

# 简述图卷积神经网络\*

柳明

School of CEI/AI, Nanjing Normal University

## 摘要

卷积神经网络因其强大的建模能力引起研究者的广泛关注，在图像处理、自然语言处理等领域得到了成功应用。然而，传统卷积神经网络只能处理欧式空间数据，而在许多应用场景如社交网络、交通网络、蛋白质网络中，数据都是以图数据的形式呈现的。本文首先介绍了图卷积神经网络的背景，之后就图卷积神经网络的两类主流方法展开简要叙述和梳理：谱方法基于谱图理论和卷积定理在谱域定义图卷积，空间方法通过直接在结点域定义结点相关性实现图卷积。最后，本文介绍了图卷积神经网络的相关应用，并对图卷积神经网络的发展做了一些总结。

## 1. 引言

卷积神经网络及其催生的一系列方法在过去近十年中得到了快速发展 [11, 15, 19, 37]，它通过稀疏交互 (sparse interactions)、参数共享 (parameter sharing)、等变表示 (equivariant representations) 这三个重要的思想来帮助改进神经网络模型。卷积神经网络较全连接神经网络而言，通过减少连接数、共享参数等方法削弱了网络的表达能力上限，但使得整个网络更容易被训练。卷积神经网络在图像处理和自然语言处理等领域的一些具体问题上如机器翻译 [14]、图像识别 [11] 和语音识别 [12] 等表现出了强大的建模能力。

但是，传统的卷积神经网络只能处理欧式结构的数据 (Euclidean data) 如图像、文本等，这些领域的数据具有平移不变性。平移不变性 (translation invariance) 使得我们可以在输入数据的数据空间定义全局共享的卷积核，从而定义卷积神经网络。以图像数据为例，一张图片可以表示为欧式空间一组网格状规则散布的像素点，那么平移不变性则表示以任意像素点为中心，可以获取到相同尺寸的局部结构。基于这些性质，卷积神经网络通过在训练过程中学习被每个像素点共享的卷积核参数 (权重和偏置) 来对局部连接进行建模，从而学习到丰富的隐层表示来体现图片的不同模式。

非欧式结构的数据广泛存在于科学研究和社会研究等领域，典型的例子有社交网络、交通网络、物体形状识别 [27] 的网状表面 (meshed surfaces) 和通信的传感器网络等。这些例子中的数据都可以将其拓扑结构抽象为图数据，不同于图像和文本数据，对图中的每个结点来说，其局部结构都各不相同，这就使得此类数据不具有平移不变性 [36]。平移不变性的缺失对在图数据上重新定义卷积神经网络提出了挑战。

由于非欧数据的普遍性，越来越多的研究者尝试对已有的深度学习模型进行改造或类比已有方法、概念重新在图上定义某些操作，以满足对此类数据进行处理的需求。图数据上普遍存在的结点相邻和依赖关系与卷积神经网络对局部结构的强大建模能力让图卷积神经网络成为图深度学习领域相当活跃和重要的一个社区。

---

\*本文的编写基于 CVPR'11 论文 [2] 的 L<sup>A</sup>T<sub>E</sub>X 模板。

在模式识别课程的文献阅读过程中，我们尝试结合几篇较有代表性的文章，在本文中对图卷积神经网络的发展历程进行一点简单的整理和总结。

### 1.1. 图卷积神经网络面临的挑战

前面部分介绍过，图数据因为不满足平移不变性而难以直接用传统的卷积神经网络进行处理，除此以外，将卷积神经网络迁移到图数据上进行应用还存在其他难点。构建图卷积神经网络过程中的挑战主要来源于以下几个方面：

**图数据是非欧空间数据。** 图数据作为非欧空间数据，不满足平移不变性，即每个结点的局部结构各异。传统卷积神经网络的两个基本算子（卷积和池化）都依赖于数据的平移不变性。如何在图数据上重新定义卷积和池化算子成为了首要的挑战性工作。

**图数据具有多样的特性。** 实际生活中的多种场景都可以用图数据自然地表示，这使得这些具体场景下的图数据具有各种各样的特性，如社交网络中用户之间的有向连接，文献引用网络中作者和引文的异质连接，政治关系网络中的正负倾向带符号连接等。图的多样性使得图数据中蕴含的信息更加丰富，但是对多种特性进行建模也要求图卷积神经网络更加复杂精细，这也是图卷积神经网络面临的一个挑战。

**图数据的规模往往很大。** 在实际的应用场景下，图的规模可能极大，含有百万甚至千万级别的结点，如推荐系统中的用户商品网络，社交网络中的用户网络。如何在可接受的空间和时间范围要求内，在大规模图数据上构建图卷积神经网络也是一大挑战。

### 1.2. 图卷积神经网络的任务

图数据建模所针对的应用场景非常广泛，这也使得图数据建模所处理的任务也多种多样。对图数据建模可以求解图级别（graph-level）、结节点级别（node-level）和边级别（edge-level）三种层面的问题。

图级别的任务可以描述为输入图数据，输出每个图的类别，包括图生成、图分类等，如蛋白质网络中的蛋白质分类、分子毒性预测。结节点级别的任务的要求是输入结点不带类别的图，输出每个结点的级别，包括节点分类等，如引用网络中的文章分类。边级别的任务则用来预测结点之间的相互关系，包括链接预测（link prediction）等，如推荐系统中用户对商品的偏好推断。

### 1.3. 图卷积神经网络方法及分类

图数据建模的历史由来已久。最初，研究人员更加关注统计分析的方法，如谷歌创始人提出的网页排序算法 PageRank [31]等，机器学习模型在初期并没有参与到图数据处理的发展中。此外，谱图理论（spectral graph theory）也常被应用到其中，如吴恩达等提出的谱聚类方法 [28]可以用拉普拉斯矩阵的特征值和特征向量做社区分析或人群聚类等。随着近年来深度学习的崛起，研究者们也尝试将深度学习模型与图数据结合起来，这一阶段受 Mikolov 等提出的 word2vec 学习词嵌入 [26]的启发，代表性的研究工作是图嵌入（graph embedding），即通过结点的邻近性为每个结点学习固定长度的低维、稠密的嵌入表示，如 DeepWalk [33]、LINE [38]、node2vec [7]等。这一时期，在解决具体的应用问题时，建模过程通常被分为两阶段：第一阶段，学习每个结点的统一长度的表示（即 embedding）；第二阶段，将学习到的结点表示作为输入，训练相应的分类模型。而近年来，研究重心也逐渐转移到如何将深度学习模型与图数据结合，直接进行端到端的学习。

图卷积神经网络的本质目的是用来提取拓扑图的空间特征，拓扑图源于从实际问题中抽取出的关系网络。实现这个目标有两种主流方式：最早提出的方法借助谱图理论将图上信息通过图傅里叶变换变换到谱域(spectral domain)，结合卷积定理(convolution theorem)在谱域定义并实现卷积操作，谱方法的主要挑战是这样处理之后，定义在谱域的卷积核在空间上没有局部连接的性质；第二种方法是在空间域(spatial domain)或称顶点域(vertex domain)中直接提取邻居结点信息进行处理，空间方法将卷积算子重新定义为对目标顶点的邻居结点信息进行加权平均的函数，这种方法主要的挑战是不同结点的邻居数量可能差异巨大，例如有的图中顶点度数大致服从幂律分布(power-law degree distribution)。

