物品关联建模等。其中，基于用户的标签建模主要是通过对用户行为数据的分析，从中提取出用户与标签之间的关系，并构建 <用户-标签> 的映射关系；基于物品的标签建模则是针对不同物品之间的标签差异进行建模，从而分析物品与标签之

间的关系；而基于标签的 < 用户-物品 > 关联建模则是将标签作为桥梁，分析用户和物品之间的关系。

尽管已有多种方法提高了标签感知推荐算法的推荐性能，但由于数据的稀疏性和标签的一次多义性与多词同义性，推荐系统仍然面临一些无法接受的问题^[11]。稀疏性指大多数用户只为他们少数交互过的物品标注标签，导致标签数据不够充分。一次多义性和多词同义性的问题来自于上下文信息的缺失以及用户表达习惯的不确定性。图 2.1 展现了一些例子。一些具有不同形式的标签有着相同的含义，图中的“第二次世界大战”和“二战”在社会标签数据中是不同的标签，但通常被不同的用户分配给同一部电影“辛德勒的名单”。另一些相同的标签却有着不同的含义，图中的“苹果”在科技爱好者群体当中，会被视为是一家科技公司，但大多数人则认为它是一种水果。

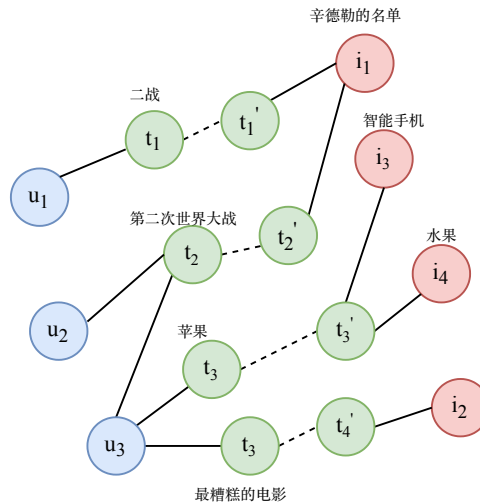


图 2.1 用户、标签与物品

本文针对社会标签数据进行了建模定义，提出了一种基于标签的用户-物品关联建模方法。具体而言，通过对用户在社交网络平台上添加的标签进行分析，提取出用户和标签之间的关系，并将其映射到用户-标签矩阵中，以此构建起用户-标签图；同时，通过对物品的标签进行分析，提取出物品与标签之间的关系，并将其映射到物品-标签矩阵中，以此构建起物品-标签图。在此基础上，通过对用户-标签图和物品-标签图进行分析，建立起基于社会标注图，从而实现对用户个性化需求的精准推荐。

2.1.2 社会标注图

本文研究标签感知推荐系统中，利用大众标注记录的两种主要方式。目前主要存在两种方法，一种是将标签编码成隐向量的形式^[32-33]，另一种是构建包含多种关系的异构图^[35]。本文与上述方法不同之处在于，利用大众标注记录中的 $a = (u, t, i) \in \mathcal{A}$ 两组边 $\mathcal{E}_{u,t}$ 和 $\mathcal{E}_{i,t}$ 构建了两个含义不同的二部图。其中，集合 $\mathcal{E}_{u,t}$ 表示用户主动标注标签的行为，而集合 $\mathcal{E}_{i,t}$ 表示物品被动接受标签的行为。在标签感知推荐系统当中，集合 $\mathcal{E}_{u,t}$ 和 $\mathcal{E}_{i,t}$ 中的边可以被定义为：

$$e_{u,t} = \begin{cases} 1, & \text{if } (u, t) \in \mathcal{E}_{u,t} \\ 0, & \text{otherwise} \end{cases}, \quad (2-1)$$

$$e_{i,t} = \begin{cases} 1, & \text{if } (i, t) \in \mathcal{E}_{i,t} \\ 0, & \text{otherwise} \end{cases}. \quad (2-2)$$

基于上述边的定义，本文可以根据整个社会标注数据 \mathcal{A} ，构建一个社会标注图。该图由用户-标签的标注行为和物品-标签的被标注行为组成。每种边的集合可以使用两个二部图 $\mathcal{G}_{UT} = (\mathcal{U}, \mathcal{T}, \mathcal{E}_{u,t})$ ， $\mathcal{G}_{IT} = (\mathcal{I}, \mathcal{T}, \mathcal{E}_{i,t})$ 来表示。

2.2 标签感知推荐算法

推荐系统是一个涉及计算机科学、心理学、社会学等多个学科领域的综合性问题，包含了许多内容。然而，推荐算法的模型可以被抽象为“学习推荐”(learning to recommend)的问题。在这个问题中，我们的目标是学习一个模型，该模型可以根据用户历史行为和物品属性等信息，预测用户可能喜欢的物品并向其进行推荐。本节将介绍如何对标签感知推荐算法进行数学建模。

推荐系统是一个涉及多个学科领域，如计算机科学、心理学和社会学等的广泛领域。推荐算法的模型可以被抽象为“学习推荐”的问题。在此背景下，本节介绍如何将标签感知推荐算法数学化。标签感知推荐算法的问题可以抽象为学习预测函数 f 。该函数以社会标签数据 \mathcal{A} 中用户 u 的信息集合 \mathcal{U} 、物品 i 的信息集合 \mathcal{I} 和标签 t 的信息集合 \mathcal{T} 作为输入，输出预测分数 \hat{y} ，表示用户与物品之间发生交互的概率， $f: \mathcal{U} \times \mathcal{T} \times \mathcal{I}$ 。接

着, 对每个用户候选物品按交互概率从高到低排列, 从中选择 K 个物品组成用户的个性化推荐列表, 这一任务被称为 Top-K 推荐任务。根据所使用的数据方式不同, 现有的推荐算法研究可以划分为基于协同过滤的方法和基于富信息的方法^[37]。两者的区别在于, 基于协同过滤的方法通常只使用用户和物品信息中的 ID 特征, 即用户在推荐系统中的唯一标识, 而基于富信息的方法则会引入如用户性别、用户年龄、物品评论等信息。与之前将标签信息作为富信息的方法不同^[32-33,38], 本文主要探讨基于协同过滤方法, 仅使用标签的 ID 特征而不考虑标签内容, 即:

$$\hat{y}_{ui} = f(u, i, t), \quad (2-3)$$

在构建了推荐模型之后, 通常会使用优化目标函数来优化推荐模型的参数, 即:

$$\min_{\Theta} \mathbf{E}_{\mathcal{A}^+} \mathcal{L}(f), \quad (2-4)$$

其中, $\mathcal{L}(\cdot)$ 是定义在 f 上的损失函数, \mathcal{A}^+ 是观测到的社会标签数据集合, Θ 是模型参数。

由于本文专注于为标签感知推荐系统中用户提供个性化推荐列表。通过“学习推荐”得到推荐模型后, 标签感知推荐系统将会生成 Top-K 推荐列表:

$$Top(u, K) = \underset{i \in I}{\operatorname{argmax}}^{(K)} \hat{y}_{u,i}. \quad (2-5)$$

2.3 本章小结

本章探讨了标签感知推荐系统数据和算法两个方面。在数据部分, 本章梳理了构建标签数据的研究现状, 并基于现有研究的缺陷给出了社会标注图的定义。在算法部分, 本章给出了标签感知推荐系统的模型构建和优化目标的形式化定义。

3 基于图神经网络的标签感知推荐算法

本章首先将介绍本文的核心方法——图神经网络，接着介绍推荐系统是怎样更好地利用图神经网络的。最后本章提出轻量化的社会标注图协同过滤模型 (Light Folksonomy Graph Collaborative Filtering, LFGCF) 的设计。

3.1 图神经网络

本节将介绍图神经网络的一般形式。初始设定的图神经网络通常用于进行节点分类、整图分类等任务。如何合理的调整模型的组件是图神经网络在推荐系统应用的研究重点与热点。

3.1.1 针对图表征学习的图神经网络

近年来，深度学习在处理符合欧氏结构 (euclidean structure) 的数据类型，如图片和文本等方面取得了显著的成果。这些数据通常被视为高维张量 (Tensor) 进行处理，具有信息量丰富、结构稠密等特点。然而，在实际场景中，非欧式结构 (non-euclidean structure) 的数据类型却更加普遍，其中最常见的便是图数据。推荐系统中的数据就包含大量的图数据，例如由大量用户和物品组成的用户-物品二部图、用户与用户之间的社交网络以及物品与物品之间的实体关系知识图谱等。这些数据通常是极度稀疏的，这使得深度学习模型在推荐系统领域并没有很快取代传统算法。而图神经网络的出现，则有效地解决了这一窘境^[37]。

随着图数据的广泛应用，图神经网络逐渐成为了研究热点。图神经网络是一种可优化的神经网络结构，专门用于处理图上的节点、边和整个图的属性，并通过邻接矩阵的表述方式保持图的对称性。目前，主流的图神经网络主要基于消息传递神经网络 (Message Passing Neural Network, MPNN) 构建^[39]。此外，图卷积神经网络 (Graph Convolutional Network, GCN) 是一种在图神经网络领域广受欢迎的模型，由 Thomas 等人在 2017 年提出^[40]，并已经成为许多图机器学习任务的最优解。图卷积神经网络将卷积操作引入图神经网络，可以逐层更新节点的特征表征而不改变节点之间的连接关系。图神经网络接收一张图作为输入，并通过逐步转换节点、边和整个图的特征表征，实现对图数据的有效处理。图数据通常包括两个方面的信息：

- (1) 图的结构，该数据通常以图的邻接矩阵 A 组成;

(2) 图上节点上的特征, 通常以 $N \times D$ 的特征向量 X 描述, 其中 N 为节点总数, D 为输入特征向量的维度。类似的, 边和整图也可以拥有类似的特征向量。

其中, 一类是反映图拓扑结构的信息, 另一类是反映附加特征的信息。通过图神经网络, 可以获得节点级别的输出 $Z \in \mathbf{R}^{N \times F}$, 其中, F 是输出向量的维度。此外, 通过使用不同的读出结构, 如节点分类、链接预测以及整图分类, 可以实现不同的图机器学习任务。与其他神经网络类似, 图神经网络同样具备特征变换和非线性激活等神经网络的特点。

通常神经网络层可以被记为一个非线性方程:

$$H^{(k+1)} = f(H^{(k)}, A) \quad (3-1)$$

其中, $H^{(0)} = X$ 且 $H^{(K)} = Z$, K 为网络的层数。神经网络的模型架构仅仅由于 $f(\cdot, \cdot)$ 的不同而不同。

本小节首先考虑一个简单的消息传播情形:

$$f(H^{(k+1)}, A) = \sigma(AH^{(k)}W^{(k)}) \quad (3-2)$$

其中, $W^{(k)}$ 代表神经网络的第 k 层权重, 而 σ 则为该层神经网络的非线性激活函数, 例如 $ReLU$ 。虽然这种图神经网络已经实现了特征变换、非线性激活等等基本特性, 但它有两个非常关键限制。首先, 邻接矩阵 A 计算乘法时, 会丢失节点自身^[40]。这使得在传递信息时只考虑了邻居节点的信息。其次, 邻接矩阵 A 未被归一化, 这导致乘法运算会改变特征向量的数值尺度。这两个问题限制了图神经网络在某些任务上的表现。

图卷积神经网络通过强制添加自循环 (self-loops) 与归一化邻接矩阵克服了以上问题。自循环是在邻接矩阵 A 上增加与 A 维度相同的对角矩阵 I :

$$\hat{A} = A + I \quad (3-3)$$

图的邻接矩阵 A 可以将矩阵内的每一行求和来实现, 即矩阵的度。因此, 定义 D 为邻接矩阵 A 的对角度矩阵。则对邻接矩阵进行归一化计算可以被定义为:

$$Norm(A) = D^{-1}A \quad (3-4)$$

在实践中，为计算方便，会进一步使用对称的方式进行归一化计算：

$$Norm(A) = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \quad (3-5)$$

结合以上两个方法，便可以得到图卷积神经网络的定义：

$$f(H^{(k)}, A) = \sigma(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(k)} W^{(k)}) \quad (3-6)$$

图神经网络也可以从图的空域上进行理解。与本节先前的设定相同， $H^{(k)}$ 为第 k 层图卷积层上，所有节点的特征向量。且第 k 层的特征向量由第 $k-1$ 层通过卷积操作获得。第一层的特征向量 $H^{(0)} = X$ 。图卷积神经网络的计算过程由三个步骤实现，分别是消息传播、特征变换与非线性激活。

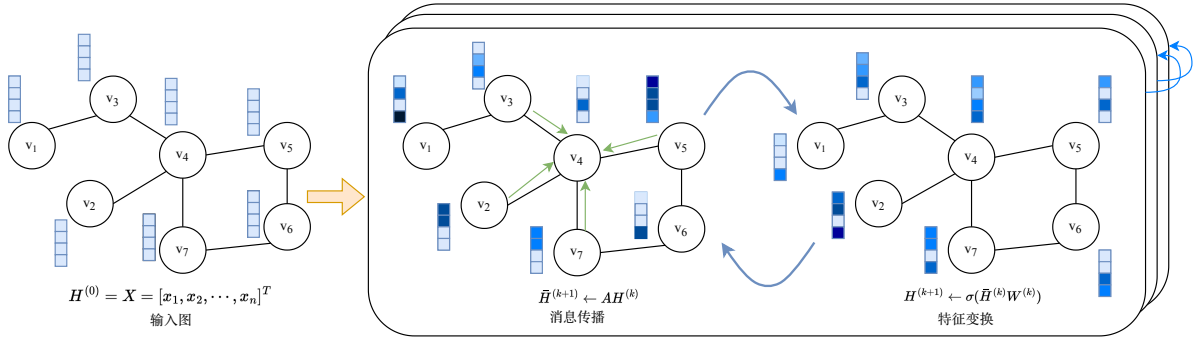


图 3.1 空域视角下的图神经网络

在消息传播阶段，每一个节点都会获取上一层邻居节点的向量。如图 3.1 节点 v_4 从邻居节点 v_2 、 v_3 、 v_5 、 v_7 以及自身 v_4 处获得了上一层图卷积的向量。在在特征变换阶段，每一个节点会通过邻居节点进行归一化后，与一个被训练的权重参数相乘。在非线性激活阶段，以上结果将会通过一个非线性函数从而获得非线性表达能力。因此，对于任意的节点 v_i 而言，在第 k 层获得的特征向量 $h_i^{(k)}$ 可以被定义为：

$$h_i^{(k)} = f^{(k)}(W^{(k)} \cdot \frac{\sum_{u \in \mathcal{N}(i)} h_u^{(k-1)}}{|\mathcal{N}(i)|} + B^{(k)} \cdot h_i^{(k-1)}) \quad (3-7)$$

其中 $\mathcal{N}(i)$ 为节点 v_i 的所有邻居节点， $h_u^{(k-1)}$ 为节点 v_i 的一个邻居节点 v_u 在第 $k-1$ 层的表达， $h_i^{(k-1)}$ 为节点 v_i 在第 $k-1$ 层的表达。 $|\mathcal{N}(i)|$ 表示对求和后特征向量进行归一化，以防止向量的数值在经过数层计算后发生尺度上的变化。 $B^{(k)}$ 是对节点自身特征的整合操作。

图神经网络通过聚合邻居节点的信息来生成当前节点的表达，其核心思想在于可以通过堆叠多个卷积层来增大单个节点的感受域 (receptive field)，从而使节点可以触及多跳邻居的信息，具备高阶连接性。然而，高阶图神经网络会带来过平滑 (oversmoothing) 的问题^[41]。为了解决这个问题，出现了一些优秀的模型对 GCN 进行改进，例如 GraphSAGE^[42]和 GAT^[43]。GraphSAGE 主要从空域切入，通过采样的方式选择一定数量的邻居，避免了计算时将所有邻居节点参数同时加载到内存造成的内存溢出问题，从而避免了复杂度接近 $O(|V|)$ 的极端情况，并且极大的扩展了图神经网络的尺度，使得图神经网络首次引入到工业推荐系统 Pinterest^① 中。GAT 对邻居节点重要性进行改进，与 GCN 使用邻接矩阵的度作为权重不同，GAT 使用注意力机制来衡量邻居节点相对于目标节点的重要性，从而可以为更加重要的邻居分配更大的权值。

3.1.2 基于图神经网络的推荐算法

在过去几年中，有许多关于图神经网络的推荐算法工作被提出。最直接的原因在于，大多数推荐系统中的数据基本都有图结构，而图神经网络技术已经被证明在各个领域的图表示学习中是非常强大的^[44]。用户和物品之间的交互数据可以用用户和物品节点之间的二部图 3.2a 来表示，其中用户与物品之间的边表示了用户和物品之间的交互。此外，用户与多个物品的交互也可以转换为序列图，表示了物品交互的先后关系 3.2b。除此之外，社交网络 3.2c。与知识图谱的例子 3.2d。

虽然利用高阶的协同过滤信息在推荐系统中并不是全新的想法。例如，Koren 等人的 SVD++ 算法^[45] 结合了被交互的物品等表征来丰富用户表征。另外，Gori 等人的 ItemRank 算法^[46] 通过随机游走算法从物品-物品图中学习表征，并根据用户的偏好对物品进行排名。其中，SVD++ 可以视为是利用一阶邻居来改善用户表征，而 ItemRank 则是利用二阶邻居来改善物品表征。与这些传统算法对比，图神经网络更为灵活，可以更方便地对用户-物品交互中的多跳连接进行建模，并且这些高阶的协同过滤信息已经被证明对推荐系统是有效的^[47]。对于推荐系统而言，其中的图数据通常包括两个方面的信息：

