

基于多层次多视角的图注意力 Top-N 推荐方法



刘志鑫 张泽华 张杰

太原理工大学信息与计算机学院 太原 030024

(itsliuzhixin@163.com)

摘要 推荐系统是当前数据挖掘领域的研究热点,海量数据的涌现促使多源信息融合的推荐方法得到极大的关注。但是,现有的基于异质信息融合的推荐方法在进行特征表示时往往忽略了用户和项目之间的交互信息以及元路径之间的相互影响。因此,考虑到属性节点嵌入和结构元路径的不同视角,提出了一种多层次图注意力的网络推荐方法。该方法通过构建不同的元路径,将多源信息网络结构粒化为多个独立的粗粒度网络,然后基于图注意力机制结合局部节点属性嵌入,来分别学习用户和项目的潜在特征,最终给出融合后的细粒度网络推荐。在现实大规模数据集上进行横向和纵向评测,实验结果表明该方法能够有效地提升推荐性能。

关键词: 层次粒化; 多源信息融合; 图注意力网络; Top-N 推荐

中图法分类号 TP183

Top-N Recommendation Method for Graph Attention Based on Multi-level and Multi-view

LIU Zhi-xin, ZHANG Ze-hua and ZHANG Jie

School of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China

Abstract Recommendation system is a research hotspot in the field of data mining. Due to the emergence of massive data, the recommendation methods of multi-source information fusion receive great attention. However, the existing recommendation methods based on heterogeneous information fusion often ignore the interaction information between users and items, as well as the interaction between meta-paths in feature representation. Therefore, considering the influence of different perspectives of attribute node embedding and structural meta-paths, a network recommendation method with multi-level graph attention is proposed. This method granulates the multi-source information network structure into multiple independent coarse-grained networks by constructing different meta-paths. Then, based on graph attention mechanism and local node attribute embedding, this method can learn the potential features of users and items separately. Finally, it gives a fine-grained network recommendation after fusion. The horizontal and vertical evaluations are conducted on real large-scale data sets, and the experimental results show that this method can effectively improve the recommendation performance.

Keywords Hierarchical granulation, Multi-source information fusion, Graph attention network, Top-N recommendation

1 引言

随着在线信息量的不断增长,人们很容易被电影、书籍和音乐等大量信息实体所淹没,为了给用户推荐与其兴趣匹配的信息实体,迄今为止已经提出了许多推荐方法。传统的推荐方法通常只针对一种类型的数据(如用户-项目交互记录)来推断用户的喜好。然而,现实生活中的大多数实际系统是由大量相互作用、类型不同的组件构成,其信息的丰富程度不言而喻。最近,越来越多的研究人员开始将这些互连的多类型网络化数据建模为多源信息网络,并通过利用网络中丰富的对象和关系信息来设计结构分析方法,在这样的多源信息

网络中做推荐已经成为当前的一大趋势^[1-3]。在这种由多种类型的节点和边构成的多源信息网络中已经有许多先进的建模方法^[4-5]被提出。由于多源信息网络可以灵活地表征各种异质数据,因此在推荐系统中其常用于为丰富的数据建模^[1,3,6]。其中,元路径(meta-path)是定义在网络结构上链接两类对象的一条路径,它不仅刻画了对象之间的语义关系,而且能够抽取对象之间的特征信息^[7]。图1给出了一个电影推荐的例子,用于说明多源信息网络在推荐系统中的应用。大多数基于多源信息网络的推荐方法^[1,8]都是对元路径的相似度进行一些转换(如矩阵分解、网络嵌入等操作变换)来学习有效的潜在特征,从而增强用户和项目的原始特征,以此达到

到稿日期:2020-08-04 返修日期:2020-09-31 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61503273,61702356);教育部产学研合作协同育人项目;山西省回国留学人员科研资助项目

This work was supported by the National Natural Science Foundation of China(61503273,61702356), Industry-University Cooperation Education Program of the Ministry of Education and Shanxi Scholarship Council of China.

通信作者:张泽华(zehua_zhang@163.com)

提高推荐效果的目的,如图 1(b)所示。

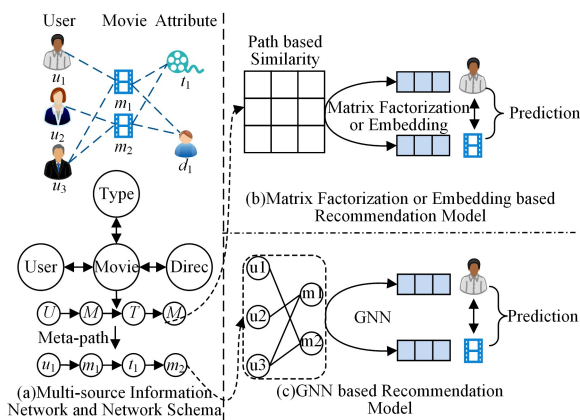


图 1 基于多源信息网络的推荐过程

Fig. 1 Recommendation process based on multi-source information network

尽管上述方法在一定程度上提高了推荐的性能,但是现有的基于多源信息网络的方法仍存在两个问题:

1) 对于元路径上得到的用户-项目交互矩阵,现有方法大都利用矩阵分解或网络嵌入来获取用户和项目的潜在特征。但矩阵分解或网络嵌入无法捕获用户和项目的交互信息,从而忽略了用户和项目交互的重要性。

