

互信息与多条元路径融合的异质网络表示学习方法*

贾霄生, 赵中英, 李超, 栾文静, 梁永全



(山东大学 计算机科学与工程学院, 山东 青岛 266590)

通信作者: 赵中英, Email: zzysuin@163.com

摘要: 异质信息网络能够对真实世界的诸多复杂应用场景进行建模, 其表示学习研究也得到了众多学者的广泛关注. 现有的异质网络表示学习方法大多基于元路径来捕获网络中的结构和语义信息, 已经在后续的网络分析任务中取得很好的效果. 然而, 此类方法忽略了元路径的内部节点信息和不同元路径实例的重要性; 仅能捕捉到节点的局部信息. 因此, 提出互信息与多条元路径融合的异质网络表示学习方法. 首先, 利用一种称为关系旋转编码的元路径内部编码方式, 基于相邻节点和元路径上下文节点捕获异质网络的结构和语义信息, 采用注意力机制来建模各元路径实例的重要性; 然后, 提出一种互信息最大化与多条元路径融合的无监督异质网络表示学习方法, 使用互信息捕获全局信息以及全局信息和局部信息之间的联系. 最后, 在两个真实数据集上进行实验, 并与当前主流的算法进行比较分析. 结果表明, 所提方法在节点分类和聚类任务上性能都有提升, 甚至和一些半监督算法相比也表现出强劲性能.

关键词: 异质网络表示学习; 元路径; 注意力机制; 互信息; 无监督学习

中图法分类号: TP18

中文引用格式: 贾霄生, 赵中英, 李超, 栾文静, 梁永全. 互信息与多条元路径融合的异质网络表示学习方法. 软件学报, 2023, 34(7): 3256–3271. <http://www.jos.org.cn/1000-9825/6535.htm>

英文引用格式: Jia XS, Zhao ZY, Li C, Luan WJ, Liang YQ. Heterogeneous Network Representation Learning Method Fusing Mutual Information and Multiple Meta-paths. Ruan Jian Xue Bao/Journal of Software, 2023, 34(7): 3256–3271 (in Chinese). <http://www.jos.org.cn/1000-9825/6535.htm>

Heterogeneous Network Representation Learning Method Fusing Mutual Information and Multiple Meta-paths

JIA Xiao-Sheng, ZHAO Zhong-Ying, LI Chao, LUAN Wen-Jing, LIANG Yong-Quan

(School of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China)

Abstract: Heterogeneous information networks can be used for modeling several applications in the real world. Their representation learning has received extensive attention from scholars. Most of the representation learning methods extract structural and semantic information based on meta-paths and their effectiveness in network analysis have been proved. However, these methods ignore the node internal information and different degrees of importance of meta-path instances. Besides, they can capture only the local node information. Thus, this study proposes a heterogeneous network representation learning method fusing mutual information and multiple meta-paths. First, a meta-path internal encoding method called relational rotation encoding is used, which captures the structural and semantic information of the heterogeneous information network according to adjacent nodes and meta-path context nodes. It uses an attention mechanism to model the importance of each meta-path instance. Then, an unsupervised heterogeneous network representation learning method fusing mutual information maximization and multiple meta-paths is proposed and mutual information can capture both global and local information. Finally, experiments are conducted on two real datasets. Compared with the current mainstream algorithms as well as some semi-supervised algorithms, the results show that the proposed method has better performance on node classification and clustering.

Key words: heterogeneous network representation learning; meta-path; attention mechanism; mutual information; unsupervised learning

* 基金项目: 国家自然科学基金 (62072288, 61702306)

收稿时间: 2021-06-05; 修改时间: 2021-09-05; 采用时间: 2021-11-16; jos 在线出版时间: 2022-03-24

CNKI 网络首发时间: 2022-11-15

真实世界的许多应用场景都可以用网络建模, 如属性网络^[1,2]、社交网络^[3,4]、知识图谱^[5,6]、生物网络^[7,8]等. 异质信息网络 (heterogeneous information network, HIN) 能够直观清楚地表示网络中多种类型节点以及节点之间的联系. 因此, 异质信息网络表示学习引起学术界与工业界的广泛关注, 并成功应用到节点分类^[9,10]、聚类^[11,12]、个性化推荐^[13,14]等任务中.

异质信息网络表示学习 (heterogeneous information network embedding, HINE) 是将网络中的节点映射为欧式空间的低维稠密向量, 同时最大限度地保持网络中的拓扑结构与异质信息. 主流的方法大多都是基于元路径的, 例如石川教授团队提出的 HERec^[15] 基于元路径的随机游走生成节点序列以学习节点的表示, 该方法本质上是将异质网络看成同质网络, 会造成不可逆的信息损失. Dong 等人^[9]提出的 Metapath2vec 将单条元路径引导的随机游走序列放到 skip-gram^[16]模型中学习节点表示, 但是单条元路径限制了模型处理异质网络中多种语义信息的能力. 也有部分学者提出了基于神经网络的方法, 如 Wang 等人^[11]提出的 HAN 使用多条元路径, 提出节点级别和语义级别的注意力机制来聚合节点邻居信息和语义信息, 但是该方法丢弃了元路径上的所有中间节点, 造成了信息损失. Wang 等人^[17]提出的 HHNE 基于元路径进行随机游走, 计算节点间的相似性, 并将其作为监督信息指导节点表示的学习; Wang 等人^[18]提出一种新的自引导随机游走方法, 将异质网络嵌入双曲空间, 并在链接预测和网络重构任务中取得很好效果. 以上方法仅考虑了局部信息, 并没有考虑到全局信息对节点表示的作用. SDNE^[19]使用自动编码器, 同时优化一阶和二阶相似度, 在保留局部和全局结构的同时对稀疏网络也具有鲁棒性, 但是没有考虑节点的特征信息. 以上模型都是有监督的方法, 模型训练需要有标签的数据, 但是真实世界的的数据通常没有足够多的标记数据, 而人工标记则需要的大量的人力物力, 因此, 有监督的方法在真实应用场景中无法推广使用.

一些无监督学习方法, 如 PGRA^[20]使用特定关系投影和关系特征注意力网络考虑节点特征以及连接关系和目标关系之间的兼容性来学习节点表示. DGI^[21]将 DIM^[22]的最大化互信息思想应用到图领域, 通过最大化局部特征和全局特征间的互信息来学习节点表示. HDGI^[23]进一步推广到 HINE 领域, 结合元路径的概念, 利用层次聚合的方法融合学习节点表示, 但是也存在着忽略元路径内部节点和各元路径实例重要性的问题. 现有的基于元路径的 HINE 算法对元路径实例的处理如图 1(a) 所示, 假设给定 $A_1-P_1-A_2$ 、 $A_3-P_2-T_1-P_3-A_4$ 两个元路径实例, 大多数算法只保留元路径实例的首尾两个节点 A_1-A_2 和 A_3-A_4 , 忽略了元路径中间节点的信息. 本文采用的方法如图 1(b) 所示, 使用关系旋转编码对元路径实例中的所有节点层层聚合, 最终得到整个元路径实例的表示 (具体操作可参照第 3.1.2 节中的关系旋转编码).

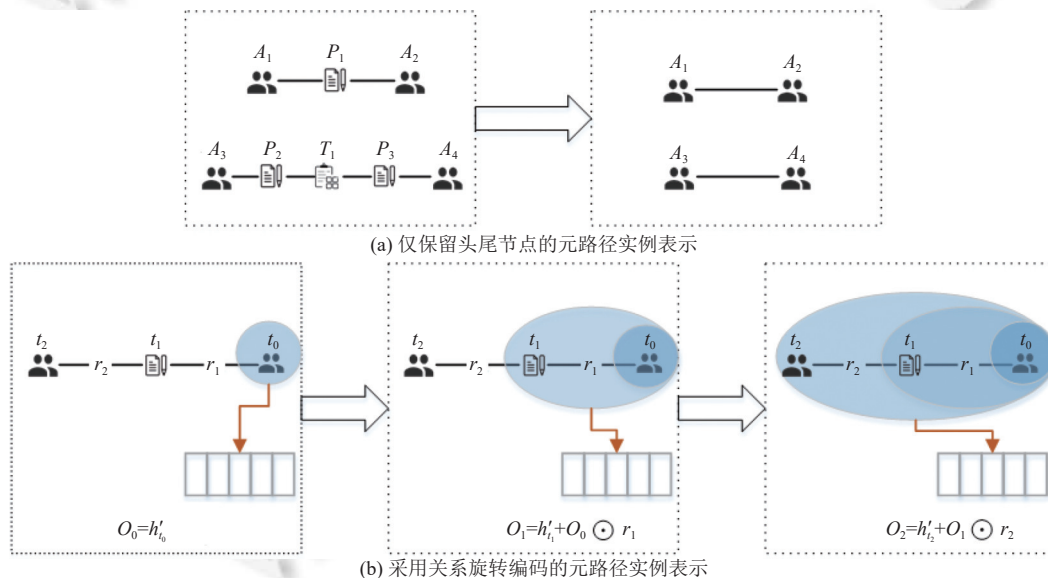


图 1 元路径不同处理方式的对比

综上所述, 现有的 HINE 仍然存在着以下亟待解决的问题: (1) 没有充分挖掘元路径内部节点的信息, 忽略了不同元路径实例具有不同重要性的特点; (2) 仅能捕捉到节点的局部信息. 为了解决上述问题, 本文提出一种新的互信息与多条元路径融合的异质网络表示学习方法, 主要贡献如下.

