Detailed Explanation of GCN Mathematical Theory

——(math matters!)

SIGMA Group

Wei Chen

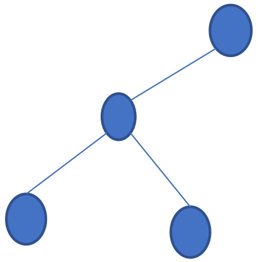
2022/06

Ref： gcn.pdf (ml-researcher.github.io)

- **Outline**
  - **1、Warmup（基础介绍）**
  - **2、Spectral graph theory（谱图理论基础知识）**
  - **3、Fourier transformation（傅里叶变换基础知识）**
  - **4、GCN（公式推导）**

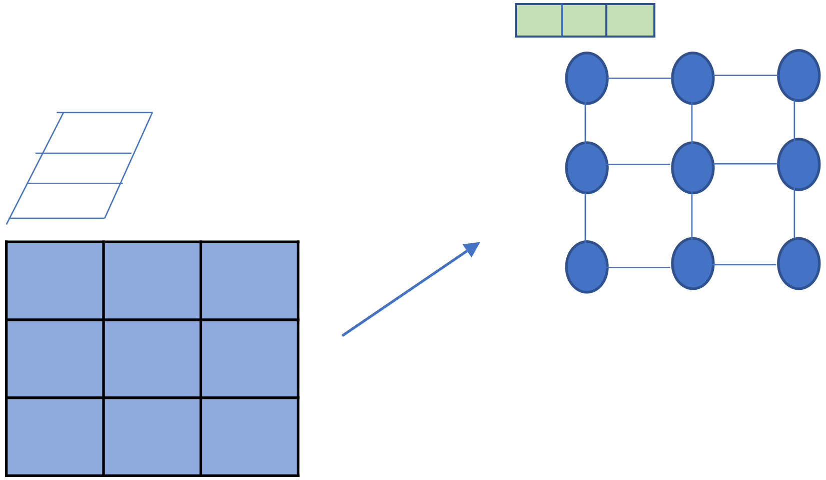
1、 Warmup

- Graph vs Image ?



Differences?

Connections?



It is composed of points and edges, indicating the relationship between any two entities

Represent the relationship between any two entities in a Euclidean space

Image is a special case of graph in European Space

So, the graph is more general and more important!

But…,what is graph neural network?

1、Warmup

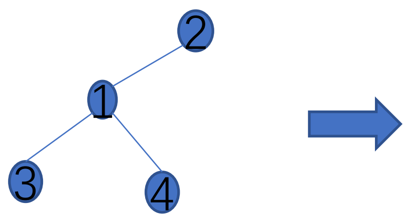
- Basic of GCN

  - GNN: $H^{(l+1)} = f\left(A, H^{(l)}\right)$

    - Note: For A (the adjacency matrix) and $H^{(l)}$ (input feature), the feature of the next layer is obtained through some mapping on the graph
    - All GNN related models are designing f, Different f get different GNN architectures

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

    - Note: here, we only consider undirected simple graph (No self loop, No duplicate edge).
    - Def: A: adjacency matrix D: degree matrix $\hat{A} = A + I$ $\widehat{D} = D + I$



A

0	1	1	1
1	0	0	0
1	0	0	0
1	0	0	0

D

3	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

1. Add self loop $\hat{A} = A + I$
2. Symmetric normalization $\widehat{D}^{-\frac{1}{2}}\hat{A}\widehat{D}^{-\frac{1}{2}}$
3. Aggregation of input features $\sigma(\widehat{D}^{-\frac{1}{2}}\hat{A}\widehat{D}^{-\frac{1}{2}}H^{(l)}\theta)$

•What’s more?

- Only Intuitive understanding?
  - Not enough!
- Know what is and why? (math matters!)
  - Spectral graph theory + Fourier transform

2、Spectral graph theory

- 什么是谱图理论?

