

| 2025W GCN 理论推导

| GCN 理论推导

在 B 站发现了一个宝藏 up 主，讲的 GCN 理论推导非常清晰，大家可以去学习一波！

[图卷积神经网络（GCN）的数学原理详解——谱图理论和傅立叶变换初探](#)

我把视频内容整理成文档了，大家可以看下。

| 1-warmup

$$\text{GNN} = H^{(l+1)} = f(A, H^{(l)}) \quad (1-1)$$

$$\text{GCN} : H^{(l+1)} = \sigma(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} \Theta) \quad (1-2)$$

这个图是一个 undirected simple graph。

A 是 adjacency matrix, D 是 degree matrix,

$$\hat{A} = A + I \quad (1-3)$$

$$\hat{D} = D + I \quad (1-4)$$

设 D 为

$$\begin{bmatrix} 3 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \quad (1-5)$$

添加自环后的 \hat{D} 为

$$\begin{bmatrix} 4 & & & \\ & 2 & & \\ & & 2 & \\ & & & 2 \end{bmatrix} \quad (1-6)$$

对 \hat{D} 进行幂操作, $\hat{D}^{-\frac{1}{2}}$ 为

$$\begin{bmatrix} 4^{-\frac{1}{2}} & & & \\ & 2^{-\frac{1}{2}} & & \\ & & 2^{-\frac{1}{2}} & \\ & & & 2^{-\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & & & \\ & \frac{1}{\sqrt{2}} & & \\ & & \frac{1}{\sqrt{2}} & \\ & & & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (1-7)$$

| 2-spectral graph theory

线性代数回顾

特征值与特征向量

$$A\vec{x} = \lambda\vec{x}, |\vec{x}| \neq 0 \quad (2-1)$$

λ 是 A 的一个特征值, \vec{x} 是 A 的一个特征向量。

实对称矩阵

如果一个矩阵是实对称阵, 那么它一定有 n 个特征值, 这 n 个特征值有 n 个互相正交的特征向量。

$$A = U\Lambda U^T, UU^T = I \quad (2-2)$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \quad (2-3)$$

半正定矩阵

如果一个矩阵的所有的特征值都大于等于 0, 那么这个矩阵是半实对称阵。

二次型

$$\vec{x}^T A \vec{x}, \forall i, \lambda_i \geq 0 \quad (2-4)$$

瑞利熵

$$\frac{\vec{x}^T A \vec{x}}{\vec{x}^T \vec{x}} \quad (2-5)$$

如果 \vec{x} 是一个矩阵的特征向量, 那么瑞利熵为该矩阵的特征值。

$$\frac{\vec{x}^T A \vec{x}}{\vec{x}^T \vec{x}} = \frac{\vec{x}^T (\lambda \vec{x})}{\vec{x}^T \vec{x}} = \frac{\lambda (\vec{x}^T \vec{x})}{\vec{x}^T \vec{x}} = \lambda \quad (2-6)$$

因此瑞利熵是我们研究特征值的重要手段,

谱图理论

图的拉普拉斯矩阵

$$L = D - A \quad (2-7)$$

图的拉普拉斯矩阵的对称化

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \quad (2-8)$$

L 和 L_{sym} 是半正定矩阵。

可从瑞利熵 $\frac{\vec{x}^T A \vec{x}}{\vec{x}^T \vec{x}} \geq 0$ 证得。 $\vec{x}^T \vec{x}$ 肯定是大于等于 0，只需证明分子大于等于 0。

****证明 L 是半正定矩阵**

先定义一个矩阵，第 i 行的第 i 列和第 j 行的第 j 列是 1，第 i 行的第 j 列和第 j 行的第 i 列是 -1，其余为 0。

$$G_{(i,j)} = \begin{bmatrix} \ddots & & & & \\ & 1 & & -1 & \\ & & \ddots & & \\ & -1 & & 1 & \\ & & & & \ddots \end{bmatrix} \quad (2-9)$$

L 可以改写为

$$L = D - A = \sum_{(i,j) \in E} G_{(i,j)} \quad (2-10)$$

$G_{(i,j)}$ 的二次型为

$$\vec{x} G_{(i,j)} \vec{x} = \vec{x} \begin{bmatrix} \vdots \\ x_i - x_j \\ \vdots \\ x_j - x_i \end{bmatrix} = x_i(x_i - x_j) + x_j(x_j - x_i) = (x_i - x_j)^2 \quad (2-11)$$

