

分 数:	
评卷人:	

華 中 科 技 大 學

数 据 中 心 技 术 课 程 论 文 （ 报 告 ）

题 目：具有超参数重要性的可解释自动图表示学习

学 号 M202173855

姓 名 刘瑞祺

专 业 电子信息

课程指导教师 施展 童薇

院（系、所） 计算机科学与技术学院

2021 年 1 月 5 日

图嵌入技术综述

刘瑞祺¹⁾

¹⁾(华中科技大学计算机科学与技术, 武汉 430074)

摘 要 现实中很多数据都是非结构化的图结构, 例如社交网络、经济网络、生物医学网络、信息网络(互联网站、学术引用)、互联网、神经网络。而图数据是它们的通用语言, 因此具备极大的研究价值, 分析图数据可以洞察社会结构、语言和不同的交流模式。目前学术界已经提出了许多方法用于分析图数据, 最近使用向量方法进行图节点表示的方法已经获得了广泛的研究关注。在本文的综述报告中, 首先介绍了图嵌入领域的主要研究任务及其当前面临的挑战, 如可扩展性、维度选择和保留特征, 以及对应的解决方案, 简要介绍现有方法的原理及其优势。随后我们提出了图嵌入领域的几种主流嵌入方案, 基于矩阵分解方法、基于随机游走的方法和基于深度学习模型, 举例说明了每种方法中的代表性算法, 并分析了它们在各种任务中的性能。最后通过分析当前研究问题提出了一些可能存在的潜在应用和未来发展方向。

关键词 图嵌入; 矩阵分解; 随机游走; 深度学习; 图神经网络

中图法分类号 TP302

Overview of Graph Embedding

Liu RuiQi¹⁾

¹⁾(Computer science and technology, Huazhong University of Science and Technology, Wuhan 430074)

Abstract In reality, many data are unstructured graph structures, such as social networks, economic networks, biomedical networks, information networks (Internet websites, academic references), the Internet and neural networks. Graph data is their common language, so it has great research value. Analyzing graph data can gain insight into social structure, language and different communication modes. At present, many methods have been proposed to analyze graph data. Recently, the method of graph node representation using vector method has attracted extensive research attention. In the overview report of this paper, firstly, the main research tasks and current challenges in the field of graph embedding are introduced, such as scalability, dimension selection and retention features, as well as the corresponding solutions, and the principles and advantages of existing methods are briefly introduced. Then we propose several mainstream embedding schemes in the field of graph embedding, which are based on matrix decomposition method, random walk method and depth learning model. We illustrate the representative algorithms of each method, and analyze their performance in various tasks. Finally, by analyzing the current research problems, some possible potential applications and future development directions are put forward.

Key words graph embedding; Matrix decomposition; random walk; deep learning; GNN

1 引言

真实世界中的复杂系统通常都可以建模为图的结构, 用图中的节点或节点属性表示真实网络系

统中的实体或实体标签, 用图中的边表示真实网络中的实体关系, 通过图嵌入(Graph Embedding, GE)研究, 可更好地分析真实网络系统的结构和性质, 如在微信网络中为用户推荐好友或感兴趣的内容和广告等, 鉴于以上应用的重要价值, 近年来图嵌

入引起了学界和业界的广泛关注。

大数据背景下, 图结构或图数据呈现出了海量、高维、稀疏、异构、复杂和动态等特点, 图嵌入也面临以下五个方面的挑战: 节点海量性、节点属性信息的融合、图的异构性、节点动态增量性和模型的非线性。在图数据分析的研究中, 可采用包括图嵌入、模型构建以及问题求解这三个步骤的支撑技术, 图嵌入是最关键的步骤, 是一种图表示学习(Graph Representation Learning)方法, 需要将高维向量映射到低维空间, 图嵌入方法根据其对象不同可分为两类, 一类是整个图的嵌入、只涉及小规模图, 另一类是大规模图中节点或边的嵌入, 因此图嵌入方法需实现以下两方面的目标: 一是在低维空间中学习得到的嵌入向量可重构原图结构, 二是学习得到的嵌入向量可以支持图的推理分析任务, 对一般的图嵌入任务流程如图 1 所示。

