

图神经网络前沿进展与应用

吴 博 梁 循 张树森 徐 睿

(中国人民大学信息学院 北京 100872)

摘 要 图结构数据是现实生活中广泛存在的一类数据形式. 宏观上的互联网、知识图谱、社交网络数据, 微观上的蛋白质、化合物分子等都可以用图结构来建模和表示. 由于图结构数据的复杂性和异质性, 对图结构数据的分析和处理一直是研究界的难点和重点. 图神经网络(Graph Neural Network, GNN)是近年来出现的一种利用深度学习直接对图结构数据进行学习的框架, 其优异的性能引起了学者高度的关注和深入的探索. 通过在图中的节点和边上制定一定的策略, GNN 将图结构数据转化为规范而标准的表示, 并输入到多种不同的神经网络中进行训练, 在节点分类、边信息传播和图聚类任务上取得优良的效果. 与其他图学习算法相比较, GNN 能够学习到图结构数据中的节点以及边的内在规律和更加深层次的语义特征. 由于具有对图结构数据强大的非线性拟合能力, 因此不同领域的图相关问题上, GNN 都表现出更高的准确率和更好的鲁棒性. 本文在现有 GNN 研究的基础上, 首先概述了 GNN 的出现历程, 并介绍了相关概念和定义. 之后本文着重讨论和对比了 GNN 中的各种算法框架, 包括核心思想、任务划分、学习方式、优缺点、适用范围、实现成本等. 此外, 本文对 GNN 算法在多个不同领域下的应用场景进行了详细的阐述, 将 GNN 与其他图学习算法的优缺点作了联系和比较. 针对存在的一些问题和挑战, 本文勾画了 GNN 的未来方向和发展趋势, 最后对全文进行了全面而细致的总结.

关键词 图神经网络; 深度学习; 图结构数据; 拉普拉斯矩阵; 谱分解; 节点特征聚合; 图生成

中图法分类号 TP18

DOI 号 10.11897/SP.J.1016.2022.00035

Advances and Applications in Graph Neural Network

WU Bo LIANG Xun ZHANG Shu-Sen XU Rui

(School of Information, Renmin University of China, Beijing 100872)

Abstract As is known to all, Graph-structure data is a kind of data form widely existing in real life. Internet network, knowledge graph, social network data in macro perspective, together with protein, compound molecules data in micro perspective, are all can be modeled and represented by graph-structure. Because graph-structure data has complexity and heterogeneity attribute, the analysis and processing of graph-structure data have always been a difficulty in research community. The researchers have been studying the property information and topological structure information in graph and try to find out a way or method to learn and explore the graph automatically. In order to solve the problems above, Graph Neural Network (GNN) appears as a kind of framework that uses deep learning to learn graph-structure data directly in recent years. On the one hand, the excellent performance of GNN has aroused high attention and deep exploration in research community. GNN transforms the graph-structure data into standard representation by making a series of certain strategies on the multifarious nodes and edges of the graph, and then the representation can be input into a variety of different artificial neural networks for train-

收稿日期: 2019-11-24; 在线发布日期: 2020-02-04. 本课题得到国家自然科学基金(No. 62072463, No. 71531012)、国家社会科学基金(No. 18ZDA309)、北京市自然科学基金(No. 4172032)、京东商城电子商务研究项目(No. 413313012)、北大方正集团有限公司数字出版技术国家重点实验室开放课题资助. 吴 博, 博士研究生, 主要研究领域为图神经网络、深度学习. E-mail: wubochn@ruc.edu.cn. 梁循(通信作者), 博士, 教授, 计算机学会(CCF)会员, 主要研究领域为神经网络、社会计算、自然语言处理. E-mail: xliang@ruc.edu.cn. 张树森, 博士研究生, 主要研究领域为数据挖掘、社会计算. 徐 睿, 博士研究生, 主要研究领域为图像处理、机器学习.

ing, and achieves excellent results in tasks such as node classification, edge information dissemination, graph clustering and so on. On the other hand, when it is compared with other graph learning algorithms, GNN can learn the internal rules and semantic features of node and edge features in graph-structure data. Because it has a strong nonlinear fitting ability to graph-structure data, GNN has higher accuracy and better robustness on graph-structure related problems in different fields. To make it more suitable and efficient for specific applications, there is a great deal of variants of GNN algorithm and framework are proposed in past few years. Based on the existing GNN research, this paper first summarizes the history of GNN, and introduces the related concepts and definitions. After an overview of GNN theory, we then focus on the discussion and comparison of various algorithms in GNN, including the core idea, task division, types of graphs, activation function, different dataset, advantages and disadvantages, the scope of application, implementation costs, learning methods and benchmark network. We give a novel classification and divide GNN into five different artificial neural networks. In addition, the application of GNN algorithm in many different fields is described in details such as natural language processing, molecule graph generation and so on. This paper gives an introduction to the other kinds of graph learning algorithm which are recognized as network embedding and graph kernel. We compare the advantages and disadvantages between GNN and network embedding as well as graph kernel. Although GNN has been very popular over past years, these two kinds of graph learning algorithm are also to be proved competitive in some tasks. In view of the existing problems and challenges, this paper outlines the future direction and development trend of GNN, which includes depth of artificial neural network, dynamics, receptive field of GNN, the fusion of multi artificial neural networks, and the combination between artificial network embedding and GNN. Last but not least, we make a comprehensive and detailed summary of the full text.

Keywords graph neural network; deep learning; graph-structure data; laplacian matrix; spectral decomposition; node feature aggregating; graph generating

1 引 言

近年来,深度学习^[1]在多个领域取得明显优异的效果,特别是在计算机视觉、音频识别以及自然语言处理三个方面取得突破性进展.深度学习通过建立人工神经网络,对输入的信息和数据逐层进行特征的提取和筛选,最终获得分类和预测等任务的结果.相较于统计机器学习等浅层学习模式,深度学习所使用的神经网络架构具有多个功能各异的复杂网络层,其特征提取和识别的数量和质量显著提高,并且能够自底向上生成更加高级的特征表示.这使得机器能够获得抽象概念,具备更强的表征学习能力^[2].诸如多层感知机(Multilayer Perceptron, MLP)^[3]、卷积神经网络(Convolutional Neural Network, CNN)^[4]、循环神经网络(Recurrent Neural Network, RNN)^[5]、生成对抗网络(Generative Adversarial Network,

GAN)^[6]和自编码器(Auto-encoder, AE^[7]等性能优异的神经网络已经成为许多研究领域解决问题的通用网络框架.

但是随着研究的深入,研究人员发现深度学习并不能适应和解决所有的情况和问题.在过去十多年的发展中,深度学习取得的成就主要限定在了计算机视觉、自然语言处理和音频分析领域上.这些领域上的数据和信息有着比较显著的特点.文本、图像、音频、视频的数据格式在形式上有着统一而规整的尺寸和维度,它们也被称作欧式结构(Euclidean Structure)或者网格结构(Grid Structure)数据.除此之外,现实生活中存在大量的非欧式结构的图数据,例如互联网、知识图谱、社交网络、蛋白质、化合物分子等.尽管深度学习在欧式结构数据上取得巨大的成功,但在图结构数据上,基于神经网络的深度学习表现得并不好.在图结构数据中,节点与节点之间的边连接可能是均匀分布的,也可能是不均匀的.节点与节点之间没有严格意义上的先后顺序.对于

神经网络的输入端而言,这些数据没有固定的输入尺寸.在数学表达上,这些数据与欧式结构数据相比,每一个区块的特征矩阵维度都不是统一的,如图 1 所示.由于无法使用统一规整的算子对数据编排,导致 CNN 等神经网络不能再直接对其进行诸如卷积和池化等操作,也就不再有局部连接、权值共享、特征抽象等性质^[8].如何将 CNN 等深度学习算法用于分析图结构数据上成为一个有挑战性和前沿性的课题.近年来 Gori 等人^[9]用 RNN 来压缩节点信息和学习图节点标签,首次提出图神经网络(Graph Neural Network, GNN)这一概念.之后文献^[10]提出图卷积网络(Graph Convolutional Network, GCN),正式将 CNN 用于对图结构数据建模. GCN 通过整合中心节点和邻居节点的特征和标签信息,给出图中每个节点的规整表达形式,并将其输入到 CNN 中.这样一来 GCN 就能利用多尺度的信息,组合成更高层次的表达.其有效地利用了图结构信息和属性信息,为深度学习中其他神经网络迁移至图上提供了标准的范式.在新的研究思路的基础上,各种 GNN 架构相继被构造出来,在多个领域的图结构数据中发挥了独特的作用,并促进了图相关的

人工智能推理任务的发展.

本文针对近年来出现的 GNN 学习方法和研究现状进行了系统的归纳和梳理,并对它们的主要思想、改进以及局限性做了详尽分析.目前已有 Xu 等人^[11]关于图卷积神经网络的综述,本文在全面对比分析的基础上,对目前主要的 GNN 算法进行了更加合理的分类和介绍.除了图卷积神经网络, GNN 主流算法还包括有图自编码器、图生成网络、图循环网络以及图注意力网络.本文对每类 GNN 算法都给出了其定义和典型方法,将 GNN 中每种算法的机制、优势、缺点、适用范围、实现成本等进行了提炼总结.在进行了相应的数据实验基础上,与其他基准图算法进行了比对.本文在第 2 节中给出关于 GNN 的基本概念和定义;在第 3 节分门别类的给出 GNN 的主要模型和算法;在第 4 节,对比和分析 GNN 与网络嵌入(Network Embedding)以及图核(Graph Kernel)方法的特性和优势.在第 5 节中,阐述目前 GNN 在多个领域图数据上的具体应用;在第 6 节归纳和总结现有 GNN 模型缺陷和不足,并对未来发展方向和趋势进行展望.最后在第 7 节对全文所述进行总结.

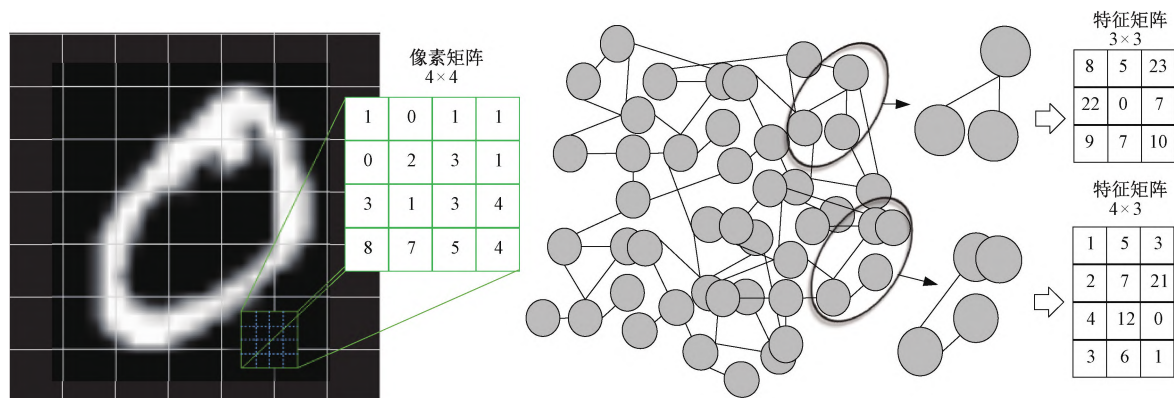


图 1 欧式结构和图结构数据比较

2 概念与定义

2.1 符号与定义

GNN 是指使用神经网络来学习图结构数据,提取和发掘图结构数据中的特征和模式,满足聚类、分类、预测、分割、生成等图学习任务需求的算法总称. GNN 的历史最早可以追溯到 2005 年, Gori 等人^[9]第一次提出 GNN 概念,用 RNN 来处理无向图、有向图、标签图和循环图等.在这之后, Scarselli 等人^[12]和 Micheli 等人^[13]继承和发展了该模式的

GNN 算法,并做了一定程度的改进.早期阶段的 GNN 主要是以 RNN 为主体框架,通过简单的特征映射和节点聚集为每个节点生成向量式表达,不能很好地应对现实中复杂多变的图数据.针对此情况,文献^[10]提出将 CNN 应用到图上,通过对卷积算子巧妙的转换,提出了图卷积网络(Graph Convolutional Network, GCN),并衍生了许多变体. GCN 实现了 CNN 在图上的平移不变、局部感知和权值共享^[14],为接下来其他 GNN 框架的构造和改进提供思想上的指导和借鉴.下面给出 GNN 中相关符号的说明和定义.

定义 1. 图(Graph)是由节点(Vertex)和连接节点的边(Edge)所构成,通常记为 $G=(V,E)$. 其中 $V=\{v_1, v_2, \dots, v_n\}$ 代表节点集合, $E=\{e_1, e_2, \dots, e_m\}$ 代表边集合. 通常节点也被称为顶点或者交点. 边也被称为链接或者弧. 通用的图表示是一个五元组: $G(V, E, A, X, D)$. 其中 $A^{N \times N}$ 代表图的邻接矩阵, $X^{N \times F}$ 代表节点的特征矩阵, $D^{N \times N}$ 代表度矩阵, N 和 F 分别代表节点的数量和节点的特征维度.

定义 2. 组合拉普拉斯矩阵, 又称标准拉普拉斯矩阵, 由对角矩阵和邻接矩阵组合而成:

$$L = D - A \quad (1)$$

该矩阵只在中心节点和一阶相连的节点上有非零元素, 其余之处均为零.

定义 3. 对称标准化拉普拉斯矩阵:

$$L^{\text{sym}} = I - D^{-1/2} A D^{-1/2} \quad (2)$$

其元素值由式(3)给出:

$$L^{\text{sym}}_{i,j} = \begin{cases} 1 & i = j \text{ 且 } \deg(v_i) \neq 0 \\ \frac{-1}{\sqrt{\deg(v_i)\deg(v_j)}} & i \neq j \text{ 且 } v_i \text{ 与 } v_j \text{ 相连} \\ 0 & \text{其他} \end{cases} \quad (3)$$

由于需要对邻居节点的数量进行统一并对特征值求平均, 所以一般在谱分解图卷积中多用定义 3 的对称标准化拉普拉斯矩阵进行傅里叶变换和特征向量分解.

2.2 图的类型

不管是数学中的拓扑网络还是现实生活中的网络, 都可以统一抽象为图. 图的形式多种多样, 按照图的边是否具有方向可以划分为有向图和无向图. 边的属性称作权值, 根据边是否具有权值可以将图划分为带权图和非带权图. 此外还有异质图 and 同质图. 同质图是指图中所有数据都为同一类型, 即节点类型和边类型单一. 而图中的数据类型不尽相同的为异质图, 即具有多种类型的节点和边.

事实上, GNN 不仅仅局限于图数据处理. 从广义上看, 由于可以在赋范空间内建立拓扑关联, 现实世界的任何数据都可以被视作图结构数据, 包括最为常见的文字、图像、音频和视频. 因此在转变为图结构形式后, 欧式结构的数据也可以输入到 GNN 中进行训练和问题求解. 这在下文介绍的实验数据集上有所体现.

3 图神经网络模型

3.1 图卷积网络

图卷积网络(Graph Convolutional Network, GCN)进行卷积操作主要有两种方法: 一种是基于谱分解, 即谱分解图卷积. 另一种是基于节点空间变换, 即空间图卷积. Bruna 等人^[10]第一次将卷积神经网络泛化到图数据上, 提出两种并列的图卷积模型——谱分解图卷积和空间图卷积. Bruna 等人对比分析了一般图结构数据和网格数据共有的特点和不同之处, 综合运用了空间图卷积和谱分解处理图像聚类问题. 下面本文对谱分解图卷积和空间图卷积进行详细的梳理和介绍.

3.1.1 谱分解图卷积

尽管文献^[10]最早提出用 CNN 分析图数据, 但是谱分解图卷积与图信号处理(Graph Signal Processing)^{[15][16][17][18]}有着密不可分的关联关系. 文献^[19]提出处理图上离散信号方法(Discrete Signal Processing, DSP), 该方法把节点特征视作图信号, 用邻接矩阵表示图中信号的转移和扩散, 利用傅里叶变换扩展图信号并进行过滤. 在 GCN 中, 过滤器、卷积、傅里叶变换、谱分解等理念都在该文献中有涉及. 在傅里叶变换中, 文献^[20]将图上的特征向量类比为拉普拉斯算子的特征函数, 为傅里叶变换和卷积运算推广到图上奠定基础. 拉普拉斯矩阵是半正定对称矩阵, 因此可以进行特征分解, 由此傅里叶变换中的特征函数可以转变为拉普拉斯矩阵的特征向量. 所以谱分解其实是利用图的拉普拉斯矩阵进行特征分解. Chung^[21]提出的谱图理论(Spectral Graph Theory)使得通过谱分解构造卷积核成为可能, 其谱乘法器的平滑性支撑起了空间局部性. 文献^[10]通过操作图的拉普拉斯矩阵的特征向量来进行卷积操作, 构造了卷积函数:

$$H^l = \sigma(\mathbf{U} \mathbf{G}_\theta(\Lambda) \mathbf{U}^T \mathbf{X}) \quad (4)$$

$$\mathbf{G}_\theta(\Lambda) = \begin{bmatrix} \theta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_n \end{bmatrix} \quad (5)$$

其中 $\sigma(\cdot)$ 是 sigmoid 激活函数, H^l 是每层的输出, $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ 是网络层参数, \mathbf{X} 是节点特征向量. 通过初始化权值, 在数据训练中进行信息前向传播, 并对损失函数求偏导将误差进行反向传播. 但是缺点在于需要进行特征分解, 在大图上特征分解的

计算代价很高. 而且每一次前向传播需要计算 U 、 $G_\theta(\Lambda)$ 以及 U^T 三者的乘积, 其计算复杂度与节点数量的平方成正比, 即 $O(N^2)$. 由于 Bruna 等人提出的谱分解图卷积中卷积核参数多, 在网络层传播过程中计算复杂度成幂指数增加, 导致 Bruna 提出的图卷积算法主要限定于只具有低维特征的节点以及小规模图上, 因此如何对参数缩减成为一个迫切需要解决的问题.

在图信号处理领域, 文献^[22]使用切比雪夫多项式近似拟合参数, 在图拉普拉斯矩阵上进行快速小波变换. 文献^[23]则使用 Lanczos^[24]算法将拉普拉斯矩阵进行正交相似变换, 构造对角矩阵来生成 krylov^[24]子空间的正交基, 并推广到矩阵函数. 这些图信号快速过滤模型为谱分解图卷积加速计算提供了理论和方法上的指导. 在此基础上, Defferrar 等人^[25]提出一种基于快速滤波递推公式的局部卷积算子, 通常记作 Chebyshev 卷积算子. 其思想是应用切比雪夫多项式对特征矩阵求解, 卷积核可以被定义成如下形式:

$$g_\theta(\Lambda) = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\Lambda}) \quad (6)$$

其中 θ 是切比雪夫系数, T_k 是切比雪夫多项式中的第 k 项, 该项由递归定义产生: $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$. 其中 $T_0(x) = 1$ 而 $T_1(x) = x$. $\tilde{\Lambda} = 2\Lambda/\lambda_{max} - \mathbf{I}$ 是标度特征值的对角矩阵, 取值在 $[-1, 1]$. 由此卷积层输出表达为

$$H^l = \sigma\left(\sum_{i=0}^K \theta_k T_k(\tilde{\mathbf{L}}) \mathbf{X}\right) \quad (7)$$

中 $\tilde{\mathbf{L}} = 2\mathbf{L}/\lambda_{max} - \mathbf{I}$, 这样卷积核的参数就变成只有 K 个. 每次卷积时将 K 阶邻居上的特征向量进行加权求和, 而 θ_k 就相当于权重系数. 计算复杂度为 $O(K|E|)$, 这使得训练参数的计算复杂度大幅度下降, 特别是在稀疏图处理中尤其明显. 文献^[26]认为 Chebyshev 卷积算子的一个主要缺点是 $\tilde{\Lambda}$ 只能取值在 $[-1, 1]$ 之间, 只适用于特征值较小的情况. 当拉普拉斯矩阵的特征值簇群分布在更大范围上时, 这种缺陷尤其明显. 因此 Levie 等人^[26]提出 CayleyNet 算法. CayleyNet 基于参数化有理复函数多项式(Cayley Polynomial), 引入了谱变焦参数并调节其大小, 使其能更好地适应不同拉普拉斯矩阵特征值. Levie 等人证明了如果用雅可比置换公式, Chebyshev 卷积算子是 Cayley 的一个特例. Cayley-Net 能输出更好的平滑谱转移函数表示, 并产生高

质量的局部空间滤波器, 并且时间复杂度与稀疏连通图的边数成线性相关.

