

## 基于课程学习和图嵌入的协同推荐

黄飞虎, 帅剑波, 彭舰

引用本文

黄飞虎, 帅剑波, 彭舰. 基于课程学习和图嵌入的协同推荐[J]. 计算机科学, 2023, 50(11A): 221100030-8.

HUANG Feihu, SHUAI Jianbo, PENG Jian. Collaborative Recommendation Based on Curriculum Learning and Graph Embedding [J]. Computer Science, 2023, 50(11A): 221100030-8.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于异构信息网络的信贷反欺诈研究](#)

Study on Credit Anti-fraud Based on Heterogeneous Information Network

计算机科学, 2023, 50(11A): 221100173-9. <https://doi.org/10.11896/jsjcx.221100173>

[一种基于Meta-learning改进的特征交互算法](#)

Improved Feature Interaction Algorithm Based on Meta-learning

计算机科学, 2023, 50(11A): 230100087-8. <https://doi.org/10.11896/jsjcx.230100087>

[基于图嵌入的正交局部保持投影无监督特征选择](#)

Orthogonal Locality Preserving Projection Unsupervised Feature Selection Based on Graph Embedding

计算机科学, 2023, 50(11A): 220900003-9. <https://doi.org/10.11896/jsjcx.220900003>

[一种基于时效近邻可信选取策略的协同过滤推荐方法](#)

Time-effective Nearest Neighbor Trusted Selection Strategy Based Collaborative Filtering Recommendation Method

计算机科学, 2023, 50(11A): 220800199-11. <https://doi.org/10.11896/jsjcx.220800199>

[基于动态负采样的图卷积协同过滤推荐模型](#)

Dynamic Negative Sampling for Graph Convolution Network Based Collaborative Filtering Recommendation Model

计算机科学, 2023, 50(11A): 230200149-7. <https://doi.org/10.11896/jsjcx.230200149>

# 基于课程学习和图嵌入的协同推荐

黄飞虎<sup>1,2</sup> 帅剑波<sup>2</sup> 彭 舰<sup>2</sup>

1 四川中电启明星信息技术有限公司 成都 610000

2 四川大学计算机学院 成都 610065

(huangfh@scu.edu.cn)

**摘要** 推荐系统主要基于用户信息提供个性化服务。然而,用户对数据隐私泄露的广泛关注,给当前推荐算法提出了新的挑战。现有工作主要从差分隐私、匿名化、密码学和联邦学习的角度解决隐私泄露问题,但存在数据扰动和计算复杂的缺点。不同于现有工作,文中提出了基于课程学习和图神经网络的协同过滤模型(CLG-CF),充分利用评分表在隐式反馈的场景学习用户和物品嵌入。CLG-CF利用二部图建模评分表,基于图卷积网络实现用户和物品的表示学习,然后通过多层神经网络完成(用户,物品)对的预测。CLG-CF模型在训练过程中,采用负采样计算增强样本,为了解决样本的真伪问题,创新地引入课程学习指导模型学习。在3个真实的大规模数据集上进行了实验,结果表明 CLG-CF模型在不使用用户和物品信息的情况下,能够实现不错的推荐效果。

**关键词:** 推荐系统;协同过滤;课程学习;图嵌入;数据隐私

**中图法分类号** TP389

## Collaborative Recommendation Based on Curriculum Learning and Graph Embedding

HUANG Feihu<sup>1,2</sup>, SHUAI Jianbo<sup>2</sup> and PENG Jian<sup>2</sup>

1 Aostar Information Technologies Co., Ltd., Chengdu 610000, China

2 College of Computer Science, Sichuan University, Chengdu 610065, China

**Abstract** Recommendation system mainly provide personalized services based on user information. However, users are widely concerned about data privacy issues, which poses new challenges to current recommendation algorithms. Existing works mainly address this problem based on the perspectives of differential privacy, anonymization, cryptography, and federated learning. Data disturbance and computational complexity are the main shortcomings of existing methods. Different from existing work, this paper proposes a collaborative filtering model based on curriculum learning and graph neural network(CLG-CF), which makes full use of rating information to learn the embedding of user and item in implicit feedback scenarios. CLG-CF utilizes a bipartite graph modeling scoring table, then realizes the representation learning of users and items based on graph convolutional networks, finally completes the prediction of (user, item) pairs through a multi-layer neural network. During the training process of the CLG-CF model, negative sampling is used to enhance training samples. In order to solve the problem of samples' authenticity, the curriculum learning is innovatively introduced to guide the model learning. Extensive experiments are conducted on three real large-scale datasets, and the results show that the CLG-CF model can achieve good recommendation results without using user and item information.

**Keywords** Recommendation system, Collaborative filtering, Curriculum learning, Graph embedding, Data privacy

## 1 引言

推荐系统是解决信息过载的主要方法,它依据用户与项目的交互历史,主动为用户提供具有个性化的服务<sup>[1]</sup>。传统的推荐系统算法包括基于内容的推荐算法和协同过滤推荐算法<sup>[1-2]</sup>。随着神经网络的发展,基于深度学习构建推荐算法是目前的主流思路<sup>[3-4]</sup>。

现有大部分推荐系统是在用户和商品信息都非常丰富的情况下,研究如何进行高效的特征嵌入和组合来提升模型的推荐能力。近几年,互联网公司不断有“大数据杀熟”的新闻报道,引发了人们对数据隐私泄露与滥用的担忧<sup>[5]</sup>。尤其是推荐系统这类对用户信息依赖很强的应用,成为了讨论数据隐私保护的热点场景。因此,如何减少推荐系统对信息的依赖,同时保证一定的用户体验,是目前推荐算法面临的主要

基金项目:四川省重点研发计划(22ZYFG0034);四川省重点实验室开放课题(SCITLAB-20001);四川大学博士后交叉学科基金(10822041A2137);四川大学宜宾市合作项目(2020CDYB-30)

This work was supported by the Sichuan Science and Technology Program(22ZYFG0034), Intelligent Terminal Key Laboratory of Sichuan Province(SCITLAB-20001), Post Doctoral Interdisciplinary Innovation Fund(10822041A2137) and Sichuan University and Yibin Cooperation Program(2020CDYB-30).

