搜索引擎的PageRank模型

浙江大学 谈之奕

Google

- Google的创立

Brin和Page进入斯坦福大学攻读博士学位

Brin和Page创立Google公司

Google公司在Nasdaq上市

Google公司总资产达到3200亿美元

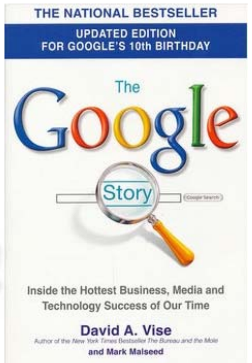
1995 1996 1998 2004 2020

建立了Google搜索引擎的核心模型与算法**PageRank**

Googol

10^{100}

Google!



Lawrence Edward Page (1973-)
Google创始人

Sergey Mikhaylovich Brin (1973-)
Google创始人

Vise DA, Malseed M, *The Google Story: Inside the Hottest Business, Media, and Technology Success of Our Time*, Delacorte Press, 2005. (中译本：撬动地球的Google，中信出版社，2006)

搜索引擎

- 搜索引擎（search engine）
  - 一种用于帮助用户查找因特网信息的技术工具。其工作原理是先以一定的策略在因特网中搜索，发现信息，然后对信息进行理解、提取、组织和处理，最后为用户提供信息和服务
- 互联网信息获取的特点与难点
  - 互联网上有海量信息，网页及其内容处于动态变化中，缺乏统一管理
  - 具有相同关键词的网页数量众多，如何展示搜索结果，提升用户搜索体验
  - 如何给出一种合理、客观、定量、可操作的网页排序规则，使“重要”的网页在搜索结果中靠前显示
- PageRank的网页排序机制
  - 互联网中的网页通过超链接（hyperlink）关联
  - 网页的重要程度由网页之间的链接关系决定

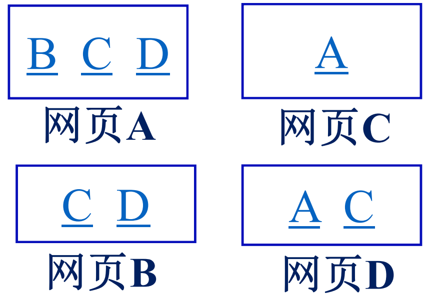
网页重要度

数学
建模
MATH T

- 网页重要度的原则与假设
  - 某网页重要，是因为有重要的网页链接到它
    - 对任一网页A，确定一数值为其重要度，作为网页排序的依据
    - 链接到网页A的所有网页对网页A的重要度均有贡献，贡献大小与这些网页自身的重要度有关
  - （传递性）重要度大的网页链接到网页A时对网页A的重要度的贡献比重要度小的网页链接到网页A时对网页A的重要度的贡献大
    - 某网页对其它网页重要度的贡献之和等于它的重要度
  - （等效性）网页对它所链接的每个网页的重要度的贡献相等
    - 某网页对其它网页的重要度贡献与它所链接的网页数量呈反比
  - （叠加性）链接到网页A的网页越多，网页A越重要
    - 网页A的重要度是所有链接到A的网页对网页A的重要度的贡献之和
  - （无关性）网页链接其它网页的多少，与其本身的重要度无关

网页重要度

- 网络链接图
  - 互联网中网页之间的链接关系可用图表示，称为网络链接图
    - 顶点：网页 $v_1, v_2, \cdots, v_n$
    - 弧：网页间的链接关系
      - 若网页 v_i 上有链接指向网页 v_j，则网络链接图中有一条以 v_i 为起点，v_j 为终点的弧
    - 出度：以某顶点为起点的弧的总数，即该网页链接的网页数量



- 网页重要度
  - 网页 v_i 的重要度记为 x_i，出度记为 q_i
    - 网页 v_i 对其它网页重要度贡献之和为 x_i
    - 网页 v_i 对它链接的 q_i 个网页中的任一个的重要度贡献为 $\frac{x_i}{q_i}$
    - 若链接到网页 v_j 的网页有 $v_{j_1}, v_{j_2}, \cdots, v_{j_k}$，则

