

学号 1908040126

年级 2019 级

河海大学

## 本科毕业设计

# 结合超图卷积网络和图注意力网络的会话推荐模型

专 业 计算机科学与技术

姓 名 李怡飞

指导教师 许卓明

评 阅 人 姜渊胜

2023 年 6 月

中国 南京

# **BACHELOR'S DEGREE THESIS OF HOHAI UNIVERSITY**

## **Session-based Recommendation Model Combining Hypergraph Convolutional Networks with Graph Attention Networks**

College : College of Computer Science and Information

Subject : Computer Science and Technology

Name : Yifei Li

Directed by : Zhuoming Xu Professor

NANJING CHINA

学术声明：

## 郑 重 声 明

本人呈交的毕业设计，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本设计（论文）的研究成果不包含他人享有著作权的内容。对本设计（论文）所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本设计（论文）的知识产权归属于培养单位。

本人签名： 李心

日期： 2023.06.19

## 摘 要

近年来，基于会话的推荐（也有简称为“会话推荐”）已然成为推荐模型的新范式，其典型任务是根据当前会话中的项目序列预测下一个最有可能的用户-项目交互（即下一项推荐）。如何从不同层面协同学习用户偏好并有效预测用户偏好是下一项推荐任务中一个重要且具有挑战性的问题。

目前主流的会话推荐模型是基于图神经网络的会话推荐模型。然而，大多数这样的推荐模型仅仅建模各对项目之间的过渡关系，忽略了项目之间的高阶关系。超图提供了一种自然的方式来捕捉复杂的高阶关系。为了获得项目间的高阶关系，本文引入超图卷积网络对项目层的用户偏好进行建模和处理，以更好地捕获项目节点之间的相关性。图注意力网络可以自适应地学习节点之间的注意力权重并决定特征的聚合程度。基于上述技术背景与动机，本文引入图注意力网络对根据所有会话序列构建的会话图进行处理，获得会话层的用户偏好。本文的研究问题是如何结合超图卷积网络和图注意力网络来更有效地捕获用户偏好，由此提高下一项推荐的性能（准确性）。

本文的主要工作是设计并实现一种结合超图卷积网络和图注意力网络的会话推荐模型（Session-based Recommendation Model Combining Hypergraph Convolutional Networks with Graph Attention Networks，简称 SR-HCGAT）。SR-HCGAT 利用超图卷积网络（HGCN）从根据所有会话信息构建的超图中捕获项目层的用户偏好；利用图注意力网络（GAT）从根据所有会话信息构建的会话图中捕获会话层的用户偏好；对包含互补信息的上述两种用户偏好进行联合学习以更准确地捕获用户偏好；最后，利用学得的项目层的用户偏好产生下一项推荐，以改善推荐性能。

本文采用性能评价指标精度（Precision@K）和平均倒数排名（Mean Reciprocal Rank@K）在三个真实世界数据集 Tmall、Nowplaying、Diginetica 上开展了实验评价，实验结果表明本文提出的结合超图卷积网络和图注意力网络的会话推荐模型（SR-HCGAT）的推荐性能优于七个具有代表性的会话推荐模型。本文成果具有科学意义与实用价值。

**关键词：**基于会话的推荐；超图卷积网络；图注意力网络；联合学习；推荐系统

# ABSTRACT

In recent years, session-based recommendation has emerged as a new paradigm of recommendation models. Its typical task is to predict the next most probable user-item interaction (i.e., next-item recommendation) based on the item sequence in the current session. In such a recommendation task, how to effectively learn user preferences and predict user preferences from different perspectives is an important and challenging problem.

The current mainstream models for session-based recommendation are based on graph neural networks. However, most of these models only model the transition between each pair of items and overlook the high-order relations among items. Hypergraphs provide a natural way to capture complex high-order relations. To capture the high-order relations among items, this thesis introduces hypergraph convolutional networks to model and process user preferences at the item level, in order to better capture the correlations between item nodes. Graph attention networks can adaptively learn attention weights between nodes and determine the degree of feature aggregation. Based on the above technical background and motivation, the thesis introduces graph attention networks to process the session graph constructed from all session sequences and obtain user preferences at the session level. The research question in this thesis is how to combine hypergraph convolutional networks with graph attention networks to capture user preferences more effectively, thus improving the performance (accuracy) of next-item recommendation.

The main objective of this thesis is to design and implement a novel model called Session-based Recommendation Model Combining Hypergraph Convolutional Networks with Graph Attention Networks (SR-HCGAT). SR-HCGAT utilizes Hypergraph Convolutional Networks (HGCN) to capture item-level user preferences from the hypergraph constructed from all session information, and utilizes Graph Attention Networks (GAT) to capture session-level user preferences from the session graph constructed from all session information. The SR-HCGAT model jointly learns the above two user preferences containing complementary information to capture the user

preferences more accurately. Finally, the learned item-level user preferences are used to generate next-item recommendations to improve performance.

In this thesis, the performance evaluation metrics Precision@K and Mean Reciprocal Rank@K were used to carry out experimental evaluation on three real-world data sets: Tmall, Nowplaying and Diginetica. The experimental results show that the proposed SR-HCGAT model outperforms seven representative session-based recommendation models. The achievement of this thesis holds scientific significance and practical value.

**Key words:** session-based recommendation; hypergraph convolutional network; graph attention network; joint learning; recommender system

# 目 录

摘 要 .....	I
ABSTRACT .....	II
目 录 .....	IV
第 1 章 绪论 .....	1
1.1 课题背景和意义 .....	1
1.2 研究现状 .....	2
1.2.1 基于 RNN 的会话推荐模型 .....	2
1.2.2 基于 GNN 的会话推荐模型 .....	3
1.3 本文工作概述 .....	3
1.4 论文组织结构 .....	4
第 2 章 背景知识与支撑技术 .....	5
2.1 超图卷积网络 .....	5
2.2 图注意力网络 .....	8
2.3 基于会话的推荐 .....	9
2.3.1 会话推荐的基本概念 .....	9
2.3.2 会话推荐技术与模型 .....	10
第 3 章 结合 HGCN 和 GAT 的会话推荐模型及其实现 .....	12
3.1 引言 .....	12
3.2 问题陈述及符号定义 .....	12
3.3 SR-HCGAT 会话推荐模型 .....	13
3.3.1 模型概述 .....	13
3.3.2 超图和会话图的构建 .....	14
3.3.3 用户偏好捕获层 .....	15
3.3.4 联合学习层 .....	17
3.3.5 概率预测层 .....	18
3.3.6 模型训练 .....	18

3.4 模型实现 .....	18
3.4.1 编程语言与开发工具 .....	18
3.4.2 用户偏好捕获层的实现 .....	19
3.4.3 联合学习层的实现 .....	20
3.4.4 概率预测层的实现 .....	21
3.4.5 模型训练的实现 .....	21
3.5 本章小结 .....	22
<b>第 4 章 推荐性能实验评价 .....</b>	<b>23</b>
4.1 引言 .....	23
4.2 实验设置 .....	23
4.2.1 数据集 .....	23
4.2.2 性能评价指标 .....	24
4.2.3 基线模型 .....	25
4.2.4 超参数设定 .....	26
4.3 实验结果与讨论 .....	26
4.4 本章小结 .....	28
<b>第 5 章 总结与展望 .....</b>	<b>29</b>
5.1 工作总结 .....	29
5.2 未来工作展望 .....	29
<b>参考文献 .....</b>	<b>30</b>
<b>致 谢 .....</b>	<b>35</b>



# 第1章 绪论

## 1.1 课题背景和意义

在信息社会和数字化经济时代,为了满足用户对智能推荐的需求,人们提出了推荐系统(Recommender System, RS)<sup>[1]</sup>以准确响应用户需求并结合用户历史访问记录提供可靠的推荐结果。推荐系统是缓解信息过载或选择过载的有力工具,在消费、服务和决策等方面发挥着举足轻重的作用。遗憾的是,协同过滤(Collaborative Filtering)、基于内容的推荐(Content-based Recommendation)、混合推荐<sup>[2,3]</sup>等传统推荐系统只关注用户的长期、静态偏好而忽略了用户的短期兴趣,它们提供的推荐结果并不准确。因此,基于会话的推荐系统(Session-Based Recommender System, SBRS)<sup>[4,5,6]</sup>应运而生,它能够捕获用户的短期、动态偏好,也顺理成章成为推荐系统的新范式。

基于会话的推荐通过对用户行为序列进行建模并综合分析用户在一段时间内的行为和用户的历史行为,为用户提供个性化的推荐服务,增强用户的满意度和信任感,同时也能提高推荐的效果和效率。例如,用户在浏览电影推荐网站时,从首页开始浏览,逐渐深入到某一电影详情页,最后可能选择了一部电影观看。在对用户所喜爱的电影进行推荐的过程中,基于会话的推荐会综合用户当前选择的电影和曾经选择的电影记录,更精准地推荐用户所喜爱的电影。

实现基于会话的推荐关键在于如何对用户行为序列进行建模,一般采用循环神经网络(Recurrent Neural Network, RNN)<sup>[7]</sup>或其变体进行建模,如长短期记忆网络(Long Short-Term Memory, LSTM)<sup>[8]</sup>或门控循环单元(Gate Recurrent Unit, GRU)<sup>[9]</sup>等。这些模型可以从用户的历史行为中学习到用户的兴趣演化趋势,预测用户下一步可能发生的行为。基于会话的推荐已经被广泛应用于电商、社交网络、新闻推荐等领域,取得了不错的效果,尤其在面对用户兴趣变化较为频繁的场景下,表现更加出色。

总的来说,基于会话的推荐在现代商业和互联网应用中扮演着非常重要的角色,利用各种算法和技术来分析用户的行为和偏好,以此为用户提供个性化的服务推荐,这不仅能够提高用户满意度、优化商业模式、还能降低成本、提高品牌忠诚度等等,对企业 and 用户都有重大意义。

## 1.2 研究现状

近年来，深度学习技术与应用蓬勃发展，在计算机视觉<sup>[10]</sup>和语音识别<sup>[11]</sup>等领域取得了卓越成就。在基于会话的推荐领域<sup>[12]</sup>中，基于深度神经网络的会话推荐模型利用深度神经网络的强大能力从用户的历史行为中捕获用户行为序列的复杂关系和模式，从而改善推荐性能。基于循环神经网络（Recurrent Neural Networks, RNNs）<sup>[12,13]</sup>的会话推荐模型和基于图神经网络（Graph Neural Networks, GNNs）的会话推荐模型近年来一直是深度学习应用至基于会话的推荐系统的研究热点。

### 1.2.1 基于 RNN 的会话推荐模型

RNN 是一种常用于处理序列数据的神经网络。图 1.1 是循环神经网络的示例，其中“延时器”为一个虚拟单元，记录神经元的最近一次（或几次）活性值。考虑到 RNN 及其变体的循环结构在建模会话序列数据方面的潜在优势，研究者将 RNN 用于基于会话推荐的任务。

