

文章编号:1008-1542(2020)02-0133-15

信息网络表示学习方法综述

鲁军豪,许云峰

(河北科技大学信息科学与工程学院,河北石家庄 050018)

摘要:网络表示学习方法将信息网络表示为低维稠密携带网络节点特征信息的实数向量,应用于下游机器学习任务的输入,随着机器学习与深度学习的发展,网络表示学习拥有强大的建模能力且应用广泛。对网络表示学习方法、应用进行了归纳总结。首先,对当前国内外网络表示学习方法进行梳理归类,分为传统方法、基于网络结构的嵌入、融入属性信息的嵌入,以及基于谱域的图卷积、基于空间的图卷积和图 attention 网络,按类别对各类模型详细阐述,对比模型之间的适用性和方法特点;其次,介绍了网络表示学习的相关应用,包括推荐系统领域、生物医药领域等,整理常用的数据集、开源实现的表示学习模型和强大的图深度学习库供研究者参考调用;最后,对网络表示学习的发展趋势进行了总结与展望。未来可在深层的图神经网络学习、动态和异构网络的表示、网络模型的泛化能力等方面继续开展研究。

关键词:神经网络;网络;表示学习;图神经网络;图卷积;图深度学习库

中图分类号:TP311.13 文献标识码:A doi:10.7535/hbkd.2020yx02004

A survey of information network representation learning

LU Junhao, XU Yunfeng

(School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China)

Abstract: The network representation learning algorithm represents the information network as a low-dimensional dense real vector carrying the characteristic information of network nodes, and is applied to the input of downstream machine learning tasks. With the development of machine learning and deep learning, network representation learning has been widely used due to its powerful modeling capabilities and extensive applications. The network representation learning methods and their application were summarized. Firstly, the current network representation learning methods at home and abroad were categorized into different groups, including traditional methods, network structure-based embedding, embedding with attribute information, and spectral-based Graph Convolutional Networks, spatial-based Graph Convolutional Networks and Graph Attention Net-

收稿日期:2019-12-31;修回日期:2020-03-17;责任编辑:陈书欣

基金项目:中国留学基金委地方合作项目(201808130283);教育部人工智能协同育人项目(201801003011);河北科技大学校立基金(82/1182108);河北科技大学雾霾与大气污染防治科研项目(82/1182169)

第一作者简介:鲁军豪(1995—),男,河南周口人,硕士研究生,主要从事网络表示学习方面的研究。

通讯作者:许云峰副教授。E-mail:hbkd_xyf@hebust.edu.cn

鲁军豪,许云峰.信息网络表示学习方法综述[J].河北科技大学学报,2020,41(2):133-147.

LU Junhao, XU Yunfeng. A survey of information network representation learning[J]. Journal of Hebei University of Science and Technology, 2020, 41(2): 133-147.

works. Then various models were expounded, and the applicability and characteristics of the models were compared. Secondly, the related applications of network representation learning, including recommender system and biomedical field, were introduced. Commonly used data sets are also given, and open source implementations of representation learning models and powerful graph deep learning libraries for reference were organized. Finally, the developing trend of network representation learning was summarized and forecasted. Deep graph neural network, dynamic and heterogeneous network representation, and generalization ability of network model will need to be further studied.

Keywords: computer neural network; network; representation learning; graph neural network; Graph Convolutional Networks; graph deep learning library

网络是表达实体与实体之间联系的一种数据形式,广泛存在于人们的生活中,例如:人们与周边人形成的社交网络;论文作者之间形成的引文网络;生物医药中的蛋白质网络和药物网络;甚至人脸扫描的点云网络等。网络表示学习是衔接网络与原始数据与网络应用任务的桥梁^[1],如图1所示,网络表示学习方法从复杂的信息网络中学习每个实体的特征信息,将其表示为低维稠密的实数向量,以应用于下游的机器学习任务。

网络中蕴含丰富信息,对社会生产中的信息网络进行分析与研究具有非常高的学术与应用价值。例如,在电子商务中,一个基于信息网络的学习系统能够利用用户和产品^[2-5]之间的交互做出准确推荐;在化学研究中,分子被建模为图网络,它们的生物活性需要被识别,以发现药物^[6-7];在社交网络中,对网络进行社区划分与链路预测,可对客户人群进行分组与推荐^[8-9]。神经网络与深度学习在欧几里德数据上取得了很大成功,但信息网络属于不规则的非欧几里德数据,如何进行有效地信息提取成为一个值得研究的课题。最初的网络表示学习方法受到经典降维技术的影响^[10-11],主要集中在矩阵分解方法上^[12],随后受到 Word2vec 技术^[13-14]和 DeepWalk^[15]的启发,涌现出许多利用随机游走产生节点序列并输入给 skip-gram 模型生成节点嵌入的表示方法。随着深度学习的兴起,许多研究者将卷积与 self-attention 机制^[16]引入网络表示中,在网络的谱域与空间域进行端到端的网络计算,取得了优异的任务效果。

目前对包括图神经网络的网络表示学习综述数量有限,BRONSTEIN 等^[17]给出了非欧几里德领域的深度学习方法的概述,包括图和流形,但忽略了几个重要的基于空间的方法。WU 等^[18]给出图神经网络领域的模型介绍,但缺少了对传统方法以及基于随机游走方法的介绍。涂存超等^[1]的综述中仅简单提到谱域的图卷积模型,缺少对基于空间的图卷积以及图 attention 网络的介绍。本文系统并详细地介绍了网络表示学习相关方法模型,将其分为 2 大类,分别为网络节点嵌入方法和图神经网络方法,细分为 6 个类别,分别是传统方法、基于网络结构的嵌入、融入属性信息的嵌入,以及基于谱域的图卷积、基于空间的图卷积和图 attention 网络,并对比各模型的适用性和优缺点,为网络表示学习领域的学者提供全面的综述参考。同时,本文整理了开源实现的表示学习模型和图深度学习库,以及常用的数据集供研究者参考使用。网络表示学习方法类别如图 2 所示。

本文给定了表示学习方法公式中常见的符号定义,对网络表示学习方法进行详细的阐述,整理对比了各模型的适用性与优缺点,给出了基础数据集和开

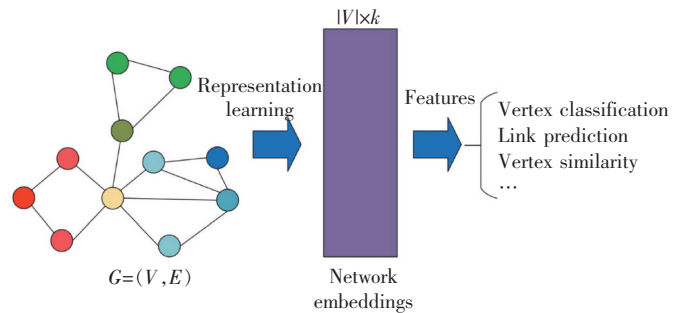


图 1 网络表示学习流程图

Fig.1 Flow chart of network representation learning

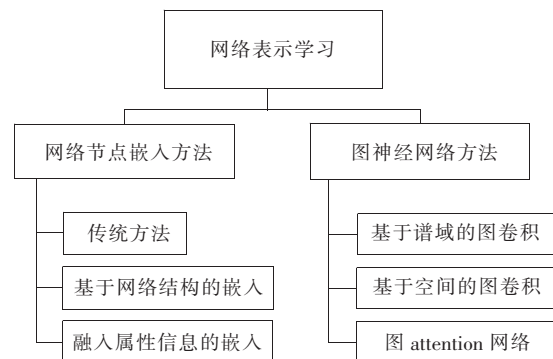


图 2 网络表示学习方法类别

Fig.2 Classification of network representation learning methods

源实现的表示学习模型和图深度学习库,对网络表示学习的应用进行了介绍,并对网络表示学习的未来进行了展望。

1 网络表示学习的基本定义

网络表示学习公式中常见的符号及其含义如表 1 所示。

表 1 常用符号及其含义

Tab.1 Commonly used notations and their meaning

符号	含义	符号	含义
G	网络或图	N, M	节点和边的总数
V	网络中节点的集合	$N_k(i), N(i)$	节点 i 的 k 阶或 1 阶邻居
E	网络中边的集合	$X_1 \odot X_2$	Element-wise product
v_i	节点 i	$\mathbf{H}^{(l)}$	l^{th} 层的特征表示矩阵
e_{ij}	节点 i 与节点 j 连接的边	$\sigma(\cdot)$	sigmoid 激活函数
$\mathbf{A} \in \mathbb{R}^{ V \times V }$	邻接矩阵	$\sigma_h(\cdot)$	Tanh 激活函数
$\mathbf{D} \in \mathbb{R}^{ V \times V }$	度矩阵		

