

多通道图注意力解耦社交推荐方法*

洪明利^{1,2}, 王靖^{1,2}, 贾彩燕^{1,2}

(1. 北京交通大学计算机与信息技术学院, 北京 100044;
2. 北京交通大学交通数据分析与挖掘北京市重点实验室, 北京 100044)

摘要:提出具有解耦能力的多通道图注意力社交推荐模型,该模型主要包括深度聚类模块、多通道图注意力聚合模块和评分预测模块。其中,深度聚类模块用于对用户和项目进行分组,并利用聚类结果将用户社交图和用户项目图拆分成多个用户社交子图及用户项目子图,以学习用户兴趣分组及用户对不同类项目的兴趣;多通道图注意力聚合模块学习不同子图对预测结果的注意力;评分预测模块将学习到的用户表示向量和项目表示向量输入多层感知机进行评分预测。在多个真实数据集上的实验结果表明:提出的方法优于其他社交推荐算法。与最新的用于社交推荐的图神经网络方法相比,在 Ciao 和 Epinions 数据集上,均方根误差分别降低了 2.26% 和 2.07%,平均绝对误差分别降低了 2.58% 和 3.06%。

关键词:推荐系统;社交网络;图神经网络;注意力网络;深度聚类

中图分类号:TN95 文献标志码:A 开放科学(资源服务)标识码(OSID):

文章编号:1001-2486(2022)03-001-09



听语音
与作者互动
聊科研

Multi-channel graph attention network with disentangling capability for social recommendation

HONG Mingli^{1,2}, WANG Jing^{1,2}, JIA Caiyan^{1,2}

(1. School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China;
2. Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China)

Abstract: A Multi-channel graph attention network social recommendation model with disentangling capability was proposed. This model mainly included three modules: the deep clustering module, the aggregation module based on multi-channel graph attention network, and the rating prediction module. Among them, the deep clustering module was used to group users and items. The clustering results can be used to split user-user social graph and user-item interaction graph into multiple subgraph to learn user interest groups and users' interests in different types of items. The aggregation module learns the attention of different sub-graphs to the prediction results. The rating prediction module input the learned user representation vector and item representation vector into the multilayer perceptron for rating prediction. Extensive experiments on multiple real-world datasets demonstrate that the proposed method is better than other social recommendation algorithms. Specifically, compared with the latest graph neural networks method for social recommendation, the root mean square error is respectively reduced by 2.26% and 2.07% on the Ciao and Epinions datasets, and the mean absolute error is respectively reduced by 2.58% and 3.06%.

Keywords: recommendation system; social network; graph neural network; attention network; deep clustering

随着互联网技术的快速发展,网上的信息也日益丰富。人们难以从大量的信息中提取到有用的信息,这就是著名的信息过载现象。推荐系统^[1]作为一种信息过滤技术,可以有效解决信息过载问题。但传统的协同过滤推荐算法仅使用用户-项目交互信息进行推荐,数据稀疏、分布不均匀等问题,常导致推荐效果不佳^[2]。针对上述问题,人们通过引入社交关系来缓解数据稀疏性等问题,提升了推荐方法的性能。根据社交影响力现象,用户在

购买项目时会受到社交朋友的影响。因此,社交推荐^[3-4]是近年来推荐领域的重要研究方向之一,一系列社交推荐模型^[5-15]相继被提出。

尽管社交推荐在近年取得比较大的成功,但是还存在许多待解决的问题。第一,如何有效提取和融合社交信息和用户-项目交互信息?已有的社交推荐系统使用简单相加、拼接或者多层感知机来融合用户的历史交互信息和社交信息,无法区分用户购买是因为用户本身的兴趣、还是受

* 收稿日期:2021-09-13

基金项目:国家自然科学基金资助项目(61876016,61632004);中央高校基本科研业务费专项资金资助项目(2019JBZ110)

作者简介:洪明利(1997-),男,浙江台州人,硕士研究生,E-mail:19120361@bjtu.edu.cn;

贾彩燕(通信作者),女,教授,博士,博士生导师,E-mail:cyyia@bjtu.edu.cn

到朋友的影响。第二,如何区分出用户的购买动机?目前常见的神经网络会学习用户的嵌入表示,但是用户的多种兴趣纠缠在一起,影响推荐的结果。多成分图卷积协同过滤(multi-component graph convolutional collaborative filtering, MCCF)^[16]使用多个图注意力来提取用户的多个表示,有利于学习用户对不同物品的可能喜好;但是 MCCF 无法区分多个通道表示之间的区别,缺乏可解释性和真正的解耦能力。

因此,本文提出多通道图注意力解耦社交推荐方法(multi-channel graph attention network with disentangling capability for social recommendation, MGAT-D)。该模型主要有三个部分:深度聚类模块、多通道图注意力聚合模块和评分预测模块。其中:深度聚类模块用于对用户和项目进行分组,并利用聚类结果将用户社交图和用户-项目图拆分成多个用户社交子图及用户-项目子图;聚类模块的目的是将同一类项目划分在一个用户-项目子图中,同一类用户划分在一个项目-用户子图中,从而完成用户兴趣的分组。多通道图注意力聚合模块融合不同子图上的信息,其目的是挖掘用户对不同类项目的兴趣以及项目更适合哪一类人群;用户的每个通道的嵌入表示指用户对某类项目的兴趣,从而完成用户兴趣的解耦表示。评分预测模块使用多层感知机进行评分预测。在两个基准数据集上的实验结果表明:本文的方法优于其他同类社交推荐方法,包括 GraphRec^[13]、DANSER^[14]等。

1 相关工作

1.1 社交推荐

随着社交平台的快速发展,社交推荐在近些年越来越受到关注。基于矩阵分解的推荐模型在矩阵分解的基础上融合社交信息以捕捉更具有表达性的用户偏好向量,它的特点是灵活性高、推荐性能较好等。基于矩阵分解的推荐模型可以分为两类:基于矩阵分解的协同分解法和基于矩阵分解的社交正则化法。协同分解法^[6,17]同时分解用户-项目评分矩阵和社交矩阵,两者共享用户特征向量。基于概率矩阵分解的社会推荐模型(social recommendation using probabilistic matrix factorization, SoRec)^[6]是一种基于矩阵分解的协同分解法,该方法共享一个由用户-项目评分矩阵和社会关系矩阵分解的共同的潜在用户特征向量。基于矩阵分解的社交正则化法^[8-9]分解用户-项目评分矩阵,并且用社交矩阵限制用户特征