2) 大多数现有的方法是不加区别地对潜在特征加以推荐,但是并非所有的潜在特征都是有用的,有些无用的特征可能会产生噪声从而干扰推荐。

为了解决上述问题,本文采用图 1(c)所示的图神经网络方法对元路径进行处理,结合粒计算理论(Granular Computing),从层次粒化的角度^[9]出发,提出了基于多层次多视角的图注意力 Top-N 推荐方法(MMGRec)。首先通过构造元路径的方式粒化多源信息网络空间,利用图注意力网络学习最粗粒度用户和项目的潜在特征,通过粗粒度的特征融合得到细粒度的特征表示,最终对所得到的细粒度用户和项目的潜在特征通过一系列的特征交互进行 Top-N 预。

本文的主要工作如下:

1) 提出了一种基于注意力机制的图神经网络嵌入方法,学习最粗粒度下用户和项目的潜在特征。

2) 设计了一种基于注意力机制的特征融合方式,采用局部注意力机制对粗粒度用户和项目的潜在特征附加个性化权重,从而得到细粒度特征表示。

3) 将提出的考虑属性节点级和结构元路径的不同算法与其他常用的推荐方法进行横向和纵向的评测比较,在两个现实世界的数据集上进行实验,证明了 MMGRec 的性能优于几种先进的推荐方法。

2 相关工作

粒计算(Granular Computing)是当前人工智能领域的一种新的概念和计算范式,其强调对现实世界问题进行多层次、多视角的理解和描述,从而得到对问题的粒结构表示。Zadeh^[10]首先提出并讨论了模糊集与信息粒化。Miao 等^[9]将粒计算视为一种粒化的思维方式及方法论,并采用多层次与

多视角的思想来解决问题。Hu 等^[11]提出了邻域信息系统和邻域决策表模型,并基于此模型构造了数值型属性的选择算法。Qian 等^[12]为解决大数据有限的标记属性以及属性约简中的过拟合问题,给出了具有线性时间复杂度的上下文近似和属性约简算法。Zhao 等^[13]则考虑社区漂移问题,提出了一种结合网络粗糙粒化的网络嵌入社区发现方法。

由于在多源信息网络中元路径可视为定义在网络模式上的连接两类对象的一类路径,它不仅刻画了对象之间的语义关系,而且能够抽取对象之间的特征信息。Shi 等^[14]利用基于元路径的相似性和双重正则化的方法提出了 SimMF。但由于元路径结构的局限性无法利用某些复杂语义,Zhao 等^[6]提出了异质信息网络上基于元图的推荐融合框架 FMG; Dai 等^[15]提出了基于元图的注意力机制推荐方法 MGAR。另一方面,在异质信息网络中,网络嵌入能够有效地保留网络的结构,充分挖掘用户和项目的潜在特征。Shi 等^[1]针对异质信息网络的推荐提出了异质网络嵌入方法 HERec。考虑到元路径与用户-项目对之间的影响,Hu 等^[8]提出了基于丰富元路径的上下文协同注意力推荐模型 MCRec。

图神经网络^[16]作为一种基于深度学习的强大的图表示技术,在诸多数据挖掘领域的应用中已展现出卓越的性能。在推荐系统中,Berg 等^[17]从图上的链路预测角度考虑了构建推荐模型,提出了一种基于双向交互图上传递可区分消息的图自动编码器框架 GC-MC。Zheng 等^[18]利用图在频谱域中体现图连通性这一特性,提出了谱协同过滤模型 Spectral CF。

多源信息网络和图神经网络相结合,成为了融合结构信息和辅助信息的一种新思路^[19-20]。Fan 等^[21]提出了意图推荐的以元路径为导向的嵌入方法 MEIRec。对于不同类型且信息比较稀疏的数据来说,多源信息网络和图神经网络结合可以有效地挖掘信息并进行特征表示。

3 基本概念

定义 1(属性图) 给定属性图 $\mathcal{G} = \{V, \epsilon, \mathcal{T}, \mathcal{R}, \varphi\}$, 其中 V 表示节点的集合, ϵ 表示节点之间边的集合, \mathcal{T} 表示节点的类型, \mathcal{R} 表示边的类型, φ 为 $\epsilon \rightarrow \mathcal{R}$ 的映射函数。当 $|\mathcal{T}| + |\mathcal{R}| > 2$ 时,表明属性图 \mathcal{G} 的节点和边具有多重异质特性架构 $T_G = (\mathcal{T}, \mathcal{R})$ 。在多源信息网络中,网络架构被用来描述局部信息交互的图元结构,其中包含了节点对象的属性及其之间的交互关系。

图 1(a)列举了关于电影推荐的多源信息网络的例子和相关的网络架构,从中可以发现,这个网络包含了多种类型的对象(如用户、电影和导演)和对象之间的语义关系(用户和电影之间的观看关系、用户之间的朋友关系和电影与导演之间的指导关系)。

将图 \mathcal{G} 中的每个节点视为基本粒,衡量粒的“大小”的概念是粒度(粗细)。基本粒为最细粒度的粒,图的粒化过程就是将多个粒组合为一个更粗粒度的粒操作。可通过构建元路径的方式对基本粒进行结构粒化。

定义 2(元路径^[22]) 其可表示为 $\rho = \{\mathcal{A}_1 \xrightarrow{r_1} \mathcal{A}_2 \xrightarrow{r_2} \dots \xrightarrow{r_n} \mathcal{A}_{n+1}\}$ 。其中, $r \in \mathcal{R}$ 是描述 \mathcal{A}_i 和 \mathcal{A}_{i+1} 之间信息传播的复