通信作者:彭舰(penguest@scu.edu.cn)

挑战。基于隐私保护技术是解决该问题的主要思路。比如,文献[6]基于差分隐私构建推荐算法,寻求在保证数据集隐私安全的前提下获取推荐质量最优的算法。也有基于匿名化和密码学的推荐模型,所谓匿名技术就是用户在获得个性化推荐时,不要求其真实身份,可以以匿名的方式使用系统<sup>[7]</sup>。密码学方法就是采用加密算法实现对原始数据的隐私保护。此外,采用联邦学习也可以实现隐私保护<sup>[8]</sup>。如,文献[9]提出了面向联邦学习的基于矩阵分解的推荐模型,本地客户端仅上传梯度至服务端进行聚合优化,避免了直接上传用户信息,从而达到保护隐私的目的。

本文认为差分隐私方法属于数据扰动方法,容易导致数据可用性低。匿名化方法尽管简单,但容易受到差分攻击等危险,密码学方法则是计算复杂度高。基于联邦学习的模型,在执行过程中需要考虑客户端与服务端的通信开销,而且不同客户端的数据分布不一致问题容易导致模型收敛性差。为了解决现有推荐模型面临的隐私保护问题,本文创新地提出了基于图嵌入技术的推荐模型,设计了一种不使用用户和物品信息的推荐模型,即基于课程学习和图嵌入的协同推荐模型(简称为 CLG-CF)。该模型主要利用评分表在隐式反馈的场景学习用户和物品嵌入。限制不使用用户特征和物品特征之后,仅仅依靠隐式反馈表中的信息,也很难达到一个好的推荐效果,因此需要从评分表中挖掘更多的有用信息,并找出潜在的辅助信息来提升用户和物品嵌入的质量。由于评分表和图可以相互转换,受此启发,本文将评分表转化成一种特殊的异质图,即二部图。在二部图中,只有两种类型的节点,但是每种类型节点的邻居都是另一种类型的节点,这表明没有两个相同类型的节点直接相连。通过图卷积挖掘二部图中的结构信息,实现对用户和物品特征的学习。在训练阶段,应用课程学习来正则化损失,解决“伪正样本”和“伪负样本”的问题,主要思路是按照样本损失进行排序并筛选参与学习指导的样本。课程学习(Curriculum Learning, CL)是近几年逐渐兴起的一个前沿方向,它是一种训练策略,模仿人类的学习过程,主张让模型先从容易的样本开始学习,并逐渐进阶到复杂的样本和知识。

本文的主要创新有两点:1)模型不使用用户和物品的特征,而是使用专门针对二部图的图嵌入方法来聚集邻居信息并更新中心节点,使节点的嵌入携带图的局部结构信息;2)使用课程学习的方法,对模型的最终损失进行正则化,使得节点的嵌入朝着更准确的方向进行更新。

## 2 相关工作

协同过滤模型按照不同的实现方法分为基于内存和基于模型的方法。基于内存的方法是一种启发式的方法,其主要思想是相似的用户会喜欢相同的物品以及相似的物品会被同一个用户喜欢。基于模型的算法最早也借鉴了这种思想,利用聚类模型对用户或者物品进行划分,如 Ling 等<sup>[10]</sup>为缓解冷启动提出了一种基于用户聚类的方法, Zhou 等<sup>[11]</sup>提出的功能矩阵分解模型,将矩阵分解与决策树模型相结合,能够在用户冷启动过程中协助选择物品。

基于内容的推荐是根据用户交互过的物品,选择其他相似属性的物品来进行推荐。这类模型起源于信息检索的方法,物品通过特征提取的方法进行表示,系统基于用户交互过

的物品特征,学习用户的兴趣,然后计算用户与其他未交互物品的匹配程度。因此涌现了许多关于物品特征表示和用户兴趣建模的模型,如 Zhang 等<sup>[12]</sup>提出的自适应过滤器可以用于动态建模用户兴趣,Robertson 等<sup>[13]</sup>提出的自适应过滤系统可以更好地提取物品特征, Zhou 等<sup>[14]</sup>提出利用注意力机制的 DIN 模型来进行用户兴趣的建模,后续考虑到用户兴趣随时间改变,又提出了基于门控循环网络的 DIEN 模型<sup>[15]</sup>。

随着神经网络和深度学习的发展,深度推荐系统开始出现。相比传统的推荐系统,一方面可以更好地对用户和物品的交互进行建模,因为在矩阵分解模型中使用用户和物品的内积进行交互的建模,无法准确地表达出用户与物品之间复杂的关系,为了解决这个问题,最早由 Salakhutdinov 等<sup>[16]</sup>提出了一种两层限制玻尔兹曼机来模拟用户对项目的评分;另一方面也可以更好利用其他辅助信息来加强模型的性能,因为传统的协同过滤模型仅仅能够利用历史交互信息,如 Wang 等<sup>[17]</sup>提出的 DKN 模型,将知识图谱融入到新闻推荐中。

基于图卷积网络的推荐算法是目前热门的研究点, Pinsage<sup>[18]</sup>、IGMC<sup>[19]</sup>、以及 GCMC<sup>[20]</sup>是几个具有代表性的算法。Pinsage 延续了 GraphSAGE 模型提出的局部卷积操作,只采样部分邻居节点来进行聚合操作,主要解决传统的 GCN 模型计算时间随着图的增大会不断提高,无法应用于大型图的问题。GCMC 主要应用于用户对于物品有具体评分的场景。IGMC 模型与其他模型最大的不同是利用用户和物品节点的局部图来进行评分预测,而不是直接利用用户和物品的嵌入。不同于现有方法, CLG-CF 模型中的图嵌入部分,针对二部图的特殊结构进行了特殊处理:在卷积过程中,为获取同类型节点的信息,一次聚合操作不再是只单纯聚合一阶邻居信息,而是每次都聚合到二阶邻居,并且聚合过程考虑了二部图的结构特性,从而增强了对用户和物品特征的学习能力。

## 3 CLG-CF 模型

### 3.1 相关定义

首先对本文模型使用的场景、数据源和问题进行定义。用  $G=(U, V, E)$  来表示由评分表生成的二部图,其中  $U$  表示用户节点集合,  $V$  表示物品节点集合,  $E \in U \times V$  表示边的集合,即边的两个端点一定是来自于不同节点集的两个节点。  $A \in \{0, 1\}^{|U| \times |V|}$  表示原始的隐式反馈评分表,也表示邻接矩阵。其中  $A_{i,j}$  表示用户  $u_i \in U$  和物品  $v_j \in V$  是否有交互,取值为 1 表示有,取值为 0 表示没有。模型的目标就是以二部图  $G$  和邻接矩阵  $A$  为输入,输出每一个用户和物品的嵌入。为了便于表示,用  $u_i$  和  $v_i$  直接作为第  $i$  个用户  $u_i$  和第  $i$  个物品  $v_i$  的嵌入表示,其中  $u_i \in R^d, v_i \in R^d$ , 且  $d \ll \min\{|U|, |V|\}$ 。