(1) 利用关系旋转编码, 基于相邻节点和元路径上下文节点建模元路径内部节点, 使用注意力机制建模各元路径实例的不同重要性, 以融合多种语义信息.

(2) 提出一种互信息最大化与多条元路径融合的异质信息网络节点表示学习方法, 通过最大化局部表示和全局表示之间的互信息, 捕获全局信息以及全局信息和局部信息之间的联系.

(3) 在真实数据集上进行实验并与主流的方法进行比较, 实验结果表明, 本文提出的方法具有很好的表征能力, 在分类和聚类任务上均优于主流的方法.

本文第 1 节介绍相关工作. 第 2 节给出相关概念的解释. 第 3 节介绍互信息与多条元路径融合的异质网络表示学习方法. 第 4 节介绍实验设置和结果分析. 第 5 节总结全文.

1 相关工作

1.1 基于元路径的异质网络表示学习方法

元路径未提出之前, 基于边的代表性模型如 DeepWalk^[24]、node2vec^[25]和 TADW^[26]等算法将随机游走得到的序列和 skip-gram^[16]模型结合起来学习异质网络的节点表示, 但是上述方法本质上将异质网络看成同质网络. 元路径这一概念被提出之后, 因其能更好地提取网络中的模式信息而被广泛应用于 HINE 领域. 代表算法如 Metapath2vec^[9]将由单条元路径引导的随机游走序列放到 skip-gram 模型中学习节点表示, 但是单条元路径不能完整地处理 HIN 中的多种语义关系. HIN2Vec^[27]同时考虑不同类型节点及其之间复杂多样的关系, 将网络表示转化为多分类问题以学习节点及元路径的表示; HERec^[15]将异质网络转换为基于元路径的同质网络并应用 DeepWalk 来学习节点表示. 然而将异质网络转换为同质网络, 不可避免地会造成信息损失; HEAD^[28]提出一种基于对抗解耦器的 HINE 框架, 使用元路径解耦器将每个元路径下的节点表示分别映射至固有和特有空间中, 设计对抗学习器保证固有表示和特有表示的充分解耦; mSHINE^[29]设计了一个新颖的基于元路径的 HINE 框架, 能够同时学习不同元路径的多个节点表示, 还提出一组用于选择初始元路径的标准, 降低了最佳元路径选择的成本; HGCN^[30]设计了对对象级聚合和类型级聚合的分层聚合架构, 在长度限制内自动评估所有可能的元路径, 并为每个目标对象发现和利用最有效的元路径, 提高了元路径的可解释性.

1.2 基于深度神经网络的异质网络表示学习方法

深度模型能够更好地捕捉非线性关系, 因此最近一些工作也开始利用深度模型来对异质网络进行建模. 基于深度神经网络的异质网络表示学习大致可以分为 4 类: 基于自动编码器^[19]、基于生成对抗网络^[31]、基于强化学习^[32]和基于图神经网络^[33]的方法.

基于自动编码器的方法, 如 SDNE 使用深度自动编码器优化一阶和二阶相似度, 在保留局部和全局结构的同时, 抽取网络结构的非线性特征; SHINE^[34]先对异质信息进行压缩编码得到特征表示, 再通过聚合函数进行聚合. 基于生成对抗网络的方法, GraphGAN^[31]通过广度优先搜索生成虚拟节点; HeGAN^[35]设计了具有关系感知能力的鉴别器和生成器, 通过学习节点的潜在分布反向改进负采样; 基于深度强化学习的模型将下游任务性能作为奖励来优化整体框架, 从而在规避元路径选择的同时学得节点表示, 如 MPDRL^[32]基于任务准确度发现长度不等的语义丰富元径, 并基于该元路径集合学习节点表示. 然而上述基于神经网络的方法都没有考虑节点复杂的特征属性信息.

基于图神经网络的异质网络表示学习方法的核心思想是消息传递机制, 将邻域信息聚合传递给目标节点, 在特定域进行卷积操作捕捉高阶非线性的依赖关系从而学习节点的表示; HAN^[11]提出一种新的涉及到节点级别和语义级别的异质信息网络分层注意力机制, 考虑不同的权重来聚合邻居和元路径信息. MAGNN^[36]采用元路径内部聚合和元路径间的聚合, 在建模过程中加入节点内容特征的信息; ConCH^[37]将半监督学习和自监督学习相结合,

使用元路径捕获对象之间的语义关系, 通过图卷积和注意力机制融合学习对象嵌入和上下文嵌入; HetSANN^[38]在不使用元路径的前提下, 使用 GNN^[39]和注意力机制^[40]聚合邻居间的各种关系信息; 基于图神经网络的 HINE 过于强调邻域节点, 忽略了全局信息的重要性, 无论是从结构还是从基于任务的偏好去学习节点的表示向量都是有局限性的。

1.3 基于互信息的异质网络表示学习方法

近年来, 由于互信息可以捕捉到变量间非线性的统计相关性, 基于互信息的网络表示学习方法引起学者的广泛关注. MINE^[23]利用神经网络的梯度下降法实现了高维连续随机变量之间互信息的估计, 通过反向传播进行训练, 提升了对抗生成网络的重建和推断能力; DGI 将 DIM 的最大化互信息 (infomax) 思想应用到图领域, HDGI 和 DMGI^[10]将此思想运用到 HIN 领域, 保留了 DGI 最大化互信息的同时, 结合异质网络中元路径的概念, 利用层次聚合的方法, 学习到不同元路径的权重, 进而通过融合得到最终的节点表示; EGLN^[41]通过互信息最大化提出了一种增强图学习网络的方法, 在没有任何特征输入的情况下, 利用增强图学习模块和节点嵌入模块迭代地学习节点表示, 并设计了一个局部全局一致性优化函数来捕获增强图学习过程中的全局属性。

2 基本定义

本文用到的主要符号及其含义如表 1 所示。

表 1 符号及相关说明

| 名称 | 描述 |
|-----------------|-----------------------|
| \mathbb{R}^n | n 维的欧式空间 |
| V | 图中的节点集合 |
| v | 节点 $v \in V$ |
| P | 元路径 |
| $P(v, u)$ | 节点 v 和 u 之间的元路径实例 |
| N_v | 节点 v 的邻居集合 |
| h_v^P | 节点 v 在元路径 P 下的表示 |
| M | 元路径数量 |
| \vec{s} | 节点的全局表示 |
| W | 权重矩阵 |
| α, β | 标准化注意力参数 |
| $\sigma(\cdot)$ | 激活函数 |
| \odot | 逐元素乘法 |
| \parallel | 向量连接符 |
| C | Shuffle()操作 |
| R | 平均编码函数 |
| D | 鉴别器 |

定义 1. 异质信息网络^[42]. 给定图 $G = (V, E)$, 其中 V 为节点集合, E 为边集合; 存在节点类型映射 $\varphi: V \rightarrow A$ 与边类型映射 $\psi: E \rightarrow R$, 其中 A 是节点的类型集合, R 是边的类型集合. 图中的每个节点 v 属于节点类型集合 A 中的一个特定节点类型, 表示为 $\varphi(v) \in A$; 每条边 $e \in E$ 属于边类型集合 R 中的一个特定边类型, 表示为 $\psi(e) \in R$. 如果在一个信息网络中节点类型的数量 $|A| > 1$ 或边类型数量 $|R| > 1$, 那么这样的信息网络称为异质信息网络. 如图 2 所示的异质信息网络, 存在 Author、Paper 和 Term 这 3 种不同类型的节点, 节点类型数量 $|A| > 1$.