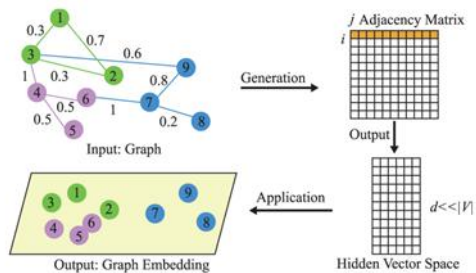


图 1 图嵌入流程示意图

基于图嵌入的分析任务可以大致抽象为以下四类: 1. 节点分类, 2. 链接预测, 3. 聚类, 4. 可视化。节点分类的目的是根据其他被标记的节点和网络的拓扑确定节点的标签。链接预测是指预测原始图数据中缺失的链接或未来可能发生的链接的任务。聚类是用来找到相似节点的子集, 并将它们组合在一起; 最后, 可视化有助于洞察网络的结构。而在过去的几十年里, 学术界已针对上述任务提出了许多方法。对于节点分类, 大致有两类方法——使用随机游走传播标签的方法, 以及从节点提取特征并对其应用分类器的方法。链接预测方法包括基于相似度的方法、极大似然模型和概率模型。聚类方法包括基于属性分析的模型和直接最大化属性见距离的方法或者最小化集群间距离的方法。

通常, 传统意义上的图嵌入的方法一般在原始图的邻接矩阵上操作, 这种传统基于邻接矩阵的方法如 PCA 分解和 SVD 分解不能适用于大型网络, 计算开销非常大。后来基于随机游走的策略结合 NLP 领域的方法与模型, 成功在图嵌入任务上提供

了灵活的解决方案。近年来主要研究采用深度学习模型实现图嵌入方法, 如图卷积神经网络和图注意力网络成功采用深度学习模型对图结构特征进行自动化提取。这些在向量空间中表示网络的方法, 能够实现在保持其原始图性质的同时, 挖掘领域性信息和社区特征, 因此基于向量空间表示方法已经成为图嵌入的广泛流行方案。将嵌入结果作为特征输入到下游任务的计算模型中, 根据训练数据学习参数, 可以消除直接应用于图上的复杂分类模型的需要。

设计图嵌入方法需要考虑以下三个方面的问题: 1. 属性的选择: 一个优良的节点向量表示应该保留图的结构和各个节点之间的连接。所以第一个挑战是选择嵌入应该保留的图形的属性。考虑到原始图形定义的距离度量和属性数量庞大, 这种选择可能会很困难, 算法的性能却决于具体任务的具体设计。2. 可扩展性: 大多数真实的网络都很大, 包含数百万个节点和链接——所以嵌入方法应该是可扩展的, 能够处理巨大规模的图数据。所以如何设计一个可扩展的模型同样是一个挑战, 特别是当任务需要保持网络的全局属性时, 模型的设计将会变得及其困难。3. 嵌入的维数: 嵌入表达的维数同样也是一个问题, 在不同的任务背景下找到嵌入表示的最佳维数同样很困难, 不同维度的嵌入表达各有其优点所在。例如, 较高的维数可以提高重建精度, 但会有较高的时间和空间复杂度。但如果所选的模型只捕获节点之间的局部连接, 则更低的维数可能会导致更好的链接预测精度。

接下来, 本文将系统的介绍一些与图嵌入相关的理论, 以及一些经典模型, 最后将详细的介绍该领域内三篇文章的贡献点及不足。

2 理论

根据上述介绍的三种主流图嵌入方法, 本节将简要介绍不同方法的原理及其优势。

2.1 矩阵分解

2.1.1 定义

基于矩阵分解的算法将节点间的连接以矩阵的形式表示出来,并将该矩阵因式分解得到图嵌入表达。用于表示连接的矩阵包括节点邻接矩阵、拉普拉斯矩阵、节点转移概率矩阵和 Katz 相似矩阵等。对代表矩阵进行因式分解的方法因矩阵的性质而异。如果得到的矩阵是正半定的(例如拉普拉斯矩阵),则可以使用特征值分解。对于非结构化矩阵,可以使用梯度下降法在线性时间内获得嵌入。

2.1.2 拉普拉斯特征值映射

为什么在网络分区的情况下,系统在一致性和可用性之间只能二选一呢?

