

基于全局注意力机制的属性网络表示学习

许营坤 马放南 杨旭华 叶蕾

浙江工业大学计算机科学与技术学院 杭州 310023

(xyk@zjut.edu.cn)

摘要 属性网络不仅具有复杂的拓扑结构,其节点还包含丰富的属性信息。属性网络表示学习方法同时提取网络拓扑结构和节点的属性信息来学习大型属性网络的低维向量表示,在节点分类、链路预测和社区识别等网络分析技术方面具有非常重要和广泛的应用。文中首先根据属性网络的拓扑结构得到网络的结构嵌入向量;接着通过全局注意力机制来学习相邻节点的属性信息,先用卷积神经网络对节点的属性信息作卷积操作得到隐藏向量,再对卷积的隐藏向量生成全局注意力的权重向量和相关性矩阵,进而得到节点的属性嵌入向量;最后将结构嵌入向量和属性嵌入向量连接得到同时反映网络结构和节点属性的联合嵌入向量。在3个真实数据集上,将提出的新算法与当前的8种知名网络表示学习模型在链路预测和节点分类等任务上进行比较,实验结果表明新算法具有良好的属性网络表示效果。

关键词: 全局注意力;卷积神经网络;结构嵌入;属性嵌入;联合嵌入

中图分类号 TP391

Attribute Network Representation Learning Based on Global Attention

XU Ying-kun, MA Fang-nan, YANG Xu-hua and YE Lei

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Abstract The attribute network not only has complex topology, its nodes also contain rich attribute information. Attribute network represent learning methods simultaneously extracts network topology and node attribute information to learn low-dimensional vector embedding of large attribute networks. It has very important and extensive applications in network analysis techniques such as node classification, link prediction and community identification. In this paper, we first obtain the embedded vector of the network structure according to the topology of the attribute network. Then we propose to learn the attribute information of adjacent nodes through global attention mechanism, use convolutional neural network to convolve the attribute information of the node to obtain the hidden vectors, and the weight vector and correlation matrix of global attention are generated from the hidden vectors of convolution, and then the attribute embedding vector of nodes is obtained. Finally, the structure embedding vector and the attribute embedding vector are connected to obtain the joint embedding vector which reflects the network structure and the attribute simultaneously. On three real data sets, the new algorithm is compared with the current eight network embedding models on tasks such as link prediction and node classification. Experimental results show that the new algorithm has good attribute network embedding effects.

Keywords Global attention mechanism, Convolutional neural network, Structure embedding, Attribute embedding, Joint embedding

1 引言

当前是一个网络的时代,网络数据挖掘和分析技术(包括节点分类^[1-2]、社区划分^[3-5]、推荐算法^[6-8]、链路预测^[9-10]和图分类^[11-12]等)在现实社会中得到了极其广泛的应用,例如,社交网络中的用户共同兴趣社区发现、通过引文网络揭示学科发展的规律,以及蛋白质相互作用网络中的蛋白质功能预测

等。在网络科学中一个重要的问题就是如何更加合理和高效地表示网络。近年来,学者们对网络嵌入进行了广泛研究。网络的高维度和高稀疏性使网络表示学习充满了挑战性。

当前网络表示学习领域已经有了大量的研究,大体上可以将这些方法分为3类:基于因子分解的方法、基于随机游走的方法和基于深度学习的方法。

基于因子分解的方法:LLE(Locally Linear Embedding)^[13]

到稿日期:2021-01-26 返修日期:2021-04-08

基金项目:国家自然科学基金项目(61773348);浙江省自然科学基金(LY20F020029)

This work was supported by the National Natural Science Foundation of China(61773348) and Zhejiang Province Natural Science Foundation of China(LY20F020029).

通信作者:叶蕾(yelei@zjut.edu.cn)

假设每个节点都是嵌入空间中相邻节点的线性组合,使用邻居节点表示的加权和向量与中心节点表示向量之间的距离作为损失函数,通过最小化损失函数得到节点的嵌入向量。Laplacian Eigenmaps^[14]通过在目标函数中添加惩罚项,使两个相邻的节点在嵌入的向量空间中尽可能靠近,从而保持网络的结构。GraRep^[15]使用矩阵分解的方法作为解决网络嵌入的方案,通过定义 k 阶近邻损失函数,同时捕获网络的局部和全局拓扑结构信息。

因子分解方法由于计算量大、耗时长,很难在大型网络中使用。因此计算复杂度更低和性能更好的网络表示学习方法成为当前时代的需求。受 Word2vec 方法^[16]的启发,DeepWalk 方法^[1]通过随机游走得到节点的序列,用 Word2vec 模型学习得到节点的嵌入向量。Node2vec 方法^[9]采用有偏好的随机游走策略,在广度优先和深度优先搜索之间权衡随机游走得到的序列中节点出现的概率来保持节点之间的高阶邻近性,从而产生比 DeepWalk 质量更高和信息更多的嵌入向量。为摆脱训练容易陷入局部最优的情况,HARP^[17]使用图形粗化聚合层次结构中上一层的节点来创建节点的层次结构,先生成粗糙的图嵌入,通过递归的粗粒化方式,将原网络的节点和边通过合并划分成一系列分层结构更小的网络图,然后利用 DeepWalk 或 Node2vec 方法进行不断的特征提取,从而得到高性能的嵌入向量。

