计算方法

刘景铖

计算机软件新技术国家重点实验室
南京大学

回顾与补充

上节课: 随机游走 (Random walk), Page Rank

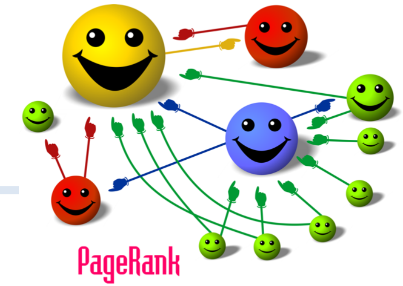
- 记X_t为时间t随机游走所处的状态, 则转移矩阵 $P_{i,j} = \Pr[X_{t+1} = j | X_t = i]$
- 记$p_t(i)$为时间t在状态i的概率, 则有$\overrightarrow{p_{t+1}} = \overrightarrow{p_t} \cdot P$

- **不可约: 强连通**
- **周期**：$\gcd\{\, t \mid (P^t)_{i,i} > 0\}$，周期为1称为非周期性的

对于有限，不可约，非周期的马尔可夫链，有以下事实：

1. 存在一个稳态分布$\vec{\pi}$.
2. **随着** $t \to \infty$， $\overrightarrow{p_t}$ **都会收敛到**$\vec{\pi}$，无论从什么样的$\overrightarrow{p_0}$开始.
3. 稳态分布是唯一的
4. $\pi(i) = \dfrac{1}{h_i}$

谱图理论 (Spectral graph theory)



谱分析：特征值 + 特征向量 + 相关的线性代数

图论与组合结构:

- 连通性 (Cheeger不等式)
- 图染色
- 聚类(Clustering)
- Mixing of random walks
- Expander graphs (efficient network, superconcentrators)

计算机科学:

- Pagerank
- 稀疏化Sparsification
- 迭代法解线性方程
- 电阻网络
- Expander codes (LDPC, Tanner codes)
- 不可近似性(Dinur's proof of the PCP theorem)
- 去随机化 (Derandomization)

矩阵的幂

回顾已经多次出现的矩阵的幂

- 对于可以对角化的矩阵A，收敛只需要特征值的幂是收敛的（特别地，谱半径$\rho(A) = 1$也有可能是收敛的）

$$\boldsymbol{x} = \boldsymbol{\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n}$$
$$\boldsymbol{A^k x = \lambda_1^k \alpha_1 v_1 + \lambda_2^k \alpha_2 v_2 + \cdots + \lambda_n^k \alpha_n v_n}$$

- 但一般来说，可能需要使用：谱半径$\rho(A) < 1$当且仅当 $\lim_{k\to\infty} A^k = 0$
  - 课后练习：考虑$\begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$的特征值？$\begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}^k$ =?
  - 注意：随机游走的转移矩阵一定不会是$\begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$，对应的只会是$\begin{bmatrix} \mathbf{1/2} & \mathbf{1/2} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$

- 随机游走：$\overrightarrow{p_{t+1}} = \overrightarrow{p_t} \cdot P$，转移矩阵$P := D^{-1}A$，及其相似的对角阵$W = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$
  - 这节课：马尔可夫链基本定理在无向图上成立：对于<u>连通的</u>，<u>非二分图</u>，存在唯一的稳态分布，且会收敛
  - 完整证明可参照Olle Häggström的Finite Markov chains and algorithmic applications

Graph spectrum

考虑图的邻接矩阵，它的代数性质（如特征值与特征向量）能否与图本身的组合性质（连通性，是否二分图）相对应？

图的邻接矩阵的例子：完全图 $A = J - I = \vec{1}\,\vec{1}^T - I$

J **的特征值和特征向量是什么？**

$$J\,\vec{1} = n\,\vec{1}$$

Graph spectrum

考虑图的邻接矩阵，它的代数性质（如特征值与特征向量）能否与图本身的组合性质相对应？

图的邻接矩阵的例子：二分图

- 引理：对于二分图G，如果 α 是$A(G)$ 的一个特征值，且重数为k，那么$-\alpha$也是$A(G)$ 的一个特征值，重数也是k

- 特征值的**重数**(multiplicity)
  - **代数重数**(algebraic multiplicity)：特征多项式里面，根的重复次数
  - 几何重数(geometric multiplicity)：特征值对应的特征空间的维度
  - 对于可对角化的矩阵（特别地，对于无向图的邻接矩阵，它们是一样的）

Graph spectrum

引理：对于二分图G，如果α是$A(G)$ 的一个特征值，且重数为k，那么$-\alpha$也是$A(G)$ 的一个特征值，重数也是k

证明：把A表示为$\begin{array}{c|cc} & U & V \\ \hline U & 0 & B \\ V & B^T & 0 \end{array}$，即 $\begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix}$