定义 1 无向图^[18]:无向网络或图定义为 $G=(V, E)$ 。其中 V 表示网络中所有节点的集合,任一节点表示为 $v_i, \forall v_i \in V$; E 表示网络中所有边的集合,任一条边表示为 $e_{ij}, e_{ij}=(v_i, v_j) \in E$,即节点 v_i 与 v_j 之间存在一条连接的边。网络的邻接矩阵定义为 $\mathbf{A} \in \mathbb{R}^{|V| \times |V|}$,如果 $e_{ij} \in E$,则 $\mathbf{A}_{ij}=1$,否则 $\mathbf{A}_{ij}=0$,有向图的邻接矩阵是一个对称矩阵。另外,网络的度矩阵 \mathbf{D} 是一个对角矩阵, \mathbf{D}_{ii} 表示第 i 个节点的度且 $\mathbf{D}_{ii}=\sum_j \mathbf{A}_{ij}$,即与第 i 个节点相连的节点的数量。

定义 2 有向图^[18]:节点之间连接的边有指向性的网络或图被称为有向图,定义为 $G=(V, E)$ 。由于节点边的指向性,有向图的邻接矩阵 \mathbf{A}_{ij} 不一定等于 \mathbf{A}_{ji} 。

2 网络表示学习方法

网络表示学习方法将非欧几里德结构的网络节点表示为低维稠密的特征向量,供下游机器学习任务使用,是一项特征工程任务。随着近年来深度学习的兴起,图神经网络成为网络表示学习领域研究的热点。与网络节点的嵌入表示不同,图神经网络多为端到端的训练框架。

2.1 网络节点嵌入方法

2.1.1 传统方法

早期的网络表示学习方法被作为降维技术的一部分,将网络节点嵌入到低维空间的思想是让互相连接的节点在嵌入后的低维向量空间中彼此保持更近的距离。LLE(locally linear embedding^[19])算法认为每一个数据点都可以由其近邻点的线性加权组合得到。该方法先寻找每个样本点的 k 个近邻点,由每个样本点的近邻点计算出该样本点的局部重构权重矩阵,然后用局部权重矩阵和其近邻点计算出该样本点的输出值,LLE 算法的目标函数表示为

$$\phi(\mathbf{Y}) = \frac{1}{2} \sum_i \left| \mathbf{Y}_i - \sum_j \mathbf{W}_{i,j} \mathbf{Y}_j \right|^2. \quad (1)$$

LE(Laplacian eigenmaps)^[11]用边权重 \mathbf{W}_{ij} 衡量节点的相似性,在原始空间越相似的节点,嵌入到低维空间以后也越相似,LE 方法的目标函数如下:

$$\phi(\mathbf{Y}) = \frac{1}{2} \sum_{i,j} \left| \mathbf{Y}_i - \mathbf{Y}_j \right|^2 \mathbf{W}_{i,j} = \text{trace}(\mathbf{Y}^T \mathbf{L} \mathbf{Y}), \quad (2)$$

通过最小化该目标函数,可以得到节点的低维嵌入向量表示,其中 \mathbf{L} 是构建图的拉普拉斯矩阵。

GF(graph factorization)^[20]是第 1 个时间复杂度为 $O(|E|)$ 的网络嵌入算法^[21]。GF 算法对图的邻接矩阵进行因式分解来得到节点的嵌入表示,其目标函数如下:

$$\phi(\mathbf{Y}, \lambda) = \frac{1}{2} \sum_{(i,j) \in E} (\mathbf{W}_{i,j} - \langle \mathbf{Y}_i, \mathbf{Y}_j \rangle)^2 + \frac{\lambda}{2} \sum_i \|\mathbf{Y}_i\|^2, \tag{3}$$

式中 λ 是一个正则化系数。

2.1.2 基于网络结构的嵌入方法

Google 在 2013 年推出一个用于获取词向量的工具包 Word2vec^[13-14]，它的简单高效引起了很多人的关注，同时也给网络表示学习提供了很好的思路。Word2vec 是根据大量语料库中词语的共现关系来得出每个单词的向量嵌入，有 CBOW 和 skip-gram 两种模型。前者根据上下文预测中心词，后者是根据中心词预测上下文，这里只介绍 skip-gram 模型，如图 3 所示。

用核心词 w 与其上下文 $\text{context}(W)$ 组成训练样本，skip-gram 模型用核心词 w 预测上下文 $\text{context}(W)$ ，通过大量语料库的词语共现关系来不断更新词语的向量表示，用负采样来加快训练速度，最终得到每个词语的向量表示。在 Word2vec 被提出之后，研究者在其基础上提出了大量基于随机游走的网络表示学习方法，如 DeepWalk^[15]，Node2vec^[22] 等，这些方法利用随机游走产生的节点序列作为 skip-gram 的输入，从而生成节点的低维向量表示。

PEROZZI 等^[15]观测到节点在短随机游走中的分布和词语在自然语言中的分布都满足幂律分布，并在游走的过程中获取了节点的局部结构信息，从而将 Word2vec 模型引入网络表示学习，提出 DeepWalk 算法，利用节点截断随机游走产生的类似语料中句子的序列，借助词向量模型生成节点的嵌入向量，使得效果有了较大的提升。DeepWalk 对 Karate 数据集的嵌入效果如图 4 所示。

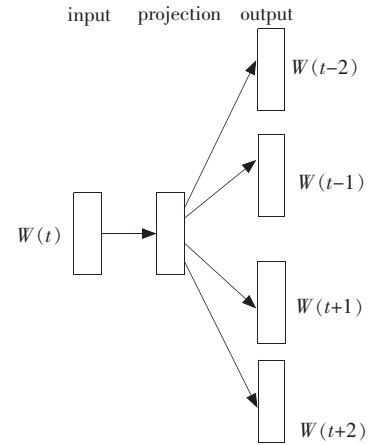


图 3 窗口大小为 2 的 skip-gram 模型^[13]

Fig.3 Skip-gram model with window size 2^[13]

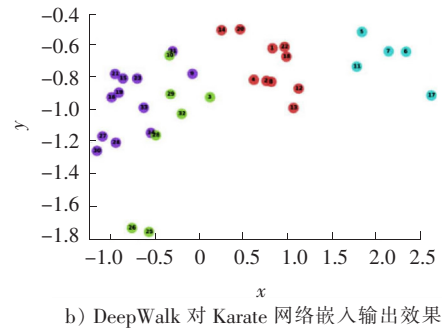
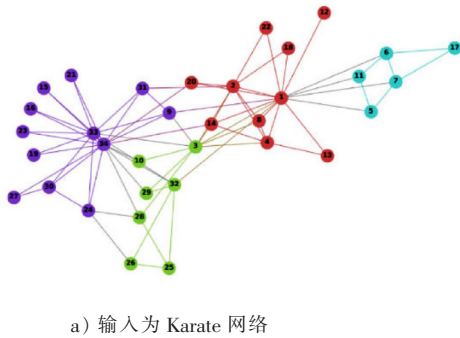


图 4 DeepWalk 对 Karate 数据集的嵌入效果

Fig.4 Embedding effect of Karate dataset by using the DeepWalk

GROVER 等^[22]将 BFS 与 DFS 融入节点的随机游走，提出了 Node2vec 算法。相比于 DeepWalk 算法中节点的随机游走，加入 BFS 和 DFS 策略的 Node2vec 可以更好地挖掘网络中的拓扑信息，如图 5 所示。