定义 2. 网络模式^[42]. 网络模式记为 $T_G = (A, R)$, 是带有节点类型映射 $\varphi: V \rightarrow A$ 与边类型映射 $\psi: E \rightarrow R$ 的信息网络 $G = (V, E, \varphi, \psi)$ 的元模式, 是定义在节点类型集合 A 和边类型集合 R 上的有向图, 如图 3 为引文网络的网络模式。

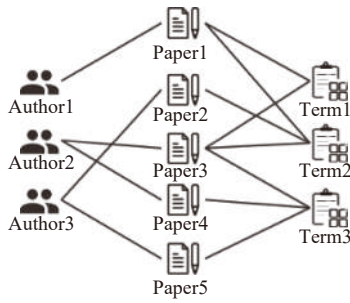


图 2 异质信息网络示例

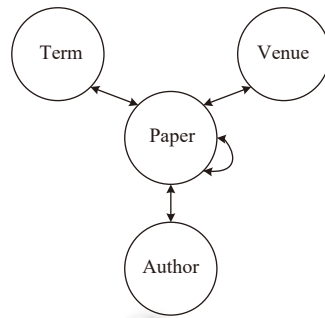


图 3 引文网络的网络模式

定义 3. 异质网络表示学习^[27]. 给定异质信息网络 $G = (V, E)$, 目标是为网络中的节点 $v \in V$ (或边、子图等) 学习一个映射关系 $f: v \rightarrow r_v \in \mathbb{R}^d, d \ll |V|$, 转换函数 f 用于捕获定义在原网络中的相似度信息, r_v 是为节点 v 学得低维稠密向量, 其中包含了节点在原网络中的异质信息.

3 互信息与多条元路径融合的异质网络表示学习方法

互信息与多条元路径融合的异质网络表示学习方法如图 4 所示, 其中包含了 4 个主要组件: (1) 基于多条元路径融合的局部表示; (2) 异质网络的全局表示编码; (3) 生成负样本; (4) 基于互信息最大化的鉴别器. 下面将对其进行详细介绍.

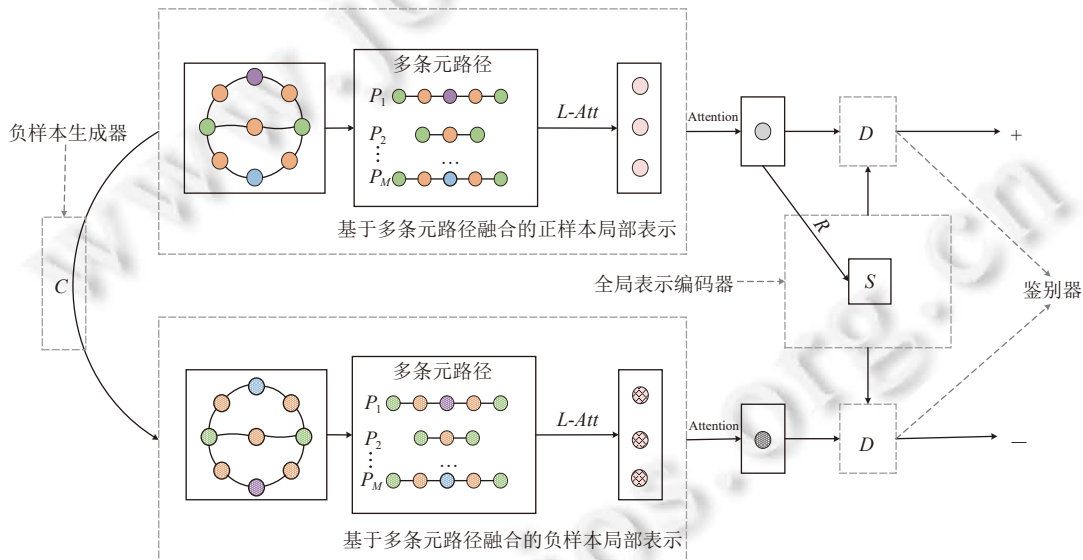


图 4 互信息与多条元路径融合的异质网络表示学习方法

3.1 基于多条元路径融合的局部表示

3.1.1 节点特征转换

HIN 中不同类型的节点有不同的属性, 因此不同的节点类型具有不同的特征向量维度, 属于不同的特征空间. 为了便于统一处理, 本文对特定类型的节点进行线性变换, 将特征向量投影到相同的特征空间中:

$$h'_v = W_A \cdot x_v^A \tag{1}$$

其中, $x_v^A \in \mathbb{R}^{d_A}$ 是原始特征向量, $h'_v \in \mathbb{R}^d$ 是节点 v 映射后的向量, $W_A \in \mathbb{R}^{d \times d_A}$ 是类型为 A 的节点参数化权重矩阵,

$A = \{A_1, A_2, \dots, A_M\}$ 是节点类型集合. 应用此节点特征转换操作, 所有节点的投影特征共享相同的维度, 便于完成元路径聚合过程.

3.1.2 元路径内部聚合

给定元路径 P , 基于目标节点、相邻节点和元路径上下文节点来建模元路径内部节点, 使用元路径内部聚合编码对 P 的每一个元路径实例 p 来学习节点的结构信息和语义信息.

$$h_{P(v,u)} = f_{\theta}(P(v,u)) = f_{\theta}(h'_v, h'_u, \{h'_t, \forall t \in \{m^{P(v,u)}\}\}) \quad (2)$$

其中, $h_{P(v,u)} \in \mathbb{R}^d$, $P(v,u)$ 表示单个实例, 中间节点定义为 $\{m^{P(v,u)}\} = P(v,u) \setminus \{u,v\}$, 两个节点之间可能有多个元路径实例. f_{θ} 表示元路径实例编码方式, 以下多种编码方式可供选择:

- 求平均值编码. 对元路径实例 $P(v,u)$ 的节点向量的元素取均值:

$$h_{P(v,u)} = MEAN(\{h'_t, \forall t \in P(v,u)\}) \quad (3)$$

- 线性编码. 对求平均值的编码器进行扩展, 添加一个线性变换:

$$h_{P(v,u)} = W_P \cdot MEAN(\{h'_t, \forall t \in P(v,u)\}) \quad (4)$$

• 关系旋转编码^[43]. 考虑节点表示在元路径顺序结构中的信息, 基于复杂空间中关系旋转对元路径实例进行编码, 具体操作如图 1 元路径处理方式对比所示, 其公式为:

$$\begin{cases} O_0 = h'_0 = h'_u \\ O_i = h'_i + O_{i-1} \odot r_i \\ h_{P(v,u)} = \frac{O_n}{n+1} \end{cases} \quad (5)$$

其中, $P(v,u) = (t_0, t_1, \dots, t_n)$, $t_0 = u$, $t_n = v$, 设定 r_i 为节点 t_{i-1} 和 t_i 之间的关系向量, r_i 和 h'_i 都是复数向量, \odot 表示元素内积, 关系旋转编码定义为在复数向量空间中从源节点到目标节点的旋转, 采用该方法可以建模元路径的结构信息. 本文实验中采用关系旋转编码方式对元路径实例进行编码.

对目标节点 v 在给定元路径 P 引导下的所有实例使用注意力机制加权求和. 因为不同的元路径实例对目标节点 v 的表示学习过程中权重不一样, 所以本文通过学习每个元路径实例的归一化权重 α_{vu}^p , 然后对所有实例进行加权求和来对此进行建模:

$$\begin{cases} e_{vu}^p = \text{LeakyReLU}(a_p^T \cdot [h'_v \parallel h_{P(v,u)}]) \\ \alpha_{vu}^p = \frac{\exp(e_{vu}^p)}{\sum_{s \in N_v^p} \exp(e_{vs}^p)} \\ h_v^p = \sigma \left(\sum_{u \in N_v^p} \alpha_{vu}^p \cdot h_{P(v,u)} \right) \end{cases} \quad (6)$$

其中, $\alpha_{vu}^p \in \mathbb{R}^{2d}$ 是元路径实例 $P(v,u)$ 的参数化注意力向量, \parallel 表示向量连接运算符. 对不同元路径实例的表示赋予不同的权重, 经过激活之后得到的 e_{vu}^p 表示元路径实例 $P(v,u)$ 对节点 v 的重要性, 使用 Softmax 函数对 e_{vu}^p 进行规范化获得所有节点的归一化权重 α_{vu}^p , 最后通过激活函数输出关于节点元路径实例表示的加权组合 h_v^p .

此外, 本文将注意力机制扩展为多头, 有助于提高学习过程中的稳定性, 减小网络的异质性带来的高方差, 具体如下:

$$h_v^p = \parallel_{k=1}^K \sigma \left(\sum_{u \in N_v^p} [\alpha_{vu}^p]_k \cdot h_{P(v,u)} \right) \quad (7)$$

其中, $[\alpha_{vu}^p]_k$ 表示元路径实例 $P(v,u)$ 对节点 v 在第 k 个注意力头处的归一化重要性. 给定映射后的特征向量 h'_v 以及

元路径集合 $P_A = \{P_1, P_2, \dots, P_M\}$ 和得到的特征向量, 单条元路径内部聚合就是为目标节点 v 生成 M 个针对特定元路径的向量表示, 记为 $\{h_v^{P_1}, h_v^{P_2}, \dots, h_v^{P_M}\}$, M 是类型为 A 的节点的元路径数量。

3.1.3 多条元路径间的聚合

每条元路径在 HIN 中的重要性是不一样的. 受 HAN 算法的启发, 本文设计了语义注意力层 $L-Att$, 为不同的元路径分配不同的权重:

$$\{\beta_{P_1}, \beta_{P_2}, \dots, \beta_{P_p}\} = L-Att(h_1^{P_A}, h_2^{P_A}, \dots, h_v^{P_A}) \quad (8)$$

具体地, 语义注意力层 $L-Att$ 包含以下步骤.