- (1) 交互信息，记录用户域与物品域之间的交互数据。例如，在标签感知推荐系统

① <https://www.pinterest.com/>

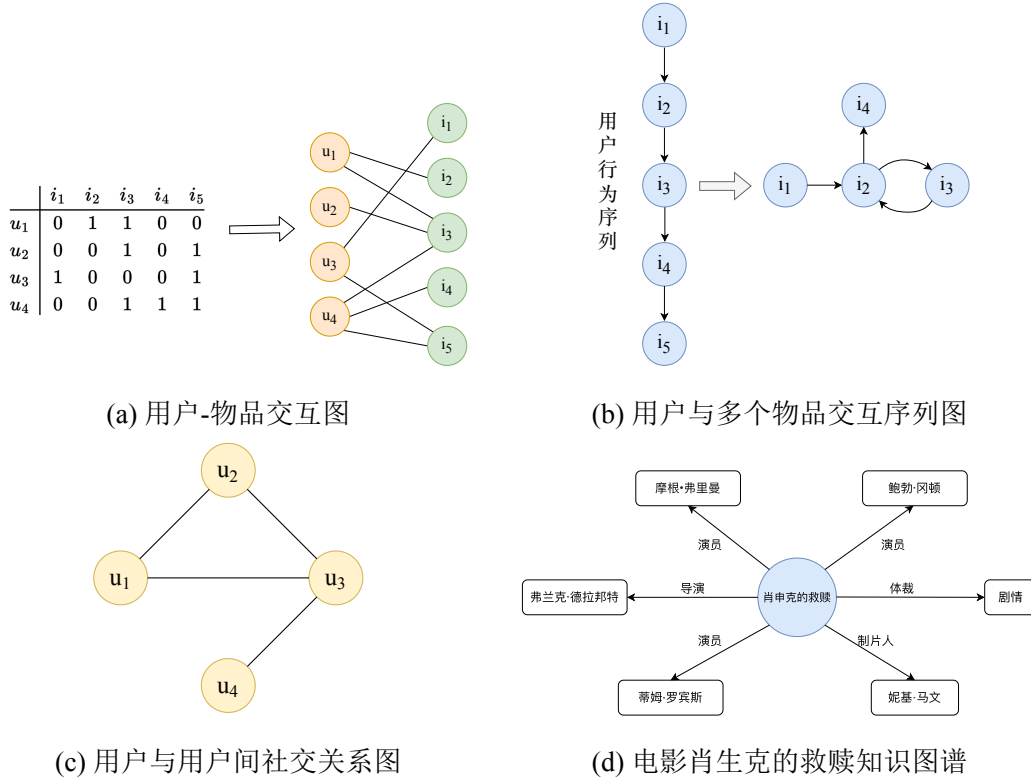


图 3.2 推荐系统中的图数据

中，社会标签数据是一种交互信息。用户使用标签标注物品，这种交互形成了三种边，即用户-标签、物品-标签和用户-物品之间的边；

(2) 节点特征，记录某个域自身附带的信息。例如，用户的年龄、性别等信息，物品的标签、评论等信息，这些信息可以作为一类附加信息（side information）在训练时使用。

而基于协同过滤的推荐方法通常只使用第一类信息，这类信息由用户与物品的 ID 特征构成的三元组组成。而并不去使用用户、物品自身的信息。由于图神经网络非常契合推荐系统的数据形式，因此一些算法算法，如 Wang 等人^[26]尝试将完整的图神经网络结构引入推荐系统，提出了 NGCF 模型。但早期的应用图神经网络的算法依照 GCN 相同的消息传播来调优模型的参数：特征变换（feature transformation）、邻居聚合（neighborhood aggregation）与非线性激活（nonlinear activation）。尽管这样的算法展示了图神经网络在推荐系统中大有作为，但更新的研究则证明，完整的图卷积神经网络对于推荐系统而言过于冗余^[27]。原始的图卷积神经网络最初是设计用于带有附加特征的图（attributed graph）上的节点分类、整图分类等任务。这类图数据中的节点通常都附带有额外的特征

信息作为输入模型的嵌入特征。而在推荐系统中的用户-物品二部图中，每个节点（用户或物品）仅由节点 ID 作为输入，缺乏额外的语义信息贡献，仅用于区分不同的节点。因此，去除图卷积神经网络中的某些组件，将有助于图神经网络在推荐系统中的应用，He 等人^[27] 由此提出了 LightGCN 模型。本小节将继续介绍 NGCF 与 LightGCN 这两个算法。

3.1.2.1 NGCF 简介

对于基于协同过滤的推荐方法，用户、物品的 ID 通常与唯一的一个嵌入表征表示。令用户 u 的表征记为 $e_u^{(0)}$ ，物品 i 的表征则为 $e_i^{(0)}$ 。与图卷积神经网络的结构类似，NGCF 在用户-物品二部图上的消息传播可以被定义为：

$$\begin{aligned} e_u^{(k)} &= \sigma(W_1 e_u^{(k-1)} + \sum_{i \in \mathcal{N}_u} \frac{1}{\sqrt{|\mathcal{N}_u||\mathcal{N}_i|}} (W_1 e_i^{(k-1)} + W_2 (e_i^{(k-1)} \odot e_u^{(k-1)}))) \\ e_i^{(k)} &= \sigma(W_1 e_i^{(k-1)} + \sum_{u \in \mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_i||\mathcal{N}_u|}} (W_1 e_u^{(k-1)} + W_2 (e_u^{(k-1)} \odot e_i^{(k-1)}))) \end{aligned} \quad (3-8)$$

其中， $\sigma(\cdot)$ 为非线性激活函数。 $e_u^{(k-1)}$ 和 $e_i^{(k-1)}$ 为用户 u 和物品 i 的嵌入表征经过 $k-1$ 次消息传播得到的嵌入表征。 \mathcal{N}_u 为用户 u 交互过的所有物品， \mathcal{N}_i 为物品 i 交互过的所有用户，这些交互关系也可以被视为节点的一阶邻居。 W_1 与 W_2 为特征变换所用到的可训练参数。通过 K 次消息传播，最终可以获得 $K+1$ 条用于描述用户的嵌入表征 $(e_u^{(0)}, e_u^{(1)}, \dots, e_u^{(K)})$ 与用于描述物品的嵌入表征 $(e_i^{(0)}, e_i^{(1)}, \dots, e_i^{(K)})$ 。最终 NGCF 使用拼接（concatenates）的方式，将这些嵌入表征合并，作为最终读出用于推荐的嵌入表征。

NGCF 在很大程度上继承了标准图卷积神经网络的框架，包括使用非线性激活函数 $\sigma(\cdot)$ 和特征变换参数 W_1 和 W_2 。然而，这些组件已经被证明在基于协同过滤的推荐算法中没有起到作用^[27]。对于节点分类、整图分类等任务，每个节点或整个图都拥有丰富的输入特征。例如，在学术引用图中，每个论文节点都可以使用标题和摘要作为输入特征来训练模型^[48]。然而，在基于协同过滤的推荐算法中，用户-物品的二部图只包含用于区分节点的 ID 信息。在这种情况下，使用复杂的非线性激活和特征变换无助于学习更优质的特征，甚至会丢失部分协同过滤信息并增加训练难度。

3.1.2.2 LightGCN 简介

比较于 NGCF, LightGCN 去除了特征变换与非线性激活。LightGCN 在用户-物品上的二部图上的消息传播可以被定义为:

$$\begin{aligned} e_u^k &= \sum_{i \in \mathcal{N}_u} \frac{1}{\sqrt{|\mathcal{N}_u||\mathcal{N}_i|}} e_i^{(k-1)} \\ e_i^k &= \sum_{u \in \mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_i||\mathcal{N}_u|}} e_u^{(k-1)} \end{aligned} \quad (3-9)$$

其中, 用于归一化的 $\frac{1}{\sqrt{|\mathcal{N}_u||\mathcal{N}_i|}}$ 于消息传播在其中得到保留。而 $\sigma(\cdot)$ 非线性激活函数与特征变换 W_1 与 W_2 从 NGCF 的模型被去除。此外, LightGCN 也设计了不同的读出函数, 当用户与物品的初始嵌入表征通过 K 次消息传播得到了描述用户的嵌入表征 $(e_u^{(0)}, e_u^{(1)}, \dots, e_u^{(k)})$ 与用于描述物品的嵌入表征 $(e_i^{(0)}, e_i^{(1)}, \dots, e_i^{(k)})$ 后, 使用简单的加权平均和得到最终用于推荐的嵌入表征。

LightGCN 的优越性在于其简化了 NGCF 中的特征变换和非线性激活函数, 从而减少了模型的参数量, 避免了过拟合现象, 并且保持了模型的高效性。同时, LightGCN 能够高效地利用用户-物品交互矩阵中的信息, 使得该算法适用于大规模的推荐系统场景。因此, LightGCN 已成为基于协同过滤算法的推荐系统领域的一种重要参考模型, 其简洁而有效的设计为推荐系统研究提供了新的思路和方向。

3.2 基于轻量化图卷积的标签感知推荐模型 LFGCF

本节首先提出一种基于轻量化图卷积的标签感知推荐模型 LFGCF, 该模型使用简化后的图神经网络, 在易于训练的同时提高了推荐性能。

3.2.1 模型架构

在目前主流的图神经网络模型中, 如 GCN^[40] 和 GAT^[49]等模型通常是节点、边或整图等表征学习而设计的, 这些图通常都存在附加特征。具体而言, 每个节点都具有输入的嵌入表征, 因此需要一个初始特征变换将原始特征转换为统一的格式。接下来, 节点上的嵌入表征和其他邻居节点的信息会通过一个聚合函数 (aggregation function) 进行聚合, 最终得到聚合后的嵌入表征。最后, 聚合后的嵌入表征通过非线性激活函数更新为新的嵌入表征。然而, 在推荐系统中的二部图中, 每个节点 (包括用户或物品) 通

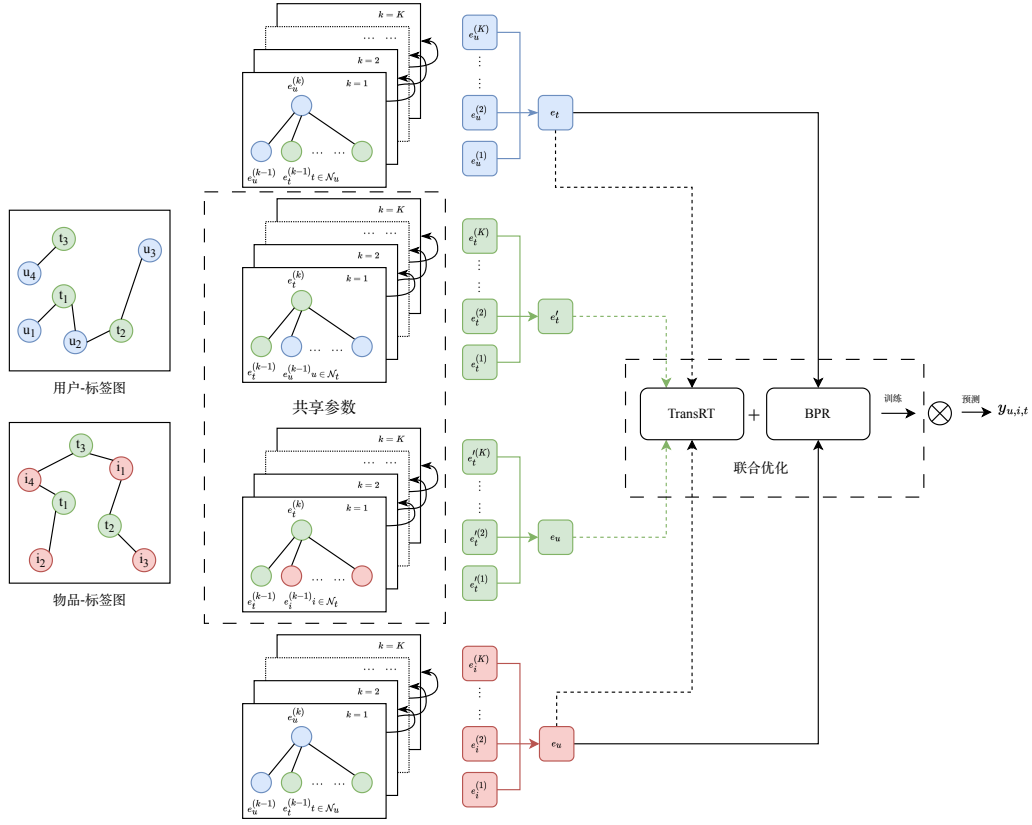


图 3.3 LFGCF 的模型结构

常仅拥有唯一的编号，而没有任何有意义的语义信息作为输入特征。在这种情况下，将初始化后的唯一编号嵌入表征作为输入，通过多层特征变换和非线性激活函数来进行推荐系统任务是现代神经网络的关键^[50]。然而，这样的操作在图数据上会增加表征学习的难度，从而降低推荐任务的性能^[27]。

$$Top(u, K) = \underset{i \in \mathcal{I}}{argmax}^{(K)}(\hat{y}_{u,i}) \quad (3-10)$$

本节提出的 LFGCF 如图 3.3 所示。LFGCF 模型的输入图为社会标注图，且分别对于用户-标签图、物品-标签图进行消息传播操作，由于模型去除了特征变换与非线性激活操作，因此节点的嵌入表征被简单地加权和运算整合。最后，模型使用 BPR 损失函数与基于知识图谱的 TransRT 正则函数联合优化。

3.2.2 消息传播操作

一个基础的图神经网络结构会通过消息传递学习到节点的表征^[40]。新的节点表征将会从其邻居特征中传播而来。这里的聚合函数可以被定义为：

$$\begin{aligned} e_u^{(k+1)} &= \mathcal{AGG}(e_u^{(k)}, e_t^{(k)} : t \in \mathcal{N}_u) \\ e_i^{(k+1)} &= \mathcal{AGG}(e_i^{(k)}, e_t^{(k)} : t \in \mathcal{N}_i) \end{aligned} \quad (3-11)$$

其中, $e_u, e_i \in \mathbf{R}^d$ 分别代表着 G_{UT} or G_{IT} 中的用户和物品节点的表征。此外, \mathcal{AGG} 是邻居聚合函数, 其考虑到目标节点与其 k 阶邻居之间的聚合关系。多跳的邻居聚合函数被用于缓解标签感知推荐系统中的稀疏性^[35]。许多研究为邻居聚合操作提出了不同的邻居聚合函数, 例如在 GIN^[35] 中使用的加权和聚合, 在 GraphSage^[51] 中使用的 LSTM 聚合, 以及 BGCNN^[52] 中的双线性交互聚合。然而大部分研究都依靠带有复杂的特征变换或者非线性激活的聚合函数, 例如 TGCN 中复杂的注意力机制与卷积神经网络设计。尽管这些研究在带有语义特征输入的节点或整图分类任务中取得了非常好的性能, 但是标签感知推荐系统仅仅只将编号过的嵌入表征作为输入特征, 这样复杂的操作对于 Top-K 推荐任务而言是冗余的。

3.2.3 轻量化聚合操作

本小节对图 G_{UT} 和图 G_{IT} 组成的社会标注图进行轻量化后的图神经网络操作。在轻量化的社会标注图卷积神经网络当中, 本文主要关注与图卷积神经网络中对于推荐系统最关键的组件。为了轻量化的设计, 本文使用了简单的加权和聚合操作, 并舍弃了复杂的特征变换以及非线性激活函数。这样的轻量化聚合函数在图 G_{UT} 可以被定义为:

$$\begin{aligned} e_u^{(k+1)} &= \sum_{t \in \mathcal{N}_u} \frac{1}{\sqrt{|\mathcal{N}_u|} \sqrt{|\mathcal{N}_t|}} e_t^{(k)} \\ e_t^{(k+1)} &= \sum_{u \in \mathcal{N}_t} \frac{1}{\sqrt{|\mathcal{N}_t|} \sqrt{|\mathcal{N}_u|}} e_u^{(k)} \end{aligned} \quad (3-12)$$

其中对称的归一化项 $\frac{1}{\sqrt{|\mathcal{N}_u|} \sqrt{|\mathcal{N}_t|}}$ 服从了标准 GCN 的设计, 这是为了避免嵌入表征随着图卷积操作在数量级上的变化。类似于图 G_{UT} 上的操作, 在图 G_{IT} 种的聚合函数可以

被定义为:

$$\begin{aligned} e_i^{(k+1)} &= \sum_{t \in \mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_i|} \sqrt{|\mathcal{N}_t|}} e_t^{(k)} \\ e_t^{(k+1)} &= \sum_{u \in \mathcal{N}_t} \frac{1}{\sqrt{|\mathcal{N}_t|} \sqrt{|\mathcal{N}_u|}} e_u^{(k)} \end{aligned} \quad (3-13)$$

3.2.4 卷积层合并与模型预测

本文提出的 LFGCF 模型的训练参数仅仅只有第 0 层的嵌入表征。一旦嵌入表征被初始化, 后续层便可以用轻量化聚合函数合计算得出。经过 K 层轻量化聚合函数计算后, 本文提出一个层合并函数, 可以进一步合并每一层的嵌入表征并构建出最终的节点表征。层合并函数可以被定义为:

$$e = \text{COMB}(e^{(k)} : k \in K). \quad (3-14)$$

其中, COMB 是层的合并函数, 它使用了特定节点类型的所有层上的表征。 K 是层的数量。不同层的嵌入表征可以捕获到社会标注图中不同的语义信息。例如, 第一层通过用户(物品)和标签之间嵌入表征平滑化, 第二层则会让用户(物品)和标签存在重叠的嵌入表征变得平滑, 更高阶的层可以捕获节点之间更高阶的相似性^[26]。因此, 本文并未过度设计特殊的合并函数操作。本文的层合并函数可以被进一步定义为:

$$e_u = \sum_{k=0}^K a_k e_u^{(k)}, \quad e_i = \sum_{k=0}^K a_k e_i^{(k)}, \quad e_t = \sum_{k=0}^K a_k e_t^{(k)} \quad (3-15)$$

其中, $a_k \geq 0$ 贡献了第 K 层的嵌入表征的对于最终嵌入表征的重要性。这个参数可以被作为一个手动调整的超参数, 或者一个模型参数在模型训练时自动学习。本文将 a_k 统一的设置成 $1/(K+1)$, 这里的 K 代表了层数。本文设计该层合并函数有两个原因: 1) 随着层的提升, 嵌入表征将会变得愈加过平滑^[53], 因此只使用最后一层最为最终嵌入表征是存在问题的; 2) 以不同权重合并不同层的嵌入表征视为是图卷积神经网络中的自连接效应^[27]。