向量。基于社会正则化的推荐系统(recommender systems with social regularization, SoReg)^[8]是一个带社交正则化的矩阵分解模型,它利用社交矩阵对用户特征向量进行正则化约束。

由于基于矩阵分解的任务只提取了线性信息,所以研究人员又提出了基于深度神经网络(deep neural network, DNN)的非线性推荐方法。基于社会关系深度建模的推荐模型(deep modeling of social relations for recommendation, DeepSoR)^[10]是一种基于深度神经网络的社交推荐模型,该模型首先将社交关系输入到图嵌入模型 node2vec^[18]来预训练用户节点嵌入,然后使用深度神经网络来提取社交关系矩阵中的用户非线性特征向量,并将其集成到概率矩阵分解中进行评分预测。深度社交协同过滤模型(deep social collaborative filtering, DSCF)^[11]是另一种基于深度神经网络的社交推荐模型,它能提取高阶的社交邻居的信息,从而更好提取社交信息和用户-项目交互信息。

1.2 图神经网络

图神经网络^[19-20]是基于图结构的广义神经网络,该网络提供了一种从相邻节点聚合信息以捕获节点的网络拓扑信息的方法。近年来,图神经网络被研究人员用在推荐领域^[21-22]并取得很好的效果。新近提出的 MCCF^[16]在用户物品交互图上使用多个图注意力来提取用户和物品的多个表示。基于图神经网络的社交推荐模型(graph neural networks for social recommendation, GraphRec)^[13]提出了一种基于图注意力网络的社交推荐系统,使用图注意力网络同时在用户物品交互图中捕获交互信息、在社交图中提取社交信息。双图注意力用于推荐系统中多方面社交影响的深度潜在表示(dual graph attention networks for deep latent representation of multifaceted social effects in recommender systems, DANSER)^[14]是一种对偶图注意力神经网络,可以同时捕获用户和物品的社交同质性和社会影响力。其中,社交同质性用邻居的向量表示来聚合,社会影响力用邻居的上下文感知偏好来聚合。社会推荐的神经影响扩散模型(neural influence diffusion model for social recommendation, DiffNet)^[23]使用 GCNs 来捕捉用户偏好在社交网络中的社交扩散过程。但现有模型缺乏学习用户购买动力的能力,不能对用户不同类商品的兴趣及用户兴趣组进行端到端的建模。鉴于近年来兴起的深度聚类方法(deep

embedding clustering, DEC)^[24-25]可以在学习表示向量的同时对数据进行分组。因此,将深度聚类引入图神经网络模型中,以提高模型学习用户对不同类商品兴趣等的的能力,从而提高模型的解耦能力。

2 MGAT-D 模型

本节提出社交推荐问题的形式化定义和模型 MGAT-D 及模型参数的学习过程。

2.1 定义

设有 M 个用户和 N 个项目,令 $U = \{u_1, u_2, \dots, u_M\}$ 和 $V = \{v_1, v_2, \dots, v_N\}$ 分别表示用户集合和项目集合。设用户项目交互矩阵为 $R = [r_{ui}]_{M \times N}$,其中 r_{ui} 表示用户 u 给项目 i 的评分。设 $R_1(u)$ 表示用户 u 评论的项目集合, $R_U(i)$ 表示评论项目 i 的用户集合。设用户社交矩阵为 $S = [s_{ij}]_{M \times M}$, $s_{ij} = 1$ 表示用户 i 与用户 j 是朋友关系。 $S_U(u)$ 表示用户 u 的社交邻居。

通过项目的相似度来构造项目隐式矩阵 $T = [t_{ij}]_{N \times N}$,当同时评论项目 i 和项目 j 的用户数目大于阈值 $h = 5$ 时,设 $t_{ij} = 1$,否则为 0。 $T_1(i)$ 表示项目 i 在项目隐式矩阵上的相邻节点。

社会化推荐的任务可以形式化定义为:给定

用户 u 和项目 i ,基于用户交互矩阵 R 、用户社交矩阵 S 和项目隐式矩阵 T 来预测评分。

2.2 模型架构

图 1 表示模型 MGAT-D 的整体结构图,其核心主要由三部分组成:深度聚类模块、多通道图注意力聚合模块和评分预测模块。深度聚类模块利用聚类技术对用户和项目进行分组,目的是将同一类项目划分在一个用户-项目子图中,同一类用户划分在同一类项目-用户子图中,从而完成用户兴趣的分组。多通道图注意力聚合模块融合不同子图上的信息,其目的是挖掘用户对不同类项目的兴趣以及找到项目更适合哪一类人群。以用户为例,用户每个通道的嵌入表示用户对某类项目的兴趣从而完成兴趣的解耦表示。评分预测模块将学习到的用户表示向量和项目表示向量输入多层感知机进行评分预测。

2.2.1 深度聚类模块

嵌入表示:设 $P = [p_i]_{D \times M}$ 为用户表示向量, $Q = [q_i]_{D \times N}$ 为项目表示向量,其中 D 为表示向量的维度。 P 和 Q 均随机初始化。