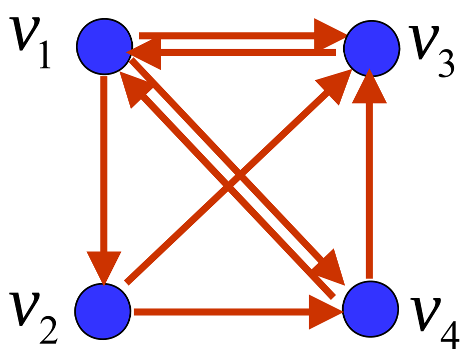
$$x_j = \frac{x_{j_1}}{q_{j_1}} + \frac{x_{j_2}}{q_{j_2}} + \cdots + \frac{x_{j_k}}{q_{j_k}}$$

传递性

等效性

叠加性

无关性



网页重要度

- 网页重要度
  - 线性方程组

$$x_j = \frac{x_{j_1}}{q_{j_1}} + \frac{x_{j_2}}{q_{j_2}} + \cdots + \frac{x_{j_k}}{q_{j_k}}, \quad j = 1, \cdots, n$$

$$= p_{i1}x_1 + p_{i2}x_2 \cdots + p_{in}x_n = \sum_{j=1}^{n} p_{ij}x_j,$$

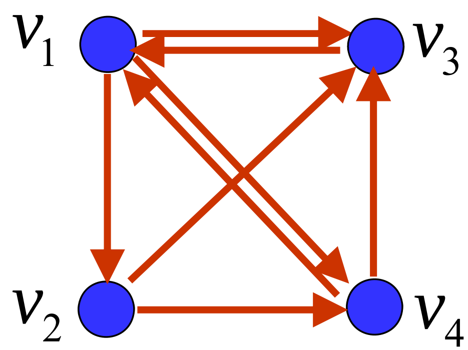
  - 矩阵表示
    - 令 $p_{ij} = \begin{cases} \frac{1}{q_j}, & 若有链接自v_j链向v_i \\ 0, & 其他 \end{cases}$
    - 记矩阵 $\mathbf{P} = (p_{ij})_{n\times n}$ 为（初始）链接矩阵，$\mathbf{X} = (x_1, \cdots, x_n)^{\mathrm{T}}$为网页重要度向量
    - $\mathbf{X}$ 为线性方程组 $\mathbf{X} = \mathbf{P}\mathbf{X}$ 的解
      - $\mathrm{rank}(\mathbf{I} - \mathbf{P}) < n$
      - 矩阵 $\mathbf{P}$ 有特征值1

线性方程组是否一定有非零解？

$$\begin{cases} x_1 = & & x_3 + \frac{1}{2}x_4 \\ x_2 = \frac{1}{3}x_1 & & \\ x_3 = \frac{1}{3}x_1 + \frac{1}{2}x_2 & & + \frac{1}{2}x_4 \\ x_4 = \frac{1}{3}x_1 + \frac{1}{2}x_2 & & \end{cases}$$

$$\mathbf{P} = \begin{pmatrix} 0 & \boxed{0} & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}$$

$$\mathbf{I} - \mathbf{P} = \begin{pmatrix} 1 & 0 & -1 & -1/2 \\ -1/3 & 1 & 0 & 0 \\ -1/3 & -1/2 & 1 & -1/2 \\ -1/3 & -1/2 & 0 & 1 \end{pmatrix}$$



	1	2	3	4
1			√	√
2	√			
3	√	√		√
4	√	√		
出度	3	2	1	4

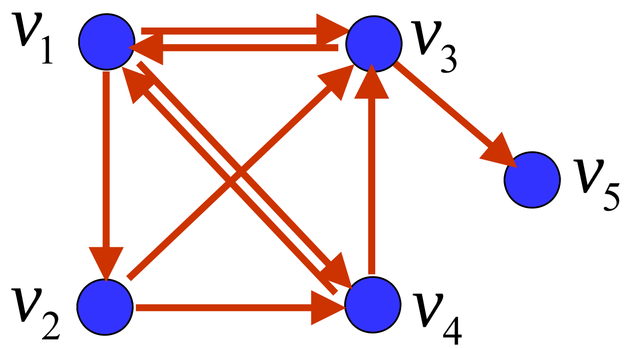
$$x_1 = \frac{12}{31}, x_2 = \frac{4}{31}, x_3 = \frac{9}{31}, x_4 = \frac{6}{31}$$