合关系 $r_1 \circ r_2 \circ \dots \circ r_l$, \circ 是关系之间的组合运算符。

如图 1(a) 所示, 用户 u_1 和电影 m_2 可以被多条元路径连接, 如 $u_1 - m_1 - u_3 - m_2$ (UMUM) 和 $u_1 - m_1 - t_1 - m_2$ (UMTM)。对于 u_1 和 m_2 之间的交互, 不同的元路径传达不同的语义信息。元路径 UMUM 和 UMTM 可以分别获得粗粒度网络 $\mathcal{G}^{(UMUM)}$ 和 $\mathcal{G}^{(UMTM)}$ 。因此, 通过获取不同的粗粒度网络, 我们能够从中捕获不同的语义信息。

定义 3(结构粒化) 给定具有异质架构 $T_G = (\mathcal{T}, \mathcal{R})$ 的属性图 \mathcal{G} , 其结构粒化过程可描述为将基本粒以元路径 ρ 所表达的异质关系组合为粗粒度网络 $\mathcal{G}^{(\rho)}$ 。

定义 4(二部图) 一个用于推荐的用户-项目二部图定义为 $\mathcal{B} = \{U, I, \epsilon\}$, 其中 U 和 I 是用户和项目两个不相交的顶点集。如果存在一条边 $e \in \epsilon$ 连接一对节点 $\langle u, i \rangle$, 则 $u \in U$ 和

$i \in I$ 存在一定的交互关系。图 1(c) 为由粗粒度网络 $\mathcal{G}^{(UMTM)}$ 所得到的用户-项目交互二部图。

问题描述如下: 给定属性图 $\mathcal{G} = \{\mathcal{V}, \epsilon, \mathcal{T}, \mathcal{R}, \varphi\}$, 本文的目标是从节点的结构信息和属性信息方面学习节点的潜在特征, 从而通过特征交互得到用户节点对某一项目的评分。即通过学习函数 $f: (\mathcal{V}_u, \mathcal{V}_i) \rightarrow y_{ui} \in \mathbb{R}$, 对所有未与用户节点 \mathcal{V}_u 有交互记录的项目节点 \mathcal{V}_i 进行评分, 从而得到 $\{y_{u_1}, y_{u_2}, \dots, y_{u_n}\}$, 评分最高的前 N 个项目节点为最终的 Top-N 推荐。

4 MMGRec 模型

本文提出的基于多层次多视角的图注意力 Top-N 推荐方法的整体框架如图 2 所示, 该框架主要包含特征嵌入层、图注意力层和特征融合层 3 个部分。

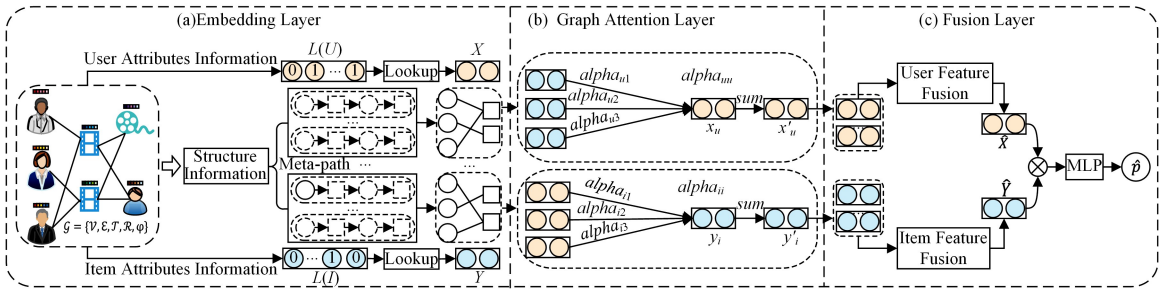


图 2 基于多层次多视角的图注意力 Top-N 推荐方法框架

Fig. 2 Framework of multi-view based attention graph method for top-N recommendation

4.1 特征嵌入层

如图 2(a) 所示, 多源信息网络所构成的属性图 $\mathcal{G} = \{\mathcal{V}, \epsilon, \mathcal{T}, \mathcal{R}, \varphi\}$ 被划分为属性信息和结构信息两部分, 属性信息又分为用户属性信息和项目属性信息。为了学习用户和项目在多源信息网络中的属性信息, 本文采用基于 Lookup 的嵌入方法^[21]。首先, $L^{(U)} \in \mathbb{R}^{|U| \times |U|}$ 和 $L^{(I)} \in \mathbb{R}^{|I| \times |I|}$ 分别表示用户和项目的 one-hot 嵌入。由于 one-hot 嵌入的结果是稀疏的, 因此采用 Lookup 层将用户和项目的 one-hot 嵌入转换为低维密集向量。 $\mathbf{X} = \{x\}^{|U|}$ 和 $\mathbf{Y} = \{y\}^{|I|}$ 分别表示用户和项目的最终嵌入结果。Lookup 层包含两个参数矩阵 $\mathbf{P} \in \mathbb{R}^{|U| \times d}$ 和 $\mathbf{Q} \in \mathbb{R}^{|I| \times d}$, 这两个参数矩阵分别对应用户和项目的潜在因子。这里的 d 是用户和项目嵌入维度的大小, $|U|$ 和 $|I|$ 分别表示用户的个数和项目的个数。Lookup 层的实现如下:

$$x_u = \mathbf{P}^T \cdot l_u \quad (1)$$

$$y_i = \mathbf{Q}^T \cdot l_i \quad (2)$$

4.2 图注意力层

图 2(a) 中, 对于多源信息网络所划分出的结构信息, 以构建元路径的方式将基本粒结构粒化为多个粗粒度网络, 每一个粗粒度网络 $\mathcal{G}^{(\rho)}$ 得到一个用户-项目交互二部图 $\mathcal{B}^{(\rho)}$ 。