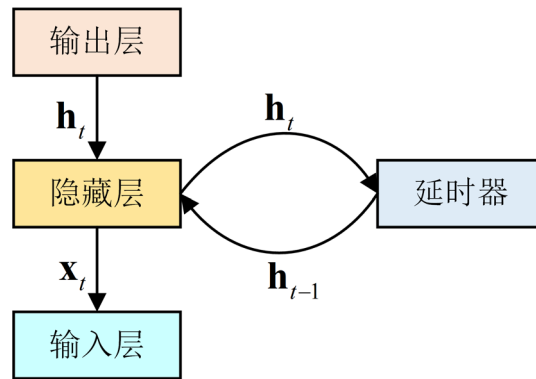


图 1.1 循环神经网络示例

Balazs Hidasi 等人<sup>[9]</sup>第一次在会话推荐中使用了 RNN，提出了一种基于 RNN 的会话推荐模型 GRU4Rec，该模型使用门控循环单元对用户的会话序列进行建模，以捕捉会话中的上下文信息和用户的兴趣演化。Li 等人<sup>[14]</sup>提出了一种会话推荐模型 NARM，该模型使用注意力机制（attention）建模用户的兴趣演化，通过将注意力机制集成到 RNN 中来建模用户的序列化行为和主要意图，使用双线性匹配方案计算推荐分数。Liu 等人<sup>[15]</sup>为捕捉用户在会话中的短期兴趣，使用一个短期兴趣编码器替换掉了 NARM 中的递归编码器，提出了短期注意力优先模型 STAMP。STAMP 模型能够利用注意力机制和短期记忆优先机制（Short-Term Attention/Memory Priority）对会话序列的重要元素进行加权编码和优先级更新。Huang 等人<sup>[16]</sup>提出了一种知识增强的序列推荐器 KSR，该模型将 RNN 与由知识库

(Knowledge Base, KB) 语义增强的键值记忆网络 (Key-Value Memory Networks) 相结合, 但是这种方法无法捕获 KB 中编码的实体之间高阶依赖。

总之, 基于 RNN 的会话推荐模型都存在短期记忆问题, 忽略了远距离项目间的复杂过渡关系, 尤其是对于较长的会话, 无法有效地建模项目之间的长期依赖关系, 无法捕捉到更早的历史行为对当前推荐的影响。

### 1.2.2 基于 GNN 的会话推荐模型

GNN 使用图数据结构对用户和项目信息进行建模, 在捕捉上下文信息、处理稀疏性和融合多种信息源复杂关系等方面具有突出的表现, 因而在基于会话的推荐领域受到了越来越多的关注。

Wu 等人<sup>[17]</sup>第一次在会话推荐中使用了 GNN, 提出了一种基于 GNN 的会话推荐模型 SR-GNN。该模型将会话中的项目序列建模为会话图, 并且利用图神经网络捕获项目之间的复杂过渡关系, 以提高会话推荐的准确性。Qiu 等人<sup>[18]</sup>通过构建会话图来研究项目过渡模式, 并提出了一种协同考虑会话图中序列顺序和潜在顺序的全图神经网络模型 FGNN, 该模型使用多个加权图注意力网络来学习会话中项过渡模式的固有顺序。Xu 等人<sup>[19]</sup>提出了一种图上下文化自注意力网络模型 GC-SAN, 该模型利用自注意机制和 GNN 分别学习局部依赖和全局依赖。Yu 等人<sup>[20]</sup>考虑了预测目标项, 提出了一种目标意图神经网络模型 TAGNN, 该模型首先使用图神经网络学习项目嵌入, 然后针对不同的目标项目自适应地学习不同的用户兴趣。

然而, 在基于会话的推荐中, 传统的图神经网络只能通过节点和边建模一对一的成对关系, 不能有效地建模多对多的超成对关系, 从而不能捕获项目间的高阶复杂关系, 限制了推荐性能的提高。

## 1.3 本文工作概述

本文的研究课题是“结合超图卷积网络和图注意力网络的会话推荐模型”, 核心任务是设计并实现一种结合超图卷积网络和图注意力网络的会话推荐模型, 并对该模型的推荐性能进行实验评价, 以验证推荐性能的改善。具体工作如下:

(1) 本文提出了一种结合超图卷积网络和图注意力网络的会话推荐模型 (Session-based Recommendation Model Combining Hypergraph Convolutional Networks with Graph Attention Networks, 简称 SR-HCGAT)。该模型利用 HGCN 从

根据所有会话信息构建的超图中捕获项目层的用户偏好，利用 GAT 从根据所有会话信息构建的会话图中捕获会话层的用户偏好，对包含互补信息的上述两种用户偏好进行联合学习，以更准确地捕获用户偏好。利用学得的项目层用户偏好产生推荐，最终达到改善推荐模型性能的目的。

(2) 本文在 Tmall、NowPlaying、Diginetica 三个真实世界数据集上采用标准性能度量指标对 SR-HCGAT 模型进行推荐性能实验评价，通过将 SR-HCGAT 与 SR-GNN、FGNN、DHCN 等几个前沿的基线模型进行性能对比，验证了 SR-HCGAT 模型具有更优的推荐性能。

## 1.4 论文组织结构

本文一共分为五章：

第一章，绪论。本章介绍了课题的背景和意义，介绍了会话推荐模型研究现状，概述了本文主要工作。

第二章，背景知识与支撑技术。本章介绍了超图卷积网络的概念和工作原理；介绍了图注意力网络的概念和工作原理；介绍了会话推荐的基本概念、技术和目前流行的会话推荐模型。

第三章，结合 HGCN 和 GAT 的会话推荐模型及其实现。本章详细描述了本文提出的 SR-HCGAT 模型，包括引言、问题陈述及符号定义、模型的总体框架、模型各层的工作原理以及模型的实现。

第四章，推荐性能实验评价。本章介绍了 SR-HCGAT 模型的实验设置，给出并讨论了实验结果，验证了推荐性能的改善并得出科学结论。

第五章，总结与展望。本章对本文工作进行总结，并对下一步工作进行展望。

## 第2章 背景知识与支撑技术

本章将介绍：超图卷积网络的概念和工作原理；图注意力网络的概念和工作原理；会话推荐的概念、技术和目前流行的会话推荐模型。

### 2.1 超图卷积网络

在普通的图结构中，每条边只能连接两个节点。在超图中，每条超边可以连接多个节点。相比传统图卷积网络只能处理二元关系的图数据，超图卷积网络（Hypergraph Convolutional Network, HGCN）<sup>[21]</sup>可以更好地捕捉多元关系数据中的复杂交互模式。

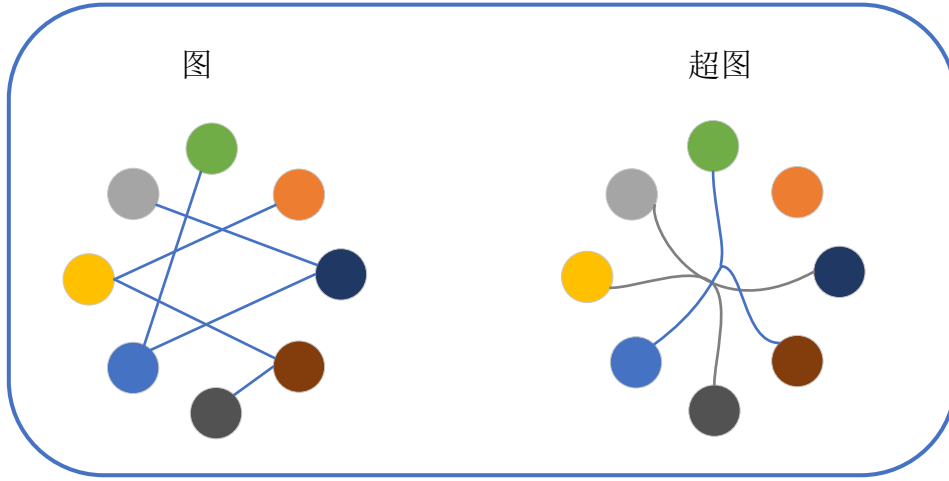


图 2.1 图和超图的示例图

一个超图可以定义为  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ ，其中  $\mathcal{V}$  表示节点的集合， $\mathcal{E}$  表示超边的集合， $\mathbf{W}$  是超边权重的对角矩阵，它为每一条超边都分配了一个权重。这样，超图  $\mathcal{G}$  可以表示为一个  $|\mathcal{V}| \times |\mathcal{E}|$  的关联矩阵  $\mathbf{H}$ ，其中的每一项由公式（2.1）定义：

$$h(v, e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases} \quad (2.1)$$

在超图中，一个节点  $v \in \mathcal{V}$  的度可以定义为  $d(v) = \sum_{e \in \mathcal{E}} \omega(e) h(v, e)$ ，一条边  $e \in \mathcal{E}$  的度可以定义为  $\delta(e) = \sum_{v \in \mathcal{V}} h(v, e)$ ， $\mathbf{D}_e$  与  $\mathbf{D}_v$  可用来分别表示边度和节点度的对角矩阵。有了公式（2.1）的定义，超图上的节点分类问题可以被表述为一个正则化框架，如公式（2.2）所示：

$$\operatorname{argmin}_f \{ \mathcal{R}_{emp}(f) + \Omega(f) \} \quad (2.2)$$

其中  $\Omega(f)$  为超图上的正则化,  $\mathcal{R}_{emp}(f)$  为监督经验损失,  $f(\cdot)$  是分类函数。正则化  $\Omega(f)$  可由公式 (2.3) 定义:

$$\Omega(f) = \frac{1}{2} \sum_{e \in \mathcal{E}} \sum_{u, v \in \mathcal{V}} \frac{\omega(e) h(u, e) h(v, e)}{\delta(e)} \left( \frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right) \quad (2.3)$$

令  $\theta = \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2}$ ,  $\Lambda = \mathbf{I} - \Theta$ , 则归一化的  $\Omega(f)$  可由公式 (2.4) 表示:

$$\Omega(f) = f^\top \Lambda \quad (2.4)$$

其中  $\Lambda$  是半正定的, 通常被称为超图拉普拉斯算子<sup>[22]</sup>。

由于  $\Lambda$  是一个  $n \times n$  的半正定矩阵, 因此可以利用特征分解  $\Lambda = \Phi \Lambda \Phi^\top$  得到标准正交特征向量  $\Phi = \text{diag}(\phi_1, \phi_2, \dots, \phi_n)$  和包含相应非负特征的对角矩阵  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ , 超图中信号  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  的傅里叶变换可以被定义为  $\hat{\mathbf{x}} = \Phi^\top \mathbf{x}$ , 特征向量被视为傅里叶基, 特征值则被视为频率。信号  $\mathbf{x}$  与滤波器  $\mathbf{g}$  的频谱卷积可表示为公式 (2.5):

$$\mathbf{g} * \mathbf{x} = \Phi \left( (\Phi^\top \mathbf{g}) \odot (\Phi^\top \mathbf{x}) \right) = \Phi \mathbf{g}(\Lambda) \Phi^\top \mathbf{x} \quad (2.5)$$