在节点分类任务中, 有可能部分节点没有标签, 因此有监督学习不再适用此情况. 针对此问题, Kipf 等人^[27]提出半监督 GCN 节点分类模型, 该模型在图的边数上进行线性缩放并学习隐藏层表示. 这些表示既编码局部图结构, 也编码节点特征, 并且将图拉普拉斯正则项添加进损失函数上:

$$\mathcal{L} = \mathcal{L}_0 + \lambda \mathcal{L}_{reg} = \mathcal{L}_0 + \lambda f(\mathbf{X})^T \mathbf{L} f(\mathbf{X}) \quad (8)$$

\mathcal{L}_0 是图中带标签节点的有监督损失函数, λ 是 \mathcal{L}_{reg} 的权重参数, $f(\cdot)$ 是神经网络可微函数. 改进 $f(\cdot)$ 蕴含的条件从而允许模型从 \mathcal{L}_0 中分配梯度信息. 该方法不再假定相连的节点更倾向于具有相似类别, 而是使用数据中部分有标签的节点数据对 GCN 训练, 使网络模型对其余无标签的数据进行分类. Kipf 等人在卷积核上进行一阶近似生成网络层间传播规则, 通过稀疏密集矩阵乘法作用于信息前向传播上. 式(7)中令 $K=1$, 并近似让 $\lambda_{max} \approx 2$, 即每层卷积只考虑直接邻域, 类似于 CNN 中 3×3 卷积核, 并运用归一化技巧得到输出:

$$\mathbf{Z} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X} \Theta \quad (9)$$

其中 $\Theta \in R^{C \times F}$ 为简化后的由单参数构成的滤波器参数矩阵, $\mathbf{Z} \in R^{N \times F}$ 是卷积后的信号矩阵. 其时间开销跟边的数量成线性相关, 计算复杂度为 $O(|E|FC)$. 该方法只考虑一阶邻域, 并通过堆叠多层增加感受野. Kipf 等人的主要贡献在于对网络的深度加深, 减少邻域宽度, 进一步提升了准确率和学习能力. 但是该方法仍然限定在无向图和无权图上, 并且要加载整个图才能进行训练, 对于机器存储空间要求较高, 因此超大规模的图可能加载和运算时间比较长.

在谱分解图卷积的进一步改进中, Li 等人^[28]认为目前大部分谱分解图卷积的过滤器被设计用来处理具有共享和固定的图结构数据, 没有考虑到在大小和连通性上的变化. 因此 Li 等人提出一个概括性的动态图卷积网络(Adaptive Graph Convolution Network, AGCN). 为了能刻画不同的拓扑结构, AGCN 给每组数据中的单个样本订制一个拉普拉斯矩阵, 使得过滤器能根据图的拓扑结构组合邻居节点特征, 将组内节点特征和组间节点特征联系起来. Li 等人认为在图上工作的卷积器不能保证提取出全部有意义的特征, 因此 AGCN 会训练一个残差图来探索图中的残差子结

构.同时为了减少训练多个拉普拉斯矩阵的时间,AGCN用马氏距离(Mahalanobis Distance)学习矩阵间最优距离度量参数,将原来的计算复杂度从与节点个数相关变成与网络层维度相关.文献^[29]提出一种图胶囊网络(Graph Capsule Network, GCAPS-CNN),借鉴了文献^[30]中局部胶囊(Local Capsules)的思想——封装复杂的隐含实体计算,并输出具有实例化参数的信息向量.由于标准化卷积层中的局部性会导致聚合函数得到的节点特征有信息丢失^[31],因此 GCAPS-CNN 将图卷积运算的标量输出改成信息向量,其包含了节点特征的高阶统计信息.同时固定每层输出的节点特征维数以避免维度爆炸.针对置换等变性问题,GCAPS-CNN 构造基于协方差矩阵的置换不变层,在胶囊函数中使用双射的多项式系数保存局部邻居特征值,从而脱离了对节点顺序的依赖,并使用图谱距离簇(Family of Graph Spectral Distances)^[32]学习图全局拓扑结构特征.由于管理图胶囊向量维度的需要额外的信息,因此学习成本会有一定上升.此外,为了同时学习到局部一致性和全局一致性,Zhuang 等人^[33]设计了一种双图卷积神经网络.在双图卷积神经网络中,有两个图卷积神经网络存在,分别代表全局知识和局部知识.通过随机游走构建正点互信息(Positive Pointwise Mutual Information, PPMI)矩阵,利用邻接矩阵和 PPMI 矩阵分别嵌入基于局部一致性和全局一致性的图知识,并引入一个无监督损失函数整合来自两个网络的不同数据:

$$\mathcal{L}_0 + \lambda(t) \mathcal{L}_{\text{reg}}(\text{Conv}_A, \text{Conv}_P) \quad (10)$$

其中 $\lambda(t)$ 是一个随时间变化的动态参数, Conv_A 代表相连的节点趋向具有相同标签的知识, Conv_P 代表在相同上下文环境下出现的节点趋向具有相同标签的知识.

传统 CNN 为了降低网络层后端的训练参数,往往会设计一个或者多个池化层,这在 GCN 中也十分有必要.在图池化处理上, Henaff 等人^[34]将卷积算子扩展到高维数据上,使用多谱聚类在不同尺度上构造邻域,在图卷积网络架构中加入最大池化层和平均池化层,逐步抽取更高级别的语义特征,并且降低了模型复杂度. Henaff 等人假设图的结构没有先验信息,结合高斯扩散核(Gaussian Diffusion Kernel)^[35]以及自校正扩散核(Self-tuning Diffusion Kernel)^[36]计算节点特征相似度矩阵,完成无监督和有监督的图估计(Graph Estimation)目标.在大

规模文本分类、生物信息预测以及图像识别任务上验证了该方法的有效性.此外,文献^[37]提出一种包含 SortPooling 网络层结构的深度图卷积神经网络(Deep Graph Convolutional Neural Network, DGCNN). DGCNN 先用图卷积层处理输入的图结构数据,然后将产生的图节点排列成固定序列.借鉴 Weisfeiler-Lehman^[38]图核算法用来对具有同一属性的节点进行聚类,其输出在形式上为固定维度大小的矩阵.之后再使用 CNN 中的一维的卷积层进行卷积运算,最后与全连接层相连.该体系结构允许对原始图进行端到端的梯度下降训练,无需将图转换为向量. DGCNN 通过置换不变量的图结构任务增强池化网络,从而提高 GCN 的特征缩减和特征抽象的性能.而文献^[39]更进一步地认为 GCN 等代表性 GNN 学习函数在本质上是扁平的,没有学习到图的层次表示.这一限制影响了图分类任务中图标签的预测精度,因此提出一种端到端的可微分图池化方法——DiffPool. DiffPool 为每一层节点学习一个可微分的软簇分配,这种软簇分配是将已经学习好的节点嵌入表示映射到一组聚类集合中,然后输入到下一个神经网络层中去.即 GCN 下一层的输入对应上一层中的聚类集合,通过循环迭代多次就可以生成图的层次表示,相应的图表示也会越来越抽象.最终输出的节点向量可以用作 Softmax 层的输入,并且使用随机梯度下降来进行端到端的训练.同时,每经过一层池化处理,图的输入规模也会相应的减小,显式地降低了计算量. DiffPool 不仅仅可以嵌套在 GCN 网络框架上,并且可以与各种 GNN 端到端地相结合.实验结果表明, DiffPool 结合现有的 GNN 方法平均提高了 5% 到 10% 的精度.在图分类问题上, DiffPool 在五个基准数据集中的四个上实现了最佳性能.

表 1 从模型、核心思想、图种类、学习模式、激活函数、数据以及任务对谱分解图卷积模型进行了汇总.

3.1.2 空间图卷积

空间图卷积从图结构数据的空间特征出发,探讨邻居节点的表示形式,使得每个节点的邻居节点表示变得统一和规整,方便卷积运算.空间图卷积方法主要有三个关键问题,一是中心节点的选择;二是感受域的大小,即邻居节点个数的选取;三是如何处理邻居节点的特征,即构建合适的邻居节点特征聚合函数.

表 1 谱分解图卷积模型

模型/文献	核心思想	图种类	学习模式	激活函数	数据	任务
文献 ^[10]	傅里叶变换	同质无向	有监督	ReLU	①	图聚类
Chebyshev ^[25]	切比雪夫多项式近似拟合	同质无向	有监督	ReLU Softmax	①②	图分类
CayleyNet ^[26]	Cayley 多项式	同质无向	有监督 半监督	ReLU	①③④	图分类 节点分类 矩阵补全
文献 ^[27]	拉普拉斯正则项 一阶邻域近似	同质无向 同质有向	半监督	ReLU Softmax	⑤⑥⑦⑧	节点分类
AGCN ^[28]	单样本拉普拉斯矩阵 残差图 马氏距离	异质无向	有监督	—	⑮⑯⑰⑱⑲ ⑳㉑㉒ ㉓	图分类 图预测
GCAPS-CNN ^[29]	高阶统计向量 协方差矩阵 图谱距离簇	异质无向 同质无向	半监督	Softmax	⑨⑩⑪⑫ ㉔㉕㉖㉗	图分类
文献 ^[33]	双图卷积 PPMI 矩阵	异质无向	半监督	Sigmoid Softmax	⑤⑥⑦⑧	节点分类
文献 ^[34]	高斯扩散核 自校正扩散核	异质无向 同质无向	有监督 无监督	Relu	⑭⑮⑯	图分类 图预测
DGCNN ^[37]	SortPooling Weisfeiler-Lehman 图核	异质无向	有监督	Softmax Tanh Relu	⑨⑩⑪⑫⑬ ㉔㉕㉖㉗	图分类
DiffPool ^[39]	微分软聚类 节点嵌入映射	异质无向 同质无向	有监督	Softmax Relu	⑨⑬⑭⑮⑯⑰	图分类
数据集	①Mnist;②20News;③Cora;④MovieLens;⑤Cora;⑥Citeseer;⑦Pubmede;⑧NELL;⑨COLLAB;⑩IMDB-BINARY;⑪IMDB-MULTI;⑫REDDIT-BINARY;⑬REDDIT-MULT;⑭Reuters;⑮Delaney solubility;⑯Az-logD;⑰Hydration-free energy;⑱Tox21;⑲ClinTox;⑳Sider;㉑Toxcast;㉒Sydney Urban Objects;㉓PTC;㉔PROTEIN;㉕NCI1;㉖NCI109;㉗D&D;㉘DPP4;㉙ENZYMES;㉚ImageNet;㉛MUTAG					

针对上述三个关键问题,文献^[40]首先提出在空间上的卷积操作方法——PATCHY-SAN. PATCHY-SAN 的处理流程可以分为三个步骤:节点中心度量、节点邻域集合收集以及子图规范化. 节点中心度量是为了确定节点次序. PATCHY-SAN 提供了两种可供选择的方法——节点中心度或者 Weisfeiler-Lehman^[41]算法,用于测量节点的中心性以获得节点排名,并且按照固定的间隔从排序中选取节点. 在节点邻域集合收集

利用广度优先搜索扩展中心节点的邻居节点,和中心节点一起构成一个固定大小的邻域集合. 若某个节点缺少邻居节点,则在其一阶邻居节点的基础上继续扩展,直到收集到足够的邻域集合节点数量. 在子图规范化中,对邻域集合中的节点按照一定的标号函数进行排序,形成规范化子图. 每个节点属性值构成了一个输入通道,边的属性值也构成了一个输入通道,再用标准的 CNN 完成空间图卷积. 该过程如图 2 所示.

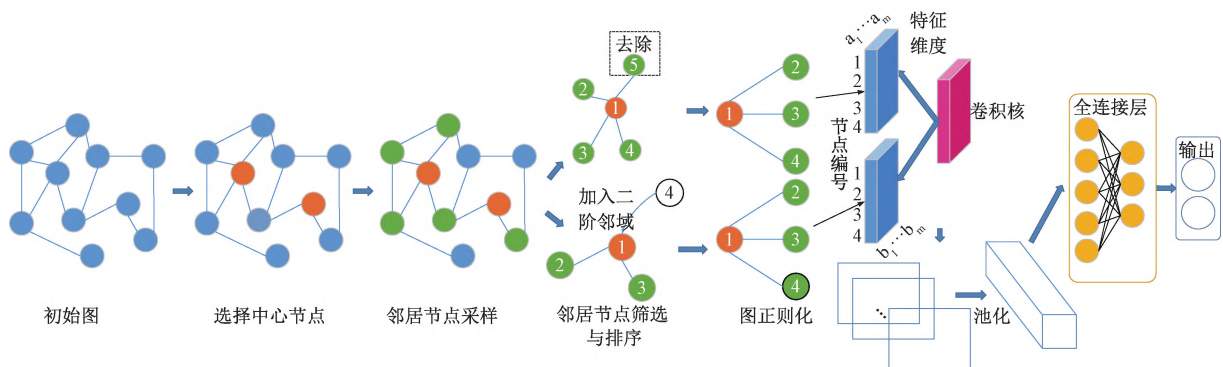


图 2 一个标准的空间图卷积处理流程

PATCHY-SAN 模型有序地组织了中心节点和邻居节点,并对它们的特征加以了利用. 子图规范

化使得每个子图都是固定大小的,方便参数共享. 但是其缺点在于节点中心度量函数不够确定,中心节

点排序的好坏直接影响模型的性能. 邻居节点的数目需要根据训练量和图结构进行调整. 模型在较小规模的图数据中有过拟合的情况发生.

除了 PATCHY-SAN 构造规则欧式结构数据的方法, 文献^[42]提出扩散卷积神经网络(Diffusion Convolutional Neural Network, DCNN). DCNN 基于扩散核的思想, 考虑节点之间链接重要性的不同, 将中心节点的邻居节点特征进行映射. 扩散卷积核的操作使得同质图的输入会得到同一个预测结果, 因而具有平移不变性. 对于图 t 中每一个节点 i , 每跳 j 后的特征 k , 其激励函数可以概括为

$$\mathbf{Z}_{ijk} = f(\mathbf{W}_{jk}^c \cdot \sum_{l=1}^{N_t} \mathbf{P}_{ijl}^* \mathbf{X}_{tlk}) \quad (11)$$

其中 \mathbf{P}_l 为节点概率转移矩阵, \mathbf{X}_l 为特征矩阵, \mathbf{W}^c 为权值矩阵. DCNN 的核心是概率转移矩阵, 在一定程度上可以识别同构图. 但是面对稠密图时, 每跳内访问的节点数目会非常多, 需要保存大量的张量, 这对于计算机内存空间有较大的需求. 模型能较好地获取图局部信息, 但是欠缺捕获长距离节点信息的能力. 针对此问题, 文献^[43]使用随机游走(Random Walk)对图进行预处理. 在图结构已知的情况下, 对于图中每个节点, 根据与中心节点的邻近相关性大小排序, 并选取 p 个邻居节点. 其排序值由转移矩阵计算完成. 这样时间复杂度由 DCNN 的 $O(N^2F)$ 下降为 $O(Np)$, 其中 $p \ll N$. 而在图结构未知的情况下, 用图相关矩阵作为相似矩阵来估计转移概率. 之后用卷积算子的权值与邻居节点内积作为卷积输出, 权重值根据邻居节点与中心节点的距离调整, 距离越远值越低, 代表相关性越小.

感受域大小一直是空间图卷积中的关键参数, 其涉及到邻居节点数目的选取. 由于空间图卷积通常都是递归地计算邻居节点, 导致在网络层数线性增加的同时, 感受域的大小指数递增. 为了减少训练的复杂程度, 相应的采样方法被提了出来. 这其中比较主流的采样方法可以分为两类: 一类是修改邻居节点的采样策略, 另一类是修改中心节点的采样策略. 在邻居节点采样上, 文献^[44]提出一种归纳式节点嵌入算法 GraphSAGE(Graph Sample and Aggregate). GraphSAGE 中心节点的感受域由多轮迭代产生, 在每轮迭代中抽取不同标准数目的邻居节点. 对于邻居节点数目不足的, 采取重复采样策略, 并生成中心节点的特征聚集向量. 相较于传统直推式节点嵌入算法, GraphSAGE 并不是为图中每个节点生成固定的表示, 而是学习一个为节点产生嵌

入表示的映射函数. 在中心节点的采样上, 文献^[45]提出基于重要性采样(Importance Sampling)的 FastGCN 算法. 它的核心思想是将图卷积解释为概率测度下嵌入函数的积分变换, 即图中的节点都被认定是独立同分布的样本, 把损失函数和卷积层当作是节点嵌入函数的积分, 通过定义样本损失和样本梯度的蒙特卡罗(Monte Carlo)近似来计算积分. 由于对中心节点进行了选择性采样, 模型训练不再是对全图进行, 而是部分节点批量迭代, 这使得训练过程更加有效. 相较于 GraphSAGE, FastGCN 可以在新的节点和边不断加入进来的情况下, 预测精准度仍旧保持较高水准.

尽管邻居节点随机采样和节点中心重要性采样能够在一定程度上控制感受域的规模, 但是存在估计上的偏差和采样的邻居节点消失的缺陷. 针对此问题, 文献^[46]提出基于方差缩减的随机训练方法对其进行优化. 其将节点的历史激活信息 \bar{h}_v^l 作为控制的变量, 卷积层激活函数为

$$(\mathbf{P}\mathbf{H}^l)_u = \sum_{v \in n(u)} \mathbf{P}_{uv} \nabla h_v^l + \sum_{v \in n(u)} \mathbf{P}_{uv} \bar{h}_v^l \quad (12)$$

其中 \mathbf{P} 是正则化后的传播矩阵, \mathbf{H} 是 l 层的激活矩阵. 在模型训练中, 权重参数没有发生较大改变时, 近似地用上轮的激活信息 \bar{h}_v^l 代替 h_v^l , 而 $\nabla h_v^l = h_v^l - \bar{h}_v^l$. 在多个数据集上的评测表明, 该方法的收敛性和准确率与没有使用采样方法的 GCN 不相上下, 但大幅度减少了计算量. 此后, 采样策略从对节点操作优化逐渐转变为对网络层更新函数的优化. 文献^[47]提出一种网络层间自适应采样策略(Adaptive Sampling). 在自顶向下分层构造网络过程中, 对底层的采样会在一定程度上受限于顶层. 其中采样的邻居节点对于上一层网络是可见的, 由不同的父节点共享, 避免了邻域因为固定大小数值而导致的过度扩展. 自适应采样策略将层间更新函数重新定义为期望的形式, 使得邻居节点服从一定的概率分布, 并引入蒙特卡罗算法, 采样器会根据训练过程中方差减少而动态适应. 每个节点有一个自相关函数, 用来确定其对采样的重要性. 通过层间跳跃连接(Skip Connection)的方式保存二阶近似邻居节点, 让信息在不同网络层间传播, 从而增强训练的有效性和经济性. 该方法总结了 GraphSAGE 和 FastGCN 的采样, 并指出它们都是自适应采样的一个特例.