近年来,以深度学习为基础的网络表示学习已成为解决网络分析任务的关键技术。SDNE^[10]使用深度自动编码器来保持一阶和二阶网络节点的相似性,利用高度非线性函数来获取嵌入向量。SDNE 包含一个自动编码器和拉普拉斯特征映射,其中自动编码器利用节点间的二阶相似度来保留网络全局结构,拉普拉斯特征映射^[14]利用一阶相似性来保留网络局部结构。DNGR^[18]结合随机游走和深度自动编码器,在网络上使用随机游走模型生成概率共现矩阵,再将该矩阵转化为 PPMI 矩阵,然后输入到叠加的去噪自动编码器中以获得网络嵌入向量。GCN^[2]通过在网络上定义卷积算子来解决网络嵌入问题,通过迭代聚合节点的邻域嵌入,并结合前一次迭代中获得的嵌入来获得新的嵌入。VGAE^[19]采用了图卷积网络(GCN)编码器和内积解码器,并根据网络的邻接矩阵,用 GCN 来学习节点之间的高阶依赖关系,使用内积解码器来提高嵌入性能。

这些图表示学习模型大多只将网络结构信息进行编码得到节点嵌入向量,而没有考虑现实网络中节点的大量属性信息。为解决属性网络的表示学习问题,学者们提出了一系列属性网络表示学习方法。例如,Yang 等^[20]提出节点属性关联深度游走,在矩阵分解框架下,改进了属性信息深度游走策略,将节点的属性特征引入网络表示学习。Tu 等^[21]提出了最大边距深度游走方法,利用节点的属性信息来学习网络表征。Tu 等^[22]考虑到节点的属性信息,将网络结构和节点的属性信息作为节点嵌入,提出了上下文网络感知嵌入方法。

本文考虑了网络节点的属性信息,对节点的属性编码,通过引入全局注意力机制,解决节点属性信息超载的问题,改善节点属性的长距离依赖关系,以此来学习节点的属性嵌入,再

联合网络的结构嵌入得到节点的联合嵌入向量表示形式。

本文第 2 节介绍了相关工作;第 3 节提出了新的属性网络嵌入模型;第 4 节给出了数值实验和结果分析;最后总结全文。

2 相关工作

2.1 网络嵌入

当前已有大量的网络表示学习模型被提出用于学习有效的节点低维表示^[15,21,23]。例如,DeepWalk^[1]将随机游走和 Skip-Gram 模型^[24]相结合来学习网络嵌入。LINE^[25]对网络中所有的一阶相似度和二阶相似度的节点对进行概率建模,通过最小化该概率分布和经验分布之间的 KL 距离来学习节点表示。Node2vec^[9]综合考虑深度优先和宽度优先邻域的随机游走,有效保留了网络结构。SDNE^[10]使用一个深度自动编码器同时优化一阶和二阶相似度,学习得到的嵌入向量能够保留局部和全局结构,并且对稀疏网络具有鲁棒性。GCN^[2]提出了一种用于非欧网络数据的卷积神经网络,通过对网络的局部结构和节点的特征进行编码,从而得到节点的嵌入向量。

在现实网络中,很多节点具有丰富的属性,例如, Twitter 中有数百万活跃和已连接的用户,他们都可以发布多条推文,这些推文的文本、图像或视频等都是属性信息。节点属性对于网络挖掘任务具有重要价值,如解决大型网络过滤器的泡沫问题^[26]、预测人群的社会行为^[27]等。节点属性包含丰富的信息,将其与网络结构联合考虑可以提高节点嵌入的质量。然而,许多表征方法只将网络结构信息进行编码得到节点嵌入向量,而没有考虑到现实社会网络中节点的属性信息,这会影响嵌入向量的准确度从而产生很多问题。例如,一个节点在与不同的邻居交互时很难展示出自己不同的属性特点,而相邻但属性差异较大的节点的嵌入向量比较靠近等。对此,学者们提出了一系列属性网络表征学习方法。例如,Chen 等^[28]引入分组增强型网络嵌入(GEN),对网络嵌入中现有的分组信息进行整合,以加强网络嵌入。Sun 等^[29]将文本内容视为一种特殊的节点,通过利用结构和文本信息来学习网络嵌入,进而提出了上下文增强的网络嵌入(CENE)。与之前提出的算法不同,本文算法结合注意力机制、网络结构信息和节点属性信息得到节点嵌入向量,以下为相关的研究基础。

2.2 注意力机制

注意力机制可以使神经网络具备专注于其输入特征子集的能力,即选择关注特定的输入。其近几年在图像处理、语音识别、自然语言处理和网络表征学习等不同领域和类型的任务中得到了广泛应用。

注意力机制是一种通用的方法,不依赖于特定的框架,但目前大多数注意力模型都包含 Encoder-Decoder 框架^[30-31]。NMT 模型^[32]将注意力机制应用到机器翻译中的 Encoder-Decoder 中,使得编码器的输出为 encoder 过程中各隐藏层的加权和,对于很长的输入来说,其可以缓解长距离的依赖问题。随着注意力机制的广泛应用,已经涌现出了许多注意力