## 2. 谱方法及基于谱方法的图卷积神经网络

图数据上平移不变性的确实，给在结点域定义卷积神经网络带来困难。谱方法基于谱图理论和卷积定理在谱域定义图卷积。本小节我们先介绍相关的背景知识和基本原理。

### 2.1. 谱方法的基本原理

为了便于后文叙述，首先我们在表1中给出本文中一些常见符号及其基本含义。

符号	含义
$G$	图
$V$	结点集合
$E$	连边集合
$n$	结点个数
$I_n \in R^{n \times n}$	单位矩阵
$A \in R^{n \times n}$	邻接矩阵
$D \in R^{n \times n}$	度矩阵(对角阵)
$L \in R^{n \times n}$	拉普拉斯矩阵
$U \in R^{n \times n}$	特征向量矩阵
$\Lambda \in R^{n \times n}$	特征值矩阵(对角阵)
$u_i \in R^n$	第 $i$ 个特征向量
$X \in R^{n \times d}$	结点特征矩阵
$X_i \in R^d$	第 $i$ 个结点的特征
$f \in R^n$	信号

表 1. 符号定义。

我们用  $G = (V, E, A)$  表示无向图，其中  $V$  表示图中结点组成的集合， $n = |V|$  表示图上的结点数， $E$  表示边集合， $A$  表示邻接矩阵，定义了结点之间的连接关系，且在无向图中有  $A_{i,j} = A_{j,i}$ 。 $D$  是图的度数矩阵， $D_{i,i}$  即为第  $i$  个结点的度且  $D_{i,i} = \sum_j A_{i,j}$ 。

谱图理论中常用的拉普拉斯矩阵(Laplacian matrix)定义为  $L = D - A$ 。另外，在图表示学习领域，常用的拉普拉斯矩阵及其变体有三种。组合拉普拉斯矩阵(combinatorial Laplacian)即为刚才提到的由度数矩阵和邻接矩阵相减定义的形式  $L = D - A$ ；随机游走领域常用的随机游走归一化拉普拉斯矩阵(random walk normalized Laplacian)定义为

$$L^{rw} = D^{-1}L = I_n - D^{-1}A \quad (1)$$

还有在图卷积神经网络领域应用广泛的对称规一化拉普拉斯矩阵 (symmetric normalized Laplacian), 定义为

$$\mathbf{L}^{sys} = \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}} = \mathbf{I}_n - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \quad (2)$$

显然  $\mathbf{L}$  是一个实对称矩阵, 可以被正交对角化, 即对  $\mathbf{L}$  做特征分解  $\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$ , 其中  $\mathbf{U} = \{\mathbf{u}_i\}_{i=1}^n = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  表示  $n$  个相互正交的特征向量 (eigenvector) 组成的变换矩阵,  $\mathbf{\Lambda} = \text{diag}[\lambda_1, \dots, \lambda_n]$  是一个对角阵,  $\lambda_i$  表示与特征向量  $\mathbf{u}_i$  相对应的特征值。

$\mathbf{X}$  表示图  $G$  上的结点特征 (features), 其中  $\mathbf{X} \in R^{n \times d}$ , 每个结点有  $d$  个特征, 行向量  $\mathbf{X}_i \in R^d$  是第  $i$  个结点的特征向量, 列向量  $\mathbf{x}_j \in R^n$  表示图上结点的第  $j$  个特征 (即为定义在结点上的一个信号), 元素  $\mathbf{X}_{i,j}$  表示矩阵  $\mathbf{X}$  的第  $i$  行第  $j$  列, 即第  $i$  个结点的第  $j$  个特征,  $\mathbf{f}, \mathbf{x}, \mathbf{y} \in R^n$  表示图上的信号,  $\mathbf{f}_i$  表示结点  $i$  在信号  $\mathbf{f}$  上的取值。

图信号处理领域的学者将传统信号处理中使用的傅里叶变换 (Fourier transform) 推广到了图上 [36]。传统的傅里叶变换定义为

$$F(\omega) = \mathcal{F}[f(t)] = \int f(t) e^{-i\omega t} dt \quad (3)$$

即为信号  $f(t)$  与基函数  $e^{-i\omega t}$  乘积的积分。而时频傅里叶的基函数  $e^{-i\omega t}$  是拉普拉斯算子的特征函数,  $\omega$  和特征值有关。广义的特征方程定义为  $AV = \lambda V$ , 其中  $A$  与本文中的邻接矩阵无关, 只是一种变换,  $V$  是特征 (eigen-) 向量或特征函数,  $\lambda$  是特征值。函数  $e^{-i\omega t}$  满足

$$\Delta e^{-i\omega t} = \frac{\partial^2}{\partial t^2} e^{-i\omega t} = -\omega^2 e^{-i\omega t} \quad (4)$$

根据广义的特征方程定义,  $e^{-i\omega t}$  就是  $\Delta$  的特征函数。

在谱图理论 [3] 中有这样一个结论: 一个图的拉普拉斯矩阵, 就等价于定义在这个图上的拉普拉斯算子 (二阶微分)。那么回到广义特征方程  $AV = \lambda V$ , 用拉普拉斯矩阵  $\mathbf{L}$  代替变换  $A$ , 得到  $\mathbf{L}V = \lambda V$ 。类比时频傅里叶的推导过程, 我们就可以用拉普拉斯矩阵的特征向量作为图傅里叶变换的基。以拉普拉斯矩阵的特征向量作为谱空间下一组基底, 那么图上信号  $\mathbf{x} \in R^n$  的傅里叶变换为

$$\hat{\mathbf{x}} = \mathbf{U}^T \mathbf{x} \quad (5)$$

其中,  $\mathbf{x}$  为信号在结点域的原始表示,  $\hat{\mathbf{x}}$  为信号  $\mathbf{x}$  变换到谱域之后的表示,  $\mathbf{U}^T$  表示特征向量矩阵的转置, 用于做傅里叶变换。同样, 信号  $\mathbf{x}$  的傅里叶逆变换为

$$\mathbf{x} = \mathbf{U} \hat{\mathbf{x}} \quad (6)$$

在此基础上, 谱方法利用卷积定理在谱域上定义了图卷积。关于卷积定理可以叙述为: 信号卷积的傅里叶变换等价于信号傅里叶变换后的乘积, 形式化描述为

$$\mathcal{F}(\mathbf{x} * \mathbf{y}) = \mathcal{F}(\mathbf{x}) \cdot \mathcal{F}(\mathbf{y}) \quad (7)$$

其中,  $\mathbf{x}$  和  $\mathbf{y}$  表示两个原始信号,  $*$  表示卷积算子,  $\cdot$  表示乘积算子。对式 7 等号两侧同时做傅里叶逆变换, 可以得到

$$\mathbf{x} * \mathbf{y} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{x}) \cdot \mathcal{F}(\mathbf{y})) \quad (8)$$

其中,  $\mathcal{F}^{-1}(\mathbf{x})$  表示信号  $\mathbf{x}$  的傅里叶逆变换。

根据以上理论基础，由式5、式6和式8，可将图上的卷积写为

$$\mathbf{x} *_G \mathbf{y} = \mathbf{U} ((\mathbf{U}^T \mathbf{x}) \odot (\mathbf{U}^T \mathbf{y})) \quad (9)$$

其中， $*_G$  表示图上的卷积算子， $\odot$  为哈达玛积 (Hadamard product) 即将向量、矩阵对应位置元素相乘，可以将信号  $\mathbf{x}$  视为输入信号，信号  $\mathbf{y}$  视为卷积核，在谱空间的卷积核是  $\mathbf{U}^T \mathbf{y}$ 。