拉普拉斯特征映射的目标是在保证链接权值 W_{ij} 较高时两个节点在嵌入之后空间应当更加接近。具体来说,它通过最小化以下目标函数完成嵌入:

$$\phi(Y) = \frac{1}{2} \sum_{i,j} |Y_i - Y_j|^2 W_{ij} = \text{tr}(Y^T L Y)$$

其中 L 为图 G 的拉普拉斯算子,目标函数受 $Y^T D Y = I$ 的约束。这个问题的解可以通过取归一化拉普拉斯矩阵的 d 个最小特征值对应的特征向量,即 $L_{\text{norm}} = D^{-1/2} L D^{-1/2}$ 。

2.1.2 柯西图嵌入

柯西图嵌入通过将拉普拉斯损失函数进行变化保留节点相似性信息,经过重新设计,目标函数的最大值变为:

$$\phi(Y) = \sum_{i,j} \frac{W_{ij}}{|Y_i - Y_j|^2 + \sigma^2}$$

这种新的目标设计为距离的反函数,重点放在表达节点的相似性而非节点的差异性上。

2.2 随机游走

2.2.1 定义

在矩阵分解方法之后,基于 random walk 的随机游走思想被广泛采用,图中的各种属性包括节点中心性和相似度可以通过随机游走策略进行有效表达。这种方法的优势在于可扩展性非常强,这类算法不需要掌握图的整体结构,仅需要了解图的局部信息就可以进行嵌入表达,因此可以适用于动态的大规模图数据,同时由于 random walk 算法的高并行性,算法的训练过程也能通过并行化处理进行加速。

2.2.2 Deepwalk

DeepWalk 方法是基于随机游走算法中最具代表性的一种算法,采用 NLP 领域的经典词嵌入方法 Skip-gram 模型,以随机游走到的节点序列作为语句输入,采用滑动窗口思想得到不同节点之间的相似性度量。具体地说,该方法通过随机游走得到一系列节点序列,基于反向传播,以节点 v_i 为中心最大化 v_i 与其它滑动窗口内的节点共现的概率,也使得 v_i 的前 k 个节点和后 k 个节点的共现概率最大化,其中 $2k+1$ 是随机游走策略设定的长度。该模型通过生成多个长度为 $2k+1$ 的随机游走,并对每个随机游走的对数概率优化,经过 Skip-gram 模型训练得到节点嵌入矩阵,将高维图节点向量压缩为低维稠密向量表示。

2.2.3 node2vec

node2vec 模型与 DeepWalk 方法类似,通过最大化后续节点在固定长度随机漫步中出现的概率,保持节点之间更高的接近度。node2vec 与 DeepWalk 的关键区别在于,node2vec 采用了带有超参数有偏随机漫步,对宽度优先(BFS)和深度优先(DFS)的不同图搜索策略之间提供了一种权衡,因此产生了比 DeepWalk 更高质量、更有信息的嵌入。因此,选择合适的超参数可以使 node2vec 保持网络社区结构和节点之间的结构等效性,从而根据不同问题得到最有效的嵌入方案。

2.3 深度学习

随着对深度学习研究的不断深入,研究发现深度学习自编码器能够对数据中的非线性结构进行建模,因此基于深度神经网络的方法开始被大量应用于图嵌入领域。

2.3.1 深度自学习编码器

Wang et al.提出使用深度学习自编码器来保持网络的一阶和二阶邻近性。通过共同优化这两个邻近特性来实现这一点。该方法利用高度非线性函数来获取嵌入。该模型由无监督模块和监督模块组成。无监督部分通过自动编码器自动寻找节点的嵌入,从而重构节点的邻域信息。后者是基于拉普拉斯特征映射方法,通过设置一个损失函数降低相似节点在拉普拉斯映射空间内的距离以此训练获得图嵌入的非线性函数的映射方式。

2.3.2 GCN

GCN 模型可以说是开创了图数据结构上实现卷积算法的第一个深度学习模型。图卷积网络

(GCNs)通过谱域变换在图上定义一个卷积算子,该模型对节点的邻域嵌入情况进行迭代聚合,并将得到的嵌入情况与前一次迭代的嵌入情况作为函数,得到新的嵌入情况。因为方法只对局部邻域进行聚合嵌入所以该方法具有可扩展性,每次迭代使感受野的区域增大获得节点的多阶邻域信息,可以通过多次迭代可以使学习到的节点嵌入表征图的全局信息。