  - 参考wiki:

Spectral graph theory

From Wikipedia, the free encyclopedia

In mathematics, **spectral graph theory** is the study of the properties of a graph in relationship to the characteristic polynomial, eigenvalues, and eigenvectors of matrices associated with the graph, such as its adjacency matrix or Laplacian matrix.

The adjacency matrix of a simple undirected graph is a real symmetric matrix and is therefore orthogonally diagonalizable; its eigenvalues are real algebraic integers.

While the adjacency matrix depends on the vertex labeling, its spectrum is a graph invariant, although not a complete one.

Spectral graph theory is also concerned with graph parameters that are defined via multiplicities of eigenvalues of matrices associated to the graph, such as the Colin de Verdière number.

  - 简单总结：研究与图的邻接矩阵相关的一些矩阵性质的领域

如何研究矩阵性质？线性代数

因此，谱图理论就是将线性代数研究矩阵的性质限定在与图的邻接矩阵相关的一些矩阵上（线代子领域）

2、Spectral graph theory

- 线代部分知识复习：
  - 1、特征值与特征向量
    - 定义：对于一个矩阵A，乘一个向量x，等于一个标量λ乘向量x $A\vec{x}=\lambda\vec{x}$ $|\vec{x}|\neq 0$

      则x是矩阵A的特征向量，λ是矩阵A的特征值

$A=U\Lambda U^T$ $UU^T=I$ $\Lambda=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}$

  - 2、实对称矩阵
    - 性质：若一个矩阵(n*n)是一个实对称阵，则它一定有n个特征值，并且这n个特征值对应着n个互相正交的特征向量
  - 3、半正定矩阵
    - 定义：所有特征值大于等于0的矩阵 $\forall i,\ \lambda_i\geq 0$
  - 4、二次型→瑞利熵
    - 二次型定义：给定一个矩阵A，右边乘向量x，左边乘向量x的转置，那么这个就是向量x对于矩阵A的二次型 $\vec{x}^TA\vec{x}$
    - 瑞利熵定义：一个向量关于矩阵A的二次型 与 这个向量关于单位矩阵的二次型 的比值 $\frac{\vec{x}^TA\vec{x}}{\vec{x}^T\vec{x}}$

      为什么研究二次型和瑞利熵？因为它与矩阵的特征值有着密切的联系

假定这里x是A的一个特征向量，我们可以证明瑞利熵等于这个矩阵对应的特征值。

$$\frac{\vec{x}^TA\vec{x}}{\vec{x}^T\vec{x}}$$

when $\vec{x}=U_i$

$$=\frac{\vec{x}^T\cdot(\lambda\vec{x})}{\vec{x}^T\vec{x}}=\frac{\lambda(\vec{x}^T\vec{x})}{\vec{x}^T\vec{x}}=\lambda$$

因此，瑞利熵是研究特征值的一个重要手段！

2、Spectral graph theory

- 谱图理论部分知识介绍:
  - 1、这里介绍和GCN密切相关的两个矩阵性质:

$$\begin{cases} L = D - A & \text{（拉普拉斯矩阵）} \\ L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} & \text{（拉普拉斯矩阵的对称规一化矩阵）} \end{cases}$$

  - 2、为什么研究这两个矩阵？具有优良的性质:
    - 2.1、这两个矩阵都是实对称阵，因此有n个特征值和n个互相正交的特征向量

$A = U \Lambda U^T$　$UU^T = I$　$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$

    - 2.2、这两个矩阵都是半正定阵，因此n个特征值都大于等于0　$\forall i,\ \lambda_i \geq 0$

证明:　positive semi-definite $\Leftarrow \dfrac{\vec{x}^T L \vec{x}}{\vec{x}^T \vec{x}} \geq 0$　　分母是向量的模长形式，恒大于0，因此只要证明分子大于0

这里我们构造一个辅助矩阵$G_{(i,j)}$　$G_{(i,j)} = \begin{bmatrix} \ddots & & & \\ & 1 & -1 & \\ & -1 & 1 & \\ & & & \ddots \end{bmatrix}$