用于对用户和项目进行分组,并利用聚类结果将用户-项目图拆分成 K_1 个用户-项目交互子图和 K_2 个项目-用户交互子图,以学习用户和

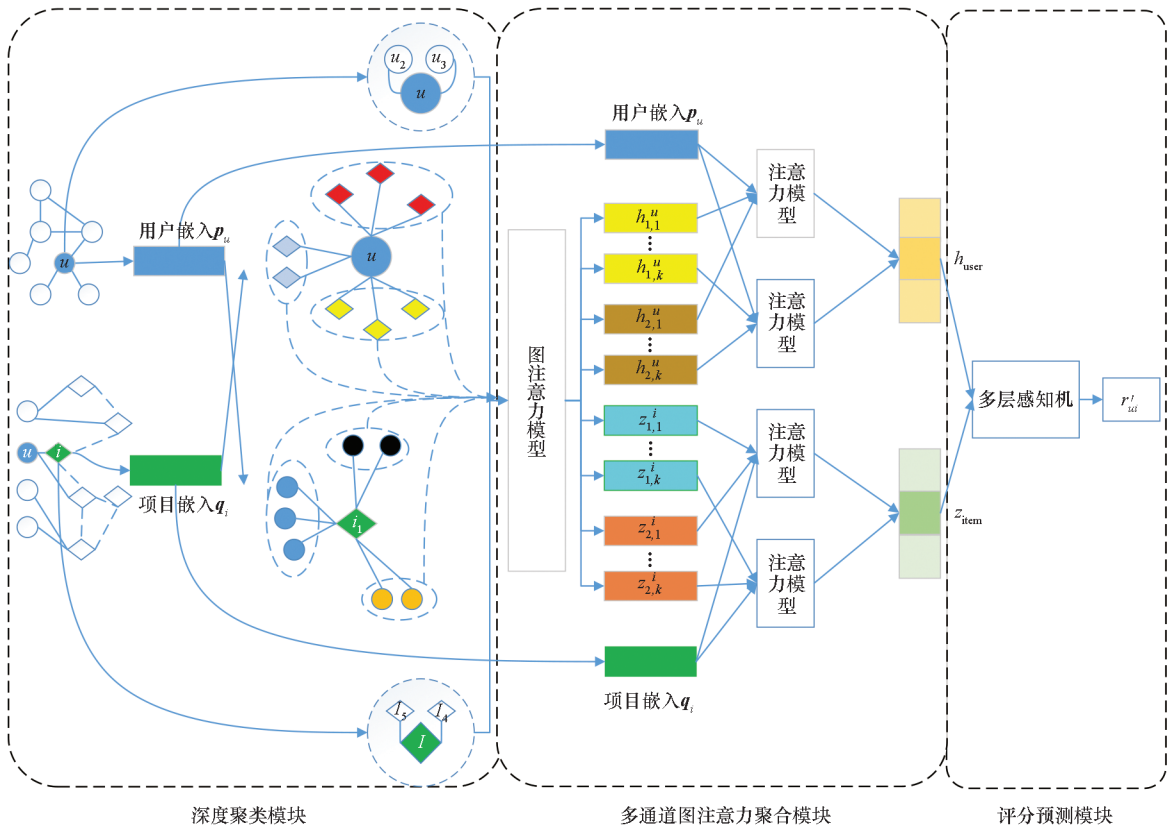


图 1 模型架构

Fig. 1 Model framework

项目的特征分组、用户对不同类项目的兴趣。聚类模型的目的是将同一类项目划分在一个用户 - 项目子图,将同一类用户划分在一个项目 - 用户子图。如图 1 中红色的项目属于同一类,被划分在一个用户 - 项目子图中,黑色的用户为同类用户被划分到同一项目 - 用户子图中。训练用户和项目的嵌入表示 \mathbf{P} 和 \mathbf{Q} ,然后使用聚类算法将用户和项目分别聚成 K_2 和 K_1 类。由于需要多次聚类,所以选用复杂度较低的 K -means 聚类算法^[26]。

设对 M 个用户节点进行聚类的聚类损失为:

$$L_{ucls} = \frac{1}{M} \sum_{u=1}^M \|\mathbf{p}_u - \boldsymbol{\mu}_{c_u}\|^2 \quad (1)$$

式中, c_u 代表用户 u 被分配的聚类索引值, $c_u \in \{1, 2, \dots, K_2\}$, $\boldsymbol{\mu}_{c_u}$ 代表用户 u 归属的类对应索引值的聚类中心。

对 N 个项目节点进行聚类的聚类损失为:

$$L_{icls} = \frac{1}{N} \sum_{i=1}^N \|\mathbf{q}_i - \boldsymbol{\nu}_{d_i}\|^2 \quad (2)$$

式中, d_i 代表项目 i 被分配的聚类索引值, $d_i \in \{1, 2, \dots, K_1\}$, $\boldsymbol{\nu}_{d_i}$ 代表项目 i 归属的类对应索引值的聚类中心。

设用户和项目的聚类结果分别为 $\{u_{cls}\}_{n=1}^{K_2}$ 和 $\{i_{cls}\}_{m=1}^{K_1}$ 。根据聚类结果进行子图划分。将用户 u 交互过的项目中属于一类的拆分成一个用户项目交互子图,同理拆分其他子图。用户 u 的 K_1 个子图为 $\{\mathbf{B}_m^u\}_{m=1}^{K_1}$, 项目 i 的 K_2 个子图为 $\{\mathbf{C}_n^i\}_{n=1}^{K_2}$ 。

2.2.2 多通道图注意力聚合模块

图注意力网络通过注意力机制来衡量相邻节点的不同权重,使得模型能够过滤出图上的噪声并关注重要的相邻节点。因此,使用了图注意力模型来更好地提取用户 - 项目交互图、用户社交图和项目隐式图的信息,从而使得社交影响和用户 - 项目交互信息影响可以通过图传播到相邻的节点。多通道图注意力聚合模块的目的是挖掘用户对不同类项目的兴趣以及项目更适合哪一类人群。以用户为例,用户每个通道的嵌入表示指用户对某类项目的兴趣从而完成兴趣的解耦表示为了实现这个目标,设计了多通道线性变换和两个聚合操作(即用户聚合操作和项目聚合操作)来分别捕获用户和项目的特征表示。

多通道线性变换:

假设用户购买项目是出于多种动机,所以提取用户和项目在不同聚类上的特征表示,每个类都表示一个通道。因此,首先分别为用户和项目

设计了 K_2 和 K_1 个线性变换。

对于用户 u , 设他在第 n 个类的表示向量为:

$$\mathbf{x}_n^u = \mathbf{W}_{nu} \mathbf{p}_u \quad (3)$$

式中, \mathbf{W}_{nu} 是神经网络的权重, \mathbf{p}_u 为初始的用户向量。

对于项目 i , 设它在第 m 个类的表示向量为:

$$\mathbf{y}_m^i = \mathbf{W}_{mi} \mathbf{q}_i \quad (4)$$

式中, \mathbf{W}_{mi} 是神经网络的权重, \mathbf{q}_i 为初始的项目向量。

在用户聚合模型中使用了两个图注意力模型来挖掘不同图上的信息。第一个图注意力模型的目的是挖掘子图的嵌入表示。第二个注意力模块用来融合社交子图和用户 - 项目子图从而获得用户一个通道的嵌入表示。由于用户 - 项目子图可能信息稀疏,融合社交关系可以极大缓解这一问题。