传统的图神经网络是作用在同质网络上的, 而对于二部图, 用户和项目明显不是同一类型。为了学习每个粗粒度网络中用户和项目节点的潜在特征, 本文对 GAT^[23] 进行了改进, 提出了一种适用于二部图的图注意力网络层。引入图注意力层的目的是在充分学习粗粒度网络结构信息的基础上融入属性信息, 使其能够更加全面地学习用户和项目的潜在特征。

网络层的输入是用户和项目的节点特征, $\mathbf{h} = \{\mathbf{X}, \mathbf{Y}\}$, $\mathbf{X} =$

$\{x_1, x_2, \dots, x_m\}$, $x_i \in \mathbb{R}^d$, m 是用户节点的个数, $\mathbf{Y} = \{y_1, y_2, \dots, y_n\}$, $y_i \in \mathbb{R}^d$, n 是项目节点的个数, d 是每个节点的特征数量。网络层的输出是用户和项目的节点特征, $h' = \{X', Y'\}$, $X' = \{x'_1, x'_2, \dots, x'_m\}$, $x'_i \in \mathbb{R}^{d'}$, $Y' = \{y'_1, y'_2, \dots, y'_n\}$, $y'_i \in \mathbb{R}^{d'}$, d 和 d' 可能不同。为了处理二部图, 本文设计了一对类型转换矩阵 $\mathbf{T}_U \in \mathbb{R}^{d \times d}$ 和 $\mathbf{T}_I \in \mathbb{R}^{d \times d}$, \mathbf{T}_U 是将用户类型转换为项目类型, \mathbf{T}_I 是将项目类型转换为用户类型。然后运用自注意力(self-attention)实现节点间的注意力机制 $att: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ 。

图 2(b) 所示为节点间注意力机制的运作方式。以获取用户的潜在特征为例, 用户 u 的项目邻居被定义为 \mathcal{N}_u , 对于每个项目 $i \in \mathcal{N}_u$, 用户 u 和项目 i 的注意力系数为 e_{ui} 。为了使节点间的系数更容易比较, 采用 $softmax$ 函数使其标准化为注意力值 α_{ui} 。

$$e_{ui} = att(x_u, \mathcal{T}_I y_i) \quad (3)$$

$$\alpha_{ui} = \frac{\exp(e_{ui})}{\sum_{k \in \mathcal{N}_u} \exp(e_{uk})} \quad (4)$$

其中的注意力机制 att 是一个单层的前馈神经网络, 拥有一个注意力权重参数 $a \in \mathbb{R}^{2d}$, 最后采用 $LeakyReLU$ 非线性激活函数来增加模型的表达能力。

$$att(x_u, \mathbf{T}_I y_i) = LeakyReLU(a^T [x_u \parallel \mathbf{T}_I y_i]) \quad (5)$$

其中, \cdot^T 表示的是矩阵的转置, \parallel 是级联操作。为了避免过拟合, 我们对注意力值进行了 dropout 操作。最终, 用户的节点特征输出如下:

$$x'_u = \sigma \left(\sum_{i \in \mathcal{N}_u} \alpha_{ui} \mathbf{T}_I y_i \right) \quad (6)$$

其中, σ 为非线性激活函数, 本文实验采用的是 $ReLU$ 。因此,

最终项目的节点特征输出如下:

$$\mathbf{y}_i' = \sigma\left(\sum_{u \in \mathcal{N}_i} \alpha_{iu} \mathbf{T}_U \mathbf{x}_u\right) \quad (7)$$

4.3 特征融合层

对于不同元路径构成的粗粒度网络 $\mathcal{G}^{(p)}$,其用户和项目的潜在特征对推荐的贡献不同,若不对潜在特征的重要性进行区分,则可能会导致次优推荐。因此,本文提出了一种基于注意力机制的特征融合方式来区分不同潜在特征的重要性。图 2(c)所示为整个网络的特征融合层,分别对用户和项目在图注意力层所得到的粗粒度特征进行融合,从而得到用户和项目最细粒度的特征表示 $\hat{\mathbf{X}}$ 和 $\hat{\mathbf{Y}}$ 。

用户的粗粒度融合框架如图 3 所示。

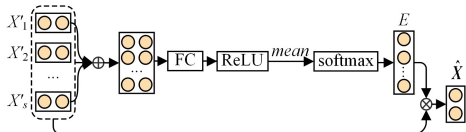


图 3 用户特征粗粒度融合

Fig. 3 Coarse-grained fusion of user features

以用户的特征融合为例,融合框架的输入为由 s 个粗粒度神经网络图注意力层得到的用户潜在特征 $\{\mathbf{X}_1', \mathbf{X}_2', \dots, \mathbf{X}_s'\}$, $\mathbf{X}_i' \in \mathbb{R}^{m \times d'}$, 输出为融合后最细粒度的用户潜在特征 $\hat{\mathbf{X}}$, 具体表示如下:

$$\hat{\mathbf{X}} = \sum_{i=1}^s \mathbf{E}_i \mathbf{X}_i' \quad (8)$$

其中, $\mathbf{E} = \{E_1, E_2, \dots, E_s\}$ 为用户注意力权重, s 为粗粒度网络个数。 \mathbf{E}_i 是粗粒度网络 $\mathcal{G}^{(p)}$ 所得到的用户潜在特征权重, 被一个注意力网络层学习, 由于注意力权重的范围不确定, 因此采用 *softmax* 函数对其进行标准化。

$$E_i = \frac{\exp(\text{mean}(\text{ReLU}(\mathbf{W}\mathbf{X}_i' + b)))}{\sum_{k=1}^s \exp(\text{mean}(\text{ReLU}(\mathbf{W}\mathbf{X}_k' + b)))} \quad (9)$$

