

# 《Compressed Interaction Graph based Framework for Multi-behavior Recommendation》复现报告

论文作者：Wei Guo, Chang Meng, Enming Yuan, et al.

## 摘要

随着互联网技术的飞速进步，推荐系统已经逐渐渗透到人们的日常生活中，成为处理和过滤海量数据，帮助用户迅速发现其所需信息的关键工具。由于数据的极端稀疏性，对于多行为推荐系统的研究构成了重大挑战。以往的研究在利用多行为数据方面存在不足，难以有效捕捉多种行为之间的相互作用和高阶关系。当采用多任务学习 (MTL) 框架对多行为推荐系统进行建模时，存在一些关键性的挑战，例如梯度冲突问题。本文研究与复现最近的工作 CIGF，文章提出的压缩图交互框架 (CIGF) 一定程度地缓解了数据的分布不均匀和多任务学习中的梯度冲突的问题。复现结果与原论文相近，最后对模型的多任务模块尝试了改进，并对实验结果分析。

**关键词：**多行为推荐；多任务

## 1 引言

### 1.1 选题背景

随着互联网技术的飞速进步，推荐系统已经逐渐渗透到人们的日常生活中，成为处理和过滤海量数据，帮助用户迅速发现其所需信息的关键工具。它不仅极大地提高了信息获取的效率，也增强了个性化体验。然而，传统推荐系统主要依赖于协同过滤 (Collaborative Filtering, CF) 技术，该技术通过挖掘用户与物品之间的历史交互数据来学习隐性特征表示，进而利用这些学习到的特征进行预测。当前，大多数协同过滤 (CF)<sup>[6,7,13,16]</sup> 方法主要集中于基于单一交互行为的建模。然而，在实际应用中，用户与物品的互动通常包含多种不同类型的行为。以在线购物为例，用户与商品的交互不仅限于点击，还可能包括收藏、购买等多种行为形式。这些不同的交互类型各自承载了不同的用户偏好和意图，因此，在建模过程中考虑这些多样化的交互方式对于提高推荐系统的准确性和效率至关重要。

### 1.2 选题依据

在现实世界的应用场景中，数据的分布通常表现出显著的不平衡，尤其是在用户行为和交互类型方面。这种不平衡性尤其在目标行为（比如在线购买中的购买）中表现得尤为明显，由于数据的极端稀疏性，对于多行为推荐系统的研究构成了重大挑战。早期的研究在利用多行为数据方面存在不足，难以有效捕捉多种行为之间的相互作用和高阶关系。此外，当采用

多任务学习 (MTL) 框架对多行为推荐系统进行建模时, 存在一些关键性的挑战, 例如梯度冲突问题。这种问题在任务间的目标不一致时尤为突出, 导致模型优化过程中的各个任务相互干扰, 从而对整体推荐系统的性能产生负面影响。因此, 上述两个问题共同构成了限制多行为推荐系统发展的主要障碍。

### 1.3 选题意义

针对以上问题, 本文复现的论文提出一个能够有效利用多行为数据的推荐系统框架具有重要的研究价值和实际应用意义。该论文提出的压缩交互图框架 (CIGF) 不仅能更有效的捕捉多种行为之间的相互作用和高阶关系, 而且缓解了将多行为数据作为标签时的梯度冲突问题。CIGF 通过压缩交互图卷积网络 (CIGCN) 和多专家分离输入网络 (MESI) 的设计, 提高了推荐系统对用户偏好的理解深度, 同时也增强了模型对稀疏数据的处理能力, 从而提升了推荐效果, 并对推荐系统的发展方向提供了新的视角和理论支持。通过实验证明了 CIGF 在多个真实世界数据集上的有效性。通过对该篇文章的研究和复现对我将来的研究有很大的帮助。

## 2 相关工作

本节主要回顾在推荐系统领域中, 一些代表性的图神经网络 (Graph Neural Network, GNN) 方法和多任务学习 (Multi-task Learning, MTL) 方法。