由于拉普拉斯矩阵恰好就是所有i，j的边对于这个$G_{(i,j)}$求一个和，从而L可以重写为: $L = D - A = \sum_{(i,j)\in E} G_{(i,j)}$

为什么构造这样的辅助矩阵$G_{(i,j)}$
因为它的二次型很有特点

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

从而:

$$\begin{aligned} & \vec{x}^T L \vec{x} \\ &= \vec{x}^T \left(\sum G_{(i,j)}\right) \vec{x} \\ &= \sum \vec{x}^T G_{(i,j)} \vec{x} \\ &= \sum (x_i - x_j)^2 \geq 0 \end{aligned}$$

此外:

$$\begin{aligned} & \vec{x}^T L_{sym} \vec{x} \\ &= (\vec{x}^T D^{-\frac{1}{2}}) L (D^{-\frac{1}{2}} \vec{x}) \\ &= \sum \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}}\right)^2 \geq 0 \end{aligned}$$

2、Spectral graph theory

- 谱图理论部分知识介绍：
  - 1、这里介绍和GCN密切相关的两个矩阵性质：

$$L = D - A \quad \text{（拉普拉斯矩阵）}$$

$$L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} \quad \text{（拉普拉斯矩阵的对称规一化矩阵）}$$

  - 2、为什么研究这两个矩阵？具有优良的性质：
    - 2.2、这两个矩阵都是半正定阵，因此n个特征值都大于等于0，性质不够强，我们进一步证明其特征值∈ [0,2]

证明： positive semi-definite. $\Leftarrow \frac{\vec{x}^T L \vec{x}}{\vec{x}^T \vec{x}} \geq 0$ 分母是向量的模长形式，恒大于0，因此只要证明分子大于0

这里我们构造一个辅助矩阵$G_{(i,j)}$ $G^{pos}_{(i,j)} =$

我们定义一个postiva版本的拉普拉斯矩阵，即： $L^{pos} = D + A = \sum_{(i,j)\in E} G^{pos}_{(i,j)}$

同理辅助矩阵$G_{(i,j)}$它的二次型： $\vec{x}^T G^{pos}_{(i,j)} \vec{x} = (x_i + x_j)^2$

从而： $\vec{x}^T L^{pos} \vec{x} = \sum (x_i + x_j)^2 \geq 0$

此外： $L^{pos}_{sym} = D^{-\frac{1}{2}} L^{pos} D^{-\frac{1}{2}} = I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$

$\vec{x}^T L^{pos}_{sym} \vec{x} = \sum \left(\frac{x_i}{\sqrt{d_i}} + \frac{x_j}{\sqrt{d_j}}\right)^2 \geq 0$

$\vec{x}^T (I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) \vec{x} \geq 0$ $\Downarrow$ $\vec{x}^T \vec{x} + \vec{x}^T D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \vec{x} \geq 0$

$\vec{x}^T \vec{x} \geq -\vec{x}^T D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \vec{x}$ $\Downarrow$ $2\vec{x}^T \vec{x} \geq \vec{x}^T \vec{x} - \vec{x}^T D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \vec{x}$

$2\vec{x}^T \vec{x} \geq \vec{x}^T (I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) \vec{x}$ $\Downarrow$ $2\vec{x}^T \vec{x} \geq \vec{x}^T D^{-\frac{1}{2}} (D - A) D^{-\frac{1}{2}} \vec{x}$

$2\vec{x}^T \vec{x} \geq \vec{x}^T L_{sym} \vec{x}$ $\Downarrow$ $2 \geq \frac{\vec{x}^T L_{sym} \vec{x}}{\vec{x}^T \vec{x}}$ ☆

3、Fourier transformation

- 什么是傅里叶变换?

