Graph Convolutional Network

Lexa

Rec-sys 04
March 21, 2022

Outline

Outline

What is the convolution?

Let $f, g : \mathbb{R} \to \mathbb{R}$. For $x \in \mathbb{R}$,

$$(f * g)(x) := \int_{-\infty}^{\infty} f(x-y)g(y)dy = \int_{-\infty}^{\infty} g(x-y)f(y)dy.$$

What is the Fourier transform?

Let $f, g : \mathbb{R} \to \mathbb{R}$. For $x, \xi \in \mathbb{R}$,

$$\mathcal{F}[f](\xi) := \int_{-\infty}^{\infty} f(x) e^{-2\pi i x\xi} dx, \quad \mathcal{F}^{-1}[f](x) := \mathcal{F}[f](-x)$$

Note : The convolution and Fourier transform are operators, that means,

$$\text{input, output} = \text{functions.}$$

Properties of FT and convolution

Let $f, g : \mathbb{R} \to \mathbb{R}$.

(i) The composition of $\mathcal{F}$ and $\mathcal{F}^{-1}$ are identity. That means,

$$\mathcal{F}[\mathcal{F}^{-1}[f]] = f, \quad \mathcal{F}^{-1}[\mathcal{F}[g]] = g.$$

(ii) The Fourier transform of convolution is a pointwise multiplication.

$$\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g].$$

Outline

Laplacian matrix

Let $G = (V, E)$ be a graph. Here, V is the set of nodes and E is the set of edges. A graph laplacian $L_G = D - A$, where D is the degree matrix and A is the adjacency matrix.

Laplacian matrix

We say that $N \times N$ matrix L is a *symmetric* matrix if

$$L = L^T.$$

Theorem

Let $G = (V, E)$ be a undirected graph with $|V| = N$, $|E| = M$.

1. *The graph laplacian L_G is a real symmetric matrix.*
2. *There exists a $M \times N$ matrix B such that.*

$$L_G = B^T B.$$

Laplacian matrix

Therefore, for $x \in \mathbb{R}^N$,

$$x^T L_G x = x^T B^T B x = \|Bx\|_2^2 \geq 0.$$

This implies that L_G is positive semidefinite, thus,

$$0 = \lambda_1 \leq \cdots \leq \lambda_N,$$

where λ_i's are eigenvalues of L_G.

Remark

We can easily find the smallest eigenvalue λ_1 and its eigenvector v_1. Let

$$v_1 := (1/\sqrt{N}, \cdots, 1/\sqrt{N}).$$

Then,

$$L_G v_1 = (0, \cdots, 0).$$

Basic concepts on linear algebra

We say that a set of vectors $\{e_1, \cdots, e_m\} \subset \mathbb{R}^d$ is orthonormal if $\|e_i\|_2 = 1$ for all $i = 1, \cdots, m$ and

$$\langle e_i, e_j \rangle = 0 \text{ for } i \neq j.$$

Theorem

Let A be a $N \times N$ matrix. The followings are equivalent.

1. *A is real and symmetric.*
2. *All eigenvalues are real and a set of eigenvectors $\{v_1, \cdots, v_N\}$ is orthonormal and linearly independent.*

Consequently, the graph laplacian L_G is represented by $Q\Lambda Q^T$ where $\Lambda = diag\{\lambda_1, \cdots, \lambda_N\}$ and $Q = [v_1, \cdots, v_N]$. (v_i is a column vector)

Graph Fourier transform

We say that $f : V \to \mathbb{R}$ is a graph *signal*. Any graph signal f can be represented by

$$f : V \to \mathbb{R} \quad \Longleftrightarrow \quad F = (f(V_1), \cdots, f(V_N)) \in \mathbb{R}^N.$$

A *graph Fourier transform* of signal F are defined by

$$\mathcal{F}[F] := Q^T F$$

A graph FT of F can be represented by

$$\mathcal{F}[F] = (v_1^T F, \cdots, v_N^T F) \in \mathbb{R}^N \quad \Longleftrightarrow \quad \mathcal{F}[f] : \{\lambda_1, \cdots, \lambda_N\} \to \mathbb{R},$$

where

$$\mathcal{F}[f](\lambda_i) := v_i^T F.$$

A *graph inverse Fourier transform* of $g : \{\lambda_1, \cdots, \lambda_N\} \to \mathbb{R}$ are defined by

$$\mathcal{F}^{-1}[g] := QG, \quad G := (g(\lambda_1), \cdots, g(\lambda_N)).$$

Graph Fourier transform

Summarizing this, GFT, IGFT : $\mathbb{R}^N \to \mathbb{R}^N$.