我们令  $\mathbf{U}^T \mathbf{y} = [\theta_0, \dots, \theta_{n-1}]^T$ ， $g_\theta = \text{diag}([\theta_0, \dots, \theta_{n-1}])$ ，那么就可以用对角阵  $g_\theta$  代替列向量  $\mathbf{U}^T \mathbf{y}$ ，之后就可以将哈达玛积运算转化为简洁的矩阵乘法形式，从而将式9化成

$$\mathbf{x} *_G \mathbf{y} = \mathbf{U} g_\theta \mathbf{U}^T \mathbf{x} \quad (10)$$

经过以上推导，我们得到了图上卷积算子的定义，基于此定义，谱方法一支陆续出现了一些图卷积神经网络模型。

## 2.2. 基于谱方法的图卷积神经网络

本小节基于前述的基本原理简单介绍几个具有代表性的基于谱方法的图卷积神经网络工作。

**Spectral CNN.** 谱卷积神经网络 (Spectral CNN) [1]由 LeCun 组的 Bruna 等人在 2014 年提出，是最早被提出的构建图卷积神经网络的方法，他们从基本的谱图理论出发并结合卷积定理，在谱空间定义了图卷积。该方法利用卷积定理在每一层定义图卷积算子，并堆叠多层组成神经网络。

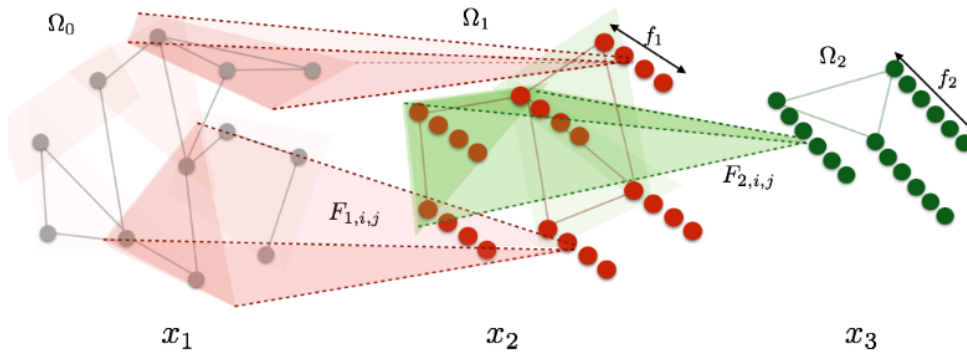


图 1. 谱卷积神经网络。

谱卷积神经网络第  $m$  层的前向传播过程 (公式和符号等均采用本文记法，与原文中有不同) 如下

$$\mathbf{x}_j^{m+1} = h \left( \mathbf{U} \sum_{i=1}^p F_{i,j}^m \mathbf{U}^T \mathbf{x}_i^m \right), \quad j = 1, \dots, q \quad (11)$$

其中， $p$  和  $q$  分别是输入特征 (features) 和输出特征的维度，即第  $m$  层和第  $m+1$  层的通道数， $\mathbf{x}_i^m \in R^n$  表示图上结点在第  $m$  层的第  $i$  个输入特征， $F_{i,j}^m$  表示谱空间下第  $m$  层的可学习的卷积核，是一个对角阵， $h$  表示非线性激活函数。在谱卷积神经网络中，这样一层结构将特征 (features) 从  $p$  维转化到  $q$  维，且基于卷积定理通过学习卷积核的参数实现了图卷积。

总的来说，谱卷积神经网络将卷积核作用在谱空间的输入信号上，并利用卷积定理实现图卷积，以完成结点之间的信息聚合，然后将非线性激活函数作用在信息聚合的结果上，堆叠多层形成神经网络。但是，谱



卷积神经网络也存在一些缺陷：首先，谱卷积神经网络方法要求对图的拉普拉斯矩阵进行特征分解（eigendecomposition），因为图卷积的求解过程中显式地使用了拉普拉斯矩阵的特征向量，并且每一次前向传播，都要计算  $U$ 、卷积核（对角阵）和  $U^T$  三者的矩阵乘积，计算代价非常高，论文中也提到这会有  $O(n^3)$  的计算复杂度；其次，卷积核有  $n$  个参数，而数  $n$  直接受到图规模的影响，尤其对于大规模图来说；最后，谱卷积神经网络不满足局部连接（spatial localization）的性质，因为卷积核作用在信号  $\mathbf{x}_i$  上，而信号  $\mathbf{x}_i$  含有全部结点第  $i$  个特征的信息，意味着每次信息聚合其实都有全局信息的参与。

**ChebNet.** Spectral CNN 提出之后出现了一些工作致力于实现图卷积神经网络的局部性和加速计算，Defferrard 等人提出的切比雪夫网络（ChebNet）[4]通过参数化卷积核实现局部性，也降低了参数复杂度和计算复杂度。

在式10中， $g_\theta$  是需要学习的卷积核，在谱卷积神经网络中， $g_\theta$  是一个有  $n$  个待学习参数的对角阵。Defferrard 等在文章中提出，对于卷积核  $g_\theta$ ，我们如果对其用一个  $K$  阶多项式进行拟合：

$$g_\beta(\Lambda) \approx \sum_{k=0}^{K-1} \beta_k \Lambda^k \quad (12)$$

其中， $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ 。将式12代入式10，图卷积可化为

$$\begin{aligned} \mathbf{x} *_G \mathbf{y} &= U g_\beta(\Lambda) U^T \mathbf{x} \\ &= U \sum_{k=0}^{K-1} \beta_k \Lambda^k U^T \mathbf{x} \\ &= \sum_{k=0}^{K-1} \beta_k (U \Lambda^k U^T) \mathbf{x} \\ &= \sum_{k=0}^{K-1} \beta_k \mathbf{L}^k \mathbf{x} \end{aligned} \quad (13)$$

这样就将参数个数从  $n$  个减少到  $K$  个，一般  $K$  远远小于  $n$ ；同时也不需要进行矩阵的特征分解，之前用傅里叶变换的基  $U^T$  乘信号  $\mathbf{x}$  的计算复杂度是  $O(n^2)$ ，而现在由于式13中拉普拉斯矩阵  $\mathbf{L}$  是一个比较稀疏的矩阵，这一步的运算量与结点数  $n$  无关，只和边数有关，所以这一步的计算复杂度可以被降低到  $O(|E|)$ 。

具体地，ChebNet 引入了切比雪夫多项式（Chebyshev polynomial） $T_k$  对  $g_\theta$  进行拟合。切比雪夫多项式的递推定义为

$$\begin{cases} T_0(x) = 1 \\ T_1(x) = x \\ T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x) \end{cases} \quad (14)$$

$g_\beta$  的  $K$  阶近似为

$$g_\beta(\Lambda) \approx \sum_{k=0}^{K-1} \beta_k T_k(\tilde{\Lambda}) \quad (15)$$

其中， $\tilde{\Lambda} = 2\Lambda/\lambda_{max} - \mathbf{I}_n$ ， $\lambda_{max}$  是拉普拉斯矩阵最大的特征值，拉普拉斯矩阵是半正定矩阵，其特征值都非负（有特征值 0），这个操作将特征值压缩到  $[-1, 1]$  的范围中，可以防止梯度爆炸。那么根据式15和式10，

