华南理工大学
South China University of Technology

研 究 生 课 程 论 文

(2018-2019 学年第一学期)

图卷积网络原理介绍

研究生：钟嘉杰

提交日期： 2018 年 1 月 1 日　　　　　　　　研究生签名：

学　　号	201720140049	学　　院	计算机科学与工程学院
课程编号	S0812006	课程名称	计算机科学高级专题
学位类别	硕士	任课教师	吕建明 副教授

教师评语：

成绩评定：　　分　　　　任课教师签名：　　　　年　月　日

图卷积网络原理介绍

钟嘉杰

摘要：图卷积网络与卷积神经网络类似，用于自动发现并提取图结构中的局部特征。由于图中局部邻域大小的不确定性，我们不能使用固定大小的卷积核对整个图进行卷积操作。本文将介绍一种基于图傅立叶变换的图卷积操作，并使用这种卷积操作构建图卷积网络，最后以一种直观的角度理解卷积操作的物理含义。

关键词：图卷积；谱空间；傅立叶变换

1. 介绍

卷积神经网络可以高效、准确地自动提取大规模数据集中一维或多维输入的特征，目前已被大规模用于图像分类、自然语言处理等领域。准确地说，卷积神经网络从输入中提取大量局部特征，这些提取到的局部特征通常被称为卷积核。卷积核对输入的数据的局部区域做线性组合产生输出，而线性组合的权重在不同的区域是共享的。但是，传统的卷积网络很难推广到图的输入上，对于一维或二维的输入信号，一个位置的邻域大小是固定的，而对于一个任意的图，一个节点的邻接节点的数量是不固定的，很难使用一个统一的卷积核对图的所有节点做卷积操作。对此，[1][2]利用图傅立叶变换[6]的技巧，把空域上的卷积操作替换为频域上的乘积操作，解决图上卷积运算的问题，从而产生了适用于任意图结构的图卷积网络。本文将对[1][2]所提出的图卷积网络的原理作简单介绍。

2. 预备知识

图的 Laplacian 矩阵：对于一个给定的有权图$G=<V,E>$，W为这个图的

邻接矩阵，W_{ij}为连边(V_i, V_j)的权重，对角矩阵D为这个图的广义度数矩阵，$D_{ii} = \sum_k W_{ik}$。该图的 Laplacian 矩阵L定义如下：

$$L = D - W \tag{1}$$

图的谱空间：对于一个给定的图G，图 Laplacian 矩阵为L，我们使用矩阵L的特征空间（Eigenspace）作为图的谱空间。设矩阵L的特征值分解为$L = U\Lambda U^T$，$\Lambda = \mathrm{diag}([\lambda_1, \dots, \lambda_i])$为特征值（Eigenvalue）矩阵，$U = [U_1, \dots, U_i]$为特征向量（Eigenvector）矩阵，则矩阵$U$为标准正交矩阵，$U_1, \dots, U_i$为该图的谱空间的基向量。

图傅立叶变换：对于一个给定的图$G =< V, E >$，图的 Laplacian 矩阵为 L，图的信号$X = [x_1, \dots, x_i]$，其中x_i为节点V_i上该信号的值。图上的傅立叶变换指的是把图的信号X变换到图的谱空间中。图傅立叶变换的计算过程可由以下公式给出：

$$X^* = U^T X$$

其中X_i^*为变换后的谱空间信号。对应地，图傅立叶逆变换的计算过程如下：

$$X = UX^*$$

由于矩阵 U为标准正交矩阵，即$UU^T = I$，所以$U = (U^T)^{-1}$为矩阵U^{T}的逆矩阵。

3. 图上的卷积操作

对于卷积神经网络来说，卷积操作是对输入信号的局部邻域的值的线性组合。由于图中节点的邻域大小是不固定的，所以我们不能使用一个统一大小的卷积核对全图所有节点的邻域做卷积操作。为此，我们可以通过图的傅立叶变换来解决这一问题。

定义：对于给定的图$G = (V, E)$，G的 Laplacian 矩阵的特征值分解为$L = U\Lambda U^T$，图上的信号为X。设一个卷积核在谱空间的形式为$g \in \mathbb{R}^{|V| \times |V|}$，则该卷积

核g与图的信号X的卷积操作$g * X$定义如下：

$$g * X = UgU^T X$$

此处卷积定义借用了信号处理领域中卷积的“时域卷积等价于频域乘积”的性质。先把输入信号变换到谱空间得到$X^* = U^T X$，在谱空间与卷积核g相乘得到谱空间的乘积$\hat{X}^* = gX^*$，最后再对该乘积做傅立叶逆变换得到该信号在原空间中的形式$\hat{X} = U\hat{X}^*$。把以上步骤结合起来就能得到上式。