### 2.1 图神经网络推荐

图神经网络因其在捕获和处理关系数据方面展现出了卓越的潜力, 已经成为推荐领域中一个关键的研究方向。近年来, 为应对推荐系统中的多行为数据问题, 已经提出了众多基于 GNN 的深度学习方法。这些方法主要针对两类不同的图结构进行了优化设计, 即对应单行为和多行为的推荐。在单行为的情况下, 例如, NGCF<sup>[13]</sup> 利用图神经网络架构捕捉用户和物品之间的交互, 通过层级传播和非线性变换来有效建模这些复杂的用户-物品关系。另一方面, LR-GCCF<sup>[2]</sup> 引入了一种层级聚合策略, 通过整合不同层级的图卷积输出, 从而提升了用户和物品的表示能力。此外, LightGCN<sup>[6]</sup> 则在简化传统图卷积网络的基础上, 去除了权重和非线性激活函数, 专注于通过邻居的直接累积来强化特征的表示学习。对于多行为的情况, HGNN<sup>[17]</sup> 针对包含多种类型节点和边缘的图提出了一种新颖的建模策略, 能够有效地整合不同类型的信息。R-GCN<sup>[11]</sup> 通过引入关系类型特定的转换权重, 为异质图中不同类型的关系建模, 从而在保持关系特异性的同时, 也能捕捉用户和物品间的复杂交互。而 HGAT<sup>[9]</sup> 则利用注意力机制来动态地聚焦于异质图中不同类型的边缘, 进而更细粒度地捕获用户-物品之间的多样化关系。此外, MBGCN<sup>[8]</sup> 设计了一种复杂的图卷积结构, 用于同时学习用户在不同行为下的偏好。该方法通过建立用户-物品间的多重关系图, 利用图卷积网络深入挖掘这些多维度的交互数据, 以更全面地捕获用户的行为模式和偏好。

### 2.2 多任务学习方法

在多任务学习领域, 已经有一系列经典方法, 如 MMOE<sup>[10]</sup> 和 PLE<sup>[12]</sup>, 这些方法已被成功应用于推荐系统中。这些方法的核心思想是通过共享隐藏层来同时学习多个相关任务, 从

而提高每个任务的性能，并捕获不同任务间的潜在关联。例如，MMOE 在处理具有多种行为的用户数据时，通过设置多个专家门来捕获不同任务间的复杂关系，使得网络能够针对每个特定任务学习特定的特征表示。PLE 则进一步扩展了这一概念，通过逐层提取和组合特征，实现了对任务间层次结构的更细致建模。

在推荐系统中，多任务学习方法被用来同时优化多个相关的推荐任务，例如，评分预测和点击率预测。NMTR<sup>[3]</sup> 是一种多任务学习框架，它结合了神经协同过滤 (NCF) 模块，以独特的方式学习每种行为下用户的偏好。NMTR 通过级联的方式关联各个行为的预测，有效捕获了不同类型行为之间的层次关系。DIPN<sup>[4]</sup> 则通过深度学习方法对用户的多元兴趣进行建模，能够同时处理多个与用户兴趣相关的任务。DIPN 的特点在于它如何利用深层神经网络来捕获用户兴趣的多样性和复杂性。MATN<sup>[14]</sup> 采用了注意力机制来集中处理多个推荐相关任务。MATN 通过关注不同任务中关键的特征，能够更精确地对用户的多样化需求做出响应。这些多任务学习方法的共同点在于，它们都试图通过共享知识和特征来提高多个推荐任务的整体性能，同时保持对每个任务特定需求的敏感性和适应性。

### 2.3 多行为推荐

最新的研究成果中，GHCF<sup>[1]</sup> 方法通过处理异质图，即包含多种类型节点和边缘的图，来全面捕捉用户与物品之间的关系。该方法采用了创新的多任务学习策略，通过共享用户-物品交互信息作为输入，以并行优化与不同用户行为相关的多项任务。另一方面，MB-GMN<sup>[15]</sup> 采用了图神经网络的一种变体，并融合了元学习机制，不仅关注用户与物品之间的直接互动，而且还旨在揭示多种行为类型之间的复杂相互依赖关系。尽管 GHCF 和 MB-GMN 在多任务处理方面都使用了相同的输入数据源，但它们均未充分考虑到不同行为数据之间的潜在差异。CIGF<sup>[5]</sup> 的研究人员指出，现有的多行为推荐方法大致可分为两类。第一类基于图结构的方法，将多种行为作为特征的一部分，这些方法尽管能够处理特征层面的行为信息，却往往无法捕捉到个体层面的细微行为关系。第二类以多任务学习为基础的方法，倾向于将不同的用户行为视为独立的任务标签，并采用统一的权重方案来共同优化这些任务，但这种方法易受样本差异所引起的任务冲突的影响。为了缓解这一问题，CIGF 论文提出了 CIGCN 网络和 MESI 网络来缓解这一问题。