$\odot$  表示元素间的 Hadamard 积,  $\mathbf{g}(\Lambda) = \text{diag}(g(\lambda_1), g(\lambda_2), \dots, g(\lambda_n))$  是傅里叶系数的函数。

但傅里叶计算的计算代价为  $\mathcal{O}(n^2)$ , 为了解决计算代价高的问题, 可以使用一个  $K$  阶多项式对  $\mathbf{g}(\Lambda)$  进行参数化, 使用截断的切比雪夫展开作为多项式。 $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ , 初始  $T_0(x) = 1, T_1(x) = x$ 。于是  $\mathbf{g}(\Lambda)$  就可以参数化, 如公式 (2.6) 所示:

$$\mathbf{g} * \mathbf{x} \approx \sum_{k=0}^K \theta_k T_k(\tilde{\Lambda}) \mathbf{x} \quad (2.6)$$

其中  $T_k(\tilde{\Lambda})$  为  $k$  阶切比雪夫多项式, 其拉普拉斯式  $\tilde{\Lambda} = \frac{2}{\lambda_{\max}} \Lambda - \mathbf{I}$ 。

上式中只考虑了矩阵幂、加法和乘法，无疑增加了计算复杂度。可以进一步让  $K=1$  来限制卷积运算的阶数，因为超图中的拉普拉斯算子已经可以很好地表示节点之间的高阶相关性，如此卷积操作可进一步简化为公式 (2.7)：

$$\mathbf{g} * \mathbf{x} \approx \theta_0 \mathbf{x} - \theta_1 \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \mathbf{x} \quad (2.7)$$

其中， $\theta_0$  和  $\theta_1$  是所有节点滤波器的参数，可以进一步使用单参数  $\theta$  来避免过拟合问题，如公式 (2.8) 所示：

$$\begin{cases} \theta_1 = -\frac{1}{2} \theta \\ \theta_0 = \frac{1}{2} \theta \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \end{cases} \quad (2.8)$$

如此一来，卷积操作可进一步简化为公式 (2.9)：

$$\begin{aligned} \mathbf{g} * \mathbf{x} &\approx \frac{1}{2} \theta \mathbf{D}_v^{-1/2} \mathbf{H} (\mathbf{W} + \mathbf{I}) \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \mathbf{x} \\ &\approx \theta \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \mathbf{x} \end{aligned} \quad (2.9)$$

$(\mathbf{W} + \mathbf{I})$  可视为超边的权值， $\mathbf{W}$  可初始化为单位矩阵，即意味着所有超边权值相等。

当存在一个具有  $n$  个节点和  $C_1$  维特征的超图信号  $\mathbf{X} \in \mathbb{R}^{n \times C_1}$ ，超边卷积可公式化为公式 (2.10)：

$$\mathbf{Y} = \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \mathbf{X} \Theta \quad (2.10)$$

将这些超边组连接起来生成超图邻接矩阵  $\mathbf{H}$ 。构造超边卷积层  $f(\mathbf{X}, \mathbf{W}, \Theta)$  的过程可以用公式 (2.11) 进行描述：

$$\mathbf{X}^{(l+1)} = \sigma(\mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-1/2} \mathbf{X}^{(l)} \Theta^{(l)}) \quad (2.11)$$

式中， $\mathbf{X}^{(l)} \in \mathbb{R}^{N \times C}$  是超图在第  $l$  层的信号， $\mathbf{X}^{(0)} = \mathbf{X}$ ， $\sigma$  代表了非线性激活函数。

总的来说，超图卷积网络可以通过学习滤波器矩阵  $\Theta^{(l)}$  对初始节点特征  $\mathbf{X}^{(l)}$  进行处理，提取出  $C_2$  维特征；然后超边捕获节点特征，并形成超边特征  $\mathbb{R}^{C_2 \times E}$ ；最后将它们相关的超边特征聚合得到输出节点特征，通过与矩阵  $\mathbf{H}$  相乘得到输出节点特征，以表明在上式中  $\mathbf{D}_v$  和  $\mathbf{D}_e$  起归一化作用。据此，通过对节点和超边进行变换，超图卷积网络可以更好地捕获项目间的超成对关系。



## 2.2 图注意力网络

注意力机制是一种用于有序处理事件的方法，它能够使个体在面对事物时聚焦于最重要的部分，从而提取目标特征。图注意力网络（Graph Attention Network, GAT）<sup>[23]</sup>是一种使用注意力机制的图神经网络变体，引入注意力机制计算图中边的权重意味着在计算相邻节点之间的关系时，可以充分根据其重要性进行学习，帮助神经网络更有效地筛选信息，减少计算时的数据量，从而提高结果的准确性。下面将先描述单个图注意力层的构造。

图注意力层的输入是一组节点特征  $\mathbf{h} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}$ ， $\vec{h}_i \in \mathbb{R}^F$ ，其中  $N$  是节点数量， $F$  是每个节点的特征数量。该层将产生一组新的（可能具有不同基数  $F'$ ）节点特征  $\mathbf{h}' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_N\}$ ， $\vec{h}'_i \in \mathbb{R}^{F'}$  作为其输出。

首先，将权重矩阵  $\mathbf{W} \in \mathbb{R}^{F' \times F}$  参数化的共享线性变换应用至各节点，以使线性变换输入特征转换为更高级的特征。然后在这些节点上进行一种共享注意力机制  $a: \mathbb{R}^{F'} \times \mathbb{R}^F \rightarrow \mathbb{R}$ ，其返回注意力系数，计算公式如（2.12）所示：

$$e_{ij} = a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j) \quad (2.12)$$

注意力系数可用于表明节点  $j$  的特征对节点  $i$  的重要性，可以将其视为邻边的每个权重，它指明了应该给予特定节点多少关注。然而，不同节点在计算注意力系数时可能会得到差异很大的值，为了使系数在不同节点之间易于比较，需要对它们进行归一化，如（2.13）所示：

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})} \quad (2.13)$$

softmax 函数使得值分布在 0 至 1 之间，并且保持其总和为 1，归一化后的值用  $\alpha$  表示。既可以对所有节点对计算注意力系数，也可以只关注邻居节点，因后者操作屏蔽了其余所有节点故而被称作掩蔽注意力，下面将探讨注意力机制如何工作。

注意力机制  $\alpha$  是一个单层前馈神经网络，由权重向量  $\vec{\mathbf{a}} \in \mathbb{R}^{2F'}$  参数化，并应用 LeakyReLU 非线性激活函数。完整注意力机制公式如（2.14）所示：

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\vec{\mathbf{a}}^T [\mathbf{W}\vec{h}_i \parallel \mathbf{W}\vec{h}_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\vec{\mathbf{a}}^T [\mathbf{W}\vec{h}_i \parallel \mathbf{W}\vec{h}_k]))} \quad (2.14)$$



其中,  $\cdot^T$  表示转置,  $\parallel$  表示连接运算。式中识别出作为输入传入的两个转换后的节点嵌入, 将两个节点特征向量进行连接运算, 当它们通过一个单层神经网络时与权重向量相乘。此外, 每个输出都应用了 LeakyReLU 非线性激活函数以强调节点间的正相关性, 即剔除负值而仅保留正值。

为了将图注意力系数包含在图神经网络层中, 可以将节点邻居状态与该边对应的注意力系数相乘, 该过程可视为对每个节点重要程度加权的节点特征向量的线性组合, 其结果是重要元素被放大, 次要元素被抑制。经过归一化处理的注意力系数将被视为节点的最终输出特征, 如公式 (2.15) 所示:

$$\vec{h}_i' = \sigma \left( \sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W} \vec{h}_j \right) \quad (2.15)$$

总之, 图注意力机制通过计算注意力权重来聚焦于节点之间的关系, 提取关键特征, 具有多头机制和局部性约束, 适用于各种图结构数据分析任务。它是处理图数据的重要工具, 对于图分析、预测和推荐等应用具有广泛的应用前景。

## 2.3 基于会话的推荐

### 2.3.1 会话推荐的基本概念

基于会话的推荐 (Session-Based Recommendation, SBR) [4] 考虑到用户在一个会话中的行为和兴趣变化, 更加关注用户当前的需求和上下文, 能够更好捕捉用户的短期兴趣和意图, 向用户提供更准确、更符合需求的推荐结果。

基于会话的推荐通过分析用户的历史行为会话信息来预测下一项[24,25]。用户 (User) 是对商品进行操作 (如: 点击、购买) 并接受推荐结果的主体, 所有的用户共同构成用户集  $U = \{u_1, u_2, \dots, u_{|U|}\}$ 。项目 (Item) 是被推荐的实体 (如: 一本书、一部电影)。数据集中的所有项目共同组成项目集  $V = \{v_1, v_2, \dots, v_{|V|}\}$ 。动作 (Action) 通常是用户对项目进行的操作, 用  $a$  表示一个动作, 其与一个唯一的 ID 和一系列属性相关联以提供自身属性信息。交互是会话中最基本的单元, 用  $o$  进行表示。一次交互由一个用户  $u$ , 一个项目  $v$  和该用户对该项目所采取的动作  $a$  组成, 即  $o = \langle u, v, a \rangle$ 。会话 (Session) 是一段连续时间内生成的非空有界交互列表, 用  $s$  表示, 即  $s = \{o_1, o_2, \dots, o_{|s|}\}$ 。基于会话的推荐模型的基本输入是用于推荐的部分已知

会话信息，其输出是预测的交互或预测的交互列表<sup>[26]</sup>。

在推荐系统中应用会话推荐的方式也都不尽相同，不同的推荐系统有各自的特征和形式化，此处根据权威综述<sup>[4]</sup>对会话推荐技术给出一个抽象级的形式化。

使用  $l = \{o_1, o_2, \dots, o_j, \dots, o_n\}$  表示包括  $n$  个交互的列表，每个交互由一个项目和相应的动作组成，在基于单一类型动作会话的推荐系统中，可以将每个交互简化为一个项目列表  $l_v$ ，即  $l_v = \{v_1, v_2, \dots, v_j, \dots, v_n\} (v_j \in V)$ ， $L$  是列表集，其中包含从候选项目集  $V$  和动作集  $A$  派生出的所有可能的交互列表。 $c$  是输入，由用于推荐的所有会话信息组成，所有会话上下文共同构成会话上下文集  $C$ 。若用  $f$  表示一个效用函数用于返回给定会话上下文  $c$  的候选列表集  $l$  的效用分数，则可以通过最大化效用分数选择推荐的交互列表  $\hat{l} \in L$ ，如公式 (2.16) 所示：

$$\hat{l} = \operatorname{argmax} f(c, l) \quad (2.16)$$

其中， $c \in C$ ， $l \in L$  效用函数可以指定为多种形式，如似然、条件概率等。在交互列表上使用效用函数来优化整个候选列表，而不是优化单个候选交互（项），可以从多个方面对整个列表进行表征和评估。

神经网络在处理序列数据和图结构数据<sup>[27]</sup>方面表现突出，是实现会话推荐技术的强大工具，在建模会话内的项目之间的依赖和提取实体高阶特征表示方面具有天然的优势，下小节将介绍主流的基于深度神经网络的会话推荐模型。