随机矩阵

- 随机矩阵
  - 各行（列）元素之和均为 1 的非负方阵称为行（列）随机矩阵（row（column） stochastic matrix）
  - 各行与各列元素之和均为 1 的非负方阵称为双随机矩阵（doubly stochastic matrix ）
- 随机矩阵的模最大特征值
  - 任一随机矩阵的模最大特征值为 1
    - 设 λ 是行随机矩阵 $\mathbf{P}=(p_{ij})_{n\times n}$ 的特征值，非零向量 $\mathbf{X}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 为属于特征值 λ 的特征向量。设 $|x_i|=\max\limits_{1\le j\le n}|x_j|>0$
    - 由 $\mathbf{PX}=\lambda\mathbf{X}$，可得 $\lambda x_i=\sum\limits_{j=1}^{n}p_{ij}x_j$
    - 两边取模，$|\lambda||x_i|=|\lambda x_i|=\left|\sum\limits_{j=1}^{n}p_{ij}x_j\right|\le\sum\limits_{j=1}^{n}|p_{ij}||x_j|\le|x_i|\sum\limits_{j=1}^{n}|p_{ij}|=|x_i|$，即 $|\lambda|\le 1$

链接矩阵

- 链接矩阵
  - 链接矩阵 $\mathbf{P}$ 的每列元素之和为 1 ，为列随机矩阵
  - 悬挂网页（dangling link）
    - 若某网页不链接任意其它网页，将链接矩阵 $\mathbf{P}$ 中对应列的所有元素由 0 修改为 $\frac{1}{n}$ ，得到（修正）链接矩阵 $\overline{\mathbf{P}}$



$$\overline{\mathbf{P}} = \mathbf{P} + \frac{1}{n}\mathbf{1}\mathbf{d}^{\mathrm{T}}$$

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 & 1/5 \\ 1/3 & 0 & 0 & 0 & 1/5 \\ 1/3 & 1/2 & 0 & 1/2 & 1/5 \\ 1/3 & 1/2 & 0 & 0 & 1/5 \\ 0 & 0 & 1/2 & 0 & 1/5 \end{pmatrix}$$

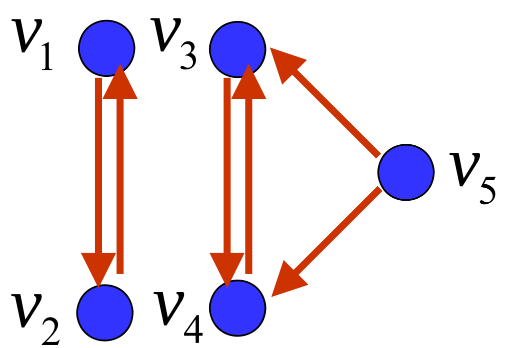
$$\overline{\mathbf{P}} = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & \frac{1}{5} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{5} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & \frac{1}{5} \end{pmatrix} = \mathbf{P} + \frac{1}{5}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

线性方程组 $\mathbf{X} = \mathbf{P}\mathbf{X}$ 是否只有唯一的线性无关的解

矩阵 $\mathbf{P}$ 属于特征值 1 的线性无关的特征向量是否是唯一的

链接矩阵

- 链接矩阵
  - 若 $\overline{\mathbf{P}}$ 有两个属于特征值 1 的线性无关的特征向量，重要度向量排序不唯一
  - 将 $\overline{\mathbf{P}}$ 修改为（最终）链接矩阵 $\overline{\overline{\mathbf{P}}} = \alpha\overline{\mathbf{P}} + (1-\alpha)\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}$，其中参数 $\alpha = 0.85$
    - $\overline{\overline{\mathbf{P}}}$ 为完全正矩阵（totally positive matrix）与列随机矩阵
      - $\mathbf{1}^{\mathrm{T}}\overline{\overline{\mathbf{P}}} = \alpha\mathbf{1}^{\mathrm{T}}\overline{\mathbf{P}} + (1-\alpha)\frac{1}{n}\mathbf{1}^{\mathrm{T}}\mathbf{1}\mathbf{1}^{\mathrm{T}} = \alpha\mathbf{1}^{\mathrm{T}} + (1-\alpha)\mathbf{1}^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}}$