L 的二次型为

$$\vec{x}^T L \vec{x} = \vec{x}^T \left(\sum_{(i,j) \in E} G_{(i,j)} \right) \vec{x} = \sum_{(i,j) \in E} \vec{x}^T G_{(i,j)} \vec{x} = \sum_{(i,j) \in E} (x_i - x_j)^2 \geq 0. \quad (2-12)$$

所以 L 是一个半正定的矩阵。

****证明 L_{sym} 是半正定矩阵**

$$\vec{x} L_{sym} \vec{x} = \left(\vec{x}^T D^{-\frac{1}{2}} \right) L \left(D^{-\frac{1}{2}} \vec{x} \right) = \sum_{(i,j) \in E} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2 \geq 0. \quad (2-13)$$

所以 L_{sym} 是一个半正定的矩阵。

L_{sym} 有一个更好的性质，它的特征值范围为 [0,2]。

$$G_{(i,j)}^{pos} = \begin{bmatrix} \ddots & & & & \\ & 1 & & 1 & \\ & & \ddots & & \\ & 1 & & 1 & \\ & & & & \ddots \end{bmatrix} \quad (2-14)$$

$$\vec{x}^T G_{(i,j)}^{pos} \vec{x} = (x_i + x_j)^2 \quad (2-15)$$

$$L^{pos} = D + A = \sum_{(i,j) \in E} G_{(i,j)}^{pos} \quad (2-16)$$

$$\vec{x}^T L^{pos} \vec{x} = \sum_{(i,j) \in E} (x_i + x_j)^2 \geq 0. \quad (2-17)$$

代入 $L^{pos} = D + A$

$$L_{sym}^{pos} = D^{-\frac{1}{2}} L^{pos} D^{-\frac{1}{2}} = I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \quad (2-18)$$

结果代入下式

$$\vec{x}^T L_{sym}^{pos} \vec{x} = \sum_{(i,j) \in E} \left(\frac{x_i}{\sqrt{d_i}} + \frac{x_j}{\sqrt{d_j}} \right)^2 \geq 0. \quad (2-19)$$

$$\begin{aligned} \vec{x}^T \left(I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) \vec{x} &\geq 0 \\ \vec{x}^T \vec{x} + \vec{x}^T D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \vec{x} &\geq 0 \\ \vec{x}^T \vec{x} &\geq -\vec{x}^T D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \vec{x} \\ 2\vec{x}^T \vec{x} &\geq \vec{x}^T \vec{x} - \vec{x}^T D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \vec{x} \\ 2\vec{x}^T \vec{x} &\geq \vec{x}^T \vec{x} \left(I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) \vec{x} \\ 2\vec{x}^T \vec{x} &\geq \vec{x}^T D^{-\frac{1}{2}} (D - A) D^{-\frac{1}{2}} \vec{x} \\ 2\vec{x}^T \vec{x} &\geq \vec{x}^T L_{sym} \vec{x} \\ 2 &\geq \frac{\vec{x}^T L_{sym} \vec{x}}{\vec{x}^T \vec{x}} \end{aligned} \quad (2-20)$$

L_{sym} 的瑞利熵小于等于 2，即特征值小于等于 2，前面已经证明大于等于 0，所以 L_{sym} 的特征值范围为 [0,2]。

3-fourier transformation

什么是傅里叶变化，从不同的域研究数据，并在域间进行转换。

[傅里叶分析之掐死教程（完整版）更新于2014.06.06](#)

例子 1

$f(x)$ 是男女声混合音频，通过傅里叶变换可将男女声分离到不同频率。

例子 2

FFT

$$\begin{aligned} f(x) &= a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \\ g(x) &= b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n \\ f(x)g(x) &= c_0 + c_1 x + c_2 x^2 + \cdots + c_{2n} x^{2n} \end{aligned} \quad (3-1)$$

转换为自变量对应的值的点对域

$$\begin{aligned}
f(x) &\iff (x_1, f(x_1))(x_2, f(x_2)) \cdots (x_{n+1}, f(x_{n+1})) \\
g(x) &\iff (x_1, g(x_1))(x_2, g(x_2)) \cdots (x_{n+1}, g(x_{n+1})) \\
f(x)g(x) &\iff (x_1, f(x_1)g(x_1))(x_2, f(x_2)g(x_2)) \cdots (x_{n+1}, f(x_{n+1})g(x_{n+1}))
\end{aligned} \tag{3-2}$$