(1) 对所有节点转换后的特定元路径的节点向量进行取平均操作:

$$s_{P_i} = \frac{1}{|V_A|} \sum_{v \in V_A} \tanh(M_A \cdot h_v^{P_i} + b_A) \quad (9)$$

其中, $M_A \in \mathbb{R}^{d_m \times d}$ 和 $b_A \in \mathbb{R}^{d_m}$ 都是可学习的参数.

(2) 使用注意力机制融合节点 v 的特定元路径节点向量, 如下所示:

$$e_{P_i} = q_A^T \cdot s_{P_i} \quad (10)$$

其中, $q_A \in \mathbb{R}^{d_m}$ 是节点类型为 A 的参数化注意向量. 对 e_{P_i} 进行归一化, 得到的 β_{P_i} 就是元路径 P_i 对于 A 类型节点的重要性.

$$\beta_{P_i} = \frac{\exp(e_{P_i})}{\sum_{p \in P_A} \exp(e_p)} \quad (11)$$

(3) 使用 β_{P_i} 对节点 v 的所有针对特定元路径的向量进行加权求和:

$$h_v^{P_A} = \sum_{p \in P_A} \beta_p \cdot h_v^p \quad (12)$$

(4) 使用线性变换函数和一层非线性函数, 将节点表示映射到输出所需维度的向量空间中:

$$h_v = \sigma(W_o \cdot h_v^{P_A}) \quad (13)$$

其中, $W_o \in \mathbb{R}^{d_o \times d}$ 是权重矩阵, $\sigma(\cdot)$ 是激活函数, 此处映射是根据具体任务不同而有所变化的.

语义注意力层 $L-Att$ 的具体算法描述见算法 1.

算法 1. 语义注意力层 $L-Att$ 算法.

Input: A 类型节点的 $|V_A|$ 组潜在向量 $\{h_v^{P_1}, h_v^{P_2}, \dots, h_v^{P_M}\}$; 节点类型集合 $A = \{A_1, A_2, \dots, A_M\}$; 元路径集合 $P_A = \{P_1, P_2, \dots, P_M\}$;

Output: 融合不同元路径权重的节点表示 $h = \{h_v, \forall v \in V\}$.

for $P_A \in \{P_1, P_2, \dots, P_M\}$ **do**:

for $v \in V$ **do**:

 对元路径特定的节点向量取平均获得 s_{P_i} ; /*公式 (9)*/

 使用注意力机制融合节点向量获得 e_{P_i} ; /*公式 (10)*/

 归一化得到元路径 P_i 对于 A 类型节点的重要性权重 β_{P_i} ; /*公式 (11)*/

 重要性权重与特定元路径的向量进行加权求和得到 $h_v^{P_A}$; /*公式 (12)*/

 线性变换将节点表示映射到所需的向量空间得到 h_v ; /*公式 (13)*/

end

end

return $h = \{h_v, \forall v \in V\}$;

3.2 异质网络的全局表示编码

本文方法需要通过最大化局部表示和全局表示之间的互信息来进行模型的优化, 首先通过多条元路径间的聚

合得到节点 v 的局部表示 h_v . 然后采用最大池化与平均编码函数来计算 HIN 全局表示 \vec{s} .

(1) 最大池化编码函数: 将每个节点向量单独输入到一个全连接层. 使用元素级的 max-pooling 操作 (对应维度的所有元素取最大值), 汇总出节点集的信息:

$$\vec{s}_{\text{pool}} = \max(\sigma(W_{\text{pool}}h_v + b), v \in \{1, 2, \dots, N\}) \quad (14)$$

(2) 平均编码函数 R : 取节点表示的平均值作为图级的汇总表示:

$$\vec{s} = \sigma\left(\frac{1}{N} \sum_{i=1}^N h_v\right) \quad (15)$$

3.3 生成负样本

为了获得高质量的负样本, 本文采用 DGI 和 HDGI 中改变节点特征的方法来获取负样本:

$$\left(\tilde{X}, \{A^{P_1}, A^{P_2}, \dots, A^{P_M}\}\right) = C\left(X, \{A^{P_1}, A^{P_2}, \dots, A^{P_M}\}\right) \quad (16)$$

此方法保证所有基于元路径的邻接矩阵不变, 使得异质信息网络 G 的整体结构保持稳定, 使用 shuffle() 函数, 打乱初始节点特征矩阵 X 的每一行的顺序, 从而改变节点的特征. 根据谱论^[44], 整个图的结构没有变化, 只是每个节点对应的初始特征发生了变化.

3.4 基于互信息最大化的鉴别器

互信息可以捕捉到变量间非线性的统计相关性, 因此其能够度量真实的依赖性. 互信息表示当给定变量 Z 时, 观察者对变量 X 的不确定度的减少量, 具体计算公式如下:

$$I(X; Z) = H(X) - H(X|Z) = \int_{XZ} \log \frac{dP_{XZ}}{dP_X \otimes dP_Z} dP_{XZ} \quad (17)$$

其中, P_X 以及 P_Z 分别是两个变量的联合分布和它们各自的边缘分布.

由公式 (17) 可以得到互信息与 KL 散度的关系:

$$I(X; Z) = D_{KL}(P_{XZ}P_X \otimes P_Z) \quad (18)$$

由于互信息计算难度高的缺点, Belghazi 等人^[23]在文献 [23] 中利用互信息和 KL 散度之间的密切联系, 通过神经网络中的梯度下降来估计互信息, 使用 DV (Donsker-Varadhan)^[45]表示来逼近 KL 散度, 即:

$$D_{KL}(PQ) = \sup_{U: \Omega \rightarrow R} E_P[U] - \log(E_Q[e^U]) \quad (19)$$

其中, U 是一族函数, 定义域是 P 或 Q , 值域是 R , 可以看成是对于输入的打分. 即从一族函数中, 找出使等式右边最大的函数 U , 则相当于通过计算 P 和 Q 的 KL 散度来估计互信息, 如公式 (20):

$$I(X; Z) = D_{KL}(P_{XZ}P_X \otimes P_Z) = \sup_{U: \Omega \rightarrow R} E_{P_{XZ}}[U] - \log(E_{P_X \otimes P_Z}[e^U]) \quad (20)$$

然而, 函数族的搜索空间非常大, 公式 (20) 中的函数 U 很难被找出, 但是将 U 参数化为神经网络 U_θ , 可以找到互信息的较为紧致的下界: $E_{P_{XZ}}[U_\theta] - \log(E_{P_X \otimes P_Z}[e^{U_\theta}])$, 然后通过梯度更新不断抬高该下界, 就变相增大了 X 和 Z 间的互信息.

根据这一理论, 本文通过训练鉴别器 D 区分 (h_i, \vec{s}) 和 (\tilde{h}_i, \vec{s}) 来估计互信息, 样本 (h_i, \vec{s}) 表示为正样本集, (\tilde{h}_i, \vec{s}) 表示负样本集, 鉴别器 D 是一个二分类器:

$$D(h_i, \vec{s}) = \sigma(h_i^T W \vec{s}) \quad (21)$$

在本文中, 基于 Jensen-Shannon 散度和互信息间的单调关系, 可以采用针对鉴别器 D 的二元交叉熵损失最大化互信息:

$$\mathcal{L}(H, \tilde{H}, \vec{s}) = \frac{1}{2N} \left(\sum_{i=1}^N E_{(X)}[\log D(h_i, \vec{s})] + \sum_{j=1}^N E_{(\tilde{X})}[\log(1 - D(\tilde{h}_j, \vec{s}))] \right) \quad (22)$$

上述损失可以通过梯度下降进行优化.

综上所述, 我们给出互信息与多条元路径融合的异质网络表示学习方法的具体操作步骤, 见算法 2.

算法 2. 互信息与多条元路径融合的异质网络表示学习方法.

Input: 异质图 $G = (V, E)$; 节点类型 $A = \{A_1, A_2, \dots, A_M\}$; 元路径集合 $P_A = \{P_1, P_2, \dots, P_M\}$; 节点特征 $\{X_v, \forall v \in V\}$;

Output: 互信息与多条元路径的融合节点表示 $h = \{h_v, \forall v \in V\}$.

```

for  $v \in V$  do:
    对所有节点特征使用 shuffle() 函数生成负样本/*公式 (16)*/
end
for 节点类型  $A \in \{A_1, A_2, \dots, A_M\}$  do:
    特定类型节点映射  $h'_v \leftarrow W_A \cdot x_v^A$ ; /*公式 (1)*/
end
for  $A \in \{A_1, A_2, \dots, A_M\}$  do:
    for  $P_A \in \{P_1, P_2, \dots, P_M\}$  do:
        for  $v \in V$  do:
            对元路径实例进行编码得到  $h_v^P$ ; /*公式 (2), 公式 (5)–公式 (7) */
        end
        利用 L-Att 算法得到融合不同元路径权重的节点表示  $h$ ; /*公式 (13) */
        利用 readout 函数得到全局表示  $\vec{s}$ ; /*公式 (15)*/
    end
    使用基于互信息最大化的鉴别器优化节点表示; /*公式 (21), 公式 (22)*/
end
return 互信息与多条元路径融合的节点表示  $h = \{h_v, \forall v \in V\}$ ;

```

4 实验设置与结果分析

4.1 数据集

本文采用以下两种数据集进行实验, 数据集的主要描述如表 2 所示.