最终模型预测可以被定义成用户和物品最终表征之间的内积, 这里的得分用于生成最终的推荐结果。:

$$\hat{y}_{ui} = e_u^T e_i \quad (3-16)$$

3.2.5 基于知识图谱的嵌入层

基于知识图谱的表示学习是一种常用的方法，用于缓解知识图谱嵌入中的稀疏性，同时有效地将节点参数化为向量表示，保持图谱的拓扑结构。本文提出了一种基于变换思想的新型知识图谱嵌入方法。具体而言，本文提出了一个新的正则化函数 TransRT，该函数基于在知识图谱中广泛使用的 TransR 方法^[54]。具体来说，TransRT 通过优化转换原则 $e_u + e_t \approx e_i$ 来学习每个节点的嵌入。这里， $e_u, e_i, e_t \in \mathbf{R}^d$ 分别是用户 u 、物品 i 和标签 t 的最终嵌入，而 e_u, e_i 代表它们在标签空间的投影表示。因此，对于一条给定的社会标签数据 (u, t, i) ，其可信度分数可以定义如下：

$$g(u, t, i) = \|e_u + e_t - e_i\|_2^2 \quad (3-17)$$

这里，由于 e_u, e_i, e_t 存在于相同的维度空间，但是并不存在于相同的语义空间，因此该分数 $g(u, t, i)$ 越低，说明该大众分类记录越有可能是真实的。

3.2.6 模型预测输出与联合优化

LFGCF 的可训练参数仅仅只有第 0 层的嵌入表征，它合并了社会标注图中的用户、物品和标签与标准的矩阵分解模型的参数复杂度相同。为了获得更好的推荐排序任务，本文使用 BPR 损失函数作为推荐任务的优化函数，该损失函数是一种基于匹配的损失函数，其鼓励被观察到的样本出现的概率要高于没有被观察到，而是被随机采样到的样本。BPR 损失函数可以被定义为：

$$\mathcal{L}_{rec} = \sum_{(u, i, i') \in O} -\ln(\sigma(\hat{y}_{u, i} - \hat{y}_{u, i'})) \quad (3-18)$$

其中， $O = \{(u, i, i') | (u, i) \in \mathcal{A}, (u, i') \notin \mathcal{A}\}$ 表示配对 (u, i) 在社会标注数据中出现过，而配对 (u, i') 意味着用户 u 和物品 i' 在记录中没有出现过，但是由未观察到的配对中随机采样中获得。

为了训练基于知识图谱的嵌入 TransRT，本文以最小化该似然得分为优化方向：

$$\mathcal{L}_T = \alpha ReLU(g_u(u, t_u, i) - g_i(u, t_i, i)). \quad (3-19)$$

这里的 α 控制着知识图谱正则化的强度，并且 $g_u(u, t_u, i)$ 和 $g_i(u, t_i, i)$ 由社会标注数据计算得到。为了更清楚的展示本文设计的意图，本文简单回顾 BPR 损失函数。BPR 损失

通过梯度优化 $\nabla e_u = -\eta(1 - \sigma(\hat{y}_{ui} - \hat{y}_{ui'}))(e_i - e_{i'})$ 。由于大多数推荐系统都服从长尾分布, e_u 被训练后会越来越与流行的物品 i 的嵌入表征 e_i 相似。这意味着 e_u 和 e_i 会分别从 G_{UT} and G_{IT} 中不同的标签表征嵌入聚合而成。本文的设计确保了标签作为连接用户和物品的桥梁, 而不是停留在局部类似的地方, 并且通过 TransRT 平滑的标签嵌入可以缓解模糊性和冗余性。

为了使推荐的学习参数更有效, 并保持大众标注记录之间的正则化关系, 本文使用联合学习框架将 Top-K 推荐任务和 TransRT 结合起来。最后, LFGCF 的总目标函数定义如下:

$$\mathcal{L} = \mathcal{L}_{rec} + \mathcal{L}_T + \gamma \|\Theta\|_2. \quad (3-20)$$

这里, γ 控制着正则化的强度。本文使用 Adam^[55] 优化 \mathcal{L}_{rec} 和 \mathcal{L}_T 并在小批量策略进行训练。最后, 本文使用早停策略来防止模型被过度训练。

3.3 本章小结

本章深入介绍了图神经网络, 并从空间域和矩阵两个时间理解消息传播的本质。接着介绍了 NGCF 和 LightGCN 两个推荐模型, 其中 LightGCN 由于使用了轻量化的图神经网络, 从而对比 NGCF 有着显著的性能提升。之后, 本章提出了 LFGCF。模型由轻量化的社会标注图卷积神经网络与基于知识图谱的标签关系映射两个核心模块组成, 最后使用 Top-K 推荐任务多任务学习优化目标函数, 得到最后的推荐模型。

4 基于对比学习的标签感知推荐算法

本章将介绍模型所使用的对比学习范式，并讨论推荐系统的存在的公平性与数据偏见。之后，本章将提出 TAGCL 的设计架构，讨论了该框架是如何使用多任务学习进行训练并调优。

4.1 对比学习

本小节将简要回顾对比学习，并讨论其最为重要的 InfoNCE 损失函数是如何从 NCE 损失函数发展而来。

4.1.1 对比学习回顾

对比学习 (Contrastive Learning, CL) 是自监督学习 (Self-Supervised Learning, SSL) 中的一种方法，其旨在利用未标记的大量数据来学习知识。图 4.1 展示了对比学习的经典案例^①，以便更好地理解这种学习方法。具体而言，图 4.1a 是人工绘制的一张 1 美元纸币的图像，而图 4.1b 则是一张真实的 1 美元纸币的图像。如果对这两张图像进行标注，我们会认为它们都属于“1 美元”这个类别。然而，如果在像素级别上进行比较，可以明显看到两张图像存在巨大的差异：图 4.1a 不仅丢失了许多边缘细节，而且人脸朝向和“one dollar”的位置等关键信息也完全不同。

在使用监督学习模型的情况下，模型会将这两个数据样本视为独立的样本，并将它们分别输入模型进行学习。然而，由于这两个数据样本属于相同的标签，模型容易感到困惑，使得模型优化损失函数变得不稳定。相比之下，对比学习将这两张图像视为一对样本，并要求模型最大化它们在投影空间中的互信息，以便促进模型捕捉两张图像的相似性。

对比学习通过采用数据增强方法来构造相似和不相似的数据，并要求模型能够区分这些数据，从而使得相似的数据在投影空间中更为接近，而不相似的数据在投影空间中距离较远。对于对比学习模型来说，除了必备的编码器模块 (Encoder) 之外，还需要数据增广模块来生成单条数据样本的不同视图，以及对比学习模块，该模块通常采用噪声对比估计 (Noise Contrastive Estimation, NCE) 作为目标函数，以最大化同一数据样本不同视图之间的一致性以及不同样本的视图之间的差异性。本小节将以语言模型为例，

^① 摘录自：<https://aeon.co/essays/your-brain-does-not-process-information-and-it-is-not-a-computer>

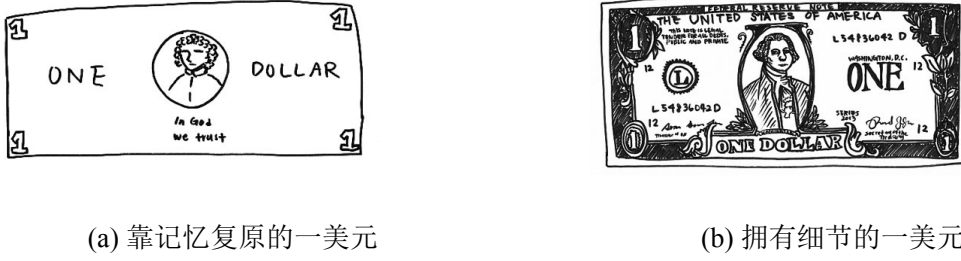


图 4.1 一个经典的对比学习例子

深入介绍噪声对比估计，为后续内容提供理论基础。

4.1.2 噪声对比估计

本小节将以语言模型为例，深入介绍噪声对比估计。在语言模型中，给定一个序列，模型的目标是预测该序列的下一个单词。由于该模型在对参数进行极大似然估计时，需要存在较高的计算复杂度，因此噪声对比估计最初为了简化计算而被提出。

4.1.2.1 语言模型

语言模型 (language model) 是一种用于计算句子或文本的概率模型。它假设一门语言的所有可能的单词组合服从一定的概率分布。若将句子 s 看成是单词 w 的序列 $s = \{w_1, \dots, w_m\}$ ，那么语言模型就是构建一个参数为这些单词的概率模型预测句子成立的概率。同时，可以中间任意一个单词 w_i 以外其余的单词视为第 i 个单词的上下文 c_i 。那么，这个语言模型以链式法则展开后，也可以以条件概率的形式表示：

$$\begin{aligned}
 p(w_1, w_2, \dots, w_m) &= p(w_1) \times p(w_2|w_1) \times p(w_3|w_1, w_2) \times \dots \times p(w_m|w_1, \dots, w_m) \\
 &= \prod_{i=1}^m p(w_i|w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_m) \\
 &= \prod_{i=1}^m p(w_i|c_i),
 \end{aligned} \tag{4-1}$$

由上式可知，语言模型是条件概率 $p(w|c)$ 的集合。然而，直接计算每个单词 w 在整体语料库中的条件概率需要很大计算量。因此，在真实世界运用的统计学语言模型通常会引入马尔可夫假设，即“一个词出现的概率只取决于它之前有限的 n 个词有关”，这便是 n -gram 模型，它将公式 4-1 简化为：

$$p(w_1, w_2, \dots, w_m) = \prod_{i=1}^m p(w_i|w_{i-n+1}, \dots, w_{i-1}) \tag{4-2}$$

在建模统计语言模型时，利用极大似然估计，根据公式 4-1，可以得到其似然函数的对数形式：

$$\mathcal{L}_{MLE} = \sum_{w_i \in S} \log p_{\theta}(w_i | c_i) \quad (4-3)$$

这样，当最大化这个似然函数 \mathcal{L}_{MLE} 时，实际上就是将 $p(w|c)$ 视为 w 和 c 的函数， θ 为待定的模型参数：

$$p_{\theta}(w|c) = F(w, c; \theta) \quad (4-4)$$

当模型通过数据训练获得到最优参数 θ^* 时，函数 F 将会被确定，那么任意的 $p(w|c)$ 都快可以使用这个模型 $F(w, c; \theta^*)$ 计算得出。

由于神经网络能够拟合任何函数，因此被引入到统计语言模型中。Bengio 等人^[56]提出了神经概率语言模型（Neural Probabilistic Language Model, NPLM）。其可以在包含任意大小的上下文中建模单词 w 出现的概率。具体而言，它将语言模型视为一个多分类问题。若将单词库记为 $V = \{v_1, v_2, \dots, v_{|V|}\}$ ，将 (w, c) 作为一对训练样本，在通过神经网络和 SoftMax 函数归一化后，可以输出一个向量 $\hat{y} = [\hat{y}_{i,1}, \hat{y}_{i,2}, \dots, \hat{y}_{i,|V|}]$ ，其中每一维 $\hat{y}_{i,j}$ 表示当上下文为 c_i 时，第 i 个单词为单词库中第 j 个单词 v_j 的概率。训练时会迫使模型将最后单词库中概率最大的单词作为训练样本对中的 w_i 。这样，通过一定的训练后，给神经网络一个上下文 c'_i ，神经网络就可以端到端的预测下一个单词 w'_i 。

假设输入到 SoftMax 函数前的结果为 $s_{\theta}(w, c)$ 。实际上该输出可以表示单词 w 在这个上下文 c 中的匹配程度。那么单词 w 出现的条件概率可以表示为：

$$\begin{aligned} p_{\theta}(w|c) &= \frac{\exp(s_{\theta})(w, c)}{\sum_{w' \in V} \exp(s_{\theta})(w', c)} \\ &= \frac{u_{\theta}(w, c)}{Z(c)} \end{aligned} \quad (4-5)$$

其中， $u_{\theta}(w, c) = \exp(s_{\theta})(w, c)$ 表示得到单词 w 的概率。令 $Z(c) = \sum_{w' \in V} \exp(s_{\theta})(w', c)$ 表示当前单词库内所有单词的概率累和。由于单词库的数量一般十分巨大，因此计算 $Z(c)$ 是非常昂贵、耗时的一件事。在语言模型引入噪声对比估计则为了解决这个问题。

4.1.2.2 NCE 损失函数

噪声对比估计的核心思想是通过学习数据分布样本和噪声样本之间的区别^[57]，以此发现数据中的一些特性。具体而言，噪声对比估计将问题从多分类问题转化为二分类

问题。模型只需要可以区分数据样本和噪声样本就可以学习到最优参数 θ^* 。假设特定上下文 c 的数据分布为 $\tilde{p}(w|c)$ ，从该概率取出的样本为正样本，记类别为 $D = 1$ ，而另一个与 c 无关的噪声分布为 $q(w)$ ，从该概率取出的样本为负样本，记类别为 $D = 0$ 。设从中取出 k_d 个正样本和 k_n 个负样本，再将这些正负样本混合成一个分布 $p(w|c)$ 。可以分别得到以下概率：

$$\begin{aligned} p(D = 1) &= \frac{k_d}{k_d + k_n} \\ p(D = 0) &= \frac{k_n}{k_d + k_n} \\ p(w|D = 1, c) &= \tilde{p}(w|c) \\ p(w|D = 0, c) &= q(w) \end{aligned} \quad (4-6)$$

根据这些概率表达式，且记负样本和正样本之间的比例 $k_n/k_d = k$ ，可以计算后验概率：

$$\begin{aligned} p(D = 0|w, c) &= \frac{k \times q(w)}{\tilde{p}(w|c) + k \times q(w)} \\ p(D = 1|w, c) &= \frac{\tilde{p}(w|c)}{\tilde{p}(w|c) + k \times q(w)} \end{aligned} \quad (4-7)$$

根据噪声对比估计的假设，将 $Z(c)$ 作为一个参数 z_c 进行估计，由公式 4-5 可得：

$$\begin{aligned} p_\theta(D = 0|w, c) &= \frac{k \times q(w)}{u_\theta(w, c) + k \times q(w)} \\ p_\theta(D = 1|w, c) &= \frac{\tilde{p}(w|c)}{u_\theta(w, c) + k \times q(w)} \end{aligned} \quad (4-8)$$

假设取出的样本 D_t 服从伯努利分布，其用于优化模型的损失函数则为：

$$\begin{aligned} \mathcal{L}_{NCE} &= \sum_{t=1}^{k_d+k_n} [D_t \log p(w|D = 1, c) + (1 - D_t) \log p(w|D = 0, c)] \\ &= \sum_{t=1}^{k_d} \log \frac{u_\theta(w, c)}{u_\theta(w, c) + k \times q(w)} + \sum_{t=1}^{k_n} \log \frac{k \times q(w)}{u_\theta(w, c) + k \times q(w)} \end{aligned} \quad (4-9)$$

而噪声对比估计还需要在公式 4-9 上除以正样本数量，当数量很大时，便可以得到噪声对比估计的损失函数：

$$J_{NCE}^c = \mathbf{E}_{w \sim \tilde{p}(w|c)} \log \frac{u_\theta(w, c)}{u_\theta(w, c) + k \times q(w)} + k \mathbf{E}_{w \sim q(w)} \log \frac{k \times q(w)}{u_\theta(w, c) + k \times q(w)} \quad (4-10)$$

因此，本文可以得到 NCE 损失函数。对于使用噪声对比估计训练一个语言模型，从上下文 c 中取出单词作为正样本，从噪声分布中取出单词作为负样本，当正负样本比为 $1:k$ 时，通过训练一个二分类器，从而可以在少样本的前提下完成训练。如果在取正样本数量为 1 时，那么噪声对比估计的目标函数则等价于交叉熵损失函数。

4.1.2.3 InfoNCE 损失函数

本小节将继续介绍 InfoNCE 损失函数是如何从 NCE 损失函数推动得出。InfoNCE 损失函数将互信息引入 NCE 损失函数当中^[58]。继续以语言模型为例，当构建预测任务模型时，通过最大化当前上下文 c_i 和未来的数据 w_{i+k} 之间的互信息来辅助模型训练：

$$I(w_{i+k}; c_i) = \sum_{x, c} p(w_{i+k}, c_t) \log \frac{p(w_{i+k}|c_i)}{p(w_{i+k})} \quad (4-11)$$

若要最大化公式 4-11 中的互信息 $I(w_{i+k}; c_i)$ ，需要最大化 $\frac{p(w_{i+k}|c_i)}{p(w_{i+k})}$ 。其中分子 $p(w_{i+k}|c_i)$ 表示在该上下文 c_i 下获得目标单词 w_{i+k} 的概率，该概率相当于最终想要获得的目标分布。分母 $p(w_{i+k})$ 则表示在任意情况下，获得单词 w_{i+k} 的概率，因此被视为噪声分布。根据公式 4-5 可以获得根据 c_i 预测 w_{i+k} 的表达式：

$$p(w_{i+k}|c_i) = \frac{f_k(w_{i+k}, c_t)}{\sum_{w_j \in X} f_k(w_j, c_t)} \quad (4-12)$$

最终，得到以交叉熵形式的 InfoNCE 损失函数：

$$\mathcal{L}_N = -\mathbf{E}_X [\log \frac{f_k(w_{i+k}, c_t)}{\sum_{w_j \in X} f_k(w_j, c_t)}] \quad (4-13)$$

此外，还可以简单的推导得出，最小化 InfoNCE 损失函数，等价于最大化前上下文 c_i 和未来的数据 w_{i+k} 之间的互信息：

$$\begin{aligned} \mathcal{L}_N^{opt} &= -\mathbf{E}_X \log \left[\frac{\frac{p(w_{i+k}|c_i)}{p(w_{i+k})}}{\frac{p(w_{i+k}|c_i)}{p(w_{i+k})} + \sum_{w_j \in X_{neg}} \frac{p(w_j|c_i)}{p(w_j)}} \right] \\ &= \mathbf{E}_X \log \left[1 + \frac{p(w_{i+k})}{p(w_{i+k}|c_i)} \sum_{w_j \in X_{neg}} \frac{p(w_j|c_i)}{p(w_j)} \right] \\ &\approx -\mathbf{E}_X \log \left[1 + \frac{p(w_{i+k})}{p(w_{i+k}|c_i)} (N-1) \mathbf{E}_{x_j} \frac{p(w_j|c_i)}{p(w_j)} \right] \\ &= \mathbf{E}_X \log \left[1 + \frac{p(w_{i+k})}{p(w_{i+k}|c_i)} (N-1) \right] \\ &\geq \mathbf{E}_X \log \left[\frac{p(w_{i+k})}{p(w_{i+k}|c_i)} N \right] \\ &= -I(w_{i+k}; c_i) + \log(N) \end{aligned} \quad (4-14)$$