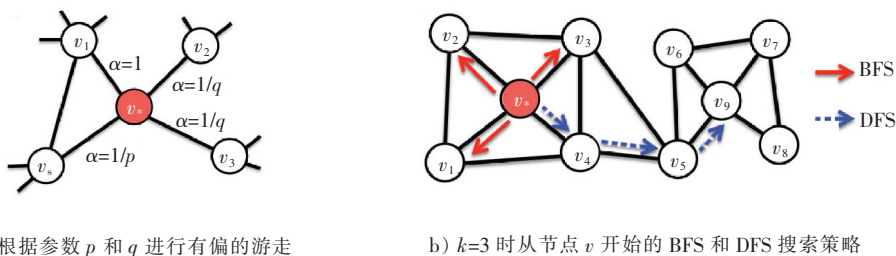


图 5 Node2vec 算法

Fig.5 Node2vec algorithm

图 5 a)是 Node2vec 算法控制二阶随机游走的示意图,该算法设置了 2 个超参数 p 和 q 来平衡游走方式, α 是关于当前节点 v_s 搜索下一个节点 v_x 的函数,表达式如下:

$$\alpha(v_s, v_x) = \begin{cases} \frac{1}{p}, & \text{if } d_{v_s v_x} = 0, \\ 1, & \text{if } d_{v_s v_x} = 1, \\ \frac{1}{q}, & \text{if } d_{v_s v_x} = 2. \end{cases} \quad (4)$$

从当前节点 v 游走到下一个节点 v_x 的概率就是其归一化形式:

$$p(v_x | v) = \frac{\alpha(v_s, v_x)}{\sum_{(y,v) \in E} \alpha(v_s, y)}, \quad (5)$$

从式(5)可以看出,参数 p 可以调节游走回到上一个节点的概率,参数 q 可调节游走偏向于 BFS 还是 DFS,当 q 较小时,游走倾向于 DFS,当 q 较大时,游走倾向于 BFS。特别地,当 p, q 参数都为 1 时,等价于随机游走。

需要注意的是在嵌入向量的更新计算中,DeepWalk 采用分层 Softmax 来计算归一化因子,使用二叉树结构加速计算,而 Node2vec 算法则采用负采样方法,每个训练样本只更新部分模型权重,从而提高嵌入的训练速度。

DeepWalk^[15] 和 Node2vec^[22] 利用随机或有偏随机游走来获得节点的邻域信息,从而生成嵌入向量,使在网络结构中相连和距离较近的节点在对应的低维空间中也具有相近的距离。但是在网络中存在结构一致性(structural identify)节点,如图 6 所示,顶点 u 与顶点 v 的度数分别是 5 和 4,分别连接 3 个和 2 个三角形网络结构,并通过 2 个顶点($d, e; x, w$)与外界相连,这样的节点虽然没有直接相连也没有共同邻域,但具有空间结构一致性。RIBERIO 等^[23] 观察到 DeepWalk 和 Node2vec 等算法构造的节点序列不能识别相隔较远的具有结构一致性的节点,为了解决此问题,提出 Struc2vec 算法。

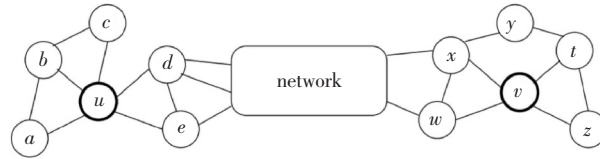


图 6 网络中节点 u 与 v 存在结构一致性^[23]

Fig.6 Two nodes (u and v) that have structural consistency^[23]

Struc2vec 算法的设计思想:如果 2 个节点度相同,则它们的结构相似;若它们的邻居也拥有相同的度,则相似性更高。考虑节点 u ,令 $R_k(u)$ 表示与节点 u 距离为 k 的节点集合,即以 u 为根的 BFS 树上第 k 层的节点集合, $s(S)$ 表示集合 $S \in V$ 的有序度(degree)序列,定义 $f_k(u, v)$ 为

$$f_k(u, v) = f_{k-1}(u, v) + g(s(R_k(u)), s(R_k(v))), \quad (6)$$

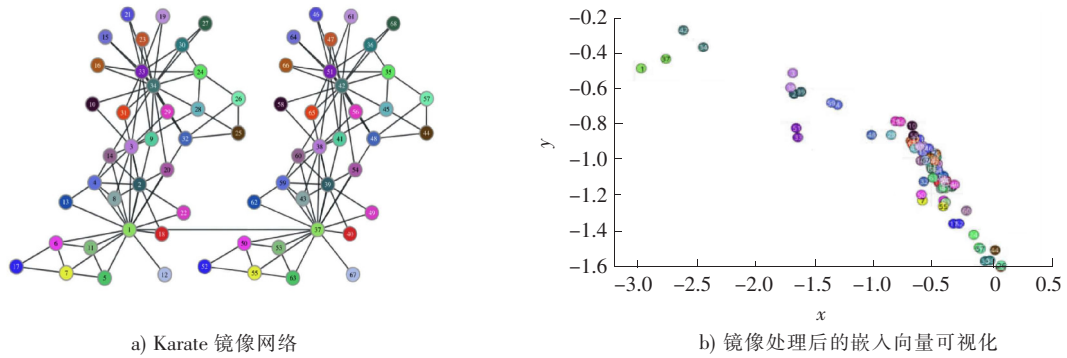
$$k \geq 0 \text{ and } |R_k(u)|, |R_k(v)| > 0,$$

其中 $g(D_1, D_2) \geq 0$ 是衡量有序度序列 D_1 和 D_2 的距离函数,一般为动态时间规整算法(dynamic time warping),并且 $f_{-1} = 0$ 。利用 $f_k(u, v)$ 构造分层图,同层之间节点的权值为 $w_k(u, v) = e^{-f_k(u, v)}$,相邻层对应的权值公式为

$$\begin{cases} w(u_k, u_{k+1}) = \log(\Gamma_k(u) + e), \\ \Gamma_k(u) = \sum_{v \in V} 1(w_k(u, v) > \bar{w}_k), \\ w(u_k, u_{k-1}) = 1, \end{cases} \quad (7)$$

之后通过随机游走构建序列,利用 skip-gram 模型得到节点的向量嵌入。如图 7 所示,在对 Karate 数据^[24] 进行镜像处理后的向量可视化对比中可以发现,镜像节点在空间中聚集在一起,相比于 DeepWalk 与 Node2vec 算法,Struc2vec 算法能准确捕捉拥有结构一致性的节点。

Struc2vec 应用于无权无向图(带权图的权重不会被用到,有向图会当作无向图处理),主要关注的是图中顶点的空间结构相似性,蚂蚁金服的风控模型在应用了 Struc2vec 后与 Node2vec 相比有质的提升。



注:为了更好地说明结构相似性,笔者将 Karate 原网络复制一份,使得每一个节点都有一个与之结构相同的镜像节点,并在两网络之间添加了一条连接的边。

图 7 Struc2vec 算法对镜像网络嵌入向量的可视化

Fig.7 Visualization of embeddings of mirror network by Struc2vec algorithm

TANG 等^[25]提出了 LINE 算法,使得在小时范围内单机学习百万级顶点网络表示成为了可能。LINE 算法提出了一阶相似度与二阶相似度,一阶相似度描述了直接相连的节点之间的相似性,如图 8 中节点 6 和节点 7 直接相连,两节点相似度为边的权重;二阶相似度描述为一对节点之间的接近程度和邻居网络结构之间的相似性,如图 8 中节点 5 与节点 6 之间没有直接的边关系,但有共同的网络邻居。

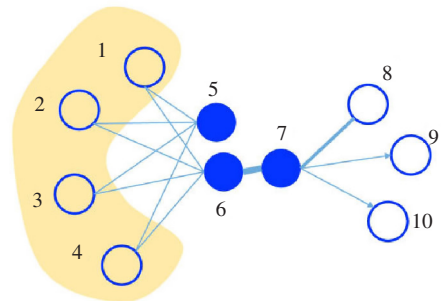


图 8 信息网络图^[25]

Fig.8 A toy example of information network^[25]

一个网络中的边关系往往是非常稀疏的,所以有必要进一步刻画二阶相似度关系来考虑虽然并不直接相连但是共同邻居较多的节点对,从而对第一阶相似度的信息予以补充^[1]。LINE 算法适用于大规模网络,并且网络的边不限制是否有向和是否带权,在节点分类等任务中表现出不错的效果。

GraRep 算法^[26]在 LINE 定义的一阶和二阶相似性的启发下,将这种相似性推广到更高阶,定义了 k 阶相似性。GraRep 中也使用了像 LINE 中一样将高阶的图中的相似关系在低维向量空间中用条件概率表示,并且使用了负采样优化的策略。

2.1.3 融入属性信息的嵌入方法

上述方法模型只利用了网络中拓扑结构的相似性来生成节点嵌入向量,真实世界的网络通常在节点和边上附属有标签文本等丰富的属性信息,若能充分利用网络中的属性信息,会得到更好的嵌入效果。