在邻居节点数目确定后, 需要考虑对邻居节点特征进行聚合处理. 在邻居节点特征聚合上, 一般的

做法是对邻居节点进行线性或者非线性变换,并与中心节点特征进行整合.在 GraphSAGE 中,设计了三种不同的聚合策略:Mean Aggregator——将邻居节点特征值取平均;LSTM Aggregator——邻居节点按随机排列输入 LSTM,并取其隐层输出;Pooling Aggregator——取邻居节点特征经过线性变换后各个位置上最大值.除了对一阶邻居节点的聚合,二阶及以上的邻居节点特征也不能被忽视.因此文献^[48]提出 CoN-GCN(Core Neighbors-GCN).CoN-GCN 算法分为两个步骤:步骤一为待编码节点选择重要的邻居节点,邻居节点由结构紧密度函数计算得出.步骤二是从候选邻居节点中分级采样得到固定数量的核心邻居节点.经过编码后,将节点及其邻域的特征矩阵送入卷积层.由于每个节点的一阶邻居节点和二阶邻居节点都被聚合,因此 CoN-GCN 能够学得中心节点更有效的表示.CoN-GCN 算法核心在于结构紧密度量函数,可以使用共同邻居数,Jaccard 系数和 Adamic/Adar 系数.如果出现结构度量相同的邻居节点,则再使用欧式距离区分,从而得到邻居节点的等级排序.Gao 等人^[49]提出可学习图卷积层(Learnable Graph Convolutional Layer, LGCL).LGCL 提出一种子图训练方法.在选取初始节点后,用宽度优先搜索动态随机挑选外围节点数量,在达到规定数目后,停止扩展.在形成子图后,LGCL 根据原始邻居节点特征值排名构造新的邻居特征向量,特征向量的维数相当于通道数,通过一维卷积神经网络进行卷积运算,最终将中心节点与邻居节点聚集的多维特征压缩为一维格式.LGCL 在一定程度上减少了图卷积的内存和计算时间,适用于大规模图训练.

此外,一些学者从边的角度出发,在邻居节点特征聚合中利用特殊的边信息,能够对邻居节点特征聚合产生更加有效的作用.例如在社交媒体等现实网络中,边的符号是必须被考虑的事情.将 GCN 的模型框架从无符号图上应用到有符号图上需要面对两个挑战:一是如何处理负链接;二是如何将负链接和正链接有机结合,从而学习有效的节点表示.针对上述问题,Derr^[50]提出基于平衡理论(Balance Theory)的符号图卷积网络(Signed Graph Convolutional Network, SGCN).在 SGCN 中,一个平衡路径被定义成由偶数个负链接组成,一个不平衡路径由奇数个负链接组成.由此,聚合函数分为两个部分:一个对应中心节点的平衡邻居节点集合,一个对应不平衡的邻居节点集合.将递归定义得到的平衡

路径节点集合和不平衡路径节点集合与中心节点特征进行拼接并激活输出,达到在不同神经网络层之间传播正负极性连接信息的目标.在四个真实世界数据集上的实验表明,SGCN 能够高效地学习到节点的潜在表达,并且在长路径上的聚合函数能够提升节点表达效果. Such 等人^[51]提出 Graph-CNN. Graph-CNN 构建了一个具有图池化操作的异质图过滤器,能够同时学习图中节点和边的特征值. Graph-CNN 可以识别具有边缘特征的图,通过计算卷积过滤器与输入节点的内积并求和,得到节点输出. Graph-CNN 实际上是把节点和边的特征加入进邻接矩阵,通过线性变换得到新的节点特征值.在权重参数的初始化上,由于图数据的输出依赖于卷积核的权重系数、输入节点值和邻接矩阵,高斯分布随机初始化方法可能会造成梯度爆炸和梯度消失.因此 Graph-CNN 采用 Xavier^[52]初始化和批量规范化(Batch Normalization)^[53]方法避免此情况出现.

表 2 从模型、核心思想、图种类、学习模式、激活函数、数据以及任务对空间图卷积模型进行了汇总.

3.1.3 谱分解图卷积和空间图卷积小结

GCN 的核心原理在于利用边的连接属性对原始节点信息进行聚合和整理,以便生成新的节点表示.从本质上看,谱分解图卷积和空间图卷积都是对邻居节点特征聚合,产生一个固定维度大小的输出表示,进而达到传播邻居节点信息的目的.基于谱分解的方法在数学上给出了卷积操作的严格定义和公式证明,但是缺点在于傅里叶变换对于每一个图都是唯一的,所以只能在固定的图上进行卷积运算,对于图结构依赖性很强.由于拉普拉斯矩阵中的邻接矩阵和度矩阵都是基于无向图定义的,除非将有向图转换为无向图,因此谱分解的方法并不能直接用于有向图和带权图.此外,谱分解图卷积对图中所有节点同时进行变换,计算和存储空间的要求较高,难以适应大规模图数据集的分析.空间图卷积通过规定一定顺序直接对图中的每个节点操作.由于是在节点上进行局部卷积,因此卷积核权值可以被共享.空间图卷积使用了一定的采样策略,可以并行批量化计算,降低了计算时间和存储空间.但是其缺点在于中心节点顺序需要确定,因此具有位置依赖性.邻居节点数目的选择是一个不确定的过程,不同方法构造的感受域大小一般不尽相同,这给不同网络的参数比较带来一定的困难.

表 2 空间图卷积模型汇总

模型/文献	核心理念	图种类	学习模式	激活函数	数据	任务
PATCHY-SAN ^[40]	节点中心度量 节点邻域集合收集 子图规范化	异质无向	无监督 有监督	Softmax	①②③④⑤⑥	运行时间分析 特征可视化 图分类
DCNN ^[42]	扩散卷积核	同质无向 异质无向	有监督 半监督	Softmax	①③④⑩⑫⑯⑰	节点分类 图分类
文献 ^[43]	随机游走 转移矩阵	同质有向 异质无向	—	Sigmoid	⑦⑨	回归 图分类
GraphSAGE ^[44]	特征聚集向量 聚合策略	同质无向	无监督 有监督	Sigmoid	⑧⑭⑮	节点分类
FastGCN ^[45]	蒙特卡罗近似 嵌入函数积分变换	同质无向	有监督 半监督	—	⑩⑫⑮	节点分类 图预测
文献 ^[46]	控制变量估计器 方差缩减	同质无向 异质无向	半监督	—	⑧⑩⑪⑫⑬⑮	预测
文献 ^[47]	自适应采样 蒙特卡罗近似 跳跃连接	同质无向	有监督	LeakyReLU Softmax	⑩⑪⑫⑮	节点分类 预测
CoN-GCN ^[48]	邻居节点排序 分级采样	同质无向	半监督	—	⑩⑪⑫	节点分类
LGCL ^[49]	特征压缩 宽度搜索	同质无向 异质无向	有监督	Softmax	⑧⑩⑪⑫	节点分类
SGCN ^[50]	平衡理论 平衡路径	同质无向	有监督	—	⑮⑰⑱⑲	链接预测
Graph-CNN ^[51]	异质图过滤器 Xavier 初始化 批量规范化	异质有向	有监督	Softmax Relu	③⑥⑦②③④⑤	图分类 节点分类
数据集	①MUTAG; ②PCT; ③NC1; ④NC109; ⑤PROTEIN; ⑥D&D; ⑦DPP4; ⑧PPI; ⑨Mnis; ⑩Cora; ⑪Citeseer; ⑫Pubmed; ⑬NELL; ⑭Web of Science; ⑮Reddit; ⑯Bitcoin-Alpha; ⑰Bitcoin-OTC; ⑱Slashdot; ⑲Epinions; ⑳PTC; ㉑ENZYMES; ㉒CIFAR-10; ㉓ImageNet; ㉔fMRI; ㉕Bosphorus 3D Face					

3.2 图自编码器

在深度学习领域,自编码器(Auto-encoder, AE)是一类将输入信息进行表征学习的人工神经网络。自编码器一般包含编码器和解码器两个部分,基于自编码器的 GNN 被称为图自编码器(Graph Auto-encoder, GAE),可以半监督或者无监督地学习图节点信息,如图 3 所示。

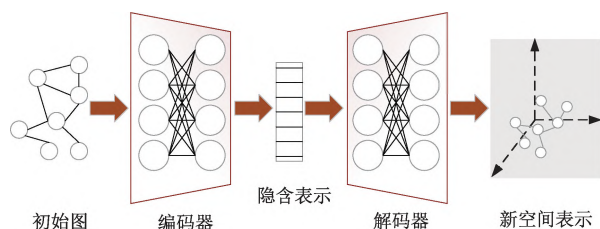


图 3 图自编码器

在图自编码器上,文献^[54]提出基于深度神经网

络的表示模型(Deep Neural Network for Graph Representations, DNGR)。DNGR 采用随机游走模型(Random Surfing Model)获取图结构信息,生成概率共现矩阵,并在概率共现矩阵的基础上计算 PPMI 矩阵。在图节点嵌入表示学习上, DNGR 设计了一个叠加去噪自编码器(Stacked Denoising Auto-encoder, SDA),输入 PPMI 矩阵学习图节点低维表示,并且输入的一部分会被随机置零以提高模型的鲁棒性。DNGR 的优点在于能学习到有向图中更多的结构信息,其生成的低维嵌入表示可以用于不同的下游任务。但缺点是忽略了图属性信息,没有将图属性和图结构信息一并纳入到模型框架中,因此图结构的轻微变化就会影响节点表示的好坏。针对节点内容信息的收集, Wang 等人^[55]提出一种边缘图自编码器(Marginalized Graph Autoencoder, MGAE)算法。其在自编码器中使用基于谱分解的图卷积网络层,整合节点属性特征和图结构信息,使

得它们之间能进行数据交互. MGAE 堆叠多层图形自编码器,以建立一个深层次的架构来学习有效的节点表示. Wang 等人认为在训练中随机噪声引起的干扰可能会提供更有效的输出表示,因此会在节点内容特征中动态地加入一些干扰项. 通过将某些特征值置为零,获得在大规模图上学习的能力. MGAE 构建了优化器以确保编码的节点属性信息和真实属性信息之间的误差最小化. 在得到每个节点的表示后, MGAE 使用谱聚类算法得到图聚类结果.

在一定条件下,自编码等价于正则拉普拉斯矩阵的奇异值分解(Singular Value Decomposition, SVD)^[56]. 因此文献^[57]提出一种 AutoGCN 模型,其模拟一阶卷积算子,使用了多个级联自编码器替代卷积矩阵,这相当于在图上构建了一个动态谱分解图卷积,在一定程度上缓解了增加网络深度而导致的参数过平滑的现象. 非线性的自编码器可以学习到更加复杂的局部图结构表达. 同时 AutoGCN 应用多任务目标函数,缓解了冷启动问题,可以更有效地学习到稀疏图中节点特征的隐含表达.

在 AE 中,由编码器产生的中间低维隐含向量表示必须要贴合真实样本,否则通过解码器解码出来的新样本表示与真实的数据难以近似. 因此变分自编码器(Variational Auto-encode, VAE)^[58]被提了出来. 变分自编码器中,编码器学到的不是样本的低维向量表示,而是低维向量表示的分布. 通过将编码器生成的中间隐含向量约束在一个正态分布上,然后在该正态分布中采样得到中间低维向量表示,并经过解码器还原出原始样本. 基于此优点,文献^[59]提出一种变分图自编码器(Variational Graph Auto-encoder, VGAE). VGAE 使用两层 Kipf 等人^[27]提出的 GCN 充当编码器,输入图的邻接矩阵 A 和节点的特征矩阵 X,学习节点隐含向量表示的均值和方差,并用隐含向量内积充当解码器. 损失函数由两部分组成:第一部分是生成图和原始图之间的距离度量差;第二部分是节点隐含向量分布和标准正态分布的 Kullback-Leibler 散度. 尽管 VGAE 能生成贴合原始图数据的表示,但是需要迭代训练图卷积网络来进行优化,因此训练时间会比较漫长. 除此之外,文献^[60]提出了半隐式图变分自编码器(Semi-implicit Graph Variational Auto-encoder, SIG-VAE). 由于图中可能存在重尾、多模式,偏斜和丰富的依赖结构, SIG-VAE 引导半隐式分层变分推理,允许对给定图隐式数据的后验分布进行更加

贴合地建模. SIG-VAE 基于半隐式变分推理^[61]和正则流形(Normalizing Flow)^[62]理论,其整体是一个层次化的变分框架,并使用伯努利-泊松链接解码器^[63]. 这可以使邻居节点共享,从而更好地对图依赖结构进行生成建模. 这种层次结构提供了更灵活的图生成范式,能更好地捕获真实世界的图属性. 与 VGAE 相比, SIG-VAE 是一个更加强大的图生成模型,具有半隐式构造所支持的更可靠推理,其派生图的隐含表示更具可解释性. SIG-VAE 不仅在图数据建模上扩展了 VAE 的灵活性,同时在节点分类、图聚类、图生成和链接预测任务的准确率上取得更好的表现. 文献^[64]提出适用于生成小图的 GraphVAE 算法. GraphVAE 核心思想是用编码器将图转为连续向量表达,通过解码器直接输出一个预先定义的最大概率完全连通图,而节点和边的属性被建模为独立的随机变量. 此外, GraphVAE 设计了图匹配算法来计算图之间的相似性.

在隐含表示的分布上,为了进一步提高图嵌入生成的有效性,采用对抗化和正则化技术的图自编码器被提了出来. Pan 等人^[65]提出了一种对抗正则图嵌入生成框架,其包含两种对抗方法:ARGA (Adversarially Regularized Graph Autoencoder)和 ARVGA (Adversarially Regularized Variational Graph Autoencoder). 该框架将图中的拓扑结构和节点属性编码成一个紧凑的表示,然后训练解码器重构图结构. 为了使学习到的图节点嵌入更具有鲁棒性, Pan 等人使用 GCN 探索图的结构和节点属性信息,将整个图编码进一个隐含空间. 在解码重构图的同时,添加对抗训练模型规范隐含节点,并学习到更具鲁棒性的图表达因式. 对抗训练模型用来区分隐含节点是来自于真正的先验分布还是图编码器产生的. 图编码器和对抗正则解码器被置于一个联合训练框架下,这样它们能够同时被优化. 文献^[66]认为假设数据服从先验的正态分布并不是一个最合理的方法,因此提出基于重启随机游走(Random Walk with Restart)^[67]的正则化技术. 在编码器和解码器中间应用随机游走和 SkipGram^[68]学习拓扑图的局部信息,从而获得节点序列,并最大化节点之间的共现概率. 在此基础上建立了 RWR-GAE (Random Walk Regularized Graph Autoencoder)和 RWR-VGAE (Random Walk Regularized Variational Graph Autoencoder)两种网络正则框架.

总体上看,对抗图自编码器使用了正则化和对抗性训练等手段. 一方面这可以让编码器从不同的

角度学习原始数据特征,产生的低维隐含嵌入能更加有效地表示原始图,并可以用于其他的下游任务中.另一方面,解码器具有更加强大的还原能力,模

型的鲁棒性进一步增强.表3从模型、核心思想、图种类、学习模式、激活函数、数据以及任务对图自编码器模型进行了汇总.

表3 图自编码器模型汇总

模型/文献	核心思想	图种类	学习模式	激活函数	数据	任务
DNGR ^[54]	随机游走 叠加去噪自编码器 PPMI 矩阵	异质无向	—	Sigmoid	①②③④⑤⑥⑦	节点聚类 可视化 相似度检测
MGAE ^[55]	堆叠图自编码 边缘化 结构内容学习	异质无向	无监督	Sigmoid Relu	③⑧⑩	图聚类
AutoGCN ^[57]	级联自编码器	异质无向	半监督	Softmax Relu	⑧⑨	节点分类
VGAE ^[59]	Kullback-Leibler 散度 期望损失 变分编码器	异质无向	无监督	Sigmoid Relu	⑧⑩⑪	链接预测
SIG-VAE ^[60]	半隐式变分推理 伯努利-泊松链接解码器	异质无向	半监督 无监督	—	⑧⑩⑪⑫⑬⑭⑮⑯	节点分类 图聚类 图生成
GraphVAE ^[64]	概率连通图 图匹配	异质无向	—	Sigmoid Relu Softmax	⑰⑱	图生成
ARGA&ARVGA ^[65]	对抗训练 图编码	异质无向	无监督	Sigmoid Relu	⑧⑩⑪	链接预测 节点聚类 图可视化
RWR-GAE & RWR-VGA E ^[66]	随机游走 正则节点表示 SkipGram	异质无向	无监督	Relu	⑧⑩⑪	节点聚类 链接预测
数据集	①20News;②Wine;③Wikipedia;④WordSim353;⑤WordSim Similarity;⑥WordSim Relatedness;⑦MC;⑧Cora; ⑨JD Finance;⑩Citeseer;⑪PubMed;⑫USAir;⑬NS;⑭Router;⑮Power;⑯Yeast;⑰QM9;⑱ZINC					

3.3 图生成网络

建模和生成图是研究生物工程和社会科学网络的基础.图生成网络(Graph Generative Network, GGN)是一类用来生成图数据的GNN,其使用一定的规则对节点和边进行重新组合,最终生成具有特定属性和要求的目标图.然而,在图上模拟复杂分布,并从这些分布中有效地采样是比较困难的.因为有些图数据具有非唯一性、高维性质,图中边缘之间存在复杂的非局部依赖性.因此不能假设所有的图数据都来自于同一个先验分布,尤其是对于异质图,模型在识别过程中必须要具有平移不变性.因此GGN着重用来解决这类问题和克服其中的难点.GGN的输入可以是节点或者边向量,也可以是给定的图嵌入表示,然后对采样的数据学习后合成各种任务所需要的图.

文献[69]提出图形生成对抗网络(Graphical Generative Adversarial Network, Graphical-GAN).

Graphical-GAN用贝叶斯网络学习随机变量之间的依赖结构,用期望传播算法(expectation Propagation)^[70]联合训练生成器和判别器. Graphical-GAN最小化了每个局部因素的差异,并估计图中的局部分歧,以处理隐式似然性. Graphical-GAN有两种实例变体:高斯混合图生成网络(Gaussian Mixture GGN)和状态空间图生成网络(State Space GGN),分别用来学习图中的离散结构和时间特征. Graphical-GAN结合了深度学习模型学习复杂隐含特征和概率模型学习数据先验知识的能力.但是在概率图模型中,模型依赖于很强的独立性假设,只能获取特定的图属性,而图语法模型(Graph Grammars Model)^[71]对图的内容语义信息限制严格,针对上述不足.文献[72]介绍了一个强大的新方法用来学习生成图.该方法聚集节点的邻域并使用GRU(Gated Recurrent Unit)^[73]更新节点的向量表示.在线性迭代地添加节点和边的过

程中,结合 MLP 中每轮输出结果的概率分布,为节点和边分配不同的生成和连接的决策序列。通过学习图上的分布可以准确地得到图的结构和属性,从而摆脱向量和序列式知识表示的限制。节点状态的初始化会利用图属性特征,同时加入一定的条件信息使得模型能灵活模拟图中的一些关系。在训练和评估中,用蒙特卡罗近似估计方法和均匀随机排序处理图样本数据,使元素的生成具有合理性和有序性。但是节点和边的顺序参数在训练和测试中是固定不变的,不能很好地适应具有不同顺序关系的图。此外,在迭代生成节点和边的过程中,决策序列过长,一方面使得概率估计变得不准,另一方面会使得训练过程变得更加困难,可能会产生梯度消失和梯度爆炸的现象。

现有的图分析方法对于图中的缺省值过于敏感,不能处理拓扑结构混杂的情况。针对此问题,文献[74]在 GAN 的基础上提出图拓扑插值器(Graph Topology Interpolator,GTI),重新构建了基于边权重的邻接矩阵,用于量度不同的边对整个图的重要程度。GTI 框架由六个层模块组成——分级识别层模块、分割层模块、GAN 层模块、再生成层模块、汇集层模块和阶段识别模块。其中分级识别层用于制定规则并决定子图是否被输入下一个模块;分割层模块用于区分非重叠的子图;GAN 层模块由反卷积生成器和卷积判别器组成,用于特征学习;再生成层模块用于构建新的子图;汇集层模块用于整合所有层的权重邻接矩阵;块识别模块用于缩减边权重矩阵大小并将图划分成包含不同边的块集合。文献[75]提出 Graph-GAN 模型,Graph-GAN 同样由生成器和判别器构成。生成器用来生成类似于真实存在的,具有边关系的节点,并用策略梯度进行优化。Graph-GAN 在生成器中使用一种全新的激活函数——Graph Softmax。Graph Softmax 具备规范化、图形结构感知和高效计算三个要素,因此能高效地捕获图拓扑结构特征。判别器用来区分输入节点是从真实的连通性分布图中选取的,还是在生成器中产生的模拟节点。Graph-GAN 构建了一个最大最小化的目标函数,使得生成器和判别器经过多轮迭代循环后达到稳定的状态。