通过以上讨论，可以看出噪声对比估计是如何结合互信息以得到 InfoNCE 损失函数的。目前 InfoNCE 主要作为自监督学习中的一个对比学习损失函数，本文也依靠 InfoNCE 发掘难例的能力以得到更好的推荐效果，并克服推荐系统中常见的流行性偏差。

4.1.3 推荐算法的优化方向

现有的推荐算法通常将其问题转化为有监督学习任务，以此来训练模型，但是这种方法并不总是能够取得令人满意的效果。逐点型损失函数（point-aware）是早期推荐算法研究中使用的一种方式。在有监督学习的设定下，通过模型拟合用户对物品的打分来优化模型，它鼓励模型的预测值去拟合用户对物品打分的真实值来估计用户的偏好。逐点型损失函数的缺点是忽略了用户之间的相似性，这可能会导致模型的泛化性能下降。为了解决这个问题，成对型损失函数（pair-aware）被引入推荐算法中，这类损失函数将训练数据组织成用户、正样本物品和负样本物品构成的三元组，然后通过模型使得用户对正样本的得分高于负样本。它优化的是观测到的交互矩阵与重构的交互矩阵之间的一致性。然而，随着推荐算法的发展，研究者们发现这些传统的损失函数在面对长尾数据时，效果并不理想。为了解决这个问题，软最大化损失函数（SoftMax）被提出作为推荐算法未来的优化方向。与成对型损失函数不同，软最大化损失函数利用更多负样本物品，而不是只选取一部分负样本，以此来提高模型的泛化性能。并且理论证明优化 SoftMax 损失函数等价于最大化评估排序的 NDCG^[59]。但目前由于该损失函数的计算复杂度较高，在实际的个性化推荐中并不常用。

本文延续使用成对型损失函数，并在多任务学习框架下，应用结合了噪声对比估计的 SoftMax 函数对模型进行优化，从而兼顾推荐性能与长尾数据的挖掘能力。

4.2 推荐系统的偏差

推荐系统是一种广泛应用于在线信息推荐和个性化服务的算法，它通过对用户的历史行为和偏好进行分析，为用户提供个性化的信息推荐。然而，由于各种原因，推荐系统中可能存在各种偏差。本节将介绍推荐系统中存在的各类偏见，尤其是以流行性偏差为主为推荐系统带来的挑战。

4.2.1 推荐系统中存在的偏差

推荐系统中的数据反馈回路可以被抽象为三个关键组件：用户、模型和数据。该反馈回路通常包括三个阶段^[9]：1）数据的收集（Collection）阶段，该阶段通过用户收集数据；2）学习（Learning）阶段，该阶段通过数据训练模型；3）服务（Serving）阶段，该

阶段通过训练得到的模型为用户提供合适的推荐内容。通过以上三个阶段的反馈流程，用户和推荐系统处在一个互相演进的动态平衡中，用户的个人兴趣和行为将会通过推荐得到更新，而推荐系统可以通过更新的用户数据训练新的模型，系统从而在这个回路中不断更新，并自我完善。

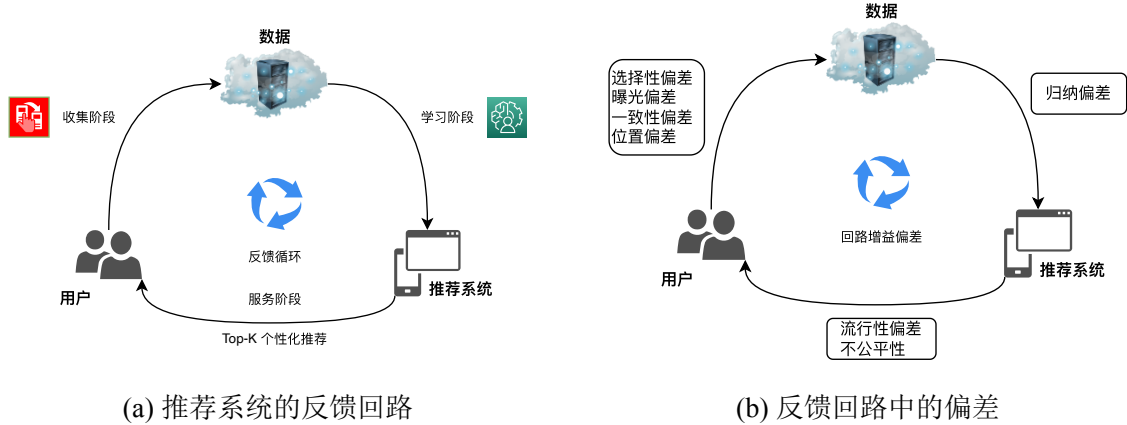


图 4.2 推荐系统中的反馈回路与偏差

通常而言，推荐系统在收集、学习、服务三个阶段中通常存在八类偏差^[9]。其中，与用户相关的偏差（bias in user）有选择性偏差（selection bias）、曝光偏差（exposure bias）、一致性偏差（conformity bias）、位置偏差（position bias）；与数据有关的偏差（bias in data）：流行性偏差（popularity bias）、不公平性（unfairness）；与模型有关的偏差（bias in model）有归纳偏差（inductive bias）；最后，在整个数据反馈回路中还存在回路增益偏差（bias amplification in loop）。本节将逐个介绍这些偏差。

(1) 选择性偏差。选择性偏差通常出现在用户可以自由地为物品打分时。由于被系统采集到的打分并不能代替所有用户对该物品的打分，因此这里的打分数据通常是非随机性缺失（Missing Not at Random, MNAR）的。因此，系统可能会在数据中忽略某些用户的真实评价，进而导致推荐偏差；

(2) 曝光偏差。曝光偏差通常出现在用户只能接触到一部分特定物品，而其余部分被默认的视为用户不喜欢的物品。这样的偏差可能导致用户错失一些可能符合其偏好的物品，进而降低了推荐系统的准确性；

(3) 一致性偏差。一致性偏差通常出现在大多数用户倾向于与自己所属于的群体中其他人相似的行为，即使这样做违背了用户自己的偏好。这意味着反馈并不总是标志着

用户的真实偏好。这种偏差可能导致用户无法得到真正符合自己需求的推荐结果；

(4) 位置偏差。位置偏差通常发生在用户倾向于与推荐列表中位置靠前的物品进行交互，而不考虑物品实际的相关性。这可能导致一些物品始终处于推荐列表的前列，而其他物品则很难得到用户的关注；

(5) 归纳偏差。归纳偏差是指模型在训练过程中，为了更好的被优化。引入某些假设，从而作出超越训练数据进行归纳。这种偏差可能导致模型出现过拟合，使得其在真实场景中的表现不佳；

(6) 流行性偏差。流行性偏差是推荐系统中最常见的偏差之一，其具体含义是当某个物品被推荐的频率高于其受欢迎程度时，就会出现流行性偏差。这种偏差可能导致推荐系统的推荐结果过于倾向于那些已经广受欢迎的物品，而忽略了那些相对不太流行但仍然有一定用户喜好的物品，进而降低了系统的多样性；

(7) 不公平性。不公平性是推荐系统中另一个重要的问题。这种偏差通常源于交互数据中的歧视行为，可能导致系统有意或无意地歧视某些个人或群体，并偏向于其他个人或群体。例如，一个基于性别的推荐系统可能会倾向于向女性用户推荐化妆品或服装，而向男性用户推荐电子产品或运动装备，从而对某些用户造成不公平的对待；

(8) 回路增益偏差。回路增益偏差指出，推荐系统中的偏差在数据反馈回路中将被放大，从而进一步影响推荐结果。其中，与用户相关的偏差会影响用户未来的行为，导致数据中存在更多的偏差；与数据相关的偏差会放大数据的不平衡性，导致用户收到的推荐结果更加偏差。此外，模型中存在的偏差会对整个模型的学习产生系统性影响，进而在收集数据和推荐结果中引入更多的偏差。因此，应当采取措施来减少或消除这些偏差，以确保推荐结果的公平性和多样性。

4.2.2 推荐系统的流行性偏差

本文重点探讨推荐系统中流行性偏差现象的解决方法。在推荐系统中，长尾现象非常普遍：大部分用户会与少量热门物品进行交互。当模型在这种长尾数据上进行训练时，通常会高估受欢迎物品的分数，将不受欢迎的物品预测为负面。由于训练数据的分布不均，头部物品会更频繁地出现在训练数据中，对模型参数更新的贡献更大，从而导致严重的流行度偏差问题。因此，受欢迎的项目被推荐的频率甚至超过了它们在数据集

中的真实受欢迎程度。忽视流行偏向会带来许多问题^[60]。1) 流行偏向会降低个性化推荐的水平, 削弱了推荐系统的多样性。由于不同用户的偏好是不同的, 总是推荐受欢迎的项目会伤害用户体验, 特别是那些喜欢小众项目的用户。2) 流行偏向会降低推荐结果的公平性。流行的物品并不总是高质量的。过度推荐热门项目会降低其他项目的曝光率, 即使它们是很好的匹配, 这是不公平的。3) 流行偏见会进一步增加流行物品的曝光机会, 使得流行物品变得更加流行—这会导致未来训练数据变得更加不平衡, 引发所谓的“马太效应”问题。

4.3 基于图对比学习的标签感知推荐模型 TAGCL

本节首先提出一种基于图对比学习的模型 TAGCL, 该模型使用对比学习以克服数据中存在的流行度偏差, 获得了优秀的推荐性能。

4.3.1 模型架构

TAGCL 延续使用了 LFGCF 结合用户-标签图与物品-标签图的设计, 如图 4.3。其中用户-标签图被用于捕获用户的偏好, 而物品-标签图被用于获取物品的特质。在这种情况下, 用户和项目并没有直接联系。虽然这可能有一些反直觉, 但这有助于帮助简化图结构和训练过程。

与主流的推荐模型相似, 每一个实体都被一个维度为 d 的嵌入特征 (embedding) 描述。 e_u 和 e_i 分别代表着用户和物品的嵌入表征, 同时, 与 LFGCF 不同, TAGCL 进一步区分了用户-标签图与物品-标签图所获得的标签嵌入特征, $e_{t(u)}$ ($e_{t(i)}$) 为用户-标签 (物品-标签) 图所获得的嵌入表征。

4.3.2 嵌入表征的数据增强

近年来, 由于对比学习在计算机视觉和自然语言处理任务中表现出从大量无标签数据中提取有效特征的潜力, 引起了广泛的研究兴趣^[61-64]。对比学习在推荐系统中的应用需要两个关键组件: 数据增强和对比学习任务。本节主要关注如何有效地进行数据增强以创建差异最大化的对比视图。目前有两种主流的图数据增强, Wang 等人^[65]提出了推荐模型 SGL, 图结构使用边丢弃、节点丢弃、随机游走等进行变换, 来创建不同的子图。Yu 等人^[66]提出了推荐模型 SimGCL, 对嵌入表征加入随机噪声, 从而创建不同的视图。

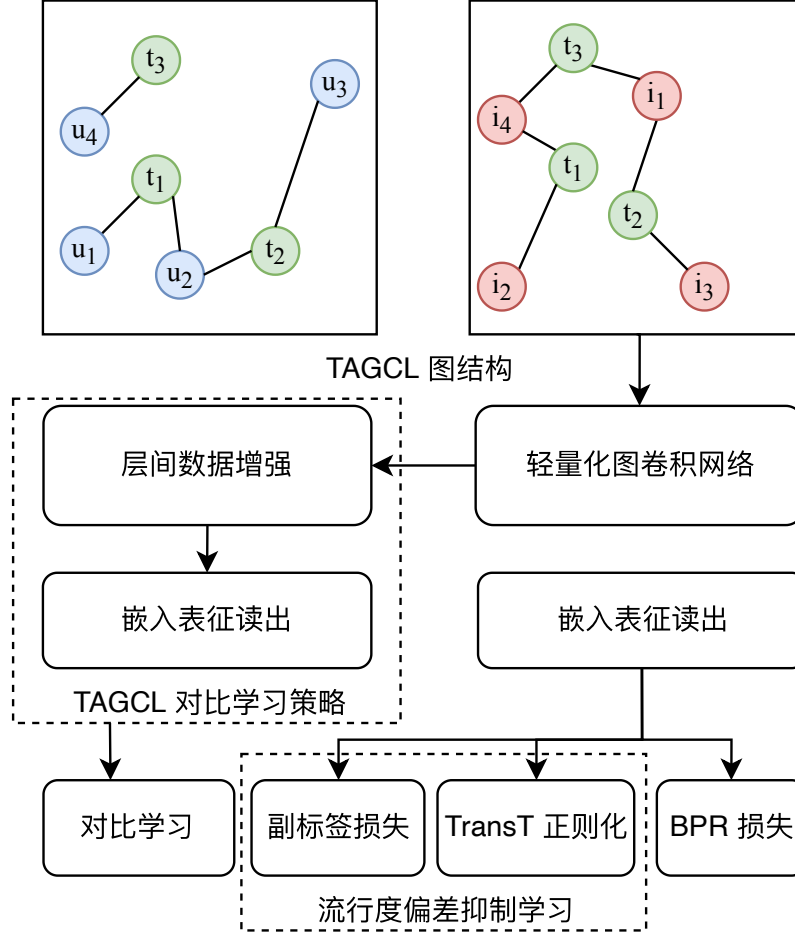


图 4.3 TAGCL 的模型架构

实验证明，SimGCL 添扰动的方式更为简单有效。受到 SimGCL 设计的启发，TAGCL 中数据增强是通过对嵌入表征添加小的扰动来进行的。与应用各种操作修改图结构相比，对嵌入表征进行轻微扰动为每个节点创造了相似的视图。通过增加小的扰动，TAGCL 实现了更好的性能，并提高训练效率。以用户 u 的嵌入表征为例。嵌入表征的数据增强可以被定义为：

$$e'_u = e_u + \epsilon \delta'_u \quad e''_u = e_u + \epsilon \delta''_u \quad (4-15)$$

其中，扰动 δ'/δ'' 是服从正态分布的归一化噪声， ϵ 则控制扰动的强度。最终，噪声使用 SoftMax 函数进行归一化。

4.3.3 多任务学习框架

TAGCL 采用了多任务学习的方式，主要包括推荐任务、对比学习任务 and 流行度抑制学习任务。本文采用了 BPR 作为优化函数。BPR 损失函数通过预测更高的分数来

记录交互记录，对于个性化排名推荐特别有效。其数学表达式为：

$$\mathcal{L}_{rec} = - \sum_{u \in \mathcal{U}} \sum_{i \in N_u} \sum_{i' \notin N_u} \ln \sigma(\hat{y}_{ui} - \hat{y}_{ui'}) + \lambda \|\Theta\|^2 \quad (4-16)$$

其中， λ 控制着 L_2 正则函数的强度。下文将阐述其余学习任务损失函数的概念和设计。

4.3.3.1 对比学习任务

经过数据增强后，每个节点在图中被构建为两种视图。对比学习的任务就是最大化提高同一节点视图的一致性。对比学习损失函数通过调节 SoftMax 函数中的温度参数 τ ，可以对数据中的长尾标签进行不同程度的挖掘。最终，我们使用 InfoNCE^[67] 作为对比学习损失函数，以用户-标签图为例可以被表征为：

$$\begin{aligned} \mathcal{L}_{cl}^{user} &= \sum_{u \in \mathcal{U}} -\ln \frac{\exp(e_u'^T e_u''/\tau)}{\sum_{v \in \mathcal{U}} \exp(e_u'^T e_v''/\tau)} \quad u \neq v \\ \mathcal{L}_{cl}^{tag(u)} &= \sum_{t \in \mathcal{T}} -\ln \frac{\exp(e_{t(u)}'^T e_{t(u)}''/\tau)}{\sum_{s \in \mathcal{T}} \exp(e_{t(u)}'^T e_{s(u)}''/\tau)} \quad t \neq s \\ \mathcal{L}_{cl}^{\mathcal{G}_{UT}} &= \mathcal{L}_{cl}^{user} + \mathcal{L}_{cl}^{tag(u)} \end{aligned} \quad (4-17)$$

其中的超参数 τ 为 SoftMax 中的温度系数。类似的，在物品-标签图也使用了类似的对比学习损失函数 $\mathcal{L}_{cl}^{\mathcal{G}_{IT}}$ 。因此，TAGCL 的完整对比学习损失函数被定义为：

$$\mathcal{L}_{cl} = \mathcal{L}_{cl}^{\mathcal{G}_{UT}} + \mathcal{L}_{cl}^{\mathcal{G}_{IT}} \quad (4-18)$$

4.3.3.2 流行度偏差抑制学习任务

为了平衡性能和推荐系统的公平性，本文提出了两个目标函数：负标签损失和 TransT 正则化。在训练过程中，每条社会标注数据都是形如 $\langle u, t, i \rangle \in \mathcal{A}$ 的三元组。由于推荐系统中只有很少量的正向互动，因此通常需要从未发生交互的样本中随机采样一部分用于模型训练。在 TAGCL 中，我们将物品采样进一步扩展到标签中，因此训练集 \mathcal{D} 可以扩展为五元组 (u, t, t', i, i') 。

$$\mathcal{D} = \{(u, t, t', i, i') | (u, t, i) \in \mathcal{A} \cap (u, t', i) \notin \mathcal{A} \cap (u, t, i') \notin \mathcal{A}\} \quad (4-19)$$

为了在标签领域进行采样，本文首先从那些没有被任何用户分配的标签中进行采样。受到标签与物品之间交互分布差异的启发，提出了负标签损失作为一种优化推荐系

统公平性的损失。对于一个五元组 (u, t, t', i, i') ，负标签损失函数定义为：

$$\mathcal{L}_{nt} = - \sum_{(u, t, t', i, i') \in \mathcal{D}} \ln[\sigma(e_{t(i)}^T e_i - e_{t'(i')}^T e_{i'})] \quad (4-20)$$

其中 σ 是激活函数。由于标签的选择集中于少数标签，因此标签和物品之间的关系相对不受数据偏差的影响。因此，负标签损失将物品和标签的嵌入向量的内积作为输入，最大化正配对和负配对之间的乘积差。更高的内积值意味着标签-物品对之间的相关性更高。