YANG 等^[27]证明了 DeepWalk 算法其实质上是分解矩阵 M 和计算出 M 的闭式解,并提出了一种融入文本信息的网络表示方法 TADW。TADW 方法的思想是在矩阵分解过程中,加入对文本特征信息的考虑,如图 9 所示。其中 $W \in R^{k \times |V|}$, $H \in R^{k \times f_t}$ 为参数矩阵, $T \in R^{f_t \times |V|}$ 为文本特征矩阵,加入文本信息矩阵后算法的损失函数如式(8)所示,TADW 算法使用共轭梯度下降法迭代更新 W 矩阵和 H 矩阵求解参数。

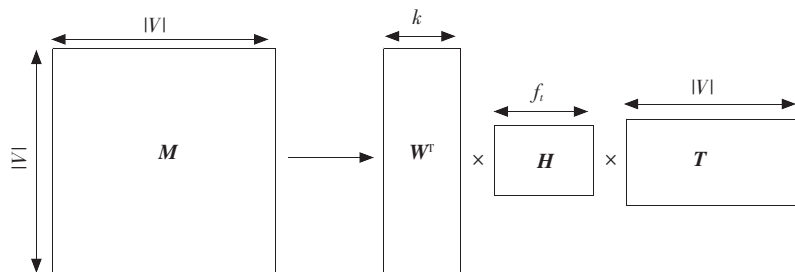


图 9 TADW 模型框架示意图^[27]

Fig.9 Framework of TADW^[27]

$$\min_{\mathbf{W}, \mathbf{H}} \|\mathbf{M} - \mathbf{W}^T \mathbf{H} \mathbf{T}\|_F^2 + \frac{\lambda}{2} (\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2). \quad (8)$$

TU 等^[28]提出的 CANE 算法在兼顾网络结构相似性的同时,利用 attention 机制自适应计算节点文本内容的相似性。图 10 为 CANE 方法的框架示意图,该方法利用卷积神经网络对一条边上 2 个节点的文本信息进行编码。在文本表示生成的过程中,利用相互注意力机制,选取 2 个节点彼此最相关的卷积结果构成最后的文本表示向量。

文献[29]中将文本也转化为特殊的节点,形成两种连接的边,即节点-节点和节点-文档,对两种边一起建模嵌入得到节点表示。针对属性网络的表示学习方法有 ANRL^[30], SIGNet^[31], TransNet^[32]和 SANE^[33],它们结合网络结构和节点的属性信息使学习到的嵌入表示具有更多的信息性。BiasedWalk^[34]则考虑兼顾有偏随机游走的优势以及网络节点的属性信息,以学习到信息更加丰富的网络嵌入表示。

当信息网络中的节点或边类型数量 $n \geq 2$ 时,这种网络被称为异构网络。针对异构网络的表示学习方法也有大量研究者做出很多探索,典型的方法有 metapath2vec^[35], HIN2Vec^[36], Tri-DNR^[37], HNE^[38], HEBE^[39], EOE^[40],以及基于 Attention^[41]机制的 HAN^[42]等。

2.2 图神经网络方法

2.2.1 基于谱域的图卷积

BRUNA 等^[43]首先提出了在图网络中基于谱域的卷积方法,函数卷积的傅里叶变换等于函数傅里叶变换的乘积,表达式如下:

$$F(f * h) = \hat{f}(\omega) \cdot \hat{h}(\omega), \quad (9)$$

式中 $*$ 表示卷积,这样在傅里叶域做乘积,然后做傅里叶逆变换,即等于在原函数空间直接做卷积。BRUNA 等^[43]的谱卷积方法具体定义如下:

$$x_{k+1,j} = h\left(\mathbf{V} \sum_{i=1}^{f_{k-1}} \mathbf{F}_{k,i,j} \mathbf{V}^T x_{k,i}\right), \quad j = 1, 2, \dots, f_k, \quad (10)$$

式中: x_k 为输入图信号; \mathbf{F} 为包含可学习参数的对角矩阵; h 为非线性的激活函数。

该方法存在 3 个明显的缺点:1)每次卷积都要进行矩阵相乘使得计算复杂度为 $O(N^3)$,并且需要特征分解,计算量非常大,不适合用于大型图;2)每次卷积都要考虑所有的节点,没有空间局部性;3)参数个数为 $O(N)$ 。

针对上述方法的不足,BRUNA 等^[43]在 DEFFERRARD 等^[7]提出 ChebNet 方法的基础上进一步完善,概括为 $y = \sigma(\mathbf{U} g_\theta(\mathbf{A}) \mathbf{U}^T x)$,其中 \mathbf{A} 是特征值组成的对角矩阵,在 ChebNet 方法中, $g_\theta(\mathbf{A})$ 被定义为

$$g_\theta(\mathbf{A}) = \sum_{k=0}^{K-1} \theta_k \mathbf{A}^k, \quad (11)$$

代入式(11)中并将 \mathbf{L}^k 使用切比雪夫展开得到:

$$y = \sigma\left(\sum_{k=0}^{K-1} \theta_k \mathbf{L}^k x\right) = \sigma\left(\sum_{k=0}^{K-1} \theta_k \mathbf{T}_k(\tilde{\mathbf{L}})\right), \quad (12)$$

式中, $\tilde{\mathbf{L}} = 2 \frac{\mathbf{L}}{\lambda_{\max}} - \mathbf{I}_N$,这里用到拉普拉斯矩阵的特征分解的性质: $\mathbf{L} = \mathbf{U} \mathbf{A} \mathbf{U}^T$ 。ChebNet 方法将计算的复杂度降低为 $O(K|E|)$,并且只考虑 K 阶邻域,有一定的空间局部性。

KIPF 等^[44]在 ChebNet 方法的基础上,令 $K=1$,每层卷积只考虑直接邻域,并使深度加深,提出分层传播机制的多层图卷积网络 GCN,获得了非常显著的效果。GCN 模型整理的计算过程可以概括为

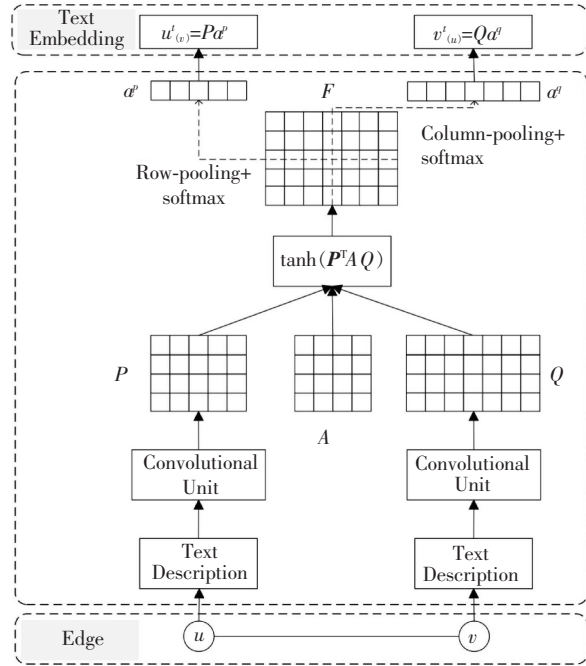


图 10 CANE 算法框架示意图^[28]

Fig.10 An illustration of CANE framework^[28]

$$\mathbf{H}^{(l+1)} = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}^{(l)}), \tag{13}$$

式中： $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, \mathbf{A} 为图的邻接矩阵, \mathbf{I} 为单位阵; $\tilde{\mathbf{D}}$ 为图的度矩阵。

GCN 算法的计算架构如图 11 所示。

2.2.2 基于空间的图卷积

GNNs^[45] 是最早研究图神经网络的著作之一, 采用 AlmeidaPineda 算法训练其模型, 算法将节点的表示定义为

$$\begin{cases} x_n = f_w(l_n, l_{co[n]}, x_{ne[n]}, l_{ne[n]}), \\ o_n = g_w(x_n, l_n), \end{cases} \tag{14}$$

其核心思想是运行传播过程以达到固定点, 然后在给定收敛解的情况下执行反向过程。算法主要分为 2 个步骤: Forward 和 Backward。Forward 具体计算如图 12 所示, f_w 根据当前父节点、相邻子节点、相连边的特征及状态对当前父节点的状态进行更新, 然后迭代 t 次达到