## 3 本文方法

### 3.1 本文方法概述

论文中提出的 CIGF 框架，以端到端的方式将多行为数据”作为特征”和”作为标签”。该模型图如图 1 所示所示，由 3 个主要部分组成：(1) 输入层，将用户和物品参数化为嵌入向量；(2) 压缩交互图卷积网络 (CIGCN)，通过将多行为数据”视为特征”，显式地从多行为数据中提取实例级高阶关系；(3) 多专家独立输入 (MESI) 网络，将具有独立输入的多行为数据作为标签来挖掘多任务监督信号。

多行为推荐任务的形式化定义。将用户集合和物品集合分别表示为  $\mathcal{U} = \{u_1, u_2, \dots, u_M\}$  和  $\mathcal{I} = \{i_1, i_2, \dots, i_N\}$ ，用户-物品交互行为矩阵为  $\mathcal{Y} = \{\mathbf{Y}^1, \mathbf{Y}^2, \dots, \mathbf{Y}^K\}$ 。

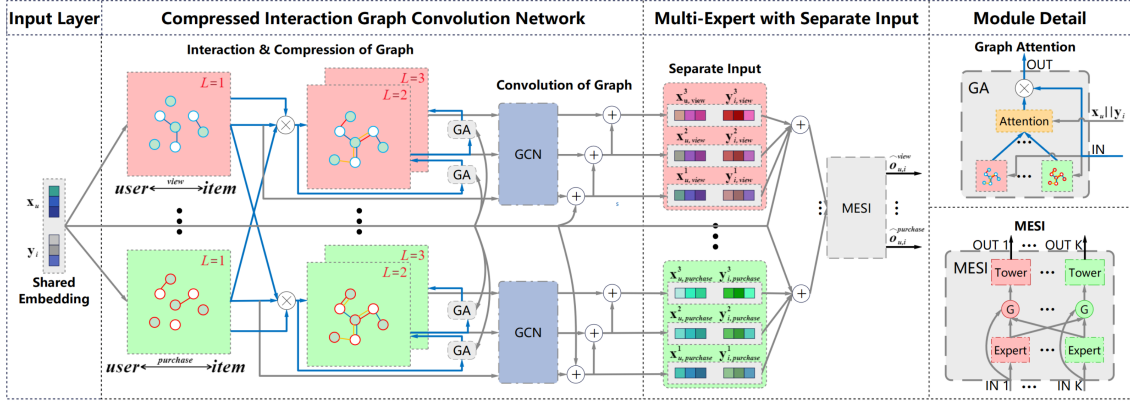


图 1. 模型示意图

### 3.2 输入

首先使用共享嵌入层将用户和物品的独热 ID 转换为低维稠密嵌入向量。形式化地给定一个用户-物品对  $(u, i)$ ，用户  $u$  和物品  $i$  的嵌入向量查找操作可以表示为如下形式：

$$\mathbf{x}_u = \mathbf{P}^T \cdot \mathbf{p}_u, \quad \mathbf{y}_i = \mathbf{Q}^T \cdot \mathbf{q}_i \quad (1)$$

### 3.3 压缩交互图卷积

如图 1 所示，首先将 MBG 划分为若干个行为的子图  $\mathcal{G}^1, \mathcal{G}^2, \dots, \mathcal{G}^K$ 。相对应的邻接矩阵为  $\mathbf{A}^1, \mathbf{A}^2, \dots, \mathbf{A}^K$ ，可以表示为：

$$\mathbf{A}^k = \begin{pmatrix} 0 & (\mathbf{Y}^k)^T \\ \mathbf{Y}^k & 0 \end{pmatrix} \quad (2)$$