尽管上述模型将 GAN 扩展到了图上,并提出了相应的模拟图结构和生成图数据的模型,但是仍然存在一定的缺陷和不足。文献[74]对整个邻接矩阵的操作导致计算量庞大,尤其是在大图上时间缓

慢。文献[75]虽然在链接预测和节点分类上取得较好的效果,但是在诸如化学分子式等真实世界数据集上的学习中,并没有保存其中物理或者化学意义的图范式。针对此问题,文献[76]提出模拟现实世界网络的隐式图生成模型——NetGAN。NetGAN 从邻接矩阵中生成随机游走路程,并使用有偏二阶随机游走[77]采样方式将学习过程从模拟图的拓扑结构转变为学习在图上的分布。NetGAN 构造了一个端到端的生成器和解码器,其生成器用于生成离散输出样本,解码器用来区分输入目标是真实图还是合成图。NetGAN 在训练时运用随机神经网络 Wasserstein GAN[78]框架,提出 Val-criterion 和 EO-criterion 两种提前终止策略,用来控制图生成过程的有效性。在训练完成后,新生成的图可以由一组随机游走节点的规范化共生矩阵得到。在实验中,NetGAN 除了能够显式地生成真实世界存在的网络图,在链接预测上也具有优良的性能表现,从侧面反应了其较强的泛化特性。文献[79]提出化学分子式生成对抗网络(Molecular Generative Adversarial Network,MolGAN)。MolGAN 基本框架是由一个生成器和一个判别器组成的改进型 WGAN[80]。类似于 NetGAN,MolGAN 从先验向量中采样,并传递给生成器,后者输出化学分子式的图结构表示。判别器则用来区分是来自真实数据集还是生成器产生的分子图,并引入基于 ORGAN[81]的强化学习来构建一个奖励网络,使用外部软件评估化学分子式的相关属性。为了保持训练的稳定性,MolGAN 还使用小批量鉴别方式[82]防止模型崩溃。MolGAN 避免了基于似然的方法需要图匹配或启发式节点排序的过程,有助于寻找具有特定属性的化学分子式。其模型框架如图 4 所示。

但是,NetGAN 和 MolGAN 不能生成带有标签节点的图,对于图大小也过于敏感导致扩展性不佳。此外,仅仅只是为了得到最低的判别器损失,导致生成器产生的分子式以及某些网络图等结果很单一。在涉及图数据样本多样性方面有较大的缺陷,容易发生模式坍塌(Mode Collapse)。为了避免这种现象,有时候不得不提前终止训练。针对此问题,文献[83]提出标记图生成对抗网络(Labeled-graph Generative Adversarial Network,LGGAN)。LGGAN 的生成器可以为节点生成邻接矩阵和标签,它的判别器则使用 Kipf 等人[27]提出的 GCN 结构,并加入残差连接,这样可以抽取自适应的、结构感知

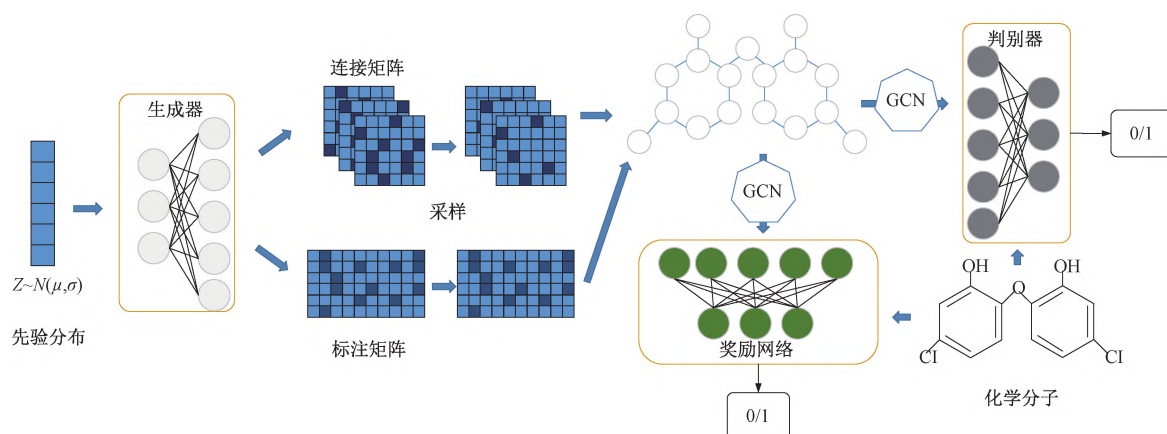


图4 MolGAN生成化学分子图

的高级图特征来识别真实图. 此外, LGGAN 探讨了三种不同的框架, 分别是标准 GAN^[6], Conditional-GAN^[84] 和 AC-GAN^[85]. Conditional-GAN 使用了图标签信息判断图真假, AC-GAN 则在使用图标签信息的基础上, 让判别器区分真假的同时判断其类别. 这样可以产生较多种类的图, 在一定程度上避免了生成器产生的图类别唯一. 针对图生成网络需要大量带有标签样本的不足, 文献[86]提出一种仅需要少量有标签图数据的半监督图生成网络——GraphSGAN. GraphSGAN 核心思想是在数据图中不同子图的低密度区域生成假样本, 这能减少这块

区域其他节点对于待分类样本的干扰, 提高分类的准确性和扩展性. 在 GraphSGAN 的输入预处理上, 先使用网络嵌入的方法生成节点的隐含嵌入表示, 之后与节点的原始特征向量拼接成一个完整的节点输入向量. 在生成器上先利用数据的高斯噪声输出假样本, 并在训练过程中使用批量规范化^[53]和权重规范化 (Weight Normalization)^[87]技巧, 全连接层中的参数通过权重规范化来约束. 另外在输入层和全连接层之后加入一个随机层, 以达到平滑的目的.

表4从模型、核心思想、图种类、学习模式、激活函数、数据以及任务对图生成网络模型进行了汇总.

表4 图生成网络模型汇总

模型/文献	核心思想	图种类	学习模式	激活函数	数据	任务
Graphical-GAN ^[69]	期望传播 高斯混合 GGN 状态空间 GGN	同质无向	无监督	Softmax Tanh Relu	⑦⑧⑨⑩⑪	图聚类
文献 ^[72]	GRU 更新 蒙特卡罗近似 均匀随机排序	同质无向 异质无向	—	Sigmoid Softmax	①②③	图生成
GTI ^[74]	拓扑插值	异质无向	无监督	leakyReLU	③④⑤⑥⑦⑧⑨⑩	图生成
Graph-GAN ^[75]	Graph Softmax	异质有向	—	Sigmoid Softmax	⑪⑫⑬⑭⑮	节点分类 链接预测 兴趣推荐
NetGAN ^[76]	有偏二阶随机游走 Wasserstein GAN 规范共生矩阵	异质无向	—	Softmax Tanh	⑯⑰⑱⑲⑳㉑	图生成 链接预测
MolGAN ^[79]	ORGAN 强化学习 小批量鉴别	异质无向	—	Sigmoid Softmax Tanh	㉒㉓	图生成
LGGAN ^[83]	残差连接 Conditional-GAN AC-GAN CT-GAN	异质无向	—	—	⑰⑱㉔㉕	图生成 图分类
GraphSGAN ^[86]	假样本生成	异质无向	半监督	Softmax	⑰⑱㉖	节点分类 实体抽取
数据集	①Cycles; ②Trees; ③BA; ④ER; ⑤WS; ⑥Kronecker; ⑦Facebook; ⑧Wiki-vote; ⑨RoadNet; ⑩P2P-Gnutella; ⑪arXiv-AstroPh; ⑫arXiv-GrQc; ⑬BlogCatalog; ⑭Wikipedia; ⑮MovieLens; ⑯Cora-ML; ⑰Cora; ⑱Citeseer; ⑲Pubmed; ⑳DBLP; ㉑POL. BLOGS; ㉒QM9; ㉓GDB-17; ㉔ENZYMES; ㉕PROTEINS; ㉖DIEL; ㉗MNIST; ㉘SVHN; ㉙CIFAR10; ㉚CelebA; ㉛3D Chairs					

3.4 图循环网络

图循环网络(Graph Recurrent Network, GRN)是最早出现的一种 GNN 模型. 相较于其他的 GNN 算法, GRN 通常将图数据转换为序列, 在训练的过程中序列会不断地递归演进和变化. GRN 模型一般使用双向循环神经网络(Bidirectional RNN, Bi-RNN)和长短期记忆网络(Long Short-Term Memory Network, LSTM)作为网络架构.

文献[9]将 RNN 用于学习图的拓扑结构和节点分类问题, 每个节点的状态向量取决于参数化的转移函数. 该转移函数以上一轮的节点状态、节点标签、邻居节点标签作为输入, 再将得到的当前状态向量和该节点标签整合得到依赖输出向量. 对于有向图, 则是根据边的方向决定是否将邻居节点聚集给中心节点. Scarselli 等人^[12]运用信息扩展的思想, 通过函数 $\tau(G, n) \in R^m$ 将图 G 和某个节点 n 映射到 m 维的向量空间上. 图的一个节点对应一个分析单元, 通过多组单元处理对给定的图进行分析. 这些单元整合节点特征和邻居节点特征, 相互间进行信息交互并更新各自的状态直到一个稳定点, 最后根据单元的不同状态决定是否激活输出. 在参数估计时, 对网络权重的雅可比矩阵施加惩罚项, 并使用 Almeida-Pineda^[88-89]算法训练模型.

文献[9, 12]都是迭代地激活和计算节点状态, 在传播函数上反复应用压缩映射使得节点状态最终趋向于稳定. 文献[90]在 Scarselli 等人^[12]前期工作的基础上, 提出门控图序列神经网络(Gated Graph Sequence Neural Network, GGS-NN). GGS-NN 将 GRU 优化技术加入 Scarselli 等人的模型中, 通过编码图序列特征, 产生隐含表示向量. GGS-NN 用节点注释初始化网络的隐层状态, 图中不同的边决定节点信息发送的方向和内容. 在信息传播后, 节点根据邻域的变化更新自己的状态. 此外, 设置了一个全局状态表示, 该全局状态的聚合输出值取决于所有节点的状态. GGS-NN 使用两种训练设置: 一是在给定所有节点注解表示后训练; 二是在只给定一个节点注解表示上进行端到端的训练. 相对于 LSTM 等纯序列模型, GGS-NN 模型具有良好的归纳偏差性质. 在程序验证等人工智能推理的图任务上具有灵活的学习能力和泛化能力. 文献[91]提出门控图变换神经网络(Gated Graph Transformer Neural Network, GGT-NN). GGT-NN 在 GGS-NN 基础上添加新的节点状态更新函数, 根据节点对的状态和外部输入更新边属性, 利用注意力机制来挑选不同

的节点并输出一个全局图. 针对问答任务, GGT-NN 寻找输入问题序列中的隐藏状态用来构造内部图, 并构建一维细胞自动机和具有四个不同状态的图灵机来寻找问题中的潜在的规则. 相较于 GGS-NN, GGT-NN 在对话问答任务^[92]上表现出更佳的性能. 但缺点是时间复杂度和空间复杂度比较大, 随着对话问题任务复杂性加深, 所需计算资源增长程度也会更加剧烈. You 等人^[93]提出了一个深度自回归模型——GraphRNN. 在对图的结构进行最小假设的情况下, GraphRNN 将多组有代表性的图转化成序列并学习, 近似地描述了图的任意分布. GraphRNN 构造了两个学习层级: 运作在图层面的 RNN 获取图的状态信息并生成新的节点, 并使用宽度优先搜索动态扩展节点, 然后对节点排序; 运作在边层面的 RNN 用来给新节点生成边, 并且会严格限定新生成边的数量, 避免形成稠密连通图. 为了定量地评估 GraphRNN 的性能, You 等人引入度分布、聚类系数分布和最大平均值差异三个指标, 用于判断合成图的质量好坏.

结构化序列可以表示视频中的帧、传感器网络上的时空测量值, 以及自然语言词汇图上随机游走路径. 针对这种图空间结构随时间变化的问题, Seo 等人^[94]提出一种能预测结构化序列数据的深度学习模型——图卷积递归网络(Graph Convolutional Recurrent Network, GCRN). GCRN 是标准 RNN 对任意图结构化序列数据的应用. GCRN 模型由 Defferrari^[25]提出的 Chebyshev 卷积算子和 RNN 组成, 其中 Chebyshev 卷积算子用于提取图节点特征, RNN 用于识别空间结构和动态模式. Seo 等人研究了两种可能的 GCRN 体系结构, 一种是用叠加图卷积提取特征, 提取到的特征输出给 LSTM 进行序列学习; 另一种是将 RNN 中的矩阵乘法换成了图卷积操作, 完成序列学习. Seo 等人用两个实际问题测试这两种结构: 一是变化的 Mnist 数据预测; 二是基于 Penn Treebank 数据集的自然语言建模. 实验表明, 同时利用图的空间信息和时间信息可以提高任务学习的精度和速度. 在实际应用中, 图的拓扑结构和节点属性都会随着时间的推移而变化, 即静态图会变成时空属性图. 时间属性图中的节点分类问题有两个方面的挑战. 一是很难有效地修改时空语境信息. 二是由于时间维度和空间维度具有互相纠缠的性质, 为了学习一个目标节点的特征表示, 区分不同邻域、不同时间段等各种因素的相对重要性过于困难. 因此文献[95]提出了一种时空注意递归

网络 (Spatio-temporal Attentive Recurrent Network, STAR) 模型来解决上述问题. STAR 算法通过采样和聚集局部邻居节点来提取邻域向量表示, 之后将邻域表示和节点属性输入到时空 GRU 中, 学习图时空上下文信息. 在此基础上, 用时空注意力算

法对中间隐含向量的重要性进行分析. 在真实数据集上的大量实验证明了 STAR 模型对于时空图语义分析的有效性.

表 5 从模型、核心思想、图种类、学习模式、激活函数、数据以及任务对图循环网络模型进行了汇总.

表 5 图循环网络模型汇总

模型/文献	核心思想	图种类	学习模式	激活函数	数据	任务
文献 ^[9]	迭代激活节点 压缩映射	任意	—	Sigmoid	—	图连通性 分类 图生成
文献 ^[12]	迭代激活节点 压缩映射	任意	有监督	Sigmoid	—	子图匹配 突变检测 网页排序
GGs-NN ^[90]	门控递归单元 全局状态表示	有向	—	Sigmoid Softmax	—	文本推理 最短路径 欧拉环路
GGT-NN ^[91]	全局图 细胞自动机 图灵机	有向	有监督	Sigmoid Softmax Tanh	—	文本推理 规则挖掘
GraphRNN ^[93]	序列学习 宽度优先搜索	异质无向	—	Sigmoid Relu	①②③④⑤	图生成
GCRN ^[94]	图时空学习	同质无向 异质无向	—	Sigmoid Tanh	⑥⑦⑧	视频预测 自然语言建模
STAR ^[95]	时空 GRU	异质无向	有监督	Softmax Tanh LeakyRelu	⑨⑩⑪⑫	节点分类 可视化
数据集	①ER; ②Grid; ③BA; ④Protein; ⑤Ego; ⑥Moving-Mnist; ⑦Penn; ⑧Treebank; ⑨Brain; ⑩Reddit; ⑪DBLP-5; ⑫DBLP-3					

3.5 图注意力网络

注意力机制可以让一个神经网络只关注任务学习所需要的信息, 它能够选择特定的输入^[96]. 在 GNN 中引入注意力机制可以让神经网络关注对任务更加相关的节点和边, 提升训练的有效性和测试的精度, 由此形成图注意力网络 (Graph Attention Network, GAT).

根据注意力机制的不同, 可以分为自注意力、多头注意力、层次注意力等. 在自注意力使用上, 文献^[97]首次提出将自注意力机制应用至 GNN. 在空间图卷积的基础上, 该方法将屏蔽的自注意力层堆叠到空间图卷积的聚合函数中. 注意力互相关系数的计算公式如下:

$$e_{ij} = a(\mathbf{W}\mathbf{h}_i, \mathbf{W}\mathbf{h}_j) \quad (13)$$

其中 $a(\cdot)$ 表示注意力计算函数, 这里使用单层前向传播神经网络, \mathbf{W} 代表权重矩阵. 该公式表达了邻居节点 j 对节点 i 的重要程度. 其隐式地为中心节点的邻居节点分配权重因子, 具有不同度的节点会被指定不同的权重参数. 这表示邻居节点对中心节

点的不同重要性, 可以理解为空域上的结构信息. 该方法的好处在于不需要预先知道图的拓扑结构, 并且可以并行地计算节点-邻居对, 同时也避免了诸如求逆等代价高昂的矩阵运算. 但是 GAT 矩阵乘法的操作仅限于二阶张量, 在一些具有多个图的数据集中, 这无疑会限制注意力网络层的批处理能力, 导致 GPU 的处理性能会受到影响. 并且某些图的邻域可能会高度重叠, 在并行化过程中会增加额外的计算成本.

此外, 文献^[98]在 GAE 基础上引入自注意力机制, 提出了一种无监督表示学习的图注意力自编码器 (Graph Attention Auto-encoder, GATE). GATE 在编码器和解码器层同时引入自注意力, 重新构建节点属性和图结构信息. 在编码器中, 将节点属性作为节点的初始表达. 每一层都使用注意力机制, 根据邻域中节点之间的相关性来生成新的节点表示. 在解码器中, 则反转编码重建节点属性. 解码器的每一层逆向操作都与编码器正向操作对应. 此外, 对节点表示进行了正则化以重构图结构. 由于

GATE 框架中不需要预先知道图的结构,因此可以应用于归纳学习. GATE 在基准数据集的节点分类任务中表现优秀,对于直推式和归纳式任务都具有较强的竞争力. Busbridge 等人^[99]提出一种比较特殊的关系型图注意力网络(Relational Graph Attention Network, RGAT),以关系型图卷积网络(Relational Graph Convolutional Network, RGCN)^[100]为基础,将注意力机制扩展到关系型图域中,并发展成两种变体:内部关系图注意力机制(Within-Relation Graph Attention, WIRGAT)和交叉关系图注意力机制(Across-Relation Graph Attention, ARGAT). WIRGAT 和 ARGAT 使用了附加注意力机制和乘性注意力机制. WIRGAT 用来聚集节点的邻居并使用注意力系数作用于节点聚合矩阵. ARGAT 用来阐述不同节点表示上实现单个概率分布和关系的重要性,并将图局部先验知识进行编码. 由于在实际表现中, RGAT 在归纳式图分类任务上产生了边际效应,因此 Busbridge 等人使用完全累积分布函数(Cumulative Distribution Function, CDF)搜索合适的超参数,并做了统计假设检验.

多个独立注意力的计算可以作为一个集成的作用,防止过拟合. 同样地,为了更好地利用邻居节点信息, Zhang 等人^[101]使用了多头注意力(Multi-head Attention Mechanism)^[96],提出一种门控注意力网络(Gated Attention Network, GaAN),以子网卷积的方式,构建 GGRU(Graph Gated Recurrent Unit). Zhang 等人认为在应用多注意力机制的同时,滥用注意力会在子空间内聚集不够重要的关系信息给单个节点,最终导致预测准确率的降低. 因此设计了一个软门(Soft Gate)控制多个注意力端,探究不同注意力端在子空间内生成的节点特征对于网络的重要性程度,从而构造了一种关键价值关注机制(Key-value Attention Mechanism). 在层次注意力机制上, Do 等人^[102]提出一种多标签学习图注意力模型(Graph Attention Model for Multi-Label Learning, GAML). 在多标签分类问题中,分类的准确性和可解释性是一个难点,这与标签和子图的未发掘关系之间存在某种关联. 因此 GAML 将所有标签视作一个个的节点,称为标签节点. 普通图节点称作数据节点,然后将标签节点与数据节点一起作为神经网络的输入端. 之后 GAML 运用消息传递算法(Message Passing Algorithm)^[100,103-104]构建了 MPGNN(Message Passing Graph Neural Network). 在更新每个节点的局部结构过程中,

MPGNN 会在标签节点和数据节点之间传递子结构消息,并且采用一种分层注意力机制来探究不同子结构消息的重要性. 该分层注意力机制使用中间注意力因素(Intermediate Attentional Factors)保存计算结果,作用于标签节点,使其可以抽取最相关的子结构来更新状态,该状态随后用来预测标签节点最合适类别. 在状态更新步骤中,节点状态更新公式为

$$\mathbf{x}_i^t = g^u(\mathbf{x}_i^{t-1}, \mathbf{m}_i^t) \quad (14)$$

其中 \mathbf{x}_i^t 是第 t 轮节点 i 的状态向量, \mathbf{m}_i^t 是节点的消息向量. $g^u(\cdot)$ 表示网络框架,这里 Do 等人使用了 Highway Network^[105]用来获取数据之间的长距离依赖特征,更方便地表达标签和输入子图在不同尺度上的关系,使得 GAML 具有优良的表现. 更重要的是 GAML 可以提供较为直观的可视化结果,这有助于了解不同标签子结构关系以及解释模型的内在原理.