我们可以将切比雪夫多项式代入化简，具体地，图卷积可以写成

$$\begin{aligned}
\mathbf{x} *_G \mathbf{y} &= \mathbf{U} g_\beta(\Lambda) \mathbf{U}^T \mathbf{x} \\
&= \mathbf{U} \sum_{k=0}^{K-1} \beta_k T_k(\tilde{\Lambda}) \mathbf{U}^T \mathbf{x} \\
&= \sum_{k=0}^{K-1} \beta_k \left( \mathbf{U} T_k(\tilde{\Lambda}) \mathbf{U}^T \right) \mathbf{x} \\
&= \sum_{k=0}^{K-1} \beta_k T_k(\mathbf{U} \tilde{\Lambda} \mathbf{U}^T) \mathbf{x} \\
&= \sum_{k=0}^{K-1} \beta_k T_k(\tilde{\mathbf{L}}) \mathbf{x}
\end{aligned} \tag{16}$$

其中，和  $\tilde{\Lambda}$  类似地有  $\tilde{\mathbf{L}} = 2\mathbf{L}/\lambda_{max} - \mathbf{I}_n$ 。式16中第三步到第四步的推导  $\mathbf{U} T_k(\tilde{\Lambda}) \mathbf{U}^T = T_k(\mathbf{U} \tilde{\Lambda} \mathbf{U}^T)$  用数学归纳法可以很容易证得。

基于以上内容，切比雪夫网络第  $m$  层的前向传播过程可以记为

$$\begin{aligned}
\mathbf{x}_j^{m+1} &= h \left( \mathbf{U} \sum_{i=1}^p \left( \sum_{k=0}^{K-1} \beta_k T_k(\tilde{\Lambda}) \right) \mathbf{U}^T \mathbf{x}_i^m \right), \\
&= h \left( \sum_{i=1}^p \sum_{k=0}^{K-1} \beta_k T_k(\tilde{\mathbf{L}}) \mathbf{x}_i^m \right), \quad j = 1, \dots, q
\end{aligned} \tag{17}$$

与前文记法相同， $p$  和  $q$  分别是输入特征（features）和输出特征的维度，即第  $m$  层和第  $m+1$  层的通道数， $\mathbf{x}_i^m \in R^n$  表示图上结点在第  $m$  层的第  $i$  个输入特征， $h$  表示非线性激活函数。

切比雪夫网络利用特征值矩阵的多项式参数化卷积核实现了谱卷积神经网络，而且巧妙地利用了  $\mathbf{L} = \mathbf{U} \Lambda \mathbf{U}^T$  的关系引入拉普拉斯矩阵，从而避免了对拉普拉斯矩阵的特征分解，同时参数的复杂度从  $O(n \times p \times q)$  下降到  $O(K \times p \times q)$ 。

此外，参数化的卷积核很好地保证了局部性，式13中的  $\mathbf{L}^k$  就体现了局部连接特性，在拉普拉斯矩阵中，当且仅当结点  $i, j$  满足  $K$  跳可达时， $\mathbf{L}_{i,j}^K \neq 0$ ，这里和离散数学或图论中提到过的邻接矩阵  $k$  次幂的意义类似，代表着  $k$  阶可达关系。特别地， $K$  就控制着卷积核的感受野（receptive field）大小，也就是说每次卷积会将距离中心顶点  $K$  跳之内的所有邻居上的特征（features）信息进行加权求和，权值就是在模型中学习的参数  $\beta_k$ 。 $K=1$  和  $K=2$  的情形如图2和图3所示。

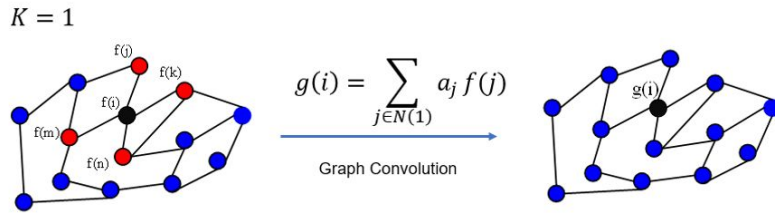


图 2. 切比雪夫网络  $K=1$  时的图卷积示意。

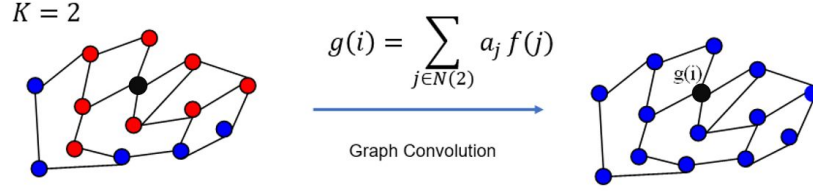


图 3. 切比雪夫网络  $K = 2$  时的图卷积示意。

**Graph Wavelet Neural Network.** 图小波神经网络 [46] 由中科院计算所徐冰冰等人在 ICLR 2019 上发表，不同于切比雪夫网络对谱卷积神经网络中的卷积核进行多项式参数化而达到实现局部性和加速计算的方式，小波神经网络更换了傅里叶变换的基底，用小波变换代替傅里叶变换，同样也满足了局部性和加速计算的要求。

小波神经网络指出，小波变换也定义了一种将信号从结点域变换到谱域的方法 [10]。这里用

$$\Psi_s = \{\psi_{s1}, \psi_{s2}, \dots, \psi_{sn}\} \quad (18)$$

表示小波变换的基底，其中  $\psi_{si}$  表示从第  $i$  个结点出发的能量扩散，刻画了第  $i$  个结点的局部结构。小波基底的定义依赖于拉普拉斯矩阵的特征向量，即  $\Psi_s = \mathbf{U} \mathbf{G}_s \mathbf{U}^T$ ，其中  $\mathbf{G}_s = \text{diag}(\{g_s(\lambda_i)\}_{i=1}^n)$ ，对角线元素由  $g$  函数作用在特征值上得到。选取不同的  $g$  函数会赋予小波基底不同的性质，在小波神经网络中，作者使用了热核 (heat kernel) 函数，即  $g_s(\lambda_i) = e^{-s\lambda_i}$ 。

以  $\Psi_s$  为谱空间的基底，则图上小波逆变换的变换矩阵为  $\Psi_s^{-1} = \mathbf{U} \mathbf{G}_{-s} \mathbf{U}^T$ ，其中  $\mathbf{G}_{-s}$  表示将上述  $g$  函数替换为  $g_{-s}(\lambda_i) = e^{-s\lambda_i}$ 。

那么，图卷积的定义式 9 可以被改写成

$$\mathbf{x} *_G \mathbf{y} = \Psi \left( (\Psi^{-1} \mathbf{x}) \odot (\Psi^{-1} \mathbf{y}) \right) \quad (19)$$

图小波神经网络第  $m$  层的前向传播过程可以记为

$$\mathbf{x}_j^{m+1} = h \left( \Psi_s \sum_{i=1}^p F_{i,j}^m \Psi_s^{-1} \mathbf{x}_i^m \right), \quad j = 1, \dots, q \quad (20)$$

同样地， $p$  和  $q$  分别是输入特征 (features) 和输出特征的维度， $\mathbf{x}_i^m \in R^n$  表示图上结点在第  $m$  层的第  $i$  个输入特征， $F_{i,j}^m$  表示谱空间下第  $m$  层的可学习的卷积核，是一个对角阵， $h$  表示非线性激活函数。