同样, 项目的特征融合可以表示为:

$$\hat{\mathbf{Y}} = \sum_{i=1}^s \mathbf{F}_i \mathbf{Y}_i' \quad (10)$$

$$\mathbf{F}_i = \frac{\exp(\text{mean}(\text{ReLU}(\mathbf{W}\mathbf{Y}_i' + b)))}{\sum_{k=1}^s \exp(\text{mean}(\text{ReLU}(\mathbf{W}\mathbf{Y}_k' + b)))} \quad (11)$$

4.4 模型优化和预测

用户 u 和项目 i 间存在交互的概率为:

$$\hat{p}(e_{u,i} | \mathcal{G}) = \text{Sigmoid}(\mathbf{W}(\mathbf{x}_u \odot \mathbf{y}_i) + b) \quad (12)$$

其中, \mathbf{x}_u 和 \mathbf{y}_i 分别为用户和项目的最细粒度特征表示, \mathbf{x}_u 和 \mathbf{y}_i 之间的交互还可以用全连接神经网络来替换^[15], \odot 为向量按元素乘积, \mathbf{W} 和 b 分别为线性层的权重和偏置, *Sigmoid* 为非线性激活函数, $\hat{p}(e_{u,i} | \mathcal{G})$ 定义为多源信息网络 \mathcal{G} 中存在边 $e_{u,i}$ 的概率, 简写为 $\hat{p}_{u,i}$ 。如果某用户与一个项目存在交互, 则 $p_{u,i} = 1$, 否则 $p_{u,i} = 0$ 。为优化模型, 损失函数采用交叉熵函数。

$$\mathcal{L} = -\left(\sum_{(u,i) \in \mathcal{Z}} p_{u,i} \ln \hat{p}_{u,i} + (1 - p_{u,i}) \ln(1 - \hat{p}_{u,i})\right) + \lambda \omega(\theta) \quad (13)$$

其中, $\omega(\theta)$ 是正则化项, λ 是权重, $\mathcal{Z} = \mathcal{Z}^+ \cup \mathcal{Z}^-$, \mathcal{Z}^+ 为现有边的集合, 而 \mathcal{Z}^- 为用户和项目之间未观察到交互的集合。该损失函数需要用户和项目交互的正例 (u, i) 和负例 (u, j) , 负例从用户 u 尚未与之交互的项目中抽样, 抽样个数对实验的影

响会在实验部分进行分析。

5 实验设计与分析

在两个推荐系统的大规模真实数据集上进行实验, 并与通用的推荐方法进行横向评测, 与本文提出的节点级的算法 MMGRec_{nd}、路径级的算法 MMGRec_{MF} 以及两者最终的融合算法 MMGRec_{total} 进行纵向比较。

5.1 数据集设置

实验采用了两个不同领域的真实数据集, 即 Movielens-100K 电影数据集和 Yelp 商业数据集。对于这两个数据集, 将评分视为交互记录^[24] 来表示用户是否对项目进行了评分。表 1 列出了这两个数据集的详细内容, 每个数据集的第一行对应于用户、项目和交互的数量, 而其他行对应于其他关系的统计信息。在元路径的选择上, 基于每个数据集的属性信息和实验任务的要求, 选择了如表 2 所列的用于构建粗粒度网络的元路径。在元路径长度设置上, 只选择步长最多为 4 的短元路径, 因为长元路径可能会引入噪声^[8]。

表 1 Movielens-100k 和 Yelp 数据集的统计信息

Table 1 Statistics of Movielens-100k and Yelp datasets

Datasets	relations(A-B)	# A	# B	# A-B
Movielens (100k)	User-Movie	943	1682	100000
	User-User	943	943	47150
	Movie-Movie	1682	1682	82789
	Movie-Genre	1682	18	2861
Yelp	User-Business	16239	14284	198397
	User-User	16239	16239	158590
	Business-City(Ci)	14267	47	14267
	Business-Category(Ca)	14180	511	40009

表 2 每个数据集中所选择的元路径

Table 2 Selected meta-paths in each dataset

Datasets	Meta-paths
Movielens(100k)	UM, UMUM, UMGM
Yelp	UB, UBUB, UBCaB, UBCiB

5.2 评价标准

为了评估推荐性能, 实验将每个数据集的整个用户隐式反馈记录随机分为训练集和测试集, 80% 的反馈记录作为训练集, 20% 的反馈记录作为测试集。由于在评估过程中对每个用户的所有项目进行排名非常耗时, 因此对于测试集中的每个正项, 随机抽取 50 个与目标用户没有交互记录的负样本与之结合, 将组成的列表进行排名^[8]。在信息检索中, Top-N 推荐任务通常采用类似的评估指标, 本实验使用前 K 个精准率(Prec@K)、前 K 个召回率(Recall@K)和归一化折损累计增益(NDCG@K)作为评估指标。NDCG@K 通过为前 K 个排名较高的相关项目分配更高的分数来说明相关项目的位置, NDCG@K 的定义如下:

$$\text{NDCG}@K = \frac{1}{\text{IDCG}} \times \sum_{i=1}^K \frac{2^{r_i-1}}{\log_2(i+1)} \quad (14)$$

如果位置 i 的项目属于相关项目, 则 r_i 为 1, 否则为 0。IDCG 表示推荐结果的列表为完美排名, 如果所得到的排名列表为完美排名, 则 NDCG 值为 1。为了保证实验结果的有效性, 实验采用不同随机分割的训练集和测试集运行 10 次, 并记录平均结果。

5.3 实验设置与评测

本文实验使用 pytorch 的 python 库实现 MMGRec, 其中图

神经网络框架使用 torch-geometric。对于模型参数细节,实验使用高斯分布随机初始化模型参数,并使用自适应矩估计(Adam)^[25]优化模型,学习率 lr 设置为 0.01,正则化项的权重 λ 设置为 0.0001,用户和项目的嵌入维度 d 为 64。对于其他对比方法,实验采用 10% 的训练数据作为验证集来优化其参数。