在对图注意力网络模型进一步研究中,噪声干扰会使得模式分析局限在图上一小块范围. 针对此问题,文献^[106]提出图注意力模型(Graph Attention Model, GAM). GAM 使用注意力机制引导的随机游走采样节点,强迫游走区域限定在任务相关区域,并忽视无用数据节点,从而捕获图结构中关键的区域连通信息. GAM 会动态选择有效的节点序列,在不需要对整个图进行全局分析的情况下,其计算复杂度和存储开销也大幅度缩短. 此外,为了将基于注意力的 GNN 推广到更大规模,更加复杂并包含噪声的图中, Knyazev 等人^[107]研究 GNN 中节点之间的注意力机制以及影响其有效性的因素. Knyazev 认为在一些分类任务中,注意力的引入可以使得模型的表现大幅度提升. 但是在实践中达到最优效果却不是那么容易,往往需要有监督地进行注意力学习. 受到图同构网络(Graph Isomorphism Network, GIN)^[108]和 Kipf 等人^[27]研究工作的启发, Knyazev 等人设计了 ChebyGIN 网络,并推出两个简洁的图推理任务用来验证图注意力的有效性. 第一个任务是用不同的颜色标记图中的不同节点,第二个任务是计数图中三角形的数量. 在基本事实受控的环境中进行上述两个任务,这能保证不受到其他因素影响,从而探索注意力机制在神经网络中的深层作用. 在池化层中, Knyazev 设定了一个阈值,只有注意力系数大于给定阈值的节点才会被保留. 尽管在舍弃节点时,图结构发生了改变,并且可能会出现孤立的点. 但是在训练中,模型为相近的节

点预测了相同的系数值,因此整个局部邻域都可以被合并或者丢弃.在多个数据集上的测试表明,图分类结果的准确性和注意力系数的初始化有很大关系.特别是在无监督学习中,如果注意力系数初始即

最优,那么将会使得模型的准确率有较大的提升.

表 6 从模型、基准网络、注意力机制、图种类、学习模式、激活函数、数据以及任务对图注意力网络模型进行了汇总.

表 6 图注意力网络模型汇总

模型/文献	基准网络	注意力机制	图种类	学习模式	激活函数	数据	任务
文献 ^[97]	GCN	自注意力	异质无向	有监督	Sigmoid LeakyReLU	①②③④	节点分类
GATE ^[98]	GAE	自注意力	异质无向	无监督	Sigmoid Softmax	①②③	节点分类
RGAT ^[99]	GCN	附加注意力 乘性注意力	异质无向	半监督	Sigmoid LeakyReLU	⑤⑥⑦	节点分类 图分类
GaAN ^[101]	GCN	多头注意力	异质无向	—	Sigmoid Softmax	④⑩	节点分类
GAML ^[102]	MPGNN	分层注意力	异质有向	有监督	Sigmoid Relu	⑧⑨⑩⑪	多标签分类 图预测
GAM ^[106]	RNN	注意力引导 随机游走	无向有向	有监督	Softmax	⑭⑮⑯⑰⑱	图分类
ChebyGIN ^[107]	GCN	标准注意力	异质无向	弱监督	Softmax	⑳㉑㉒㉓㉔㉕	图分类
数据集	①Cora;②Citeseer;③Pubmed;④PPI;⑤AIFB;⑥MUTAG;⑦Tox21;⑧Media_mill;⑨Bookmarks;⑩Corel5k; ⑪NUS-WIDE;⑫TACRED;⑬Semeval-10;⑭HIV;⑮NCI-1;⑯NCI-33;⑰NCI-83;⑱NCI-123;⑲Reddit; ⑳Triangles;㉑Mnist;㉒Proteins;㉓D&D;㉔Collab;㉕Colors						

4 与网络嵌入和图核方法联系比较

通过前文的归纳和分析,从总体上看,图神经网络可以分为五类:图卷积网络、图自编码器、图生成网络、图循环网络和图注意力网络.每种图神经网络都有自己对图结构数据处理的一套算法和体系,其中的原理和适用的范围也有一定差别.当然它们之间不是相互孤立和排斥的,例如文献[59,65]的图自编码器中包含图卷积层,文献[91,95]的图循环网络为了图序列学习更有效,也会加入注意力模块.而图注意力网络也大多以其他图神经网络框架为基础,构建合适的节点、边以及图注意力网络层.因此在实际操作当中,需要根据图的分布和特征信息,以及任务的实际需求,选择合适的图神经网络,来更加有效地学习图结构数据.表 7 是 GNN 机制、优点、缺点、适用范围及实现成本汇总表.

从总体上看,上述 GNN 的算法和框架都利用到了图拓扑结构信息和节点属性信息.GCN 通过卷积操作,将节点的特征聚合起来,达到信息传播的目的.空间图卷积和谱分解卷积分别从两个不同的角度阐述了上述过程.空间图卷积把图看作图片,节点当作图片中的像素点,节点属性值的个数相当于像素点通道个数.通过在节点上直接定义卷积操作来对图进行卷积.在空间图卷积中,PATCHY-SAN

利用图标签等信息通过排序函数对邻居节点排序,并用了一维卷积网络对排序后的节点进行卷积运算.因此 PATCHY-SAN 对于图拓扑结构和节点属性信息的捕获依赖于排序函数.只要排序函数不发生变化,PATCHY-SAN 就具有图上的平移不变性.而在 DCNN 中,它利用概率转移矩阵 P_i 对每个节点的邻居特征信息求加权和, P_i 为度规范化邻接矩阵的幂矩阵.随着幂指数增加,远距离连接的节点的权重值会逐渐变小.所以在 DCNN 中,节点主要获取的是图中近距离的拓扑结构和属性,越近的邻居节点对中心节点影响也越大.GraphSAGE 则是在中心节点的每一阶邻居节点中进行采样处理,对于第 k 个网络层的节点向量 \mathbf{v} ,其计算函数式可以简写为 $\mathbf{h}_v^k = \sigma(\mathbf{W}^k \odot \mathbf{h}_v^{k-1} || f(\mathbf{h}_u^{k-1}, \forall \mathbf{u} \in \mathcal{N}(\mathbf{v})))$.其中 $\mathcal{N}(\mathbf{v})$ 是节点 \mathbf{v} 的随机采样邻居节点集合. $f(\cdot)$ 代表聚合操作, $||$ 代表向量拼接操作, \odot 是向量内积, \mathbf{W}^k 是网络层参数矩阵, $\sigma(\cdot)$ 为激活函数.GraphSAGE 没有直接使用邻接矩阵,而是使用邻居节点采样,因此某些邻居节点会被舍弃,所以它每一阶的邻居不一定是完整的.在谱分解图卷积中,从最开始的利用拉普拉斯矩阵特征向量做卷积运算,经过 Defferrar 等人^[25]和 Kipf 等人^[27]的发展,在使用切比雪夫多项式简化后,谱分解卷积具有了局部连接性和参数共享性,运算复杂度也大为下降.每一次图卷积的过程相当于将更高一阶的邻居特征

聚集起来,使得节点的属性信息通过图拓扑连接结构由近及远地传播到全图。

表 7 GNN 机制、优点、缺点、适用范围及实现成本

GNN 种类	机制	优点	缺点	适用范围	实现成本	
GCN	谱分解图卷积	拉普拉斯矩阵特征分解	参数共享	拉普拉斯矩阵唯一,已训练图结构不能应用于其他图;训练时要将全图加载进内存	固定的小规模图结构	节点数目越多,图规模越大,训练成本越高
	空间图卷积	在图上直接卷积,定义空间近邻节点卷积权重共享操作	参数共享	中心节点、感受域、聚合函数不确定,相互制约依赖	较大规模的图结构	空间图卷积操作难度低,较易实现
GAE	编码器将图转为低维连续表示,解码器重构图的结构和属性	提取最具有代表性图信息;缩减输入量,中间隐含表示可用于下游图任务	需要手动设置隐层维度;多为无监督学习,比有监督性能差	小规模图结构	取决于编码器对原始图降维程度	
GGN	交替生成节点和边,或者对抗训练生成完整图表示	对图结构和属性具有强大的模拟学习能力	对噪声敏感,需要精确的图标签信息;图质量检查要求专家级知识;可能发生模式坍塌	领域相关的图生成任务	取决于相关领域的图复杂程度	
GRN	将图转为序列,用 Bi-RNN、LSTM、GRU 等循环神经网络训练	迭代压缩节点信息,能学习图在时间维度上的特征	具有顺序依赖性	具有时空特征的图	不能并行化,实现成本较高	
GAT	引入注意力机制,关注对任务有影响的邻居节点信息,从而分配不同的权重	鲁棒性和可解释性更强;无需全图信息	增加了额外计算邻居节点信息的时间和内存消耗,注意力的作用跟网络初始化有关	归纳式图任务	可并行化,适用于较大规模的图	

GCN 的方法比较依赖于原始的邻接矩阵,一般节点间的连接权重就是节点的度或者边上的权值。但是在 GAT 中,则是利用注意力机制对图拓扑结构和节点属性进行学习和整合。在 GAT 中,第 k 个网络层的中心节点 v ,其表达式可简写为 $\mathbf{h}_v^k = \sigma(\sum_{u \in \mathcal{N}(v)} \alpha_{vu} \mathbf{W}^{k-1} \mathbf{h}_u^{k-1})$ 。其中 α_{vu} 为中心节点和邻居节点的权重参数,由神经网络学习得来。一般地,谱分解图卷积给中心节点 v 的邻居节点 u 分配的权重 α_{vu} 与节点的度 $\deg(v)$ 有关,例如在 Kipf 等人^[27]的方法中 $\alpha_{vu} = \frac{1}{\sqrt{\deg(v)\deg(u)}}$ 。而对于 GAT 来说,

每个中心节点与邻居节点的权重 α_{vu} 应该由神经网络计算得出。这样可以使得与任务相关的节点获得更大的权重,让模型能够捕获更多关键信息。而在 GRN 中,节点和图都被视作序列输入到 RNN 中。从节点向量生成来看,每一步的输入包括上一步的状态、中心节点、邻居节点和节点标签等。由于 RNN 具有“遗忘”和“记忆”的功能,因此在每一步计算中会隐式地保留或者舍去部分节点信息,最终生成的节点向量反映了图的局部结构和属性。

在 GNN 的损失函数中,一般只利用到节点标签信息进行计算。但由于聚集操作的存在,预测标签也包含了其他邻居节点标签信息,因此模型隐式地判断了对图拓扑结构信息的利用程度。除此之外,GAE 和 GGN 还重构了图邻接矩阵,通过矩阵内积的方式显式地计算了模型对于图拓扑结构的掌握程度。GAE 和 GGN 也具有一定的相似性,其中一部分算法的目标不是生成固定具体的节点向量或者邻接矩阵表示,而是生成它们的概率分布,因此模型具有更好的拟合和生成数据的能力。

不难看出,GNN 对图结构数据处理的有效性来源于对图拓扑结构和节点属信息的利用程度。因此 GNN 中或多或少、或隐或显地构造了处理上述信息的聚集和融合函数,简称聚合函数或者聚合方式。在谱分解图卷积中,拉普拉斯矩阵是连接图拓扑结构和节点属性信息的“桥梁”,它将两种信息融合在一起,并通过参数学习获得节点嵌入向量表示。而在空间图卷积中,中心节点度量和感受域大小的确定为聚合函数的输入对象和操作范围奠定基础。在 GraphSAGE 中,作者提出了取平均值、取最大值和 LSTM 变换处理三种基础聚合方法,取得了良好的

效果. 之后一些学者还从其他的角度发展了聚合函数, 例如 LGCL 的子图训练法、SGCN 的符号边组合法等, 进一步补充了对于图中信息的利用方式. 在 GAE 中, 为了在降维中减少信息损失, DNGR 采用了随机游走、概率共现矩阵等聚合方式. MGAE、AutoGCN、VGAE 等没有直接构造聚合函数, 而是直接使用 GCN 编码图节点, 因此这些模型的聚合函数与 GCN 相同. 在 GGN 中, 聚合方式多种多样. Graphical-GAN 将节点视作变量, 用贝叶斯网络计算节点依赖结构, 是一种隐式的聚合. 文献[72]使用 GRU 进行节点特征迭代聚集, 这与 GRN 有些类似, 区别在于文献[72]用 GRU 聚集的节点表示是为了模型后续能够生成新的节点和边. 除此之外, NetGAN 使用有偏随机游走构建节点序列, 该聚合方式除了能捕获局部图结构, 也能捕获一定全局图结构. 文献[75]先通过宽度优先搜索构建节点二叉树, 再使用 Graph Softmax 分配节点间的权重, 这种聚合方式有利于给生成器产生有效的概率分布, 并具有结构感知以拟合正确的节点度分布. GRN 的聚合操作来自于 RNN 网络信息前向传播过程的计算. 从多个图序列来看, GRN 会留下有用的节点信息, 对于无效的节点信息则选择“遗忘”. 在 GAT 中, 聚合函数来自于各种各样的注意力机制. 通过在节点、边上施加不同的注意力机制, 聚合后的节点表示在质量上得到了一定的提升. 在保持信息完整性的前提下对任务相关的关键节点以及边给与更大的权重, 模型的可解释性更强.

除了 GNN 之外, 网络嵌入和图核方法也是两种比较重要的图学习方法. 网络嵌入的中心思想是通过生成映射函数将网络中的每个节点转换为低维度的隐含表示, 所学习到的特征表示可以用作基于图的各种任务^[109]. 传统的网络嵌入有矩阵分解、随机游走和深度学习三种方法. 网络嵌入的面向对象主要是图中的节点, 将其转为低维表达后, 图结构信息有了一定程度的压缩. 由于为每个节点生成单独的嵌入表示, 节点之间的参数不会得到共享, 参数会随着节点增长而增加. GNN 框架一般较为显式地提取特征表示, 多个节点特征在聚合后直接被输入, 较为完整地保留了图拓扑结构信息和内容属性信息. 在 DeepWalk^[110]中, 算法利用随机游走为每个节点生成序列, 类似于深度优先搜索为节点提取上下文信息. 而在 GNN 中, 邻居节点采样类似于宽度优先搜索, 聚集起来后在形式上相当于一个子图. 在文献[66, 76, 86]的图分析问题中, 也不乏将网络嵌入的

输出通过拼接整合节点特征向量后作为 GNN 的输入, 以期获得更好的训练效果.

图核方法被描述成定义在图结构空间上的半正定函数, 该函数将图映射到希尔伯特空间, 用支持向量机 (Support Vector Machine, SVM) 等核函数在映射空间内进行标准分类处理^[111], 是解决图分类问题的传统算法. 针对标签图、权重图、有向图等不同类型的图结构存在不同的图核, 主要有随机游走核^[112]、限制大小子图^[113]和基于子树模式^[114]三种代表性方法. 图核方法易于训练, 且具有可证明的理论保证, 但核函数往往依赖于手工组合图特征, 操作复杂且耗时长, 实际表达能力受到限制. 其唯一任务在于探寻相似图结构, 用于不同图之间的匹配, 是一种有效的图结构相似度量方式. 相较于图核方法, GNN 能将提取的图表示直接用于图分类. 对于节点、边、图等特征提取在 GNN 中全部由神经网络端到端地自动完成, 可以完成分类、预测、图生成等更广泛的任务和目标.

表 8 是 GNN 主要算法的实验数据, 有 Cora、Citeseer、Pubmed 三个引文网络数据集, 以及蛋白质数据集 PPI 和知识图谱数据集 NELL. 横向对比了节点分类、图聚类、链接预测和节点聚类四个基准任务. 在节点分类任务中, 谱分解 GCN、空间 GCN、GGN 和 GAT 的典型算法的准确率都能保持在 61% 以上. 在 Cora 数据集上, 相较于 DeepWalk^[110], Chebyshev^[25]模型提升了 14%; 文献[33]模型提升了 16.3%. 对比其他网络嵌入算法, GNN 在 Citeseer、Pubmed 数据集上提升也比较显著. 这表明在图结构数据处理上, 谱分解 GCN、空间 GCN、GGN、GAT 等 GNN 模型要明显优于网络嵌入算法.

5 应 用

由于 GNN 能较好地学习图结构数据的特征, 因此在许多图相关的领域有着广泛的应用. 若按照应用中图的层次结构划分, 则大体可以分为节点、边和图层面. 在节点层面, 常见的有节点分类、节点聚合、节点表示学习. 在边层面, 则有边分类、边聚类以及链接预测. 在图层面, 图分类、图生成、子图划分、图相似度分析等应用较为广泛. 按照图的种类划分, 可以分为引文网络、社交网络、交通网络、图像、化合物分子结构、蛋白质网络等. 按照应用领域划分, 可以分为自然语言处理、图像处理、轨迹预测、物理化学和药物医学等. 为了方便说明和阐述, 本文从

GNN 的主要应用领域这一角度出发,对近年来出现的 GNN 应用实例进行分类归纳。

表 8 GNN 主要算法在不同数据集上的任务分析

任务类型	模型	数据集					类别
		Cora	Citeseer	Pubmed	PPI	NELL	
节点分类	DeepWalk ^[110]	67.2%	43.2%	65.3%	—	58.1%	网络嵌入
	SemiEmb ^[115]	59.0%	59.6%	71.1%	—	26.7%	
	Planetoid ^[116]	75.7%	64.7%	77.2%	—	61.9%	
	文献 ^[27]	81.5%	70.3%	79.0%	—	66.0%	谱分解 GCN
	Chebyshev ^[25]	81.2%	69.8%	74.4%	—	—	
	CayleyNets ^[26]	81.9%	—	—	—	—	
	文献 ^[33]	83.5%	72.6%	80.0%	—	74.2%	空间 GCN
	GraphSAGE ^[44]	—	—	—	61.2%	—	
	文献 ^[46]	82.0%	70.9%	78.7%	97.8%	64.9%	
	LGCL ^[49]	83.3%	73.0%	79.5%	77.2%	—	GGN
	GraphSGAN ^[86]	83.0%	73.1%	—	—	—	
	文献 ^[97]	83.0%	72.5%	79%	97.3%	—	
	GaAN ^[101]	—	—	—	98.7%	—	GAT
	GATE ^[98]	83.2%	71.8%	80.9%	—	—	
	MGAE ^[55]	68.1%	66.9%	—	—	—	
图聚类	NetGAN ^[74]	68.5%	—	—	—	—	GGN
链接预测	RWR-GAE &	R-G	92.7%	91.5%	96.3%	—	GAE
	RWR-VGAE ^[66]	R-V	92.5%	92.4%	95.2%	—	
节点聚类	RWR-GAE &	R-G	66.9%	61.6%	72.6%	—	GAE
	RWR-VGAE ^[66]	R-V	68.5%	61.3%	73.6%	—	

5.1 图神经网络与自然语言处理

在自然语言处理中,较为常见的任务有词嵌入、实体和关系抽取、阅读理解、对话问答等。在许多情境下,文本中的单词和句子一般可以用向量或者序列来表示,但是信息的表达在一定程度上被压缩。面对更加复杂的文本理解和阅读问题,这种嵌入表示无法在解释性和准确率之间达到平衡。通过对输入文本中的元素重构为节点、边以及图表达,GNN 可以构造各种语义图或者依赖图,在此基础上可以进行更加有效的模式分析和推理。

词嵌入是自然语言处理的基础部分,词向量的好坏直接影响了诸如情感分析、句法分析等下游任务的精准度。目前大部分基于神经网络的词嵌入工作都是利用的词的顺序上下文,但是会忽略句子的语法特征。而少部分利用词句法上下文的模型会导致词汇量的爆炸式增长。针对此问题,文献[117]提出基于 GCN 的词嵌入生成模型。该模型分为两个部分,SynGCN 和 SemGCN。在不增加词汇表大小的情况下,SynGCN 将单词的上下文语法依赖信息添加进词嵌入中。SemGCN 框架则用于整合诸如同义、反义等不同的语义知识,用来增强所学习的词嵌入表示。由于 GCN 同时整合单词的语法和语义知识,因此产出的词向量具有较高的质量,在词语相似度、概念分类和词语类比三个任务上均达到最优。

关系抽取是自然语言处理中另外一个重要的子领域,目的是从文本中识别实体并抽取实体之间的语义关系,从而支持问答、搜索、推理等上层应用。在传统方法中,依赖树具有丰富的结构化信息,因此通常会利用依赖树进行关系抽取。但是基于规则的硬剪枝策略无法很好地选出语句依赖结构中的关键信息,并忽略依赖树中的无关信息。针对此问题,文献[118]提出注意力引导的图卷积网络(Attention Guided Graph Convolutional Network, AGGCN)。AGGCN 由多个相同的块组成,每个块分成注意力引导图卷积层、密集连接层和线性联合层。在输入神经网络前,将原始依赖树转变为全连接的带权图,权重表示句子中单词之间的关系影响因子。每个块自动学习相关的子结构。对于关系抽取任务来说,该模型可以理解作为一种软修剪的方法。在大规模跨句多元关系抽取中的实验结果表明,AGGCN 能够更好地利用依赖树的信息,提升关系抽取的准确率。另外,文献[119]提出了一种学习命名实体和关系抽取的模型 GraphRel。GraphRel 的关系抽取分为两个阶段。在第一阶段,GraphRel 堆叠多个 Bi-LSTM 句子编码器和 GCN 依赖树编码器,自动抽取句子和词汇特征,标记实体中的单词并对关系三元组进行预测。第二阶段,考虑关系三元组之间的相互作用,Graphrel 加入关系加权 GCN,在实体损失和关系损

失的双重约束下,沿着依赖链提取实体节点隐藏特征.同时建立一个新的具有关系加权边的全连通图.第二阶段的 GCN 有效地考虑了实体和关系之间的相互作用,其中的线性结构和依赖结构都用来提取文本的序列特征和区域特征,极大地提高了对重叠关系的区分能力.由于 GNN 一般只能在预先给定的图上对关系进行抽取,这无疑限制了关系推理的深度.针对此问题,文献[120]提出了一种基于生成参数的图神经网络关系抽取——GP-GNN. GP-GNN 首先构建了基于文本序列输入的全连接层,用来接收文本数据.之后构建了三个不同层面的模块:能够对语句中的信息进行编码的边模块、节点关系传播模块和使用节点表示进行预测的分类模块.通过对文本的编码,使得图神经网络能够从自然语言建模图中学习有用的参数,并进行多跳关系推理,从而提取文本中更加复杂的关系.