本节进一步提出了一种新的 TransT 正则化函数，旨在以一种公平的方式推进推荐系统。由于负标签损失只考虑物品-标签图 \mathcal{G}_{IT} 中标签的嵌入表征 $e_{t(i)}$ ，而摒弃了用户-标签图 \mathcal{G}_{UT} 中的嵌入表征。不同于 TransTag^[35] 和 LFGCF 中提出的 TransRT。本节将 \mathcal{G}_{UT} 和 \mathcal{G}_{IT} 之间，标签嵌入的差异（即 $e_{t(u)} - e_{t(i)}$ ），视为沟通用户和物品之间的桥梁。具体来说，在记录 $(u, t, i) \in \mathcal{A}$ 的情况下，物品 i 的嵌入表征 e_i 应该接近于 $e_u + e_{t(u)} - e_{t(i)}$ 。由于大多数数据遵循长尾分布，当物品 i 受欢迎时， e_u 会朝着 e_i 的方向更新。同时， e_u 和 e_i 是在信息聚合过程中从标签嵌入中聚合出来的，这使得它们局部相似。虽然将 e_t 假定为实体间关系的嵌入与我们的直觉一致，即标签是连接用户和项目的桥梁，但当标签在社会标注图结构中用户和项目具有相同的重要性时，这种方法并不实际。本文的独特设计 $e_{t(u)} - e_{t(i)}$ 可以充分利用 e_t ，并揭示出用户和物品之间的实质性关系。最终，TransT 正则化函数可以定义为：

$$\mathcal{L}_T = \sum_{(u, t, i) \in \mathcal{A}} \|e_u + (e_{t(u)} - e_{t(i)}) - e_i\|_2 \quad (4-21)$$

4.3.4 优化目标函数

综合本节提出的 BPR 推荐损失、负标签损失、对比学习损失和 TransT 正则化函数被整合成 TAGCL 的最终优化目标函数：

$$\mathcal{L} = \mathcal{L}_{rec} + \alpha \mathcal{L}_{cl} + \beta \mathcal{L}_{nt} + \gamma \mathcal{L}_T \quad (4-22)$$

其中， α, β 和 γ 分别控制着每一个损失函数的强度。由于 TAGCL 仅有第 0 层的嵌入表征为可训练参数，即 $\Theta = \{e_u^{(0)}, e_{t(u)}^{(0)}, e_{t(i)}^{(0)}, e_i^{(0)}\}$ 。最后模型使用 Adam 优化，并在小批量策略进行。本文使用早停策略来防止模型被过度训练。

4.4 本章小结

本章主要介绍了对比学习及其在推荐系统中的应用。对比学习是一种有效的难例挖掘方法，可以克服推荐数据中存在的流行度偏差。本章将对比学习引入到推荐系统中，并由 NCE 损失函数推导出 InfoNCE 损失函数。相比于传统的 NCE 损失函数，InfoNCE 损失函数能够更好地捕捉物品间的互信息，提升了推荐系统的性能。在本章的研究中，本文提出了一种新的模型设计架构 TAGCL。TAGCL 模型采用了多任务学习的框架，通过同时学习多个任务，提高了模型的泛化能力。该模型采用了深度神经网络的思想，并采用了类似于 TransE 的则化函数 TransT，以促进推荐系统的公平性。

5 标签感知推荐的算法实验评估

推荐系统是一种常见的信息过滤技术，其评估通常采用公开数据集进行实验，并比较不同模型的实验性能。本章首先将介绍标签感知推荐系统中常用的数据集，以及用于评估的指标。在给出基线模型的对比之后，本章对 LFGCF 和 TAGCL 模型的实验结果分别进行分析。此外，在真实的推荐系统数据上评估了 TAGCL 模型的性能。

5.1 数据集

本节简要介绍了研究所使用到的数据集。除去三个常用的评估数据意外，本文还在一个真实运行的标签系统数据上评估实验，以佐证模型的有效性。

本文基于三个公开的数据集 MovieLens、Last.FM 和 Delicious 进行实验。这些数据集都在 HetRec2011^[68] 中发布。为了充分验证本文提出的模型 LFGCF 和 TAGCL 在标签感知推荐系统中的表现，本文在一个真实应用的社会化书签和出版物系统 BibSonomy 中进一步验证大规模场景下的模型性能。

(1) **MovieLens** 是一个电影推荐数据集，来源于电影推荐系统 MovieLens^①。平台上的用户可以为电影进行打分并且使用任意标签标注电影。MovieLens 数据集中，每一个用户都有一个喜欢的电影的标注列表，本文将电影作为用户的个性化推荐物品。数据集由用户 ID，电影 ID 和标签 ID 组成；

(2) **Last.fm** 是一个音乐人推荐数据集，来源于著名的电台网站 Last.fm^②。平台上的用户可以搜索、收藏、评论自己喜欢的音乐，并且可以随意为自己喜欢的音乐标注上个性化的标签。MovieLens 数据集中，每一个用户都有一个喜欢的歌手的标注列表，本文将歌手作为用户的个性化推荐物品。数据集由用户 ID，歌手 ID 和标签 ID 组成；

(3) **Delicious** 是一个网页书签推荐数据集，来源于提供网络书签管理服务的社交平台 Del.icio.us^③。平台上的用户可以和其他用户分享、交流网络书签，也可以保持整理自己的私人书签，并鼓励用户为书签进行标注。数据集由用户 ID，书签 ID 和标签 ID 组成；

(4) **BibSonomy** 是一个真实标签系统 BibSonomy^④ 的数据集，以 SQL 的形式提供

① <https://movielens.org/>

② <https://www.last.fm/>

③ <https://del.icio.us/>

④ <https://www.bibsonomy.org/>

给研究人员。平台为用户提供了便捷出版物和网页书签管理方式，并通过社交网络协助寻找潜在的研究方向。数据集可以分为 BibSonomy-BM 与 BibSonomy-BT 两个部分。BibSonomy-BM 由用户 ID，书签 ID 和标签 ID 组成。BibSonomy-BT 由用户 ID，出版物 ID 和标签 ID 组成。

为了降低用户随意标注为数据引入的噪声，本文保持了^[32]相同的设置，对于 Last.Fm 数据集，过滤频次少于 5 次的标签；对于 Delicious 数据集，过滤频次少于 15 次的标签；对于 BibSonomy，过滤频次少于 15 次的标签。经过过滤后的数据如表 5.1：

表 5.1 数据集统计

数据集	用户	物品	标签	标注记录	稀疏率 (%)
Last.FM	1808	12212	2305	175641	99.20%
MovieLens	1651	5381	1586	36728	99.59%
Delicious	1843	65877	3508	330744	99.73%
BibSonomy-BM	5996	576232	8092	1622320	99.95%
BibSonomy-BT	9721	750514	9721	1972556	99.97%

由表 5.1 可以看出，Last.FM、MovieLens 与 Delicious 这类的研究型公开数据集，数据规模较小，而真实的运行的标签系统数据规模更大，并且更为稀疏。

5.2 实验方法

本节将介绍本文进行推荐算法实验的实验方法，首先介绍整体实验设计思路，给出实验评估时使用的评估指标，最后简要介绍需要对比的基线模型。

5.2.1 实验设计

推荐系统的设计和实现涉及到多个领域，如机器学习、信息检索、数据挖掘和人工智能等。在推荐系统研究中，实验是不可或缺的一步，实验结果的准确性和可重复性对研究的质量至关重要。本文的实验在一台服务器上进行，硬件配置如下：2 颗 8 核 E5-2620V4 2.0GHz，DDR4 ECC REG 内存 128G，2400MHz；系统硬盘：1 块 256G SATA SSD；数据硬盘：4TB 企业级硬盘；2 块 GPU 卡（NVIDIA TITAN Xp Pascal）。根据 Dacrema 等人^[69]的调研，同一个模型算法在不同平台上实现可能会导致结果不同，从而

增加了实验重复性的难度。因此，本文选择使用一个基于 PyTorch 实现的推荐系统实验平台 RecBole^[70]，版本为 1.0.1。RecBole 是一个面向研究者的、易于开发和复现的推荐系统实验平台，提供了统一、全面、高效的推荐系统实验环境。使用 RecBole 进行实验，可以保证实验结果的可重复性和准确性。最终实验设计流程如下：

(1) 数据集选择：在本文实验中，本文选取了 HetRec2011 发布的三个数据集作为实验数据，同时在一个真实运行的推荐系统数据上进行实验；

(2) 数据预处理：为了减少冷启动问题和数据噪声对模型训练的影响，我们对数据进行了预处理，其中包括 ID 重映射和低频标注记录过滤等操作；

(3) 数据划分：本文将处理后的数据随机划分为训练集、验证集和测试集，其中训练集、验证集和测试集的比例分别为 6:2:2；

(4) 模型训练：本文采用 RecBole 实验框架下的基线模型、LFGCF 和 TGCN 进行模型训练，直至损失函数收敛；

(5) 模型验证与测试：在验证集上选取效果最优的模型，并在测试集中得到最终推荐结果；

(6) 超参数实验：本文对模型进行超参数实验，遍历探索模型最优超参数，并分析超参数灵敏度；

(7) 实验分析与总结。最后，本文对实验结果进行了分析与总结，得出了一些有价值的结论和启示。

5.2.1.1 评估指标

为了充分验证和对比模型性能，本文选取了四个在个性化排序推荐领域常见的评估指标：召回率 (Recall@K)、准确率 (Precision@K)^[71]、归一化累计折损增益 (Normalized Discounted Cumulative Gain, NDCG@K)^[72]，平均逆排名 (Mean Reciprocal Rank, MRR@K)^[73]，以及一个用于评估推荐系统流行度的指标——平均推荐流行度 (Average Recommendation Popularity, ARP@K)^[6]。这些指标与标签感知推荐系统的性能和最终的 Top-K 推荐列表的质量密切相关，其中除去平均推荐流行度以外的指标数值越高，则表示模型性能越优秀。

召回率和准确率的计算方法源自于评估分类状况的混淆矩阵 (Confusion Matrix)。混淆矩阵是机器学习中用于评估分类器性能的一种工具。它是一个二维的表格，横轴

代表预测值，纵轴代表实际值。每个单元格代表预测结果为横轴类别，实际结果为纵轴类别的样本数。混淆矩阵通过计数真正的阳性（True Positive, TP）、假阴性（False Negative, FN）、假阳性（False Positive, FP）和真正的阴性（True Negative, TN）来评估分类器的准确性。这些信息可以计算出多种分类指标，例如准确率（Accuracy）、召回率（Sensitivity）、等。表5.2为一个二分类的混淆矩阵。

表 5.2 二分类混淆矩阵

实际值	预测值	
	正类	负类
正类	TP	FN
负类	FP	TN

基于混淆矩阵，召回率和准确率可以被定义为：

$$R = \frac{TP}{TP + FN}, \quad (5-1)$$

$$P = \frac{TP}{TP + FP}$$

在此基础上，个性化推荐系统的召回率和准确率根据具体应用场景有着进一步改进。其中召回率衡量的是系统推荐给用户的物品中，用户实际发生交互的物品数量与系统推荐给用户的物品的比值。该指标为推荐系统最为重要的指标，本文在验证集上选取召回率最高的模型作为测试模型。召回率可以被定义为：

$$Recall@K = \frac{|R^N(u) \cap T(u)|}{|T(u)|} \quad (5-2)$$

其中， $T(u)$ 表示系统推荐给用户的物品的数量， $R^N(u)$ 用户实际发生交互的物品数量， $Recall@N$ 中的 K 指个性化推荐列表的长度。

准确率则描述的是是系统推荐给用户的物品中，用户实际发生交互的物品数量与用户实际发生交互物品的比值。准确率可以被定义为：

$$Precision@K = \frac{|R^N(u) \cap T(u)|}{N} \quad (5-3)$$

其中， N 表示用户在系统中实际交互过的物品数量， $R^N(u)$ 用户实际发生交互的物品数量， $Precision@N$ 中的 K 指个性化推荐列表的长度。

归一化累计折损增益是一种评估排名效果的指标，常用于搜索引擎、推荐系统等领域。它通过计算排名中每个项目的得分权重，以评估排名的质量。NDCG 的优点在于它可以评估排名的相对质量，并且可以适用于任意数量的排名项目。它是一种规范化的指标，因此可以在不同的应用场景中进行比较。

$$nDCG@K = \frac{1}{\mathcal{U}} \sum_{u \in \mathcal{U}} \frac{\sum_{n=1}^N \frac{I(R_n^N(u) \in T(u))}{\log(n+1)}}{\sum_{n=1}^N \frac{1}{\log(n+1)}} \quad (5-4)$$

其中 $R_n^N(u)$ 指 Top-K 推荐中的第 n^{th} 项用户发生交互的物品 $R^N(u)$ ， $NDCG@N$ 中的 K 指个性化推荐列表的长度。

平均逆排名是一种常用于信息检索与推荐系统负评估指标，用于评估查询结果的质量。它通过计算正确检索结果在检索列表中的排名来评估系统性能。

$$MRR@K = \frac{1}{\mathcal{U}} \sum_{u \in \mathcal{U}} \frac{1}{rank_u^*} \quad (5-5)$$

其中， $rank_u^*$ 表示对用户的推荐中第一个相关项目的排名位置。

为评估推荐系统中存在的流行度偏差带来的不公平现象，本文额外使用平均推荐流行度是否降低了对热门物品的推荐偏见。其通过计算模型推荐给用户的物品，其在训练数据中流行度的平均值，以评估模型是否对于热门物品更为偏好。

$$AveragePopularity@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{\sum_{i \in R_u} \phi(i)}{|R_u|} \quad (5-6)$$

其中 $\phi(i)$ 指物品 i 在训练数据的被交互次数。与本文其他用到的指标不同，平均推荐流行度越低，代表模型的推荐能力越优秀。

5.2.2 对比模型

为了综合对比评估推荐模型的性能，本文比较了通用的推荐算法模型和标签感知推荐算法模型。其中，通用推荐算法模型采用了基于图神经网络的推荐算法中常用的基准模型 LightGCN^[27]和基于图对比学习的 SimGCL^[66]；标签感知的推荐算法模型则采用了 BPR-T^[17]和基于图神经网络算法的 TGCN^[35]。模型具体介绍如下：

(1) **LightGCN** 是一种广泛使用的基准模型，因其简单性和有效性而受到推荐系统领域的青睐。它采用了一种去除非线性激活函数的简单图神经网络进行表征学习，从而对推荐系统中的用户和物品进行建模；

(2) **SimGCL** 与常见的对图数据进行增强以进行对比学习任务的方法不同，它对嵌入表征进行扰动，以构建对比学习任务。该模型的主干网络采用了 **LightGCN** 的设计，进一步提升了模型的性能；

(3) **BPR-T** 通过改进 **BPR** 损失函数，将用户标注行为引入协同过滤模型中，提高了推荐系统的性能。具体而言，**BPR-T** 将标签信息作为用户行为的一部分来进行建模，从而更准确地预测用户对物品的偏好；

(4) **TGCN** 构建了一个复杂的图卷积模型，利用注意力机制和卷积神经网络增强了模型的特征学习能力，使得标签感知推荐系统第一次运用了图神经网络。该模型能够准确地捕捉用户和物品之间的关系，从而更好地进行推荐。

5.3 实验评估

本节将介绍基线模型、LFGCF 和 TAGCL 在 MovieLens、Last.FM 和 Delicious 数据集上的实验结果，并对 LFGCF 和 TAGCL 进行详细的消融实验，以进一步比较模型设计的优劣。

5.3.1 实验设置

为了保证实验的公平性，本文使用基于 Minibatch 设置下的 Adam 优化器^[55]来训练所有模型，BatchSize 设置为 2048，并将模型的最大训练轮数设置为 500 轮，同时采用早停机制^[74]来避免过拟合的问题。本文使用网格法搜索最佳超参数，其中学习率的搜索范围为 {0.0005、0.001、0.005、0.01}，正则化权重的搜索范围为 {1e-5、1e-4、1e-3、1e-2}。对于所有需要嵌入表征的模型，嵌入表征维度设置为 64，并通过最常用的 Xavier 方式^[75]进行随机初始化。推荐序列长度均设置为 20。其余基线模型的性能按照原始论文的报告进行调整。

5.3.2 性能对比实验分析

最终模型的个性化推荐性能如表 5.3、5.4、5.5 所示。其中，imp. 指提高的百分比，黑体高亮的数字指所有模型中最好的性能，下划线指标签感知推荐系统中最高的性能。结果表明，LFGCF 和 TAGCL 在大部分指标中都优于基线模型。在标签感知推荐的模型中，LFGCF 依靠轻量化设计的图神经网络结构，比起直接使用原始图神经网络的 TGCN，

有着更强的性能。由于模型在更短的训练轮次内就收敛了，所以可以获得更稳健的嵌入表征，从而得到更稳定的表现。同时，LFGCF 也依靠图神经网络的设计，在消息传播过程中获取了丰富的上下文信息，缓解了标签感知推荐系统这一特有问题中的一次多义和多次同义的问题。而 TAGCL 则进一步优化了当数据存在大量偏差时模型的性能，尤其是在 Last.FM 和 MovieLens 这两个数据集中存在大量偏见时。但是在数据集 Delicious 中存在较少偏差时，TAGCL 的性能稍有损失。

对于基线模型而言，通用推荐模型 SimGCL 在大多数指标中都优于 LightGCN。这是因为 SimGCL 使用对比学习改进了 LightGCN 在训练时，损失函数偏向于拟合数据集的结果而不是寻找用户和物品之间的交互规律。而在标签感知推荐算法中，虽然 TGCN 使用了图神经网络这一更新深度学习技术，但并没有针对标签推荐做出优化。因此，在部分指标中，TGCN 和 BPR-T 相比并没有显著的优势。

5.3.2.1 MovieLens 结果分析

表 5.3 为数据集 MovieLens 下的模型对比实验。本表中列出了在 MovieLens 数据集上的六种不同的推荐模型，包括通用推荐模型（LightGCN 和 SimGCL），标签感知推荐模型（BPR-T 和 TGCN），LFGCF 和 TAGCL。从评估指标来看，TAGCL 在五个指标中均取得了最佳的结果。其中，TAGCL 在 Pre. 上的表现最好，达到了 0.0405；在 NDCG 上，TAGCL 取得了 0.2338 的分数，排名第一；在 MRR 上，TAGCL 的得分为 0.2356，排名第二；在 Rec. 和 ARP 指标上，TAGCL 分别排名第二和第一。此外，LFGCF 在 Rec. 和 ARP 指标上的表现相对较差，分别排名第四和第五，但在 Prec. 上表现不错，排名第二。SimGCL 在 Prec. 上的表现相对较差，排名第四，但在 NDCG 和 MRR 上的表现较好，排名第二和第三。BPR-T 在五个指标中排名第三，相对而言表现较为平衡。LightGCN 和 TGCN 在五个指标中排名较中等，表现不如其他模型。

