

图卷积网络

刘家锋

哈尔滨工业大学

图卷积网络

- 1 图与卷积
- 2 拉普拉斯算子和傅里叶变换
- 3 图卷积网络

图与卷积

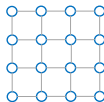
Euclidean Structure数据

● Euclidean Structure数据

- 机器学习处理的一大类数据具有潜在的欧氏结构，例如图像、视频、语音等
- 欧氏结构的特点是数据“排列整齐”，数据之间形成一种规整的网格邻接关系
- 卷积神经网络对这类数据的处理取得了极大的成功

● 图像的网格表示

- 像素-节点，颜色或灰度-节点的属性



Non Euclidean Structure数据

- **Non Euclidean Structure**数据

- 机器学习处理的另外一类数据不具有潜在的欧氏结构，数据的排列不整齐
- 图结构数据是一种常见的非欧结构数据

● 社交网络数据

- 人-节点，相互关注-连接



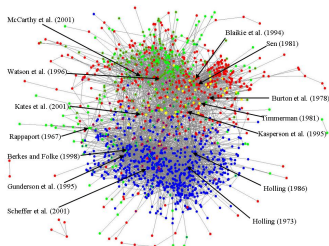
Non Euclidean Structure数据

- 科学文献数据

- 文献-节点，文献的文本内容-节点属性
- 文献之间的引用关系-节点之间的连接

- 学习的任务

- 半监督学习：已知少量文献的类别，推理其它文献的类别



Non Euclidean Structure数据

- 其它的非欧结构数据

- 化学成分结构数据
- 生物基因蛋白数据
- 知识图谱数据

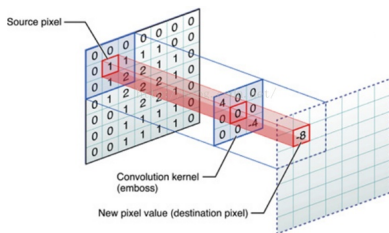
- 非欧结构数据的特点

- 数据是无序的，但数据之间存在着关联，数据之间的关系可以用图的方式来描述
- 每个节点的信息不仅与节点属性有关，也与相邻节点有关，甚至与节点局部的网络结构有关
- 图的结构是不固定的，节点之间的连接数量也是不确定的，可多可少

卷积

● 数据局部特征的提取

- 卷积的本质就是利用共享权值的滤波器，通过计算中心像素（节点）以及相邻像素（节点）的加权和来实现空间（局部）特征的提取
- 欧氏结构数据可以很容易地利用CNN来实现局部特征的提取与学习
- 在图结构的数据上是否能够实现同样的卷积运算？



节点域实现卷积

● 图卷积的困难

- 与欧氏结构数据不同，图结构数据上每个节点的邻域节点数不同
- 不可能定义一个统一的、共享权值的卷积核，适用于图上的所有节点
- 卷积核也不可能具有平移不变性

● 节点域卷积

- 在每一个节点上定义不同的卷积核
- 需要学习的网络参数量大，无法保证泛化能力
- 网络的通用性差，图的结构发生变化，需要重新学习网络

频率域实现卷积

- 数学上的卷积

- 信号 $g(t)$ 与滤波器 $h(t)$ 的卷积定义为：

$$g(t) * h(t) = \int_{-\infty}^{+\infty} g(\tau)h(t - \tau)d\tau$$

- 离散卷积：

$$g(n) * h(n) = \sum_{i=-\infty}^{+\infty} g(i)h(n - i)$$

- 欧氏结构数据上的卷积，实质上就是有界滤波器 $h(n)$ 的离散卷积

频率域实现卷积

- 卷积定理

- 令信号和滤波器的Fourier变换分别为：

$$G(\omega) = \mathcal{F}(g(t)), \quad H(\omega) = \mathcal{F}(h(t))$$

- 则有：

$$\mathcal{F}(g(t) * h(t)) = G(\omega)H(\omega)$$

$$g(t) * h(t) = \mathcal{F}^{-1}(G(\omega)H(\omega))$$

- 是否可以在频率域定义图上的卷积？

拉普拉斯算子和傅里叶变换

Laplacian Operator

- 平方可积函数空间

- 令 $\Omega \subset R^d$, $L^2(\Omega)$ 是定义在 Ω 上的平方可积函数空间, 即 $\forall f(\mathbf{x}) \in L^2(\Omega)$, 成立:

$$\int_{\Omega} f^2(\mathbf{x}) d\mathbf{x} < \infty$$

- $L^2(\Omega)$ 为 Hilbert 空间 (内积空间), 内积定义为:

$$\langle u, v \rangle = \int_{\Omega} u(\mathbf{x})v(\mathbf{x})d\mathbf{x}, \quad u(\mathbf{x}), v(\mathbf{x}) \in L^2(\Omega)$$

- 相应的范数:

$$\|u(\mathbf{x})\|_2^2 = \int_{\Omega} u^2(\mathbf{x})d\mathbf{x}$$

Laplacian Operator

- 算子

- 算子是定义在Hilbert空间上的一个映射
- 梯度算子：函数到矢量场的映射

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \partial f(\mathbf{x})/\partial x_1 \\ \vdots \\ \partial f(\mathbf{x})/\partial x_d \end{pmatrix}$$

- Laplacian算子：** $C^2(\Omega) \rightarrow C^2(\Omega)$

$$\Delta f(\mathbf{x}) = \sum_{i=1}^d \frac{\partial^2 f(\mathbf{x})}{\partial x_i^2} \in C^2(\Omega), \quad f(\mathbf{x}) \in C^2(\Omega)$$

其中， $C^2(\Omega)$ 为二阶连续可微函数空间

Laplacian Operator

- 图像的Laplacian算子

- 算子:

$$\Delta = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

- 算子作用于图像:

$$\begin{aligned} \Delta I(m, n) = & 4I(m, n) - I(m-1, n) - I(m, n-1) \\ & - I(m+1, n) - I(m, n+1) \end{aligned}$$

Laplacian算子的谱

- 算子的谱

- 如果存在 (λ, u) , $\lambda \in R, u(\mathbf{x}) \in C^2(\Omega)$, 成立:

$$\Delta u(\mathbf{x}) = \lambda u(\mathbf{x}), \quad \|u(\mathbf{x})\|_2 = 1$$

则称 λ 为算子 Δ 的特征值, $u(\mathbf{x})$ 为相应的特征函数

- Laplacian算子的谱

- 1维情形: $e^{-i\omega x}$ 为特征函数, ω^2 为特征值

$$\begin{aligned}\Delta e^{-i\omega x} &= \frac{d^2 e^{-i\omega x}}{dx^2} \\ &= \frac{d}{dx} \left(\frac{de^{-i\omega x}}{dx} \right) \\ &= \frac{d}{dx} (-i\omega e^{-i\omega x}) = \omega^2 e^{-i\omega x}\end{aligned}$$

Laplacian算子的谱

- Laplacian算子的谱

- 2维情形: $e^{-i(\omega_x x + \omega_y y)}$ 为特征函数, $\omega_x^2 + \omega_y^2$ 为特征值

$$\begin{aligned}\frac{\partial e^{-i(\omega_x x + \omega_y y)}}{\partial x} &= -i\omega_x e^{-i(\omega_x x + \omega_y y)}, & \frac{\partial^2 e^{-i(\omega_x x + \omega_y y)}}{\partial x^2} &= \omega_x^2 e^{-i(\omega_x x + \omega_y y)} \\ \frac{\partial e^{-i(\omega_x x + \omega_y y)}}{\partial y} &= -i\omega_y e^{-i(\omega_x x + \omega_y y)}, & \frac{\partial^2 e^{-i(\omega_x x + \omega_y y)}}{\partial y^2} &= \omega_y^2 e^{-i(\omega_x x + \omega_y y)}\end{aligned}$$

$$\begin{aligned}\Delta e^{-i(\omega_x x + \omega_y y)} &= \frac{\partial^2 e^{-i(\omega_x x + \omega_y y)}}{\partial x^2} + \frac{\partial^2 e^{-i(\omega_x x + \omega_y y)}}{\partial y^2} \\ &= (\omega_x^2 + \omega_y^2) e^{-i(\omega_x x + \omega_y y)}\end{aligned}$$