表 2 实验数据集描述

| 数据集 | 节点类型 | 节点数量 | 边类型 | 边数量 | 元路径 |
|------|--------------|--------|-------------------|----------------------------|-----------------------|
| DBLP | Author (A) | 4 057 | A-P P-T P-V | 19 645 85 810 14 328 | APA APTPA APVPA |
| | Paper (P) | 14 328 | | | |
| | Term (T) | 20 | | | |
| | Venue (V) | 20 | | | |
| IMDB | Movie (M) | 4 278 | M-D M-A | 4 278 12 828 | MDM MAM DMD |
| | Actor (A) | 5 257 | | | |
| | Director (D) | 2 081 | | | |

DBLP 是一个计算机科学文献网站数据集, 包含 4 057 位作者、14 328 篇论文、7 723 个术语和 20 个出版地点. 作者分为 4 个研究领域 (数据库、数据挖掘、人工智能和信息检索). DBLP 数据集的网络模式如图 5(a) 所示.

IMDB 是一个关于电影和电视节目的在线数据库, 包括演员、导演和电影等信息, 包含 4 278 部电影、2 081 名导演和 5 257 名演员. 经过数据预处理, 电影根据其类型信息被分为 3 类 (动作片、喜剧片和戏剧片). IMDB 数据集的网络模式如图 5(b) 所示.

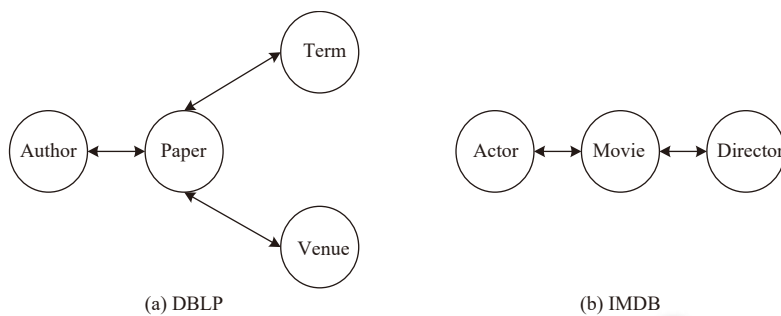


图5 数据集的网络模式

4.2 评价指标

实验采用了 *Micro-F1* 和 *Macro-F1* 作为分类任务的评价指标, 其定义如公式 (23) 所示. 评价指标中包含了准确率 (*precision, pre*) 和召回率 (*recall, rec*), 其定义如公式 (24) 所示:

$$\begin{cases} Micro-F1 = 2 \frac{rec_{mi} \times pre_{mi}}{rec_{mi} + pre_{mi}} \\ Macro-F1 = 2 \frac{rec_{ma} \times pre_{ma}}{rec_{ma} + pre_{ma}} \end{cases} \quad (23)$$

$$\begin{cases} rec_{mi} = \frac{\sum_{i \in T} TP_i}{\sum_{i \in T} (TP_i + FN_i)} \\ pre_{mi} = \frac{\sum_{i \in T} TP_i}{\sum_{i \in T} (TP_i + FP_i)} \end{cases}, \begin{cases} rec_{ma} = \frac{1}{|T|} \sum_{i \in T} \frac{TP_i}{TP_i + FN_i} \\ pre_{ma} = \frac{1}{|T|} \sum_{i \in T} \frac{TP_i}{TP_i + FP_i} \end{cases} \quad (24)$$

其中, T 表示类别的集合; TP 表示预测正确的正样例; FP 表示预测错误的正样例; FN 表示预测错误的负样例.

此外, 实验采用 *NMI* 和 *ARI* 作为聚类任务的评价指标, *NMI* 的定义如公式 (25) 所示:

$$NMI(U, V) = \frac{MI(U, V)}{F(H(U), H(V))} \quad (25)$$

其中, U 为真实标签向量, V 为预测标签向量, $F()$ 选取算术平均函数. $MI(U, V)$ 定义如公式 (26) 所示:

$$MI(U, V) = \sum_{i=1}^R \sum_{j=1}^C p_{i,j} \log \left(\frac{p_{i,j}}{p_i \times p_j} \right) \quad (26)$$

其中, $p_{i,j} = \frac{|U_i \cap V_j|}{N} = \frac{m_{ij}}{N}$, $p_i = \frac{|U_i|}{N}$, $p_j = \frac{|V_j|}{N}$.

ARI 的定义如公式 (27) 所示:

$$ARI = \frac{RI - E|RI|}{\max(RI) - E|RI|} \quad (27)$$

其中, $RI = \frac{a+b}{C_2^{n_{samples}}}$, a 表示被划分为同一簇的实例对数量, b 表示被划分为不同簇的实例对数量.

4.3 对比算法

我们将本文方法与目前的主流算法进行比较, 包括无监督和半监督的算法.

4.3.1 无监督算法

Raw Feature: 对节点原始特征组成的词袋 (*bag-of-words*) 学习节点的表示, 并直接在任务中测试.

DeepWalk^[24]: 通过随机游走采样节点的序列, 利用 *skip-gram* 结合负采样学习网络的节点表示.

DeepWalk+Raw Feature (DeepWalk+F): 将从 *DeepWalk* 中学到的节点表示和依据节点原始特征组成的词袋学

习的表示连接起来作为节点的最终表示。

Metapath2vec^[9]: 基于元路径的随机游走生成节点序列, 使用异质的 skip-gram 学习节点表示, 但只能处理特定的一条元路径, 所以本文对使用到的所有元路径都进行了测试, 并保留了结果最好的一个。

DGI^[21]: 针对同质网络提出的一种基于互信息的无监督网络表示学习方法, 通过最大化局部和全局之间的互信息学习节点的表示。

HDGI^[12]: 使用元路径建模异质网络结构中的语义信息, 使用图卷积模块和语义级别的注意力机制捕获节点的局部表示, 通过最大化局部和全局的互信息学习高阶的节点表示。

4.3.2 半监督算法

GCN^[46]: 是一种同质图神经网络, 在图形傅里叶域中执行卷积操作, 通过聚集来自邻居的信息学习节点表示。本文在基于元路径的同质网络中测试 GCN, 并保留最优效果。

GAT^[47]: 也是一种同质图神经网络, 在卷积的过程中应用注意力机制为邻居节点分配不同的权重来学习节点表示。本文在基于元路径的同质网络中测试 GAT, 并保留最优效果。

HAN^[11]: 一种基于层次注意力机制的异质图神经网络, 使用节点级别和语义级别的注意力机制来聚合邻居信息和语义信息, 进而学习节点的表示。

4.4 实验结果与分析

我们通过实验不断地调整参数, 将各参数设置如下: 丢弃率 (dropout_rate) 为 0.3、学习率 (learning_rate) 为 0.001、权重衰减 (weight_decay) 为 0.001、批量大小 (batch-size) 为 8, 使用早停法降低模型过拟合并设置耐心值为 30。对比算法中的模型参数则采用了原文中作者建议的参数设置。对于 GCN、GAT 和 HAN, 本文使用相同的训练集、验证集和测试集, 使用 Adam 优化器, 学习率和权重衰减设置为 0.001, 注意力头数设置为 8。

4.4.1 节点分类

在节点分类任务中, 本文使用 *Micro-F1* 和 *Macro-F1* 两个常见的节点分类评价指标对基线算法和本文提出的方法进行对比, 结果如表 3 所示。

表 3 中, A 表示模型只使用网络的拓扑信息, X 表示模型只使用网络的特征信息, X+A 表示模型使用网络的拓扑信息和特征信息。通过表 3 可以看出在节点分类任务上本文方法均优于基线算法, 这是因为: (1) DeepWalk、Metapath2vec 和 Raw Feature 在学习节点表示的过程中只使用了网络的部分信息, 不能完整地得到节点的表示; (2) DGI 和 HDGI 两个算法使用元路径和互信息建模异质信息网络考虑了全局结构信息, 但是忽略了元路径内部节点信息和不同元路径实例的重要性; (3) 半监督 GNNs 算法 GCN、GAT 和 HAN 过于强调邻域节点, 忽略了节点的全局结构信息, 这使其节点表示具有局限性。