实验考虑了两种代表性的推荐方法,即基于协同过滤仅利用隐式反馈的推荐方法(BPR, NeuMF) 和基于多源信息网络利用丰富异质信息的推荐方法(FMG_{rank}, MCRec)。另外,实验提供了 MMGRec_{total}, MMGRec_{nd} 和 MMGRec_{MF} 这 3 种模型来验证双注意力机制的有效性。评测方法的详细内容如下。

(1) BPR^[26]: 贝叶斯个性化排名模型最大程度地减少了隐式反馈的成对排名损失。

(2) NeuMF^[24]: 仅使用隐式反馈的用于 Top-N 推荐的神经网络方法,由矩阵分解部分和全连接层部分组成。

(3) FMG_{rank}^[6]: 用于评分预测的基于多源信息网络的推荐模型。对于 Top-N 推荐,本实验将其优化目标改为 BPR^[26] 中的成对排名损失。

(4) MCRec^[8]: 多源信息网络中用于 Top-N 推荐的具有协同注意力机制的深度网络模型。

(5) MMGRec_{nd}: 保留了节点间的注意力,消除了粗粒度网络间的注意力,并为每个网络 $G^{(p)}$ 分配了相同的重要性。

(6) MMGRec_{MF}: 保留了粗粒度网络间的注意力,消除了节点间的注意力,并为每个邻居分配了相同的重要性。

(7) MMGRec_{total}: MMGRec_{nd} 与 MMGRec_{MF} 两种方法的融合。

5.4 实验结果

表 3 列出了 MMGRec 和基准方法在两个数据集上的结果比较,图 4 则给出了不同层次注意力机制的性能比较。实验结果总结如下:

(1) MMGRec 在 3 个数据集上的精准率、召回率和归一化折损累计增益始终优于所有的基准方法。结果表明了 MMGRec 对 Top-N 推荐的有效性,其采用图神经网络的方法合理利用用户和项目的交互信息,提高了推荐性能。

(2) 鉴于 MMGRec 的两个变体,总体性能为 MMGRec_{total} > MMGRec_{nd} > MMGRec_{MF}。结果表明,双注意力机制在两个不同层次粒度大小的网络上学习节点特征能够有效提高推荐的性能。首先,基本粒邻居节点的重要性和粗粒度网络的重要性都应取决于特定的交互作用,而不应视为同等重要;其次,基本粒节点之间的交互比粗粒度网络之间的交互更加重要,基本粒节点之间的交互反映了更深层次的交互信息,对节点特征的生成更加重要。

(3) 在两种类型的基准方法中,基于多源信息网络的方法(FMG_{rank}, MCRec) 优于协同过滤的方法(BPR, NeuMF),表明了异质信息的有效性。NeuMF 利用多层感知机来建模用户和项目之间的复杂交互,这意味着深层神经网络通过捕获复杂交互关系来进行推荐的表现卓越。在多源信息网络的基准方法中, MCRec 的表现最好,说明引入协同注意力机制能够对不同的交互有特定的重要性区分,从而能够学习对推荐更有效的特征表示。

表 3 两个数据集上的实验结果

Table 3 Experimental results on two datasets

Datasets	Meterics	BPR	NeuMF	FMG _{rank}	MCRec	MMGRec _{MF}	MMGRec _{nd}	MMGRec _{total}
Movielens (100k)	Prec@10	0.3010	0.3293	0.3256	0.3451	0.3427	0.3636	0.3678
	Recall@10	0.1946	0.2090	0.2165	0.2256	0.2231	0.2337	0.2377
	NDCG@10	0.6459	0.6587	0.6682	0.6900	0.6530	0.6983	0.7075
Yelp	Prec@10	0.1474	0.1504	0.1538	0.1686	0.1681	0.1842	0.1891
	Recall@10	0.5504	0.5857	0.5951	0.6326	0.6015	0.6504	0.6647
	NDCG@10	0.5549	0.5713	0.5861	0.6303	0.5967	0.6455	0.6593

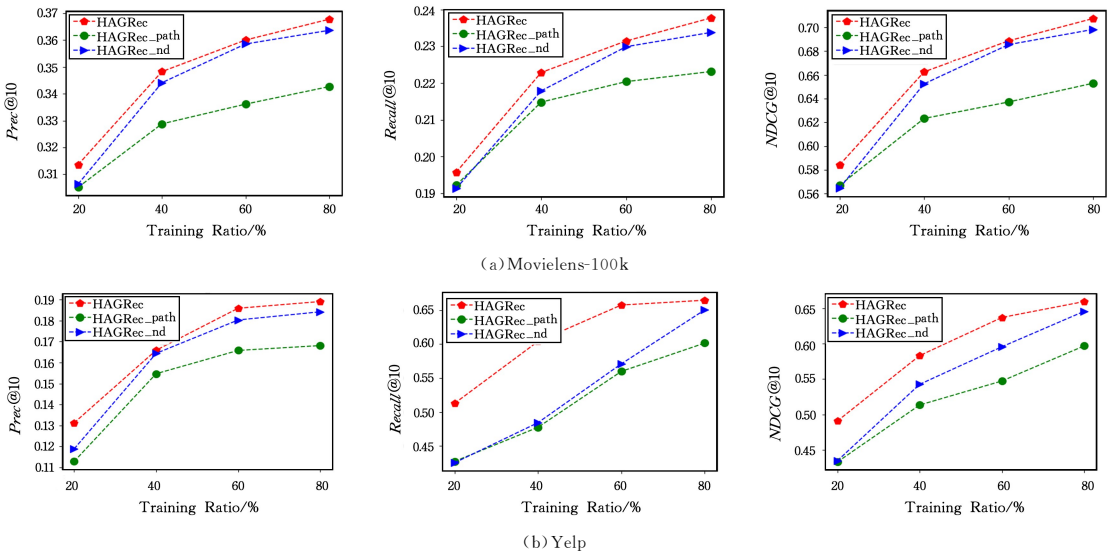


图 4 两个数据集上不同层次注意力机制的性能比较

Fig. 4 Performance comparison of different levels of attention mechanism on two datasets

