2.2 Example

谱卷积神经网络 Spectral Convolutional Neural Networks (Spectral CNN)

谱卷积神经网络【3】假设滤波器 g_θ 是一组可学习的参数。考虑将公式 (2.18) $diag\left(\hat{h}(\lambda)\right)$ 替换为参数对角矩阵 $diag(\theta_l)$，于是得到的卷积层可表示为

$$y_{out} = f(U g_\theta U^T x) \tag{2.21}$$

其中

$$g_\theta = \begin{pmatrix} \theta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_n \end{pmatrix} \tag{2.22}$$

这里 $f(\cdot)$ 是一个激活函数，$\Theta = (\theta_1, \dots, \theta_n)$ 是待优化参数。若考虑多通道的输入输出，则记 $g_\Theta^{(k)} = \left\{\Theta_{i,j}^{(k)}\right\}, i = 1,2, \dots, N_{k-1}, j = 1,2, \dots, N_k$，这里 k 是卷积层索引，N_k 表示节点表示的通道数，$\left[\Theta_{i,j}^{(k)}\right]_{i=1}^{N_{k-1}}$ 表示第 k 层卷积层的第 j 个卷积通道。其中：

$$\Theta_{i,j}^{(k)} = \begin{pmatrix} \theta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_n \end{pmatrix}_{i,j}^{(k)} \tag{2.23}$$

于是多通道的 Spectral CNN 的图卷积层为

$$h_{:,j}^{(k)} = \sigma\left(\sum_{i=1}^{N_{k-1}} U \Theta_{i,j}^{(k)} U^T h_{:,i}^{(k-1)}\right), \qquad j = 1,2, \dots, N_k \tag{2.24}$$

Spectral CNN 所面临的问题：

• 每一次前向传播都需要计算 U, g_θ, U^T 三者的乘积，计算复杂度达 $O(n^3)$；

• 卷积核是全局的，并非有 Spatial Localization 性质；

• 单个卷积通道就需要 n 个参数，计算代价巨大。

局部连接的谱卷积神经网络 GCN with fast localized spectral filtering

为了解决（2.24）中 Spectral GCN 不具有局部感知的缺点，【4】提出了一种该网络的改进，其用$\sum_{j=0}^{K} \alpha_j \lambda_i^j$ 逼近（2.21）中 g_θ，这里的上标 j 表示特征值的次幂，于是单个卷积操作为

$$y_{out} = f(U g_\alpha(\Lambda) U^T x) \tag{2.25}$$

其中

$$\begin{aligned} g_\alpha(\Lambda) &= \begin{pmatrix} \sum_{j=0}^{K} \alpha_j \lambda_1^j & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{j=0}^{K} \alpha_j \lambda_n^j \end{pmatrix} \\ &= \sum_{j=0}^{K} \alpha_j \Lambda^j \end{aligned} \tag{2.26}$$

这里

$$\Lambda^j = \begin{pmatrix} \lambda_1^j & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n^j \end{pmatrix} \tag{2.27}$$

为了简化计算，考虑

$$\begin{aligned} & U g_\alpha(\Lambda) U^T \\ & = U \sum_{j=0}^{K} \alpha_j \Lambda^j U^T \\ & = \sum_{j=0}^{K} \alpha_j U \Lambda^j U^T \\ & = \sum_{j=0}^{K} \alpha_j L^j \end{aligned} \tag{2.28}$$

故

$$y_{out} = f\left(\sum_{j=0}^{K} \alpha_j L^j x\right) \tag{2.29}$$

其中 $\mathrm{A} = (a_0, \dots, \alpha_K)$ 是待优化参数。该卷积核的多通道卷积层的表示与（2.24）类似。

如此设计的卷积核的优点为

• 卷积核只有 K 个参数，在较大的图中，往往 $K \ll n$，参数的复杂度较低。

• 不需要进行特征分解，只需要使用 L 做变换。

• 具有 Spatial localization，K 就是卷积核的 Receptive Field，每次卷积只会将中心顶点的 K-hop neighbor 上的 feature 进行加权求和，权系数是α_k。

//

关于 Spatial Localization 的说明

常见的 Laplacian 有 以下三种形式

1. $L = D - A$: Combinatorial Laplacian;

2. $L^{sys} = D^{-1/2} L D^{-1/2}$: Symmetric normalized Laplacian;

3. $L^{rw} = D^{-1} L$: Random walk normalized Laplacian.

这里的 D 为度对角矩阵，A 为邻接矩阵。对于无向图来说 $A = A^T$，其行或列表示对应节点与其他节点的连接情况，同时注意到 D 为非负对角矩阵，于是考虑矩阵乘积

$$y^{(1)} = Lx \tag{*1}$$

向量 $y^{(1)}$ 中某一节点信息将只与其本身和邻居节点的表示有关。进一步地，考虑

$$y^{(2)} = Ly^{(1)} = L^2 x \tag{*2}$$

这里的 $y^{(2)}$ 只整合了 $y^{(1)}$ 中相应节点及其一阶（1-hop neighbor）邻居的信息，于是也包含了图信号 x 中二阶邻居节点的信息（2-hop neighbor）。

以此类推，则可以证明式（2.28）恰好整合了其 K 阶邻居的信息，亦可以称该卷积核拥有一个大小为 K 的感受野。另外应该注意到 $K = 0$ 时，$L^0 = I_n$，此时卷积核只对相对应的节点做线性映射。