Laplacian算子的特征函数与Fourier基

- Laplacian算子的谱

- 多维情形: $e^{-i\omega^t \mathbf{x}}$ 为特征函数, $\|\omega\|^2$ 为特征值

- Fourier变换

- 1维Fourier变换:

$$G(\omega) = \mathcal{F}(g(t)) = \int_{-\infty}^{+\infty} g(t) e^{-i\omega t} dt$$

- 多维Fourier变换:

$$G(\omega) = \mathcal{F}(g(\mathbf{x})) = \int_{-\infty}^{+\infty} g(\mathbf{x}) e^{-i\omega^t \mathbf{x}} d\mathbf{x}$$

- Laplacian算子的特征函数即为Fourier变换的基: $\{e^{-i\omega^t \mathbf{x}}\}$

图上的Laplacian

● 图的Laplacian矩阵

- 无向图 $G = \{E, V\}$ 上的Laplacian矩阵 $L = D - W$ 是一个映射 $L : R^N \rightarrow R^N$ ， N 为图上的节点数
- 对于任意矢量 $\mathbf{f} \in R^N$ ，映射之后的第 i 个元素：

$$\begin{aligned}(L\mathbf{f})(i) &= [(D - W)\mathbf{f}](i) = \sum_j w_{ij}\mathbf{f}(i) - \sum_j w_{ij}\mathbf{f}(j) \\ &= \sum_j w_{ij}[\mathbf{f}(i) - \mathbf{f}(j)]\end{aligned}$$

图上的Laplacian

- 与图像的Laplacian的一致性

- 将图像 I 表示成图的形式，每个像素对应一个节点，每个节点只与4邻域像素有连接，连接权重为1
- 定义矢量 $\mathbf{f}(m, n) = I(m, n)$ ，像素排列成矢量的形式：

$$\begin{aligned}(L\mathbf{f})(m, n) &= 4I(m, n) - I(m+1, n) - I(m-1, n) \\ &\quad - I(m, n+1) - I(m, n-1) \\ &= \Delta I(m, n)\end{aligned}$$

图上的Fourier变换

● 定义图的Fourier变换

- 令图 G 的Laplacian矩阵 L 的特征值和特征矢量：

$$L\mathbf{u}_l = \lambda_l \mathbf{u}_l, \quad 0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_N$$

- \mathbf{f} 是将所有节点上的属性值排成的矢量（以1维属性为例，多维属性类似）
- 图 G 上的Fourier变换定义为：

$$F(l) = \langle \mathbf{f}, \mathbf{u}_l \rangle = \sum_{n=1}^N f(n) u_l(n)$$

- 图 G 上的Fourier反变换定义为：（ L 为对称矩阵，因此无需考虑复共轭）

$$f(n) = \sum_{l=1}^N F(l) u_l(n)$$

图上的Fourier变换

- 图**Fourier**变换的对应关系
 - 图的Laplacian矩阵 $L \rightarrow$ Laplacian算子 Δ
 - L 的特征矢量 $\{\mathbf{u}_l\} \rightarrow$ Fourier基 $\{e^{-i\omega t}\}$
 - L 的特征值 $\{\lambda_l\} \rightarrow$ 频率 $\{\omega\}$

图卷积网络

频域上实现卷积

● 频率域上的滤波器

- 给定图 G ，Laplacian矩阵 L 描述了图的拓扑结构，对应的特征值分解：

$$L = U\Lambda U^t$$

其中， $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$ ， $U = (\mathbf{u}_1, \dots, \mathbf{u}_N)$