5.5 模型分析

MMGRec 的一个显著特征是层次注意力机制。为了更好地理解不同层次粒度对推荐的重要性,本文对层次粒度的注意机制进行了详细分析。

5.5.1 基本粒节点间的注意力分析

对于每个粗粒度网络,MMGRec 能够学习用户及其项目邻居之间的注意力值,一些对交互有用的项目邻居往往具有更大的注意力值。如图 5(a)所示,以 Movielens-100k 数据集

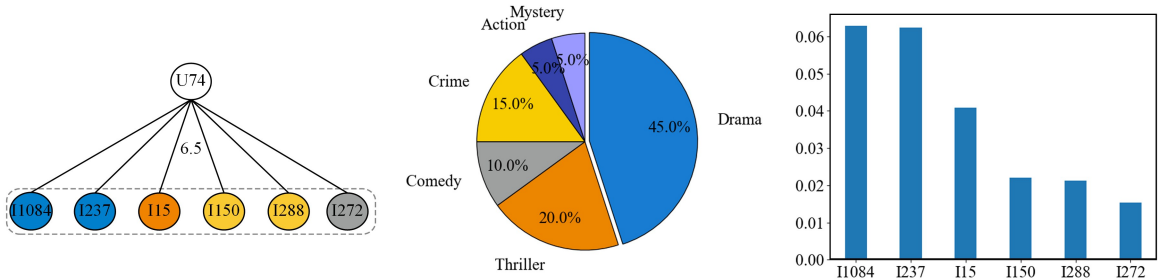


图 5 用户 U74 的喜好分析
Fig.5 Preferences analysis of user U74

5.5.2 粗粒度网络间的注意力分析

MMGRec 可以学习不同粗粒度网络对于推荐的重要性。为了验证粗粒度网络间的注意力对于推荐的重要性,以 Movielens-100k 和 Yelp 为例,图 6 给出了由各个元路径构成的粗粒度网络的推荐精准率和相应的注意力值。显然,单个粗粒度网络的性能与其注意力值之间存在正相关关系。可以看到,两个元路径 UM 和 UB 所构成的粗粒度网络的注意力值最高,意味着真实数据本身在学习各个节点的表示中起着最重要的作用。这是有道理的,因为通常挖掘真实数据所得到的信息被视为一种补充信息。

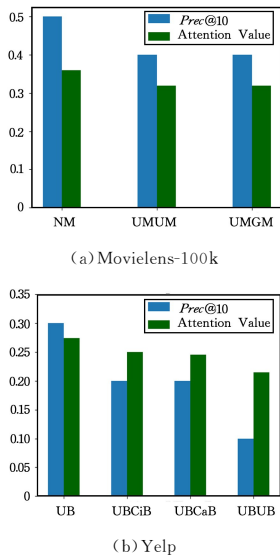


图 6 两个数据集上单个粗粒度网络的性能和相关注意力值
Fig.6 Performance and attention value of a single coarse-grained network on two datasets

5.5.3 重要参数对模型的影响

MMGRec 包含一些重要的参数,列举其中两个参数对推荐精准率的影响(以 Movielens-100k 数据集为例),例如基本

中序号为 74 的用户为例,给定一个粗粒度网络 $G^{(UMGM)}$,枚举了其中 6 个 U74 用户的项目邻居,其中 I1084 和 I237 是戏剧片,I115 是惊悚片,I1150 和 I288 是犯罪片,I272 是戏剧片,其注意力值如图 5(c)所示。可以看到,I1084 和 I237 在节点级注意力中获得了最高的注意力值,因为在图 5(b)中用户 U74 在电影类型的选择上最喜欢戏剧类型,其次是惊悚片。基于以上分析,节点级注意力机制可以分辨邻居之间的差异,并为某些有意义的邻居分配更高的权重。

粒间注意力系数的 dropout 参数和训练时负样本抽样的数量。对于 dropout 参数,我们在 $\{0, 0.3, 0.5, 0.8\}$ 范围内进行变化;对于负样本抽样的数量,在 $\{1, 3, 5, 7, 9\}$ 范围内进行变化。如图 7 所示,MMGRec 对这两个参数不太敏感,最终在 dropout 为 0.5 和负抽样本数量为 5 时获得了最佳性能。

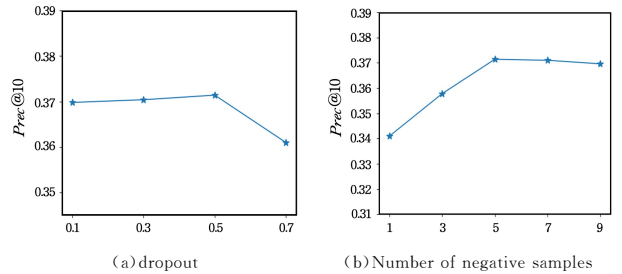


图 7 Movielens-100k 数据集上 MMGRec 的参数调整
Fig.7 Parameter tuning of MMGRec on Movielens-100k dataset