### 3.2 CLG-CF 模型概述

本节提出的 CLG-CF 模型如图 1 所示。模型以原始评分表构造的初始图为输入,图中的每个节点都是高维稀疏的表示;经过嵌入层,得到初始化嵌入图;然后通过图卷积网络的聚合和更新操作,得到最终的嵌入图,图中每个节点的嵌入都已经携带丰富的信息;最后将每一对的  $(u, v)$  的嵌入通过一个三层的神经网络将其中的隐藏信息学习并表达出来,再将神经网络的输出通过 sigmoid 函数变为模型的最终输出。值得注意的是,图中的负采样,是为了在隐式反馈场景下更好地

更新模型的参数,因此仅在模型训练时使用。下面将详细介绍模型训练的各个步骤。

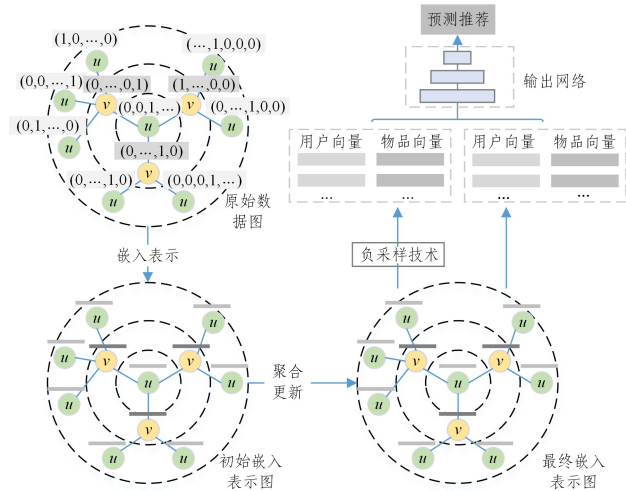


图1 CLG-CF模型概览

Fig.1 Overview of CLG-CF

### 3.3 输入和初始嵌入

隐式反馈的场景下,输入只有原始的评分表和由评分表构造出的初始二部图,用户和物品唯一可用的信息就是它们的编号,因此通常先将用户和物品进行 one-hot 编码:

$$\begin{aligned} & \underbrace{(100\dots 0)}_{u_1} \underbrace{(010\dots 0)}_{u_2} \dots \underbrace{(000\dots 1)}_{u_m} \\ & \underbrace{(100\dots 0)}_{v_1} \underbrace{(010\dots 0)}_{v_2} \dots \underbrace{(000\dots 1)}_{v_n} \end{aligned} \quad (1)$$

其中,  $m$  表示用户的数量,  $n$  表示物品的数量,  $u_i \in R^m$ ,  $v_i \in R^n$  且  $\text{sum}(u_i) = 1, \text{sum}(v_i) = 1$ 。因此每一个物品和用户的嵌入都是高维稀疏的,这样的表示很难用于学习和预测,需要用嵌入技术将其转换为低维稠密的向量,嵌入方式如下:

$$\begin{aligned} u_i &= U \times u_i \\ v_i &= V \times v_i \end{aligned} \quad (2)$$

其中,  $U \in R^{m \times d}$  和  $V \in R^{n \times d}$  分别是用户嵌入矩阵和物品嵌入

矩阵,  $d$  是嵌入维度。等式右边的  $u_i$  和  $v_i$  表示式(1)所示的 one-hot 向量。

### 3.4 聚合器和更新器

聚合操作和更新操作是一个图神经网络的重要组成部分,也是区分不同模型的关键之一。CLG-CF 模型的聚合和更新操作如图 2 所示。图 2(a) 表示对于用户节点的更新,首先让中心节点的一阶邻居成为次中心节点,先对次中心节点进行聚合和更新操作:

$$h_i = \sigma \left( \frac{1}{|\mathcal{N}(v_i)|} \sum_{u_i \in \mathcal{N}(v_i)} W_u^1 \cdot u_i \right) \quad (3)$$

$$v_i^{\text{new}} = h_i$$

其中,  $\sigma$  表示 leakyReLU 激活函数,  $W_u^1 \in R^{d_1 \times d}$  表示参数矩阵,  $d_1$  表示聚合信息  $h_i$  的维度,  $\mathcal{N}(v_i)$  表示节点  $v_i$  的一阶邻居。在更新操作时,直接使用聚合信息作为节点更新后的嵌入  $v_i^{\text{new}}$ ,而没有加入节点原始的嵌入  $v_i$ ,这是因为物品节点的一阶邻居都是用户节点,因此聚合信息  $h_i$  可以看作是和这个物品有交互的用户群体信息,而  $v_i$  表示物品的嵌入,因此不参与用户群体的表示。次中心节点更新之后,再更新中心节点:

$$\bar{h}_i = \sigma \left( \frac{1}{|\mathcal{N}(u_i)|} \sum_{v_i^{\text{new}} \in \mathcal{N}(u_i)} W_u^2 \cdot v_i^{\text{new}} \right) \quad (4)$$

$$u_i^{\text{new}} = \text{ReLU}(W_u^3 \cdot [\bar{h}_i | u_i])$$

其中,  $\sigma$  表示 leakyReLU 激活函数,  $W_u^2 \in R^{d \times d_1}$  和  $W_u^3 \in R^{d \times 2d}$  都表示参数矩阵,  $\mathcal{N}(u_i)$  表示节点  $u_i$  的一阶邻居,  $\bar{h}_i$  表示聚合的信息。中心节点更新时使用聚合信息和节点最原始的嵌入  $u_i$ ,因为中心节点本也是用户嵌入,并且是和所有次中心节点表示物品都有交互的用户。在次中心节点全部更新后,携带的信息已经从一个物品变成了和这个物品有过交互的用户群体的表示。实际上,中心节点的信息在每一个次中心节点的更新表示  $v_i^{\text{new}}$  中都存在,但是模型依旧在更新操作中使用了中心节点的原始嵌入,是一个类似残差连接的操作,一方面是增强模型的表达能力,另一方面是防止图卷积次数太多,导致节点嵌入过平滑问题。