阅读理解一直是文本分析中一项重要而富有挑战性的任务,文档中的实体、句子之间的复杂关系,以及跨文档之间的联系无疑给机器理解文本带来很大的困扰.在此问题上,Tu 等人^[121]引入了具有不同类型节点和边的异构图,提出异构文档—实体图模型(Heterogeneous Document-Entity Graph, HDE-graph)用于解决跨文档的多跳阅读理解. HDE-graph 包含了实体、上下文候选信息、文档等不同粒度级别的信息. Tu 等人用编码器对 HDE-graph 节点初始化,并使用了共同注意力机制和自注意力机制. HDE-graph 的优点在于能不断地对文本中的证据积累,形成的证据链可以用来在多个文档中推理,从而得到问题的最终答案.在维基文章上的对比测试表明 HDE-graph 具有最高的准确率.文献[122]提出动态融合图网络(Dynamically Fused Graph Network,DFGN)模型,将 GNN 应用在跨文档的阅读理解中. DFGN 包含一个动态融合层,它会从给定查询中提到的实体开始,沿着实体链不断进行实体搜索.在此过程中,会构建动态文本实体图,并逐步从给定文档中找到支持答案的相关联实体.在公共数据集上的评估表明,DFGN 能够产生具有可解释的推理链.除了较为常规的通用文本外,一些特殊领域的文本还需要更加有针对性的分析和处理.例如在新闻文本中,很多具有倾向性的观点等信息都被隐藏在词汇和语句中,因此发掘作者的观点是一件比较困难的事情.针对此问题,Li 等人^[123]把新闻文章和社交圈嵌入到同一个空间,构建社交融合的文本表示(Socially-infused Textual Repre-

sentation). 由于社交网络中存在大量的用户关系、用户身份以及个人观点等信息,因此可以将文本、用户的社交活动以及用户之间的关系构建一个图,用 GCN 学习它们的表示. Li 等人捕获了新闻事件中媒体的政治视角,强调了将新闻文本置于情景化社交网络的重要性,揭示了这些信息如何在社交网络中传播.

在对话问答任务中,推理能力、可解释性和扩展性是对话问答系统面临的主要挑战.在基于 GNN 的对话问答建模上,文献[124]提出 web 级文档多跳认知图(Cognitive Graph QA, CogQA)模型.为了在给出问题答案的同时提供可解释的推理路径,该框架以认知科学的对偶理论为基础,通过协调隐式实体抽取模块和显式语义推理模块,在迭代过程中逐步构建认知图.其中推理模块的功能是用 GNN 构建的,它的功能是抽取问题中的相关实体,并对语义信息进行编码,生成隐含特征表示,然后生成推理证据路径,使得对话系统提供的答案更具有说服力.

5.2 图神经网络在物理化学和药物医学中的应用

在物理化学以及药物医学三个自然科学领域中,原子、分子、药物化合物等实体数据都可以在逻辑上被表示成图结构.由于一些分子化合物和药物等有机体是异构非线性的,数据中存在大量的噪声.在标签稀缺的情况下,将相关信息集成起来学习是一个挑战.如果能够对这些数据进行建模和分析,则可以缩短研究人员探测原子结构、化学键属性、药物靶分子结合机理等任务的时间,为其潜在的属性和分布提供一定的参考和支持.

Duvenaud 等人^[125]用 GCN 对分子指纹信息采集,设计了一种数据驱动的端到端预测模型.其中的输入是任意分子的图结构. Duvenaud 等人定义了分子指纹相似性度量的方法,每一层网络用哈希函数整合原子和原子邻居结构信息,用索引将所有节点特征转变成固定长度的分子指纹向量. Duvenaud 等人的贡献在于将计算分子指纹向量的神经网络层替换为可微神经网络,进而可以循环提取分子指纹特征,忽略分子结构中非重要变化. Duvenaud 等人的方法侧重于减少分子非结构性改变对分子指纹的影响,而文献[126]希望更多地利用到原子属性信息,因此描述了一种分子图卷积方法.通过对分子图中原子、化学键、原子间和分子间距离等项目简单编码,自顶向下的提取原子特征和分子级特征.在每个网络层中对相同的原子施以相同的操作,以保持排列不变性.在分子级特征抽取前,再对原子做一次卷

积来扩展深度,使得原子特征向量的每个维度包含的信息更多,让分子级特征化过程中丢失的信息更少,更好地利用分子图的属性信息.在分子属性信息的进一步探究中,文献[104]提出消息传递神经网络(Message Passing Neural Network, MPNN)用于化学分子式的属性预测. MPNN 将之前用来学习化学分子式的多个 GNN^[90,125-126]整合进一个框架中,着重解决两个评估基准问题——量子力学模拟的平均错误率和化学精度. MPNN 的信息函数定义了矩阵乘法、边特征编码网络和成对消息传递机制. MPNN 将单个节点向量转变为多个分向量,并对每个分向量进行信息前向传播来得到临时的嵌入表达,以提高 MPNN 对于化学分子式属性学习的可扩展性.此外,前文提到的 MolGAN^[79]也是生成化学分子式的有效模型. MolGAN 与 MPNN 区别在于其利用生成对抗网络,模拟的是化学分子式的先验分布,属于隐式学习. MPNN 则利用了多种 GNN 来对化学键信息进行传播,属于显式学习.

在药物发现的早期阶段,预测药物分子对靶点的药效以及结合亲和力可以用两种方法.一种是基于分子力场的经验评分函数,一种是基于化学信息描述符的分子对接技术.但是这需要专家级知识来对药物进行编码,需要人工调整某些参数和特征,耗时耗力.针对此问题,文献[127]开发了一个通用的三维空间图卷积运算,直接用图卷积网络学习原子级蛋白质化学作用,称作原子卷积神经网络(Atomic Convolutional Neural Network, ACNN). ACNN 首先学习了结合态的晶体结构,通过计算配体和蛋白质相关联的能量来给出结合亲和力的数值. ACNN 优点在于能识别蛋白质配体亚结构中缺失复合物的非共价相互作用.与以往蛋白质配体预测系统不同,ACNN 是端到端且完全可微的,代表了基于数据驱动和物理结构探测的深度学习新范例,为基于结构的药物活性预测提供了坚实的基础.而在药物临床应用中,模拟生成的药物分子图必须能被现有的化学理论分析,并与相关的病理研究相匹配,所以模型必须要同时保证精确性和可解释性.基于此目的,文献[128]利用 GCN 对小分子药物操作,其步骤包括节点特征提取、邻居节点聚合、图池化和图聚集,最终生成单个小分子的向量表达.之后结合残差 LSTM 网络对小分子向量进行训练,对小分子向量之间相似距离进行度量,有效地预测小分子药物的活性,并与现有的临床医学病理特征建立密切的联系.

此外,药物相似性测量是现代生物医学的基础任务之一,药物相似性研究可以支持很多下游的临床任务.一方面,开发药物与现有药物之间相似性代表了一种隐含关系,有时候需要从旧药物已知特性推断新药物的新特性.另一方面,新型药物的新特征给药物相似性学习带来全面性和准确性的挑战.面对精确的药物相似性度量问题, Ma 等人^[129]提出了一种多视图的 GAE 建模方式.该方式将每种药物都建模为药物关联网络中的一个节点,并扩展了 GCN 来嵌入多视图节点特征和图边缘特征.在药物节点学习中添加注意力机制,确定每个视图中相关任务和特征的权重,保证对数据变化的动态适应. Ma 等人将未知标签视作无特征的隐含变量,构造了半监督和无监督两种学习方式.实验结果表明该方法有较好的药物预测精度,在药物节点嵌入特征上具有较强的可解释性.

5.3 图神经网络与图像处理

图像处理是计算机应用领域的一个热门方向,得益于强大的 GPU 计算能力和数据的张量表示形式,基于深度学习的模型和方法能高效地进行图像处理.尽管 CNN 在图像分类等问题上达到较高的精度,但面对更加复杂的语义分割、图像问答等任务, CNN 还有一些不足和缺陷.针对这些问题,图神经网络可以融合多种异质内容和结构信息,构建相应的知识图 and 对象图等,利用额外的信息来指导具体任务的进行.

文献[130]根据人体关节和骨骼之间的运动相关性,将骨骼数据表示为有向无环图,从而对图像中的动作进行识别.该模型设计了一种新的有向 GNN,用于提取关节、骨骼及其相互关系的信息,并根据提取的特征预测动作类别.为了更好地适应动作识别任务,在训练过程上对图的拓扑结构进行了自适应调整.在 TSCN(Two-stream Convolutional Network)^[131]框架下,将骨架序列的运动信息与空间信息相结合地学习,进一步提高了模型的性能.

在图像问答上,传统方法仅仅基于图像内容来做出相关的回答,缺乏必要的知识使得模型可能会答非所问.针对此问题,文献[132]开发了一个实体图,把所有实体用 GCN 囊括进来.这样在问题检索中,可以避免对单个事实的依赖.在给定图像和问题后,通过它们之间的相似度评分将对应的事实抽取出来.在推理正确答案时,该模型可以从候选答案列表中综合选择最相符的答案.此外,通过对全局词向量进行排序,选择知识库中支持事实的列表,以消除

同义词和同形词带来的混淆影响。

在图像语义分割中,传统的语义分割方法主要在二维平面进行处理.文献[133]提出了一种在三维点云上建立 k 近邻图的三维图神经网络(3D Graph Neural Network, 3DGNN). 图中的每个节点对应于点云中的一组点集合,与物体外观特征初始化表示向量相关联,其中物体外观特征由 CNN 从二维图像中提取.点云中的每个节点都有一个递归函数,该函数会根据当前状态和邻居消息动态更新节点的隐含表示,最终每个节点都能用来预测图像中每个像素的语义类.此外,文献[134]认为三维点云可以通过一种超点图的结构有效地捕获,用 GCN 生成了包含上下文关系的对象部件表示,解决了海量点云语义分割的难题.

在图像分类中,传统的研究思路是只要有足够多的数据样本,那么就可以用大量的样本来训练模型,达到非常高的准确率.但是在实际情况中,不一定有足够多的数据给模型来训练,因而催生了小样本学习.同样地,图神经网络可以被应用至小样本学习中.针对样本类别标签不足的问题, Garcia 等人^[135]将小样本学习作为一种有监督的消息传递任务,利用 GNN 进行端到端的训练. Garcia 等人将节点与数据集中的图像相关联,用相似度函数来生成边,在经过 GCN 训练后对图像进行分类.此外 Garcia 等人还将该模型框架扩展到半监督学习和主动学习上,并取得了良好的效果.

5.4 图神经网络在交通流量和轨迹预测中的应用

交通问题是现代城市的热点民生问题,及时准确的交通预测对于城市道路的控制和引导至关重要.特别是车辆速度、交通量和道路密度等统计量需要有较为精确估计,这对于城市交通网络中的路线规划和流量控制具有重要意义.由于交通流具有高度的非线性和复杂性特点,传统机器学习方法难以学习空间和时间的依赖关系,不能满足中长期交通流预测任务的要求.针对此问题,文献[136]提出一种时空图卷积网络(Spatio-Temporal Graph Convolutional Network, STGCN),用于解决交通领域的时间序列预测问题.不同于使用反复卷积和递归单元,STGCN 构造和堆叠了多个时空图卷积模块,用来抽取和融合时空特征.这种多尺度可扩展网络建模的方式,让 STGCN 能够学习到交通数据中较为全面的时空相关性特征.此外,交通中的空间依赖关系会随着时间的推移而改变,而固定的拉普拉斯矩阵无法捕捉这种变化.为了跟踪交通数据之间的

空间依赖关系,文献[137]提出了一种动态时空图卷积网络(Dynamic Graph Convolutional Neural Network, DGCNN)用于精确的交通预测.为了能够降低复杂度并进行实时学习,DGCNN 在网络框架中构造并预训练了张量分解层,将交通数据张量分解成低阶张量和稀疏张量.其中低阶张量对应稳定的全局张量,稀疏张量对应可变的局部张量.全局张量反映道路交通中长期的时空关系网络,局部张量用来捕获交通流量的日常波动和起伏. DGCNN 使用动态图估计器来学习全局张量和局部张量的拉普拉斯矩阵.相较于 STGCN, DGCNN 能够随时捕捉到交通流量中一些细微因素变化的影响,因此在训练误差和测试误差上都要更低一些.针对交通流中存在瞬时高峰和事故发生突然性等问题,文献[138]将道路传感器视作节点,用传感器之间的相近度作为边,从而构造一个有向图.通过扩散卷积运算来捕捉空间相关性,构建了扩散卷积循环网络(Diffusion Convolutional Recurrent Neural Network, DCRNN). DCRNN 在图上进行双向随机游走来捕获空间依赖性,运用序列到序列的学习框架,并使用具有定时采样的图编码器-解码器架构来捕获时间依赖性.

另外,针对道路行人轨迹预测和追踪, Kosaraju 等人^[139]提出了一种使用 GAT 结构的 Social-BiGAT 模型. Social-BiGAT 中构建了一个具有对抗性的递归图编码器-解码器体系结构. Kosaraju 等人使用了 Bicycle-GAN^[140]作为编码器训练框架,模拟不同场景下行人的社会交互行为.通过对场景图中人物之间的互动信息进行编码,得到行人的未来路径预测结果.其中每一个场景与其潜在噪声向量之间形成可逆变换,明确解释行人轨迹预测问题的多模态性质,提高了人物多模态轨迹预测精度.

5.5 其 他

除了上述应用外, GNN 还在诸如知识图谱与知识库、信息检索、动态网络异常检测、医保欺诈分析、网络图分析等其他领域发挥重要的作用.

在知识图谱上, Wang 等人^[141]在知识图谱中使用 GCN 完成实体对齐任务. Wang 等人将每种语言的实体用 GCN 嵌入到一个统一的向量空间中,从实体的结构信息和属性信息中学习等价关系.此外,文献[142]将 GNN 应用于知识库外实体知识转移.在传播阶段,通过邻居节点聚合和池化操作得到知识库外实体的嵌入.在输出阶段,在节点向量上定义了面向任务的目标函数,使用 TransE^[143]作为预期输出模型.该模型的优势在于仅利用测试时提供的

有限辅助知识,不需要复杂的再训练.在信息检索中,文献[144]提出特征交互图神经网络(Feature Interaction Graph Neural Network, Fi-GNN)用于点击率预测. Fi-GNN 中的节点对应不同领域特征,边对应特征之间的交互影响. Fi-GNN 利用图的强表征性,灵活、明确地对复杂的用户特征交互进行建模,为点击率预测提供良好的模型解释.在推荐系统上,文献[145]把用户-推荐项关系转变为图表达,用 GCN 从用户和项目中学习有意义的统计图模式,并用 RNN 将参数扩散应用于评分矩阵,完成对推荐系统矩阵的补全.在关于动态网络异常检测方面,文献[146]提出一种面向动态网络的无监督学习算法 Dynamic-DGI. Dynamic-DGI 框架分为四个部分,分别是图神经网络,用来对图结构和属性特征提取;动态网络表示学习;图时间变化特征提取;流数据异常检测.其中图神经网络使用 Kipf 等人^[27]的卷积核对图节点特征提取和聚合.为了将边信息直接考虑进来,Dynamic-DGI 把原始图转为线图,再次用 Kipf 等人^[27]的卷积核对线图进行边信息提取,同时将原图和线图提取的特征进行拼接整合,作为图节点和边的隐含向量表示. Dynamic-DGI 优点在于能学习到图结构和属性的双重变化,捕捉到图动态变化中的异常部分.在医保欺诈分析中, Yi 等人^[147]提出了一种基于图卷积和变分自编码的单

分类医保欺诈检测模型-OCGVAE. OCGVAE 框架由 GCN、VAE 和标准偏差向量三个模块组成. Yi 提取病人就诊记录构建了医生和病人之间的有权图.用 GCN 提取医生-患者关系网络节点特征,其中的隐藏层设计了参数共享结构,用来降低医患网络的数据维度.实验结果表明,OCGVAE 模型对于医保欺诈标签的预测准确率达到 87%,大幅度优于传统算法,有效地提高了医保欺诈筛查精度.在网络图分析中, Zhang 等人^[148]将 GCN 用于现实网络的链接预测,提出 γ 衰减理论,将大量启发式策略整合进一个框架之中,并且证明了这些启发式策略近似等价于局部子图模式.然后使用 GCN 从局部子图中学习,并进行链接预测.文献[149]对 GNN 性能进行理论分析,重点探讨了 GNN 框架结构和参数优化对图数据分析的有效性,提出了最小 GNN 结构的平均场理论用于解决图分割问题.

可以看出来,在科学和生活的多个领域,凡是能将数据表示成图结构的,都能用图神经网络来进行学习.通过对图结构信息以及图属性信息的有效捕捉,图神经网络在各种图任务上达到了较高的精度,成为解决图相关问题的有效手段,因此在不同的领域都有着良好的应用前景.表 9 是 GNN 主要应用领域实例列表.