假设$\begin{pmatrix} x \\ y \end{pmatrix}$是$A$ 的一个特征向量，对应特征值为α：

$$\begin{pmatrix} By \\ B^T x \end{pmatrix} = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \alpha\begin{pmatrix} x \\ y \end{pmatrix}$$

因此 $B^T x = \alpha y, By = \alpha x$。再考虑$\begin{pmatrix} x \\ -y \end{pmatrix}$

$$A\begin{pmatrix} x \\ -y \end{pmatrix} = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix}\begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -By \\ B^T x \end{pmatrix} = \begin{pmatrix} -\alpha x \\ \alpha y \end{pmatrix} = -\alpha\begin{pmatrix} x \\ -y \end{pmatrix}$$

因此 $-\alpha$也是A 的特征值

最后，注意到：α的重数为 $k \Leftrightarrow$ 存在 k个线性无关的特征向量对应的特征值α

对每一个分别取反后依然是线性无关的，因此$-\alpha$的重数也为 k

Graph spectrum

事实上，引理的反方向也是对的。

引理：设$A(G)$的特征值为$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$，如果$\forall i,\ \alpha_i = -\alpha_{n-i+1}$，那么$G$一定是二分图

证明：首先证明：对于任意奇数k, $\sum_i \alpha_i^k = 0$

A的特征值为$\alpha_1, \alpha_2, \ldots, \alpha_n$则$A^k$的特征值为$\alpha_1^k, \alpha_2^k, \ldots, \alpha_n^k$，因此对于任意奇数k，

$$trace(A^k) = \sum_i \lambda_i^k = 0$$

组合含义: $(A^k)_{i,j}$ = 长度为k的从i到j的行走方案的数目

$$trace(A^k) = \sum_i (A^k)_{i,i} = 0\text{，又因为}(A^k)_{i,i} \geq 0\text{，所以}(A^k)_{i,i} = 0$$

即，对于任意奇数k，长度为k的环路不存在。因此，一定是二分图

Largest eigenvalue of adjacency matrix

图的邻接矩阵A的最大特征值$\alpha_{max} \leq \deg_{max}(G)$

证明：设v为对应最大特征值的特征向量，有$Av = \alpha_1 v$

令 $v_j = \max_i v_i > 0$，则$(Av)_j = (\alpha_1 v)_j$

$$\Rightarrow \alpha_1 v_j = \sum_i A_{j,i} v_i \leq \deg_{max}(G) \cdot v_j$$

$$\Rightarrow \alpha_{max} \leq \deg_{max}(G)$$

Laplacian matrix

给定图G，它的拉普拉斯矩阵$L(G)$定义为$L(G) := D(G) - A(G)$, 其中 $D(G)$是对角线上为度数的矩阵

对于d-正则图， $L(G) = dI - A(G)$，因此正则图的拉普拉斯矩阵和邻接矩阵的特征空间是一样的。但一般图上并不成立。

可以写出$L(G) = \sum_{e \in E} L_e$，其中$L_e$是只有e这一条边的图的拉普拉斯矩阵

另外：$x^T L\, x = \sum_{ij \in E} (x_i - x_j)^2$

Laplacian matrix

给定图G，它的拉普拉斯矩阵$L(G)$定义为$L(G) := D(G) - A(G)$, 其中 $D(G)$是对角线上为度数的矩阵

性质：$\vec{1}$ 是$L(G)$的一个特征向量，对应特征值是0

性质：$L(G) \succcurlyeq 0$

Connectedness

定理：给定图G，它是连通的当且仅当$L(G)$的特征值0的重数为1

证明(<=)：先考虑G是不连通的情况。此时我们需要证明$L(G)$的特征值0的重数大于1.

$$L(G) = \begin{array}{c c} & \begin{array}{cc} V_1 & V_2 \end{array} \\ \begin{array}{c} V_1 \\ V_2 \end{array} & \begin{bmatrix} L(G_1) & 0 \\ 0 & L(G_2) \end{bmatrix} \end{array}$$

有2组线性无关的特征向量，对应特征值为0

Connectedness

定理：给定图G，它是连通的当且仅当$L(G)$的特征值0的重数为1

证明(=>)：接下来考虑G是连通的。假设$L\,x = 0$，则$x^T L\,x = \sum_{ij\in E}(x_i - x_j)^2 = 0$

因此 $(x_i - x_j)^2 = 0, \forall ij \in E$，即$x_i = x_j, \forall ij \in E$

因为 G是连通的，所以也有$x_i = x_j, \forall i,j \in V$

因此 $x = \mathrm{c}\,\vec{1}$

Second eigenvalue

把$L(G)$的特征值记为$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$

刚刚的定理证明了： G是非连通的当且仅当$\lambda_2 = 0$, **或者说** G是连通的当且仅当$\lambda_2 > 0$

更一般地（课后练习）：

$\lambda_k = 0$当且仅当 G有至少k个连通分量。

Robust connectedness

连通性和二分图在组合上也非常容易刻画，为什么需要特征值？

事实上，可以更加量化特征值的刻画：

- λ_2 非常小，当且仅当G 非常接近于不连通(i.e. 存在非常稀疏的割).
- λ_k非常小，当且仅当G 非常接近于存在k个连通分量(i.e. k个不相交的非常稀疏的分割).
- 对于邻接矩阵的特征值$\lambda_n \approx -\lambda_1$ 当且仅当G中存在接近于二分的分量