变体模型来处理更复杂的任务。Luong 等^[33]设计了适用于不同下游任务的 3 种函数,同时提出了 Global Attention 和 Local Attention 机制。Global Attention 在计算隐层向量时,会考虑编码器所有的隐藏状态;而 Local Attention 在计算隐层向量时,只关注一部分编码器的隐藏状态。Vaswani 等^[34]在注意力权重函数中添加了缩放因子,克服了常规方法中 softmax 函数可能具有的极小梯度问题,加速了模型训练。Self-Attention 方法^[35]通过关联单个输入的不同位置来计算注意力权重,以便计算输入的交互表示。GAT^[36]引入 Self-Attention 方法,通过关注网络中的邻居节点来计算每个节点的隐层状态,并训练得到嵌入向量。GeniePath^[37]是一种可扩展的能够学习自适应感受路径的图神经网络模型,其自适应路径层包括两个互补的功能单元,分别用来进行广度与深度的搜索,前者引入注意力机制来学习一阶邻域节点的权重,后者用于提取和过滤高阶邻域内汇聚的信息。

3 基于全局注意力机制的属性网络表示学习

本文考虑了属性网络节点上丰富的属性信息,提出基于全局注意力机制的节点属性嵌入(Global Attention Joint Embedding, GAJE),联合网络的结构嵌入,得到了节点更准确的联合嵌入向量形式。

3.1 问题描述

给定一个有向属性网络 $G=(V, E, T)$, 其中 V 是节点集合, 节点数为 N , E 是边的集合, T 表示节点的属性信息集合。节点 i, j 之间的边 $e_{i,j}$ 的权重为 $w_{i,j}$ 。这里, 任意节点 i 的属性信息表示为一个属性序列向量 $\mathbf{S}_i=(s_{i,1}, s_{i,2}, \dots, s_{i,g})$, 其中 g 为属性序列向量的维度。属性网络嵌入的目标是根据网络结构和节点属性信息来学习任意节点 i 的低维联合嵌入向量 $\mathbf{C}_i \in R^d$, $d \ll N$ 是低维向量 \mathbf{C}_i 的维度。

3.2 算法框架

为了充分利用网络拓扑结构和节点属性信息, 我们首先获取了网络结构的嵌入, 接着提出了基于全局注意力的节点属性嵌入网络算法, 基于结构的嵌入可以捕获网络结构的信息, 而基于节点属性的嵌入可以捕获节点自身及其邻居节点的属性信息。在此基础上, 可以获得任意节点 i 的联合嵌入向量 $\mathbf{C}_i = \mathbf{C}_i^s \oplus \mathbf{C}_i^a$, 其中 \mathbf{C}_i^s 表示节点 i 的结构嵌入向量, \mathbf{C}_i^a 表示节点 i 的属性嵌入向量, \oplus 表示相加操作。

3.3 网络结构嵌入

本文通过最大化边 $e_{i,j}$ 的对数似然函数来获取任意节点 i 的结构嵌入向量 \mathbf{C}_i^s , 如式(1)所示:

$$\sum_{e_{i,j} \in E} L_s(e_{i,j}) = w_{i,j} \log p(\mathbf{C}_i^s | \mathbf{C}_j^s) \quad (1)$$

针对网络中存在的边 $e_{i,j}$, 首先随机初始化节点 i, j 的结构嵌入向量 $\mathbf{C}_i^s, \mathbf{C}_j^s$, 然后通过不断的训练使式(1)达到最大值。当式(1)达到最大值时, 表明节点结构嵌入向量尽可能地还原了网络的拓扑结构信息, 此时 $\mathbf{C}_i^s, \mathbf{C}_j^s$ 为最优的节点结构嵌入向量。其中 $p(\mathbf{C}_i^s | \mathbf{C}_j^s)$ 表示节点 i 基于节点 j 的条件概率, 采用 Tang 等提出的 LINE^[27] 算法中的表现形式为:

$$p(\mathbf{C}_i^s | \mathbf{C}_j^s) = \frac{\exp(\mathbf{C}_i^s \cdot \mathbf{C}_j^s)}{\sum_{z \in V} \exp(\mathbf{C}_z^s \cdot \mathbf{C}_j^s)} \quad (2)$$

其中, \mathbf{C}_i^s 表示节点 i 的结构嵌入向量, \mathbf{C}_j^s 表示节点 j 的结构嵌入向量。

3.4 基于全局注意力的网络节点属性嵌入

本文用卷积神经网络来捕获节点属性的依赖关系, 并结合全局注意力机制来实现网络节点的属性信息嵌入。

本文提出的 GAJE 模型以任意连边 $e_{i,j}$ 的两端节点 i 和 j 的属性作为输入, 通过查找(looking-up)、卷积(convolution)、全局注意力(global attention)和平均池化(mean-pooling)这 4 层实现节点属性嵌入。

(1)查找: 给定一条边 $e_{i,j}$ 与两个对应节点 i, j 的属性序列向量 $\mathbf{S}_i=(s_{i,1}, s_{i,2}, \dots, s_{i,g})$ 和 $\mathbf{S}_j=(s_{j,1}, s_{j,2}, \dots, s_{j,g})$, 在该层将节点各属性 $s_{i,l}$ 和 $s_{j,l}$ 转换为对应的属性低维向量 $\mathbf{u}_{i,l} \in R^o$ 和 $\mathbf{v}_{j,l} \in R^o$, 并获得节点的属性隐层矩阵 $\mathbf{U}=(\mathbf{u}_{i,1}, \mathbf{u}_{i,2}, \dots, \mathbf{u}_{i,g})$ 和 $\mathbf{V}=(\mathbf{v}_{j,1}, \mathbf{v}_{j,2}, \dots, \mathbf{v}_{j,g})$ 。其中, o 表示属性低维向量的维度, l 表示属性序列向量的任一维度, g 表示节点属性序列向量的维度。