3 研究进展

本部分结合近年来学术界提出的图嵌入算法,具体对三种图嵌入模型算法进行展开介绍。

3.1 Node2vec

node2vec 模型是近年提出的基于随机游走的图嵌入模型代表,其模型核心是通过定义一种子图采样方案,采样到一系列节点序列,收到 NLP 领域工作的启发,通过将节点序列类比为语句,将节点类比为单词采用词嵌入经典模型 Skip-gram 训练得到节点嵌入矩阵,得到节点的稠密低维向量表示。

首先 node2vec 的子图采样方案采用 DeepWalk 的思路,随机游走的思想是每次从图中选取一个节点,以该节点作为中心节点,每次以一定概率随机选取中心节点的邻接节点,迭代上述的过程直到节点序列到达规定长度截止。相比于 DeepWalk 不同的是,作者在 Deepwalk 的基础上提出了自己的启发式的方法 2nd-order random walks,其游走方法如图 2 所示。该方法具体而言定义了 Random Walks 以及两个超参数。记开始节点为 $c_0 = u$,随机游走选择下一个节点的公式为:

$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{if } (v, x) \in E \\ 0 & \text{otherwise} \end{cases}$$

即若图 E 存在边 (v, x) ,则以一定概率选择下一个节点 x ,其中 π_{vx} 是非正则化的从 v 到 x 的转移概率, Z 是正则化常数。2nd-order random walks 对概率进行进一步改进,具体的有:

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases}$$

当算法从上一个节点 t 随机游走到节点 v ,此时随机游走原路返回的概率由超参数 p 控制;当下一个节点与节点 t 的最短距离为一跳时,游走到该节点的概率由常数控制;当下一个节点与节点 t 的最短距离为两跳时,游走到该节点概率由超参数 q 控制。根据这样的超参数灵活控制,当超参数 q 设置的相对较小时,算法更倾向于游走到更远的节点上,这种游走方式可以对应 DFS 游走策略;当超参数 p 设置的相对较小时,算法更倾向于游走到更近的节点上,这种游走方式可以对应 BFS 游走策略。

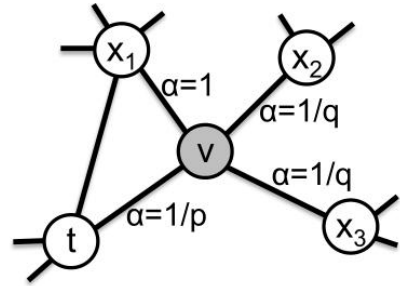


图 2 node2vec 游走策略

获得了节点序列之后 node2vec 模型采用 Skip-gram 模型将高维节点向量压缩为低维向量表示, Skip-gram 模型的原理是通过最大化节点 u 与其邻域节点共现的概率来保留节点领域信息的模型。Skip-gram 模型如图 3 所示采用简单的三层网络模型,输入为高维节点向量表示,在输入层和隐藏层之间以及隐藏层和输出层之间分别引入两个节点嵌入矩阵,采用如下损失函数进行训练:

$$\max_f \sum_{u \in V} \left[-\log Z_u + \sum_{n_i \in N_S(u)} f(n_i) \cdot f(u) \right]$$

由于该模型训练得到嵌入矩阵即为最后的节点嵌入矩阵,通过实验表明此随机游走的算法相较于矩阵分解方法以及其他随机游走算法在节点分类任务和链接预测任务上拥有更好的性能,这种随机游走策略能够适应动态变化的图,即具备可扩展性,但同时 node2vec 的游走开销也非常大,随着图规模的增大和游走长度的增加, node2vec 的性能会