//

Chebyshev Spectral CNN（ChebNet）利用切比雪夫多项式简化 Spectral GCN

利用 Chebyshev 多项式来逼近卷积核是谱卷积网络中一种被广泛应用的方法。其具体原理是通过 Chebyshev 多项式逼近（2.24）中的滤波器即

$$g_{\Theta}^{(k)} = \sum_{k=0}^{K} \alpha_k T_k(\widetilde{\Lambda}) \tag{2.30}$$

其中 $T_k(x) = 2xT_{k-1}(x) - T_{k-1}(x)$ 为 k 阶切比雪夫多项式，这里 $T_0(x) = 1$，$T_1(x) = x$；$\widetilde{\Lambda} = \frac{2\Lambda}{\lambda_{max}} - I_n$ 为 Re-Scaled 的特征值对角阵，该式将 Λ 取值范围限制在$[-1,1]$。

//

限制取值的解释

考虑第一类 Chebyshev 多项式的**解析形式**：

$$T_k(x) = \cos\big(k \cdot arcos(x)\big) \tag{$*1$}$$

因为 $\arccos(\cdot)$ 的存在，需要限制自变量的取值范围。

//

根据（2.30）卷积核的形式，图信号 x 与滤波器的卷积操作应记为

$$x_{*G}\, g_{\alpha} = U\left(\sum_{k=0}^{K} \alpha_k T_k(\widetilde{\Lambda})\right) U^T x \tag{2.31}$$

进一步考虑（2.28）中的简化计算的操作，记 $\tilde{L} = 2L/\lambda_{\max} - I_n$，得到

$$\begin{aligned} x_{*G}\, g_{\alpha} &= \left(\sum_{k=0}^{K} \alpha_k U T_k(\widetilde{\Lambda}) U^T\right) x \\ &= \left(\sum_{k=0}^{K} \alpha_k T_k(U\widetilde{\Lambda}U^T)\right) x \end{aligned} \tag{2.32}$$

上述变形归纳法可证，进而

$$
\begin{aligned}
x_{*G}\, g_{\alpha} &= \left(\sum_{k=0}^{K} \alpha_k T_k(\tilde{\mathrm{L}})\right) x \\
&= \sum_{k=0}^{K} \alpha_k T_k(\tilde{L}) x \qquad (2.33)
\end{aligned}
$$

与（2.28）相同的，这里将特征值矩阵与特征向量矩阵重新合并为拉普拉斯矩阵避免了繁杂的特征分解，其 Spatial Localization 的性质也与前者一致。

图卷积网络 举例 1

文章【5】中的 GCN 引入了 ChebNet 的一阶近似，即假设 $K=1$ 且 $\lambda_{max}=2$，其将 ChebNet 的卷积公式简化为：

$$x_{*G}\, g_{\theta} = \theta_0 x - \theta_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x \qquad (2.34)$$

为了限制参数这里假设了 $\theta = \theta_0 = -\theta_1$，一个与之等价的形式为：

$$x_{*G}\, g_{\theta} = \theta\left(I_n + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) x \qquad (2.35)$$

而其多通道的输入输出的扩展形式，即多通道卷积层为

$$H = X_{*G}\, g_{\Theta} = f(\bar{A} X \Theta) \qquad (2.36)$$

其中，$\bar{A} = I_n + D^{-1/2} A D^{-1/2}$；$\Theta = (\theta_1, \dots, \theta_n)$，这里参数中每一个元素应为一个大小为 n 的相同值的列向量；$f(\cdot)$为激活函数。另外，经验上来说，使用$I_n + D^{-1/2} A D^{-1/2}$会导致模型的数值不稳定，于是文章应用了一个归一化技巧

$$\bar{A} = \widetilde{D}^{-\frac{1}{2}} \tilde{A} \widetilde{D}^{-\frac{1}{2}} \qquad (2.37)$$

其中 $\tilde{A} = A + I_n$，$\widetilde{D}_{ii} = \sum_j \tilde{A}_{ij}$。

从空间的角度来看，GCN 可以被认为是聚合来自节点领域的特征信息，因此上述等式有等价表示：

$$h_v = f\left(\Theta^T\left(\sum_{u\in\{N(v)\cup v\}} \bar{A}_{v,u} x_u\right)\right), \qquad \forall v \in V \qquad (2.38)$$

图卷积网络 举例 2

【6】，二阶简化的 ChebNet**（待补 2.1： 整不动了）**

二阶简化的 ChebNet，其卷积层表示为

$$h^{(k+1)} = f\left(h^{(k)}, A\right) = \sigma\left(\widehat{D}^{-\frac{1}{2}}\hat{A}\widehat{D}^{-\frac{1}{2}}h^{(k)}W^{(k)}\right) \tag{2.39}$$

其中 $\hat{A} = I_n + A$，$\widehat{D}$ 是 $\hat{A}$ 的 Degree Matrix。

Adaptive Graph Convolutional Network (AGCN) 自适应图卷积网络

【7】

自适应图卷积网络(AGCN)通过图的邻接矩阵学习未知的隐藏结构关系。它通过一个以两个节点的特征为输入的可学习的距离函数来构造一个所谓的残差图邻接矩阵。

Dual Graph Convolutional Network (DGCN) 对偶图卷积网络

【8】

对偶图卷积网络引入了一种对偶图卷积结构，具有两个并行的图卷积层。虽然这两层共享参数，但它们使用归一化邻接矩阵 和正点互信息 (PPMI) 矩阵，