表 9 GNN 主要应用领域实例列表

应用领域	GNN 种类	模型框架	输入端	任务	文献
自然语言处理	GCN	语法嵌入 SynGCN; 语义嵌入 SemGCN	词汇图	词嵌入	文献 ^[117]
	GCN	注意力引导图卷积层; 密集连接网络层; 线性联合网络层	句子依赖树	关系抽取	AGGCN ^[118]
	GCN	Bi-LSTM 句子编码器; GCN 依赖树编码器; 关系加权 GCN	句子依赖树	命名实体识别; 关系抽取	GraphRel ^[119]
	GRN	文本接受全连接模块; 边信息编码模块; 节点关系传播模块; 节点预测分类模块	实体完全连通图	关系推理	GP-GNN ^[120]
	GCN	文本编码模块; 异构文档图推理模块; 累计评分模块	异构文档实体图	阅读理解	HDE-graph ^[121]
	GAT	段落选择模块; 文本查询编码模块; 推理融合模块; 答案预测模块	实体图	阅读理解	DFGN ^[122]
	GCN	文本表示与语言偏误纠正; 信息网络图卷积学习; 文本分类与推理	信息流图	新闻政治观点抽取	文献 ^[123]
	GCN	隐式实体抽取模块; 显式语义推理模块	认知图	对话问答	CogQA ^[124]

续表

应用领域	GNN 种类	模型框架	输入端	任务	文献
物理化学和药物医学	GCN	原子信息哈希整合;特征索引固定;邻居原子规范化	分子图	分子指纹采集	文献 ^[125]
	GCN	编码层;原子卷积层;分子特征抽取层	分子图	分子指纹提取	文献 ^[126]
	GCN GRN	化学键信息传播	分子图	分子属性预测	MPNN ^[104]
	GGN	ORGAN 强化学习;小批量鉴别	分子图	化学分子生成	MolGAN ^[79]
	GCN	距离矩阵与邻域列表构造;原子卷积层;径向池化层	蛋白质图	预测蛋白质配体	ACNN ^[127]
	GCN	节点特征提取;邻居节点聚合;图池化;图聚集	药物分子	药物活性预测	文献 ^[128]
	GAE GCN	药物特征图卷积提取;半监督图自编码器标签预测	药物分子图	药物相似性预测	文献 ^[129]
图像处理	GCN	人体骨骼图构建;关节、骨骼关系提取;时空序列建模	骨骼图	人体动作预测	文献 ^[130]
	GCN	问题事实获取;实体关系预测;图像实体嵌入;问题与图像信息提取;答案预测	图片和问题图	图像问答	文献 ^[132]
	GRN	三维图构造;像素特征提取;标签预测	三维点云	语义分割	3DGNN ^[133]
	GCN	点云几何分割;超点图构建;超点与超边信息提取;标签预测	三维点云	语义分割	文献 ^[134]
	GCN	图片上下文映射;初始节点特征构造;标签分布学习	图片	图像分类	文献 ^[135]
交通流量和轨迹预测	GRN GCN	时空图卷积模块	交通流图	交通流预测	STGCN ^[136]
	GCN GRN	动态拉普拉斯矩阵估计;张量分解层;时间特征抽取;空间特征抽取	交通流图	交通流预测	DGCNN ^[137]
	GRN GAE	扩散图卷积特征抽取;图自编码器时间信息捕获	道路图	交通流预测	DCRNN ^[138]
	GAT	GAT 提取社交物理特征;局部行人场景识别;全局场景识别;场景噪声编码器	场景轨迹图	行人轨迹追踪	Social-BiGAT ^[139]
知识图谱与知识库	GCN	图卷积实体嵌入;实体对齐预测	知识图谱	实体对齐	文献 ^[141]
	GRN	知识图谱信息传播模块;目标分数评估模块	知识图谱	知识库推理	文献 ^[142]
信息检索	GRN	领域特征图构建;特征交互建模;注意力传播模块	领域特征图	点击率预测	Fi-GNN ^[144]
推荐系统	GCN	图特征提取模块;基于 RNN 的矩阵扩散模块	用户-推荐项关系图	推荐系统矩阵补全	文献 ^[145]
动态网络异常检测	GCN	图神经网络层、动态网络表示学习、图时间变化特征提取模块、流数据异常检测模块	动态网络图	异常图检测	Dynamic-DGI ^[146]
医保欺诈分析	GCN	GCN、VAE、标准偏差向量	医患关系网络	医保欺诈标签预测	OCGVAE ^[147]
网络图分析	GCN	节点隐式嵌入;有监督启发式学习	网络图	链接预测	文献 ^[148]
	GCN	图信息随机块;动态平均场图分割	网络图	图分割	文献 ^[149]

6 未来研究方向

GNN 的核心在于规范化表示的图结构数据并用深度神经网络进行学习. 经过近些年的不断发展, 通过大量数学证明和实验分析后, GNN 在理论上和

实践上都被证实是对图结构数据处理的一种有效方法和框架. 尽管 GNN 在各个领域的图数据上取得了不俗的表现和较好的普适性, 但是 GNN 仍然存在一定的不足和需要完善的地方. 根据目前国内外的研究现状, 下面本文对 GNN 的一些制约因素和未来发展方向进行探讨.

6.1 网络深度

在计算机视觉、自然语言处理和音频处理中,神经网络的层数可以叠加多层.在一定范围内,神经网络层数的增加可以更好地提取数据中的特征信息.例如深层残差网络 ResNet^[150]可以达到 152 层.但是 GNN 的邻居节点聚合中,随着网络层数的增加,邻居节点的阶数会不断扩张,导致中心节点聚合特征数量成指数变多.这在大规模数据集上,尤其是节点之间的边连接数量较多时表现的非常明显.随之而来的是训练过程中计算复杂度的剧增,并可能导致过拟合的现象发生.这也就意味着随着层数的增加,GNN 模型性能会急剧下降.如果想要加深网络层数,就必须限制每层节点数量.但是这也会使得特征聚集的量变少,导致节点之间信息传播受阻.如何解决这一矛盾性问题是将来研究的重点之一.

6.2 动态性

就目前来看,现有的 GNN 大多处理的是静态齐次图.一方面,GNN 框架会假定图结构是固定的;另一方面,GNN 框架会假设图中的节点和边来自于单一源分布.然而,这两个假设在许多情况下并不能同时成立.在社交网络中,新的人可以随时进入网络,并且现有的人也可以退出网络.在推荐系统中,产品可能有不同的类型,其输入可能有不同的形式,如文本或图像.特别是在超大规模的图中,节点的个数和边的个数可能有百万、千万乃至上亿.尤其是随着数据的增加和改变,节点和边的个数以及节点和边的类型都可能发生动态的变化.在这些任务处理中,图的动态变化是不能忽视的.特别是在固定尺寸下,因为某个节点或者边发生改变而重新学习整个图将会使得代价十分昂贵.而大多数 GNN 对于大型图不具有很好的伸缩性.其主要原因是当堆叠 GNN 的多个层时,节点的最终状态涉及大量邻居的隐藏状态,导致反向传播的高复杂性.虽然目前有一定的文献^[94-95,136-137]在研究图的时空动态性,但是面对更大规模和更加复杂的动态异质图数据时还不够有效.因此如何对图的动态性进行有效的适应是未来的研究方向之一.

6.3 感受域

一个节点的感受域是指一组节点集合,包括中心节点及其邻居节点.感受域大小是决定邻居节点数量的关键参数.在大规模图数据集中,平均每个节点周围有多个邻居节点存在.随着网络层数的增加,邻居节点会递归增加数目,感受域也随之快速扩张.

这可能会超过存储空间的上限.此外,一些节点可能只有一个邻居,而另外节点可能有多达数千个邻居.邻居节点分布不均衡使得每个中心节点的感受域大小不一致.尽管可以通过添加“哑结点”和删除邻居节点的方式保持数据大小和维度的一致,但是在特征的聚集和融合中不可避免的会有信息损失现象发生,而现有的采样方法还不能完全解决该问题.

6.4 多网络的融合

由于现实世界数据的复杂性,抽象出来的图结构也会有很多的种类和变体.有向无向、异质非异质、带权不带权等等,大部分的 GNN 仅能处理其中的某一种类型.而更普遍的情况是各种各样的图混杂在一起,并且希望 GNN 能满足诸如节点分类、图分类、可视化、图生成等多种任务需求.在这种复杂的高强度的任务要求下,单一的神经网络作用过于有限.因此对于更加复杂的情况,有必要进行多网络融合.目前比较主流的多网络融合方式是 GCN 与其他 GNN 算法相结合.例如在节点属性和图拓扑结构信息的获取上,GCN 明显具有较高的性能和良好的适应性,在节点分类问题上会表现良好.鉴于其优点,在 GAE 中不乏部分模型使用 GCN 作为编码器,取得较好的效果.但如果还需要进行链接预测、节点生成或者图生成,GCN 则有点力不从心了.此时可以再增设一个 GGN,输入 GCN 处理后的节点嵌入向量,在 GGN 内生成概率分布,完成生成式任务.如果图在不断地递归演进,形成了图序列.则可以利用 GRN 来处理,以囊括多个步骤下的图信息.因此在 GNN 框架中构造不同用途的深度神经网络,从不同的侧面来提取和整合数据的特征是十分有必要的.此外可以对诸如深度置信网络(Deep Belief Network)^[151]、Transformer^[152]等神经网络进行改造,将其泛化和应用至图结构数据学习上.

6.5 与网络嵌入的结合

网络嵌入可以将原始图数据的高维稀疏矩阵转变为低维度稠密的向量,这可以大幅度压缩存储空间,并提取有效的图信息.一般图节点的原始特征矩阵是高维稀疏的,对于一个 $N \times F$ 的特征矩阵,当 F 比较大时,所需要的存储空间也相应的增加.如果矩阵比较稀疏,那么存储效率也会比较低.网络嵌入则可以利用图结构信息,生成低维连续的节点特征表示,避免存储空间浪费.其次,由于生成的节点嵌入表示包含了部分邻居节点信息,所以中心节点的感受域也可以相应的减少.对于多层图卷积和需

要迭代压缩的 GNN 来说,一定程度上可以减少网络层数和迭代压缩次数.例如 Kipf 等人^[27]半监督 GCN 复杂度为 $O(|E|FC)$, DeepWalk^[110] 的复杂度为 $O(\log(N))$. 当边连接比较密集并且节点特征维度很大时,复杂度较高.如果对节点特征降维,使得降维之后的维度 $F' \ll F$,这样总体复杂度变为 $O(\log(N)) + O(|E|F'C)$. 尽管增加了网络嵌入的计算时间,但是在图卷积层可以大幅度降低计算开销,这样可以提高训练的有效性以及降低计算复杂度.文献^[66,76,86]就使用随机游走等网络嵌入方法来为 GNN 模型构建输入序列,除此之外未来研究中也可以尝试诸如 Node2vec^[77]、LINE^[153] 等网络嵌入方法来对 GNN 的输入端进行改进.

7 总 结

从 2005 年 Gori 等人提出 GNN 概念,到 GCN 的出现为非欧式结构数据提供有效的处理范式,再到 GAE、GAT、GRN、GGN 等不同 GNN 框架变种的提出以及 GNN 在各个领域的应用,GNN 在理论和实践上经历一个从无到有、从有到优化的过程,GNN 的体系族也在不断地发展和完善.从这段历程中,可以看出许多研究人员对 GNN 算法和结构不断改进和优化.从一开始 GRN 只是对节点进行简单的特征变换和压缩,到后面 GCN 对邻居节点特征采样和聚合以及边信息传播路径的分析,GNN 利用图结构信息和属性信息的能力越来越强.在面对不同隐含数据分布时,GAE 能够对图进行非线性变换,进行半监督或者无监督学习,提取更具有代表性的特征.在面对物理化学以及药物医学等需要生成具有领域限定要求的图结构数据时,GGN 能够对图数据进行对抗性训练,获得期望的分子图、药物化合物图等.而 GAT 的出现则使得神经网络能够聚焦在任务关注的节点、边和子图上,忽视不重要的信息,增强训练的有效性和鲁棒性,提升测试结果的可解释性.

在 GNN 发展的同时,面向图结构的算法还有网络嵌入和图核方法.尽管它们也能分析和处理相关的图数据和任务,但都是从某个侧面的角度出发对图进行窥探和剖析.相较于网络嵌入和图核方法,GNN 能够直接对节点、边以及图整体进行特征聚合和提取,最大程度的保留了图的拓扑结构信息和内容属性信息,在分类、聚类、预测、生成等任务上表现更加突出.当然,在分析图数据时,这三种方法可以

综合使用,通过相互组合和融合的方式可以在特定任务上取得不俗的性能表现.

本文概括和提炼了目前主流的 GNN 框架和模型,并对不同算法进行归纳总结,分析了 GNN 在多个领域的应用以及未来发展方向.总之,作为一种图结构数据处理框架,虽然 GNN 概念的提出比较早,但是其兴起和发展的时间还不长,现有的模型算法还有进一步提高和改进的空间.在未来研究的道路上,还可能更多优秀的深度神经网络加入到 GNN 中,通过对图结构数据进行更加全面而细致的分析,达到在降低训练量的同时取得更加优异效果的目标.

参 考 文 献

- [1] Hinton, G. E. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313(5786): 504-507
- [2] JIAO Li-Cheng, YANG Shu-Yuan, LIU Fang et al. Seventy-years beyond neural networks: retrospect and prospect. *Chinese Journal of Computers*, 2016, 39(8): 1697-1716(in Chinese)
(焦李成,杨淑媛,刘芳等.神经网络七十年:回顾与展望.计算机学报,2016,39(8):1697-1716)
- [3] Gallinari P, Thiria S, Fogelman-Soulié F. Multilayer perceptrons and data analysis//*Proceedings of the IEEE International Joint Conference on Neural Networks*. San Diego, USA, 1988: 391-399
- [4] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989, 1(4): 541-551
- [5] Elman J L. Finding structure in time. *Cognitive Science*, 1990, 14(2): 179-211
- [6] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets//*Proceedings of the 28th Advances in Neural Information Processing Systems*. Montreal, Canada, 2014: 2672-2680
- [7] Rifai S, Vincent P, Muller X, et al. Contractive auto-encoders: explicit invariance during feature extraction//*Proceedings of the 28th International Conference on Machine Learning*. Washington, USA, 2011: 833-840
- [8] ZHOU Fei-Yan, JIN Lin-Peng, DONG Jun. Review of convolutional network. *Chinese Journal of Computers*, 2017, 40(6): 1229-1251(in Chinese)
(周飞燕,金林鹏,董军.卷积神经网络研究综述.计算机学报,2017,40(06):1229-1251)
- [9] Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains//*Proceedings of the International Joint Conference on Neural Networks*. Montreal, Canada, 2005: 729-734
- [10] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs//*Proceedings of the*

- 2nd International Conference on Learning Representations. Banff, Canada, 2014
- [11] Xu Bing-Bing, Cen Ke-Ting, Huang Jun-Jie et al. A survey on graph convolutional neural network. Chinese Journal of Computers, 2020, 43(5):755-780(in Chinese)
(徐冰冰,岑科廷,黄俊杰等. 图卷积神经网络综述. 计算机学报, 2020, 43(5):755-780)
- [12] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model. IEEE Transactions on Neural Networks, 2009, 20(1): 61-80
- [13] Micheli A. Neural network for graphs: a contextual approach. IEEE Transactions on Neural Networks, 2009, 20(3): 498-511
- [14] Bronstein M M, Bruna J, LeCun Y, et al. Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine, 2017, 34(4): 18-42
- [15] Tsitsvero M, Barbarossa S. On the degrees of freedom of signals on graphs//Proceedings of the 23rd European Signal Processing Conference (EUSIPCO). Nice, France, 2015: 1506-1510
- [16] Pasdeloup B, Alami R, Gripon V, et al. Toward an uncertainty principle for weighted graphs//Proceedings of the 23rd European Signal Processing Conference (EUSIPCO). Nice, France, 2015: 1496-1500
- [17] Ram I, Elad M, Cohen I. Generalized tree-based wavelet transform. IEEE Transactions on Signal Processing, 2011, 59(9): 4199-4209
- [18] Gavish M, Nadler B, Coifman R R. Multiscale wavelets on trees, graphs and high dimensional data: theory and applications to semi supervised learning//Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel, 2010: 367-374
- [19] Sandryhaila A, Moura J M F. Discrete signal processing on graphs. IEEE Transactions on Signal Processing, 2013, 61(7): 1644-1656
- [20] Shuman D I, Narang S K, Frossard P, et al. The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine, 2012, 30(3): 83-98
- [21] Chung F R K, Graham F C. Spectral graph theory. America: American Mathematical Society, 1997
- [22] Hammond D K, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory. arXiv: 0912.3848, 2009
- [23] Susnjara A, Perraudin N, Kressner D, et al. Accelerated filtering on graphs using Lanczos method. arXiv: 1509.0453, 2015
- [24] Cullum, Jane K, Willoughby, et al. Lanczos algorithms for large symmetric eigenvalue computations. Society for Industrial and Applied Mathematics, 2002, 1: Theory
- [25] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering//Proceedings of the 30th Advances in Neural Information Processing Systems. Barcelona, Spain, 2016: 3837-3845
- [26] Levie R, Monti F, Bresson X, et al. CayleyNets: graph convolutional neural networks with complex rational spectral filters. IEEE Transactions on Signal Processing, 2019, 67(1): 97-109
- [27] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks//Proceedings of 5th International Conference on Learning Representations. Toulon, France, 2017
- [28] Li R, Wang S, Zhu F, et al. Adaptive graph convolutional neural networks//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Louisiana, USA, 2018: 3546-3553
- [29] Verma S, Zhang Z. Graph capsule convolutional neural networks. arXiv: 1805.08090, 2018
- [30] Hinton G E, Krizhevsky A, Wang S D. Transforming auto-encoders//Proceedings of the 21st International Conference on Artificial Neural Networks. Espoo, Finland, 2011: 44-51
- [31] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules//Proceedings of the 31st Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 3856-3866
- [32] Verma S, Zhang Z-L. Hunt for the unique, stable, sparse and fast feature learning on graphs//Proceedings of the 31st Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 88-98
- [33] Zhuang C, Ma Q. Dual graph convolutional networks for graph-Based semi-supervised classification//Proceedings of the 27th International Conference on World Wide Web. Lyon, France, 2018: 499-508
- [34] Henaff M, Bruna J, LeCun Y. Deep convolutional networks on graph-structured data. arXiv: 1506.05163, 2015
- [35] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering//Proceedings of the 15th Advances in Neural Information Processing Systems. Vancouver, Canada, 2001: 585-591
- [36] Perona P. Self-tuning spectral clustering//Proceedings of the 18th Advances in Neural Information Processing Systems. Vancouver, Canada, 2004: 1601-1608
- [37] Zhang M, Cui Z, Neumann M, et al. An end-to-end deep learning architecture for graph classification//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 4438-4445
- [38] Shervashidze N, Schweitzer P, Leeuwen E J van, et al. Weisfeiler-lehman graph kernels. Journal of Machine Learning Research, 2010, 1(3): 1-48
- [39] Ying R, You J, Morris C, et al. Hierarchical graph representation learning with differentiable pooling//Proceedings of the 32nd Advances in Neural Information Processing Systems. Montreal, Canada, 2018: 4805-4815
- [40] Kistruck G M, Sutter C J, Lount R B, et al. Learning convolution neural networks for graphs//Proceedings of the 33rd International Conference on Machine Learning. New York,