表 3 节点分类结果 (%)

| 数据集 | 评价指标 | 训练比例 | 无监督算法 | | | | | | 半监督算法 (X+A) | | | |
|------|-----------------|------|----------|--------------|-------------|------------|-------|-------|--------------|-------|-------|-------|
| | | | A | | X | | X+A | | GCN | GAT | HAN | |
| | | | DeepWalk | Metapath2vec | Raw Feature | DeepWalk+F | DGI | HDGI | | | | Ours |
| DBLP | <i>Macro-F1</i> | 20 | 23.02 | 68.74 | 85.85 | 70.63 | 89.21 | 89.88 | 90.89 | 81.28 | 81.48 | 89.23 |
| | | 80 | 24.01 | 80.14 | 88.02 | 77.99 | 90.52 | 91.06 | 92.47 | 83.08 | 84.76 | 90.55 |
| | <i>Micro-F1</i> | 20 | 28.05 | 69.85 | 85.90 | 71.63 | 89.75 | 90.62 | 91.65 | 81.92 | 82.44 | 89.92 |
| | | 80 | 30.79 | 82.11 | 88.20 | 78.60 | 91.92 | 91.92 | 93.04 | 83.83 | 85.4 | 91.00 |
| IMDB | <i>Macro-F1</i> | 20 | 38.88 | 40.12 | 51.07 | 52.93 | 56.90 | 55.52 | 59.13 | 58.69 | 59.44 | 60.27 |
| | | 80 | 40.01 | 41.19 | 58.84 | 60.49 | 59.50 | 58.34 | 66.35 | 64.57 | 65.50 | 65.86 |
| | <i>Micro-F1</i> | 20 | 39.13 | 39.85 | 51.12 | 52.62 | 57.28 | 54.82 | 59.90 | 59.31 | 59.85 | 60.77 |
| | | 80 | 39.53 | 42.03 | 59.00 | 60.17 | 60.03 | 58.61 | 66.36 | 64.67 | 65.40 | 66.00 |

4.4.2 节点聚类

在节点聚类任务中, 本文将学习到的节点表示使用 K-means^[48]进行聚类, 聚类的簇的数量 K 设置为节点类别

的数量, 与 6 种无监督学习算法进行比较, 评价指标使用 *NMI* 和 *ARI*, 结果见表 4.

通过表 4 可以看出, 在节点聚类任务上本文方法均优于基线算法. 原因在于: (1) DeepWalk 和 Raw Feature 都是针对同质网络提出的算法, 不能很好地处理异质网络结构; (2) Metapath2vec 只能处理一条元路径, 不能同时处理不同的语义信息; (3) DGI 与 HDGI 都忽略了元路径内部节点的信息和不同元路径实例的重要性. 因此基于节点聚类的实验也验证了本文方法可以学习到节点的特征、语义和结构等有效信息, 具有很好的表征能力.

表 4 节点聚类任务结果

| 数据集 | 评价指标 | DeepWalk | Raw Feature | DeepWalk+F | Metapath2vec | DGI | HDGI | Ours |
|------|------------|----------|-------------|------------|--------------|-------|-------|--------------|
| DBLP | <i>NMI</i> | 7.40 | 11.21 | 11.98 | 34.30 | 59.23 | 60.76 | 73.98 |
| | <i>ARI</i> | 5.30 | 6.98 | 6.99 | 37.54 | 61.85 | 62.67 | 79.21 |
| IMDB | <i>NMI</i> | 1.23 | 1.06 | 1.23 | 1.15 | 0.56 | 0.8 | 1.48 |
| | <i>ARI</i> | 1.22 | 1.17 | 1.22 | 1.51 | 2.6 | 1.29 | 3.1 |

值得注意的是, 对于节点分类与聚类任务, IMDB 数据集的效果都比在 DBLP 上差, 这是因为: (1) 原始的 IMDB 数据集中每部电影都有多个标签, 实验中所使用的数据集只选择第一个标签作为该电影的类别标签; (2) IMDB 数据集中的数据特征比较稀疏. 因此相较于 DBLP 数据集, IMDB 数据集的实验效果在分类和聚类任务上相对较差.

4.5 参数分析

4.5.1 编码方式的影响

本节分析了 GRU、Bi-GRU、LSTM、Bi-LSTM、RotatE、linear、max-pooling、average 这 8 种编码方式在 DBLP 数据集下对分类结果的影响, 使用同样的实验参数设置, 得到了如图 6 所示的 *Macro-F1* 和 *Micro-F1* 的柱状图.

由图 6 可以看出, 在两个评价指标下, 使用 RotatE 编码方式的效果都是最好的. 原因在于: (1) GRU、Bi-GRU、LSTM、Bi-LSTM 的主要思想都是通过门控机制利用循环网络对顺序信息进行处理, 而本文方法生成的信息序列是一个无序的; (2) average 方法是对元路径内部节点信息使用求平均值编码, linear 方法是使用线性编码, max-pooling 方法使用最大池化编码, 这 3 种编码方式都会损失一部分信息. 而 RotatE 考虑了元路径内部的节点信息及其重要性不同的特点, 能够建模节点之间的潜在关系, 这对利用元路径来建模异质网络是非常重要的.

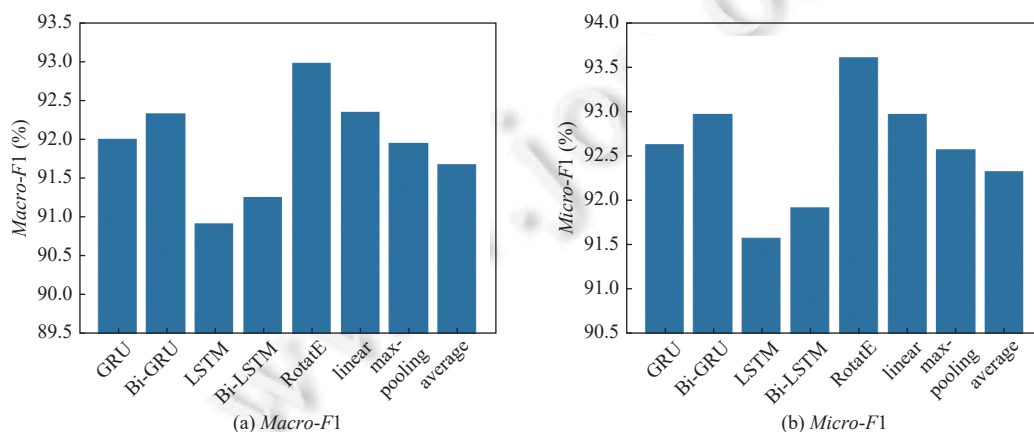


图 6 编码方式对实验结果的影响

4.5.2 batch-size 的影响

本节测试了 batch-size 的大小对实验结果的影响, 在不同规模的训练集下 $F1$ 值的变化结果如图 7 所示.

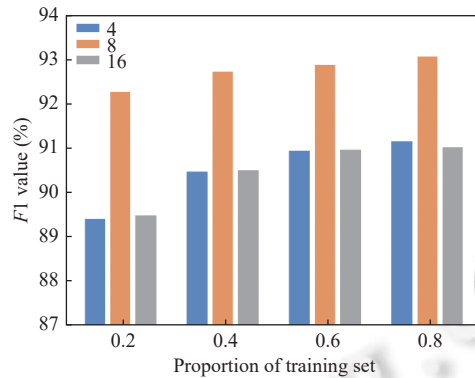


图 7 batch-size 对实验结果的影响

Batch-size 的大小影响训练速度和模型优化, 较大的 batch-size 能够得到更加精确的梯度估计, 但会导致内存溢出、收敛缓慢、泛化性差等问题; 较小的 batch-size 能够带来更好的泛化误差, 但会带来噪声, 甚至导致不收敛. 因此 batch-size 的选择也是一个重要问题. 从图 7 中可以看出, 在不同规模的训练集下, 模型在 batch-size 为 8 的时候效果最好.

4.5.3 学习率和丢失率的影响

本节测试了学习率 (learning_rate) 和丢失率 (dropout_rate) 对实验结果的影响, 在 80% 的训练集下, $Macro-F1$ 和 $Micro-F1$ 值的变化结果如图 8 所示.