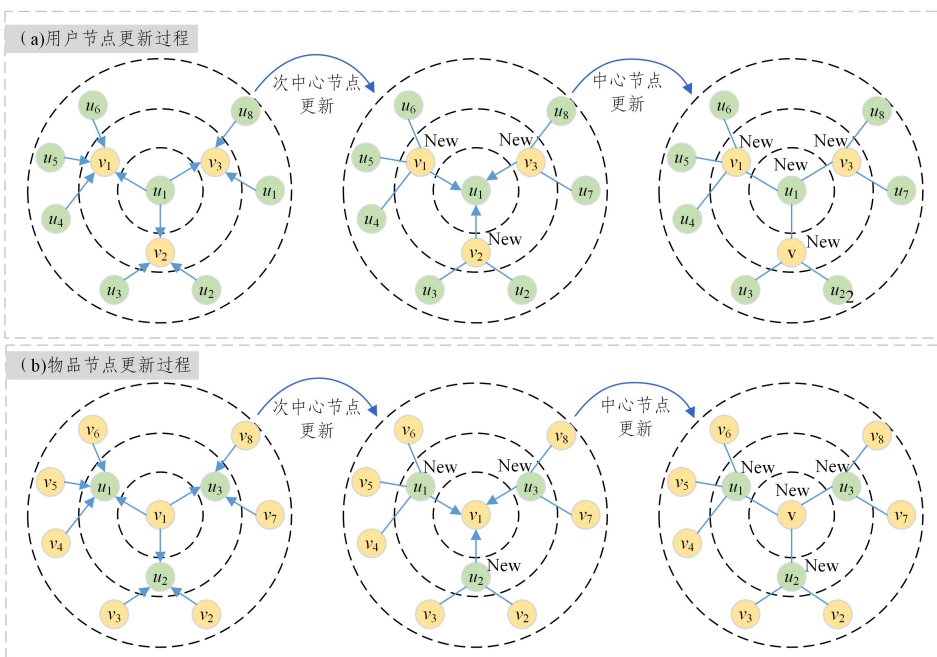


图2 聚合及更新步骤

Fig.2 Aggregation and update steps

图 2(b)给出了物品节点的更新步骤,和更新用户节点一样,首先对中心节点的一阶邻居、次中心节点进行更新:

$$\mathbf{h}_i = \sigma \left( \frac{1}{|\mathcal{N}(u_i)|} \sum_{v_j \in \mathcal{N}(u_i)} \mathbf{W}_v^1 \cdot \mathbf{v}_j \right) \quad (5)$$

$$\mathbf{u}_i^{\text{new}} = \mathbf{h}_i$$

然后,对中心节点进行更新:

$$\bar{\mathbf{h}}_i = \sigma \left( \frac{1}{|\mathcal{N}(v_i)|} \sum_{u_j^{\text{new}} \in \mathcal{N}(v_i)} \mathbf{W}_u^2 \cdot \mathbf{u}_j^{\text{new}} \right) \quad (6)$$

$$\mathbf{v}_i^{\text{new}} = \text{ReLU}(\mathbf{W}_v^3 \cdot [\bar{\mathbf{h}}_i | \mathbf{v}_i])$$

其中,  $\sigma$  表示 leakyReLU 激活函数,  $\mathbf{W}_v^1 \in R^{d_1 \times d}$ ,  $\mathbf{W}_v^2 \in R^{d \times d_1}$  和  $\mathbf{W}_v^3 \in R^{d \times 2d}$  是更新物品节点使用的参数矩阵,  $\mathbf{h}_i$  和  $\bar{\mathbf{h}}_i$  依旧分别表示更新次中心节点和中心节点时的聚合信息。和更新用户节点一样,次中心节点更新时不使用节点的原始嵌入  $\mathbf{u}_i$ , 中心节点更新时使用原始节点嵌入  $\mathbf{v}_i$ , 这样最终物品节点聚合的信息全部都来自于物品嵌入。

虽然模型在进行图卷积的过程中,并没有对不同节点进行明确区分,但是实际上隐含了协同过滤模型中的重要想法:有更多共同交互物品的用户更加相似,被更多共同用户交互过的物品更加相似。CLG-CF 模型与其他模型的聚合更新操作不同的地方在于以下两点:1) 一个中心节点的更新步骤包含两个聚合-更新的操作,首先对次中心节点进行更新,之后再对中心节点进行更新,并且由于其两次聚合-更新步骤并不一样,因此和其他模型进行两次图卷积操作依旧有区别;2) 因为在两个更新阶段采取不同的更新方式,使得基于二部图更新时,仅利用和自己相同类型节点的嵌入,更加充分地利用了二部图的结构特性。相同点则在于,也可以将聚合-更新步骤进行叠加,以聚合空间距离更远的同类型节点的信息。

### 3.5 输出

在 CLG-CF 模型中,采用一个三层的神经网络来预测(用户,物品)对的交互概率。

$$\text{hidden} = \text{ReLU}(\mathbf{W}_1 \cdot [\mathbf{u}_i | \mathbf{v}_i])$$

$$\text{hidden} = \text{ReLU}(\mathbf{W}_2 \cdot \text{hidden}) \quad (7)$$

$$\text{out} = \mathbf{W}_3 \cdot \text{hidden}$$

其中,  $\mathbf{W}_1 \in R^{2d \times l_1}$ ,  $\mathbf{W}_2 \in R^{l_1 \times l_2}$  及  $\mathbf{W}_3 \in R^{l_2 \times 1}$  分别是各个隐藏层的参数矩阵,  $l_1$  和  $l_2$  分别是第一和第二隐藏层的维度。最终将得到的输出转化为  $[0, 1]$  之间的概率,并将其作为预测值:

$$\hat{y} = \text{sigmoid}(\text{out}) \quad (8)$$

其中,  $\hat{y} \in (0, 1)$  表示对于一个  $(u, v)$  对,模型预测其有交互的概率。

### 3.6 训练

CLG-CF 模型将交互概率预测问题当作一个二分类问题来处理,因此采用交叉熵损失作为损失计算函数。隐式反馈场景中使用负采样的方法得到负样本<sup>[21]</sup>,模型的最终损失的计算式如下:

$$\text{Loss} = -(\mathbb{E}_{(u_i, v_j) \in \mathcal{B}^+} \log(\hat{y}(\mathbf{u}_i, \mathbf{v}_j))) + \mathbb{E}_{(u_i, v_j) \in \mathcal{B}^-} \log(1 - \hat{y}(\mathbf{u}_i, \mathbf{v}_j)) \quad (9)$$