5.3.2.2 Last.FM 结果分析

表 5.4 为数据集 Last.FM 下的模型对比实验。从表格可以看出，TAGCL 在五项指标中表现最好，Rec.、Pre.、NDCG、MRR 和 ARP 分别为 0.5199、0.1611、0.4949、0.5541 和 42.99。其次是 SimGCL，在 Pre.、NDCG、MRR 和 ARP 方面略逊于 TAGCL，但推荐精度最高，达到了 0.5055。LFGCF 在推荐精度和 NDCG 方面略高于 SimGCL，但在其他指标方面稍逊于 SimGCL。BPR-T 和 TGCN 在所有指标中表现最差，其中 BPR-T 在

表 5.3 数据集 MovieLens 下的模型对比实验

指标	通用推荐		标签感知推荐		LFGCF	TAGCL	imp. SOTA	imp. TRS
	LightGCN	SimGCL	BPR-T	TGCN				
Rec.	0.2788	0.2835	0.2826	0.2812	<u>0.2929</u>	0.3180	8.57%	8.57%
Pre.	0.0349	<u>0.0383</u>	0.0365	0.0372	0.0365	0.0405	5.74%	8.87%
NDCG	0.2015	<u>0.2274</u>	0.2209	0.2187	0.2140	0.2338	2.81%	5.84%
MRR	0.2101	0.2383	0.2273	0.2218	0.2183	<u>0.2356</u>	-1.13%	3.65%
ARP	26.78	18.15	22.76	25.25	<u>18.10</u>	14.96	17.07%	17.58%

推荐精度和 NDCG 方面略好于 TGCN，但在其他指标方面略逊于 TGCN。

表 5.4 数据集 Last.FM 下的模型对比实验

指标	通用推荐		标签感知推荐		LFGCF	TAGCL	imp. SOTA	imp. TRS
	LightGCN	SimGCL	BPR-T	TGCN				
Rec.	0.4835	0.5055	0.4714	0.4736	<u>0.5057</u>	0.5199	2.81%	2.81%
Pre.	0.1375	<u>0.1534</u>	0.1363	0.1332	0.1465	0.1611	5.02%	9.97%
NDCG	0.4087	<u>0.4680</u>	0.4321	0.4225	0.4482	0.4949	5.75%	10.42%
MRR	0.4664	<u>0.5263</u>	0.5083	0.4874	0.5033	0.5541	5.28%	9.01%
ARP	111.99	<u>51.67</u>	103.74	78.88	80.65	42.99	16.80%	46.70%

5.3.2.3 Delicious 结果分析

表 5.5 为数据集 Delicious 下的模型对比实验。从表格中可以看出，对于通用推荐任务，LightGCN 和 SimGCL 的表现非常接近，但在 NDCG 和 MRR 两个指标上，LightGCN 略胜一筹。对于标签感知推荐任务，BPR-T 和 TGCN 的表现也很接近，但在 Rec. 和 Pre. 两个指标上，TGCN 的表现略优于 BPR-T。同时，LFGCF 和 TAGCL 这两种改进模型相比通用推荐和标签感知推荐的基准模型都有明显的提升，其中 TAGCL 在所有指标上都取得了最佳的结果。综上所述，本实验中 LFGCF 和 TAGCL 这两种改进模型在 Delicious 数据集上表现最优，可以作为该数据集上推荐系统的首选模型。

表 5.5 数据集 Delicious 下的模型对比实验

指标	通用推荐		标签感知推荐		LFGCF	TAGCL	imp. SOTA	imp. TRS
	LightGCN	SimGCL	BPR-T	TGCN				
Rec.	0.3337	<u>0.3351</u>	0.3150	0.3284	0.3300	0.3432	2.42%	4.00%
Pre.	0.3525	<u>0.3554</u>	0.3409	0.3519	0.3498	0.3705	4.25%	5.29%
NDCG	<u>0.4213</u>	0.4177	0.3984	0.4160	0.4080	0.4385	4.08%	5.41%
MRR	<u>0.5786</u>	0.5529	0.5373	0.5589	0.5395	0.5828	0.73%	4.28%
ARP	3.11	<u>4.67</u>	6.32	6.21	4.69	5.61	-80.39%	-19.62%

5.3.3 LFGCF 模型评估

本节在 Last.FM 与 Delicious 数据集上，对 LFGCF 进行消融实验以进一步理解各个组件是如何影响其性能的。表 5.6 展示了 LFGCF 与其变体的性能比较。其中，LFGCF-L 指未使用轻量化图神经网络的变体，LFGCF-T 指未使用知识图谱的 TransRT 正则化的变体。

模型	Recall@10	Recall@20	模型	Recall@10	Recall@20
LFGCF-L	0.4208	0.4958	LFGCF-L	0.1615	0.2838
LFGCF-T	0.4336	0.5027	LFGCF-T	0.1903	0.3270
LFGCF	0.4362	0.5132	LFGCF	0.1955	0.3286

(a) 数据集 Last.FM 下的 LFGCF 消融实验

(b) 数据集 Delicious 下的 LFGCF 消融实验

表 5.6 LFGCF 消融实验

5.3.3.1 轻量化图神经网络设计的影响

在 LFGCF 模型中，本文使用了轻量化的图神经网络以获取大众标注图中的潜在特征。相较于普通的图神经网络，我们去除了非线性激活函数以及特征变换。这样的设计为 LFGCF 带来了显著的性能提升。为了验证该设计的有效性，我们使用了以 NGCF 为主干网络的模型 LFGCF-L 作为比较。表 5.6a 可见，在数据集 Last.FM 下，LFGCF-L 是未使用轻量化图神经网络的变体，它的 Recall@10 和 Recall@20 分别为 0.4208 和 0.4958，相对较低。表 5.6b 可见，在数据集 Delicious 下，LFGCF 在数据集 Delicious 中的性能更

加显著，其 Recall@10 和 Recall@20 分别为 0.1615 和 0.2838。这表明在 Delicious 数据集上，轻量化图神经网络的设计对于学习用户和物品自身的嵌入表征是非常有效的。这意味着 LFGCF 的轻量化设计可以有效地获取大众标注图中的潜在特征。

5.3.3.2 TransRT 设计的影响

在 LFGCF 模型中，本文使用了基于知识图谱的 TransRT，在优化时发挥作用。其通过标签，将用户和物品统一到同一向量空间的设计为 LFGCF 解决了标签感知推荐系统中的一次多义、多词同义问题。我们将去除了 TransRT 这一组件的 LFGCF 记为 LFGCF-T。表 5.6a 可见，在数据集 Last.FM 下，LFGCF 取得了最好的 Recall@10 和 Recall@20 分数，分别为 0.4362 和 0.5132。表 5.6b 可见，在数据集 Delicious 下，LFGCF 的性能略高于 LFGCF-T，说明 TransRT 正则化对于解决标签感知推荐系统中的多义、多词同义问题是非常有效的。总体而言，在两个数据集上，LFGCF 的性能最好，LFGCF-T 的性能略低于 LFGCF。这表明使用基于知识图谱的 TransRT 正则化可以提高 LFGCF 的性能。虽然 LFGCF 的性能略高于 LFGCF-T，但二者之间的性能差异很小，表明 LFGCF 的轻量化图神经网络设计已经非常有效。

5.3.4 TAGCL 模型评估

本节在 Last.FM 与 Delicious 数据集上，对 TAGCL 进行消融实验以进一步理解各个组件是如何影响其性能的。表 5.7 展示了 TAGCL 与其变体的性能比较。其中，TAGCL-A 等价于 LFGCF-T，是一个仅包含基本图结构的基线模型，TAGCL-CL 去掉了框架中的对比学习任务，TAGCL-NT 去掉了框架中的标签副采样过程，TAGCL-T 则去掉了基于知识图谱的 TransT 正则化。

在 Last.FM 数据集上的实验结果在表 5.7a 展示，TAGCL-NT 取得了最好的 Rec.@20 得分 (0.5212)，但是在 ARP.@20 得分方面，TAGCL 取得了最好的成绩 (42.99)。这意味着 TAGCL 在前 20 个推荐结果中的平均排名位置最佳。TAGCL-CL 在两个指标上都表现不如 TAGCL 和 TAGCL-NT，这表明对比学习任务对模型性能的贡献很大。

在 Delicious 数据集上的实验结果在表 5.7b 展示，TAGCL 取得了最好的 Rec.@20 得分 (0.3432)，但是在 ARP.@20 得分方面，TAGCL-CL 取得了最好的成绩 (5.24)。这表明在不同的数据集上，TAGCL 模型的变体会有不同的表现，但总体而言，TAGCL 在

模型	Rec.@20	ARP.@20	模型	Rec.@20	ARP.@20
TAGCL-A	0.4502	92.05	TAGCL-A	0.3188	5.33
TAGCL-CL	0.4141	89.42	TAGCL-CL	0.3301	5.24
TAGCL-NT	0.5212	43.92	TAGCL-NT	0.3425	5.48
TAGCL-T	0.5193	52.17	TAGCL-T	0.3396	5.58
TAGCL	0.5199	42.99	TAGCL	0.3432	5.61

(a) 数据集 Last.FM 下的 TAGCL 消融实验

(b) 数据集 Delicious 下的 TAGCL 消融实验

表 5.7 TAGCL 消融实验

Rec.@20 和 ARP.@20 上都取得了最佳的成绩，表明该模型在推荐任务中表现优异。

5.3.4.1 对比学习的影响

在 TAGCL 框架中，本文采用一种简单的方式实现对比学习。与常见的对比学习实现思路不同，本文没有使用复杂的图数据增强，而是通过对训练过程中的嵌入表征进行噪声扰动来构建对比学习任务。在 InfoNCE 损失函数的帮助下，TAGCL 取得了优秀的推荐性能。为了验证本文使用的特征增强方式与 InfoNCE 损失函数的有效性，我们首先移除了 TAGCL 中的对比学习任务，得到了 TAGCL-CL。同时保持 TAGCL-CL 的超参数与 TAGCL 相同，我们对 TAGCL 模型与其变体 TAGCL-CL 进行了实验，结果如表 5.7 所示。从数据可以看出，由于对比学习任务的加入，TAGCL 模型的性能会显著提高。相比于 TAGCL-A，一个包含 TAGCL 基本图结构的基线模型，TAGCL 的表现更加优秀。因此，对比学习任务对 TAGCL 模型的性能有着巨大贡献。

5.3.4.2 负标签采样的影响

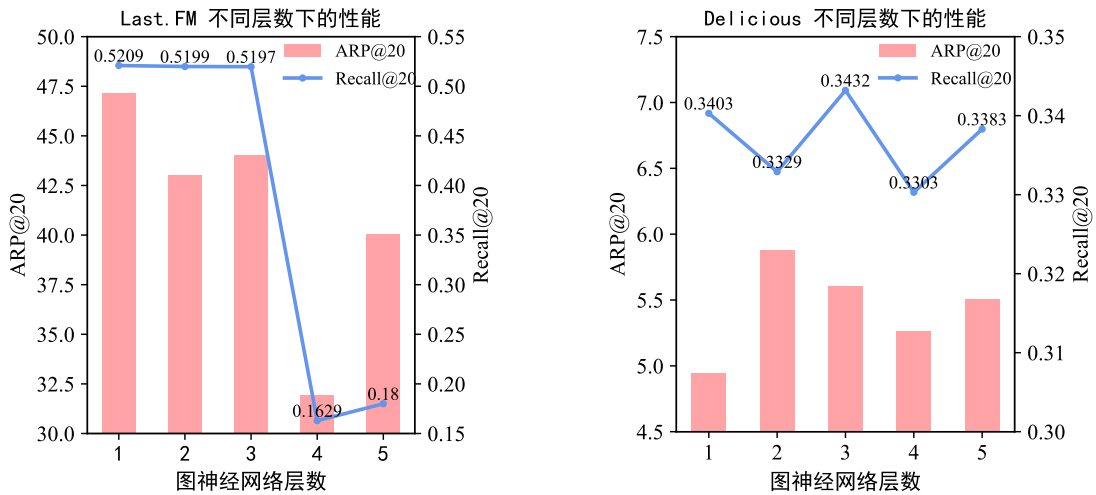
本文提出了一个新的视角，探讨了标签在标签感知推荐系统中的作用。在训练过程中，通过对标签域和物品域进行负采样，并利用物品-标签对配对损失来帮助模型从偏差较小的标签信息中学习到物品的嵌入表征。对去负标签采样的模型进行消融实验，将其记为 TAGCL-NT。实验结果显示，TAGCL-NT 略低于 TAGCL，但仍能胜过 TAGCL-CL 和 TAGCL-A。这表明负标签损失在 TAGCL 中的作用相对不那么关键。虽然标签负向采样对性能提升的贡献较小，但在数据量较大的 Delicious 数据集上，它可以大大提高模型的收敛效率，训练过程从 85 轮减少到 67 轮。

另外, 本文引入了 TransT 正则化损失函数, 作为促进推荐系统公平性的另一个组件。基于构建的图结构, TransT 正则化可以在用户和物品之间建立桥梁。通过去除 TransT 正则化损失, 建立 TAGCL-T 进行实验。结果表明 TransT 正则化是有效的, 并且比负标签损失对 TAGCL 的性能提升贡献更大。因此, TAGCL 通过将二元图之间的标签嵌入差异视为连接用户和项目的关系表示, 设法统一不同图上的用户和项目嵌入。

5.3.4.3 超参数的影响

本文对标签感知推荐系统中标签的作用进行了探究, 并对 TAGCL 方法的参数进行了灵敏度分析。在 Last.FM 和 Delicious 数据集上, 我们分别调整了 GNN 的层数和嵌入表征的维度大小, 以评估不同参数对 TAGCL 方法的影响。详细的实验结果展示在图 5.1 中。

在图 5.1 中, 我们展示了 TAGCL 方法在不同层数下在 Last.FM 和 Delicious 数据集上的准确率。结果表明, 增加 GNN 的层数并不总是能提高 TAGCL 的推荐性能。事实上, 当 GNN 层数超过三层时, Last.FM 数据集的准确率急剧下降。在 Delicious 数据集上, 虽然 Recall@20 和 ARP@20 随着层数的增加而变化, 但并没有出现明显的断崖式下降。另外, 在层数为两层时, TAGCL 方法在公平性方面表现出较优越的性能。

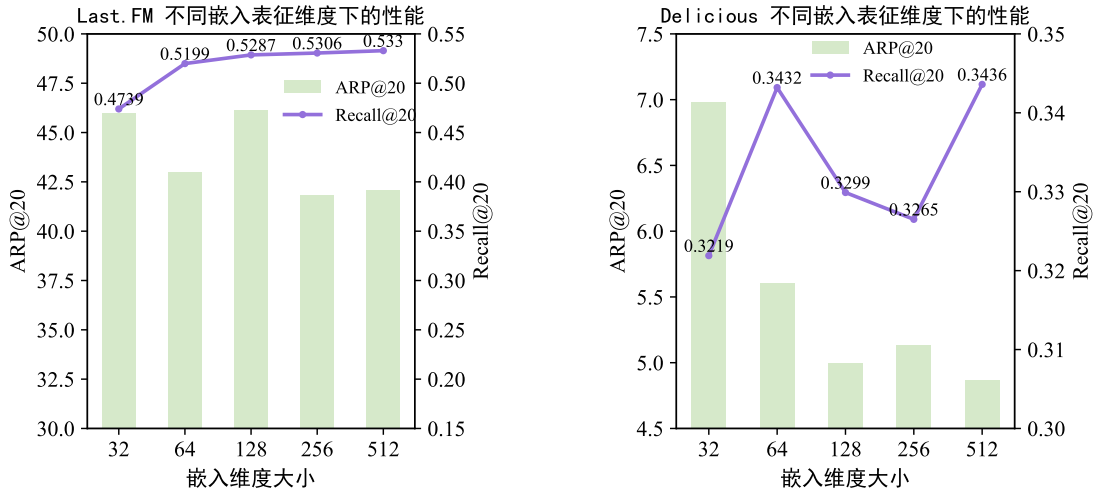


(a) Last.FM 数据层数对 TAGCL 性能的影响

(b) Delicious 数据层数对 TAGCL 性能的影响

图 5.1 图神经网络层数对 TAGCL 性能的影响

此外, 我们还研究了嵌入表征的维度大小对 TAGCL 方法的影响。实验结果展示在图 5.2 中。在 Last.FM 数据集上 (图 5.2a), 我们发现随着嵌入规模的增加, TAGCL 方



(a) Last.FM 嵌入表征维度大小对 TAGCL 性能的影响 (b) Delicious 嵌入表征维度大小对 TAGCL 性能的影响

图 5.2 嵌入表征维度大小对 TAGCL 性能的影响

法的推荐准确率也随之增加。而在 Delicious 数据集上（图 5.2b），TAGCL 的推荐性能则在嵌入规模变化时表现出较大的波动。特别地，当嵌入规模较大时，TAGCL 方法在 Delicious 数据集上表现更具有公平性。综合考虑性能和效率，我们认为使用 64 的嵌入尺寸是 TAGCL 方法的最佳选择。

5.3.4.4 个性化序列长度对模型的影响

本节旨在探索不同个性化序列长度对 TAGCL 推荐性能的影响。为此，我们分析了数据集 Last.FM 和 Delicious 下，TAGCL 在不同个性化序列长度 K 下与其他基线模型的推荐性能。我们的实验结果显示，TAGCL 在 Recall 和 NDCG 两个指标下，在任意个性化序列长度 K 下均有显著的性能提升，特别是在 NDCG 中，TAGCL 相对于次优模型有约 5% 的提升。此外，TAGCL 对流行性偏差也有较强的抑制作用，在任意个性化序列长度 K 下均表现出最低的偏差。然而，Delicious 数据集由于数据流行度偏差较小，因此该指标 TAGCL 并没有明显优势。总之，我们的实验结果表明，个性化序列长度 K 对于 TAGCL 推荐性能有着显著的影响，并且该模型能够有效地抑制流行性偏差，提高推荐质量。