其中  $\mathbf{Y}^k$  是行为  $k$  的用户-物品交互矩阵。使用这些邻接矩阵编码显式高阶图交互，针对的是每两个节点之间的实例级别关系。表示所有可能的  $l$ -th ( $1 \leq l \leq L$ ) 阶从行为  $k$  开始的交互图的集合为  $\mathcal{B}_k^l$ 。 $\mathcal{B}_k^l$  的生成可以表示为：

$$\mathcal{B}_k^l = \mathcal{B}_k^{l-1} \otimes \{\mathbf{A}^1, \mathbf{A}^2, \dots, \mathbf{A}^K\} \quad (3)$$

其中  $\mathcal{B}_k^1 = \{\mathbf{A}^k\}$ 。符号  $(\otimes)$  表示从两组矩阵中分别取出矩阵对进行矩阵乘法运算。注意，有  $K$  组高阶图  $\mathcal{B}_k^l (1 \leq k \leq K)$ 。进一步的，作者提出一个图压缩层来迭代构建高阶图集，该层通过节点级多头注意力机制来实现。针对目标节点  $v$ （节点  $v$  可以是用户节点  $u$  或物品节点  $i$ ）的图压缩层可以表示为：

$$\mathcal{B}_{v,k}^l = \mathcal{B}_{v,k}^{l-1} \otimes \{\alpha_{v,k}^{l,1}[\mathbf{A}^1, \dots, \mathbf{A}^K], \dots, \alpha_{v,k}^{l,H}[\mathbf{A}^1, \dots, \mathbf{A}^K]\} \quad (4)$$

其中  $\mathcal{B}_{v,k}^1 = \{\mathbf{A}^k\}$ ，点  $(\cdot)$  是向量乘法运算， $H$  是头的数量，而  $\alpha_{v,k}^{l,h} \in \mathbb{R}^{1 \times K}$  是为节点  $v$  在  $l$ -阶和第  $h$ -头学习到的注意力向量。同时，为了让模型能自适应地选择对用户和物品表示学习最相关的行为，节点级注意力机制设置不同行为的权重可以定义为：

$$\alpha_{v,k}^{l,h} = \sigma(\mathbf{W}_k^{l,h} x_v^{l,h} + \mathbf{b}_k^{l,h}), \quad \alpha_{v,k}^{l,h} = \sigma(\mathbf{W}_k^{l,h} y_i^{l,h} + \mathbf{b}_k^{l,h}) \quad (5)$$

其中  $\sigma(\cdot)$  是 LeakyReLU 激活函数。  $\mathbf{W}_k^{l,h} \in \mathbb{R}^{D \times d}$  和  $\mathbf{b}_k^{l,h} \in \mathbb{R}^{K \times 1}$  分别是特征变换矩阵和偏置向量，用于在注意力机制中对特征进行转换和偏移。

通过图交互和图压缩层生成图集后，之后通过图卷积来学习用户和物品的表示。每个图中的邻居信息传播可表述为：

$$\mathbf{x}_{N_u,k}^{l,s} = \text{Agg}(\mathbf{x}_u, \mathcal{B}_{u,k}^{l,s}), \quad \mathbf{y}_{N_i,k}^{l,t} = \text{Agg}(\mathbf{y}_i, \mathcal{B}_{i,k}^{l,t}) \quad (6)$$

其中  $\mathcal{B}_{u,k}^{l,s}$  和  $\mathcal{B}_{i,k}^{l,t}$  表示第  $s$ -th 和  $t$ -th 图的邻接矩阵。  $N_u$  和  $N_i$  表示  $u$  和  $i$  的邻居节点，  $x_{N_u,k}^{l,s}$  和  $y_{N_i,k}^{l,t}$  表示通过聚合来自  $s$ -th 和  $t$ -th 图的邻居信息后的输出。  $\text{Agg}(\cdot)$  是信息聚合的任意图卷积操作符。作者使用 LightGCN Aggregator [14] 模型来实现。在邻居信息传播过程之后，最终的用户和物品表示通过求和操作来实现：

$$\mathbf{x}_{N_u,k}^l = \sum_{s=1}^{\text{pow}(\mathcal{H}, l-1)} \mathbf{x}_{N_u,k}^{l,s}, \quad \mathbf{y}_{N_i,k}^l = \sum_{t=1}^{\text{pow}(\mathcal{H}, l-1)} \mathbf{y}_{N_i,k}^{l,t} \quad (7)$$