急剧下降。

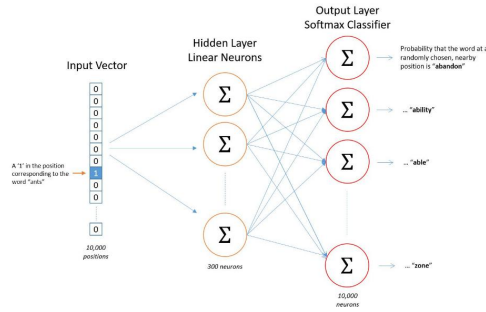


图 3 Skip-gram 模型示意图

3.2 GCN

作为深度学习领域图嵌入方法的代表作，Thomas N. Kipf 提出了 GCN 模型，首次定义了图数据结构上的卷积操作。传统卷积神经网络 (CNN) 的输入是图片等具有欧几里得结构的图结构，卷积神经网络采用局部感知区域、共享权值和空间域上的降采样，相对于位移、缩放和扭曲，具有稳定不变的特性，能够很好的提取图像的空间特征。而 GCN 模型定义的图结构不具备图片的平移不变性，因此传统的卷积方式不适用于图结构。图中每个节点的邻域节点数目不一致，无法用同样尺寸的卷积核进行提取特征。而 GCN 的本质就是定义一种卷积方式来提取图的结构特征，其关键就在于如何定义局部接受域(receptive field)。对于定义局部接受域的方法大体有以下两种思路，第一种是直接的空间域(spatial domain)上定义，第二种是通过傅里叶变化转换到谱域上进行定义。

GCN 网络模型的向前传播规则如公式所示，其中 σ 定义为非线性激活函数，在具体的实验中 σ 设定为 ReLu 函数及 softmax 函数， \tilde{A} 定义为接矩阵 $A + I_N$ ， $H^{(0)}$ 定义为每一层的节点特征向量组成的特征矩阵， $W^{(0)}$ 是可训练的参数矩阵，为卷积神经网络每层共享。

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right)$$

对于这样的向前传播模型，作者给出解释：对于谱域卷积方式，给定向量若 $x \in \mathbb{R}^n$ ， x 的傅里叶变换定义为： $\hat{x} = U^T x$ ； $x = U \hat{x}$ ，故对于传统的傅里叶变换：

$$f * g = F^{-1} \{ F\{f\} \cdot F\{g\} \}$$

将其类比到图数据结构上的傅里叶变换，我们可以在概念上将图卷积方式将其定义为：

$$x * g = U \left((U^T x) \odot (U^T g) \right)$$

其中 U 定义为图标准化的拉普拉斯矩阵 $L = I_N - D^{-1/2} A D^{-1/2} = U \Lambda U^T$ 的特征变换矩阵，设 $U^T x = [\theta_0, \theta_1, \dots, \theta_{n-1}]$ ，其中 $\theta_0, \theta_1, \dots, \theta_{n-1}$ 为自由变量为全局所共享，设 $g_\theta = \text{diag}([\theta_0, \theta_1, \dots, \theta_{n-1}])$ ，可以得到 $x *_{\theta} y = U g_\theta U^T x$ 为图卷积的公式表达形式， g_θ 为定义的卷积算子。将 $g_\theta = \text{diag}([\theta_0, \theta_1, \dots, \theta_{n-1}])$ 多项式近似为 $g_\theta(\Lambda) = \sum \beta_k \Lambda^k$ 其中 Λ 为图矩阵特征值对角阵，文中更具体的是用切比雪夫多项式来保留 K 项，同时对特征值对角阵作平滑处理，因此得到 $g_\theta *_{\theta} x = U g_\theta U^T x \approx \sum \theta'_k T_k(L) x$ ，利用节点 K -hop 邻域，且不需要对拉普拉斯矩阵进行特征值分解，作者进一步地限制切比雪夫多项式中 $K=1$ ，即卷积公式关于 L 是一个线性的变换，更具体化的解释是，作者极大简化原来的卷积推导公式，限制 $K=1$ 即卷积算子只考虑给定节点的一阶邻域信息，这样一来公式变化为如下形式：

$$g_\theta *_{\theta} x \approx \theta'_0 x - \theta'_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x$$

而在上述参数设计中，有两个自由变量 θ'_0, θ'_1 为滤波器参数变量，并且将被整个图所共享。为了进一步限制参数地数量解决过拟合和最小化每层操作涉及的数量大小，令 $\theta'_0 = -\theta'_1 = \theta$ ，得到：

$$g_\theta *_{\theta} x \approx \theta \left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) x$$

同时我们发现 $I_N + D^{-1/2} A D^{-1/2}$ 矩阵的特征值范围在 $[0, 2]$ 之间，反复使用该算子可能会产生梯度爆炸的现象，进一步将矩阵变形使得特征值在 $[-1, 1]$ 之间：令 $\tilde{A} = A + I_N$ ； $\tilde{D} = D + I_N$ ，因此最终得到 GCN 模型向前传播的公式为 $Z = (D^{-1/2} \tilde{A} D^{-1/2} X \Theta)$