### 2.3.2 会话推荐技术与模型

目前已经有不少学者将神经网络应用到基于会话的推荐中，并取得许多突破性的成果。本小节将描述三个典型的基于深度神经网络的会话推荐模型，具体包括 GRU4Rec<sup>[9]</sup>、SR-GNN<sup>[17]</sup>和 FGNN<sup>[18]</sup>。

(1) GRU4Rec (Gated Recurrent Unit for Recommender Systems)<sup>[9]</sup>：该模型是一种经典的基于循环神经网络 (Recurrent Neural Network, RNN) 的会话推荐模型，其核心是门控循环单元 (Gated Recurrent Unit, GRU)，它是一种 RNN 变体，具有门控机制，可以更好地捕捉序列数据中的长期依赖关系。GRU4Rec 首先对所有项目进行嵌入，将每个项目映射为一个低维度的向量表示。然后使用隐藏状态来表示会话的上下文信息和用户的兴趣偏好。最后 GRU4Rec 会输出当前会话中最可能交互的下一个项目。为了提高模型的训练效率，GRU4Rec 使用了会话并行的小批量

训练方法来优化模型的训练。但是 GRU4Rec 仅仅建模当前会话中项目之间的单向过渡关系，限制了推荐性能的提高。

(2) SR-GNN (Session-based Recommendation with Graph Neural Networks) [17]: 该模型是一种基于 GNN 的会话推荐模型。SR-GNN 使用 GNN 层来学习项目之间的关系和上下文信息。具体来说, SR-GNN 将每个会话图逐个处理, 并通过门控图神经网络 (Gated Graph Sequence Neural Networks, GGNN) [28]来更新会话图上的节点嵌入。接着, SR-GNN 使用注意力网络将会话的全局嵌入和局部嵌入组合成会话嵌入来提供推荐。SR-GNN 是第一个提出了将会话序列转化为会话图的思想并成功地将 GNN 运用到了基于会话的推荐中, 但是该模型仅仅利用了项目之间的过渡关系忽略了项目之间的高阶关系, 限制了推荐性能的提高。

(3) FGNN (Full Graph Neural Network) [18]: 该模型是一种基于 GNN 的会话推荐模型。具体来说, FGNN 将输入的会话的项目序列转换为具有节点的嵌入的会话图, 并提出了一种多层加权图注意力网络 (Weighted Graph Attention, WGAT) 作为会话图节点的嵌入编码器来计算会话内项目之间的信息流。经过 WGAT 处理后, 会话图中的节点的嵌入包含了不同语义信息。接着, FGNN 使用一个读出函数来获得会话嵌入用于推荐。FGNN 模型利用了全局信息, 但在每个图卷积层中只考虑节点的一阶邻居信息, 无法捕捉长距离的节点关系, 限制了推荐性能的提高。

## 第3章 结合 HGCN 和 GAT 的会话推荐模型及其实现

### 3.1 引言

近年来,图神经网络逐渐成为解决图数据结构建模问题的有效工具,但图神经网络仅仅将项目之间的过渡简单地视为两两成对关系,忽略了项目之间复杂的高阶信息。超图将节点连接关系扩展到超边(Hyperedge),一条超边可以包含多个节点,从而能够更准确地描述节点之间的关联性和语义信息<sup>[29]</sup>。因此,利用超图卷积网络能够更好地建模节点之间复杂的高阶关系。但仅使用超图卷积网络还无法充分考虑不同节点之间不同程度的重要性。图注意力网络通过自适应的注意力机制,能够动态地学习节点之间的关注度,从而更细粒度地聚合邻居节点的特征。因此,可以将图注意力网络与超图卷积网络结合,分别在项目层和会话层捕获两种互补的用户偏好,达到改善推荐性能的目的。

本章提出一种结合超图卷积网络和图注意力网络的会话推荐模型(SR-HCGAT),通过框架图和各层的描述对SR-HCGAT模型进行详细的介绍,并给出了该模型的实现方法。

### 3.2 问题陈述及符号定义

一个会话推荐系统包含了一个用户集 $U = \{u_1, u_2, \dots, u_{|U|}\}$ 和一个项目集 $V = \{v_1, v_2, \dots, v_{|V|}\}$ 。一个会话是由用户-项目交互(interaction)所组成的一个非空有界列表,即有一对开始-结束交互对来开始和结束一个会话,且会话中的交互(例如点击一个项目)是按时间顺序排序的或者是无序的。本章研究的是无序的会话,且会话中只包含一种类型的行为(action)。因此,对于一个具体的用户,该用户的每一个交互都可以被简化成该用户所交互的一个项目(item)。一个长度为 $n$ 的会话表示为一个序列 $s = [v_1, v_2, \dots, v_n]$ ,其中可能包含重复的项目,每个项目 $v_i \in V$ ,  $i = 1, 2, \dots, n$ 。进一步地,一个用户的 $m$ 个会话表示为 $S = \{s_i\}_{i=1}^m$ 。基于上述概念和符号,本章推荐任务的描述如下。

给定用户 $u$ 的 $m$ 个会话 $S = \{s_i\}_{i=1}^m$ ,下一项推荐(next-item recommendation)任务是推荐当前会话中用户可能交互的下一个项目 $v \in V$ 。具体来说,这个任务的目

标是学习一个预测函数  $\hat{y} = F(u, v | \Theta, S)$ 。其中  $\hat{y}$  表示用户  $u$  将会与项目  $v$  交互的预测概率， $\Theta$  表示函数  $F$  的模型参数。

### 3.3 SR-HCGAT 会话推荐模型

#### 3.3.1 模型概述

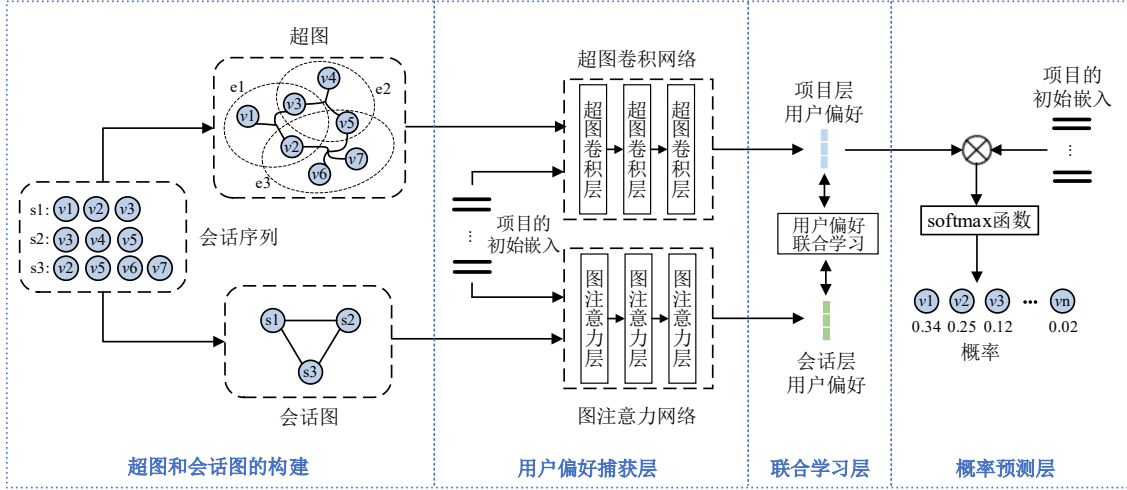


图 3.1 SR-HCGAT 模型框架图

图 3.1 展示了 SR-HCGAT 模型的总体框架图。如图所示，SR-HCGAT 模型的核心思想是利用超图卷积网络捕获项目层的用户偏好；利用图注意力网络捕获会话层的用户偏好；对两种互补的用户偏好进行联合学习以使用户偏好互信息最大化；最后使用项目层的用户偏好进行概率预测。SR-HCGAT 模型包含四个部分：超图和会话图的构建、用户偏好捕获层、联合学习层<sup>[30]</sup>和概率预测层，各层的简要描述如下：

- （1）超图和会话图的构建：该层利用历史会话序列构建超图和会话图。
- （2）用户偏好捕获层：该层利用超图卷积网络从超图中捕获项目层的用户偏好，利用图注意力网络从会话图中捕获会话层的用户偏好。
- （3）联合学习层：该层利用项目层的用户偏好和会话层的用户偏好进行联合学习，以更准确地捕获用户偏好。
- （4）概率预测层：该层利用优化后的项目层用户偏好和项目集中项目的初始嵌入进行下一项概率预测，生成每个项目的预测概率。

以下小节将对上述各层逐一进行详细描述。

### 3.3.2 超图和会话图的构建

#### 3.3.2.1 超图的构建

为了捕获项目之间的高阶关系，本文使用  $G=(V,E)$  表示一个超图，其中  $V$  是包含了  $N$  个不同节点的点集， $E$  是包含了  $M$  条超边的边集。每一条超边  $\epsilon \in E$  包含了两个或更多的节点，并且被分配了一个正权重  $W_{\epsilon\epsilon}$ ，所有的权重共同组成了一个对角矩阵  $\mathbf{W} \in \mathbb{R}^{M \times M}$ 。超图可以表示为一个关联矩阵  $\mathbf{H} \in \mathbb{R}^{M \times N}$ ，如果超边  $\epsilon \in E$  包含节点  $v_i \in V$ ，则  $H_{i\epsilon} = 1$ ，否则  $H_{i\epsilon} = 0$ 。对于每一个节点和每条超边，它们的度  $D_{ii}$  和  $B_{\epsilon\epsilon}$  分别被表示为  $D_{ii} = \sum_{\epsilon=1}^M W_{\epsilon\epsilon} H_{i\epsilon}$ ， $B_{\epsilon\epsilon} = \sum_{i=1}^N H_{i\epsilon}$ ，其中  $\mathbf{D}$  和  $\mathbf{B}$  都是对角矩阵。

在超图中，每个会话被表示为一条超边，每条超边可以表示为  $[i_{s,1}, i_{s,2}, \dots, i_{s,m}] \in E$ ，每个项目可以表示为  $i_{s,m} \in V$ 。原始会话数据被组织为线性序列，只有当用户  $i_{s,m}$  之前的项目  $i_{s,m-1}$  交互时，两个项目  $i_{s,m-1}, i_{s,m}$  才会连接。将会话数据转换成超图后，一条超边所包含的各个顶点之间是相互连接的，这也意味着在会话中用户交互的任意两个项目都是连接的，因此超图包含项目之间的高阶不成对关系，可用于全局级别的项目特征表示。

#### 3.3.2.2 会话图的构建

为了捕获会话层的用户偏好，SR-HCGAT 模型利用所有的会话序列构建会话图，使用图注意力网络来更新会话图的节点嵌入。

对于给定的会话序列，可以将其对应的会话图描述为  $S(G)=(V,E)$ 。其中， $S(G)$  的每个节点都表示一个会话，如果两个不同的会话中包含至少一个公共项目，则在  $S(G)$  中将这两个会话分别对应的两个节点用边相连，并根据它们公共节点的数量为该边分配一个权重。形式上，会话图可以定义为  $S(G)=(V_L, E_L)$ ，其中  $V_L = \{v_e : v_e \in E\}$ ， $E_L = \{(v_e, v_q) : e_p, e_q \in E, |e_p \cap e_q| \geq 1\}$ ，每条边的权重表示为  $w_{p,q} = |e_p \cap e_q| / |e_p \cup e_q|$ ，即：(两个会话图的公共节点数目) / (两个会话图总共的节点数目)。