其中,  $\mathcal{B}^+$  表示从数据集中抽取的一个正样本集合,  $\mathcal{B}^-$  表示针对  $\mathcal{B}^+$  进行负采样得到的负样本集合。式(9)直接计算损失会存在一个问题,即负采样得到的负样本并不一定是真正的负样本(“伪负样本”),可能是潜在的正样本。此外,观察到的

正样本,也并非一定都是真正的正样本(“伪正样本”),即用户有过交互的物品并不一定是用户喜欢的。

CLG-CF 使用课程学习的方法来解决这个问题<sup>[22]</sup>,即延后两种伪样本的学习。损失越大的样本说明其越难学习,课程学习就会将这个样本的学习延后。最近有研究提到,损失越大,说明这些样本最开始的分类就不正确<sup>[23]</sup>,也刚好对应了“伪正样本”和“伪负样本”。因此,应用课程学习来正则化损失需要识别样本是否为伪正(负)样本。本文设计了损失阈值的计算方法:

$$\xi = L * \min(c \cdot (1+r)^{\text{epoch}}, m) \quad (10)$$

其中,  $\xi$  表示计算出的损失阈值,  $L$  表示正负样本计算损失之后的排序,  $*$  表示取出刚好位于比例的值,  $c$  表示初始损失计算比例,  $r$  表示每一轮比例增加的程度,  $\text{epoch}$  表示轮次,  $m$  表示最大的损失计算比例。最终,损失计算式为:

$$\begin{aligned} \text{Loss} = & -(\mathbb{E}_{(u_i, v_j) \in \mathcal{B}^+} \log(\hat{y}(\mathbf{u}_i, \mathbf{v}_j))) + \mathbb{E}_{(u_i, v_j) \in \mathcal{B}^-} \log \\ & (1 - \hat{y}(\mathbf{u}_i, \mathbf{v}_j)) \\ \text{s. t. } & \mathbb{E}_{(u_i, v_j) \in \mathcal{B}^+} \log(\hat{y}(\mathbf{u}_i, \mathbf{v}_j)) \leq \xi \& \\ & \mathbb{E}_{(u_i, v_j) \in \mathcal{B}^-} \log(1 - \hat{y}(\mathbf{u}_i, \mathbf{v}_j)) \leq \xi \end{aligned} \quad (11)$$

随着训练的推进,每一轮将会有  $|\mathcal{B}^+ + \mathcal{B}^-| \cdot (1-m)$  个样本被视作伪正样本或者伪负样本而不参与损失的计算。一般会将  $m$  设置成为一个非常接近 1 的值。因为采用批量梯度下降的方法,并不一定每个批量都会存在,所以  $m$  接近 1 时遗失的正确样本更少,对模型的影响更小。

### 3.7 算法分析与描述

CLG-CF 模型的训练流程如算法 1 所示。为了和节点表示  $u, v$  区分,在算法描述中使用  $\tilde{u}, \tilde{v}$  来表示用户节点和物品节点的嵌入。首先,在步骤 3 为每个节点确定邻居时,其时间复杂度为  $O(E)$ ;在步骤 5 和步骤 6 中,在二部图上进行卷积操作,每一次全图的卷积可以视作在同一条边进行两次信息传递,总共进行  $2|E|$  次的信息传递,时间复杂度为  $O(|E|Rd)$ ;在步骤 7-步骤 10,虽然分成不同的 batch 进行计算,但是总的样本数保持不变,因此负采样和计算损失的样本总数基数都是  $|E|$ ,时间复杂度为  $O(|E|(d l_1 + l_1 l_2 + l_2))$ ;对于步骤 11,更新嵌入参数的时间复杂度为  $O((|U| + |V|) \lfloor |E| / |\mathcal{B}| \rfloor d)$ ,其中  $\lfloor \cdot \rfloor$  表示向上取整。综上所述,模型在训练时的时间复杂度为  $O(|E|(Rd d_1 + d l_1 + l_1 l_2 + l_2) + |U| |V| \lfloor |E| / |\mathcal{B}| \rfloor d)$ 。

#### 算法 1 CLG-CF 模型的训练流程

输入:二部图  $G$ ,邻接矩阵  $\mathbf{A}$ ,用户节点集合  $U$ ,物品节点集合  $V$ ,边集合  $E$ ,训练次数  $T$ ,图卷积层数  $R$ ,初始损失计算比例  $c$ ,最大损失比例  $m$ ,增长系数  $r$

输出:用户嵌入  $\tilde{u}^{\text{opt}}$ ,物品嵌入  $\tilde{v}^{\text{opt}}$ ,非嵌入参数  $\mathbf{W}^{\text{opt}}$

1. 初始化参数  $\mathbf{W}$
2. 初始化  $\tilde{u}, \tilde{v}$
3. 根据  $G$  计算每个节点的邻居集合:  $\mathcal{N}(u), \forall u \in U, \mathcal{N}(v), \forall v \in V$
4. For  $t=1$  to  $T$ ; For  $r=1$  to  $R$  do:
  5. 更新次中心节点:式(3)及式(5)
  6. 更新中心节点:式(4)及式(6),对于每一个批量  $\mathcal{B}^+ : (u, v) \in E$
  7. 负采样得到  $\mathcal{B}^- : (u, v) \notin E$
  8. 通过神经网络得到预测值,计算损失并排序得到  $L$
  9. 计算阈值:  $\xi = L * \min(c \cdot (1+r)^t, m)$
  10. 根据阈值得到计入最终损失的样本并计算损失:式(11)

11. 更新  $\tilde{\mathbf{u}}, \tilde{\mathbf{v}}, \mathbf{W}$
12.  $\mathbf{u}^{\text{opt}} = \tilde{\mathbf{u}}, \mathbf{v}^{\text{opt}} = \tilde{\mathbf{v}}, \mathbf{W}^{\text{opt}} = \mathbf{W}$

## 4 实验

### 4.1 实验设置

本文使用 DBLP, Movie Length 100K(ML-100k) 和 Movie Length 1M(ML-1M) 3 个真实数据集, 具体的统计信息如表 1 所列。对于每个数据集, 随机划分 80% 的边作为训练集, 将其余 20% 作为测试集。