- 令 $\mathbf{x} = (x_1, \dots, x_N)^t$ 是定义在图 G 上的信号（每一维元素是每个节点上的属性，多维属性依此类推）
- \mathbf{x} 的Fourier变换为：

$$\hat{\mathbf{x}} = U^t \mathbf{x}$$

频域上实现卷积

- 频率域上的滤波器

- 定义频域上的滤波器：

$$g(\Lambda) = \text{diag}(g(\lambda_1), \dots, g(\lambda_N))$$

- 频率域上信号 \mathbf{x} 的滤波输出：

$$\hat{\mathbf{z}} = g(\Lambda)\hat{\mathbf{x}} = g(\Lambda)U^t\mathbf{x}$$

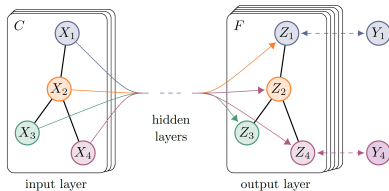
- 节点域上信号 \mathbf{x} 的滤波输出：

$$\mathbf{z} = U\hat{\mathbf{z}} = Ug(\Lambda)U^t\mathbf{x}$$

频域上实现卷积

● 频率域滤波器的特点

- 滤波器 $g(\Lambda)$ 作用在图 G 的节点上，将节点上的信号（属性值） \mathbf{x} 滤波为 \mathbf{z} ，但没有改变图 G 的拓扑结构
- 每个节点上使用的是同样的卷积核，参数是共享的



频域卷积的实现

- **Non-parametric Filter**

- 频域卷积核最直观的定义：

$$g(\Lambda) = \text{diag}(\theta_1, \dots, \theta_N)$$

- 缺点：

- 需要学习的参数比较多，等于图的节点数 N
- 滤波器不能保证具有局部性，卷积核的尺度无法控制，有可能是整个图上的卷积
- 计算复杂度高， $O(N^2)$

频域卷积的实现

● Polynomial Filter

- 定义一种特殊的频域卷积核：

$$g(\Lambda) = \sum_{k=0}^{K-1} \theta_k \Lambda^k = \text{diag} \left(\sum_{k=0}^{K-1} \theta_k \lambda_1^k, \dots, \sum_{k=0}^{K-1} \theta_k \lambda_N^k \right)$$

- 滤波器输出：

$$\mathbf{z} = U g(\Lambda) U^t \mathbf{x} = U \left(\sum_{k=0}^{K-1} \theta_k \Lambda^k \right) U^t \mathbf{x} = \left(\sum_{k=0}^{K-1} \theta_k U \Lambda^k U^t \right) \mathbf{x}$$

- 节点域的卷积核可以写成矩阵形式：

$$H = \sum_{k=0}^{K-1} \theta_k U \Lambda^k U^t = \sum_{k=0}^{K-1} \theta_k (U \Lambda U^t) \cdots (U \Lambda U^t) = \sum_{k=0}^{K-1} \theta_k L^k$$

频域卷积的实现

● Polynomial Filter的优点

- 每个频率上都是特征值 λ 的一个多项式，多项式的系数是共享的
- 需要学习的参数少，共 K 个
- 可以证明，卷积核在节点域上是局部的，每个节点的卷积只在图上与该节点最短路径 $\leq K$ 的节点集合上计算
- 滤波器可以采用Chebyshev Polynomials定义，计算复杂度 $O(K|E|)$ ，当图的连接比较稀疏时，可以简化计算

图卷积网络的半监督学习应用

- 半监督学习任务
 - Citation Network Datasets科技文献类别标注
 - 已有部分文献的Label，推断其它未知文献的类别
- 图的构造
 - 节点：Document
 - 节点属性：文本的内容，关键词或Bag of Words
 - 边：Citation Links，文献之间的引用关系

图卷积网络的半监督学习应用

● 图卷积网络

- 构造多个图卷积层，实现图上每个节点特征的提取
- 最后一个卷积层之后，构造一个线性判别层，**softmax**输出，判别每个节点的类别（每个节点上的线性判别权值共享）
- 以有Label节点的判别误差学习网络参数，包括卷积核参数和线性判别参数
- 网络收敛之后，前向预测无Label节点的类别