在图卷积层中为最终节点信息更新引入了残差操作，定义如下：

$$\mathbf{x}_{u,k}^l = \mathbf{x}_{N_u,k}^l + \mathbf{x}_{u,k}^{l-1}, \quad \mathbf{y}_{i,k}^l = \mathbf{y}_{N_i,k}^l + \mathbf{y}_{i,k}^{l-1} \quad (8)$$

最终将这些输出聚合成一个单一向量，如下所示：

$$\mathbf{x}_{u,k}^* = \sum_{l=0}^L \mathbf{x}_{u,k}^l, \quad \mathbf{y}_{i,k}^* = \sum_{l=0}^L \mathbf{y}_{i,k}^l \quad (9)$$

其中  $x_{u,k}^0 = x_u$  和  $y_{i,k}^0 = y_i$  是用户  $u$  和物品  $i$  的初始嵌入向量。

### 3.4 多专家分别独立输入

作者还提出了一种具有独立输入的多专家(MESI)网络,该网络具有新颖的独立输入设计。与之前的多任务模型不同的一点是输入是独立的，每个行为作为一个单独的输入，如图 1 中的 MESI 所示。每个专家被定义为  $\mathbf{x}_{u,k}^*$  和  $\mathbf{y}_{i,k}^*$  的信息组合，可以表示为：

$$f_{u,i}^k = \mathbf{x}_{u,k}^* \odot \mathbf{y}_{i,k}^* \quad (10)$$

其中  $\odot$  是哈达玛积运算，  $\mathbf{x}_{u,k}^*$  和  $\mathbf{y}_{i,k}^*$  是与行为  $k$  相关的输入。由于这里使用了分离输入来生成专家，总共可以获得  $K$  个专家。

使用分离输入来产生每个任务的感知门，自动选择一组对该任务预测有用的专家。任务  $k$  的门可以定义为：

$$g_{u,i}^k = \text{Softmax}(\mathbf{W}_g(\mathbf{x}_{u,k}^* \parallel \mathbf{y}_{i,k}^*) + \mathbf{b}_g) \quad (11)$$

其中  $\parallel$  是向量连接操作，  $\mathbf{W}_g \in \mathbb{R}^{K \times 2d}$  和  $\mathbf{b}_g \in \mathbb{R}^{K \times 1}$  是特征转换矩阵和偏置矩阵，  $g_{u,i}^k \in \mathbb{R}^{K \times 1}$  是用作选择器的注意力向量，用来计算所有专家的加权和。最终预测分数可以计算为：

$$\hat{o}_{u,i}^k = h^k \left( \sum_{j=1}^K g_{u,i}^k(j) \cdot f_{u,i}^j \right) \quad (12)$$

其中  $g_{u,i}^k(j)$  表示向量  $g_{u,i}^k$  的第  $j$  个元素，  $h^k(\cdot)$  是塔函数。这里使用平均操作作为塔函数。

### 3.5 损失函数

通过上文已经获得了每种行为  $k$  的预测值  $\hat{o}_{u,i}^k$ ，使用贝叶斯个性化排名（BPR）损失进行多任务学习，可以表示为：

$$\mathcal{L} = - \sum_{k=1}^K \sum_{(u,s,t) \in O_k} \ln(\sigma(\hat{o}_{u,s}^k - \hat{o}_{u,t}^k)) + \lambda \|\Theta\|^2 \quad (13)$$

其中  $O_k = \{(u, s) | (u, s) \in O_k^+, (u, t) \in O_k^-\}$  表示训练数据集。 $O_k^+$  表示在行为  $k$  下观察到的正用户-物品交互， $O_k^-$  表示未观察到的用户-物品交互。 $\Theta$  表示所有模型参数的集合， $\sigma$  是 Sigmoid 函数， $\lambda$  是  $\Theta$  的 L2 正则化系数。