基于图注意力的历史用户 - 项目交互信息聚合: 关注用户 u 和他的用户 - 项目交互子图 \mathbf{B}_m^u , 然后使用图注意力模型来聚合用户历史交互项目的信息, 就可以得到用户 u 在第 m 个通道下的项目空间用户嵌入表示 $\mathbf{h}_{1,m}^u$ 。

$$\mathbf{h}_{1,m}^u = \sigma \left[\mathbf{W} \cdot \left(\sum_{i \in \mathbf{B}_m^u} \alpha_m^{ui} \cdot \mathbf{y}_m^i \right) \right] \quad (5)$$

式中, σ 表示非线性激活函数(即 LeakyReLU 函数), \mathbf{W} 是神经网络的权重, α_m^{ui} 表示用户 u 和项目 i 的注意力权重, \mathbf{B}_m^u 是用户 u 的第 m 个聚类子图。

定义了注意力网络来获得 α_m^{ui} , 它是个两层神经网络。注意力网络的输入是目标用户 u 的向量表示 \mathbf{p}_u 和历史交互项目 i 的向量表示 \mathbf{y}_m^i 。最后通过 Softmax 函数对上述注意力得分进行标准化, 如式(6)所示。

$$\alpha_m^{ui} = \frac{\exp \{ \mathbf{W}_2 \cdot \sigma [\mathbf{W}_1 \cdot (\mathbf{p}_u \parallel \mathbf{y}_m^i)] \}}{\sum_{i \in \mathbf{B}_m^u} \exp \{ \mathbf{W}_2 \cdot \sigma [\mathbf{W}_1 \cdot (\mathbf{p}_u \parallel \mathbf{y}_m^i)] \}} \quad (6)$$

式中, \parallel 表示级联。

基于图注意力的社交信息聚合: 关注用户 u 和他的社交图 $\mathbf{S}_U(u)$, 然后使用图注意力模型来聚合社交图上的信息, 得到用户 u 在第 m 个通道下的社交空间用户嵌入表示:

$$\mathbf{h}_{2,m}^u = \sigma \left[\mathbf{W} \cdot \left(\sum_{v \in \mathbf{S}_U(u)} \alpha_m^{uv} \cdot \mathbf{p}_v \right) \right] \quad (7)$$

式中, α_m^{uv} 表示用户 u 和用户 v 的注意力权重。网络的输入是目标用户 u 的向量表示 \mathbf{p}_u 和社交邻居 v 的向量表示 \mathbf{p}_v , 如式(8)所示。

$$\alpha_m^{uv} = \frac{\exp \{ \mathbf{W}_2 \cdot \sigma [\mathbf{W}_1 \cdot (\mathbf{p}_u \parallel \mathbf{p}_v)] \}}{\sum_{v \in \mathbf{S}_U(u)} \exp \{ \mathbf{W}_2 \cdot \sigma [\mathbf{W}_1 \cdot (\mathbf{p}_u \parallel \mathbf{p}_v)] \}} \quad (8)$$

基于注意力的多通道信息融合:由于用户社交图 and 用户 - 项目交互图从不同角度提供了用户的信息,所以同时考虑用户社交图和用户 - 项目交互图上的信息从而更好地学习用户潜在嵌入向量。通过注意力网络来融合项目空间用户嵌入表示 $\mathbf{h}_{1,m}^u$ 、社交空间用户嵌入表示 $\mathbf{h}_{2,m}^u$ 和用户本身的特征 $\mathbf{h}_{3,m}^u$ ($\mathbf{h}_{3,m}^u = \mathbf{p}_u$), 得到用户潜在嵌入表示:

$$\mathbf{h}_{userm}^u = \sum_{i=1}^3 \gamma_m^i \cdot \mathbf{h}_{i,m}^u \quad (9)$$

式中, γ_m^i 是重要性系数, 当 γ_m^i 越大时, 在第 m 个通道用户受到用户 - 项目交互图的影响更大。

$$\gamma_m^i = \frac{\exp\{\mathbf{W}_2 \cdot \sigma[\mathbf{W}_1 \cdot (\mathbf{p}_u \parallel \mathbf{h}_{i,k}^u)]\}}{\sum_{i=1}^3 \exp\{\mathbf{W}_2 \cdot \sigma[\mathbf{W}_1 \cdot (\mathbf{p}_u \parallel \mathbf{h}_{i,m}^u)]\}} \quad (10)$$

设用户 u 的 K_1 个通道表示为 $\{\mathbf{h}_{userm}^u\}_{m=1}^{K_1}$, 后将所有嵌入表示连接起来当作用户的最终嵌入表示。用户的最终嵌入表示为:

$$\mathbf{h} = \mathbf{h}_{user1}^u \parallel \mathbf{h}_{user2}^u \parallel \cdots \parallel \mathbf{h}_{userK_1}^u \quad (11)$$

在项目聚合模型中, 同样使用了两个图注意力模型, 其目的与用户聚合模块类似。第一个图注意力模型的目的是挖掘子图的嵌入表示。第二个注意力模块用来融合项目隐式图和项目 - 用户子图从而获得项目一个通道的嵌入表示。

基于图注意力的历史项目 - 用户交互信息聚合: 关注项目 i 和它的项目 - 用户交互子图 \mathbf{C}_n^i , 然后使用图注意力来聚合项目历史交互用户的信息, 得到项目 i 在第 n 个通道下的用户空间项目嵌入表示 $\mathbf{z}_{1,n}^i$ 。

$$\mathbf{z}_{1,n}^i = \sigma[\mathbf{W} \cdot (\sum_{u \in \mathbf{C}_n^i} \beta_n^{iu} \cdot \mathbf{x}_n^u)] \quad (12)$$

式中, \mathbf{C}_n^i 是用户 i 的第 n 个聚类子图, β_n^{iu} 表示项目 i 对用户 u 的注意力权重。注意力网络的输入是目标项目 i 的向量表示 \mathbf{q}_i 和历史交互用户 u 的向量表示 \mathbf{x}_n^u , 如式(13)所示。

$$\beta_n^{iu} = \frac{\exp\{\mathbf{W}_2 \cdot \sigma[\mathbf{W}_1 \cdot (\mathbf{q}_i \parallel \mathbf{x}_n^u)]\}}{\sum_{u \in \mathbf{C}_n^i} \exp\{\mathbf{W}_2 \cdot \sigma[\mathbf{W}_1 \cdot (\mathbf{q}_i \parallel \mathbf{x}_n^u)]\}} \quad (13)$$