$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\overline{\overline{\mathbf{P}}} = 0.85\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.5 \\ 0 & 0 & 1 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + 0.15\cdot\frac{1}{5}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0.03 & 0.88 & 0.03 & 0.03 & 0.03 \\ 0.88 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.88 & 0.455 \\ 0.03 & 0.03 & 0.88 & 0.03 & 0.455 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \end{pmatrix}$$

数学
建模
MATH T

唯一性

数学建模 MATH T

- 唯一性
  - 完全正、列随机矩阵属于特征值 1 的特征向量分量之和不为 0
    - 设 $\mathbf{x}=(x_1,\cdots,x_n)^{\mathrm T}$ 为完全正、列随机矩阵 $\mathbf{P}$ 的属于特征值 1 的特征向量，则 $x_i=\sum_{j=1}^{n}p_{ij}x_j$
    - 若 $\sum_{i=1}^{n}x_i=0$，则 $\mathbf{x}$ 的分量有正有负，故 $|x_i|=\left|\sum_{j=1}^{n}p_{ij}x_j\right|<\sum_{j=1}^{n}p_{ij}\left|x_j\right|, i=1,\cdots,n$
    - $\sum_{i=1}^{n}|x_i|<\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij}\left|x_j\right|=\sum_{j=1}^{n}\left|x_j\right|\left(\sum_{i=1}^{n}p_{ij}\right)=\sum_{j=1}^{n}\left|x_j\right|$ 矛盾
  - 完全正、列随机矩阵仅有 1 个属于特征值 1 的线性无关的特征向量
    - 设 $\mathbf{v}=(v_1,\cdots,v_n)^{\mathrm T}$ 和 $\mathbf{w}=(w_1,\cdots,w_n)^{\mathrm T}$ 是完全正、列随机矩阵 $\mathbf{P}$ 的两个属于特征值 1 的线性无关的特征向量。令 $x_i=-\frac{W}{V}v_i+w_i, i=1,\cdots,n$，其中 $V=\sum_{k=1}^{n}v_k\neq 0$，$W=\sum_{k=1}^{n}w_k$
    - 由 $\mathbf{v}$ 和 $\mathbf{w}$ 线性无关，与 $\sum_{j=1}^{n}p_{ij}x_j=\sum_{j=1}^{n}p_{ij}\left(-\frac{W}{V}v_j+w_j\right)=-\frac{W}{V}\sum_{j=1}^{n}p_{ij}v_j+\sum_{j=1}^{n}p_{ij}w_j=-\frac{W}{V}v_i+w_i=x_i$，可知 $\mathbf{x}=(x_1,\cdots,x_n)^{\mathrm T}$ 为 $\mathbf{P}$ 的属于特征值 1 的特征向量
    - $\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}\left(-\frac{W}{V}v_i+w_i\right)=-\frac{W}{V}\sum_{i=1}^{n}v_i+\sum_{i=1}^{n}w_i=0$ 矛盾

Perron—Frobenius定理

- Perron定理
  - 若矩阵 **A** 为完全正矩阵，则
    - **A**的模最大特征值唯一，且为正实数
    - 该特征值代数重数为 1
    - 存在该特征值的一个特征向量，其分量全为正
- Perron—Frobenius定理
  - 若矩阵 **A** 为非负不可约（irreducible）矩阵，则
    - **A** 的模最大特征值为正实数
    - 该特征值代数重数为 1
    - 存在该特征值的一个特征向量，其分量全为正
- 链接矩阵与重要度向量
  - 链接矩阵为完全正、列随机矩阵，模最大特征值为 1，重要度向量唯一且分量全为正

Perron O, Zur Theorie der Matrices, *Mathematische Annalen*, 64, 248–263, 1907

Frobenius G, Ueber Matrizen aus nicht negativen Elementen, *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, 456–477, 1912