- USA, 2016; 2014-2023
- [41] Douglas B L. The weisfeiler-lehman method and graph isomorphism testing. arXiv: 1101.5211, 2011
- [42] Atwood J, Towsley D. Diffusion-convolutional neural networks//Proceedings of the 30th Advances in Neural Information Processing Systems. Barcelona, Spain, 2016; 1993-2001
- [43] Hechtlinger Y, Chakravarti P, Qin J. A generalization of convolutional neural networks to graph-structured data. arXiv: 1704.08165, 2017
- [44] William L. H, Ying Z, Leskovec J. Inductive representation learning on large graphs//Proceedings of the 31st Advances in Neural Information Processing Systems. Long Beach, USA, 2017; 1024-1034
- [45] Chen J, Ma T, Xiao C. FastGCN: fast learning with graph convolutional networks via importance sampling//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada, 2018
- [46] Chen J, Zhu J, Song L. Stochastic training of graph convolutional networks with variance reduction//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden, 2018; 941-949
- [47] Huang W, Zhang T, Rong Y, et al. Adaptive sampling towards fast graph representation learning//Proceedings of the 32nd Conference on Neural Information Processing Systems. Montreal, Canada, 2018; 4563-4572
- [48] CHEN Kejia, YANG Zeyu, LIU Zheng, et al. Graph convolutional network model using neighborhood selection strategy. Journal of Computer Applications, 2019, 39(12):3415-3419(in Chinese)
(陈可佳, 杨泽宇, 刘峥等. 基于邻域选择策略的图卷积网络模型. 计算机应用, 2019, 39(12):3415-3419)
- [49] Gao H, Wang Z, Ji S. Large-scale learnable graph convolutional networks//Proceedings of the 24th International Conference on Knowledge Discovery and Data Mining. London, UK, 2018; 1416-1424
- [50] Derr T, Ma Y, Tang J. Signed graph convolutional networks//Proceedings of the 20th International Conference on Data Mining. Singapore, 2018; 929-934
- [51] Such F P, Sah S, Dominguez M A, et al. Robust spatial filtering with graph convolutional neural networks. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(6): 884-896
- [52] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks//Proceedings of the 30th International Conference on Artificial Intelligence and Statistics. Sardinia, Italy, 2010; 249-256
- [53] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift//Proceedings of the 32nd International Conference on Machine Learning. Lille, France, 2015; 448-456
- [54] Cao S, Lu W, Xu Q. Deep neural networks for learning graph representations//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016; 1145-1152
- [55] Wang C, Pan S, Long G, et al. MGAE: marginalized graph autoencoder for graph clustering//Proceedings of the 26th ACM International Conference on Information and Knowledge Management. Singapore, 2017; 889-898
- [56] Bourlard H, Kamp Y. Auto-association by multilayer perceptrons and singular value decomposition. Biological Cybernetics, 1988, 59(4-5): 291-294
- [57] Lv L, Cheng J, Peng N, et al. Auto-encoder based graph convolutional networks for online financial anti-fraud//Proceedings of the IEEE Conference on Computational Intelligence for Financial Engineering and Economics. Shenzhen, China, 2019; 1-6
- [58] Kingma D P, Welling M. Auto-encoding variational Bayes. arXiv: 1312.6114, 2014
- [59] Kipf T N, Welling M. Variational graph auto-encoders. arXiv: 1611.07308, 2016
- [60] Hasanzadeh A, Hajiramezanali E, Duffield N, et al. Semi-implicit graph variational auto-encoders//Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver, Canada, 2019; 10711-10722
- [61] Yin M, Zhou M. Semi-implicit variational inference//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden, 2018; 5646-5655
- [62] Papamakarios G, Pavlakou T, Murray I. Masked autoregressive flow for density estimation//Proceedings of the 31st Advances in Neural Information Processing Systems. Long Beach, USA, 2017; 2338-2347
- [63] Zhou M. Infinite edge partition models for overlapping community detection and link prediction. Journal of Machine Learning Research, 2015, 38: 1135-1143
- [64] Simonovsky M, Komodakis N. GraphVAE: towards generation of small graphs using variational autoencoders//Proceedings of the 27th International Conference on Artificial Neural Networks. Rhodes, Greece, 2018; 412-422
- [65] Pan S, Hu R, Long G, et al. Adversarially regularized graph autoencoder for graph embedding//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018; 2609-2615
- [66] Vaibhav, Huang P-Y, Frederking R. RWR-GAE: random walk regularization for graph auto encoders. arXiv: 1908.04003, 2019
- [67] Pan J Y, Yang H J, Faloutsos C, et al. Automatic multimedia cross-modal correlation discovery//Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining. Washington, USA, 2004; 653-658
- [68] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. arXiv: 1301.3781, 2013
- [69] Li C, Welling M, Zhu J, et al. Graphical generative adversarial networks//Proceeding of the 32nd Conference on Neu-

- ral Information Processing Systems. Montréal, Canada, 2018; 6072-6083
- [70] Heskes T, Zoeter O. Expectation propagation for approximate inference in dynamic bayesian networks. *Computer Science*, 2013; 216-223
- [71] Taentzer G, Schween H. Movement of objects in configuration spaces modelled by graph grammars//*Proceedings of the Graph-Grammars and Their Application to Computer Science*, 4th International Workshop, Bremen, Germany, 1990; 660-675
- [72] Li Y, Vinyals O, Dyer C, et al. Learning deep generative models of graphs. *arXiv*: 1803.0332, 2018
- [73] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation//*Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar, 2014; 1724-1734
- [74] Liu W, Chen P Y, Yu F, et al. Learning graph topological features via GAN. *IEEE Access*, 2019, 7; 21834-21843
- [75] Wang H, Wang J, Wang J, et al. Graphgan: graph representation learning with generative adversarial nets//*Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. New Orleans, USA, 2018; 2508-2515
- [76] Bojchevski A, Shchur O, Zügner D, et al. NetGAN: generating graphs via random walks//*Proceedings of the 35th International Conference on Machine Learning*. Stockholm, Sweden, 2018; 609-618
- [77] Grover A, Leskovec J. Node2vec: scalable feature learning for networks//*Proceedings of the 22nd International Conference on Knowledge Discovery and Data Mining*. San Francisco, USA, 2016; 855-864
- [78] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. *arXiv*: 1701.07875, 2017
- [79] De Cao N, Kipf T. MolGAN: an implicit generative model for small molecular graphs. *arXiv*: 1805.11973, 2018
- [80] Gulrajani I, Ahmed I, Arjovsky M, et al. Improved training of Wasserstein GANs//*Proceedings of the 31st Advances in Neural Information Processing Systems*. Long Beach, USA, 2017; 5767-5777
- [81] Guimaraes G, Sanchez-Lengeling B, Outeiral C, et al. Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models. *arXiv*: 1705.10843, 2018
- [82] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training gans//*Proceedings of the 30th Conference on Neural Information Processing Systems*. Barcelona, Spain, 2016; 2226-2234
- [83] Fan S, Huang B. Labeled graph generative adversarial networks. *arXiv*: 1906.03220, 2019
- [84] Mirza M, Osindero S. Conditional generative adversarial nets. *arXiv*: 1411.1784, 2014
- [85] Odena A, Olah C, Shlens J. Conditional image synthesis with auxiliary classifier gans//*Proceedings of the 34th International Conference on Machine Learning*. Sydney, Australia, 2017; 2642-2651
- [86] Ding M, Tang J, Zhang J. Semi-supervised learning on graphs with generative adversarial nets//*Proceedings of the 27th International Conference on Information and Knowledge Management*. Torino, Italy, 2018; 913-922
- [87] Salimans T, Kingma D P. Weight normalization: a simple reparameterization to accelerate training of deep neural networks//*Proceedings of the 30th Advances in Neural Information Processing Systems*. Barcelona, Spain, 2016; 901-909
- [88] Almeida L B. Learning rule for asynchronous perceptrons with feedback in a combinatorial environment. *Artificial Neural Networks*, 1987; 102-111
- [89] Pineda F J. Generalization of back-propagation to recurrent neural networks. *Physical Review Letters*, 1987, 59(19): 2229-2232
- [90] Li Y, Zemel R, Brockschmidt M, et al. Gated graph sequence neural networks//*Proceedings of the 4th International Conference on Learning Representations*. San Juan, Puerto Rico, 2016
- [91] Johnson D D. Learning graphical state transitions//*Proceedings of the 5th International Conference on Learning Representations*. Toulon, France, 2017
- [92] Weston J, Bordes A, Chopra S, et al. Towards AI-complete question answering: a set of prerequisite toy tasks//*Proceedings of the 4th International Conference on Learning Representations*. San Juan, Puerto Rico, 2016
- [93] You J, Ying R, Ren X, et al. GraphRNN: generating realistic graphs with deep auto-regressive models//*Proceedings of the 35th International Conference on Machine Learning*. Stockholm, Sweden, 2018; 5694-5703
- [94] Seo Y, Defferrard M, Vandergheynst P, et al. Structured sequence modeling with graph convolutional recurrent networks//*Proceeding of the 25th International Conference on Neural Information Processing*. Siem Reap, Cambodia, 2018; 362-373
- [95] Xu D, Cheng W, Luo D, et al. Spatio-temporal attentive RNN for node classification in temporal attributed graphs//*Proceedings of the 38th International Joint Conference on Artificial Intelligence*. Macao, China, 2019; 3947-3953
- [96] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need//*Proceedings of the 31st Conference on Neural Information Processing Systems*. Long Beach, USA, 2017; 5998-6008
- [97] Velićević P, Cucurull G, Casanova A, et al. Graph attention networks//*Proceedings of the 6th International Conference on Learning Representations*. Vancouver, Canada, 2018
- [98] Salehi A, Davulcu H. Graph attention auto-encoders. *arXiv*: 1905.10715, 2019
- [99] Busbridge D, Sherburn D, Cavallo P, et al. Relational graph attention networks. *arXiv*: 1904.05811, 2018
- [100] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling rela-

- tional data with graph convolutional networks//Proceedings of the 15th Extended Semantic Web Conference. Heraklion, Greece, 2018: 593-607
- [101] Zhang J, Shi X, Xie J, et al. GaAN: gated attention networks for learning on large and spatiotemporal graphs//Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence. Monterey, USA, 2018: 339-349
- [102] Do K, Tran T, Nguyen T, et al. Attentional multilabel learning over graphs: a message passing approach. Machine Learning, 2019, 108(10): 1757-1781
- [103] Pham T, Tran T, Phung D, et al. Column networks for collective classification//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA, 2017: 2485-2491
- [104] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 1263-1272
- [105] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks//Proceedings of the 29th Advances in Neural Information Processing Systems. Montreal, Canada, 2015: 2377-2385
- [106] Lee J B, Rossi R, Kong X. Graph classification using structural attention//Proceedings of the 24th International Conference on Knowledge Discovery and Data Mining. London, UK, 2018: 1666-1674
- [107] Knyazev B, Taylor G W, Amer M R. Understanding attention and generalization in graph neural networks//Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver, Canada, 2019: 4204-4214
- [108] Xu K, Hu W, Leskovec J. How powerful are graph neural networks? //Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA, 2019
- [109] Chen H, Perozzi B, Al-Rfou R, et al. A tutorial on network embeddings. arXiv:1808.02590, 2018
- [110] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: online learning of social representations//Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining. New York, USA, 2014: 701-710
- [111] Mallea M D G, Meltzer P, Bentley P J. Capsule neural networks for graph classification using explicit tensorial graph representations. arXiv: 1902.08399, 2019
- [112] Borgwardt K M, Ong C S, Schönauer S, et al. Protein function prediction via graph kernels//Proceedings 30th International Conference on Intelligent Systems for Molecular Biology. Detroit, USA, 2005: 47-56
- [113] Shervashidze N, Borgwardt K M. Fast subtree kernels on graphs//Proceedings of the 23rd Advances in Neural Information Processing Systems. Vancouver, Canada, 2009: 1660-1668
- [114] Harchaoui Z, Bach F. Image classification with segmentation graph kernels//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA, 2007: 18-23
- [115] Weston Jason, Ratle Frédéric, Mobahi Hossein, et al. Deep Learning via semi-supervised embedding//Proceedings of the 25th International Conference on Machine Learning. Helsinki, Finland, 2008: 1168-1175
- [116] Yang Zhilin, W. Cohen William, Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings//Proceedings of the 33rd International Conference on Machine Learning. New York, USA, 2016: 1-9
- [117] Vashishth S, Bhandari M, Yadav P, et al. Incorporating syntactic and semantic information in word embeddings using graph convolutional networks//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 3308-3318
- [118] Guo Z, Zhang Y, Lu W. Attention guided graph convolutional networks for relation extraction//Proceedings of the 57th Conference of the Association for Computational Linguistics. Florence, Italy, 2019: 241-251
- [119] Fu T-J, Li P-H, Ma W-Y. GraphRel: modeling text as relational graphs for joint entity and relation extraction//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 1409-1418
- [120] Zhu H, Lin Y, Liu Z, et al. Graph neural networks with generated parameters for relation extraction//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 1331-1339
- [121] Tu M, Wang G, Huang J, et al. Multi-hop reading comprehension across multiple documents by reasoning over heterogeneous graphs//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 2704-2713
- [122] Qiu L, Xiao Y, Qu Y, et al. Dynamically fused graph network for multi-hop reasoning//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 6140-6150
- [123] Li C, Goldwasser D. Encoding social information with graph convolutional networks for political perspective detection in news media//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 2594-2604
- [124] Ding M, Zhou C, Chen Q, et al. Cognitive graph for multi-hop reading comprehension at scale//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 2694-2703
- [125] Duvenaud D, Maclaurin D, Aguilera-ipparraguirre J, et al. Convolutional networks on graphs for learning molecular fingerprints//Proceedings of the 29th Advances in Neural Information Processing Systems. Montreal, Canada, 2015: 2224-2232

- [126] Kearnes S, McCloskey K, Berndl M, et al. Molecular graph convolutions; moving beyond fingerprints. *Journal of Computer-Aided Molecular Design*, 2016, 30(8): 595-608
- [127] Gomes J, Ramsundar B, Feinberg E N, et al. Atomic convolutional networks for predicting protein-ligand binding affinity. *arXiv*: 1703.10603, 2017
- [128] Altae-Tran H, Ramsundar B, Pappu A S, et al. Low data drug discovery with one-shot learning. *ACS Central Science*, 2017, 3(4): 283-293
- [129] Ma T, Xiao C, Zhou J, et al. Drug similarity integration through attentive multi-view graph auto-encoders//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden, 2018: 3477-3483
- [130] Shi L, Zhang Y, Cheng J, et al. Skeleton-based action recognition with directed graph neural networks//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach, USA, 2019: 7912-7921
- [131] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos//*Proceedings of the 28th Advances in Neural Information Processing Systems*. Montreal, Canada, 2014: 568-576
- [132] Narasimhan M, Lazebnik S, Schwing A G. Out of the box; reasoning with graph convolution nets for factual visual question answering//*Proceedings of the 32nd Conference on Neural Information Processing Systems*. Montreal, Canada, 2018: 2659-2670
- [133] Qi X, Liao R, Jia J, et al. 3D graph neural networks for RGBD semantic segmentation//*Proceedings of the IEEE International Conference on Computer Vision*. Venice, Italy, 2017: 5209-5218
- [134] Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA, 2018: 4558-4567
- [135] Garcia V, Bruna J. Few-shot learning with graph neural networks//*Proceedings of the 6th International Conference on Learning Representations*. Vancouver, Canada, 2018
- [136] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks; a deep learning framework for traffic forecasting//*Proceedings of the 37th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden, 2018: 3634-3640
- [137] Diao Z, Wang X, Zhang D, et al. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting//*Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Honolulu, USA, 2019: 890-897
- [138] Li Y, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network; data-driven traffic forecasting//*Proceedings of the 6th International Conference on Learning Representations*. Vancouver, Canada, 2018
- [139] Kosaraju V, Sadeghian A, Martín-martín R, et al. Social-BiGAT: multimodal trajectory forecasting using bicycle-GAN and graph attention networks//*Proceedings of the 33rd Conference on Neural Information Processing Systems*. Vancouver, Canada, 2019:137-146
- [140] Zhu J-Y, Zhang R, Pathak D, et al. Toward multimodal image-to-image translation//*Proceedings of the 31st Conference on Neural Information Processing Systems*. Long Beach, USA, 2017: 465-476
- [141] Wang Z, Lv Q, Lan X, et al. Cross-lingual knowledge graph alignment via graph convolutional networks//*Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium, 2018: 349-357
- [142] Hamaguchi T, Oiwa H, Shimbo M, et al. Knowledge transfer for out-of-knowledge-base entities; a graph neural network approach//*Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne, Australia, 2017: 1802-1808
- [143] Bordes A, Usunier N, García-Durán A. Translating embeddings for modeling multi-relational data//*Proceedings of the 27th Advances in Neural Information Processing Systems*. Lake Tahoe, USA, 2013: 2787-2795
- [144] Li Z, Cui Z, Wu S, et al. Fi-GNN: modeling feature interactions via graph neural networks for CTR prediction//*Proceedings of the 28th International Conference on Information and Knowledge Management*. Beijing, China, 2019: 539-548
- [145] Monti F, Bronstein M M, Bresson X. Geometric matrix completion with recurrent multi-graph neural networks//*Proceedings of the 31st Conference on Neural Information Processing Systems*. Long Beach, USA, 2017: 3697-3707
- [146] Guo Jia-Yan, Li Rong-Hua, Zhang Yan, et al. Graph neural network based anomaly detection in dynamic networks. *Journal of Software*, 2020, 31(03):748-762(in Chinese)
(郭嘉琰,李荣华,张岩等.基于图神经网络的动态网络异常检测算法. *软件学报*, 2020, 31(03):748-762)
- [147] YI Dongyi, DENG Genqiang, DONG Chaoxiong, et al. Medical fraud detection algorithm based on graph. *Journal of Computer Applications*, 2020, 40(5):1272-1277(in Chinese)
(易东义,邓根强,董超雄等.基于图卷积神经网络的医保欺诈检测算法. *计算机应用*, 2020, 40(5):1272-1277)
- [148] Zhang M, Chen Y. Link prediction based on graph neural networks//*Proceedings of the 32nd Conference on Neural Information Processing Systems*. Montreal, Canada, 2018: 5171-5181
- [149] Kawamoto T, Tsubaki M, Obuchi T. Mean-field theory of graph neural networks in graph partitioning//*Proceedings of the 32nd Conference on Neural Information Processing Systems*. Montreal, Canada, 2018: 4366-4376
- [150] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 770-778
- [151] Ranzato M, Boureau Y-L, Le Cun Y. Sparse feature learn-

ing for deep belief networks//Proceedings of the 21st Advances in Neural Information Processing Systems. Vancouver, Canada, 2007: 1185-1192

- [152] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need//Proceedings of the 31st Advances in Neural Information Processing Systems. Long Beach, USA, 2017:

5998-6008

- [153] Tang J, Wang M, Zhang M, et al. LINE: large-scale information network embedding//Proceedings of the 24th International Conference on World Wide Web. Florence, Italy, 2015: 1067-1077



WU Bo, Ph. D. candidate. His research interests include graph neural network and deep learning.

LIANG Xun, Ph. D., professor, Ph. D. supervisor. His research interests include neural networks, social computing and natural language processing.

ZHANG Shu-Sen, Ph. D. candidate His research interests include data mining and social computing.

XU Rui, Ph. D. candidate. His research interests include image processing and machine learning.

Background

Over the past few years, more and more graph-structure data emerge in real life, such as Internet network, social network, protein network, compound molecules and others. Due to its complexity and heterogeneity, graph-structure data is hard to model. Based on this urgent need, GNN was proposed and try to find out the graph pattern. Through joint efforts, GNN continue to develop and improve and its excellent performance has aroused high attention and deep exploration.

As a kind of graph learning algorithm, GNN has strong ability to learn graph. By making certain strategies on the nodes and edges of the graph, GNN transforms the graph-structure data into standard representation, so as to be input into a variety of different neural networks for training, and achieve excellent results in tasks such as node classification, edge information dissemination and graph clustering. GNN can learn the internal rules and deeper graph representation of node and edge features in graph-structure data when it is compared with other graph learning algorithms. With its strong nonlinear fitting ability to graph-structure data, it makes GNN have higher accuracy and better robustness on graph-structure related problems in different fields.

This paper provides a comprehensive review of existing

research results in GNN. The authors first look back on the history of GNN and developments in recent years. Then they give an introduction of the related concepts and definitions used in GNN mathematical reasoning. The authors discuss and analyze different variants of GNN and list core ideas, learning methods and so on for each type of GNN. On this basis, the authors compare GNN with other graph learning algorithm and point out the advantages and disadvantages between them. The authors also elaborate the specific application of GNN in different scientific fields and explain how GNN works. During the end part of this paper, the authors outlines the future direction and development trend of GNN which means probable solutions to the existing problems and challenges.

This work is supported by the National Natural Science Foundation of China (No. 62072463, No. 71531012), the National Social Science Foundation of China (No. 18ZDA309), the Natural Science Foundation of Beijing (No. 4172032), the E-commerce Research Project of Jingdong Mall (No. 413313012), and the Open Project of the State Key Laboratory of Digital Publishing Technology of Peking University Founder Group Co., Ltd.