## 4 复现细节

### 4.1 与已有开源代码对比

在复现过程中用到的作者公布的源码，并尝试对原文的多专家分别独立输入模块进行了改进。具体的做法如图 2b所示，这么做的原因是在多任务学习（MTL）中，共享信息可以帮助不同任务之间的知识转移，从而提高模型的泛化能力。原始模型中的每个专家网络都有独立的输入，这可能限制了不同任务间潜在有价值信息的交流。通过引入共享专家网络，不仅可以增强任务间的信息流动，还可以提高模型的效率和效果。

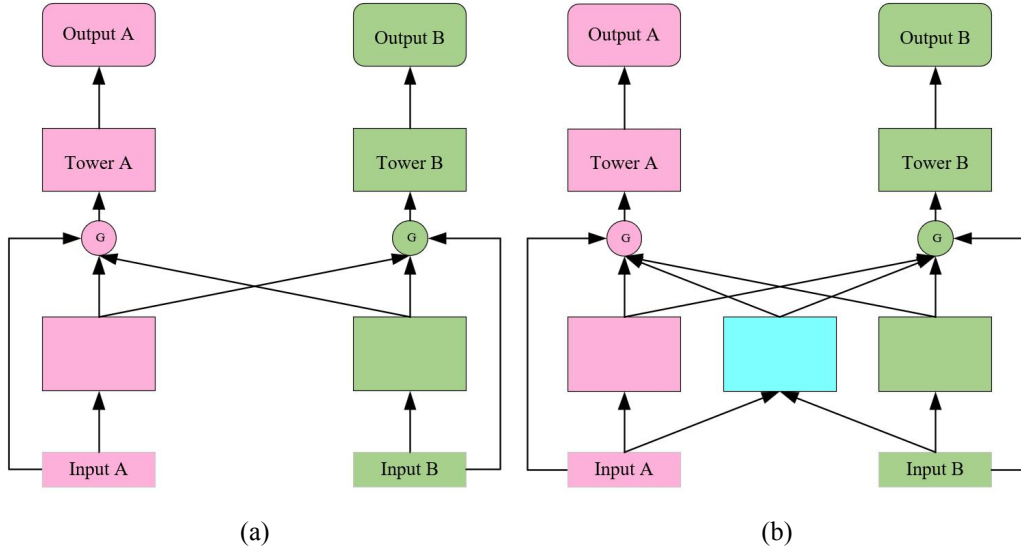


图 2. MESI 和改进后的 MESI

### 4.2 实验环境搭建

复现基于 tensorflow1.14.0 以及操作系统 windows11 完成

### 4.3 数据集和评估指标

采用 3 个公共数据集 (JD, Beibei, Taobao), 经过预处理后的数据统计数据如表1所示。使用命中率 (HR@10) 和归一化贴现累积增益 (NDCG@10) 用于评估性能。

表 1. 数据统计

Dataset	User	Item	Interaction	Behaviors
JD	10,690	13,465	335,164	View, Favorite, Buy
Beibei	21,716	7,977	3,338,068	View, Cart, Buy
Taobao	147,894	99,037	7,658,926	View, Favorite, Cart, Buy

#### 4.4 参数设置

用户和项目的物品嵌入向量大小设置为 16，并使用 Xavier 方法初始化模型参数。采用 Adam 来优化模型，并分别设置学习率为 0.001 和批量大小为 256。此外，图模型的 GCN 层数和 L2 归一化系数，epoch 设置为 100 次停止，与原文保持一致。

## 5 实验结果分析

实验结果如2所示，可以发现复现的结果与原文基本上保持一致，但是改进的效果却反而下降了，说明引入共享专家网络的改进效果并不显著。这可能有几个原因：信息冗余：共享网络可能未能捕获到对每个任务确实有用的新信息，反而引入了不必要的信息冗余。优化挑战：共享层的存在可能增加了模型的复杂性，给优化过程带来了额外挑战，尤其是在没有适当调整训练策略的情况下。任务相关性：如果不同的任务之间相关性不高，共享网络可能无法有效地学习到对各个任务都有用的通用特征。模型容量：共享专家可能限制了模型容量，使得它不能捕捉到对每个独立任务特别有用的特征。超参数调整：共享专家网络可能需要不同的超参数设置才能有效工作，而这些设置可能与原始模型的设置不同。在未来的工作中，可以考虑更细致地研究这些因素，如通过调整模型结构或训练流程来改进共享专家网络的效果。此外，更深入的分析可能需要探索数据的特性，任务间的互动以及是否存在某些任务对共享信息更加敏感。