数学
建模
MATH T

Oskar Perron
（1880－1975）
德国数学家

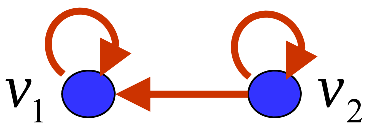
Georg Frobenius
（1849－1917）
德国数学家

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \surd$$

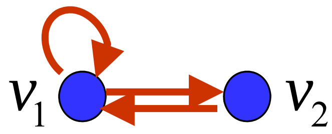
$\lambda_1 = \lambda_2 = 0 \qquad \lambda_1 = \lambda_2 = 1$

不可约矩阵

- 不可约矩阵
  - 若干个初等对换矩阵的乘积称为置换矩阵（permutation matrix）
    - 置换矩阵每行和每列都恰有一个元素为 1，其余元素都为 0
  - 若存在置换矩阵 $\mathbf{Q}$，使得 $\mathbf{Q}^{\mathrm{T}}\mathbf{A}\mathbf{Q}=\begin{pmatrix}\mathbf{X} & \mathbf{0}\\ \mathbf{Y} & \mathbf{Z}\end{pmatrix}$，其中 $\mathbf{X}$ 和 $\mathbf{Z}$ 均为方阵，则称 $\mathbf{A}$ 为可约矩阵（reducible matrix），否则 $\mathbf{A}$ 为不可约矩阵
- 不可约矩阵与有向图
  - 若对有向图中任意顶点对 v_i, v_j，既存在一条从 v_i 到 v_j 的有向路，也存在一条从 v_j 到 v_i 的有向路，则称有向图是强联通（strongly connected）的
  - 给定非负矩阵 $\mathbf{A}=(a_{ij})_{n\times n}$，构造有向图 $G(\mathbf{A})=(V,A)$，其中 $V=\{v_1,v_2,\cdots,v_n\}$，弧 $(v_i,v_j)\in A$ 当且仅当 $a_{ij}>0$
  - $\mathbf{A}$ 是不可约矩阵当且仅当 $G(\mathbf{A})$ 是强联通的

$\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}$ 

可约

$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$ 

不可约

矩阵计算

数学
建模
MATH T

- The World's Largest Matrix Computation
  - One of the reasons why Google is such an effective search engine is the PageRank™ algorithm, developed by Google's founders, Larry Page and Sergey Brin, when they were graduate students at Stanford University
  - PageRank is determined entirely by the link structure of the Web. It is recomputed about once a month and does not involve any of the actual content of Web pages or of any individual query. Then, for any particular query, Google finds the pages on the Web that match that query and lists those pages in the order of their PageRank
  - In May 2002, n was about 2.7 billion
  - The matrix G is huge, but very sparse; its number of nonzeros is the total number of hyperlinks in the pages. The matrix A is not sparse, but it is a rank one modification of a sparse matrix. Most of the elements of A are equal to the small constant δ . When $n = 2.7\times10^{9}$, $\delta = 5.5\times10^{-11}$

Cleve Barry Moler
（1939－ ）
美国数学家
Matlab创始人

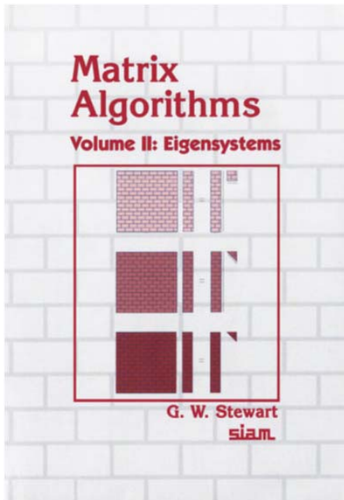
Moler CB, The World's Largest Matrix Computation, 2002.
https://ww2.mathworks.cn/company/newsletters/articles/the-world-s-largest-matrix-computation.html

幂法

数学建模 *MATH T*

- 幂法
  - 计算矩阵模最大特征值和对应的特征向量的一种迭代算法
    - 任取初始向量 $\mathbf{x}^{(0)}, \mathbf{x}^{(0)} > 0$ 且 $\mathbf{1}^{\mathrm{T}}\mathbf{x}^{(0)} = 1$
    - 计算 $\mathbf{x}^{(k)} = \mathbf{P}\,\mathbf{x}^{(k-1)}$
      - $\mathbf{1}^{\mathrm{T}}\mathbf{x}^{(k)} = \mathbf{1}^{\mathrm{T}}\mathbf{P}\,\mathbf{x}^{(k-1)} = \mathbf{1}^{\mathrm{T}}\,\mathbf{x}^{(k-1)} = 1$