根据以上推导，基于图卷积的神经网络模型可以通过将之前定义的单层卷积层进行多个叠加，每一层之后采用非线性激活函数处理，形式化的 GCN 模型示意图如下所示：

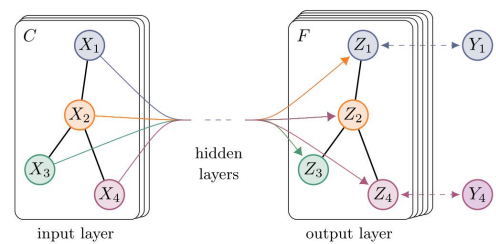


图 4 图卷积神经网络模型

在实验过程中采用两层卷积模型, 第一层隐藏层采用 ReLu 函数, 最后输出层采用 softmax 多分类模型进行处理对输入图网络进行节点分类任务, 对比了基于随机游走的 DeepWalk 算法, 迭代分类算法 ICA, 半监督图嵌入算法 SemiEmb, 在 Citeseer、Cora Pubmed、NELL 数据集上分别进行测试。在最后的节点分类准确率上相较于随机游走算法和迭代分类算法有极大提升, 同时 GCN 模型的训练速度相对于其他模型有 3 倍以上的提升, 这说明图卷积神经网络模型的设计既能够有效捕捉节点领域信息, 同时兼顾其计算效率, 更重要的是 GCN 在图嵌入领域定义了卷积算子, 成功开启了图神经网络研究的新方向。

3.3 GAT

图卷积网络 GCN 可以在图上进行卷积操作, 但是 GCN 存在一些缺陷: 比如依赖拉普拉斯矩阵, 这会使得算法不具备扩展性, 需要预先计算原始图的全局拉普拉斯矩阵, 算法的开销非常巨大; 其次同样因为采用了拉普拉斯矩阵, 因此 GCN 模型不能直接用于有向图; 此外卷积的时候没办法为邻居节点分配不同的权重, 以此来灵活控制方法。此时研究者们发现相比于 GCN 在谱域上定义卷积操作, 在空域(spatial domain)直接进行操作更具有较强的解释性, 因此 2018 年图注意力网络 GAT (Graph Attention Network)被提出, 用于解决以上 GCN 模型未能处理的问题。

GAT 模型引入注意力机制引入 masked self-attentional layers 来改进前面图卷积 graph convolution 的缺点, 对不同的相邻节点分配相应的权重, 既不需要矩阵运算, 也不需要事先知道图结构。同时相较于 GCN 的推导过程 GAT 的理论推导非常简单容易理解。首先问题定义为输入 $\mathbf{h} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}, \vec{h}_i \in \mathbb{R}^F$, 其中 N 为节点个数, F 为特征的维数, 表示输入为 N 个节点, 每个节点有 F 维的特征。输出定义为 $\mathbf{h}' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_N\}, \vec{h}'_i \in \mathbb{R}^{F'}$, 表示输出 N 个节点的 F' 维特征嵌入。

同时为了得到相应的输入与输出的转换, 还需要根据输入的特征至少一次线性变换得到输出的特征, 所以我们需要对所有节点训练一个权值矩阵: $\mathbf{W} \in \mathbb{R}^{F' \times F}$, 这个权值矩阵就是输入与输出的 F 个特征与输出的 F' 个特征之间的关系。随后针对每个节点实行自注意力 self-attention 的注意力机制,

具体机制为 $\alpha: \mathbb{R}^{F'} \times \mathbb{R}^{F'} \rightarrow \mathbb{R}$ 注意力互相关系数 attention coefficients 为: $e_{ij} = a(\mathbf{W}\vec{h}_i, \mathbf{W}\vec{h}_j)$ 这个公式表示的节点 j 对于节点 i 的重要性, 而不去考虑图的全局结构性的信息, 向量 h 就是特征向量下标 i, j 表示第 i 个节点和第 j 个节点。

作者通过 masked attention 将这个注意力机制引入图结构之中, masked attention 的含义是只计算节点 i 的相邻的节点 j 。节点 $j \in \mathcal{N}_i$, 其中 \mathcal{N}_i 为节点 i 的所有相邻节点。为了使得互相关系数更加容易计算和便于比较, 作者引入了 softmax 对所有节点的相邻节点进行正则化处理:

$$j \in \alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}$$

其中注意力机制 α 是一个单层前馈神经网络, 通过权值向量来确定 $\vec{a} \in \mathbb{R}^{2F'}$, 并且加入 LeakyReLU 的非线性激活函数, 因此注意力机制的表示如下, 形象化的注意力机制如图 5 所示:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{a}^T [\mathbf{W}\vec{h}_i \parallel \mathbf{W}\vec{h}_j]\right)\right)}{\sum_{k \in \mathcal{N}_i} \exp\left(\text{LeakyReLU}\left(\vec{a}^T [\mathbf{W}\vec{h}_i \parallel \mathbf{W}\vec{h}_k]\right)\right)}$$