### 3.3.3 用户偏好捕获层

#### 3.3.3.1 项目层用户偏好

构建超图之后，利用超图卷积网络<sup>[21]</sup>从超图中捕获项目之间的高阶关系，在超图上定义卷积操作的主要任务是如何去传播邻居节点的信息。根据 2.1 小节的内容，可以将超图的谱卷积操作定义为公式 (3.1)：

$$\mathbf{x}_i^{(l+1)} = \sum_{j=1}^N \sum_{\epsilon=1}^M H_{i\epsilon} H_{j\epsilon} W_{\epsilon\epsilon} \mathbf{x}_j^{(l)} \quad (3.1)$$

有别于 2.1 小节的介绍，此处并没有使用非线性激活函数和卷积滤波器参数矩阵，此处对所有超边赋予相同的权重，即  $W_{\epsilon\epsilon} = 1$ 。经过归一化后的矩阵表达式，如公式 (3.2) 所示：

$$\mathbf{X}_h^{(l+1)} = \mathbf{D}^{-1} \mathbf{H} \mathbf{W} \mathbf{B}^{-1} \mathbf{H}^T \mathbf{X}_h^l \quad (3.2)$$

超图卷积可以看作是对超图结构进行“节点-超边-节点”特征变换的两个阶段细化，乘法操作  $\mathbf{H}^T \mathbf{X}_h^l$  定义了从节点到超边的信息聚合，然后通过预乘  $\mathbf{H}$  将超边到节点的信息进行聚合。通过第  $L$  个超图卷积层传递  $\mathbf{X}^{(0)}$  后，对每层得到的项目嵌入进行平均即可得到最终的项目嵌入，计算公式如 (3.3) 所示：

$$\mathbf{X}_h = \frac{1}{L+1} \sum_{l=0}^L \mathbf{X}_h^{(l)} \quad (3.3)$$

为了获得更好的推荐效果，时间信息对于推荐结果的影响是不可忽略的。位置嵌入是 Transformer<sup>[31]</sup>中引入的一种有效技术，用于记忆物品的位置信息。可以通过一个可学习的位置矩阵  $\mathbf{P}_r = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_m]$  将反向位置嵌入与学习到的项目表示进行整合，其中  $m$  为当前会话的长度。会话  $s = [i_{s,1}, i_{s,2}, i_{s,3}, \dots, i_{s,m}]$  的第  $t$  项的嵌入可由公式 (3.4) 表示：

$$\mathbf{x}_t^* = \tanh(\mathbf{W}_1[x_t \parallel \mathbf{p}_{m-i+1}] + b) \quad (3.4)$$

其中  $\mathbf{W}_1 \in \mathbb{R}^{d \times 2d}$ ， $b \in \mathbb{R}^d$  是可学习的参数。会话嵌入可以通过聚合该会话中项目的表示来获得，可以通过 SR-GNN 中使用的策略<sup>[17]</sup>改进会话  $s$ ，如公式 (3.5) 所示：

$$\alpha_t = \mathbf{f}^T \sigma(\mathbf{W}_2 \mathbf{x}_s^* + \mathbf{W}_3 \mathbf{x}_t^* + c), \theta_h = \sum_{t=1}^m \alpha_t \mathbf{x}_t^* \quad (3.5)$$



其中  $\mathbf{x}_s^*$  是会话  $s$  的嵌入，这里通过对其包含的项目嵌入进行平均，如公式 (3.6) 所示：

$$\mathbf{x}_s^* = \frac{1}{m} \sum_{t=1}^m \mathbf{x}_t \quad (3.6)$$

$\mathbf{x}_t^*$  是会话  $s$  中第  $t$  个项目的嵌入。用户在会话的一般嵌入  $\mathbf{0}_h$  是通过软注意机制聚合项目嵌入进行表示的，其中的项目具有不同的优先级。 $\mathbf{f} \in \mathbb{R}^d$ ,  $\mathbf{W}_2 \in \mathbb{R}^{d \times d}$  以及  $\mathbf{W}_3 \in \mathbb{R}^{d \times d}$  是用于学习项目权重的注意参数。

### 3.3.3.2 会话层用户偏好

该层采用图注意力网络对会话图的节点嵌入进行聚合操作并更新节点嵌入，以获得聚合后的会话嵌入表示。

基于节点嵌入  $\mathbf{q}_i, \mathbf{q}_j$  和权重  $w_{i,j}$ ，图注意力网络计算一个注意力系数  $e_{i,j}$ ，该注意力系数决定了节点  $i$  对节点  $j$  的重要程度，由公式 (3.7) 定义：

$$e_{ij} = \text{LeakyReLU}\left(\mathbf{w}_{att}^T \left[ \mathbf{W}\mathbf{q}_i \parallel \mathbf{W}\mathbf{q}_j \right]\right) \quad (3.7)$$

其中， $\mathbf{W}$  和  $\mathbf{w}_{att}$  是权重参数，它们是通过参数学习进行确定的， $\parallel$  是向量拼接操作， $\cdot^T$  是矩阵转置操作。

为了直观地比较不同节点的重要程度，该层使用了一个 softmax 函数将注意力系数转化为概率形式，计算过程由公式 (3.8) 定义：

$$a_{ij} = \text{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})} \quad (3.8)$$

其中  $\mathcal{N}(i)$  是节点  $i$  的邻居节点集。该层使用一个线性组合将转换后的注意力系数与相应的邻居节点的嵌入进行组合以更新节点  $i$  的嵌入，更新过程由公式 (3.9) [18] 定义：

$$\mathbf{q}'_i = \text{ReLU}\left(\sum_{j \in \mathcal{N}(i)} a_{ij} \mathbf{W}\mathbf{q}_j\right) \quad (3.9)$$

其中，ReLU 是一个激活函数。

通过以上步骤，实现了将邻居节点的信息融合到目标节点的嵌入中并且更新了



所有节点嵌入，多层图注意力网络的迭代更新最终得到会话嵌入结果。注意力系数对不同的邻居节点赋予不同的重要性，根据节点之间的关系进行自适应的特征聚合，从而提高模型对图结构的建模能力。

### 3.3.4 联合学习层

联合学习层利用自监督信号，对项目层用户偏好和会话层用户偏好中获得的会话嵌入进行联合学习，更准确的捕获用户偏好。

创造自我监督<sup>[30]</sup>的信号：到目前为止，通过捕获项目层用户偏好和捕获会话层用户偏好的操作获得了两组特定的会话嵌入。超图卷积网络对项目层所构建的超图进行处理并得到会话嵌入；图注意力网络对会话层的会话图进行处理并得到会话嵌入；项目层（会话内）的会话嵌入和会话层（会话间）的会话嵌入可用于相互补充。SR-HCGAT 通过计算两个会话嵌入之间的相似性得分来判断它们是否表示相同的会话。如果得分高于某个阈值，则认为两个嵌入表示相同的会话，将其标记为正标签。如果得分低于阈值，则认为两个嵌入表示不同的会话，将其标记为负标签。

联合学习<sup>[30]</sup>：在每个训练批次中，SR-HCGAT 迭代重建正负伪标签（正负样本）用于对项目层的用户偏好和会话层的用户偏好之间的联合学习，迭代这个过程会产生更多提供有用信息的（informative）样本，从而使得 SR-HCGAT 可以更准确地挖掘项目层的用户偏好和会话层的用户偏好。SR-HCGAT 采用 InfoNCE 损失函数<sup>[32]</sup>作为联合学习的损失函数，它是正样本和负样本之间的标准二元交叉熵，由公式（3.10）定义：

$$\mathcal{L}_s = -\log \sigma \left( f_D \left( \theta_i^h, \tilde{\theta}_i^l \right) \right) - \log \sigma \left( 1 - f_D \left( \tilde{\theta}_i^h, \theta_i^l \right) \right) \quad (3.10)$$

其中， $\tilde{\theta}_i^h$  是负样本， $f_D(\cdot): \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  是将对一致性进行判别的区分函数，SR-HCGAT 通过向量间点积的方式来实现该区分函数，这个学习目标可以解释为最大化从超图和会话图下学习的会话嵌入之间的互信息，从而提高在项目/会话特征提取方面的性能。

最后，将推荐任务作为主任务，自监督任务作为辅任务去增强推荐任务，联合学习的目标可由公式（3.11）<sup>[30]</sup>定义：

$$\mathcal{L} = \mathcal{L}_r + \gamma \mathcal{L}_s \quad (3.11)$$

其中  $\gamma$  控制了自监督任务的强弱程度。

### 3.3.5 概率预测层

由于超图可以更好的捕获项目之间的高阶关系，因此 SR-HCGAT 模型使用联合学习后的项目层的用户偏好来预测最终的候选项目以进行推荐。给定一个会话  $s$ ，SR-HCGAT 通过项目层的用户偏好  $\theta_h$  和项目的初始嵌入  $\mathbf{x}_i$  进行点积运算来获得候选物品得分向量  $\hat{\mathbf{z}}$ ，可由公式 (3.12) 定义：

$$\hat{\mathbf{z}}_i = \theta_h^T \mathbf{x}_i \quad (3.12)$$

这意味着每个候选物品都会与学习到的物品嵌入进行内积运算，从而得到一个与该候选物品相关的得分。获得分数向量  $\hat{\mathbf{z}}$  之后，该层使用一个 softmax 函数对  $\hat{\mathbf{z}}$  进行归一化<sup>[33]</sup>，如公式 (3.13) 所示：

$$\hat{\mathbf{y}} = \text{softmax}(\hat{\mathbf{z}}) \quad (3.13)$$

从而生成项目集中的所有项目的下一项概率向量，其中每一个分量表示相应项目是用户的下一个交互项目的概率。

### 3.3.6 模型训练

SR-HCGAT 模型的学习目标被制定为交叉熵损失函数。该函数由公式 (3.14) 定义：

$$\mathcal{L}_r = -\sum_{i=1}^N \mathbf{y}_i \log(\hat{\mathbf{y}}_i) + (1 - \mathbf{y}_i) \log(1 - \hat{\mathbf{y}}_i) \quad (3.14)$$

其中， $\mathbf{y}$  是基准真实值的独热编码向量。为了简化操作，此处省略了  $L_2$  正则化项。

通过使用 Adam 算法最小化  $\mathcal{L}_r$ ，可以获得基于会话的高质量推荐。

## 3.4 模型实现

### 3.4.1 编程语言与开发工具

本文提出的 SR-HCGAT 模型的实现使用了如下编程语言、开发环境与软件工具：

- 编程语言：Python 3.10.11
- 集成开发环境（IDE）：PyCharm Professional 2023.1.2，下载地址：  
<https://www.jetbrains.com/pycharm/download/>