基于图注意力的项目隐式图信息聚合: 关注项目 i 和他的项目隐式图 $\mathbf{T}_l(i)$, 然后使用图注意力模型来聚合项目隐式图上邻居的信息, 就可以得到项目 i 在第 n 个通道下的项目隐式空间项目嵌入表示 $\mathbf{z}_{2,n}^i$ 。

$$\mathbf{z}_{2,n}^i = \sigma[\mathbf{W} \cdot (\sum_{j \in \mathbf{T}_l(i)} \beta_n^{ij} \cdot \mathbf{q}_j)] \quad (14)$$

式中, β_n^{ij} 表示项目 i 对项目 j 的注意力权重。注意力网络的输入是目标项目 i 的向量表示 \mathbf{q}_i 和项目隐式图上相邻节点的向量表示 \mathbf{q}_j , 如式(15)所示。

$$\beta_n^{ij} = \frac{\exp\{\mathbf{W}_2 \cdot \sigma[\mathbf{W}_1 \cdot (\mathbf{q}_i \parallel \mathbf{q}_j)]\}}{\sum_{j \in \mathbf{T}_l(i)} \exp\{\mathbf{W}_2 \cdot \sigma[\mathbf{W}_1 \cdot (\mathbf{q}_i \parallel \mathbf{q}_j)]\}} \quad (15)$$

基于注意力的多通道信息融合: 使用注意力网络来融合 $\mathbf{z}_{1,n}^i$ 、 $\mathbf{z}_{2,n}^i$ 和 $\mathbf{z}_{3,n}^i$ ($\mathbf{z}_{3,n}^i = \mathbf{q}_i$), 得到项目潜在嵌入表示 \mathbf{z}_{itemn}^i :

$$\mathbf{z}_{itemn}^i = \sum_{j=1}^3 \eta_n^j \cdot \mathbf{z}_{j,n}^i \quad (16)$$

式中, η_n^i 是重要性系数, 当 η_n^i 越大时, 表示在第 n 个通道项目受到项目 - 用户交互图的影响更大。

$$\eta_n^j = \frac{\exp\{\mathbf{W}_2 \cdot \sigma[\mathbf{W}_1 \cdot (\mathbf{q}_i \parallel \mathbf{z}_{j,n}^i)]\}}{\sum_{j=1}^3 \exp\{\mathbf{W}_2 \cdot \sigma[\mathbf{W}_1 \cdot (\mathbf{q}_i \parallel \mathbf{z}_{j,n}^i)]\}} \quad (17)$$

最后将所有通道的项目嵌入表示连接起来当作项目的最终嵌入表示。项目的最终嵌入表示为:

$$\mathbf{z} = \mathbf{z}_{item1}^i \parallel \mathbf{z}_{item2}^i \parallel \cdots \parallel \mathbf{z}_{itemK_2}^i \quad (18)$$

2.2.3 评分预测模块

将用户的最终嵌入表示 \mathbf{h} 和项目的最终嵌入表示 \mathbf{z} 输入多层神经网络进行评分预测。设

$$\mathbf{g}_1 = [\mathbf{h} \parallel \mathbf{z}] \quad (19)$$

则有:

$$\mathbf{g}_l = \sigma(\mathbf{W}_l \cdot \mathbf{g}_{l-1} + b_l) \quad (20)$$

$$\mathbf{r}'_{ui} = \mathbf{w}^T \cdot \mathbf{g}_l \quad (21)$$

其中, l 是隐藏层的层数 ($l = 2, 3, \dots$), \mathbf{W} 和 b 表示权重和偏置, \mathbf{r}'_{ui} 是用户 u 对项目 i 的预测评分。

2.2.4 模型训练

为了评估模型的效果, 需要指定目标函数来进行优化。损失函数衡量的是预测值与真实值之间的差异, 即

$$L_1 = \frac{1}{2|O|} \sum_{(u,i) \in O} (\mathbf{r}'_{ui} - \mathbf{r}_{ui})^2 \quad (22)$$

式中, O 是观察到的评分数据集合, $|O|$ 代表集合 O 的数量, \mathbf{r}_{ui} 是用户 u 在项目 i 上的真实评分。

深度聚类的损失包含两部分, 分别是用户的聚类损失和项目的聚类损失, 目标函数为:

$$L_2 = L_{ucfs} + L_{icfs} \quad (23)$$

在模型训练过程中, 采用交替训练法, 交替优化 L_1 和 L_2 。因此, 本方法是一种带有聚类损失和预测损失指导的深度聚类模型。由于聚类

过程的引入,可以对用户和项目进行分组聚类,用户可以同时聚合不同类的项目信息得到多个通道的用户信息,使得模型具有一定的解耦能力。

3 实验

3.1 实验设定

3.1.1 数据集

使用两个包含社交关系的推荐数据集 Ciao 和 Epinions 来测试不同算法的性能。Ciao (www.ciao.co.uk) 是一个项目评价网站, Ciao 数据集包含两部分信息:用户与项目的交互信息(评分 1 ~ 5 分),以及用户之间的直接信任关系。Epinions (www.epinions.com) 是一个包含信任关系的消费者评论网站,用户可以在网站上给项目打分。Epinions 数据集包含两部分信息:用户与项目的交互信息(评分 1 ~ 5 分),以及用户之间的直接信任关系。两个数据集的统计信息显示在表 1 中。

表 1 实验数据集统计信息

Tab.1 Statistical information on experimental data sets

| 数据集 | 用户数 | 项目数 | 交互数 | 社交数 |
|----------|--------|---------|---------|---------|
| Ciao | 7 375 | 105 114 | 284 086 | 111 781 |
| Epinions | 49 290 | 139 738 | 664 824 | 487 181 |

对于每个数据集,随机选择 80% 的评分数据作为训练集,10% 当作验证集,10% 当作测试集。将每个实验重复 10 次并取其平均值作为模型的最终结果参与评价。

3.1.2 评估指标

使用两个常用的评估指标:平均绝对误差 (mean absolute error, MAE) 和均方根误差 (root mean squared error, RMSE) 来衡量评分预测的准确性。

$$MAE = \frac{\sum_{(u,i) \in R_{\text{test}}} |r_{ui} - r'_{ui}|}{|R_{\text{test}}|} \quad (24)$$

$$RMSE = \sqrt{\frac{\sum_{(u,i) \in R_{\text{test}}} (r_{ui} - r'_{ui})^2}{|R_{\text{test}}|}} \quad (25)$$

其中, R_{test} 表示测试集, r_{ui} 和 r'_{ui} 分别是用户 u 对项目 i 的真实评分和预测评分。MAE 和 RMSE 值越小,推荐性能越好。