表 1 数据集的统计信息  
Table 1 Statistics of datasets

Datasets	U	V	E	Sparsity/%
DBLP	6001	1308	29256	99.6
ML-100K	943	1682	100000	93.7
ML-1M	6040	3682	1000000	95.5

对比的基准模型如下。

- 1) DeepWalk<sup>[24]</sup>: 利用随机游走将图结构的数据转化为序列数据进行图节点的嵌入。
- 2) PinSage<sup>[18]</sup>: 基于随机游走和图卷积网络进行节点表示学习。
- 3) GCMC<sup>[19]</sup>: 针对不同类型的边采取不同的卷积方式来提升模型的性能。
- 4) IGMC<sup>[20]</sup>: 将节点与节点之间的匹配问题转化为局部图的匹配问题, 一定程度上实现了不借助辅助信息解决冷启动的问题。
- 5) NGCF<sup>[21]</sup>: 显式地将协同信号, 即边的信息编码出来, 并在图卷积过程中加入到节点的嵌入中来提升嵌入的表达力。
- 6) BIGI<sup>[25]</sup>: 利用二部图的特性进行图卷积, 同时利用全局图嵌入和局部图嵌入的趋同性, 将全局信息加入节点的嵌入。

模型对比时的一些参数设置如下: 所有模型的节点嵌入维度固定为 128, 图卷积层数设置为 2, 图卷积聚合信息的维度为 128, 学习率设置为 0.001, dropout 设置为 0.3, 每个样本负采样数为 1。在 DBLP 和 ML-100K 数据集的实验中, 设置 batch size 为 128, 在 ML-1M 数据集的实验中将 batch size 实验设置为 512。如没有特别说明, 模型的其他特有参数根据

原始论文描述进行设置。

### 4.2 评价指标

对于本小节的 TOP-K 推荐任务, 采用如下 4 个指标, 即 F1-score, NDCG, MAP 和 MRR, 且对于后 3 个指标, 分别有 TOP-5 和 TOP-10 两个不同维度。

记 recall 表示召回率, precision 表示准确率, F1-score 的计算方式如下:

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \quad (12)$$

NDCG 的计算方式如下:

$$NDCG = \frac{1}{|U|} \sum_{j=1}^{|U|} \frac{\sum_{i \in index_j^+} 1/\log(i+1)}{\sum_{i=1}^{P_j} 1/\log(i+1)} \quad (13)$$

其中,  $index_j^+$  表示给用户  $u_j$  进行 TOP-K 推荐中有交互的物品的排名,  $P_j$  表示用户  $u_j$  的潜在交互物品的个数。

MRR 和 MAP 的计算式如下:

$$MAP = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{1/j}{K} \quad (14)$$

$$MRR = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{1}{rank_i}$$

其中,  $rank_i$  表示对用户  $u_i$  的 TOP-K 推荐中第一个正确的物品的排名。

### 4.3 对比实验

本小节进行与基准模型的性能对比实验, 表 2—表 4 分别列出了在 DBLP, ML-100K 和 ML-1M 数据集上的实验结果, 可以从中得出以下结论: 1) DeepWalk 模型相对于其他模型来说, 性能总是最差的, 因为 DeepWalk 模型虽然也利用了图结构, 但是其最终使用的是序列数据处理的方法, 并没有直接使用专门用于图结构的图神经网络; 2) CLG-CF 和 BIGI 两个模型相对于其他基于图卷积的模型性能有一定提升, 因为这两个模型是专门针对隐式反馈场景下的模型, 不需要依赖辅助信息; 3) CLG-CF 模型在第一个数量指标上劣于 BIGI, 但是在后三个排名指标上通常优于 BIGI, 这说明 CLG-CF 在推荐过程中, 虽然推荐正确的数量略少于 BIGI 模型, 但是相关物品在推荐列表中总是靠前。但在现代推荐系统中, 用户并不会明显感知到推荐列表中喜好物品的数量, 但是将用户喜欢的物品放在推荐列表前面, 可以带来更好的用户体验。

表 2 DBLP 数据集的 TOP-K 推荐性能比较

Table 2 TOP-K recommendation performance comparison on DBLP dataset

	F1@10	NDCG@5	NDGC@10	MAP@5	MAP@10	MRR@5	MRR@10
DeepWalk	6.93	6.6	9.12	4.23	5.29	10.44	11.7
PinSage	12.55	21.17	23.97	16.04	17.3	29.68	30.84
GCMC	12.02	22.18	24.62	17.98	19.12	30.65	31.7
IGMC	12.18	22.65	25.17	18.43	19.61	31.30	32.36
NGCF	12.38	23.38	25.58	18.50	19.51	32.48	33.4
BIGI	<b>14.27</b>	25.39	<b>28.28</b>	20.15	21.49	35.35	36.51
CLG-CF	12.99	<b>26.34</b>	28.16	<b>20.90</b>	<b>21.8</b>	<b>36.13</b>	<b>36.94</b>

表 3 ML-100K 数据集的 TOP-K 推荐性能比较

Table 3 TOP-K recommendation performance comparison on ML-100K dataset

	F1@10	NDCG@5	NDGC@10	MAP@5	MAP@10	MRR@5	MRR@10
DeepWalk	14.2	9.32	13.13	3.54	4.92	46.83	48.75
PinSage	21.68	14.51	20.27	6.18	9.13	64.77	65.76
GCMC	20.65	13.87	19.21	5.84	8.43	62.21	63.53

(续表)

	F1@10	NDCG@5	NDGC@10	MAP@5	MAP@10	MRR@5	MRR@10
IGMC	18.81	12.20	17.27	4.82	7.18	59.13	60.46
NGCF	21.64	14.49	20.29	6.15	9.11	64.62	65.55
BIGI	<b>23.36</b>	15.92	<b>22.14</b>	7.15	<b>10.50</b>	67.7	68.78
CLG-CF	22.38	<b>16.11</b>	21.65	<b>7.20</b>	10.17	<b>69.05</b>	<b>70.01</b>

表4 ML-1M数据集的TOP-K推荐性能比较

Table 4 TOP-K recommendation performance comparison on ML-1M dataset

(单位: %)