Graph convolution

For $x, y \in \mathbb{R}^N$ (signals), a *graph convolution* of x and y is defined by

$$x * y := \mathcal{F}^{-1}[\mathcal{F}[x]\mathcal{F}[y]] = Q[diag(Q^T x)](Q^T y).$$

Outline

Normalized laplacian matrix

For a given graph $G = (V, E)$, a *normalized graph laplacian* $\tilde{L}_G$ is defined by

$$\tilde{L}_G = D^{-1/2} L_G D^{-1/2} = I_N - D^{-1/2} A D^{-1/2}.$$

Remark

In general, we don't know upper bounds of eigenvalues of L_G. However, by mathematical computation (omit), we conclude that

$$0 = \tilde{\lambda}_1 \leq \cdots, \leq \tilde{\lambda}_N \leq 2,$$

where $\tilde{\lambda}_i$'s are eigenvalues of $\tilde{L}_G$. From now and after, U is a matrix of eigenvectors of $\tilde{L}_G$ and $\Lambda = diag\{\tilde{\lambda}_1, \cdots, \tilde{\lambda}_N\}$.

Convolution filter

For $\theta = (\theta_1, \cdots, \theta_N) \in \mathbb{R}^N$, $g_\theta : \{\tilde{\lambda}_1, \cdots, \tilde{\lambda}_N\} \to \mathbb{R}^N$

$$g_\theta(\tilde{\lambda}_i) = \theta_i.$$

Hence, we denote

$$g_\theta(\Lambda) = diag\{\theta_1, \cdots, \theta_N\}, \quad G_\theta := \mathcal{F}^{-1}[g_\theta(\Lambda)] = Qg_\theta(\Lambda).$$

The spectral convolution of signal $x \in \mathbb{R}^N$ with filter $g_\theta = g_\theta(\Lambda)$ is defined by

$$g_\theta \star x := G_\theta * x = Ug_\theta(\Lambda)U^T x$$

Note : To calculate this, we need eigendocomposition of $\tilde{L}_G$. However, this is computationally expansive. $\rightarrow$ Approximation!

Chebyshev polynomials

The Chebyshev polynomials T_n $(n = 1, 2, \cdots)$ is defined by

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

Theorem

Let $f : [-1, 1] \to \mathbb{R}$ be a piecewise smooth and continuous. Then, there exists $\{a_0, a_1, \cdots\} \subset \mathbb{R}$ such that

$$f(x) = \sum_{n=0}^{\infty} a_n T_n(x).$$

Note : Since $0 = \tilde{\lambda}_0 \leq \cdots \leq \tilde{\lambda}_N \leq 2$,

$$-1 \leq \frac{2}{\tilde{\lambda}_N}\tilde{\lambda}_i - 1 \leq 1.$$

Chebyshev polynomials

In [Wavelets on Graphs via Spectral Graph Theory, 2012], the authors proved that

$$g_\theta(\Lambda) \approx \sum_{i=0}^{K} \theta'_k T_k((2/\tilde{\lambda_N})\Lambda - I_N).$$

Therefore,

$$U g_\theta(\Lambda) U^T \approx \sum_{i=0}^{K} \theta'_k U T_k((2/\tilde{\lambda_N})\Lambda - I_N) U^T$$

and using the fact that

$$\begin{aligned} U((2/\tilde{\lambda_N})\Lambda - I_N)^k U^T = (U[(2/\tilde{\lambda_N})\Lambda - I_N]U^T)^k &= ((2/\tilde{\lambda_N})\tilde{L}_G - I_N)^k \\ &=: (\bar{L}_G)^k, \end{aligned}$$

we have

$$g_\theta \star x = U g_\theta(\Lambda) U^T x \approx \sum_{i=0}^{K} \theta'_k T_k(\bar{L}_G) x. \quad (\text{More computable!})$$

Layer-wise linear model

In this article, the authors assume that $\tilde{\lambda_N} \approx 2$.
Using first order approximation ($K = 1$),

$$g_\theta \star x \approx \theta'_0 x + \theta'_1 \bar{L}_G x = \theta'_0 x - \theta'_1 (D^{-1/2} A D^{-1/2}) x.$$

Here, we have two free parameters, θ'_0, θ'_1. The authors claim that $\theta'_0 = -\theta'_1 = \theta''$, thus,

$$g_\theta \star x \approx \theta'' (I_N + D^{-1/2} A D^{-1/2}) x.$$

For numerical stability, the authors used renormalization ttrick ;

$$g_\theta \star x \approx \theta'' (\bar{D}^{-1/2} \bar{A} \bar{D}^{-1/2}) x,$$

where $\bar{A} := A + I_N$ and $\bar{D} := D + I_N$.
Goal : Train θ''.

Layer-wise linear model

Now, we consider a signal vector $X = (X_1, \cdots, X_C)$. Each signal represents different features. That means, a signal is a matrix $X \in \mathbb{R}^{N \times C}$. Let $\Theta \in \mathbb{R}^{C \times F}$ be a matrix of filter parameters (weight matrix). Then, the convolution of X with filter Θ is defined by

$$Z := (\bar{D}^{-1/2} \bar{A} \bar{D}^{-1/2}) X \Theta \in \mathbb{R}^{N \times F}.$$