$O(n \log n)$ 的时间复杂度。

图在空间域拓扑复杂，所以要把图转换到频域，最后转换到空间域。

$$Lx = \begin{bmatrix} \sum_{(1,j) \in E} (x_1 - x_j) \\ \sum_{(2,j) \in E} (x_2 - x_j) \\ \vdots \\ \sum_{(n,j) \in E} (x_n - x_j) \end{bmatrix} = U\Lambda U^T x \tag{3-3}$$

一个向量乘以正交矩阵代表一种空间系基底的变换， $U^T x$ 是变换基底操作， $\Lambda U^T x$ 是对每一个维度进行放缩， $U\Lambda U^T x$ 是用逆变换变换回原来空间。

对拉普拉斯矩阵进行分解的时间复杂度为 $O(n^2)$ ，大规模图不适用。GCN 提出了不需要对拉普拉斯矩阵进行分解的一种复杂度与边呈线性关系的方法。

4-gcn

定义图卷积操作

$$F(A) \rightarrow L/L_{sym} \tag{4-1}$$

$$F(A) = U\Lambda U^T \tag{4-2}$$

$$g_\theta * x = U g_\theta(\Lambda) U^T x \tag{4-3}$$

对 g_θ 作限制，

$$g_\theta(\Lambda) = \theta_0 \Lambda^0 + \theta_1 \Lambda + \cdots + \theta_n \Lambda^n + \cdots \tag{4-4}$$

那么

$$U g_\theta(\Lambda) U^T = g_\theta(U\Lambda U^T) = g_\theta(F(A)) \tag{4-5}$$

不需要对 $F(A)$ 进行特征分解了。证明上式成立

$$(U\Lambda U^T)^k = U\Lambda U^T U\Lambda U^T \cdots U\Lambda U^T = U\Lambda^k U^T \tag{4-6}$$

其中 $U^T U$ 为单位矩阵。

实际操作中，并不是用公式 (4-2) 系数的形式去拟合多项式，随着 n 的变大会会有梯度消失或梯度爆炸的问题。

使用 Cheb Poly

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), T_0(x) = 1, T_1(x) = x \tag{4-7}$$

为什么它不会梯度消失或爆炸呢？它具有下面的性质，

$$T_n(\cos \theta) = \cos n\theta \tag{4-8}$$

缺点是对自变量有限制，在 $[-1, 1]$ 之间，也就是要求矩阵的特征值在 $[-1, 1]$ 之间。

L_{sym} 的特征值范围在 $[0, 2]$ 之间，减去单位矩阵就满足特征值在 $[-1, 1]$ 之间。

公式 (4-2) 可改写为

$$F(A) = U\Lambda U^T = L_{sym} - I \quad (4-9)$$

公式 (4-3) 可改写为

$$\begin{aligned} g_\theta * x &= U g_\theta(\Lambda) U^T x \\ &= U \left(\sum_{k=0}^k T_k(\Lambda) \right) U^T x \\ &= \sum_{k=0}^k \theta_k U T_k(\Lambda) U^T x \\ &= \sum_{k=0}^k \theta_k T_k(U\Lambda U^T) x \\ &= \sum_{k=0}^k \theta_k T_k(L_{sym} - I) x \end{aligned} \quad (4-10)$$

复杂度依然很高，GCN 的做法是作一阶近似

$$\begin{aligned} \sum_{k=0}^k \theta_k T_k(L_{sym} - I) x &\approx \theta_0 T_0(L_{sym} - I) x + \theta_1 T_1(L_{sym} - I) x \\ &= \theta x + \theta_1 (L_{sym} - I) x \end{aligned} \quad (4-11)$$

L_{sym} 可表示为

$$\begin{aligned} L_{sym} &= D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \\ &= D^{-\frac{1}{2}} (D - A) D^{-\frac{1}{2}} \\ &= I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \end{aligned} \quad (4-12)$$

故

$$\theta x + \theta_1 (L_{sym} - I) x = \theta_0 - \theta_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x \quad (4-13)$$

GCN 在推导过程中采用了一些正则化和 trick,

我们让参数进行共享，令 $\theta_1 = -\theta_0$

$$\theta_0 - \theta_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x \implies \theta_0 (I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) x \quad (4-14)$$

为了进一步简化上式，GCN 作者采用了一种叫做 renormalization 的 trick

$$I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \implies \hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} x \quad (4-15)$$