(2)卷积: 在 looking-up 之后, 卷积层提取节点属性隐层矩阵 \mathbf{U} 和 \mathbf{V} 的特征, 生成节点低维属性嵌入向量。具体地, 使用卷积核 $\mathbf{H} \in R^{d \times (m \times o)}$ 在长度 m 的滑动窗口上执行卷积操作, 如式(3)和式(4)所示:

$$\mathbf{P} = \mathbf{H} \cdot \mathbf{U}_{a:a+m-1} + \mathbf{b} \quad (3)$$

$$\mathbf{Q} = \mathbf{H} \cdot \mathbf{V}_{a:a+m-1} + \mathbf{b} \quad (4)$$

得到节点 i 和 j 的低维属性嵌入矩阵 $\mathbf{P} \in R^{d \times g}$ 和 $\mathbf{Q} \in R^{d \times g}$ 。其中, $\mathbf{U}_{a:a+m-1}$ 和 $\mathbf{V}_{a:a+m-1}$ 分别表示在 \mathbf{U} 和 \mathbf{V} 的第 a 个窗口中低维属性向量的连接, \mathbf{b} 是偏移向量, d 为超参数表示嵌入的维度。我们在 \mathbf{U} 和 \mathbf{V} 的边缘添加了零填充向量^[40]。

(3)全局注意力: 在生成注意力权重向量时, 考虑节点低维属性嵌入矩阵 \mathbf{P} 和 \mathbf{Q} 的所有状态。根据当前节点 i 的低维属性嵌入 \mathbf{P} 和相邻节点 j 的低维属性嵌入 \mathbf{Q} , 对比矩阵 \mathbf{P}^T 和 \mathbf{Q} 的每一行, 生成注意力权重向量 $\boldsymbol{\alpha} \in R^g$:

$$\boldsymbol{\alpha}_i = \frac{\exp(\mathbf{P}^T \mathbf{Q})}{\sum \exp(\mathbf{P}^T \mathbf{Q})} \quad (5)$$

随后计算相关性矩阵 $\mathbf{F} \in R^{g \times g}$, 将注意力权重向量对应对应的权重与 $\mathbf{P}^T \mathbf{Q}$ 的对应的行向量一一相乘, 如式(6)所示:

$$\mathbf{F} = \boldsymbol{\alpha}_i \tanh(\mathbf{P}^T \mathbf{Q}) \quad (6)$$

(4)平均池化: 为了获得节点的属性嵌入, 我们沿着相关性矩阵 \mathbf{F} 的行和列进行平均池化操作, 生成重要性向量, 分别称为行池化和列池化, 公式如下:

$$\mathbf{f}_k^p = \text{mean}(\mathbf{F}_{k,1}, \mathbf{F}_{k,2}, \dots, \mathbf{F}_{k,g}) \quad (7)$$

$$\mathbf{f}_k^q = \text{mean}(\mathbf{F}_{1,k}, \mathbf{F}_{2,k}, \dots, \mathbf{F}_{g,k}) \quad (8)$$

\mathbf{P} 的权重向量为 $\mathbf{f}^p = [f_1^p, f_2^p, \dots, f_g^p]^T$, \mathbf{Q} 的权重向量为 $\mathbf{f}^q = [f_1^q, f_2^q, \dots, f_g^q]^T$ 。接下来, 我们使用 softmax 函数将权重向量 \mathbf{f}^p 和 \mathbf{f}^q 转换为归一化权重向量 \mathbf{a}^p 和 \mathbf{a}^q 。例如, \mathbf{a}^p 和 \mathbf{a}^q 的第 k 个元素形式化如下:

$$\mathbf{a}_k^p = \frac{\exp(\mathbf{f}_k^p)}{\sum_{h \in [1, g]} \exp(\mathbf{f}_h^p)} \quad (9)$$

$$\mathbf{a}_k^q = \frac{\exp(\mathbf{f}_k^q)}{\sum_{h \in [1, g]} \exp(\mathbf{f}_h^q)} \quad (10)$$

最后, 节点 i 和 j 的属性嵌入的向量形式可以表示为:

$$C_i^f = Pa^p \quad (11)$$

$$C_j^f = Qa^q \quad (12)$$

属性嵌入算法如图 1 所示。

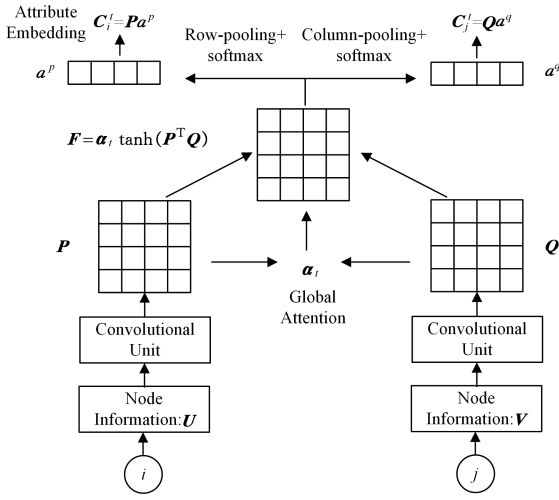


图 1 基于全局注意力的网络节点的属性嵌入示意图

Fig. 1 Attribute embedding based on global attention

3.5 网络结构和属性的联合嵌入

在任意给定一条边 (i, j) 的情况下,我们可以通过网络的结构嵌入和节点的属性嵌入来获得节点的联合嵌入 $C_i = C_i^s \oplus C_i^f$ 和 $C_j = C_j^s \oplus C_j^f$, \oplus 表示相加操作。