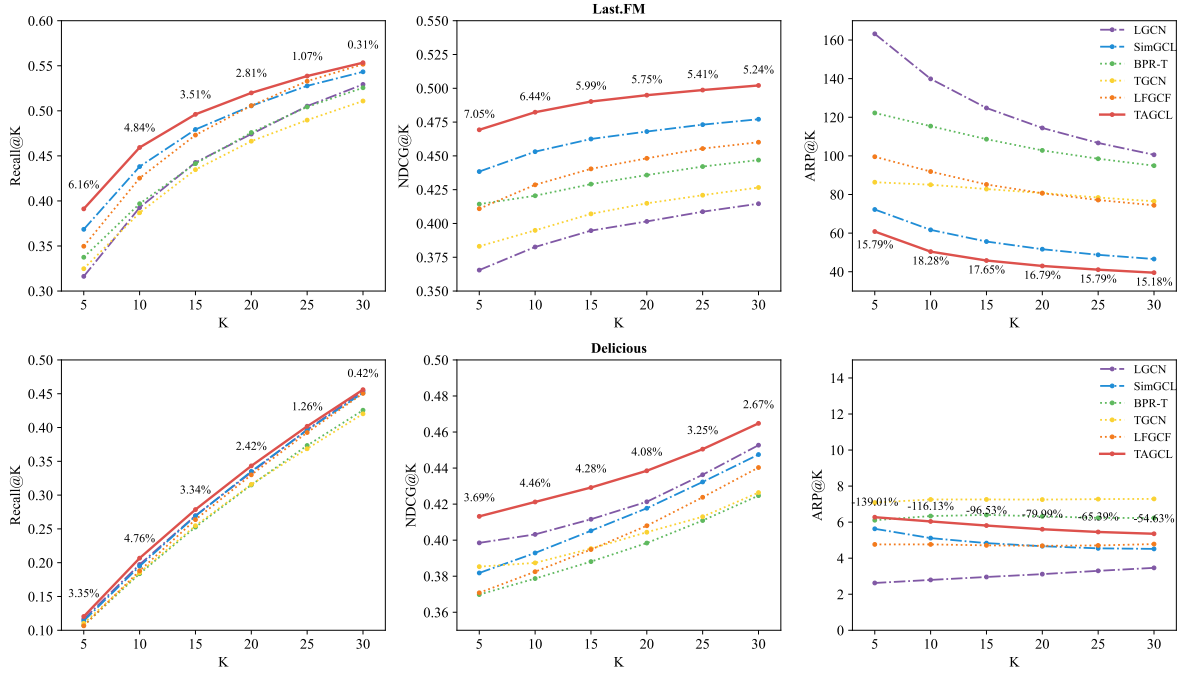


图 5.3 不同个性化序列长度下的 TAGCL 推荐性能

5.3.5 模型复杂度分析

本文量化分析了 LFGCF 的参数与推理时间复杂度，并且比较了模型参数量与推理时间。表 5.8 展示了 LFGCF 和其他基线模型的参数量与推理时间。与其他标签感知模型相比，LFGCF 的参数数量最少，因为它只需要学习用户、项目和标签的嵌入。因此，LFGCF 的参数比一般的模型多。我们的嵌入表征近包含标签嵌入，它提供了大约 0.22×10^6 的可训练参数。由于大规模的推荐系列通常对延迟有很高的要求，推理过程中的计算成本很重要。表 5.8 显示，大多数模型需要大约 0.15s 来推理 Delicious 中的所有测试集。然而，TGCN 明显比其他模型慢，因为它的重特征聚合设计，获得了注意力机制和卷积神经网络。

5.3.6 工业数据集上的应用

为了进一步验证 TAGCL 的有效性和实用性，我们在社交书签和出版物共享系统 BibSonomy 的最新版本中进行了离线测试^[76]，其中包含网站的书签数据和出版物的 BibTeX 数据。我们删除了那些标签分配少于 15 次的记录，以提高数据质量。BibSonomy-BM 和 BibSonomy-BT 分别指的是整个 BibSonomy 数据集中的书签数据和 BibTeX 数据。

表 5.8 模型在 Delicious 上的复杂度

模型	参数量	相对	推理	相对
	($\times 10^6$)	比率 (%)	时间 (s)	比率 (%)
LightGCN	~4.33	-5.18%	~0.14	+5.13%
NGCF	~4.36	-4.58%	~0.15	+13.91%
LFGCF	~4.55	-	~0.13	-
GNN-PTR	~4.57	+0.18%	~0.17	+16.33%
TGCN	~5.01	+9.15%	~2.12	+93.81%

BibSonomy-BM 包含 1622320 个交互，涵盖 5996 个用户、8092 个标签和 576232 个项目。BibSonomy-BT 包含 1972556 个交互，涵盖 9721 个用户、11313 个标签和 750514 个项目。它们的稀疏度分别为 99.95% 和 99.97%。与我们在之前使用的三个数据集相比，BibSonomy 规模更大，也更为稀疏。我们选择了 LightGCN 作为基线方法，因为它在基线方法中表现优异，并且具有简单性。表 5.9 中显示了我们提出的 TAGCL 框架与 LightGCN 的性能比较结果。结果表明，我们提出的 TAGCL 在稀疏的大规模数据集上仍然有效。在 BibSonomy 数据集上，TAGCL 的性能领先优势约为 1% 和 2%。与在 Delicious 数据集上的结果相似，TAGCL 提供了准确和高质量的推荐，但推荐的种类和多样性有所减少。我们认为这可能是由于 BibSonomy 和 Delicious 的流行性偏差性较小。

表 5.9 TAGCL 在 BibSonomy 上的应用

模型	BibSonomy-BM		BibSonomy-BT	
	Rec.@20	ARP.@20	Rec.@20	ARP.@20
LightGCN	0.6117	1.53	0.4810	1.39
TAGCL	0.6226	1.59	0.5173	1.67

5.4 本章小结

本章主要对提出的两个标签感知推荐模型 LFGCF 和 TAGCL 荆襄实验验证和性能对比分析。首先，本章介绍了实验所使用的数据集和数据预处理过程，并将数据集的统计信息与流行度偏差进行可视化。接着阐述具体的实验方法、实验平台和实验流程。在

进行多次实验后，整理数据，并对实验结果进行深入分析。实验表明，与现有的通用推荐模型和标签感知推荐模型相比，本文提出的模型在多个评估指标上都有较为显著的性能提升，在三个公开的学术数据集 MovieLens、Last.FM 和 Delicious 上，本文提出的模型 TAGCL 的召回率对比通用推荐模型有 4.6% 的提升，准确率有 5.0% 的提升，NDCG 有 4.21% 的提升，MRR 有 1.62% 的提升。对比标签感知推荐模型，TAGCL 在召回率有 5.18% 的提升，准确率有 8.0% 的提升，NDCG 有 7.22% 的提升，MRR 有 5.64% 的提升。对于数据偏差较大的 MovieLens 与 Last.FM 降低了 17% 的平均推荐流行度。最后，本文还在一个真实运行的推荐系统数据 BibSonomy 上测试 TAGCL 性能，实验结果证明 TAGCL 对比与基线模型 LightGCN 的召回率领先 1%。

6 总结与展望

本章将简要汇总本文的研究工作，并根据当前主流研究趋势对标签感知推荐系统进行未来研究展望。

6.1 工作总结

本文的研究工作主要集中在利用社交标签信息实现用户个性化物品推荐，旨在从社交标签数据中学习用户和物品的嵌入式表达，并结合个性化推荐算法实现 Top-K 推荐。本研究关注的核心问题在于如何充分利用内涵丰富且形式简洁的标签信息，以获得更好的用户和物品表达，进而提高推荐效果。同时，本文旨在合理利用图神经网络和对比学习来提高标签感知推荐的性能。

当前的研究存在三个主要问题：1) 现有的推荐系统涉及到的交互类型繁多，如何高效组织数据，并进一步建模数据中的多种交互类型是极具挑战的；2) 现有应用图神经网络的模型仅使用原始神经网络的设定，而没有根据标签感知推荐系统进行进一步调整；3) 现有模型所采用的优化机制往往忽略了图神经网络自身的特点，盲目地沿用传统推荐算法的优化目标进行模型优化。

为了解决这些问题，本文基于图神经网络和对比学习提出了两个标签感知推荐模型 LFGCF 和 TAGCL。LFGCF 模型将社交标签信息与用户和物品嵌入式表达相结合，通过利用轻量化图神经网络，使其更适应与社会标签数据。TAGCL 模型则采用了对比学习，以更好地去除推荐数据中常见的流行性偏差，并有效地提高标签感知推荐的性能。在公开数据集上，实验结果表明，本文提出的两个模型均优于现有的主流推荐算法，具有良好的推荐效果。具体来说，本文的研究工作主要包含以下几个方面：

(1) 本文旨在研究标签感知推荐系统，将其研究重点划分为数据与算法两个部分。在数据部分，本文分析过去文献对数据建模方式的问题，并针对社会标签数据提出了合理的建模方式。在算法部分，本文将标签感知推荐系统抽象为数学形式，为后续研究打下基础。

(2) 本文提出基于图神经网络的轻量化标签感知推荐模型 LFGCF，该模型针对推荐系统的特点去除了图卷积神经网络的特征变换和非线性激活组件，并采用加权和聚合函数进行消息传播，提高了模型推荐的精准度并减轻了训练难度。

(3) 本文还探讨了推荐系统中存在的流行度偏差现象，并提出了一种新的模型

TAGCL。该模型使用对比学习和知识图谱联合优化模型，在训练过程中对标签进行同时采样，有效地提高了模型推荐的精准度和公平性。

(4) 本文通过设计一系列实验并与当前主流推荐算法模型进行对比，评估了 LFGCF 和 TAGCL 的性能。此外，为验证模型的有效性和实用性，本文还在一个真实运行的系统数据中验证了模型的有效性。

综合来看，本文为推荐系统中的标签感知和不公平问题提供了新的思路和解决方案，并提出了可行的轻量化标签感知推荐模型 LFGCF 和新型的不公平感知推荐模型 TAGCL。

6.2 未来研究展望

对于标签感知推荐系统的进一步研究，本文认为有以下几点值得关注和探索：

(1) 使用超图来建模社会标注行为。超图 (hypergraph) 是一种广义的图结构，在信息科学和生命科学等领域得到了广泛应用。与传统的图不同，超图的一条边可以连接任意数量的顶点。考虑到标签可以被多个用户使用并描述多个物品，因此使用超图来进一步定义社会标注数据具有非常大的潜力。

(2) 深入研究图神经网络的更深层次。图神经网络的深度可以反映节点信息表征的来源范围。然而，随着深度的增加，过拟合 (over-fitting)、梯度消失 (gradient vanishing) 和过度平滑 (over-smoothing) 等问题变得越来越严重。在学术研究级别的数据上，使用较浅的图神经网络可以得到不错的效果。但在真实的工业场景中，小模型无法满足大规模数据的需求。因此，我们需要探索更深层次的图神经网络，并解决其面临的问题。

(3) 探索基于软最大损失函数的推荐算法优化方向。当前主流的推荐算法优化方向仍然是成对型损失函数。然而，该损失函数会引入归纳偏差，即对于用户没有交互过的物品，很多是因为推荐系统没有将其推荐给用户。因此，基于软最大损失函数的方法可以更有效地引入负样本，以减少这种偏差。

(4) 推荐系统的可解释性。对于用户来说，推荐系统的解释是非常重要的。如果用户无法理解推荐系统的决策过程，那么他们很可能对推荐结果产生怀疑，从而降低推荐系统的使用率。因此，我们需要探索如何使推荐系统更加可解释。例如，可以使用可视化技术来展示推荐系统的决策过程，并提供解释。

参考文献

- [1] 王梦琪, 唐长乐. 内容供给侧信息过载问题及优化策略——以互联网内容平台为例[J/OL]. 图书情报导刊, 2022, 7(09): 30-37 [2023-01-19]. https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDAUTO&filename=KJQB202209006&uniplatform=NZKPT&v=g6Aat-oYyJdyMGCAAK-NFT5t76naESxqQrepX64ps6OfTymEIUXhoQDo_ZJE5caY.
- [2] COVINGTON P, ADAMS J, SARGIN E. Deep Neural Networks for YouTube Recommendations [C/OL]//RecSys '16: Proceedings of the 10th ACM Conference on Recommender Systems. New York, NY, USA: Association for Computing Machinery, 2016: 191-198 [2023-01-19]. <https://doi.org/10.1145/2959100.2959190>. DOI: 10.1145/2959100.2959190.
- [3] WANG J, HUANG P, ZHAO H, et al. Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba[EB/OL]. arXiv. 2018 [2023-01-19]. <http://arxiv.org/abs/1803.02349>.
- [4] LIU Z, ZOU L, ZOU X, et al. Monolith: Real Time Recommendation System With Collisionless Embedding Table[EB/OL]. arXiv. 2022 [2023-01-19]. <http://arxiv.org/abs/2209.07663>.
- [5] GRBOVIC M, CHENG H. Real-time Personalization using Embeddings for Search Ranking at Airbnb [C/OL]//KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA: Association for Computing Machinery, 2018: 311-320 [2023-01-19]. <https://doi.org/10.1145/3219819.3219885>. DOI: 10.1145/3219819.3219885.
- [6] YIN H, CUI B, LI J, et al. Challenging the Long Tail Recommendation[EB/OL]. arXiv. 2012 [2023-01-19]. <http://arxiv.org/abs/1205.6700>.
- [7] RENDLE S, BALBY MARINHO L, NANOPOULOS A, et al. Learning Optimal Ranking with Tensor Factorization for Tag Recommendation[C/OL]//KDD '09: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France: Association for Computing Machinery, 2009: 727-736. <https://doi.org/10.1145/1557019.1557100>. DOI: 10.1145/1557019.1557100.
- [8] HOTH O A, JÄSCHKE R, SCHMITZ C, et al. Information Retrieval in Folksonomies: Search and Ranking[G/OL]//HUTCHISON D, KANADE T, KITTLER J, et al. The Semantic Web: Research and Applications: vol. 4011. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006: 411-426 [2021-11-30]. http://link.springer.com/10.1007/11762256_31. DOI: 10.1007/11762256_31.
- [9] CHEN J, DONG H, WANG X, et al. Bias and Debias in Recommender System: A Survey and Future Directions[EB/OL]. arXiv. 2021 [2023-02-28]. <http://arxiv.org/abs/2010.03240>.
- [10] ZHEN Y, LI W J, YEUNG D Y. TagiCoFi: Tag Informed Collaborative Filtering[C/OL]//RecSys '09: Proceedings of the Third ACM Conference on Recommender Systems. New York, New York,

- USA: Association for Computing Machinery, 2009: 69-76. <https://doi.org/10.1145/1639714.1639727>. DOI: 10.1145/1639714.1639727.
- [11] SHEPITSEN A, GEMMELL J, MOBASHER B, et al. Personalized Recommendation in Social Tagging Systems Using Hierarchical Clustering[C/OL]//RecSys '08: Proceedings of the 2008 ACM Conference on Recommender Systems. Lausanne, Switzerland: Association for Computing Machinery, 2008: 259-266. <https://doi.org/10.1145/1454008.1454048>. DOI: 10.1145/1454008.1454048.
- [12] PENG J, ZENG D D, ZHAO H, et al. Collaborative Filtering in Social Tagging Systems Based on Joint Item-Tag Recommendations[C/OL]//CIKM '10: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. Toronto, ON, Canada: Association for Computing Machinery, 2010: 809-818. <https://doi.org/10.1145/1871437.1871541>. DOI: 10.1145/1871437.1871541.
- [13] ZHANG Z K, ZHOU T, ZHANG Y C. Personalized recommendation via integrated diffusion on user-item-tag tripartite graphs[J/OL]. Physica A: Statistical Mechanics and its Applications, 2010, 389(1): 179-186. <https://www.sciencedirect.com/science/article/pii/S0378437109006839>. DOI: <https://doi.org/10.1016/j.physa.2009.08.036>.
- [14] ZHANG Z K, LIU C, ZHANG Y C, et al. Solving the Cold-Start Problem in Recommender Systems with Social Tags[J/OL]. Expert Systems with Applications, 2012, 39(12): 10990-11000 [2021-12-05]. <http://arxiv.org/abs/1004.3732>. DOI: 10.1016/j.eswa.2012.03.025.
- [15] FolkRank++: An Optimization of FolkRank Tag Recommendation Algorithm Integrating User and Item Information[J/OL]. KSII Transactions on Internet and Information Systems, 2021, 15(1) [2021-12-05]. <http://itiis.org/digital-library/24227>. DOI: 10.3837/tiis.2021.01.001.
- [16] RENDLE S, BALBY MARINHO L, NANOPOULOS A, et al. Learning optimal ranking with tensor factorization for tag recommendation[C/OL]//Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09. Paris, France: ACM Press, 2009: 727 [2021-11-28]. <http://portal.acm.org/citation.cfm?doid=1557019.1557100>. DOI: 10.1145/1557019.1557100.
- [17] LI H, DIAO X, CAO J, et al. Tag-aware recommendation based on Bayesian personalized ranking and feature mapping[J/OL]. Intelligent Data Analysis, 2019, 23(3): 641-659 [2021-11-25]. <https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/IDA-193982>. DOI: 10.3233/IDA-193982.
- [18] HE X, LIAO L, ZHANG H, et al. Neural Collaborative Filtering[EB/OL]. arXiv. 2017 [2023-02-03]. <http://arxiv.org/abs/1708.05031>.