由式 20 可以看出，在定义的图小波神经网络中，每层的参数的复杂度为  $O(n \times p \times q)$ ，其中  $n$  是结点数， $p$  是当前层中每个顶点的特征 (feature) 数， $q$  是下一层中每个顶点的特征数。大量的自由参数需要更大数据量的训练数据来学习，会对网络的表达能力造成影响，针对这个缺陷，作者提出可以对原先的卷积操作操作进行解耦，分解为特征变换和图卷积两步：

$$\text{feature transformation : } \mathbf{X}^{m'} = \mathbf{X}^m \mathbf{W} \quad (21)$$

$$\text{graph convolution : } \mathbf{X}^{m+1} = h \left( \Psi_s \mathbf{F}^m \Psi_s^{-1} \mathbf{X}^{m'} \right) \quad (22)$$

其中， $\mathbf{W} \in R^{p \times q}$  是特征变换矩阵， $\mathbf{X}^m \in R^{n \times p}$  为第  $m$  层的输入张量 (tensor)， $\mathbf{X}^{m+1} \in R^{n \times q}$  为第  $m$  层的输出张量， $\mathbf{F}^m$  是图卷积核 (对角阵)。这个解耦方法有效减少了参数量，将参数复杂度从  $O(n \times p \times q)$  降低到了  $O(n + p \times q)$ ，保持了模型较好的表达能力。



和傅里叶变换相比，小波基底具有一些很好的性质：（1）小波变换的基底可以通过切比雪夫多项式近似得到，也可以避免拉普拉斯矩阵特征分解带来的高计算代价；（2）小波变换的基底有局部性；（3）小波基底的局部性使得小波变换矩阵非常稀疏，大大降低了  $\Psi_s^{-1}\mathbf{x}$  的计算复杂度；（4）热核函数中的超参数  $s$  表示热量扩散的范围，通过调节超参数  $s$  可以灵活地让小波网络适应于不同的任务场景。不同  $s$  值下的小波变换基底如图4所示。图4中左图表示在  $s$  较小时，以黄色结点为中心进行的热量扩散；右图表示当  $s$  增大时，热量扩散范围也变大。

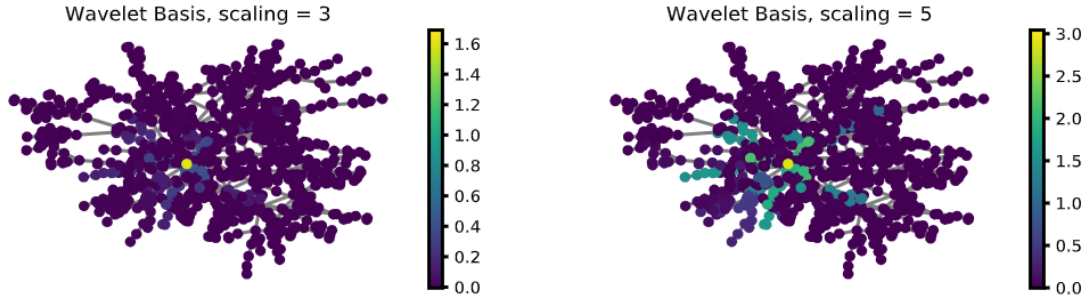


图 4. 不同  $s$  值下的小波变换基底。

文中 [46]也在顶点分类任务的 3 个数据集上对比了小波神经网络与其他方法。具体结果如表2所示，可以看出图小波神经网络的效果较优于其他谱方法。

Method	Cora	Citeseer	Pumbmed
MLP	55.1%	46.5%	71.4%
ManiReg	59.5%	60.1%	70.7%
DeepWalk	67.2%	43.2%	65.3%
ICA	75.1%	69.1%	73.9%
Spectral CNN	73.3%	58.9%	73.9%
ChebNet	81.2%	69.8%	74.4%
GCN	81.5%	70.3%	79.0%
MoNet	81.7 ± 0.5%	—	78.8 ± 0.3%
<b>GWNN</b>	<b>82.8%</b>	<b>71.7%</b>	<b>79.1%</b>

表 2. GWNN 在顶点分类任务上的结果。

### 3. 空间方法及基于空间方法的图卷积神经网络

上述方法都是从卷积定理出发在谱域定义图卷积，空间方法旨在从结点域出发，通过在结点域上直接定义聚合函数（aggregator）来聚合每个中心结点和它的邻近结点的信息，例如切比雪夫网络等可以视为以拉普拉斯矩阵或其变体形式作为聚合函数。在此启发下，开始出现一些工作通过注意力机制或其他方法直接从结点域学习聚合函数。此外，也有一部分工作通过类似传统卷积神经网络的工作方式，从空间角度定义了图卷积神经网络的通用框架并尝试解释图卷积神经网络的内部机制。

### 3.1. 空间方法的基本原理

Niepert 等人在 2016 年提出了一种可以在任意图上构造卷积神经网络的通用框架 [30]，在论文中作者对传统卷积神经网络在图片（image）上的工作流程进行拆解，重新审视了传统的卷积过程，通过类比的方法将卷积过程重新在结点域上定义。下面本小节以此为代表对空间方法的基本思路进行简单的介绍。

图5中描述了一个大小为  $3 \times 3$  的卷积核在  $4 \times 4$  的图片上滑动的过程。作者将传统卷积神经网络中的卷积操作分解为 3 个步骤进行分析：（1）确定中心结点的若干个邻居结点，图片是由规整排列的像素点构成的，可以看做是一种网格化（grid-like）的图，在示例图中，卷积核窗口里就包括了中心结点（图中标红）及其邻居；（2）给中心结点的若干个邻居结点确定一个顺序，在传统卷积神经网络中我们并未考虑过这一步，因为图片的像素点是自然有序的，或者说大家都默认了这样一种从左到右、从上到下的定序方式；（3）参数共享是卷积神经网络中的重要机制，以图5中过程为例， $3 \times 3$  的卷积核每次都都对原图像上同样  $3 \times 3$  大小的一个局部区域进行特征提取，这些局部区域中，位置相同的像素结点（如每次滑动处理时的中心结点，在图中标注了序号）在进行计算的时候的权值和偏置是共享的。

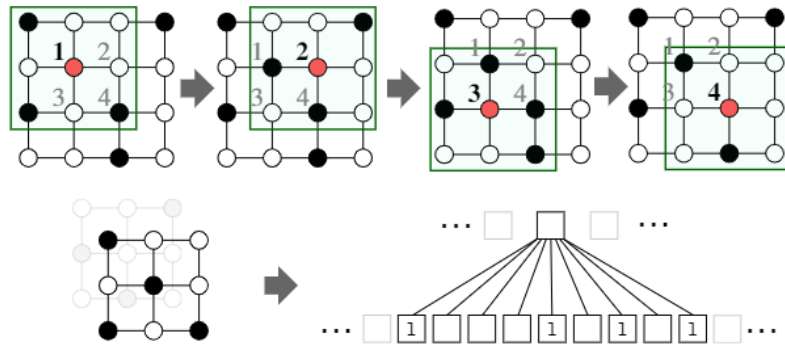


图 5. 一个卷积核大小为  $3 \times 3$  的传统卷积神经网络。（步长为 1，无填充）

为了更好地类比传统卷积神经网络的工作流程和图数据上的操作，我们再将这 3 个步骤都放到处理图数据的情景下进行分析：（1）在实际应用场景下的图数据上，每个结点的连接性可能差异非常大，有的结点邻居非常多而有的只和很少一些结点相连；（2）对提取出的结点进行定序也是一个需要解决的问题，不像图片有着天生的规整结构，在图数据中每次提取出的子图是形态各异的，这也要求研究者采取一些规则化的定序方法；（3）关于参数共享也是和第一步有相同的难处，如果前一个窗口中中心结点 a 的邻居个数为 10，后一个窗口中相对应的中心结点 b 却有 10000 个邻居，显然不可能直接让它们之间共享参数，但只要能够完成确定邻居、提取子图再定序的步骤，参数共享这一步工作也自然水到渠成。