3.1.3 基线模型

为了衡量实验精度,将 MGAT-D 模型与其他

推荐模型进行比较。本文选择的具有代表性的基线模型如下:

1) PMF^[27]: 概率矩阵分解模型,仅仅使用用户与项目评分信息。该模型基于高斯分布对用户和项目的潜在因素进行建模。

2) SoReg^[8]: 一种社交正则化模型,将社交网络信息建模为正则化项来约束矩阵分解模型。

3) SocialMF^[9]: 一种经典的社会化推荐模型,该模型将基于社交网络的信任传播机制加入矩阵分解模型中。

4) DeepSoR^[10]: 一种基于深度神经网络的社交推荐模型,使用神经网络来捕获社会关系并将其集成到概率矩阵分解模型中进行评分预测。

5) MCCF^[16]: 一种基于图神经网络的推荐模型,该模型使用了图注意力神经网络从用户 - 项目交互图中提取用户和项目的多个潜在的购买动机。

6) DANSER^[14]: 一种基于对偶图注意力网络的社交推荐模型,它使用对偶图注意力神经网络来同时捕获用户和项目的社会同质性和社会影响力。

7) GraphRec^[13]: 一种基于图注意力网络的社交推荐模型,使用图注意力神经网络来捕获用户 - 项目图中的交互信息和观点信息。

3.1.4 参数设置

使用 Pytorch 作为实验环境。为了优化目标函数,采用 RMSprop 作为优化器。设置 $K_1 = K_2$ 范围为 $\{1, 2, 3, 4\}$, 嵌入向量维度 d 设为 $\{8, 16, 32, 64, 128\}$, 学习率 lr 设为 $\{0.005, 0.001, 0.0005, 0.0001\}$, 每个节点聚合的节点数量 F 设为 $\{15, 30, 45, 60\}$, 激活函数选用 LeakyReLU 函数,斜率设为 0.2。设置提前停止法策略,如果验证集上的 RMSE 连续 5 个轮次未下降,就停止模型的训练。

3.2 实验结果和对比分析

表 2 展示 MGAT-D 模型和其他基线模型在两个数据集上的推荐效果。通过对比实验结果,可以得到以下结论:

1) 同时使用社交矩阵和用户 - 项目评分矩阵的模型优于仅使用用户 - 项目评分矩阵的模型。基于矩阵分解的社交推荐模型 SoReg 和 SocialMF 性能优于仅使用用户 - 项目评分矩阵的概率矩阵分解模型 PMF。基于图神经网络的社交推荐方法 DANSER 和 GraphRec 性能优于基于图神经网络的协同过滤模型 MCCF。这表

明使用社交信息可以有效提高推荐系统的性能。

表2 实验结果对比

Tab. 2 Comparison of experimental results

| 模型 | Ciao | | Epinions | |
|----------|----------------|----------------|----------------|----------------|
| | MAE | RMSE | MAE | RMSE |
| PMF | 0.777 8 | 1.099 3 | 0.845 8 | 1.120 0 |
| SoReg | 0.775 8 | 1.074 7 | 0.841 1 | 1.093 9 |
| SocialMF | 0.764 9 | 1.092 2 | 0.833 4 | 1.110 9 |
| DeepSoR | 0.773 9 | 1.031 6 | 0.838 3 | 1.097 2 |
| MCCF | 0.774 7 | 1.005 6 | 0.834 4 | 1.076 9 |
| DANSER | 0.749 7 | 0.990 5 | 0.811 0 | 1.068 9 |
| GraphRec | 0.730 0 | 0.970 5 | 0.806 4 | 1.060 7 |
| MGAT-D | 0.711 2 | 0.948 6 | 0.781 7 | 1.038 7 |

2) 基于深度神经网络的社交推荐模型 DeepSoR 的性能优于基于矩阵分解的社交推荐模型。这表明推荐系统中神经网络模型具有强学习能力。

3) 基于图神经网络的模型 MCCF、DANSER 和 GraphRec 性能优于其他基线模型。这说明在推荐系统中图神经网络模型具有一定的优势,其原因为:图神经网络可以学习到节点的信息,并按拓扑结构进行传递。

4) MGAT-D 优于所有基准方法,与 GraphRec 方法相比,在 Ciao 和 Epinions 数据集上 MAE 分别降低了 2.58% 和 3.06%, RMSE 分别降低了 2.26% 和 2.07%。这说明通过学习多通道的用户和项目表示,可以提取到更加丰富的用户和项目的信息,从而提高推荐算法的精度,也使得模型具有一定的解耦能力。

3.3 模型分析

本节分析模型的组件和模型的超参数对模型性能的影响。

3.3.1 通道数目的影响

为了研究 MGAT-D 模型的解耦能力,将通道数量设为 $\{1, 2, 3, 4\}$ 。表 3 总结了实验结果,其中:MGAT-D-3 表示解耦为 3 个通道的模型;MGAT-D-1 表示没有解耦能力的基模型,其不使用深度聚类方法仅提取单通道的用户和项目特征表示。表 3 中 MGAT-D-2、MGAT-D-3 和 MGAT-D-4 效果都比 MGAT-D-1 要好,从而

可知提取多通道的用户和项目的特征表示可以有效提高推荐效果。表 3 中 MGAT-D-4 效果并不优于 MGAT-D-3,这是因为随着通道数目增加,每个子图上的节点数量随之减少,反而会降低推荐的效果。

表3 不同通道数目的效果

Tab. 3 Effect of different number of channels

| 模型 | Ciao | | Epinions | |
|----------|----------------|----------------|----------------|----------------|
| | MAE | RMSE | MAE | RMSE |
| MGAT-D-1 | 0.742 1 | 0.953 9 | 0.804 8 | 1.048 6 |
| MGAT-D-2 | 0.715 4 | 0.950 0 | 0.793 7 | 1.038 0 |
| MGAT-D-3 | 0.723 6 | 0.948 0 | 0.790 9 | 1.037 6 |
| MGAT-D-4 | 0.711 2 | 0.948 6 | 0.781 7 | 1.038 7 |

3.3.2 社交图和物品隐式图对实验的影响

下面,对比分析用户社交图和物品隐式图对模型性能的影响。表 4 总结了实验结果,其中 MGAT-D-S 表示去除用户社交图的模型,MGAT-D-T 表示去除物品隐式图的模型,MGAT-D-ST 表示去除用户社交图和物品隐式图的模型。