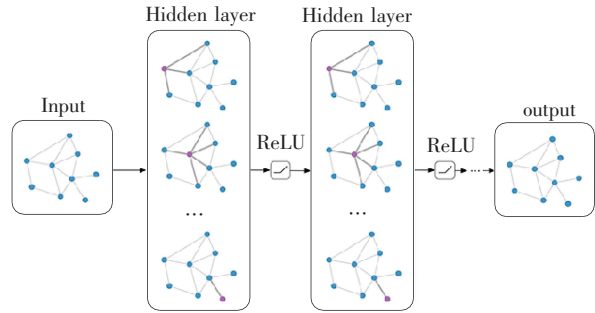


图 11 GCN 算法的计算架构

Fig.11 Framework of computing architectures of GCN algorithm

稳定状态之后, g_w 根据当前节点的状态, 输出节点的 output。Backward 过程的目的在于求解 $\frac{\partial e_w}{\partial w}$, 这里 e_w 是损失函数。跟 Forward 过程类似, 找到 t 时刻使得梯度处于一个稳定状态, 然后求解 t 时刻的梯度 $\frac{\partial e_w}{\partial w}$ 。

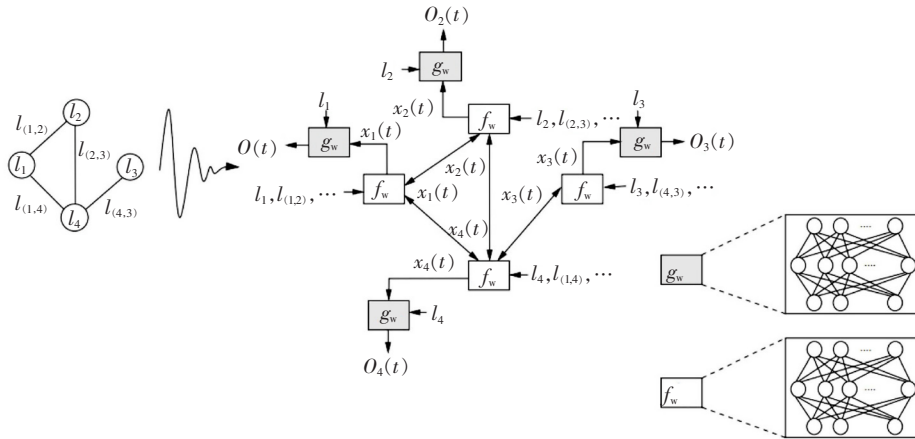


图 12 GNNs 算法 Forward 计算过程^[45]

Fig.12 Forward process of GNNs algorithm^[45]

GNN 统一了一些早期处理图数据的方法, 但也存在缺点: 为了保证公式有唯一解, $f_w(\cdot)$ 必须是压缩映射, 严重限制了模型的能力; 其次, 由于梯度下降步骤之间需要多次迭代, 因此 GNN 的计算开销较大。

GGNNs^[46] 解决了 GNN 模型存在的缺点, 即 f 不再是一个压缩映射, 同时, GGNNs 采用 BPTT(back-propagation through time) 去学习参数使迭代不需要到收敛才能输出, 可以迭代固定步长。GGNNs 主要是解决过深层的图神经网络导致过度平滑的问题, 使用 GRU 更新节点状态, 其状态更新函数定义如下:

$$\mathbf{h}_v^t = \text{GRU}(\mathbf{h}_v^{t-1}, \sum_{u \in N(v)} \mathbf{W} \mathbf{h}_u^t), \tag{15}$$

式中 \mathbf{W} 是可学习参数, 每一种边对应一个 \mathbf{W}_{edge} , 使得 GGNNs 可以处理异构图。该模型的缺点为采用 BPTT 训练牺牲了时间和记忆的效率, 由于 GGNNs 需要在所有节点上多次运行递归函数, 需要将所有节点的中间状态存储在内存中, 因此处理大型图尤其困难。DAI 等^[47] 为了提高学习效率提出了 SSE 算法, 其采用了异步方式随机更新节点的潜在表示, 用采样的批处理数据更新参数, 但算法的稳定性还存在问题。

GraphSAGE^[48]引入聚合函数概念定义图卷积,聚合函数本质上是聚集节点的邻域信息。GraphSAGE是归纳式学习的一个代表,即可以利用节点邻域信息直接学习出新增节点的嵌入表示,该算法对图的卷积定义为

$$\mathbf{h}_v^t = \sigma(\mathbf{W}^t \cdot \text{aggregate}_t(\mathbf{h}_v^{t-1}, \{\mathbf{h}_u^{t-1}, \forall u \in N(v)\})). \quad (16)$$

GraphSAGE不直接学习每个节点的表示,而是学习聚合函数,对于新增的节点利用聚合函数生成它的嵌入表示,邻居采样和信息聚合过程如图 13 所示。

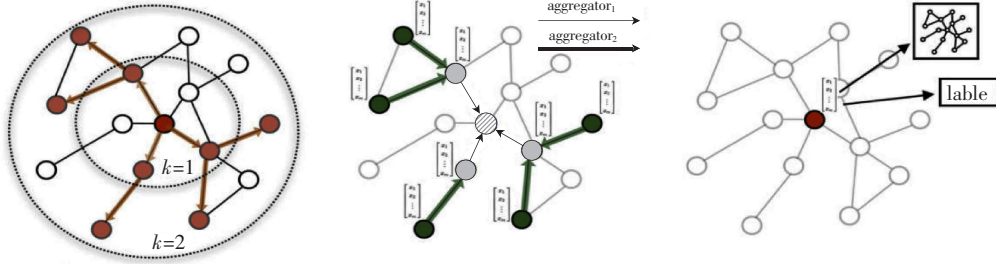


图 13 GraphSAGE 邻居采样和聚合信息方法^[48]

Fig.13 Visual illustration of the GraphSAGE sample and aggregate approach^[48]

该算法分为两步:首先是邻居节点采样,由于每个节点的度都不一样,为了计算高效,给每个节点分层并采样固定节点的邻居;其次是邻居特征聚合,通过聚集采样到的邻居特征,更新当前节点的特征。聚合函数可以采用平均聚合、归纳式聚合或者 Pooling 聚合,也可以自行设计,常用聚合函数如下。

平均聚合:

$$\mathbf{h}_{N(v)}^k = \text{mean}(\{\mathbf{h}_u^{k-1}, u \in N(v)\}), \quad (17)$$

$$\mathbf{h}_v^k = \sigma(\mathbf{W}^k \cdot \text{CONCAT}(\mathbf{h}_v^{k-1}, \mathbf{h}_{N(v)}^{k-1})). \quad (18)$$

归纳式聚合:

$$\mathbf{h}_v^k = \sigma(\mathbf{W}^k \cdot \text{mean}(\{\mathbf{h}_v^{k-1}\} \cup \{\mathbf{h}_u^{k-1}, \forall u \in N(v)\})). \quad (19)$$

Pooling 聚合:

$$\mathbf{h}_{N(v)}^k = \max(\{\sigma(\mathbf{W}_{\text{pool}} \mathbf{h}_{u_i}^k + b)\}, \forall u_i \in N(v)), \quad (20)$$

$$\mathbf{h}_v^k = \sigma(\mathbf{W}^k \cdot \text{CONCAT}(\mathbf{h}_v^{k-1}, \mathbf{h}_{N(v)}^{k-1})). \quad (21)$$

PATCHY-SAN^[49]采用图标记为每个图选择了固定数量的节点,将图的结构转换为网格结构,进而使用标准的卷积神经网络(CNN)来解决图的分类任务。不同于 PATCHY-SAN, LGCN^[50]使用标准的 CNN 生成节点级的输出。该方法对网络中每个节点的邻域进行排序,并取前 K 个组成该节点的一维网格,进而采用标准的 CNN 进行卷积计算。为了能让模型在大规模的图上训练, LGCN 提出了一个子图训练方法来减少过多的内存和计算资源的开销。

2.2.3 图 attention 网络

图 attention 网络本质上同样属于基于空间域的图神经网络,但是由于其较好的模型效果和较高的研究热度,这里单独对其概述。

GAT(graph attention network)^[41]算法的基本思想是,根据每个节点在其邻节点上的 attention 来对节点表示进行更新。具体的,该算法用 $\mathbf{h} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_N\}$ 表示图中的 N 个节点,其中 $\vec{h}_i \in \mathbb{R}^F$, F 是每个节点的特征数,目的是输出一个新的节点特征向量集 $\mathbf{h}' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_N\}$,其中 $\vec{h}'_i \in \mathbb{R}^{F'}$,公式如下:

$$\vec{h}'_i = \sigma\left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W} \vec{h}_j\right), \quad (22)$$