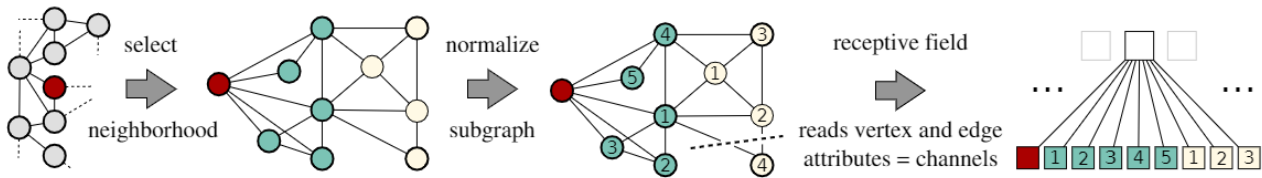


图 6. 将卷积操作解构之后迁移到图上。

作者提出的方法大致思路如图6中所示。首先，对每个中心结点，都选取  $k - 1$  ( $k$  是超参数) 个邻居结

点，从而提取出的一个大小为  $k$  的邻域集合；对于定序，可以根据具体任务选择不同的相似性度量，例如根据结点的度定序、根据到中心结点的跳数定序等度量方式；最后，就得到了长度为  $k$  的一个有序序列，这样就可以直接用经典卷积神经网络的工具进行处理，其中边的属性和顶点的属性对应不同的输入通道。

基于此类思路，有更多的基于空间方法的图卷积神经网络被提出。

### 3.2. 基于空间方法的图卷积神经网络

本小节基于前述的基本原理简单介绍几个具有代表性的基于空间方法的图卷积神经网络工作。

**GraphSAGE.** 图采样聚合网络 [9] 在 2017 年由 Hamilton 等人提出，其中“SAGE”也就直接概括了该模型对数据的处理方法：采样和聚合（SAmple and aggreGatE）。不同于之前许多模型考虑所有邻近结点的做法，图采样聚合网络采取的策略是对邻近结点做随机采样，使每个结点的邻近结点都不多于给定的采样个数。图采样聚合网络的结构和工作流程如图 7 所示。以正红色结点为目标结点，图采样聚合网络首先对其一阶邻居和二阶邻居做随机采样，并仅把采样到的结点作为相关结点，然后模型将聚合函数作用在相关结点的特征表达上，并用聚合结果更新目标结点的特征表达来完成对应任务。

其中，先对邻近结点进行随机采样可以降低计算的复杂度（图中一阶邻居采样数为 3，二阶邻居采样数为 5）。此外，聚合步骤并不是直接聚合所有的一阶邻居和二阶邻居，而是先聚合二阶邻居的特征信息得到一阶邻居的特征表示，再聚合一阶邻居的信息最后得到目标结点的特征表示。

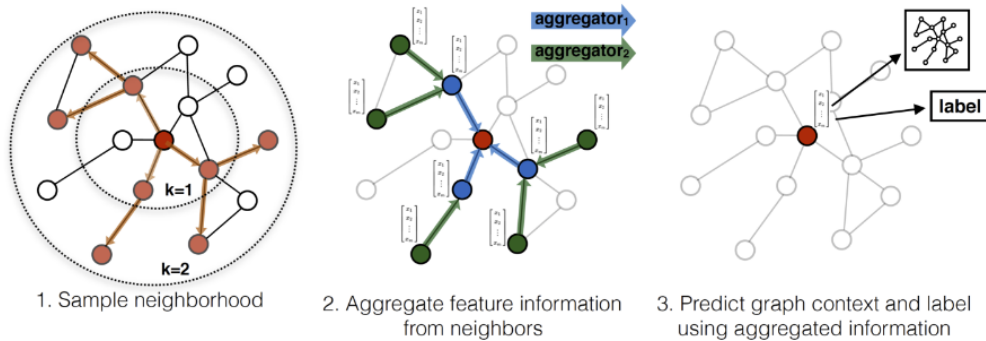


图 7. 图采样聚合网络。

其中，作者们给出了多种聚合函数的形式，分别是基于最大值的聚合，基于均值的聚合和基于长短时记忆网络（LSTM）[13]的聚合。最大值聚合和均值聚合类似卷积中的最大池化（max pooling）和平均池化（average pooling），取相关结点的最大值和均值作为聚合结果（逐个特征）；基于长短时记忆网络的聚合指将相关结点的 embedding 输入 LSTM，把 LSTM 的输出作为聚合的结果，但显然虽然 LSTM 聚合表达能力更强，但是 LSTM 的学习过程会受到输入序列顺序的影响，所以输入 LSTM 进行聚合之前需要将相关结点随机排列打乱顺序，以便让模型适应无序的输入集合。

图采样聚合网络可以做归纳学习（inductive learning），可以对训练过程中见不到的数据直接计算而不需要重新对整个图进行学习，扩大了图神经网络的落地场景。另外，图采样聚合网络的预测速度非常快，因为实现中的随机采样使它只需要选择若干跳范围内的若干个邻居即可，并且原理并不复杂，不需要复杂的理论基础也能理解。

值得一提的是，2018 年斯坦福和 Pinterest 公司合作提出了第一个工业级别（数十亿节点和数百亿边）基

于图卷积神经网络的推荐系统 PinSage [48], 并在离线评估和 AB 实验选中取得了不错的效果, 这也是图卷积神经网络在商业推荐系统首次得到成功应用。

**Graph Convolution Network.** 为了使图卷积神经网络能够适应图上半监督学习领域的任务, Kipf 等人 [18] 对切比雪夫网络进行简化并提出了 GCN (为防止混淆可以称为一阶图卷积神经网络), 他们在文中令  $K = 2$  且  $\lambda_{max} = 2$ , 将式 17 写为

$$\begin{aligned} \mathbf{x}_j^{m+1} &= h \left( \sum_{i=1}^p \sum_{k=0}^{K-1} \beta_k T_k(\tilde{\mathbf{L}}) \mathbf{x}_i^m \right) \\ &= h \left( \sum_{i=1}^p \left( \beta_0 T_0(\tilde{\mathbf{L}}) + \beta_1 T_1(\tilde{\mathbf{L}}) \right) \mathbf{x}_i^m \right) \\ &= h \left( \sum_{i=1}^p (\beta_0 + \beta_1 (\mathbf{L} - \mathbf{I}_n)) \mathbf{x}_i^m \right), \quad j = 1, \dots, q \end{aligned} \quad (23)$$

在图上半监督学习场景下, 带标签的数据很少, 为了避免模型出现过拟合的情况, Kipf 等人增加了约束  $\beta = \beta_0 = -\beta_1$  以减少模型的参数量, 并且对权重矩阵 (带权邻接矩阵) 做归一化处理, 最终得到如下的一阶图卷积神经网络

$$\mathbf{x}_j^{m+1} = h \left( \sum_{i=1}^p \beta \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{x}_i^m \right), \quad j = 1, \dots, q \quad (24)$$