由表 4 可知:MGAT-D 性能优于 MGAT-D-S、MGAT-D-T 和 MGAT-D-ST,即使用用户社交图和物品隐式图可以提高推荐的效果。

表4 社交图和物品隐式图对实验的效果

Tab. 4 Effect of social graphs and item implicit graphs on experiments

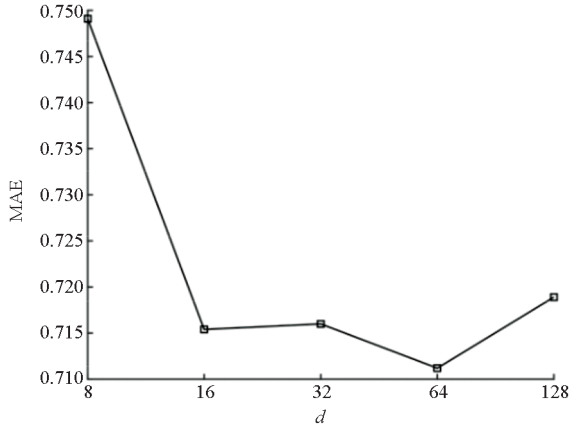
| 模型 | Ciao | | Epinions | |
|-----------|----------------|----------------|----------------|----------------|
| | MAE | RMSE | MAE | RMSE |
| MGAT-D | 0.711 2 | 0.948 6 | 0.781 7 | 1.038 7 |
| MGAT-D-S | 0.717 6 | 0.953 2 | 0.817 0 | 1.040 8 |
| MGAT-D-T | 0.728 6 | 0.960 3 | 0.799 3 | 1.049 7 |
| MGAT-D-ST | 0.727 2 | 0.967 9 | 0.808 4 | 1.052 8 |

3.3.3 其他超参数对实验的影响

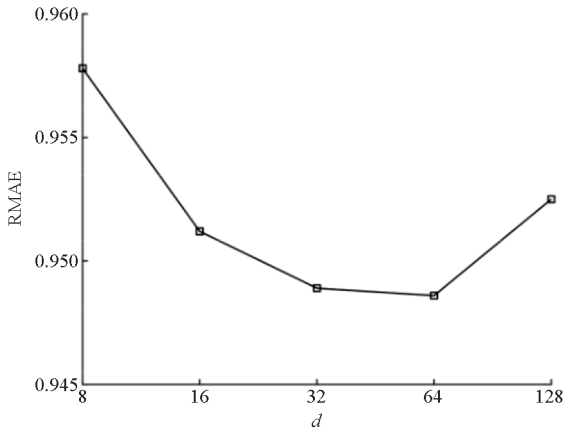
本小节对比分析模型中的一些超参数对模型性能的影响,包括学习率 lr 、向量维度 d 和每个节点聚合的节点数量 F ,它们在 Ciao 数据集上对模型的影响如图 2~4 所示。

由图 2~4 可见:学习率参数起着重要作用,如果学习率太高或者太低,模型都无法达到最优的效果,选择学习率 $lr = 1 \times 10^{-3}$ 来优化模型。嵌入向量维度 d 对模型起着重要的作用,需要适当

的嵌入向量维度。如果值太小,则模型缺乏表达性;而如果值太大,则嵌入向量会由于过于稀疏而使模型性能下降。每个节点聚合的节点数量 F 对模型起着重要的作用,随着每个节点聚合的节点数量 F 的增加,MAE 会有所下降但是 RMSE 反而上升,而且计算成本也随之增加。因此,选择每个节点聚合的节点数量 $F = 30$ 来平衡模型的复杂度和准确度。