- [19] ZHANG W, DU T, WANG J. Deep Learning over Multi-field Categorical Data: A Case Study on User Response Prediction[EB/OL]. arXiv. 2016 [2023-02-09]. <http://arxiv.org/abs/1601.02376>.
- [20] QU Y, FANG B, ZHANG W, et al. Product-based Neural Networks for User Response Prediction over Multi-field Categorical Data[EB/OL]. arXiv. 2018 [2023-02-09]. <http://arxiv.org/abs/1807.00311>.
- [21] SHAN Y, HOENS T R, JIAO J, et al. Deep Crossing: Web-Scale Modeling without Manually Crafted Combinatorial Features[C/OL]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, 2016: 255-262 [2023-02-09]. <https://dl.acm.org/doi/10.1145/2939672.2939704>. DOI: 10.1145/2939672.2939704.
- [22] SEDHAIN S, MENON A K, SANNER S, et al. AutoRec: Autoencoders Meet Collaborative Filtering[C/OL]//Proceedings of the 24th International Conference on World Wide Web. Florence Italy: ACM, 2015: 111-112 [2023-02-09]. <https://dl.acm.org/doi/10.1145/2740908.2742726>. DOI: 10.1145/2740908.2742726.
- [23] LIANG D, KRISHNAN R G, HOFFMAN M D, et al. Variational Autoencoders for Collaborative Filtering[C/OL]//WWW '18: Proceedings of the 2018 World Wide Web Conference. Republic: International World Wide Web Conferences Steering Committee, 2018: 689-698 [2023-02-09]. <https://doi.org/10.1145/3178876.3186150>. DOI: 10.1145/3178876.3186150.
- [24] HE X, HE Z, SONG J, et al. NAIS: Neural Attentive Item Similarity Model for Recommendation[J/OL]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(12): 2354-2366 [2023-02-09]. <http://arxiv.org/abs/1809.07053>. DOI: 10.1109/TKDE.2018.2831682.
- [25] CHEN J, ZHANG H, HE X, et al. Attentive Collaborative Filtering: Multimedia Recommendation with Item- and Component-Level Attention[C/OL]//SIGIR '17: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, USA: Association for Computing Machinery, 2017: 335-344 [2023-02-09]. <https://doi.org/10.1145/3077136.3080797>. DOI: 10.1145/3077136.3080797.
- [26] WANG X, HE X, WANG M, et al. Neural Graph Collaborative Filtering[C/OL]//SIGIR'19: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris, France: Association for Computing Machinery, 2019: 165-174. <https://doi.org/10.1145/3331184.3331267>. DOI: 10.1145/3331184.3331267.
- [27] HE X, DENG K, WANG X, et al. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation[C/OL]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, USA: Association for Computing Machinery, 2020: 639-648. <https://doi.org/10.1145/3397271.3401063>.

- [28] MAO K, ZHU J, WANG J, et al. SimpleX: A Simple and Strong Baseline for Collaborative Filtering [EB/OL]. arXiv. 2021 [2023-02-17]. <http://arxiv.org/abs/2109.12613>.
- [29] TANG H, ZHAO G, WU Y, et al. Multi-Sample based Contrastive Loss for Top-k Recommendation [EB/OL]. arXiv. 2021 [2023-02-17]. <http://arxiv.org/abs/2109.00217>.
- [30] WU J, WANG X, GAO X, et al. On the Effectiveness of Sampled Softmax Loss for Item Recommendation[EB/OL]. arXiv. 2022 [2023-02-28]. <http://arxiv.org/abs/2201.02327>.
- [31] XU Z, CHEN C, LUKASIEWICZ T, et al. Tag-Aware Personalized Recommendation Using a Deep-Semantic Similarity Model with Negative Sampling[J/OL]. Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, 2016: 1921-1924 [2023-02-09]. <https://dl.acm.org/doi/10.1145/2983323.2983874>. DOI: 10.1145/2983323.2983874.
- [32] ZUO Y, ZENG J, GONG M, et al. Tag-aware recommender systems based on deep neural networks [J/OL]. Neurocomputing, 2016, 204: 51-60 [2021-12-02]. <https://linkinghub.elsevier.com/retrieve/pii/S0925231216301151>. DOI: 10.1016/j.neucom.2015.10.134.
- [33] CHEN B, DING Y, XIN X, et al. AIRec: Attentive intersection model for tag-aware recommendation [J/OL]. Neurocomputing, 2021, 421: 105-114(2021-01-15) [2022-05-09]. <https://www.sciencedirect.com/science/article/pii/S0925231220312789>. DOI: 10.1016/j.neucom.2020.08.018.
- [34] HUANG R, HAN C, CUI L. Tag-aware Attentional Graph Neural Networks for Personalized Tag Recommendation[C]//2021 International Joint Conference on Neural Networks (IJCNN). 2021: 1-8. DOI: 10.1109/IJCNN52387.2021.9533380.
- [35] CHEN B, GUO W, TANG R, et al. TGCN: Tag Graph Convolutional Network for Tag-Aware Recommendation[C/OL]//CIKM '20: Proceedings of the 29th ACM International Conference on Information and Knowledge Management. Virtual Event, Ireland: Association for Computing Machinery, 2020: 155-164. <https://doi.org/10.1145/3340531.3411927>. DOI: 10.1145/3340531.3411927.
- [36] BISCHOFF K, FIRAN C S, NEJDL W, et al. Can all tags be used for search?[C/OL]//Proceeding of the 17th ACM conference on Information and knowledge mining - CIKM '08. Napa Valley, California, USA: ACM Press, 2008: 193 [2021-12-05]. <http://portal.acm.org/citation.cfm?doid=1458082.1458112>. DOI: 10.1145/1458082.1458112.
- [37] WU L, HE X, WANG X, et al. A Survey on Accuracy-oriented Neural Recommendation: From Collaborative Filtering to Information-rich Recommendation[J/OL]. IEEE Transactions on Knowledge and Data Engineering, 2022: 1-1 [2023-02-21]. <http://arxiv.org/abs/2104.13030>. DOI: 10.1109/TKDE.2022.3145690.
- [38] XU Z, CHEN C, LUKASIEWICZ T, et al. Tag-Aware Personalized Recommendation Using a Deep-

- Semantic Similarity Model with Negative Sampling[C/OL]//CIKM '16: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. Indianapolis, Indiana, USA: Association for Computing Machinery, 2016: 1921-1924. <https://doi.org/10.1145/2983323.2983874>. DOI: 10.1145/2983323.2983874.
- [39] BATTAGLIA P W, HAMRICK J B, BAPST V, et al. Relational inductive biases, deep learning, and graph networks[EB/OL]. arXiv. 2018 [2023-02-23]. <http://arxiv.org/abs/1806.01261>.
- [40] KIPF T N, WELLING M. Semi-Supervised Classification with Graph Convolutional Networks [C/OL]//5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net, 2017. <https://openreview.net/forum?id=SJU4ayYgl>.
- [41] CHEN M, WEI Z, HUANG Z, et al. Simple and Deep Graph Convolutional Networks[EB/OL]. arXiv. 2020 [2023-02-23]. <http://arxiv.org/abs/2007.02133>.
- [42] HAMILTON W L, YING R, LESKOVEC J. Inductive Representation Learning on Large Graphs [EB/OL]. arXiv. 2018 [2023-02-22]. <http://arxiv.org/abs/1706.02216>.
- [43] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[EB/OL]. arXiv. 2018 [2023-02-22]. <http://arxiv.org/abs/1710.10903>.
- [44] ZHOU J, CUI G, HU S, et al. Graph neural networks: A review of methods and applications[J/OL]. AI Open, 2020, 1: 57-81 [2023-02-24]. <https://www.sciencedirect.com/science/article/pii/S2666651021000012>. DOI: 10.1016/j.aiopen.2021.01.001.
- [45] KOREN Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C/OL] //KDD '08: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY, USA: Association for Computing Machinery, 2008: 426-434 [2023-02-23]. <https://doi.org/10.1145/1401890.1401944>. DOI: 10.1145/1401890.1401944.
- [46] GORI M. ItemRank: A Random-Walk Based Scoring Algorithm for Recommender Engines[J].
- [47] WU S, SUN F, ZHANG W, et al. Graph Neural Networks in Recommender Systems: A Survey [EB/OL]. arXiv. 2022 [2023-02-24]. <http://arxiv.org/abs/2011.02260>.
- [48] HU W, FEY M, REN H, et al. OGB-LSC: A Large-Scale Challenge for Machine Learning on Graphs [EB/OL]. arXiv. 2021 [2023-02-25]. <http://arxiv.org/abs/2103.09430>.
- [49] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[J/OL]. CoRR, 2017. arXiv: 1710.10903. <http://arxiv.org/abs/1710.10903>.
- [50] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[J/OL]. arXiv:1512.03385 [cs], 2015 [2021-01-19]. <http://arxiv.org/abs/1512.03385>.

- [51] HAMILTON W, YING Z, LESKOVEC J. Inductive Representation Learning on Large Graphs[C/OL] //GUYON I, LUXBURG U V, BENGIO S, et al. Advances in Neural Information Processing Systems: vol. 30. Curran Associates, Inc., 2017. <https://proceedings.neurips.cc/paper/2017/file/5dd9db5e033da9c6fb5ba83c7a7e9-Paper.pdf>.
- [52] ZHU H, FENG F, HE X, et al. Bilinear Graph Neural Network with Neighbor Interactions[C]//IJCAI'20: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. Yokohama, Yokohama, Japan, 2021.
- [53] LI Q, HAN Z, WU X M. Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning[EB/OL]. arXiv. 2018 [2022-05-15]. <http://arxiv.org/abs/1801.07606>.
- [54] LIN Y, LIU Z, SUN M, et al. Learning Entity and Relation Embeddings for Knowledge Graph Completion[J]. 7.
- [55] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization[EB/OL]. arXiv. 2014. <https://arxiv.org/abs/1412.6980>.
- [56] BENGIO Y, DUCHARME R, VINCENT P, et al. A Neural Probabilistic Language Model[J/OL]. Journal of Machine Learning Research, 2003, 3(Feb): 1137-1155 [2023-02-26]. <https://jmlr.csail.mit.edu/papers/v3/bengio03a>.
- [57] MNIH A, TEH Y W. A fast and simple algorithm for training neural probabilistic language models[C] //ICML'12: Proceedings of the 29th International Conference on International Conference on Machine Learning. Madison, WI, USA: Omnipress, 2012: 419-426 [2023-02-28].
- [58] OORD A V D, LI Y, VINYALS O. Representation Learning with Contrastive Predictive Coding [EB/OL]. arXiv. 2019 [2023-02-03]. <http://arxiv.org/abs/1807.03748>.
- [59] 吴剑灿. 基于图神经网络的个性化推荐[D/OL]. 中国科学技术大学, 2022 [2023-01-19]. <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CDFD&dbname=CDFDTEMP&filename=1023001839.nh&uniplatform=NZKPT&v=QUpwPQIQYPFqva-0fp7t30wlApsQoOOKKJ9RIaBmyGv66o0MAwd4j5OEacac7f5M>. DOI: 10.27517/d.cnki.gzkju.2022.000378.
- [60] ABDOLLAHPOURI H, MANSOURY M, BURKE R, et al. The Connection Between Popularity Bias, Calibration, and Fairness in Recommendation[C/OL]//Fourteenth ACM Conference on Recommender Systems. Virtual Event Brazil: ACM, 2020: 726-731 [2023-02-28]. <https://dl.acm.org/doi/10.1145/3383313.3418487>. DOI: 10.1145/3383313.3418487.
- [61] CHEN T, KORNBLITH S, NOROUZI M, et al. A Simple Framework for Contrastive Learning of Visual Representations[C/OL]//III H D, SINGH A. Proceedings of Machine Learning Research: Proceedings of the 37th International Conference on Machine Learning: vol. 119. PMLR, 2020: 1597-

1607. <https://proceedings.mlr.press/v119/chen20j.html>.
- [62] GIDARIS S, SINGH P, KOMODAKIS N. Unsupervised Representation Learning by Predicting Image Rotations[C/OL]//International Conference on Learning Representations. 2018. <https://openreview.net/forum?id=S1v4N2l0->.
- [63] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C/OL]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, 2019: 4171-4186. <https://aclanthology.org/N19-1423>. DOI: 10.18653/v1/N19-1423.
- [64] GAO T, YAO X, CHEN D. SimCSE: Simple Contrastive Learning of Sentence Embeddings[C/OL]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online: Association for Computational Linguistics, 2021: 6894-6910. <https://aclanthology.org/2021.emnlp-main.552>. DOI: 10.18653/v1/2021.emnlp-main.552.
- [65] WU J, WANG X, FENG F, et al. Self-supervised Graph Learning for Recommendation[C/OL]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021: 726-735 [2023-01-26]. <http://arxiv.org/abs/2010.10783>. DOI: 10.1145/3404835.3462862.
- [66] YU J, YIN H, XIA X, et al. Are Graph Augmentations Necessary? Simple Graph Contrastive Learning for Recommendation[C/OL]//SIGIR '22: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid, Spain: Association for Computing Machinery, 2022: 1294-1303. <https://doi.org/10.1145/3477495.3531937>. DOI: 10.1145/3477495.3531937.
- [67] OORD A V D, LI Y, VINYALS O. Representation Learning with Contrastive Predictive Coding [EB/OL]. arXiv. 2018. <https://arxiv.org/abs/1807.03748>.
- [68] CANTADOR I, BRUSILOVSKY P, KUFLIK T. Second Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec2011)[C/OL]//RecSys '11: Proceedings of the Fifth ACM Conference on Recommender Systems. Chicago, Illinois, USA: Association for Computing Machinery, 2011: 387-388. <https://doi.org/10.1145/2043932.2044016>. DOI: 10.1145/2043932.2044016.
- [69] DACREMA M F, CREMONESI P, JANNACH D. Are We Really Making Much Progress? A Wor-rying Analysis of Recent Neural Recommendation Approaches[C/OL]//RecSys '19: Proceedings of the 13th ACM Conference on Recommender Systems. Copenhagen, Denmark: Association for Com-

- puting Machinery, 2019: 101-109. <https://doi.org/10.1145/3298689.3347058>. DOI: 10.1145/3298689.3347058.
- [70] ZHAO W X, MU S, HOU Y, et al. RecBole: Towards a Unified, Comprehensive and Efficient Framework for Recommendation Algorithms[J/OL]. arXiv:2011.01731 [cs], 2021 [2022-04-17]. <http://arxiv.org/abs/2011.01731>.
- [71] ROELLEKE T. Information Retrieval Models: Foundations & Relationships[M/OL]. Springer International Publishing, 2022. <https://books.google.com.hk/books?id=YX9yEAAAQBAJ>.
- [72] WANG Y, WANG L, LI Y, et al. A Theoretical Analysis of NDCG Type Ranking Measures[EB/OL]. arXiv. 2013 [2023-02-16]. <http://arxiv.org/abs/1304.6480>.
- [73] CRASWELL N. Mean Reciprocal Rank[G/OL]//LIU L, ÖZSU M T. Encyclopedia of Database Systems. Boston, MA: Springer US, 2009: 1703-1703 [2023-02-16]. https://doi.org/10.1007/978-0-387-39940-9_488. DOI: 10.1007/978-0-387-39940-9_488.
- [74] PRECHELT L. Early Stopping —But When?[G/OL]//MONTAVON G, ORR G B, MÜLLER K R. Lecture Notes in Computer Science: Neural Networks: Tricks of the Trade: Second Edition. Berlin, Heidelberg: Springer, 2012: 53-67 [2023-02-13]. https://doi.org/10.1007/978-3-642-35289-8_5. DOI: 10.1007/978-3-642-35289-8_5.
- [75] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks [J].
- [76] BENZ D, HOTH O A, JÄSCHKE R, et al. The Social Bookmark and Publication Management System BibSonomy[J/OL]. The VLDB Journal, 2010, 19(6): 849-875. <http://www.kde.cs.uni-kassel.de/pub/pdf/benz2010social.pdf>. DOI: 10.1007/s00778-010-0208-4.

攻读硕士学位期间取得的研究成果

一、科研项目

1. 浙江省重点研发项目:“面向个性化学习的精准教育系统研制及产业化”(2017C03058)

二、已发表学术论文

1. Pursuit and evasion strategy of a differential game based on deep reinforcement learning. *Frontiers in Bioengineering and Biotechnology*, 2022, 10: 827408.
2. A fairness-aware graph contrastive learning recommender framework for social tagging systems. *Information Sciences*, 2023: 119064.

三、发明专利

1. 一种视频人员在线时长估计方法及对应系统, 专利申请号: 202110178862.3
2. 信号干扰下的超带宽精确定位方法, 专利申请号: 202210119309.7

四、科研竞赛

1. 中国研究生创新实践大赛——“华为杯”第十八届中国研究生数学建模竞赛, 三等奖
2. 第五届全国应用统计专业学位研究生案例大赛, 三等奖

致谢

三年的研究生生活转瞬即逝，恍惚还能记起刚刚收到研究生入学通知书那一刻的心情：心里的一块大石头落了地，有一丝兴奋，还有一丝迷茫。现在我站在学位论文的终点，当初一刻的心情竟然昨日重现，用“拔剑四顾心茫然”形容或许恰如其分。

生活也许就是像是一个环，研究生的人生经历也不例外，好在回到“原点”时，突然意识到自己正往前、往上迈了小小的一步。这一步着实来之不易：

首先，我要感谢我的导师董黎刚对我的孜孜教诲。您不仅是我学术道路上的引路人，而且是我人生道路上的榜样。您在我迷茫时给予我鼓励，在我困难时给予我支持。有机会受您的教诲是我学生生涯的荣幸。

其次，我要感谢我的家人和朋友们。我要感谢我的父母为我提供了研究生生活所需的物质支持，也要感谢我的朋友们为我提供了克服研究生困难所需要的精神力量。感谢我最好的室友：李啸天、李永续、谢仁华，你们一起创造了这样一个有爱有温度的宿舍环境，我难以忘记和你们一起度过的那些日日夜夜；感谢我的课题组同门：张研、李玉娥、尹彦尚、马巧巧、戴剑锋、吴先军、王程可等，和你们在实验室共同奋斗的日子如今想来弥足珍贵；感谢我从小到大的老友：夏浩瑜、徐胜亮、胡斐明、陈一炆等，虽然这些日子和你们天各一方，但幸运的是每年假期都能和你们再次相聚；感谢我两段实习经历的同事和领导，和你们共事让我收益颇丰，我不仅有学习到业界最前沿的知识，还学会了怎么样和人通力合作，完成目标。

最后，我想特别感谢吴卓璇，你给予我许多奇妙力量，让我有勇气面对生活中的重重不顺，你也给予我许多难忘记忆，让我重现发现生活中的美好瞬间。

这一步着实丰富多彩。无论怎样，希望我依旧如此幸运。

前路漫漫，“天佑勇者”。

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含本人为获得浙江工商大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名:

导师签名:

签字日期: 年 月 日

签字日期: 年 月 日

关于论文使用授权的说明

本学位论文作者完全了解浙江工商大学有关保留、使用学位论文的规定：浙江工商大学有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅，可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文，并且本人电子文档的内容和纸质论文的内容相一致。

本论文提交 ☐ 即日起/ ☐ 半年/ ☐ 一年以后，同意发布。

“内部”学位论文在解密后也遵守此规定。

学位论文作者签名:

导师签名:

签字日期: 年 月 日

签字日期: 年 月 日