为了得到同时反映网络结构和节点属性的联合嵌入向量,我们定义目标函数如下:

$$L = \alpha \cdot L_u(e) + \beta \cdot L_s(e) + \gamma \cdot L_a(e) \quad (13)$$

其中, α, β 和 γ 为控制各个部分的权重的超参,并且有:

$$L_u(e) = w_{i,j} \cdot \log p(C_i^f | C_j^f) \quad (14)$$

$$L_s(e) = w_{i,j} \cdot \log p(C_i^s | C_j^s) \quad (15)$$

$$L_a(e) = w_{i,j} \cdot \log p(C_i^a | C_j^a) \quad (16)$$

式(14)~式(16)将两种类型的节点嵌入映射到同一特征空间。类似式(2),我们使用 softmax 函数来计算条件概率 $p(C_i^f | C_j^f)$, $p(C_i^s | C_j^s)$ 和 $p(C_i^a | C_j^a)$ 。

对 $C_i \in \{C_i^s, C_i^f\}$ 和 $C_j \in \{C_j^s, C_j^f\}$, 我们的目标是使式(13)最大化,但使用 softmax 函数优化条件概率的计算代价昂贵。因此,我们采用负采样^[41],并将目标转换为以下形式:

$$\log \sigma(C_i^T \cdot C_j) + \sum_{i=1}^k E_{z \sim p(v)} [\log \sigma(-C_i^T \cdot C_z)] \quad (17)$$

其中, k 是负样本数, σ 表示 sigmoid 函数, $P(v) \propto d_v^{3/4}$ 表示节点的分布, d_v 是节点 v 的出度^[25]。之后,我们使用 Adam^[40] 优化目标函数。

4 实验

本文在 3 个真实数据集上进行了链接预测和节点分类实验,将所提方法与 8 种网络表示学习方法进行对比,实验结果表明 GAJE 算法具有高质量的网络表示效果。

4.1 数据集

本文挑选了 3 个真实网络,分别如下。

Cora: 一个典型的论文引文网络^[41]。在过滤掉没有文本信息的论文后,该网络共有 2277 篇机器学习论文,分为 7 类。

HepTh: arXiv 的另一个引文网络^[42]。我们过滤掉没有摘要的论文,最后保留了 1038 篇论文。

Zhihu: 中国最大的在线问答网站。用户在这个网站上互相关注并回答问题。我们随机从 Zhihu 抓取 10000 个活跃用户,并将其关注主题的描述作为节点属性。表 1 列出了 3 个数据集的详细统计数据。

表 1 数据集的信息

Table 1 Information of data sets

Datasets	Cora	HepTh	Zhihu
Nodes	2277	1038	10000
Edges	5214	1990	43894
Labels	7	—	—

4.2 基准算法

本文采用以下算法作为比较对象。其中,前 4 个算法为结构嵌入算法,后 4 个算法为结构和属性嵌入算法。

(1)MMB^[43]: 混合隶属度随机块模型,是一种传统的关系数据网络模型,它允许每个节点在形成边时随机选择不同的属性。

(2)DeepWalk^[1]: 在网络上执行随机游动,并使用 Skip-Gram 模型^[24]学习节点的嵌入。

(3)LINE^[25]: 使用一阶和二阶相似性来学习大规模网络中的节点嵌入。

(4)Node2vec^[9]: 提出了一种基于 DeepWalk 的有偏随机游走算法,以便于更有效地探索邻域结构。

(5)Naive Combination: 将 LINE 与基于卷积神经网络(不使用 global attention)的属性嵌入连接起来,以表示节点。

(6)TADW^[20]: 使用矩阵分解将节点的属性特征合并到网络嵌入中。

(7)CENE^[29]: 通过将文本属性视为一种特殊的节点来利用结构和属性,并优化异构连接的概率。

(8)CANE^[22]: 结构信息借助深层网络表示,并考虑到不同节点间的属性来实现嵌入。

4.3 评估指标和实验设置

对于链路预测,本文采用一个标准的评估度量 AUC(Area Under Curve)^[44],其表示预测的正例排在负例前面的概率。

对于节点分类,我们使用 L2 正则逻辑回归^[45]来训练分类器,并评估各种方法的分类精度。

为了公平起见,我们将所有方法的嵌入维度均设置为 200。在 LINE 算法中,我们将负样本数设为 5,分别学习 100 维一阶和二阶嵌入,并将它们连接起来形成 200 维嵌入。在 Node2vec 中,采用网格搜索的方法,选择性能最好的超参数进行训练。应用网格搜索法对 GAJE 的超参数 α, β 和 γ 进行设置。此外,为了加快训练过程,将 GAJE 中的负样本数 k 设为 1。

4.4 链路预测

我们分别在 Cora, HepTh 和 Zhihu 上去除不同比例的边来得到 AUC 值,结果如表 2~表 4 所列。需要注意的是,当只保留 5% 的边进行训练时,大多数节点都是孤立的,这会导致所有方法的性能都很差,没有意义。因此,我们忽略了在该训练比率下的结果。从表 2~表 4 可以得出以下结果:

(1)在不同的训练比率下,与所有不同数据集上的所

有算法相比,GAJE方法始终有显著的改进。该结果表明本文方法对于链路预测任务是有效的,并验证了其具有精确建模节点属性的能力。

表2 Cora数据集的链路预测 AUC 指标($\alpha=0.7, \beta=0.7, \gamma=0.8$)

Table 2 Link prediction AUC metric for Cora

($\alpha=0.7, \beta=0.7, \gamma=0.8$)

(单位:%)

algorithms	Training edges								
	15	25	35	45	55	65	75	85	95
MMB	54.9	57.3	60.2	61.9	65.3	68.0	71.4	72.7	76.1
DeepWalk	56.4	63.2	70.5	75.8	80.6	85.7	85.9	87.8	90.4
LINE	55.3	58.6	66.8	73.4	77.8	82.9	86.1	88.6	89.6
Node2vec	55.9	62.5	66.5	75.4	78.9	81.6	86.3	87.8	88.7
NV	73.4	82.5	85.2	87.3	88.9	91.9	92.6	93.9	94.1
TADW	87.0	88.3	90.4	91.0	89.8	93.2	91.2	93.5	93.1
CENE	72.5	86.6	84.6	88.4	89.4	89.4	93.9	95.3	96.1
CANE	86.8	91.5	92.2	93.9	94.6	94.9	95.6	96.6	97.7
GAJE	88.5	91.3	92.7	95.1	95.2	97.4	98.1	98.8	99.1

表3 HepTh数据集的链路预测 AUC 指标($\alpha=1.0, \beta=0.3, \gamma=0.3$)

Table 3 Link prediction AUC metric for HepTh

($\alpha=1.0, \beta=0.3, \gamma=0.3$)

(单位:%)

algorithms	Training edges								
	15	25	35	45	55	65	75	85	95
MMB	54.7	57.9	57.5	61.9	66.3	68.4	73.6	76.1	80.3
DeepWalk	55.2	66.1	70.1	75.8	81.3	83.4	87.6	88.9	88.4
LINE	53.7	60.6	66.5	74.0	78.5	83.9	87.5	87.8	87.8
Node2vec	57.3	63.6	69.9	76.4	84.5	87.3	88.4	89.2	89.3
NV	78.7	82.2	84.7	88.7	88.9	91.8	92.1	92.1	92.7
TADW	87.1	89.5	91.8	91.0	91.1	92.7	93.5	91.9	91.7
CENE	86.2	84.6	89.9	91.3	92.3	91.8	93.2	93.1	93.2
CANE	90.0	91.2	92.0	93.1	94.2	94.6	95.5	95.7	96.3
GAJE	89.8	91.8	92.8	94.0	94.8	95.9	96.8	96.9	97.2

表4 Zhihu数据集的链路预测 AUC 指标($\alpha=0.1, \beta=1.0, \gamma=0.1$)

Table 4 Link prediction AUC metric for Zhihu

($\alpha=0.1, \beta=1.0, \gamma=0.1$)

(单位:%)

algorithms	Training edges								
	15	25	35	45	55	65	75	85	95
MMB	51.0	51.5	53.8	58.6	61.6	66.1	68.9	68.9	72.4
DeepWalk	56.6	58.1	60.2	60.1	61.9	61.8	63.3	63.7	67.8
LINE	52.3	55.9	59.9	60.9	64.5	66.1	67.7	69.4	71.1
Node2vec	54.2	57.2	57.3	58.4	58.7	62.5	66.2	67.6	68.5
NV	55.2	56.7	58.9	62.6	64.5	68.7	68.9	69.2	71.6
TADW	52.3	54.3	55.6	57.3	60.8	62.4	65.3	63.8	69.0
CENE	56.2	57.6	60.3	63.0	66.3	66.1	70.2	69.9	73.8
CANE	56.8	59.3	62.9	64.6	68.9	70.4	71.4	73.6	75.5
GAJE	63.3	70.5	73.9	75.7	75.9	77.0	79.2	81.0	81.1

(2)需要特别注意的是,在不同的训练比例下,CENE和TADW方法表现不稳定。具体来说,CENE在较小的训练比率下表现不佳,因为它比TADW保留了更多的参数(如卷积核和词嵌入),而TADW需要更多的训练数据。与CENE不同的是,TADW在小训练比率下表现得更好,因为基于DeepWalk的方法即使在边数量有限的情况下也可以通过随机游走很好地探索稀疏网络结构。然而,由于DeepWalk的简单性和bag-of-words假设的局限性,使得TADW在大规模网络的情况下表现不佳。CANE在训练节点属性时未完全考虑节点间的属性依赖关系,因此没有达到最佳效果。相反,本文

方法GAJE在各种情况下都有稳定的表现,展示了其灵活性和鲁棒性。

(3)本文方法通过引入全局注意力机制,学习到的节点的属性嵌入比不引入注意力机制的嵌入算法获得了更好的改进,有利于完成相关的链路预测任务。

综上所述,GAJE能够学习高质量的属性嵌入,有准确学习节点属性的能力,证明了全局注意力机制,以及式(2)、式(14)一式(16)这两类条件概率的有效性。此外,在链路预测任务上的实验结果表明了GAJE的有效性和鲁棒性。