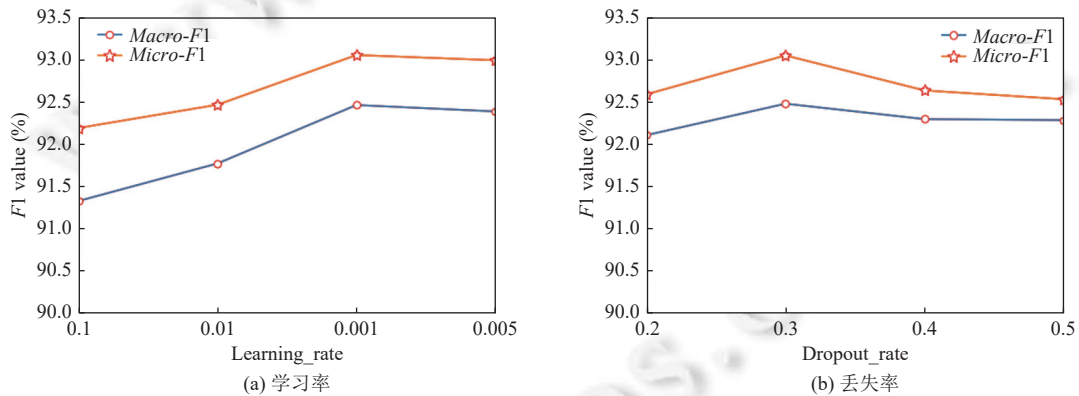


图 8 学习率和丢失率对实验结果的影响

在模型迭代过程中, 学习率会控制模型的学习进度, 决定着目标函数能否收敛到局部最小值. 当学习率设置的过小时, 收敛过程将变得十分缓慢, 而当学习率设置的过大时, 梯度可能会在最小值附近来回震荡, 甚至可能无法收敛. 由图 8(a) 可以看出, 模型在学习率为 0.001 时效果最好. 丢失率是为了防止模型过拟合, 增强模型的稳定性和鲁棒性, 由图 8(b) 可以看出, 模型在丢失率为 0.3 时效果最好.

4.5.4 聚类结果的可视化

为了进一步证明本文方法的有效性, 将本文提出的方法与 HDGI 进行聚类可视化对比实验, 结果如图 9 所示.

从图 9 展示的可视化效果来看, 本文方法的准确度要优于 HDGI. 总体而言, HDGI 虽然能识别出 4 个类簇, 但是无法准确地地区分出类簇边界. 本文提出的方法不仅可以正确识别 4 个类簇, 在相邻类簇上边界也较为明显.

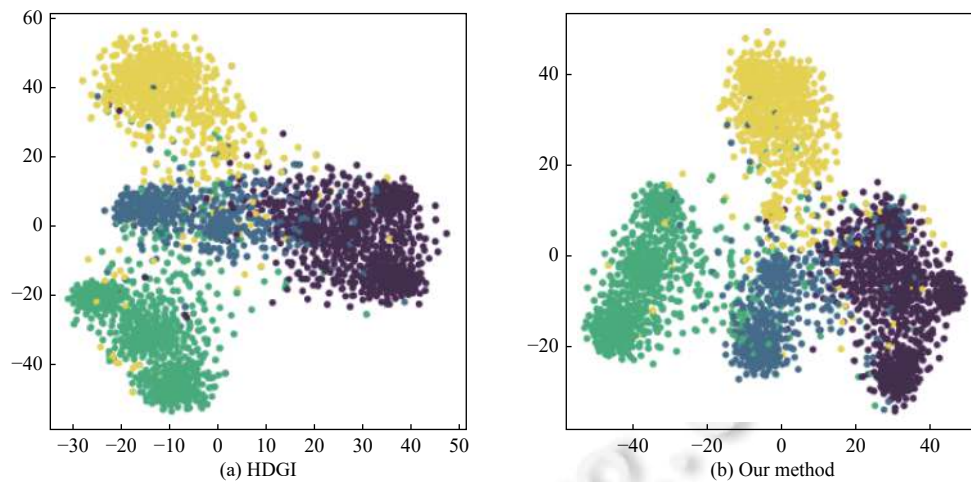


图9 聚类可视化

5 结论

本文提出一种互信息与多条元路径融合的异质网络表示学习方法. 针对大多数现有的 HINE 方法存在的元路径内部节点信息处理不充分, 忽略各元路径实例不同重要性的问题, 采用一种称为关系旋转编码的元路径内部节点编码方式, 基于相邻节点和上下文节点来获取元路径内部节点信息, 使用注意力机制建模各元路径实例的不同重要性和多种语义信息; 针对大多数现有 HINE 方法仅能捕获到节点的局部信息的问题, 提出一种互信息最大化与多条元路径融合的无监督异质网络节点表示学习方法, 利用互信息捕获全局信息以及它和局部信息之间的联系, 进而得到合理有效的节点表示. 在 DBLP 和 IMDB 两个真实数据集上进行实验并与主流算法进行对比与分析, 本文所提出的算法在节点分类和聚类任务上效果都有提升. 在下一步的研究工作中, 我们将致力于研究泛化能力更强、稳定性更好的算法, 并探索比元路径更加高效准确的异质信息提取方法.

References:

- [1] Zhao ZY, Zhou H, Qi L, Chang L, Zhou MC. Inductive representation learning via CNN for partially-unseen attributed networks. *IEEE Trans. on Network Science and Engineering*, 2021, 8(1): 695–706. [doi: [10.1109/TNSE.2020.3048902](https://doi.org/10.1109/TNSE.2020.3048902)]
- [2] Zhao ZY, Zhou H, Li C, Tang J, Zeng QT. DeepEmLAN: Deep embedding learning for attributed networks. *Information Sciences*, 2021, 543: 382–397. [doi: [10.1016/j.ins.2020.07.001](https://doi.org/10.1016/j.ins.2020.07.001)]
- [3] Bourigault S, Lamprier S, Gallinari P. Representation learning for information diffusion through social networks: An embedded cascade model. In: *Proc. of the 9th ACM Int'l Conf. on Web Search and Data Mining*. San Francisco: ACM, 2016. 573–582. [doi: [10.1145/2835776.2835817](https://doi.org/10.1145/2835776.2835817)]
- [4] Tan QY, Liu NH, Hu X. Deep representation learning for social network analysis. *Frontiers in Big Data*, 2019, 2: 2. [doi: [10.3389/fdata.2019.00002](https://doi.org/10.3389/fdata.2019.00002)]
- [5] Wang Q, Mao ZD, Wang B, Guo L. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. on Knowledge and Data Engineering*, 2017, 29(12): 2724–2743. [doi: [10.1109/TKDE.2017.2754499](https://doi.org/10.1109/TKDE.2017.2754499)]
- [6] Ji GL, He SZ, Xu LH, Liu K, Zhao J. Knowledge graph embedding via dynamic mapping matrix. In: *Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing (Vol.1: Long Papers)*. Beijing: Association for Computational Linguistics, 2015. 687–696. [doi: [10.3115/v1/P15-1067](https://doi.org/10.3115/v1/P15-1067)]
- [7] Alshahrani M, Khan MA, Maddouri O, Kinjo AR, Queralt-Rosinach N, Hoehndorf R. Neuro-symbolic representation learning on biological knowledge graphs. *Bioinformatics*, 2017, 33(17): 2723–2730. [doi: [10.1093/bioinformatics/btx275](https://doi.org/10.1093/bioinformatics/btx275)]
- [8] Jin ST, Zeng XX, Xia F, Huang W, Liu XR. Application of deep learning methods in biological networks. *Briefings in Bioinformatics*, 2021, 22(2): 1902–1917. [doi: [10.1093/bib/bbaa043](https://doi.org/10.1093/bib/bbaa043)]
- [9] Dong YX, Chawla NV, Swami A. Metapath2vec: Scalable representation learning for heterogeneous networks. In: *Proc. of the 23rd ACM*

- SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Halifax: ACM, 2017. 135–144. [doi: [10.1145/3097983.3098036](https://doi.org/10.1145/3097983.3098036)]
- [10] Park C, Kim D, Han JW, Yu H. Unsupervised attributed multiplex network embedding. In: Proc. of the AAAI Conf. on Artificial Intelligence, 2020, 34(4): 5371–5378. [doi: [10.1609/aaai.v34i04.5985](https://doi.org/10.1609/aaai.v34i04.5985)]
- [11] Wang X, Ji HY, Shi C, Wang B, Ye YF, Cui P, Yu PS. Heterogeneous graph attention network. In: Proc. of the World Wide Web Conf. San Francisco: ACM, 2019. 2022–2032. [doi: [10.1145/3308558.3313562](https://doi.org/10.1145/3308558.3313562)]
- [12] Ren YX, Liu B, Huang C, Dai P, Bo LF, Zhang JW. Heterogeneous deep graph infomax. arXiv:1911.08538, 2020.
- [13] Pazzani MJ, Billsus D. Content-based recommendation systems. In: Brusilovsky P, Kobsa A, Nejdl W, eds. The Adaptive Web. Lecture Notes in Computer Science, Vol. 4321. Berlin, Heidelberg: Springer, 2007. 325–341. [doi: [10.1007/978-3-540-72079-9_10](https://doi.org/10.1007/978-3-540-72079-9_10)]
- [14] Deshpande M, Karypis G. Item-based top-N recommendation algorithms. ACM Trans. on Information Systems (TOIS), 2004, 22(1): 143–177. [doi: [10.1145/963770.963776](https://doi.org/10.1145/963770.963776)]
- [15] Shi C, Hu BB, Zhao WX, Yu PS. Heterogeneous information network embedding for recommendation. IEEE Trans. on Knowledge and Data Engineering, 2019, 31(2): 357–370. [doi: [10.1109/TKDE.2018.2833443](https://doi.org/10.1109/TKDE.2018.2833443)]
- [16] McCormick C. Word2Vec tutorial-the skip-gram model. 2016. <http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model>.
- [17] Wang X, Zhang YD, Shi C. Hyperbolic heterogeneous information network embedding. Proc. of the 2019 AAAI Conf. on Artificial Intelligence, 2019, 33(1): 5337–5344. [doi: [10.1609/aaai.v33i01.33015337](https://doi.org/10.1609/aaai.v33i01.33015337)]
- [18] Wang LL, Gao CY, Huang CH, Liu RB, Ma WC, Vosoughi S. Embedding heterogeneous networks into hyperbolic space without meta-path. Proc. of the 2021 AAAI Conf. on Artificial Intelligence, 2021, 35(11): 10147–10155.
- [19] Wang DX, Cui P, Zhu WW. Structural deep network embedding. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 1225–1234. [doi: [10.1145/2939672.2939753](https://doi.org/10.1145/2939672.2939753)]
- [20] Chairatanakul N, Liu X, Murata T. PGRA: Projected graph relation-feature attention network for heterogeneous information network embedding. Information Sciences, 2021, 570: 769–794. [doi: [10.1016/j.ins.2021.04.070](https://doi.org/10.1016/j.ins.2021.04.070)]
- [21] Veličković P, Fedus W, Hamilton WL, Liò P, Bengio Y, Hjelm RD. Deep graph infomax. arXiv:1809.10341, 2018.
- [22] Hjelm RD, Fedorov A, Lavoie-Marchildon S, Grewal K, Bachman P, Trischler A, Bengio Y. Learning deep representations by mutual information estimation and maximization. arXiv:1808.06670, 2019.
- [23] Belghazi MI, Baratin A, Rajeshwar S, Ozair S, Bengio Y, Courville A, Hjelm RD. Mutual information neural estimation. In: Proc. of the 35th Int'l Conf. on Machine Learning. Stockholm: PMLR, 2018. 531–540.
- [24] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM, 2014. 701–710. [doi: [10.1145/2623330.2623732](https://doi.org/10.1145/2623330.2623732)]
- [25] Grover A, Leskovec J. Node2vec: Scalable feature learning for networks. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 855–864. [doi: [10.1145/2939672.2939754](https://doi.org/10.1145/2939672.2939754)]
- [26] Yang C, Liu ZY, Zhao DL, Sun MS, Chang EY. Network representation learning with rich text information. In: Proc. of the 24th Int'l Conf. on Artificial Intelligence. Buenos Aires: AAAI Press, 2015. 2111–2117.
- [27] Fu TY, Lee WC, Lei Z. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In: Proc. of the 2017 ACM on Conf. on Information and Knowledge Management. Singapore: ACM, 2017. 1797–1806. [doi: [10.1145/3132847.3132953](https://doi.org/10.1145/3132847.3132953)]
- [28] Wang RJ, Shi C, Zhao TY, Wang X, Ye YF. Heterogeneous information network embedding with adversarial disentangler. IEEE Trans. on Knowledge and Data Engineering, 2022, 34(7): 3225–3240. [doi: [10.1109/TKDE.2021.3096231](https://doi.org/10.1109/TKDE.2021.3096231)]
- [29] Zhang XY, Chen LH. mSHINE: A multiple-meta-paths simultaneous learning framework for heterogeneous information network embedding. arXiv:2104.02433, 2021.
- [30] Yang YM, Guan ZY, Li JX, Zhao W, Cui JT, Wang Q. Interpretable and efficient heterogeneous graph convolutional network. IEEE Trans. on Knowledge and Data Engineering, 2021: 1–14. [doi: [10.1109/TKDE.2021.3101356](https://doi.org/10.1109/TKDE.2021.3101356)]
- [31] Wang HW, Wang J, Wang JL, Zhao M, Zhang WN, Zhang FZ, Xie X, Guo MY. GraphGAN: Graph representation learning with generative adversarial nets. Proc. of the 2018 AAAI Conf. on Artificial Intelligence, 2018, 32(1): 2508–2515.
- [32] Wan GJ, Du B, Pan SR, Haffari G. Reinforcement learning based meta-path discovery in large-scale heterogeneous information networks. Proc. of the 2020 AAAI Conf. on Artificial Intelligence, 2020, 34(4): 6094–6101. [doi: [10.1609/aaai.v34i04.6073](https://doi.org/10.1609/aaai.v34i04.6073)]
- [33] Zhang CX, Song DJ, Huang C, Swami A, Chawla NV. Heterogeneous graph neural network. In: Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 793–803. [doi: [10.1145/3292500.3330961](https://doi.org/10.1145/3292500.3330961)]
- [34] Wang HW, Zhang FZ, Hou M, Xie X, Guo MY, Liu Q. SHINE: Signed heterogeneous information network embedding for sentiment link prediction. In: Proc. of the 11th ACM Int'l Conf. on Web Search and Data Mining. Marina Del Rey: ACM, 2018. 592–600. [doi: [10.1145/3159652.3159666](https://doi.org/10.1145/3159652.3159666)]

- [35] Hu BB, Fang Y, Shi C. Adversarial learning on heterogeneous information networks. In: Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 120–129. [doi: [10.1145/3292500.3330970](https://doi.org/10.1145/3292500.3330970)]
- [36] Fu XY, Zhang JN, Meng ZQ, King I. MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding. In: Proc. of the 2020 Web Conf. Taipei: ACM, 2020. 2331–2341. [doi: [10.1145/3366423.3380297](https://doi.org/10.1145/3366423.3380297)]
- [37] Li X, Ding DH, Kao B, Sun YZ, Mamouli N. Leveraging meta-path contexts for classification in heterogeneous information networks. In: Proc. of the 37th IEEE Int'l Conf. on Data Engineering (ICDE). Chania: IEEE, 2021. 912–923. [doi: [10.1109/ICDE51399.2021.00084](https://doi.org/10.1109/ICDE51399.2021.00084)]
- [38] Hong HT, Guo HT, Lin YC, Yang XQ, Li Z, Ye JP. An attention-based graph neural network for heterogeneous structural learning. Proc. of the 2020 AAAI Conf. on Artificial Intelligence, 2020, 34(4): 4132–4139. [doi: [10.1609/aaai.v34i04.5833](https://doi.org/10.1609/aaai.v34i04.5833)]
- [39] Zhou J, Cui GQ, Hu SD, Zhang ZY, Yang C, Liu ZY, Wang LF, Li CC, Sun MS. Graph neural networks: A review of methods and applications. arXiv:1812.08434, 2021.
- [40] Mnih V, Heess N, Graves A, Kavukcuoglu K. Recurrent models of visual attention. arXiv:1406.6247, 2014.
- [41] Yang YH, Wu L, Hong RC, Zhang K, Wang M. Enhanced graph learning for collaborative filtering via mutual information maximization. In: Proc. of the 44th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Montréal: ACM, 2021. 71–80. [doi: [10.1145/3404835.3462928](https://doi.org/10.1145/3404835.3462928)]
- [42] Sun YZ, Han JW, Yan XF, Yu PS, Wu TY. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. Proc. of the VLDB Endowment, 2011, 4(11): 992–1003. [doi: [10.14778/3402707.3402736](https://doi.org/10.14778/3402707.3402736)]
- [43] Sun ZQ, Deng ZH, Nie JY, Tang J. Rotate: Knowledge graph embedding by relational rotation in complex space. arXiv:1902.10197, 2019.
- [44] Edmunds DE, Evans WD. Spectral Theory and Differential Operators. Oxford: Oxford University Press, 2018.
- [45] Donsker MD, Varadhan SRS. Asymptotic evaluation of certain Markov process expectations for large time, I. Communications on Pure and Applied Mathematics, 1975, 28(1): 1–47. [doi: [10.1002/cpa.3160280102](https://doi.org/10.1002/cpa.3160280102)]
- [46] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907, 2017.
- [47] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. arXiv:1710.10903, 2018.
- [48] MacQueen JB. Some methods for classification and analysis of multivariate observations. In: Proc. of the 5th Berkeley Symp. on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967. 281–297.



贾霄生(1996—), 男, 硕士生, 主要研究领域为学术网络分析, 异质信息网络表示学习。



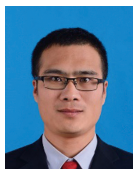
栾文静(1987—), 女, 博士, 讲师, 主要研究领域为基于位置的社交网络, 推荐系统, 机器学习。



赵中英(1983—), 女, 博士, 副教授, 博士生导师, CCF 高级会员, 主要研究领域为网络表示学习, 社交网络分析与挖掘。



梁永全(1983—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为分布式人工智能, 数据挖掘, 机器学习, 多媒体信息智能处理。



李超(1984—), 男, 博士, 副教授, 博士生导师, CCF 高级会员, 主要研究领域为异质图神经网络分析, 自然语言处理, 表示学习。