其中,  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$ , 且  $\tilde{\mathbf{D}}_{i,i} = \sum_j \tilde{\mathbf{A}}_{i,j}$ 。令  $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$ , 可以将模型写成这样的简洁形式

$$\mathbf{Z} = \text{softmax} \left( \hat{\mathbf{A}} \text{ReLU} \left( \hat{\mathbf{A}} \mathbf{X} \mathbf{W}^{(0)} \right) \mathbf{W}^{(1)} \right) \quad (25)$$

其中的  $\mathbf{W}^{(0)}$  和  $\mathbf{W}^{(1)}$  都是特征变换矩阵。

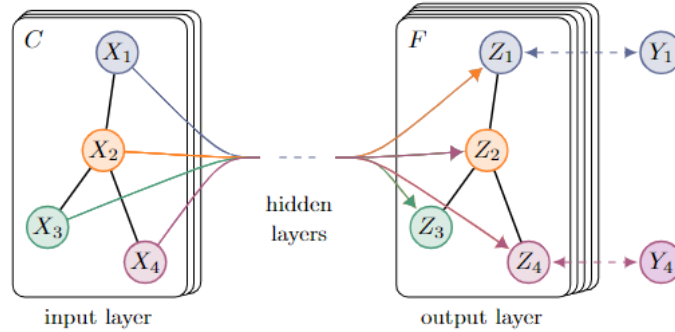


图 8. 一阶图卷积神经网络。

GCN 通过对谱方法做一阶局部近似, 实现了一个多层的图卷积神经网络。其中每一个卷积层仅处理一阶邻域信息, 通过叠加若干卷积层可以实现多阶邻域的信息传递, 但是网络层数如果太深, 会出现过平滑 (over smoothing) 的现象, 即每个顶点的输出特征都十分相似, 反而造成网络效果变差。尽管 GCN 经过推导证明了自己是谱方法的一阶近似实现, 但是实际上每个卷积层都是在做一次对所有结点的一阶邻域的信息聚合, 该方法也架起了谱方法和空间方法之间的一道桥梁。

**Graph Attention Network.** 继谷歌提出 Transformer [41]之后,注意力机制(attention mechanism)引起了各领域的研究人员的广泛关注,研究热潮中大家都在尝试结合注意力机制对自己的课题做出进一步探索。Velikovi等在 2018 年提出了图注意力网络 [42]。图注意力网络通过注意力机制定义聚合函数,在图注意力网络中,邻接矩阵仅被用来定义相关结点,而关联权重的计算则依赖于结点的特征表达。

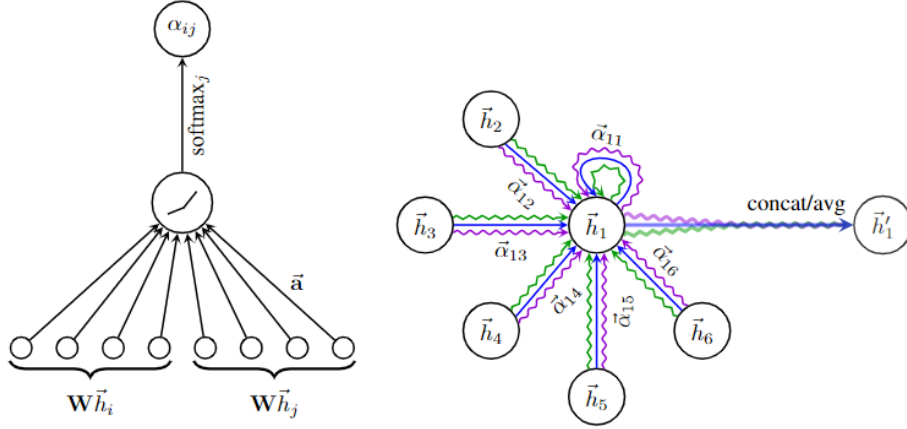


图 9. 图注意力网络。

图注意力网络每层的结构如图9所示,图9左图以结点  $i, j$  之间的注意力分数并归一化的特征表达作为输入,计算  $i, j$  之间的注意力分数并归一化,右图描述的是一个结点利用计算出的与周围结点之间的注意力分数将周围结点的表达以加权求和的形式聚合到自身。注意力打分和结点  $i$  表达的更新公式如下

$$\alpha_{i,j} = \frac{\exp(\text{LeakyReLU}(\mathbf{a} [\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\mathbf{a} [\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_k]))}$$

$$\mathbf{h}_i = \sigma \left( \sum_{j \in \mathcal{N}_i} \alpha_{i,j} \mathbf{W}\mathbf{h}_j \right) \quad (26)$$

其中,  $\mathcal{N}_i$  表示结点  $i$  的邻居所构成的集合,参数  $\mathbf{W}$  矩阵用于完成每个结点的特征维度变换,参数  $\mathbf{a}$  是一个权重参数向量,  $\|$  是拼接操作,  $\mathbf{h}_i$  即为结点  $i$  的特征表示,  $\alpha_{i,j}$  表示在参数  $\mathbf{a}$  下计算得到的结点  $i, j$  间的注意力分数,防止与结点的特征表示混淆,我们在这里用  $\sigma$  表示非线性激活函数。

此外,作者也效仿 Transformer [41]中的多头注意力 (multi-head attention) 机制,图9中右图里不同颜色就代表不同的注意力头的计算过程 (3 个注意力头),之后将结果进行拼接或平均,关于拼接和平均的具体公式如下

$$\text{concat} : \mathbf{h}'_i = \left\| \sigma \left( \sum_{j \in \mathcal{N}_i} \alpha_{i,j}^{(l)} \mathbf{W}^{(l)} \mathbf{h}_j \right) \right\|_{l=1}^L \quad (27)$$

$$\text{average} : \mathbf{h}'_i = \sigma \left( \frac{1}{L} \sum_{l=1}^L \sum_{j \in \mathcal{N}_i} \alpha_{i,j}^{(l)} \mathbf{W}^{(l)} \mathbf{h}_j \right) \quad (28)$$

其中,  $L$  为所有注意力头的头数,  $\alpha_{i,j}^{(l)}$  和  $\mathbf{W}^{(l)}$  为标号为  $l$  的注意力头中对应的注意力分数和变换矩阵。



## 4. 应用

图卷积神经网络自提出以来，受到了研究人员的大量关注，主要集中于以下几个领域：网络分析、推荐系统、生物化学、交通预测、计算机视觉、自然语言处理等。图卷积神经网络的应用领域既包括计算机科学、人工智能、信号处理等传统机器学习领域，也包括物理、生物、化学、社会科学等跨学科的研究领域。

徐冰冰等人在去年发表的综述中总结 [50] 了常见分类，见下表3。

应用领域	结点	连边	应用问题	任务	相关论文
网络分析	用户 论文	社交关系 引用关系	用户影响力预测 半监督结点分类	图回归 节点分类	DeepInf [35] GCN [18], GAT [42], GWNN [46]
推荐系统	用户, 商品	购买某商品的可能 性分数	用户偏好推荐	矩阵补全/ 链接预测	PinSage [48], RippleNet [43, 44], GraphRec [6]
生物化学	小分子  蛋白质/ 药物	化学键  相互作用	化学功能预测  副作用预测	图分类  链接预测	Duvenaud 等人 [5], Kearnes 等人 [16], GAM [21] Decagon
计算机 视觉	3D 点云图  对象	距离信息  对象之间的 关联	语义分割/ 形状分类  场景图生成/ 视觉推理	语义分割  图生成	3DGNN [34], SPG [20], RGCNN [39] Graph VQA [40], GRM [45]
自然语言 处理	概念实体  词  词 词	语义关系  依赖关系  依赖关系 共现关系	知识推理  语义角色标注/ 抽象含义表达  事件提取 文本分类	链接预测  关系提取  序列标注 图分类	GNN-for-OKB [8]  Semantic GCN [24], LSTM+GCN [25], C-GCN [49] JMEE [23], Nguyen 等人 [29] HR-DGCNN [32], TextGCN [47]