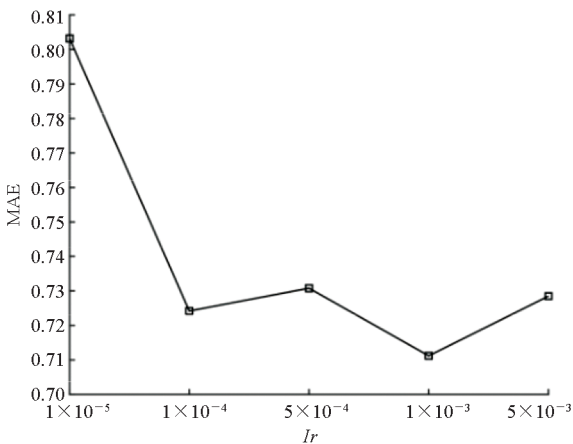
(a) MAE



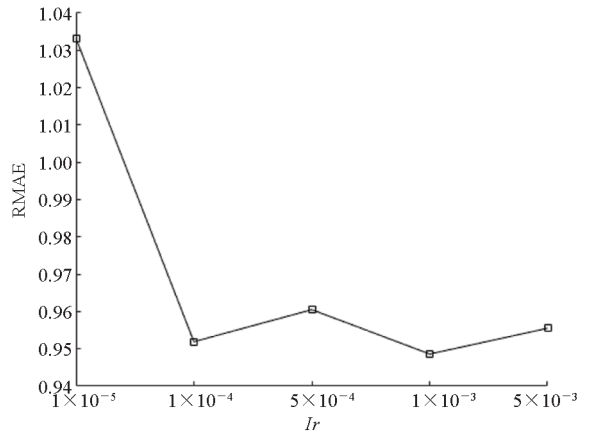
(b) RMAE

图 2 参数 d 在 Ciao 数据集上对模型的影响

Fig. 2 Influence of parameter d on Ciao dataset



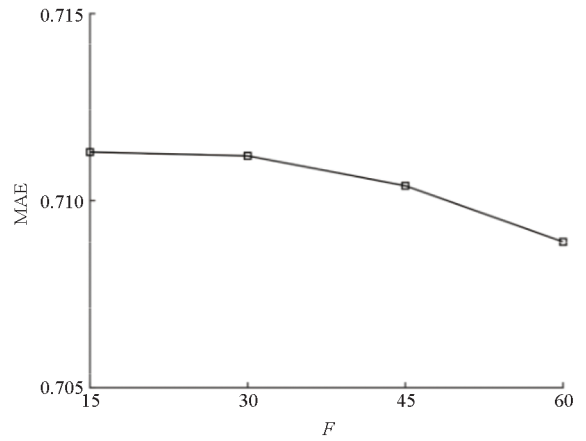
(a) MAE



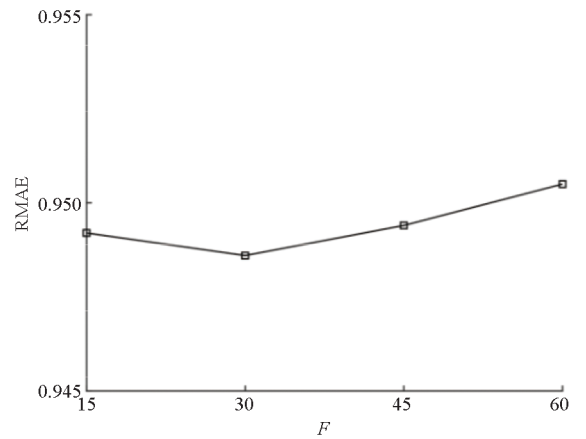
(b) RMAE

图 3 参数 lr 在 Ciao 数据集上对模型的影响

Fig. 3 Influence of parameter lr on Ciao dataset



(a) MAE



(b) RMAE

图 4 参数 F 在 Ciao 数据集上对模型的影响

Fig. 4 Influence of parameter F on Ciao dataset

4 结论

本文提出了一种具有解耦能力的多通道图注意力社交推荐模型 MGAT-D。该方法可以提取用户和项目多通道的信息来捕获更细粒度的用户偏好从而显著提高推荐的精度。具体地,使用深度

聚类对用户和项目进行分组,并利用聚类结果将用户社交图和用户-项目图拆分成多个用户社交子图及用户-项目子图,以学习用户兴趣分组及用户对不同类别项目的兴趣,从而区分用户不同的购买动机。此外,为了更好地提取和融合用户-项目交互图和社交图的信息,采取图注意力网络来提取用户社交图、用户-项目交互图、物品隐式图上的信息并且基于注意力机制来融合不同图上的信息,从而提取更丰富的用户和项目嵌入表示。MGAT-D 优于其他社交推荐算法,但是本文仅使用用户-项目评分信息和用户社交信息,未来将进一步将用户和项目的属性信息、用户和项目的知识图谱等信息进行融合,以进一步提高推荐算法的性能。

参考文献 (References)

- [1] ISINKAYE F O, FOLAJIMI Y O, OJOKOH B A. Recommendation systems: principles, methods and evaluation[J]. Egyptian Informatics Journal, 2015, 16(3): 261-273.
- [2] KHUSRO S, ALI Z, ULLAH I. Recommender systems: issues, challenges, and research opportunities [C]//Proceedings of Information Science and Applications (ICISA), 2016.
- [3] TANG J L, HU X, LIU H. Social recommendation: a review[J]. Social Network Analysis and Mining, 2013, 3: 1113-1133.
- [4] KING I, LYU M R, MA H. Introduction to social recommendation [C]//Proceedings of the 19th International Conference on World Wide Web, 2010: 1355-1356.
- [5] YANG B, LEI Y, LIU J M, et al. Social collaborative filtering by trust[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1633-1647.
- [6] MA H, YANG H X, LYU M R, et al. SoRec: social recommendation using probabilistic matrix factorization [C]//Proceedings of the 17th ACM Conference on Information and Knowledge Management, 2008: 931-940.
- [7] GU Y L, SONG J X, LIU W D, et al. CAMF: context aware matrix factorization for social recommendation [J]. Web Intelligence, 2018, 16(1): 53-71.
- [8] MA H, ZHOU D Y, LIU C, et al. Recommender systems with social regularization [C]//Proceedings of the 4th ACM International Conference on Web Search and Data Mining, 2011: 287-296.
- [9] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks [C]//Proceedings of the 4th ACM Conference on Recommender Systems, 2010: 135-142.
- [10] FAN W, LI Q, CHENG M. Deep modeling of social relations for recommendation [C]//Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), 2018, 32(1): 8075-8076.
- [11] FAN W Q, MA Y, YIN D W, et al. Deep social collaborative filtering [C]//Proceedings of the 13th ACM Conference on Recommender Systems, 2019: 305-313.
- [12] WANG X, HE X N, NIE L Q, et al. Item silk road: recommending items from information domains to social users [C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017: 185-194.
- [13] FAN W Q, MA Y, LI Q, et al. Graph neural networks for social recommendation [C]//Proceedings the World Wide Web Conference, 2019: 417-426.
- [14] WU Q T, ZHANG H R, GAO X F, et al. Dual graph attention networks for deep latent representation of multifaceted social effects in recommender systems [C]//Proceedings of the World Wide Web Conference, 2019: 2091-2102.
- [15] SONG W P, XIAO Z P, WANG Y F, et al. Session-based social recommendation via dynamic graph attention networks [C]//Proceedings of the 12th ACM International Conference on Web Search and Data Mining, 2019: 555-563.
- [16] WANG X, WANG R J, SHI C, et al. Multi-component graph convolutional collaborative filtering [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 6267-6274.
- [17] GUO G, ZHANG J, YORKE-SMITH N. TrustSVD: collaborative filtering with both the explicit and implicit influence of user trust and of item ratings [C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence, 2015, 29(1): 123-129.
- [18] GROVER A, LESKOVEC J. Node2vec: scalable feature learning for networks [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2016: 855-864.
- [19] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering [C]//Proceedings of the 30th Conference on Neural Information Processing Systems, 2016.
- [20] HAMILTON W, YING Z T, LESKOVEC J. Inductive representation learning on large graphs [C]//Proceedings of the 31st Conference on Neural Information Processing Systems, 2017.
- [21] WANG X, HE X N, 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] YING R, HE R N, CHEN K F, 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.
- [23] WU L, SUN P J, FU Y J, et al. A neural influence diffusion model for social recommendation [C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019: 235-244.
- [24] XIE J Y, GIRSHICK R, FARHADI A. Unsupervised deep embedding for clustering analysis [C]//Proceedings of the 33rd International Conference on Machine Learning, 2016.
- [25] GUO X F, GAO L, LIU X W, et al. Improved deep embedded clustering with local structure preservation [C]//Proceedings of International Joint Conference on Artificial Intelligence, 2017.
- [26] HARTIGAN J A, WONG M A. Algorithm AS 136: a K-means clustering algorithm [J]. Journal of the Royal Statistical Society: Series C (Applied Statistics), 1979, 28(1): 100-108.
- [27] SALAKHUTDINOV R, MNIH A. Probabilistic matrix factorization [C]//Proceedings of the 21st Annual Conference on Neural Information Processing Systems, 2007.