4.5 节点分类

本文方法根据节点的属性建模得到节点的属性嵌入向量,获得的属性嵌入向量有助于相关链路预测。为了演示GAJE在节点分类上的能力,本文进行了2折交叉验证,并给出了在Cora上节点分类的平均精度,如图2所示。可以观察到:

(1)本文提出的GAJE方法的性能与最先进的CENE算法相比,有明显的提升,这表明联合属性嵌入和结构嵌入能够得到高质量的节点嵌入向量,并可将其进一步应用于其他网络分析任务。

(2)引入全局注意力机制后,本文算法相比无注意力机制的算法有了不错的改进,这与链路预测的结果一致。这表明本文方法可以灵活地处理各种网络分析任务。

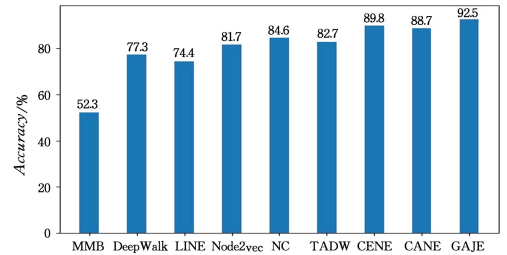


图2 Cora上各算法节点分类的结果

Fig. 2 Results of each algorithm node classification on Cora

结束语 本文提出了基于全局注意力的节点属性嵌入方法,其联合了结构嵌入,得到了高质量的且同时反映了网络结构和属性的联合嵌入向量。在真实数据集上将本文方法与8种网络表示学习方法相比较,在链路预测和节点分类两个方面,本文提出的GAJE算法均显示出准确的网络嵌入性能,同时具有良好的鲁棒性。属性网络嵌入为大型信息系统与网络分析技术之间的连接架设了桥梁,在本文中,利于节点上的文本信息得到节点的嵌入向量。未来,我们将致力于节点的图像、视频和声音等多模态信息的网络表示学习研究,我们认为多模态信息网络嵌入在社交网络、电商网络、数据资产管理和异构数据识别等领域将有广阔的前景。

参考文献

- [1] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: Online Learning of Social Representations[C] // Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York: ACM, 2014: 701-710.
- [2] KIP F T N, WELLING M. Semi-supervised Classification with Graph Convolutional Networks[C] // 5th International Confe-

- rence on Learning Representations. ShangHai,2017.
- [3] WANG C,PAN S, LONG G, et al. MGAE: Marginalized Graph Autoencoder for Graph Clustering[C]// International Conference on Information and Knowledge Management. Singapore: ACM,2017:889-898.
- [4] XIONG F,WANG X,PAN S, et al. Social Recommendation with Evolutionary Opinion Dynamics[J]. IEEE Transactions on Systems Man Cybernetics-Systems,2018,50(10):3804-3816.
- [5] ZHAO X,LI X,ZHANG Z H, et al. Community discovery algorithm combining community embedding and node embedding [J]. Computer Science,2020,47(12):279-284.
- [6] CAI H,ZHENG V,CHANG K. A Comprehensive Survey of Graph Embedding:Problems, Techniques and Applications[J]. IEEE Transactions on Knowledge and Data Engineering,2018,30(9):1616-1637.
- [7] SHI C,HU B,ZHAO W, et al. Heterogeneous Information Network Embedding for Recommendation[J]. IEEE Transactions on Knowledge and Data Engineering,2019,31(2):357-370.
- [8] ZHOU L E, YOU J G. Social Recommendation with Embedding of Summarized Graphs[J]. Journal of Chinese Computer Systems,2021,42(1):78-84.
- [9] GROVER A,LESKOVEC J. Node2vec: Scalable Feature Learning for Networks[C]// Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM,2016:855-864.
- [10] WANG D,CUI P,ZHU W. Structural Deep Network Embedding[C]// Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM,2016:1225-1234.
- [11] IVANOV S,BURNAEV E. Anonymous Walk Embeddings [C]//35th International Conference on Machine Learning. Stockholm: ACM,2018:3448-3457.
- [12] DUTTA A,RIBA P,LIADOS J, et al. Hierarchical Stochastic Graphlet Embedding for Graph-based Pattern Recognition[J]. Neural Computing & Applications,2019,32(15):11596-11597.
- [13] ROWEIS S,SAUL L. Nonlinear Dimensionality Reduction by Locally Linear Embedding[J]. Science,2000,290(5500):2323-2326.
- [14] BELKIN M,NIYOGI P. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering[J]. Advances in Neural Information Processing Systems,2001,14:585-591.
- [15] CAO S,LU W,XU Q. GraRep: Learning Graph Representations with Global Structural Information[C]// International Conference on Information and Knowledge Management. Melbourne: ACM,2015:891-900.
- [16] MIKOLOV T,CHEN K,CORRADO G, et al. Efficient Estimation of Word Representations in Vector Space[C]//1st International Conference on Learning Representations. Arizona,2013.
- [17] CHEN H,HU Y,PEROZZI B, et al. HARP-Hierarchical Representation Learning for Networks[C]// 32nd AAAI Conference on Artificial Intelligence. New Orleans: AAAI,2018:2127-2134.
- [18] CAO S,LU W,XU Q. Deep Neural Networks for Learning Graph Representations[C]// Thirtieth AAAI Conference on Artificial Intelligence. Arizona: AAAI,2016:1145-1152.
- [19] KIPF T N,WELLING M. Variational Graph Auto-Encoders [C]// Proceedings of NIPS. Barcelona: MIT Press,2016.
- [20] YANG C,LIU Z,ZHAO D, et al. Network Representation Learning with Rich Text Information[C]// IJCAI International Joint Conference on Artificial Intelligence. Buenos Aires: Morgan Kaufmann,2015:2111-2117.
- [21] TU C,ZHANG W,LIU Z, et al. Max-Margin DeepWalk: Discriminative learning of network representation[C]// IJCAI International Joint Conference on Artificial Intelligence. New York: Morgan Kaufmann,2016:3889-3895.
- [22] TU C,HAN L,LIU Z, et al. CANE: Context-Aware Network Embedding for Relation Modeling[C]// ACL 2017-55th Annual Meeting of the Association for Computational Linguistics. Vancouver: ACL,2017:1722-1731.
- [23] TU C,WANG H,ZENG X, et al. Community-enhanced Network Representation Learning for Network Analysis[J]. arXiv: 1611.06645,2016.
- [24] TOMAS M,IYYA S,KAI C, et al. Distributed Representations of Words and Phrases and Their Compositionality[C]// Proceedings of NIPS. Lake Tahoe: MIT Press,2013:3111-3119.
- [25] TANG J,QU M,WANG M, et al. LINE: Large-scale Information Network Embedding[C]// Proceedings of the 24th International Conference on World Wide Web. New York: ACM,2015: 1067-1077.
- [26] LAHOTI P,GARIMELLA K,GIONIS A. Joint non-negative Matrix Factorization for Learning Ideological Leaning on Twitter[C]// WSDM'18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Los Angeles: ACM,2018:351-359.
- [27] TAN C,TANG J,SUN J, et al. Social Action Tracking Via Noise Tolerant Time-varying Factor Graphs[C]// Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington: ACM,2010: 1049-1058.
- [28] CHEN J,ZHANG Q,HUANG X. Incorporate Group Information to Enhance Network Embedding[C]// International Conference on Information and Knowledge Management. Pollis: ACM,2016:1901-1904.
- [29] SUN X,GUO J,DING X, et al. A General Framework for Content-enhanced Network Representation Learning [J]. arXiv: 1610.02906v3,2016.
- [30] CHO K,VAN M,GULCEHRE C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[C]// EMNLP 2014—2014 Conference on Empirical Methods in Natural Language Processing. Doha: ACL,2014: 1724-1734.
- [31] SUTSKEVER I,VINYALS O,LE Q. Sequence to Sequence Learning with Neural Networks[C]// Advances in Neural Information Processing Systems. Montréal: MIT Press,2014: 3104-3112.
- [32] BAHDANAU D,CHO K,BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[C]// 3rd International Conference on Learning Representations. San Diego,2015.