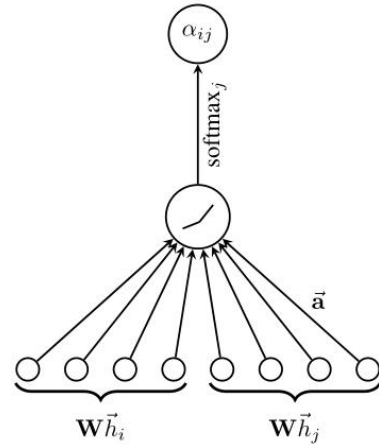


图 5 自注意力机制

得到归一化的注意系数后, 可以利用归一化的注意系数计算对应的特征的线性组合, 作为每个节点的最终输出特征。具体来说, 通过运算得到了正则化后的不同节点之间的注意力互相关系数 normalized attention coefficients, 可以用来预测每个节点的输出特征:

$$\vec{h}'_i = \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \mathbf{W}\vec{h}_j\right)$$

为了稳定自我注意的学习过程,作者发现扩展自注意力机制而使用多头注意力效果非常好。其中, K 个独立注意机制对上式进行变换,随后将其特征串接,得到如下输出特征表示:

$$\vec{h}_i' = \parallel_{k=1}^K \sigma \left(\sum_{j \in N_i} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j \right)$$

其中, \parallel 表示对向量进行直接拼接, α_{ij}^k 为第 k 个注意机制计算得到的归一化注意系数, \mathbf{W}_k 为相应的输入线性变换的权值矩阵。特别的,如果模型对网络的输出(预测)层采用多头注意,采用直接拼接的方法就不再有意义了——相反作者采用平均化处理,在最终输出层采用非线性分类函数(通常用于分类问题的 softmax 或 logistic sigmoid),得到最终的表达式:

$$\vec{h}_i' = \sigma \left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j \right)$$

相较于 GCN 模型, GAT 针对不同相邻节点的重要性进行预测,模型具有更好的性能并且对于扰动具有更好的鲁棒性,在引入注意力机制之后,只与相邻节点有关,即共享边的节点有关,无需得到整张图信息。即使丢失了 i, j 之间的链接,则不计算 α_{ij} 即将模型运用于非监督学习并且因为是在空域上直接进行定义的操作, GAT 也具备更好的解释性,即使模型输入无法反映全局信息也可以运行训练过程。

4 总结与展望

本文综述了三大类图嵌入技术:基于矩阵分解、基于随机游走和基于深度学习的方法。文中概括了图嵌入技术所面临的挑战,同时提供了各自嵌入的应用以及它们各自的评价指标。实验中采用几个公开可用的真实网络对这些算法模型进行了实证评估,并比较了它们的优缺点。本文中所介绍的随机游走模型具有非常强大的可扩展性,但问题是这种方法受游走策略的影响,很难找到一个适用于问题的最优超参数。非线性模型(如基于深度学习的模型)在捕捉图形的内在动态方面表现出了很大的潜力,它们具有近似任意函数的能力,这可以使图以低维形式表示。但这种方法的一个缺点是可解释性有限并且方法的开销不好控制。最后我认为在图嵌入领域有三个很有前途的研究方向:(1)探

索非线性模型(2)研究预测图的动态演化(3)生成具有真实世界特征的合成网络,这些方向目前的研究热点比较少,且未来伴随着知识图谱演化以及相关图算法的发展,这些任务的需求将会越来越大。

5 参考文献

- [1]. Kipf, Thomas N, and M. Welling. "Semi-Supervised Classification with Graph Convolutional Networks."
- [2]. Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks. International Conference on Learning Representations (ICLR), 2018
- [3]. Grover A, Leskovec J. node2vec: Scalable Feature Learning for Networks[C]// Acm Sigkdd International Conference on Knowledge Discovery & Data Mining. 2016.