	F1@10	NDCG@5	NDGC@10	MAP@5	MAP@10	MRR@5	MRR@10
DeepWalk	2.22	1.37	1.96	0.36	0.51	5.40	5.94
Pinsage	3.77	2.69	3.87	1.08	1.36	8.99	9.52
GCMC	3.60	2.71	3.64	1.11	1.39	8.86	9.39
IGMC	3.47	2.58	3.72	1.08	1.36	8.60	9.15
NGCF	3.82	2.79	3.99	1.17	1.48	9.36	9.91
BIGI	<b>4.06</b>	3.00	4.32	1.25	1.61	9.72	11.32
CLG-CF	<b>4.06</b>	<b>3.35</b>	<b>4.56</b>	<b>1.40</b>	<b>1.74</b>	<b>9.97</b>	<b>12.63</b>

#### 4.4 消融实验

本小节采取了以下3个模型变体进行消融实验,目的是验证CLG-CF模型所使用的课程学习方案的有效性。

1) Self-space: 自适应地选择学习的样本,但和CLG-CF模型使用的按照比例选择不同,此模型是根据一个固定的损失值进行选择。

2) Split: 在每一轮按照比例计算损失阈值时,不再是正负样本一起计算,而是分开计算。

3) Total: 不使用任何课程学习的方法,计算损失时直接使用所有样本。

所有的变体模型超参数与原始模型保持一致,消融实验的结果如表5所列。

表5 消融实验结果

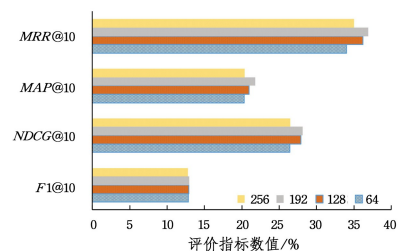
Table 5 Results of ablation experiments

		F1@10	NDCG@3	NDGC@10	MAP@3	MAP@10	MRR@3	MRR@10
DBLP	Self-space	8.69	17.63	19.46	15.94	16.73	24.82	26.12
	Split	12.9	24.21	28.10	19.70	21.65	34.31	36.38
	Total	12.85	23.86	27.70	19.40	21.40	33.53	35.67
	CLG-CF	<b>12.99</b>	<b>24.40</b>	<b>28.16</b>	<b>19.65</b>	<b>21.80</b>	<b>34.80</b>	<b>36.94</b>
ML-100K	Self-space	13.85	7.04	12.87	2.51	4.70	47.82	51.81
	Split	<b>22.78</b>	11.9	21.38	4.98	9.91	65.72	68.62
	Total	22.49	12.13	21.37	5.15	10.05	65.94	68.72
	CLG-CF	22.7	<b>12.51</b>	<b>21.65</b>	<b>5.27</b>	<b>10.17</b>	<b>67.65</b>	<b>70.01</b>
ML-1M	Self-space	2.47	1.1	2.44	0.47	0.81	4.05	6.09
	Split	3.62	2.26	4.17	1.10	1.6	9.83	12.03
	Total	4.01	2.54	4.44	1.10	1.69	9.55	12.06
	CLG-CF	4.06	2.64	4.56	1.16	1.74	9.97	12.63

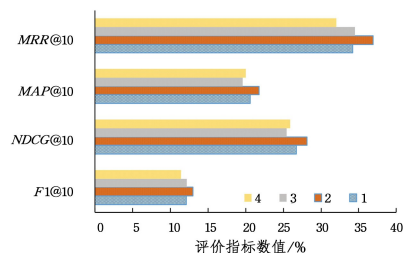
可以得出如下的结论: 1) CLG-CF原始模型在绝大多数情况下的效果都比其他3种变体模型要好(除了在ML-100K数据集中的F1@10指标),这说明提出的课程学习方法是有效的; 2) 根据比例筛选样本的方法(Split和CLG-CF)都比根据绝对阈值筛选样本的方法要好,这是因为在模型训练过程中,每个样本的损失都在动态变化,并且变化的程度不一,因此这种固定损失阈值的方法很难准确地进行样本的筛选; 3) CLG-CF原始模型比Split模型效果更好,这是因为Split模型将正样本和负样本分开来计算,表明在每轮计算中,一定会出现相同数量“伪正样本”和“伪负样本”,但实际上出现的频率并不一样; 4) Split变体模型相比不使用任何课程学习方法的模型(Total)则时优时劣,说明课程也需要和模型进行适配才能发挥出应有的效果。

#### 4.5 超参分析实验

CLG-CF模型有两个重要的超参数: 节点嵌入维度和图卷积层数。本小节就是对这两个超参数进行分析,图3给出了在DBLP数据集上这两个参数对模型性能的影响。



(a) 嵌入维度



(b) 图卷积层数

图3 嵌入维度对模型性能的影响

Fig. 3 Impact of hyper-parameter on performance

可以得到如下的信息: 1) 随着节点嵌入维度的增加,模型

的性能都是先增加后下降,这是因为嵌入维度越大,模型越容易过拟合;2)模型的性能随着图卷积层数都是先上升后下降。这是因为在卷积层数较少时,模型聚合信息的能力过于局限,只能聚合与自己紧邻的节点的信息,而当卷积层数太多时,会引入一些噪声信息,并且节点的嵌入会过于平滑,从而导致模型性能的下降。

#### 4.6 课程设计探索

在 CLG-CF 模型中,采用按比例筛选样本的课程学习方法,并且比例会最终达到一个固定的最大值,初始比例的设置就显得尤为重要。因此本小节从初始比例的设置这个方面来探索课程设计。图 4 给出了 ML-1M 数据集上,不同的初始比例对于模型评价的 4 个指标的影响,可以总结出如下信息:初始比例的设置对于模型性能的影响不再像嵌入维度和图卷积层数那样,呈现出先上升后下降的趋势,而是一种无规律的变化。