- [33] LUONG M, PHAM H, MANNING C. Effective approaches to attention-based neural machine translation [C] // Conference Proceedings—EMNLP 2015; Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015; 1412-1421.
- [34] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [C] // Advances in Neural Information Processing Systems. Long Beach: MIT Press, 2017; 5999-6009.
- [35] CHENG J, DONG L, LAPATA M. Long Short-Term Memory—Networks for Machine Reading [C] // EMNLP 2016—Conference on Empirical Methods in Natural Language Processing. Austin: ACL, 2016; 551-561.
- [36] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks [C] // 6th International Conference on Learning Representations. Vancouver, 2018.
- [37] LIU Z, CHEN C, LI L, et al. Graph Neural Networks with Adaptive Receptive Paths [C] // 33rd AAAI Conference on Artificial Intelligence. Xi'an: AAAI, 2019; 4424-4431.
- [38] HU B, LU Z, LI H, et al. Convolutional neural network architectures for matching natural language sentences [C] // Advances in Neural Information Processing Systems 27-28th Annual Conference on Neural Information Processing Systems 2014. Montreal: MIT Press, 2014; 2042-2050.
- [39] TOMAS M, IIYA S, KAI C, et al. Distributed Representations of Words and Phrases and Their Compositionality [C] // Proceedings of NIPS. Lake Tahoe: MIT Press, 2013; 3111-3119.
- [40] KINGMA D, BA J. Adam: A Method for Stochastic Optimization [C] // Proceedings of ICLR. Montreal, 2015.
- [41] MCCALLUM A, NIGAM K, RENNIE J, et al. Automating the Construction of Internet Portals with Machine Learning [J]. Information Retrieval, 2000, 3(2): 127-163.
- [42] JURE L, JON K, CHRISTOS F. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations [C] // Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. Chicago: ACM, 2005; 177-187.
- [43] AIROLDI E, BLEI D, FIENBERG S, et al. Mixed membership stochastic blockmodels [C] // Advances in Neural Information Processing Systems. Whistler: MIT Press, 2008; 1981-2014.
- [44] HANLEY J, MCNEIL B. The Meaning and Use of the Area Under a Receiver Operating Characteristic (roc) Curve [J]. Radiology, 1982, 143(1): 29-36.
- [45] FAN R, CHANG K, HSIEH C, et al. Liblinear: A Library for Large Linear Classification [J]. Journal of Machine Learning Research, 2008, 9(9): 1871-1874.



XU Ying-kun, born in 1979, Ph.D, lecturer, is a member of China Computer Federation. His main research interests include machine learning and classification algorithm.



YE Lei, born in 1979, Ph.D, associate professor. Her main research interests include machine learning and so on.