Dataset	Beibei		Taobao		JD	
Model	HR	NDCG	HR	NDCG	HR	NDCG
CIGF	0.700	0.443	0.592	0.383	——	——
CIGF (Rep)	0.698	0.440	0.590	0.380	0.556	0.423
CIGF (Imp)	0.692	0.430	0.483	0.291	0.523	0.390

表 2. 实验结果。

## 6 总结与展望

在本文中，介绍了 CIGF 框架，并对其进行了详细的实验和分析。通过实验结果，发现复现的模型与原文的结果基本一致，但增加共享专家网络的改进并没有带来预期的效果提升。这可能说明在的情景下，信息共享并未能有效促进任务之间的知识转移，或许是因为任务间

的相关性不足，或者共享层增加了模型的复杂度而未经过适当的优化。在未来的研究中，计划深入探索共享专家网络背后的动态，并尝试调整模型结构和训练策略，以便更好地利用共享信息。此外，也希望研究数据特性和任务间的互动，以便发现哪些任务可能更适合于信息共享，并据此调整模型设计。

## 参考文献

- [1] Chong Chen, Weizhi Ma, Min Zhang, Zhaowei Wang, Xiuqiang He, Chenyang Wang, Yiqun Liu, and Shaoping Ma. Graph heterogeneous multi-relational recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3958–3966, 2021.
- [2] Lei Chen, Le Wu, Richang Hong, Kun Zhang, and Meng Wang. Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 27–34, 2020.
- [3] Chen Gao, Xiangnan He, Dahua Gan, Xiangning Chen, Fuli Feng, Yong Li, Tat-Seng Chua, and Depeng Jin. Neural multi-task recommendation from multi-behavior data. In *2019 IEEE 35th international conference on data engineering (ICDE)*, pages 1554–1557. IEEE, 2019.
- [4] Long Guo, Lifeng Hua, Rongfei Jia, Binqiang Zhao, Xiaobo Wang, and Bin Cui. Buying or browsing?: Predicting real-time purchasing intent using attention-based deep network with multiple behavior. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1984–1992, 2019.
- [5] Wei Guo, Chang Meng, Enming Yuan, Zhicheng He, Huifeng Guo, Yingxue Zhang, Bo Chen, Yaochen Hu, Ruiming Tang, Xiu Li, et al. Compressed interaction graph based framework for multi-behavior recommendation. *arXiv preprint arXiv:2303.02418*, 2023.
- [6] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, pages 639–648, 2020.
- [7] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182, 2017.
- [8] Bowen Jin, Chen Gao, Xiangnan He, Depeng Jin, and Yong Li. Multi-behavior recommendation with graph convolutional networks. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 659–668, 2020.
- [9] Hu Linmei, Tianchi Yang, Chuan Shi, Houye Ji, and Xiaoli Li. Heterogeneous graph attention networks for semi-supervised short text classification. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 4821–4830, 2019.



- [10] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1930–1939, 2018.
- [11] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15*, pages 593–607. Springer, 2018.
- [12] Hongyan Tang, Junning Liu, Ming Zhao, and Xudong Gong. Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In *Proceedings of the 14th ACM Conference on Recommender Systems*, pages 269–278, 2020.
- [13] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*, pages 165–174, 2019.
- [14] Lianghao Xia, Chao Huang, Yong Xu, Peng Dai, Bo Zhang, and Liefeng Bo. Multiplex behavioral relation learning for recommendation via memory augmented transformer network. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pages 2397–2406, 2020.
- [15] Lianghao Xia, Yong Xu, Chao Huang, Peng Dai, and Liefeng Bo. Graph meta network for multi-behavior recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 757–766, 2021.
- [16] Hong-Jian Xue, Xinyu Dai, Jianbing Zhang, Shujian Huang, and Jiajun Chen. Deep matrix factorization models for recommender systems. In *IJCAI*, volume 17, pages 3203–3209. Melbourne, Australia, 2017.
- [17] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 793–803, 2019.