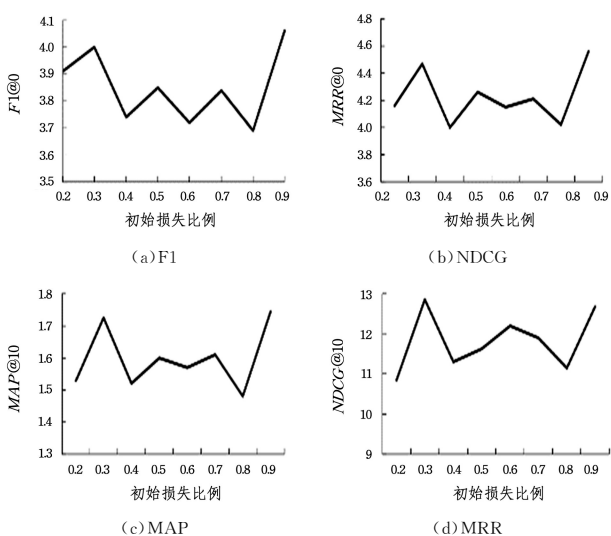


图 4 起始损失比例对模型的影响

Fig. 4 Impact of initial loss ratio on performance

综上所述,可以看出在进行课程学习时,数量和排名两个方面的性能都要兼顾。但由于课程学习的方法与模型性能之间的关系很复杂,在设计课程时需要和模型相匹配,否则就会像 4.4 节中的 Split 变体模型和 Self-space 变体模型一样,性能有时还不如没有课程加入的时候。

**结束语** 本文将 TOP-K 任务中的用户和物品匹配的问题看作一个二分类的问题,并且有如下限制:评分是隐式反馈且不使用用户和物品的属性特征。这样可使用的特征就只有用户和物品编号,并且有必要进行负采样来训练。因此提出了 CLG-CF 模型,首先是引入图卷积网络,来利用二部图中的结构信息加强节点嵌入的表达能,并且针对二部图做了优化;然后利用课程学习来缓解负采样会出现的“伪负样本”问题,同时可以缓解“伪正样本”的问题。模型在 3 个真实的数据集上进行了大量实验,证明了模型在 TOP-K 推荐任务上的优秀性能,并且通过消融实验证明了提出的课程学习方法的有效性。此外,本文还进行了超参数分析实验和课程设计探索实验,对算法的特性进行了进一步的分析。

#### 参考文献

[1] REN S L, GUO H J, HUANG W H, et al. Recommendation

Method Based on Attention Mechanism Interactive Convolutional Neural Network[J]. Computer Science, 2022, 49(10): 126-131.

- [2] HUANG F, QIAO S, PENG J, et al. STPR: A Personalized Next Point-of-Interest Recommendation Model with Spatio-Temporal Effects Based on Purpose Ranking[J]. IEEE Transactions on Emerging Topics in Computing, 2021, 9(2): 994-1005.
- [3] XU B, YI P Y, WANG J C, et al. High order Collaborative Filtering Recommendation System Based on Knowledge Graph Embedding[J]. Computer Science, 2021, 48(S2): 244-250.
- [4] XU H, HUANG F, PENG J, et al. Intent-Aware Graph Neural Networks for Session-based Recommendation [C] // International Joint Conference on Neural Networks(IJCNN). 2022:1-8.
- [5] TIAN L, HONG S. Research on legal regulation of the Big Data affinity[J]. Cyber Security and Data Governance, 2022, 41(7): 41-46.
- [6] WANG L E, LI X C, LIU H Y. News recommendation method with knowledge graph and differential privacy[J]. Journal of Computer Applications, 2022, 42(5): 1339-1346.
- [7] WEI R X, TIAN H, SHEN H. Improving  $k$ -anonymity based privacy preservation for collaborative filtering[J]. Computers & Electrical Engineering, 2018, 67: 509-519.
- [8] ZHANG H L, LI P D, WU J, et al. A Survey on Privacy-preserving Federated Recommender Systems[J]. Acta Automatica Sinica, 2022, 48(9): 2142-2163.
- [9] LIN G, LIANG F, PAN W, et al. Fedrec: Federated recommendation with explicit feedback [J]. IEEE Intelligent Systems, 2020, 36(5): 21-30.
- [10] LING Y, CUO O, CAI F, et al. User-based clustering with top-n recommendation on cold-start problem[C] // 2013 Third International Conference on Intelligent System Design and Engineering Applications. IEEE, 2013: 1585-1589.
- [11] ZHOU K, YANG S H, ZHA H. Functional matrix factorizations for cold-start recommendation[C] // Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2011: 315-324.
- [12] ZHANG Y, CALLAN J, MINKA T. Novelty and redundancy detection in adaptive filtering[C] // Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2002: 81-88.
- [13] ROBERTSON S. Threshold setting and performance optimization in adaptive filtering[J]. Information Retrieval, 2002, 5(2): 239-256.
- [14] ZHOU G, ZHU X, SONG C, et al. Deep interest network for click-through rate prediction[C] // Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 1059-1068.
- [15] ZHOU G, MOU N, FAN Y, et al. Deep interest evolution network for click-through rate prediction[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 5941-5948.
- [16] SALAKHUTDINOV R, MNIIH A, HINTON G. Restricted Boltzmann machines for collaborative filtering[C] // Proceedings of the 24th International Conference on Machine Learning. 2007: 791-798.
- [17] WANG H, ZHANG F, XIE X, et al. DKN: Deep knowledge-

- aware network for news recommendation[C]// Proceedings of the 2018 World Wide Web Conference. 2018;1835-1844.
- [18] YING R, HE R, CHEN K, et al. Graph convolutional neural networks for web-scale recommender systems[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018;974-983.
- [19] BERG R V D, KIPF T N, WELING M. Graph convolutional matrix completion[J]. arXiv:170602263, 2017.
- [20] ZHANG M, CHEN Y. Inductive matrix completion based on graph neural networks[J]. arXiv:190412058, 2019.
- [21] WANG X, HE X, WANG M, et al. Neural graph collaborative filtering[C]// Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2019;165-174.
- [22] BENGIO Y, LOURADOUR J, COLLOBERT R, et al. Curriculum learning[C]// Proceedings of the 26th Annual International Conference on Machine Learning. 2009;41-48.
- [23] WANG W, FENG F, HE X, et al. Denoising implicit feedback for recommendation[C]// Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 2021;373-381.
- [24] PEROZZI B, AL-RFOU R, SKIENA S. Deepwalk: Online learning of social representations[C]// Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. 2014;701-710.
- [25] CAO J, LIN X, GUO S, et al. Bipartite graph embedding via mutual information maximization[C]// Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 2021;635-643.



**HUANG Feihu**, born in 1990, Ph.D, is a member of China Computer Federation. His main research interests include deep learning and data mining.



**PENG Jian**, born in 1970, Ph.D, professor, Ph.D supervisor, is an outstanding member of China Computer Federation. His main research interests include big data and wireless sensor network.