- 深度学习框架：PyTorch 2.0.0，下载地址：<https://pytorch.org/get-started/pytorch-2.0/>

### 3.4.2 用户偏好捕获层的实现

#### 3.4.2.1 捕获项目层用户偏好

为了捕获项目层用户偏好，SR-HCGAT 首先通过训练获得各个节点的初始项目嵌入，将初始项目嵌入进行保存并且作为输出的一部分；然后通过输入超图的邻接矩阵以表示节点间的连接关系（稀疏矩阵）；接着创建一个结果列表（空列表）用于存储每一层的节点表示，并根据初始指定的层数进行迭代超图卷积网络的操作，初始节点表示将被添加到结果列表。第一层及之后的节点迭代传播方法如下：

（1）进行多层超图卷积的迭代，通过超图的邻接矩阵与当前的项目嵌入进行稀疏矩阵乘法操作，以得到更新后的项目嵌入。

（2）应用激活函数对更新后的项目嵌入进行非线性变换。

（3）将非线性变换后的项目嵌入添加到结果列表。

对结果列表进行求和，可得到经更新和非线性变换后的所有项目嵌入的总和，将节点表示的总和进行平均池化即可得到最终的项目嵌入。通过将项目嵌入与位置嵌入（位置嵌入可以看作是会话序列中各个项目的序列信息的表征，从会话序列长度获取）进行拼接并加权求平均最终得到会话嵌入表示。加权过程利用了位置嵌入与项目嵌入之间的关系，提高了会话嵌入的表征能力。最后，超图卷积网络返回生成的会话嵌入作为其输出结果。

#### 3.4.2.2 捕获会话层用户偏好

在捕获会话层的用户偏好时，图注意力网络接受会话图节点的特征矩阵（会话图中每个节点的特征向量）和会话图节点的邻接矩阵（会话图连接关系）作为输入，并定义了权重矩阵和注意力参数矩阵。

在单层图注意力网络中，将节点特征矩阵与权重矩阵相乘以得到节点的特征变换矩阵。通过注意力机制计算节点与邻居节点的注意力权重，注意力计算步骤如下：

（1）将节点特征变换矩阵与其自身拷贝的矩阵拼接起来，形成一个新的扩展矩阵。

（2）将扩展矩阵与注意力参数矩阵相乘，并应用 LeakyReLU 激活函数得到注

注意力系数矩阵。

(3) 为了确保只在邻接节点之间计算注意力, 需要进行邻接矩阵的处理。通过创建一个与邻接矩阵相同形状的全为负无穷大的矩阵, 用注意力系数替换对应位置邻接矩阵中的正值, 最终得到修正后的注意力矩阵。

(4) 将注意力矩阵归一化处理, 得到每个节点对邻居节点最终归一化的注意力系数矩阵。

将注意力系数矩阵与节点的特征变换矩阵相乘, 得到注意力聚合后的节点表示, 即表示节点在信息传播过程中的新特征表示, 相当于每个节点将注意力权重分配给它的邻居节点, 并将邻居节点的特征进行加权求和。总的来说, 单层图注意力网络将会话图节点特征作为输入, 并通过注意力机制对节点进行自适应的特征聚合。

通过堆叠多层图注意力网络, 可以将经过注意力聚合后的节点特征表示一层层传递下去, 进行特征聚合和传播, 使节点特征逐渐融合和更新, 最终得到一个表示整个会话的特征向量。

### 3.4.3 联合学习层的实现

在 SR-HCGAT 模型中, 联合学习主要通过自监督学习任务来实现。模型使用了自生成对比性学习作为自监督学习的方法, 通过构造正负样本对来进行联合学习, 以更好捕获项目层的用户偏好和会话层的用户偏好:

(1) 自监督学习任务使用对比损失函数衡量正样本和负样本的相似度得分。对比损失函数采用了正样本和负样本的对比, 其中, 正样本是从会话中选取的项目嵌入与对应的会话嵌入之间的相似度, 负样本是进行行列置换操作后原始会话嵌入的顺序关系遭破坏生成的。其目标是最大化正样本相似度分数, 并最小化负样本相似度分数。

(2) 通过点积运算计算正样本和负样本之间的相似度得分, 来衡量它们在嵌入空间中的相似程度。然后将正样本和负样本的相似度得分输入到交叉熵损失函数以度量相似度得分以优化模型。

(3) 将自监督学习损失与其他损失项相乘(通过乘以超参数  $\beta$ ), 以加权考虑自监督学习的贡献。

(4) 最小化联合损失(包括自监督学习损失和其他损失项)以优化模型参数,

SR-HCGAT 使用反向传播算法和 Adam 优化器<sup>[34]</sup>进行参数更新。

### 3.4.4 概率预测层的实现

通过实现以上工作，得到项目嵌入和会话嵌入表示后，SR-HCGAT 将项目层的会话嵌入与初始项目嵌入进行点积运算，然后使用 softmax 函数对点积运算结果进行归一化，将其转化为概率分布，得到项目的概率预测：

(1) 概率预测层的输入是一个相似度得分矩阵，该矩阵的大小表示为会话数  $K$  乘以项目数  $N$ ，表示每个会话中每个项目的相似度得分。

(2) 对于每个会话（每行），softmax 函数将会话中每个项目的得分转换为一个介于 0 和 1 之间的概率值。softmax 函数的计算公式如 (3.15) 所示：

$$P(i|s) = \frac{e^{S(i,s)}}{\sum_{j=1}^N e^{S(j,s)}} \quad (3.15)$$

其中， $P(i|s)$  表示给定会话  $s$  下项目  $i$  被选择的概率， $S(i,s)$  是会话  $s$  下项目  $i$  的得分， $N$  是项目数量。

(3) 概率预测层的输出是一个概率分布矩阵，大小为  $K \times N$ ，其中每个元素表示对应会话下对应项目被选择的概率。

最后，根据预测分数和真实标签计算评估指标（如 Precision、MRR 等）。

### 3.4.5 模型训练的实现

如 3.4.4 节介绍，将项目层的会话嵌入与初始项目嵌入进行点积运算，得到一个相似度得分矩阵，可以根据预测分数和目标值计算损失，并返回损失结果。

首先，创建训练测试函数用于执行训练和测试过程。该函数接收模型、训练数据和测试数据作为输入，并返回评估指标和总损失，函数内部设置了用于训练的参数（如批量大小和评估指标）。然后，生成训练数据对象的批次数据索引，在每个训练批次中进行以下步骤：

- (1) 将模型的梯度置零，以准备进行反向传播。
- (2) 调用前向传播函数计算目标值、预测分数和附加损失。
- (3) 计算总损失，包括预测分数与目标值之间的差异和附加损失。
- (4) 执行反向传播，并通过 Adam 优化器更新模型的参数。
- (5) 对总损失进行累加求和。

在整个训练过程中，模型的参数会通过优化器进行更新，以最小化损失函数。

训练过程中的总损失会被累加，并在每个训练周期结束后进行输出。同时，测试阶段的评估指标会被计算和输出，用于评估模型的性能。

### 3.5 本章小结

为了更有效地发挥超图神经网络的潜力，解决固有的数据稀疏性问题，提高会话推荐的性能，本章创新地提出了一种结合 HGCN 和 GAT 的会话推荐模型，并取得了良好的性能改善。

本章首先结合框架图对 SR-HCGAT 模型进行概述，然后分别从用户偏好捕获层、联合学习层、概率预测层以及模型训练等各个层面详细介绍了 SR-HCGAT 模型，并给出了 SR-HCGAT 每个层面的具体实现步骤和工作过程。

## 第4章 推荐性能实验评价

### 4.1 引言

推荐性能实验评价的目标是通过系统地设计和执行实验，对不同的推荐模型进行客观、准确的比较和评估。实验评价不仅可以帮助研究者深入理解各种算法的优劣势，还能为推荐系统的实际应用提供指导和决策支持。

本章首先将详细介绍用于评估 SR-HCGAT 模型的数据集、性能评价指标、基线模型以及超参数设定，以帮助读者了解如何设计和执行推荐性能实验，并正确解释和分析实验结果。4.3 小节对 SR-HCGAT 模型的实验结果与基线模型的实验结果进行对比，根据各项指标进行综合评价，以验证推荐性能的改善。通过本章的分析与评价，可以更好地理解不同推荐模型在特定场景下的表现，并为推荐系统的进一步研究和应用提供指导和参考。

### 4.2 实验设置

#### 4.2.1 数据集

本文实验采用了 Tmall、Nowplaying、Diginetica 三个真实世界数据集来评价 SR-HCGAT 模型的性能。

(1) Tmall 数据集<sup>1</sup>：该数据集来自 IJCAI-15 大赛。该数据集包含匿名用户在“双 11”前及当天近 6 个月内的购物日志，以及是否为重复购买者的标签信息。

(2) Nowplaying 数据集<sup>2</sup>：该数据集是一个常用的音乐推荐领域的数据集，综合了用户在社交媒体平台上播放音乐的行为，提供了用户、曲目和时间戳等信息。

(3) Diginetica 数据集<sup>3</sup>：该数据集基于 Deli 数据集和 TaoBao 数据集构建而成，是一个用于会话推荐领域的公开数据集，包含了用户在电子商务网站上的行为记录。

上述数据集均已移除所有仅包含 1 个项目的会话和出现次数少于 5 次的项目<sup>[14]</sup>。数据集统计结果如表 4.1 所示：

---

<sup>1</sup> <https://tianchi.aliyun.com/dataset/42>

<sup>2</sup> <http://dbis-nowplaying.uibk.ac.at/#nowplaying>

<sup>3</sup> <https://competitions.codalab.org/competitions/11161>

表 4.1 数据集信息

数据集	Tmall	Nowplaying	Diginetica
训练会话数	351,268	825,304	719,470
测试会话数	25,898	89,824	60,858
项目数	40,728	60,417	43,097
平均长度	6.69	7.42	5.12

#### 4.2.2 性能评价指标

本章实验采用精度  $P@K$  (Precision)和平均倒数排名  $MRR@K$  (Mean Reciprocal Rank)这两种流行的准确性评价指标<sup>[35]</sup>来评估推荐模型的性能。

(1) 精度：精度用于衡量模型预测的准确性，是信息检索和机器学习中常用的评价指标之一。该指标主要关注推荐列表中前  $K$  个推荐项目中真正与用户兴趣相符的数目。精度可以帮助我们了解推荐系统在给用户推荐物品时的准确程度。该指标由公式 (4.1) 定义：

$$P@K = \frac{1}{m} \sum_{u=1}^m \frac{C_u(K) \cap GT_u}{C_u(K)} \quad (4.1)$$

其中， $K$  表示推荐项目的个数， $m$  表示用户数量， $GT_u$  表示与用户相关的所有项目的集合， $C_u(K)$  表示用户  $u$  的 top- $K$  推荐列表。较高的  $P@K$  值表示在给定的返回结果中有更多的准确结果。通常会将  $P@K$  与其他指标结合使用，以提供对系统综合性能的更全面评估。