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\vec{a}^T [\mathbf{W} \vec{h}_i \parallel \mathbf{W} \vec{h}_j]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(\vec{a}^T [\mathbf{W} \vec{h}_i \parallel \mathbf{W} \vec{h}_k]))}. \quad (23)$$

上述过程可以用图 14 表示。另外,GAT 使用 multi-head attention 来稳定 self-attention 的学习过程,不同的箭头表示一个独立的 attention 计算,每个邻居节点做 K 次 attention 计算。每次 attention 计算为一个普通的 self-attention,输出结果为 \vec{h}'_i 。multi-head attention 为每个节点 i 输出 K 个不同的 \vec{h}'_i ,然后将向量合并,得到最终的 \vec{h}'_i ,两种合并规则如下:

$$\vec{h}'_i = \parallel_{k=1}^K \sigma \left(\sum_{j \in N_i} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j \right), \tag{24}$$

$$\vec{h}'_i = \sigma \left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^k \mathbf{W}^k \vec{h}_j \right). \tag{25}$$

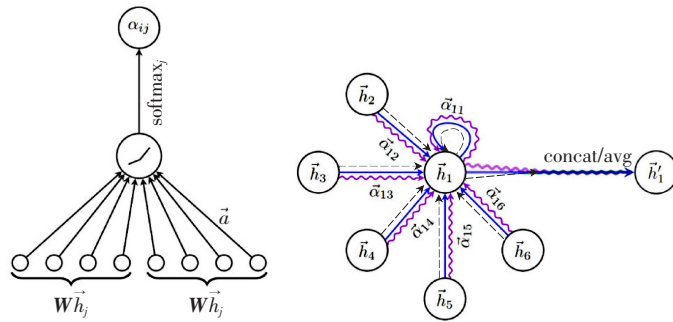


图 14 GAT 的 attention 机制与 multi-head attention($K=3$) 机制^[41]

Fig.14 Illustration of GAT attention mechanism and multi-head attention (with $K=3$ heads)^[41]

相比于其他图模型,GAT 无需使用特征值分解等复杂的矩阵运算,相比于 GCN,每个节点的重要性可以是不同的,因此,GAT 具有更强的表示能力。GAT 具有以下几个特点:1)计算速度快,可以在不同的节点上进行并行计算;2)可以同时拥有不同度的节点进行处理;3)可以被直接用于解决归纳学习问题,即可以对从未见过的图结构进行处理。

ZHANG 等^[51]提出的门控注意力网络 GAAN 同样采用了多头注意力机制来更新节点的隐藏状态,不同于 GAT 的是该算法赋予了每个 head 不同的权重,使得重要的部分能得到更大的计算权重,具体定义如下:

$$\mathbf{h}_i^t = \mathcal{O}_o \left(x_i \oplus \parallel_{k=1}^K g_i^k \sum_{j \in N_i} \alpha_k (\mathbf{h}_i^{t-1}, \mathbf{h}_j^{t-1}) \mathcal{O}_v (\mathbf{h}_j^{t-1}) \right), \tag{26}$$

式中: g_i^k 是第 K 个注意力 head 的注意力权重; $\mathcal{O}_o(\cdot)$ 和 $\mathcal{O}_v(\cdot)$ 是反馈神经网络。

另外,图 attention 模型 GAM^[52] 用一个循环神经网络模型加 attention 机制来解决图分类问题;Attention Walk^[53] 用随机游走的方式学习节点嵌入表示,不同于 DeepWalk,Attention Walk 用可微的 attention 权重来对共现矩阵进行因式分解。

3 应用

网络表示学习在社会生产中有非常广泛的应用。本文给出了网络表示学习模型适用性与特点的比较,以及研究中可能会用到的数据集及其分类统计信息,整理了部分模型的实现链接和已开源的图深度学习库,方便研究者快速复现验证并掌握模型,讨论了网络表示学习的应用方向和实例。

3.1 模型对比与数据集及开源实现

表 2 对比了网络表示学习模型的适用性和方法特点,包括模型是否支持带权图、有向图以及属性图,并给出了模型的时间复杂度和模型特点,可供研究者参考使用。表 3 给出了众多测试算法性能的数据集,数据集包括 3 类,分别是引文网络、社交网络和化学生物网络。对应于每个数据集,表中给出了数据集来源,分别统计了该数据集的子图数、节点数、边数、特征数和标签类别数量,方便研究者选择适合模型的数据集。

表 2 模型适用性与特点比较

Tab.2 Comparison of model applicability and characteristics

方法	支持带权图	支持有向图	支持属性信息	算法复杂度	算法特点
DeepWalk	0	0	0	$O(V d)$	将 skip-gram 应用到图中,为网络结构的嵌入提供了思路
Node2vec	1	0	0	$O(V d)$	加入超参数 p 和 q ,有偏随机游走
Struc2vec	1	0	0	$O(V ^3)$	侧重节点的结构信息嵌入
LINE	1	1	0	$O(E d)$	计算节点间二阶相似度,可应用到大规模网络
GraRep	1	1	0	$O(V ^3)$	学习网络的全局特征,优化部分使用 SVD
TADW	1	1	1	$O(V E)$	融入节点文本特征但复杂度高
CANE	1	0	1	—	融入节点属性信息与 attention 机制
GCN	1	0	0	$O(E d^2)$	只考虑一阶降低复杂度,通过堆叠多层获更大感受野,表示能力强,但需要加载整个静态图
GraphSAGE	0	0	0	—	归纳式学习
GAT	0	1	1	$O(V FF'+ E F')$	归纳式学习,利用 attention 具有更强表示能力

注:1 表示支持;0 表示不支持。

表 3 常用数据集整理(修改自文献[18])

Tab.3 Summary of commonly used datasets(modified from reference[18])

Category	Dataset	Source	# Graphs	# Nodes	# Edges	# Features	# Labels
Citation Networks	Cora	文献[54]	1	2 708	5 429	1 433	7
	Citeseet	文献[54]	1	3 327	4 732	3 703	6
	Pubmed	文献[54]	1	19 717	44 338	500	3
	DBLP	dblp.uni-trier.de[111] (aminer.org/citation)	1	—	—	—	—
Social Networks	BlogCatalog	文献[55]	1	10 312	333 983	—	39
	Reddit	文献[48]	1	232 965	11 606 919	602	41
	Epinions	www.epinions.cpm	1	—	—	—	—
Chemical/Biological Graphs	PPI	文献[56]	24	56 944	818 716	50	121
	NCI-1	文献[57]	4 110	—	—	37	2
	NCI-109	文献[57]	4 127	—	—	38	2
	MUTAG	文献[58]	188	—	—	7	2
	D&D	文献[59]	1 178	—	—	—	2
	QM9	文献[60]	133 885	—	—	—	13
	tox21	tripod.nih.gov/tox21/challenge/	12 707	—	—	—	12
Unstructured Graphs	MNIST	yann.lecun.com/exdb/mnist/	70 000	—	—	—	10
	Wikipedia	www.mattmahoney.net/dc/textdata	1	4 777	184 812	—	40

开源实现的模型总结如表 4 所示,图深度学习开源库总结如表 5 所示。在表 4 与表 5 中整理出了一些具有代表性模型的开源实现,以及强大的图深度学习库,可供研究者快速学习或复现验证模型效果。其中在表 5 中列出的 Euler, PGL, Plato 图深度学习库支持分布式计算,使得更大规模的网络计算成为可能。

表4 开源实现的模型总结

Tab.4 Summary of open-source implementations

模型	框架	链接
DeepWalk	Python	https://github.com/phanein/deepwalk
Node2vec	TensorFlow	https://github.com/aditya-grover/node2vec
Graphsage	TensorFlow	https://github.com/williamleif/GraphSAGE
LINE	C++	https://github.com/tangjianpku/LINE
TADW	Matlab	https://github.com/thunlp/tadw
CANE	TensorFlow	https://github.com/thunlp/cane
GCN	TensorFlow	https://github.com/kipf/gcn
GAT	TensorFlow	https://github.com/PetarV-/GAT
LGCN	TensorFlow	https://github.com/divelab/lgcn/
SDNE	Python	https://github.com/suanrong/SDNE