特别地，当卷积核g为关于矩阵L的特征值矩阵Λ的k次多项式，即$g = \sum_i^k \theta_i \Lambda^i$时，卷积结果$\hat{X} = g * X$等价于对图$G$的所有半径为$k$的邻域的线性组合。证明如下：

$$g * X = UgU^T X$$

$$= U(\sum_i^k \theta_i \Lambda^i) U^T X$$

$$= (\sum_{i=0}^k \theta_i U \Lambda^i U^T) X$$

注意到$UU^T = U^T U = I$，$U\Lambda^i U^T$可变为:

$$U\Lambda^i U^T = (U\Lambda U^T)(U\Lambda U^T) \dots (U\Lambda U^T) = L^i$$

由此可得：

$$g * X = (\sum_{i=0}^k \theta_i L^i) X$$

设矩阵$M = \sum_{i=0}^k \theta_i L^i$。由图的邻接矩阵的性质可知，对于两个节点$V_i$和$V_j$, 考虑节点$V_i$上的卷积输出$\hat{X}_i = \sum_{j=0}^{|V|} M_{ij} \cdot X_j$，若从$V_i$到$V_j$的最短路径长度大于$k$，则$M_{ij} = 0$[3]。所以，$\hat{X}_i$等价于对节点$V_i$的半径为$k$的邻域的信号的线性组合。该操作等价于卷积神经网络上的卷积操作。

4. 图卷积网络

对于一个给定的图$G=(V,E)$，邻接矩阵为W，图的 Laplacian 矩阵为L。其中每个节点带有额外的d_0维特征向量（Feature Vector），特征向量矩阵$X \in \mathbb{R}^{|V| \times d_0}$。图卷积网络的结构定义如下：

输入：图卷积网络的输入为节点特征向量矩阵$X \in \mathbb{R}^{|V| \times d_0}$。

输出：图卷积网络的输出为节点的输出特征向量矩阵$\mathrm{Z} \in \mathbb{R}^{|V| \times d_l}$。

隐含层函数：设第l层隐含层的输入为$H^{(l-1)}$，输出为H^l，隐含层函数$\mathrm{H}^{(l)}=f\left(H^{(l-1)},L\right)$具有如下形式：

$$f\left(H^{(l-1)},L\right)=\sigma(Ug_\theta(\Lambda)U^T H^{(l-1)})$$

其中卷积核$g_\theta(\Lambda)$为关于Λ的 k 次多项式，$\sigma(\cdot)$为非线性激活函数，通常可以采用 ReLU 或 sigmoid 函数。

由于计算$Ug_\theta(\Lambda)U^T X$的计算复杂度非常高，为$\mathcal{O}(|V|^2)$，在神经网络训练时需要耗费非常多的时间，为此，我们对上式做一些简化。假定卷积的邻域半径为 1，即$g_\theta(\Lambda)$的最高次数为$k=1$，并且采用图的邻接矩阵W代替 Laplacian 矩阵L，我们可以得到简化后的隐含层函数：

$$H^{(l)}=f\left(H^{(l-1)},W\right)=\sigma(WH^{(l-1)}\Theta^{(l)})$$

其中，$\Theta^{(l)} \in \mathbb{R}^{d_{l-1} \times d_l}$为该卷积层的参数。以一个包含三个隐含层的图卷积网络为例，整个网络的计算过程如下：

$$Z=\sigma(W\sigma\left(W\sigma\left(WX\Theta^{(1)}\right)\Theta^{(2)}\right)\Theta^{(3)})$$

训练方式：得到卷积网络的输出 Z 后，我们可以在 Z 的基础上添加全连接 softmax 层，使用交叉熵损失函数引入监督信号反向传播训练图卷积网络的参数。

5. 以特征传播的角度看卷积层函数

虽然图的卷积层是通过图的傅立叶变换定义、推导而来，但是我们可以通过更直观的方式理解简化后的卷积层函数。考虑一个卷积层$f(X,W)=\sigma(WX\Theta)$，该隐含层函数包括以下三个过程：

特征传播：一个节点的特征向量由该节点的半径为1的邻域中所有节点的特征向量的加权平均值产生。该过程可表示为$X'=WX$。

特征过滤：使用神经网络的全连接层对传播后的特征进行过滤。该过程可表示为$\mathrm{X}''=\mathrm{X}'\Theta$。

非线性激活：该过程与神经网络中非线性激活的步骤相同，表示为$Y=\sigma(X'')$。

通过以特征传播的角度来解读图卷积网络中的卷积层函数，我们可以更直观地理解图卷积过程的物理含义，也可以更方便地在不同的使用场景下设计并改进卷积层函数。

参考文献

[1] Kipf T N , Welling M . Semi-Supervised Classification with Graph Convolutional Networks[J]. 2016.

[2] Defferrard, Michaël, Bresson X , Vandergheynst P . Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering[J]. 2016.

[3] Hammond D K , Vandergheynst P , Gribonval, Rémi. Wavelets on Graphs via Spectral Graph Theory[J]. Applied & Computational Harmonic Analysis, 2009, 30(2):129-150.

[4] Ntorina Thanou. Graph Signal Processing: Sparse Representation and Applications[R]. 2016.

[5] Daniel A. Spielman. Spectral Graph Theory and its Applications[R].

[6] Matthew Begue. Fourier Analysis on Graphs[R].