(2) 平均倒数排名：平均倒数排名是一种用于评估信息检索或推荐系统的指标，衡量在前  $K$  个返回结果中的排序质量。它考虑了第一个准确结果出现的位置，并计算其倒数作为指标值，最终的分数为所有得分之和。平均倒数排名基于物品的排名位置计算平均倒数排名，其值越接近 1 表示排序质量越好，该指标由公式 (4.2) 定义：

$$MRR@K = \frac{1}{m} \sum_{u=1}^m \sum_{v \in GT_u, rank_v \leq K} \frac{1}{rank_v} \quad (4.2)$$

其中， $K$  表示推荐项目的个数， $m$  表示用户数量， $GT_u$  表示与用户相关的所有项目



集合,  $rank_v$  表示项目  $v$  在用户  $u$  的推荐列表排名。MRR@K 指标的优点是能够考虑到推荐物品的排名信息, 更加关注用户对于推荐结果的排序质量, 而不仅仅是是否推荐了正确的物品, 因此 MRR@K 可以作为评估推荐系统排序性能的重要指标之一。

#### 4.2.3 基线模型

为了验证 SR-HCGAT 模型的推荐性能, 本文将 SR-HCGAT 模型与七个具有代表性的会话推荐模型进行比较。这些推荐模型分为两类: 基于 RNN 的推荐模型 (GRU4REC、NARM、STAMP), 基于 GNN 的推荐模型 (SR-GNN、FGNN、DHCN、FSM)。这七个推荐模型的简要描述如下:

(1) GRU4REC<sup>[9]1</sup>: 该模型通过使用更新门和重置门来控制信息的流动和遗忘, 从而更好地捕捉长期依赖关系和序列模式、会话中的上下文信息和用户的兴趣演化。

(2) NARM<sup>[14]2</sup>: 该模型通过使用注意力机制对序列中的每个项进行加权, 可以自动地关注最相关和有意义的部分。

(3) STAMP<sup>[15]3</sup>: 该模型是一个混合推荐模型, 通过构建两个网络结构对会话序列的重要元素进行加权编码和优先级更新。

(4) SR-GNN<sup>[17]4</sup>: 该模型将会话序列表示为会话图, 并使用 GNN 对项目之间的复杂过渡关系进行建模。

(5) FGNN<sup>[18]5</sup>: 该模型协同考虑会话图中的顺序和潜在顺序, 并将会话推荐形式化为图分类问题。

(6) DHCN<sup>[30]6</sup>: 该模型是一个先进的会话推荐模型, 利用一个双通道超图卷积网络来捕获不同用户当前会话中项目之间的超成对关系。

(7) FSM<sup>[36]7</sup>: 该模型是一种用于识别序列数据<sup>[37]</sup>中频繁出现的子序列模式的

<sup>1</sup> <https://github.com/hidasib/GRU4Rec>

<sup>2</sup> [https://github.com/lijingsdu/sessionRec\\_NARM](https://github.com/lijingsdu/sessionRec_NARM)

<sup>3</sup> <https://github.com/uestcnlp/STAMP>

<sup>4</sup> <https://github.com/CRIPAC-DIG/SR-GNN>

<sup>5</sup> <https://github.com/RuihongQiu/FGNN>

<sup>6</sup> <https://github.com/xiaxin1998/DHCN>

<sup>7</sup> <https://doi.org/10.24433/CO.1542630.v1>

先进模型，能够学习会话级和全局级的项目嵌入，从而支持准确的预测。

#### 4.2.4 超参数设定

在 SR-HCGAT 模型的实验中，设置三个数据集的项目嵌入维度为 100，批大小（batch size）为 100，学习率  $\eta$  和  $L_2$  正则化系数  $\lambda$  分别为  $\eta = 0.001$ 、 $\lambda = 10^{-5}$ 。对于数据集 Tmall，采用单层网络表现较好；对于数据集 Nowplaying 和 Diginetica，采用三层网络性能较优。

### 4.3 实验结果与讨论

表 4.2、表 4.3 和表 4.4 展示了 SR-HCGAT 模型和七个对比模型在三个数据集上的推荐性能。其中，所有模型中最好的性能数字用粗体表示，“性能提升（%）”数字表示 SR-HCGAT 模型相对于对比模型中的最好性能的提升百分比，七个对比模型中最好的性能数字用下划线表示。

通过观察表 4.2、4.3、4.4 中的性能比较结果，可以发现：

（1）本文提出的 SR-HCGAT 模型在三个数据集上进行实验后，所有指标均优于七个对比的基线模型。具体来说，SR-HCGAT 的性能结果比七个对比模型中最好的性能结果分别在数据集 Tmall、Nowplaying 和 Diginetica 上提升了 5.77%、1.70% 和 1.54%，这验证了 SR-HCGAT 模型的性能优势和有效性。

（2）基于 GNN 的会话推荐模型（SR-GNN、FGNN）的推荐性能总体优于基于 RNN 的会话推荐模型（GRU4Rec），这表明 GNN 比 RNN 更适用于基于会话的推荐。

（3）SR-HCGAT 模型在 Tmall、Nowplaying 和 Diginetica 三个真实世界数据集上的推荐性能基本上都优于 DHCN 模型，这表明利用图注意力网络并结合自监督学习，对包含互补信息的两种用户偏好进行联合学习可以更准确地捕获用户偏好并提高推荐性能。

（4）可以看到 SR-HCGAT 在 Tmall 数据集的性能提升较大（5.77%），而在 Nowplaying（1.70%）和 Diginetica（1.54%）的提升较小。合理猜测 SR-HCGAT 的性能受数据集的规模影响，模型能够更好地适应该小规模数据集的特征，在大规模数据集上，其性能提升不如在小规模数据集上显著。

表 4.2 诸推荐模型在 Tmall 数据集上的性能比较

模型	Tmall			
	P@10	MRR@10	P@20	MRR@20
GRU4REC	9.38	5.69	10.84	5.82
NARM	19.08	10.35	23.24	10.64
STAMP	22.55	13.03	26.38	13.30
SR-GNN	23.35	13.38	27.49	13.67
FGNN	20.58	10.01	25.17	10.30
DHCN	27.41	15.26	<u>33.59</u>	15.66
FSM	<u>27.56</u>	<u>15.42</u>	33.29	<u>16.80</u>
SR-HCGAT (ours)	<b>29.07</b>	<b>16.31</b>	<b>34.91</b>	<b>16.95</b>
性能提升 (%)	5.48	5.77	3.93	0.89

表 4.3 诸推荐模型在 Nowplaying 数据集上的性能比较

模型	Nowplaying			
	P@10	MRR@10	P@20	MRR@20
GRU4REC	6.68	4.37	7.85	4.42
NARM	13.55	6.56	18.50	6.84
STAMP	13.15	6.49	17.58	6.80
SR-GNN	14.10	7.11	18.80	7.39
FGNN	13.81	6.72	18.68	7.07
DHCN	<u>17.19</u>	7.68	<u>22.91</u>	8.04
FSM	16.73	<u>7.72</u>	20.12	<u>8.24</u>
SR-HCGAT (ours)	<b>17.41</b>	<b>7.83</b>	<b>23.30</b>	<b>8.34</b>
性能提升 (%)	1.28	1.42	1.70	1.21

表 4.4 诸推荐模型在 Diginetica 数据集上的性能比较

模型	Diginetica			
	P@10	MRR@10	P@20	MRR@20
GRU4REC	17.79	7.24	29.29	8.24
NARM	35.36	15.04	49.51	16.08
STAMP	33.79	14.14	45.51	14.24
SR-GNN	36.70	15.43	50.59	17.48
FGNN	37.60	15.88	50.39	16.76
DHCN	<u>39.62</u>	<u>17.34</u>	<u>52.82</u>	<u>18.18</u>
FSM	39.43	17.25	51.89	17.94
SR-HCGAT (ours)	<b>39.95</b>	<b>17.56</b>	<b>53.47</b>	<b>18.46</b>
性能提升 (%)	0.83	1.27	1.23	1.54

#### 4.4 本章小结

本章首先对实验所采用的数据集 (Tmall、Nowplaying 和 Diginetica)、评价指标 (P@K, MRR@K)、用于对比的基线模型和实验超参数的设定进行解释, 对 SR-HCGAT 模型在三个数据集上的实验表明, SR-HCGAT 在 P@K 和 MRR@K 两个性能指标上明显优于七个具有代表性的会话推荐模型, 包括较先进的模型 DHCN、FSM。

## 第5章 总结与展望

### 5.1 工作总结

基于会话的推荐有广泛的应用前景。近年来，在用户数据量快速膨胀的情况下，如何为用户提供满足用户预期、符合实际要求的推荐愈发受到研究者的关注。随着深度学习技术在推荐系统领域的蓬勃发展，使用深度神经网络有望提高会话推荐的性能。本文主要工作如下：

(1) 本文提出并实现了结合超图卷积网络和图注意力网络的会话推荐模型 (SR-HCGAT)。该模型利用超图卷积网络捕获项目层的用户偏好，利用图注意力网络捕获会话层的用户偏好，并对包含互补信息的上述两种用户偏好进行联合学习以更准确地捕获用户偏好；最后，利用学得的项目层用户偏好产生推荐，以改善推荐性能。

(2) 本文在三个真实世界数据集上采用标准性能度量指标对 SR-HCGAT 进行实验评价，实验结果表明 SR-HCGAT 比几个先进的基线模型具有更好的推荐性能。

### 5.2 未来工作展望

基于会话的推荐实现了更加精准和个性化的推荐并正在飞速发展（如与社交网络<sup>[38]</sup>、异构信息网络<sup>[39]</sup>等进行结合）。未来，基于会话的推荐将会面临一些新的挑战 and 机遇。

随着移动互联网和智能终端的普及，用户的行为方式也发生了很大的变化，如何在移动设备上实现实时分析会话并提高推荐性能成为新的挑战。未来，基于会话的推荐模型需要更加注重对移动设备的适配和优化，尤其是在计算资源和带宽有限的情况下，如何实现高效的推荐将成为关键。

随着社交网络和社区的兴起，用户之间的关系和交互也变得越来越复杂。未来的基于会话的推荐系统需要考虑更多的社交因素，如用户的朋友圈、社区和兴趣组，以及用户之间的交互和合作关系等，从而更加准确地捕捉用户的兴趣和需求。