结束语 本文针对多源信息网络中的 Top-N 推荐提出了一种基于多层次多视角的图注意力 Top-N 推荐方法 MMGRec。通过节点类型转换,实现了在二部图上运用图神经网络,从而能够充分学习用户和项目之间的交互信息。为了获得高质量的特征表示,利用基本粒节点间和粗粒度网络间的注意力机制来有效避免无用信息的干扰。MMGRec 为用户和项目学习了有效的特征表示,以实现强大的交互能力。大量的实验结果证明了 MMGRec 在推荐效果方面的有效性。

此外,MMGRec 可以有效地学习用户和项目之间的交互信息,并能够学习不同层次粒度间的注意力权重。但考虑到元路径长度的局限性,我们考虑在下一步的工作中对推荐方法进行进一步的改进,以处理现实中元图、超网络等具有更复杂结构模式的多源信息网络挖掘问题。

参考文献

[1] SHI C, HU B, ZHAO W X, et al. Heterogeneous information network embedding for recommendation[J]. IEEE Transactions

- on Knowledge and Data Engineering, 2018, 31(2):357-370.
- [2] ZHU J, ZHANG J, ZHANG C, et al. CHRS: Cold start recommendation across multiple heterogeneous information networks [J]. IEEE Access, 2017, 5:15283-15299.
- [3] WANG X, HOI S C H, ESTER M, et al. Learning personalized preference of strong and weak ties for social recommendation [C] // Proceedings of the 26th International Conference on World Wide Web. 2017:1601-1610.
- [4] ZHANG J, LI T, JIANG Z, et al. A Novel Weighted Meta Graph Method for Classification in Heterogeneous Information Networks [J]. Applied Sciences, 2020, 10(5):1603.
- [5] CHEN Y, WANG C. HINE: Heterogeneous information network embedding [C] // Proceedings of the International Conference on Database Systems for Advanced Applications. Springer, Cham, 2017:180-195.
- [6] ZHAO H, YAO Q, LI J, et al. Meta-graph based recommendation fusion over heterogeneous information networks [C] // Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017:635-644.
- [7] SHI C, ZHANG Z, LUO P, et al. Semantic path based personalized recommendation on weighted heterogeneous information networks [C] // Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. 2015:453-462.
- [8] HU B, SHI C, ZHAO W X, et al. Leveraging meta-path based context for top-n recommendation with a neural co-attention model [C] // Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018:1531-1540.
- [9] 苗夺谦, 王国胤, 刘清, 等. 粒计算: 过去, 现在与展望 [M]. 北京: 科学出版社, 2007.
- [10] ZADEHL A. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic [J]. Fuzzy Sets and Systems, 1997, 90(2):111-127.
- [11] HU Q H, YU D R, XIE Z X. Numerical attribute reduction based on neighborhood granulation and rough approximation [J]. Journal of Software, 2008, 19(3):640-649.
- [12] QIAN Y, LIANG X, WANG Q, et al. Local rough set: a solution to rough data analysis in big data [J]. International Journal of Approximate Reasoning, 2018, 97:38-63.
- [13] ZHAO X, ZHANG Z H, ZHANG C W, et al. RGNE: A Network Embedding Method for Overlapping Community Detection Based on Rough Granulation [J]. Journal of Computer Research and Development, 2020, 57(6):1302-1311.
- [14] SHI C, LIU J, ZHUANG F, et al. Integrating heterogeneous information via flexible regularization framework for recommendation [J]. Knowledge and Information Systems, 2016, 49(3):835-859.
- [15] DAI F, GU X, LI B, et al. Meta-Graph Based Attention-Aware Recommendation over Heterogeneous Information Networks [C] // Proceedings of the International Conference on Computational Science. Springer, Cham, 2019:580-594.
- [16] ZHANG Z W, CUI P, ZHU W W. Deep learning on graphs: A survey [J]. arXiv:1812.04202v3, 2020.
- [17] BERG R, KIPF T N, WELLING M. Graph convolutional matrix completion [C] // Proceedings of the 24th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2018.
- [18] ZHENG L, LU C T, JIANG F, et al. Spectral collaborative filtering [C] // Proceedings of the 12th ACM Conference on Recommender Systems. 2018:311-319.
- [19] PENG H, LI J, GONG Q, et al. Fine-grained Event Categorization with Heterogeneous Graph Convolutional Networks [C] // Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. IJCAI, 2019:3238-3245.
- [20] WANG X, JI H, SHI C, et al. Heterogeneous Graph Attention Network [C] // Proceedings of the World Wide Web Conference. ACM, 2019:2022-2032.
- [21] FAN S, ZHU J, HAN X, et al. Metapath-guided heterogeneous graph neural network for intent recommendation [C] // Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019:2478-2486.
- [22] SUN Y, HAN J, YAN X, et al. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks [J]. Proceedings of the VLDB Endowment, 2011, 4(11):992-1003.
- [23] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph attention networks [C] // Proceedings of the 6th International Conference on Learning Representations. ICLR, 2018.
- [24] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering [C] // Proceedings of the 26th International Conference on World Wide Web. 2017:173-182.
- [25] KINGMA D P, BA J. Adam: a method for stochastic optimization [C] // Proceedings of the 3rd International Conference on Learning Representations. ICLR, 2015.
- [26] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. Bpr: Bayesian personalized ranking from implicit feedback [C] // Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. UAI, 2009:452-461.



LIU Zhi-xin, born in 1996, postgraduate. His main research interests include recommendation system and knowledge discovery on graph.



ZHANG Ze-hua, born in 1981, Ph.D., master supervisor, is a member of China Computer Federation. His main research interests include granular computing, uncertain reasoning and knowledge discovery on graph.