Perron-Frobenius定理

设A是非负的，不可约且非周期的矩阵（不一定对称）。

1. 最大特征值的重数是1
2. 对应的特征向量中，每个维度都是非零且同号的
3. $|\lambda_i| < \lambda_1$，对于$2 \leq i \leq n$.

证明可参见："Algebraic graph theory" Chapter 8, by Godsil and Royle

Graph conductance

Recall that G is disconnected if and only if $\lambda_2 = 0$.

Cheeger's inequality will show that G is "close" to be disconnected if and only if λ_2 is "small".

We will first define precisely what it means to be close to be disconnected.

The conductance of a set $S \subseteq V$ is defined as

$$\phi(S) \coloneqq \frac{|\delta(S)|}{\mathrm{vol}(S)},$$

where $\mathrm{vol}(S) \coloneqq \sum_{v \in S} \deg(v)$.

When the graph is d-regular, $\phi(S) \coloneqq \frac{|\delta(S)|}{d|S|}$.

Note: the expansion of a set S is defined as $\frac{|\delta(S)|}{|S|}$; For d-regular graphs, they're basically the same.

The conductance of a graph G is defined as $\phi(G) \coloneqq \min\limits_{S:\mathrm{vol}(S) \leq m} \phi(S)$.

Note that $0 \leq \phi(G) \leq 1$.

Expander graphs and sparse cuts

A graph G with constant $\phi(G)$ (e.g. $\phi(G) = 0.1$) is called an expander graph.

A set S with small $\phi(S)$ is called a sparse cut.

Both concepts are very useful.

Finding a sparse cut is useful in designing divide-and-conquer algorithms, and have applications in

- image segmentation
- data clustering
- community detection
- VLSI-design
- ....

Spectral partitioning algorithm

A popular heuristic used in practice.

1. Compute an eigenvector $x \in \mathbb{R}^n$ corresponding to the second smallest eigenvalue of $\mathcal{L}$.
2. Sort the vertices so that $x_1 \geq x_2 \geq \cdots \geq x_n$.
3. Let $S_i \coloneqq \begin{cases} \{1, \ldots, i\} & \text{if } i \leq \frac{n}{2} \\ \{i+1, \ldots, n\} & \text{if } i \geq \frac{n}{2} \end{cases}$

Return $\min\limits_{1 \leq i \leq n} \phi(S_i)$.

This is simple and can be implemented in near-linear time.

It performs very well in practice, especially in image segmentation.

Normalized matrices

To state Cheeger's inequality nicely, we introduce normalized adjacency and Laplacian matrices.

Let $\mathcal{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ be the normalized adjacency matrix.

Let $\mathcal{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ be the normalized Laplacian matrix.

Note that $\mathcal{L} = I - \mathcal{A}$.

When the graph is d-regular, $\mathcal{A} = \frac{1}{\mathrm{d}}\mathrm{A}$ and $\mathcal{L} = \frac{1}{d} L$.

Claim. Let $\alpha_1 \geq \cdots \geq \alpha_n$ be the eigenvalues of $\mathcal{A}$, and $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of $\mathcal{L}$.

Then, $1 = \alpha_1 \geq \cdots \geq \alpha_n \geq -1$, and $0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$.

Cheeger's inequality

Cheeger's Inequality [Cheeger 70, Alon-Milman 85]

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

The first inequality is called the easy direction, and the second inequality is called the hard direction. We give some intuition in the case when G is a d-regular graph.

For the easy direction: think of λ_2 as a "relaxation" of the graph conductance problem.

$$\phi(G) \approx \min_{x \perp 1 \,:\, x \text{ is binary}} \frac{\sum_{ij \in E}\left(x_i - x_j\right)^2}{d \sum_{i \in V} x_i^2} \qquad \text{and} \qquad \lambda_2 = \min_{x \perp 1} \frac{\sum_{ij \in E}\left(x_i - x_j\right)^2}{d \sum_{i \in V} x_i^2}.$$

For the hard direction: given an optimizer x for λ_2, we want to produce a set S with $\phi(S) \leq \sqrt{2\lambda_2}$. The idea is to use the <u>spectral partitioning algorithm</u>: for a "fractional" x, we try to round it to an integral (binary) solution x. This is known as "rounding".