表5 图深度学习开源库总结

Tab.5 Summary of graph deep learning library

图深度学习库	模型包含但不限于	链接
DGL	GCN, GAT, RGCN, LGNN, SSE, Tree-LSTM, DGMG, JTNN	https://github.com/dmlc/dgl
StellarGraph	GraphSage, HinSage, GAT, Attri2vec, GCN, SGC, APPNP, Node2vec, Metapath2Vec	https://github.com/stellargraph/stellargraph
OpenNE	DeepWalk, LINE, Node2vec, GraRep, TADW, GCN, HOPE, GF, SDNE, LE.	https://github.com/thunlp/OpenNE
Euler	DeepWalk, LINE, Node2vec, GCN, GraphSage, Scalable-GCN, GAT, LsHNE, LasGNN	https://github.com/alibaba/euler
PGL	GCN, GAT, GraphSAGE, unSup-GraphSage, LINE, DeepWalk, Meta-Path2Vec, Node2vec, SGC, Struc2vec, GES, DGI, GATNE.	https://github.com/PaddlePaddle/PGL
Plato	Node2vec-Randomwalk, LPA, HANP, KCore, Pagerank, Metapath-Randomwalk	https://github.com/tencent/plato
GraphVite	DeepWalk, LINE, Node2vec, TransE, RotatE, LargeVis, DisMult, ComplEx	https://github.com/DeepGraphLearning/graphvite
PyTorch Geometric	SplineCNN, GCN, NNConv, ECCConv, GAT, GraphSAGE, AGNN, Node2vec	https://github.com/rusty1s/pytorch_geometric

3.2 实际应用

网络表示学习可以应用到许多实际任务中。可以宽泛地将应用分为4类,即节点分类、社区发现、链接预测和网络可视化。

1) 节点分类 节点分类是网络节点表示为向量后最常见的任务,一般属于半监督的学习任务,即初始数据中数据标签只占一部分,学习已有标签的数据信息来标记其余数据的标签。常见的半监督学习任务有视频文档网页的类别标记、蛋白质生物功能的学习标记或者社交网络中预测部分用户的标签信息。通常先抽取节点的属性或结构特征来为节点生成嵌入信息,然后应用逻辑回归等分类器为对应节点预测标签。最近的研究评估^[15,22,25,44,48]表明这些模型对节点标签的预测或分类有很高的精度。其中GCN^[44]为代表的图卷积神经网络对节点的分类精度要高于传统方法,但它们都是对静态图进行嵌入分类。HAMILTON等^[48]提出归纳式学习模型 GraphSage,利用邻域信息学习聚合函数,以便对动态网络中的新增节点进行嵌入和分类。

2)社区发现 社区发现是在给定网络中,将物理对象或抽象对象的集合分组为类似对象组成的多个社区。社区发现是无监督的聚类,即对大量未知标注的数据集按数据的内在相似性将数据划分成多个类别,使得类别内数据相似度较大而类别间的数据相似度较小。与节点分类不同的是社区发现不需要任何标记信息,可应用范围更广,节省数据标记成本。在实际应用中,社区发现可以为药物网络划分社区,帮助推测相似药物,依据蛋白质网络中节点的联系将蛋白质自动分类。

3)链接预测 网络由实体间的交互信息构成,但在实体之间这种交互信息往往会缺失或者会在未来出现,在生物网络分析中验证节点之间是否存在链接需要复杂的实验测试和高昂的代价,因此网络的链接预测显得尤为重要。链接预测在推荐系统中应用广泛,例如 WANG 等^[61]和 OU 等^[62]预测了来自公共协作和社交网络上的节点链接。另外,链接预测在生物网络分析^[22]中也非常普遍,利用链接预测方法给出一个可能存在链接的序列是非常经济有效的。

4)网络可视化 网络表示学习为节点生成嵌入向量和网络降维可视化提供了新方法。每一个节点都被表示为一个低维稠密的向量,可以方便地利用已有的降维可视化算法如 t-SNE^[63],LargeVis^[64]等生成网络的二维或三维图形,这对于发现网络社区和其他隐藏结构有很大帮助。例如,DeepWalk^[15]的作者利用降维可视化嵌入的空手道俱乐部网络来展现 DeepWalk 方法的优越之处。Line^[25]的作者利用 LargeVis^[64]可视化 DBLP 作者网络,表明 LINE 能够将合作作者聚集在同一个领域。

4 总结与展望

本文介绍了现有的网络表示学习方法,将其分为两类并细分为 6 个类别,分别是传统方法、基于网络结构的嵌入、融入属性信息的嵌入,以及最近非常流行的基于谱域的图卷积、基于空间的图卷积和图 attention 网络。本文还对比了模型的适用性与算法特点,给出网络表示学习中的经典数据集,整理了常用模型的开源实现项目以及 8 个开源的图深度学习库。最后还介绍了网络表示学习的应用。网络表示学习发展迅速,不断取得成果,但在以下方面仍面临挑战。

1) 深层的图神经网络学习 不同于图像分类任务,实验表明,随着网络层数增加,相邻节点的嵌入表示将会逐渐靠近直至收敛到一个点,图神经网络模型性能会急剧下降^[44]。

2) 动态和异构网络的表示 真实世界中存在大量的异构网络并且随时间动态变化,大多方法模型将网络简化为同构的静止不变的网络去处理。现有的解决动态异构网络表示的模型仍有很大的提升空间。

3) 网络模型的泛化能力 实际场景中的网络往往复杂且差异较大,现有的网络模型设计都是针对某一种简化的网络,设计一个能泛化到大量网络的模型是一个值得深入研究的问题。

参考文献/References:

- [1] 涂存超,杨成,刘知远,等.网络表示学习综述[J].中国科学:信息科学,2017,47(8):980-996.
TU Cunchao, YANG Cheng, LIU Zhiyuan, et al. Network representation learning: An overview[J]. Scientia Sinica (Informationis), 2017, 47(8):980-996.
- [2] YING R, HE R N, CHEN K F, et al. Graph convolutional neural networks for web-scale recommender systems[C]//KDD'18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 974-983.
- [3] van den BERG R, KIPF T N, WELING M. Graph convolutional matrix completion[J]. Machine Learning, 2017. doi: arXiv:1706.02263.
- [4] MONTE F, BRONSTEIN M M, BRESSON X. Geometric matrix completion with recurrent multi-graph neural networks[C]//Advances in Neural Information Processing Systems. [S.l.]:[s.n.], 2017: 3697-3707.
- [5] 刘滨,张静远,刘强,等.微博分析研究综述[J].河北科技大学学报,2015,36(1):100-110.
LIU Bin, ZHANG Jingyuan, LIU Qiang, et al. Research overview of microblog analysis[J]. Journal of Hebei University of Science and Technology, 2015, 36(1):100-110.
- [6] GILMER J, SCHOENHOLZ S S, RILEY P F, et al. Neural message passing for quantum chemistry[C]//Proceedings of the 34th International Conference on Machine Learning. [S.l.]:[s.n.], 2017: 1263-1272.
- [7] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering [C]//Advances in Neural Information Processing Systems. [S.l.]:[s.n.], 2016: 3844-3852.
- [8] 郭松,张冬雯,许云峰,等.有向网络下的 CoDA 社区发现算法评估[J].河北科技大学学报,2017,38(2):169-175.
GUO Song, ZHANG Dongwen, XU Yunfeng, et al. Evaluation of the CoDA community detection algorithm based on directed network [J]. Journal of Hebei University of Science and Technology, 2017, 38(2):169-175.