  - 参考百度百科:

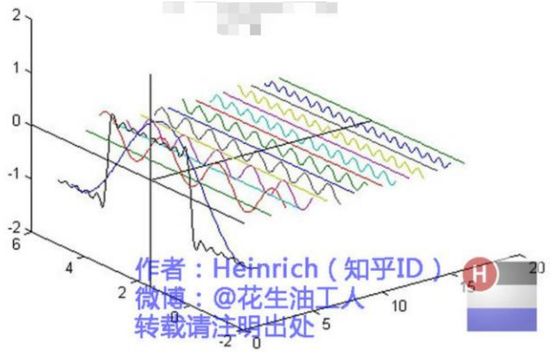
傅里叶变换 播报 编辑 讨论 23 上传视频

本词条由"科普中国"科学百科词条编写与应用工作项目 审核 。

傅里叶变换，表示能将满足一定条件的某个函数表示成三角函数（正弦和/或余弦函数）或者它们的积分的线性组合。

  - 简单解释：去研究同一个事物在不同域之间的不同视角是什么样的，以及在不同域之间的变换（信息无损）

example1



Why do this?

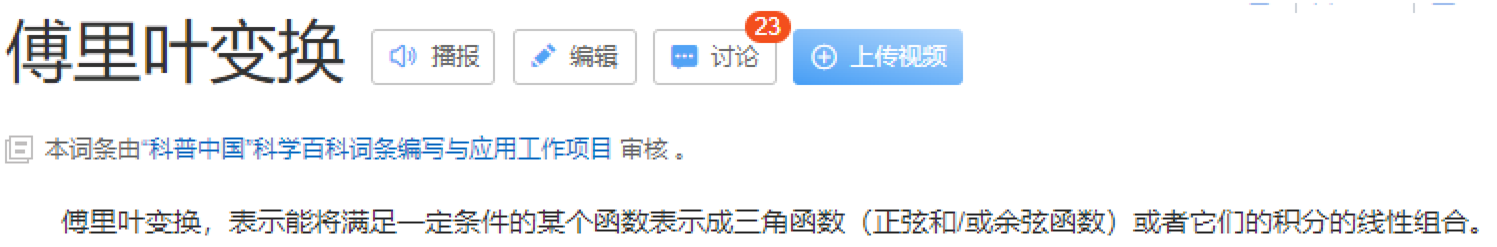
有一些操作在时域空间困难，但在频域空间十分容易!

f(x) → sound wave

3、Fourier transformation

- 什么是傅里叶变换?

  - 参考百度百科:



傅里叶变换 播报 编辑 讨论 23 上传视频

本词条由"科普中国"科学百科词条编写与应用工作项目 审核 。

傅里叶变换，表示能将满足一定条件的某个函数表示成三角函数（正弦和/或余弦函数）或者它们的积分的线性组合。

  - 简单解释：去研究同一个事物在不同域之间的不同视角是什么样的，以及在不同域之间的变换（信息无损）

example2

快速傅里叶变换(FFT)

$O(n^2)$

$f(x)=a_0+a_1x+a_2x^2+\cdots+a_nx^n$

$g(x)=b_0+b_1x+b_2x^2+\cdots+b_nx^n$

$f(x)g(x)=C_0+C_1x+C_2x^2+\cdots+C_{2n}x^{2n}$

我们知道：n次多项式可以由空间中n+1个不同的点确定

$$\overset{O(n\log n)}{\Longleftrightarrow}$$

$f(x) \Leftrightarrow (x_1,f(x_1))\ (x_2,f(x_2))\cdots(x_{n+1},f(x_{n+1}))$

$g(x) \Leftrightarrow (x_1,g(x_1))\ (x_2,g(x_2))\cdots(x_{n+1},g(x_{n+1}))$

$O(n)$

$f(x)g(x) \Leftrightarrow (x_1,f(x_1)g(x_1))\ (x_2,f(x_2)g(x_2))\cdots(x_{n+1},f(x_{n+1})g(x_{n+1}))$

3、Fourier transformation

- 什么是图上的傅里叶变换？

  - 为什么要在图上做傅里叶变换？
    - 在图像上，由于有规则地拓扑结构，因此我们可以用一个特定形状地kernel作为卷积核，但是图具有任意复杂地拓扑结构，从而我们无法给定一个特定形状地kernel。
    - 因此，在空间域做图的卷积十分困难！
    - 所以，借助傅里叶变换地思想，我们将其变换到另一个域中，在该域中，卷积十分容易，卷积之后，在进行傅里叶逆变换，变回空间域。

Fourier for Graph. : What Lx really does?