总的来说，基于会话的推荐是一个非常有前途的研究方向。未来，我们可以期待更加智能和个性化的基于会话的推荐模型的出现。

## 参考文献

- [1] Le Wu, Xiangnan He, Xiang Wang, Kun Zhang, Meng Wang. A Survey on Accuracy-Oriented Neural Recommendation: From Collaborative Filtering to Information-Rich Recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(5): 4425-4445. <https://doi.org/10.1109/TKDE.2022.3145690>
- [2] Charu C. Aggarwal. An Introduction to Recommender Systems. In: Charu C. Aggarwal, Recommender Systems: The Textbook. Springer 2016, pp. 1–28. [https://doi.org/10.1007/978-3-319-29659-3\\_1](https://doi.org/10.1007/978-3-319-29659-3_1)
- [3] Charu C. Aggarwal. Neighborhood-Based Collaborative Filtering. In: Charu C. Aggarwal, Recommender Systems: The Textbook. Springer 2016, pp. 29–70. [https://doi.org/10.1007/978-3-319-29659-3\\_2](https://doi.org/10.1007/978-3-319-29659-3_2)
- [4] Shoujin Wang, Longbing Cao, Yan Wang, Quan Z. Sheng, Mehmet A. Orgun, Defu Lian. A Survey on Session-based Recommender Systems[J]. ACM Computing Surveys, 2022, 54(7): 154:1-154:38. <https://doi.org/10.1145/3465401>
- [5] 赵海燕, 赵佳斌, 陈庆奎, 曹健. 会话推荐系统. 小型微型计算机系统[J], 2019, 40(9): 1869-1875. <http://xwxt.sict.ac.cn/CN/Y2019/V40/I9/1869>
- [6] 曾义夫, 牟其林, 周乐, 蓝天, 刘峤. 基于图表示学习的会话感知推荐模型[J]. 计算机研究与发展, 2020, 57(3):14. [https://www.nstl.gov.cn/paper\\_detail.html?id=a14be1e8a65ddf8bcd8e4909625a462b](https://www.nstl.gov.cn/paper_detail.html?id=a14be1e8a65ddf8bcd8e4909625a462b)
- [7] Balázs Hidasi, Alexandros Karatzoglou. Recurrent Neural Networks with Top-k Gains for Session-based Recommendations[C] //Twenty-Seventh ACM International Conference on Information and Knowledge Management, CIKM 2018: 843-852. <https://doi.org/10.1145/3269206.3271761>
- [8] Ian J. Goodfellow, Yoshua Bengio, Aaron C. Courville. Sequence Modeling: Recurrent and Recursive Nets. In: Deep Learning[M]. MIT Press (2016), pp. 373–420. <http://www.deeplearningbook.org/contents/rnn.html>
- [9] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, Domonkos Tikk. Session-based Recommendations with Recurrent Neural Networks[C] //Fourth International

- Conference on Learning Representations, ICLR 2016.  
<http://arxiv.org/abs/1511.06939>
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition[C] //Computer Vision and Pattern Recognition, CVPR 2016: 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [11] Alex Graves, Abdel-rahman Mohamed, Geoffrey E. Hinton. Speech recognition with deep recurrent neural networks[C]// International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013: 6645-6649.  
<https://doi.org/10.1109/ICASSP.2013.6638947>
- [12] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, Hemal Shah. Wide & Deep Learning for Recommender Systems[C] //Tenth ACM Conference on Recommender Systems, RecSys 2016: 7-10. <https://doi.org/10.1145/2988450.2988454>
- [13] Tim Donkers, Benedikt Loepp, Jürgen Ziegler. Sequential User-based Recurrent Neural Network Recommendations[C] //Eleventh ACM Conference on Recommender Systems, RecSys 2017: 152-160.  
<https://doi.org/10.1145/3109859.3109877>
- [14] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, Jun Ma. Neural Attentive Session-based Recommendation[C] //Twenty-seventh ACM on Conference on Information and Knowledge Management, CIKM 2017: 1419-1428.  
<https://doi.org/10.1145/3132847.3132926>
- [15] Qiao Liu, Yifu Zeng, Refuoe Mokhosi, Haibin Zhang: STAMP. Short-Term Attention/Memory Priority Model for Session-based Recommendation[C] //Twenty-fourth ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018: 1831-1839. <https://doi.org/10.1145/3219819.3219950>
- [16] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, Edward Y. Chang. Improving Sequential Recommendation with Knowledge-Enhanced Memory Networks[C] //Forty-first International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018: 505-514.

- <https://doi.org/10.1145/3209978.3210017>
- [17] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, Tieniu Tan. Session-Based Recommendation with Graph Neural Networks[C] //Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019: 346-353. <https://doi.org/10.1609/aaai.v33i01.3301346>
- [18] Ruihong Qiu, Jingjing Li, Zi Huang, Hongzhi Yin. Rethinking the Item Order in Session-based Recommendation with Graph Neural Networks[C] //Twenty-Eighth ACM International Conference on Information and Knowledge Management, CIKM 2019: 579-588. <https://doi.org/10.1145/3357384.3358010>
- [19] Chengfeng Xu, Pengpeng Zhao, Yanchi Liu, Victor S. Sheng, Jiajie Xu, Fuzhen Zhuang, Junhua Fang, Xiaofang Zhou. Graph Contextualized Self-Attention Network for Session-based Recommendation[C] //Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019: 3940-3946. <https://doi.org/10.24963/ijcai.2019/547>
- [20] Feng Yu, Yanqiao Zhu, Qiang Liu, Shu Wu, Liang Wang, Tieniu Tan. TAGNN: Target Attentive Graph Neural Networks for Session-based Recommendation[C] //Forty-Third International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020: 1921-1924. <https://doi.org/10.1145/3397271.3401319>
- [21] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, Yue Gao. Hypergraph Neural Networks[C] //Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019: 3558-3565. <https://doi.org/10.1609/aaai.v33i01.33013558>
- [22] Song Bai, Feihu Zhang, Philip H. S. Torr. Hypergraph convolution and hypergraph attention[J]. Pattern Recognit, 2021, 110: 107637. <https://arxiv.org/abs/1901.08150>
- [23] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, Yoshua Bengio. Graph Attention Networks[C] //Sixth International Conference on Learning Representations, ICLR (Poster) 2018. <https://openreview.net/pdf?id=rJXMpikCZ>
- [24] Huachi Zhou, Qiaoyu Tan, Xiao Huang, Kaixiong Zhou, Xiaoling Wang. Temporal Augmented Graph Neural Networks for Session-Based Recommendations[C]



- //Forty-Forth International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2021: 1798-1802.  
<https://doi.org/10.1145/3404835.3463112>
- [25] Nan Wang, Shoujin Wang, Yan Wang, Quan Z. Sheng, Mehmet A. Orgun. Modelling Local and Global Dependencies for Next-Item Recommendations[C] //Twenty-First International Conference on Web Information Systems Engineering, WISE 2020: 285-300. [https://doi.org/10.1007/978-3-030-62008-0\\_20](https://doi.org/10.1007/978-3-030-62008-0_20)
- [26] Chao Huang, Jiahui Chen, Lianghao Xia, Yong Xu, Peng Dai, Yanqing Chen, Liefeng Bo, Jiashu Zhao, Jimmy Xiangji Huang. Graph-Enhanced Multi-Task Learning of Multi-Level Transition Dynamics for Session-based Recommendation[C] // Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021: 4123-4130. <https://ojs.aaai.org/index.php/AAAI/article/view/16534>
- [27] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, Philip S. Yu. A Comprehensive Survey on Graph Neural Networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.  
<https://doi.org/10.1109/TNNLS.2020.2978386>
- [28] Yujia Li, Daniel Tarlow, Marc Brockschmidt, Richard S. Zemel. Gated Graph Sequence Neural Networks[C] //Forth International Conference on Learning Representations, ICLR 2016. <http://arxiv.org/abs/1511.05493>
- [29] Dunlu Peng, Shuo Zhang. GC-HGNN: A global-context supported hypergraph neural network for enhancing session-based recommendation[J]. Electronic Commerce Research and Applications, 2022, 52(101129).  
<https://doi.org/10.1016/j.elerap.2022.101129>
- [30] Xin Xia, Hongzhi Yin, Junliang Yu, Qinyong Wang, Lizhen Cui, Xiangliang Zhang. Self-Supervised Hypergraph Convolutional Networks for Session-based Recommendation[C] //Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021: 4503-4511. <https://ojs.aaai.org/index.php/AAAI/article/view/16578>
- [31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention is All you Need[C] //Annual Conference on Neural Information Processing Systems, NIPS 2017: 5998-6008.

- <https://arxiv.org/abs/1706.03762>
- [32] Aäron van den Oord, Yazhe Li, Oriol Vinyals: Representation Learning with Contrastive Predictive Coding. CoRR abs/1807.03748 (2018).  
<http://arxiv.org/abs/1807.03748>
- [33] Xin Xia, Hongzhi Yin, Junliang Yu, Yingxia Shao, Lizhen Cui. Self-Supervised Graph Co-Training for Session-based Recommendation[C] //Thirtieth ACM International Conference on Information and Knowledge Management, CIKM 2021: 2180-2190. <https://doi.org/10.1145/3459637.3482388>
- [34] Diederik P. Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization[C] //Third International Conference on Learning Representations, ICLR 2015.  
<https://arxiv.org/abs/1412.6980>
- [35] Charu C. Aggarwal. Evaluating Recommender Systems. In: Recommender Systems: The Textbook[M]. Springer 2016, pp. 225-254. [https://doi.org/10.1007/978-3-319-29659-3\\_7](https://doi.org/10.1007/978-3-319-29659-3_7)
- [36] Xiang Huang, Yixin He, Bin Yan, Wei Zeng. Fusing frequent sub-sequences in the session-based recommender system[J]. Expert Systems with Applications, 2022, 206: 117789. <https://linkinghub.elsevier.com/retrieve/pii/S0957417422010570>
- [37] Shoujin Wang, Qi Zhang, Liang Hu, Xiuzhen Zhang, Yan Wang, Charu Aggarwal. Sequential/Session-based Recommendations: Challenges, Approaches, Applications and Opportunities[C] //Forty-Fifth International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2022: 3425-3428.  
<https://dl.acm.org/doi/10.1145/3477495.3532685>
- [38] Tianwen Chen, Raymond Chi-Wing Wong. An Efficient and Effective Framework for Session-based Social Recommendation[C] //Fourteenth ACM International Conference on Web Search and Data Mining, WSDM 2021: 400-408.  
<https://dl.acm.org/doi/10.1145/3437963.3441792>
- [39] DRprofiling. Deep Reinforcement User Profiling for Recommendations in Heterogenous Information Networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(4): 1723-1734.  
<https://ieeexplore.ieee.org/document/9103633/>

## 致 谢

转眼间本科四年的时光已悄然过去，门卫叔叔、食堂阿姨.....一张张熟悉又陌生的脸，仿佛就在眼前。四年间，我结识了许多良师益友，他们对我的成长与发展给予了不可磨灭的帮助，在此向他们致以我最真诚的感谢。

首先，感谢我的指导教师许卓明教授和师姐孟志鑫。在这段时间里，许教授严谨的学术态度、博学的学识以及对学生的细心关怀，让我在本科最后的道路上获益匪浅、收获满满，并给予了我非常有价值的指导和建议。孟师姐无论是在课题的选择、研究方法的确定，还是在毕业论文的撰写和修改过程中，给予了我非常多的支持和帮助。

此外，感谢我的父母和家人。你们一直以来无私的关心、支持和理解是我在学习和生活上的坚实后盾。在我完成毕业论文的过程中，你们一直给我鼓励和支持，让我有了坚定的信心和动力。

感谢我的朋友们，在我学习和生活中给予了我很多帮助和支持。你们的欢声笑语陪伴了我整整四年，此次离别，不知何时再见，我始如初心。谢谢家人们，愿前程似锦，我们都有美好的未来。

最后，感谢国家。

李怡飞 2023.6