- [9] 胡旭飞, 许云峰. 基于骨干度与网络编码的链路预测模型研究[J]. 河北工业科技, 2019, 36(5): 310-313.
HU Xufei, XU Yunfeng. Research on link prediction model based on backbone degree and network coding[J]. Hebei Journal of Industrial Science and Technology, 2019, 36(5): 310-313.
- [10] KRUSKAL J B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis[J]. Psychometrika, 1964, 29(1): 1-27.
- [11] BELKIN M, NIYOGI P. Laplacian eigenmaps and spectral techniques for embedding and clustering[J]. Advances in Neural Information Processing Systems, 2002, 14(6): 585-591.
- [12] HAMILTON W L, YING R, LESKOVEC J. Representation learning on graphs: Methods and applications[J]. Social and Information Networks, 2017. doi: arXiv:1709.05584.
- [13] MIKOLOV T, CHEN K, CORRADO G S, et al. Efficient estimation of word representations in vector space[J]. Computation and Language, 2013. doi: arXiv:1301.3781.
- [14] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.
- [15] PEROZZI B, AL-FOU R, SKIENA S. DeepWalk: Online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701-710.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017. doi: arXiv:1706.03762.
- [17] BRONSTEIN M M, BRUNA J, LECUN Y, et al. Geometric deep learning: Going beyond euclidean data[J]. IEEE Signal Processing Magazine, 2017, 34(4): 18-42.
- [18] WU Zonghan, PAN Shirui, CHEN Fengwen, et al. A comprehensive survey on graph neural networks[J]. Machine Learning, 2019. doi: arXiv:1901.00596.
- [19] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.
- [20] AHMED A, SHERVASHIDZE N, NARAYANAMURTHY S, et al. Distributed large-scale natural graph factorization[C]//Proceedings of the 22nd International Conference on World Wide Web. New York: ACM, 2013: 37-48.
- [21] GOYAL P, FERRARA E. Graph embedding techniques, applications, and performance: A survey[J]. Knowledge-Based Systems, 2018, 151: 78-94.
- [22] GROVER A, LESKOVEC J. node2vec: Scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855-864.
- [23] RIBEIRO L F R, SAVERESE P H P, FIGUEIREDO D R. Struc2vec: Learning node representations from structural identity[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2017: 385-394.
- [24] ZACHARY W W. An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1977, 33(4): 452-473.
- [25] TANG Jian, QU Meng, WANG Mingzhe, et al. Line: Large-scale information network embedding[C]//Proceedings of the 24th International Conference on World Wide Web. [S.l.]:[s.n.], 2015: 1067-1077.
- [26] CAO Shaosheng, LU Wei, XU Qiongzai. Grarep: Learning graph representations with global structural information[C]//Proceedings of the 24th ACM International Conference on Information and Knowledge Management. New York: ACM, 2015: 891-900.
- [27] YANG Cheng, LIU Zhiyuan, ZHAO Deli, et al. Network representation learning with rich text information[C]//Twenty-Fourth International Joint Conference on Artificial Intelligence. [S.l.]:[s.n.], 2015: 2111-2117.
- [28] TU Cunchao, LIU Han, LIU Zhiyuan, et al. Cane: Context-aware network embedding for relation modeling[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver: ACL, 2017: 1722-1731.
- [29] SUN Xiaofei, GUO Jiang, DING Xiao, et al. A general framework for content-enhanced network representation learning[J]. Social and Information Networks, 2016. doi: arXiv:1610.02906.
- [30] ZHANG Zhen, YANG Hongxia, BU Jiajun, et al. ANRL: Attributed network representation learning via deep neural networks[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. [S.l.]:[s.n.], 2018: 3155-3161.
- [31] ISLAM M R, PRAKASH B A, RAMAKRISHNAN N. SIGNet: Scalable embeddings for signed networks[C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining. [S.l.]:[s.n.], 2018: 157-169.
- [32] TU Cunchao, ZHANG Zhengyan, LIU Zhiyuan, et al. TransNet: Translation-based network representation learning for social relation extraction[C]//Twenty-Sixth International Joint Conference on Artificial Intelligence. [S.l.]:[s.n.], 2017: 2864-2870.
- [33] LIU Weiyi, LIU Zhining, SUZUMURA T, et al. Scalable Attribute-aware network embedding with locality[J]. Machine Learning, 2018. doi: arXiv:1804.07152.
- [34] NGUYEN D, MALLIAROS F D. BiasedWalk: Biased sampling for representation learning on graphs[C]//2018 IEEE International Conference on Big Data. Seattle: IEEE, 2018: 4045-4053.
- [35] DONG Yuxiao, CHAWLA N V, SWAMI A. metapath2vec: Scalable representation learning for heterogeneous networks[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2017: 135-144.
- [36] FU Taoyang, LEE W C, LEI Zhen. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning[C]//Proceedings of the 2017 ACM Conference on Information and Knowledge Management. New York: ACM, 2017: 1797-1806.

- [37] PAN Shirun, WU Jia, ZHU Xingquan, et al. Tri-party deep network representation[C]// The 25th International Joint Conference on Artificial Intelligence.[S.l.]:[s.n.], 2016:1895-1901.
- [38] CHANG Shiyu, HAN Wei, TANG Jiliang, et al. Heterogeneous network embedding via deep architectures[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2015: 119-128.
- [39] GUI Huan, LIU Jialu, TAO Fangbo, et al. Large-scale embedding learning in heterogeneous event data[C]//2016 IEEE 16th International Conference on Data Mining. Barcelona: IEEE, 2016: 907-912.
- [40] XU Linchuan, WEI Xiaokai, CAO Jiannong, et al. Embedding of embedding (coe): Joint embedding for coupled heterogeneous networks [C]//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. New York: ACM, 2017: 741-749.
- [41] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. Machine Learning, 2017.doi: arXiv:1710.10903.
- [42] WANG Xiao, JI Houye, SHI Chuan, et al. Heterogeneous graph attention network[J]. Social and Information Networks, 2019. doi: arXiv:1903.07293.
- [43] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[J]. Machine Learning, 2013. doi: arXiv:1312.6203.
- [44] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J].Machine Learning, 2017. doi: arXiv:1609.02907.
- [45] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61-80.
- [46] LI Yujia, TARLOW D, BROCKSCHMIDT M, et al. Gated graph sequence neural networks[J]. Machine Learning, 2015. doi: arXiv:1511.05493.
- [47] DAI Hanjun, KOZAREVA Z, DAI Bo, et al. Learning steady-states of iterative algorithms over graphs[C]//Proceedings of the 35th International Conference on Machine Learning.[S.l.]:[s.n.], 2018: 1106-1114.
- [48] HAMILTON W, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//Advances in Neural Information Processing Systems. Long Beach: NIPS, 2017: 1024-1034.
- [49] NIEPERT M, AHMED M, KUTZKOV K. Learning convolutional neural networks for graphs[C]//International Conference on Machine Learning. New York: ICML, 2016: 2014-2023.
- [50] GAO Hongyang, WANG Zhengyang, JI Shuiwang. Large-scale learnable graph convolutional networks[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1416-1424.
- [51] ZHANG Jiani, SHI Xingjian, XIE Junyuan, et al. Gaan: Gated attention networks for learning on large and spatiotemporal graphs[J]. Machine Learning, 2018. doi: arXiv:1803.07294.
- [52] LEE J B, ROSSI R, KONG Xiangnan. Graph classification using structural attention[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1666-1674.
- [53] ABU-EL-HAJJA S, PEROZZI B, AL-RFOU R, et al. Watch your step: Learning node embeddings via graph attention[C]//Advances in Neural Information Processing Systems. Montréal: NIPS, 2018: 9180-9190.
- [54] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data[J]. AI Magazine, 2008, 29(3): 93-106.
- [55] TANG Lei, LIU Huan. RELATIONAL learning via latent social dimensions[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2009: 817-826.
- [56] ZITNIK M, LESKOVEC J. Predicting multicellular function through multi-layer tissue networks[J]. Bioinformatics, 2017, 33(14): 190-198.
- [57] WALE N, WATSON I A, KARYPIS G. Comparison of descriptor spaces for chemical compound retrieval and classification[J]. Knowledge and Information Systems, 2008, 14(3): 347-375.
- [58] DEBNATH A K, de COMPADRE R L L, DEBNATH G, et al. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds: Correlation with molecular orbital energies and hydrophobicity[J]. Journal of Medicinal Chemistry, 1991, 34(2): 786-797.
- [59] DOBSON P D, DOIG A J. Distinguishing enzyme structures from non-enzymes without alignments[J]. Journal of Molecular Biology, 2003, 330(4): 771-783.
- [60] RAMAKRISHNAN R, DRAL P O, RUPP M, et al. Quantum chemistry structures and properties of 134 kilo molecules[J]. Scientific Data, 2014.doi:10.1038/sdata.2014.22.
- [61] WANG Daixin, CUI Peng, ZHU Wenwu. Structural deep network embedding[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.]:[s.n.],2016: 1225-1234.
- [62] OU Mingdong, CUI Peng, PEI Jian, et al. Asymmetric transitivity preserving graph embedding[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.]:[s.n.], 2016: 1105-1114.
- [63] van der MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008(9): 2579-2625.
- [64] TANG Jian, LIU Jingzhou, ZHANG Ming, et al. Visualizing large-scale and high-dimensional data[C]//Proceedings of the 25th International Conference on World Wide Web. [S.l.]:[s.n.], 2016: 287-297.