拉普拉斯矩阵 * n行列向量(节点feature)

$$Lx = \begin{bmatrix} \sum_{(1,j)\in E}(x_1-x_j) \\ \sum_{(2,j)\in E}(x_2-x_j) \\ \vdots \\ \sum_{(n,j)\in E}(x_n-x_j) \end{bmatrix}$$

X1与x1相关节点误差地累计

对x做一个聚合邻居地操作！(卷积本质)

3、Fourier transformation

- 什么是图上的傅里叶变换?
  - 在图上做傅里叶变换?
    - 借助傅里叶变换地思想，我们将其变换到另一个域中，在该域中，卷积十分容易，卷积之后，在进行傅里叶逆变换，变回空间域。

Lx即对x做一个聚合邻居地操作!

但…这和傅里叶变换有什么联系呢?

我们前面证明，L是一个实对称的半正定矩阵，则可以分解：

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

我们又知道U是一个正交阵，则U^T也是一个正交阵

而一个向量x乘以一个正交阵，就是代表一个空间系的基底变换

因此，该公式就是先对向量x做一个基底变换，接着放缩，在基底逆变换回去
即，图上的傅里叶变换即基底变换到新的域，进行放缩,实现聚合邻居效果（在这个域中，卷积十分容易，只用放缩），再逆回去

既然图上傅里叶变换对应卷积操作这么简单，那么给定一个图，我们直接计算一下拉普拉斯矩阵，然后分解一下特征值，特征向量，不就可以实现图卷积了?

$O(n^2)$

理论上没有问题，但实际分解特征值和向量复杂度是O(n^2),对于大图不现实!

GCN

对这种带特征值分解的傅里叶变换做限制，推导出一种不需要分解，复杂度与边的数量呈线性复杂度的方法!

4、GCN

Graph Convolution.

$$F(A) \rightarrow L \,/\, L_{sym}.$$

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

$$g_\theta * x = U g_\theta(\Lambda) U^T x = U\left(\sum_{k=0}^{K} \theta_k T_k(\Lambda)\right) U^T x$$

$$g_\theta(\Lambda) = \theta_0 \Lambda^0 + \theta_1 \Lambda^1 + \cdots + \theta_n \Lambda^n + \cdots$$

$$U g_\theta(\Lambda) U^T = g_\theta(U\Lambda U^T) = g_\theta(F(A))$$

$$(U\Lambda U^T)^k = U\Lambda U^T U\Lambda U^T \cdots U\Lambda U^T = U\Lambda^k U^T$$

Cheb Poly $\quad T_n(x) = 2x T_{n-1}(x) - T_{n-2}(x) \qquad T_0(x) = 1 \quad T_1(x) = x$

$$T_n(\cos\theta) = \cos n\theta$$

$$[-1, 1]$$

$$= \sum_{k=0}^{K} \theta_k U T_k(\Lambda) U^T x = \sum_{k=0}^{K} \theta_k T_k(U\Lambda U^T) x$$

$$= \sum_{k=0}^{K} \theta_k T_k(L_{sym} - I) x$$

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

$$\approx \theta_0 T_0(L_{sym} - I)x + \theta_1 T_1(L_{sym} - I)x$$

$$= \theta_0 x + \theta_1 (L_{sym} - I) x$$

$$= \theta_0 x - \theta_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x$$

$$\theta_1 = -\theta_0 \qquad \Rightarrow \cancel{\boxed{\theta_0}}\left(I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}\right) x$$

$$\Downarrow$$

$$\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} x$$