表 3. 图卷积神经网络应用领域总结表（引自徐冰冰等综述 [50]）。

除了上述应用领域之外，在优化求解和程序推断 [17, 22] 等任务上，图卷积神经网络都开始被人们使用。由于其可以建模在现实生活中常见的图数据上，并且通过卷积、注意力或消息传播等机制，能够将网络的拓扑结构和结点属性等信息以神经网络进行捕获和建模，因此图卷积神经网络有广泛的应用前景。

## 5. 小结

图卷积神经网络是能够较高效地处理图数据的模型，在过去几年受到广泛关注，本文对图卷积神经网络的两种主流方法做了梳理和简单介绍，总结了一些较为经典的模型和方法。

图卷积神经网络方面的工作主要面临着三个挑战：图数据是非欧空间数据，图数据的多样性，图数据规模庞大。

在第2节和第3节中，我们总结了目前的两类主流方法：谱方法利用图傅里叶变换和卷积定理在谱域定义图卷积，空间方法在结点域定义加权函数再聚合中心结点及其邻居结点的特征。此外，我们还总结了一些经典的图卷积神经网络模型，并且分析了其优缺点，简要介绍了图神经网络适用的应用问题。总的来说，图卷积神经网络已经取得了一定效果，但仍有一些问题需要解决，以使得它在图数据的场景下被充分利用。

## 参考文献

- [1] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun. Spectral networks and deep locally connected networks on graphs. In *2nd International Conference on Learning Representations, ICLR 2014*, 2014. 5
- [2] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu. Global contrast based salient region detection. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, pages 409–416, 2011. 1
- [3] F. R. Chung and F. C. Graham. *Spectral graph theory*. Number 92. 1997. 4
- [4] M. Defferrard, X. Bresson, and P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 3837–3845, 2016. 6
- [5] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 2224–2232, 2015. 14
- [6] W. Fan, Y. Ma, Q. Li, Y. He, Y. E. Zhao, J. Tang, and D. Yin. Graph neural networks for social recommendation. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pages 417–426, 2019. 14
- [7] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 855–864, 2016. 2
- [8] T. Hamaguchi, H. Oiwa, M. Shimbo, and Y. Matsumoto. Knowledge transfer for out-of-knowledge-base entities : A graph neural network approach. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 1802–1808, 2017. 14
- [9] W. L. Hamilton, Z. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 1024–1034, 2017. 11
- [10] D. K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011. 8
- [11] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778, 2016. 1

- [12] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6):82–97, 2012. 1
- [13] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. 11
- [14] B. Hu, Z. Lu, H. Li, and Q. Chen. Convolutional neural network architectures for matching natural language sentences. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2042–2050, 2014. 1
- [15] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 448–456, 2015. 1
- [16] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 30(8):595–608, 2016. 14
- [17] E. B. Khalil, H. Dai, Y. Zhang, B. Dilkina, and L. Song. Learning combinatorial optimization algorithms over graphs. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 6348–6358, 2017. 14
- [18] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017. 12, 14
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pages 1106–1114, 2012. 1
- [20] L. Landrieu and M. Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 4558–4567, 2018. 14
- [21] J. B. Lee, R. A. Rossi, and X. Kong. Graph classification using structural attention. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 1666–1674, 2018. 14
- [22] Z. Li, Q. Chen, and V. Koltun. Combinatorial optimization with graph convolutional networks and guided tree search. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 537–546, 2018. 14
- [23] X. Liu, Z. Luo, and H. Huang. Jointly multiple events extraction via attention-based graph information aggregation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1247–1256, 2018. 14
- [24] D. Marcheggiani, J. Bastings, and I. Titov. Exploiting semantics in neural machine translation with graph convolutional networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 486–492, 2018. 14

- [25] D. Marcheggiani and I. Titov. Encoding sentences with graph convolutional networks for semantic role labeling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1506–1515, 2017. 14
- [26] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013. 2
- [27] M. Mirbauer, M. Krabec, J. Krivanek, and E. Sikudova. Survey and evaluation of neural 3d shape classification approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1
- [28] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3-8, 2001, Vancouver, British Columbia, Canada]*, pages 849–856, 2001. 2
- [29] T. H. Nguyen and R. Grishman. Graph convolutional networks with argument-aware pooling for event detection. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 5900–5907, 2018. 14
- [30] M. Niepert, M. Ahmed, and K. Kutzkov. Learning convolutional neural networks for graphs. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 2014–2023, 2016. 10
- [31] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999. 2
- [32] H. Peng, J. Li, Y. He, Y. Liu, M. Bao, L. Wang, Y. Song, and Q. Yang. Large-scale hierarchical text classification with recursively regularized deep graph-cnn. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018*, pages 1063–1072, 2018. 14
- [33] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: online learning of social representations. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 701–710, 2014. 2
- [34] X. Qi, R. Liao, J. Jia, S. Fidler, and R. Urtasun. 3d graph neural networks for RGBD semantic segmentation. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 5209–5218, 2017. 14
- [35] J. Qiu, J. Tang, H. Ma, Y. Dong, K. Wang, and J. Tang. Deepinf: Social influence prediction with deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 2110–2119, 2018. 14
- [36] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013. 1, 4

- [37] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. 1
- [38] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. LINE: large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web, WWW 2015, Florence, Italy, May 18-22, 2015*, pages 1067–1077, 2015. 2
- [39] G. Te, W. Hu, A. Zheng, and Z. Guo. RGCNN: regularized graph CNN for point cloud segmentation. In *2018 ACM Multimedia Conference on Multimedia Conference, MM 2018, Seoul, Republic of Korea, October 22-26, 2018*, pages 746–754, 2018. 14
- [40] D. Teney, L. Liu, and A. van den Hengel. Graph-structured representations for visual question answering. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 3233–3241, 2017. 14
- [41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008, 2017. 13
- [42] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018. 13, 14
- [43] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, and M. Guo. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, pages 417–426, 2018. 14
- [44] H. Wang, F. Zhang, J. Wang, M. Zhao, W. Li, X. Xie, and M. Guo. Exploring high-order user preference on the knowledge graph for recommender systems. *ACM Transactions on Information Systems (TOIS)*, 37(3):1–26, 2019. 14
- [45] Z. Wang, T. Chen, J. S. J. Ren, W. Yu, H. Cheng, and L. Lin. Deep reasoning with knowledge graph for social relationship understanding. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 1021–1028, 2018. 14
- [46] B. Xu, H. Shen, Q. Cao, Y. Qiu, and X. Cheng. Graph wavelet neural network. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019. 8, 9, 14
- [47] L. Yao, C. Mao, and Y. Luo. Graph convolutional networks for text classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 7370–7377, 2019. 14
- [48] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 974–983, 2018. 12, 14



- [49] Y. Zhang, P. Qi, and C. D. Manning. Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2205–2215, 2018. 14
- [50] 徐冰冰, 岑科廷, 黄俊杰, 沈华伟, 程学旗. 图卷积神经网络综述. *计算机学报*, 43(5):755–780, 2020. 14