The advantages of the power method are that it requires only matrix-vector multiplications and that it is essentially globally convergent, provided the matrix has a single, simple dominant eigenvalue. It has the disadvantages that it only computes the dominant eigenvector, and the convergence can be very slow if the dominant eigenvalue is near in magnitude to the subdominant one. Moreover, if a real matrix has a complex dominant eigenvalue, then its conjugate is also an eigenvalue with the same absolute value, and the power method cannot converge. For these reasons, the power method in its natural form is little used these days. However, it plays an implicit role in the

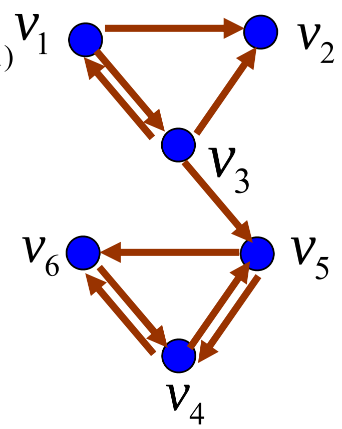


Stewart GW, *Matrix Algorithms* (two volumes). SIAM, 1998

von Mises R, Pollaczek-Geiringer H, Praktische verfahren der gleichungsauflosung. *Zeitschrift für Angewandte Mathematik und Mechanik*, 9, 58–77, 152–164, 1929.

幂法

- 链接矩阵的幂法
  - $$\mathbf{x}^{(k)} = \bar{\bar{\mathbf{P}}}\,\mathbf{x}^{(k-1)} = \left(\alpha\bar{\mathbf{P}} + (1-\alpha)\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{x}^{(k-1)} = \left(\alpha\left(\mathbf{P} + \frac{1}{n}\mathbf{1}\mathbf{d}^{\mathrm{T}}\right) + (1-\alpha)\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right)\mathbf{x}^{(k-1)}$$
    $$= \alpha\mathbf{P}\mathbf{x}^{(k-1)} + \alpha\frac{1}{n}\mathbf{1}\mathbf{d}^{\mathrm{T}}\mathbf{x}^{(k-1)} + (1-\alpha)\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{x}^{(k-1)}$$
    $$= \alpha\mathbf{P}\mathbf{x}^{(k-1)} + \alpha\frac{1}{n}\mathbf{1}\mathbf{d}^{\mathrm{T}}\mathbf{x}^{(k-1)} + (1-\alpha)\frac{1}{n}\mathbf{1}$$



$$\bar{\bar{\mathbf{P}}} = \alpha\bar{\mathbf{P}} + (1-\alpha)\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}$$

$$\bar{\mathbf{P}} = \mathbf{P} + \frac{1}{n}\mathbf{1}\mathbf{d}^{\mathrm{T}}$$

$$\bar{\bar{\mathbf{P}}} = \begin{pmatrix} 0.025 & 0.1667 & 0.3083 & 0.025 & 0.025 & 0.025 \\ 0.45 & 0.1667 & 0.3083 & 0.025 & 0.025 & 0.025 \\ 0.45 & 0.1667 & 0.025 & 0.025 & 0.025 & 0.025 \\ 0.025 & 0.1667 & 0.025 & 0.025 & 0.45 & 0.875 \\ 0.025 & 0.1667 & 0.3083 & 0.45 & 0.025 & 0.025 \\ 0.025 & 0.1667 & 0.025 & 0.45 & 0.45 & 0.025 \end{pmatrix}$$

$$\mathbf{x}^{(0)} = (0.166667, 0.166667, 0.166667, 0.166667, 0.166667, 0.166667)^{\mathrm{T}}$$
$$\mathbf{x}^{(5)} = (0.057165, 0.083312, 0.063942, 0.338898, 0.196007, 0.260676)^{\mathrm{T}}$$
$$\mathbf{x}^{(10)} = (0.052057, 0.074290, 0.057821, 0.347973, 0.199759, 0.268101)^{\mathrm{T}}$$
$$\mathbf{x}^{(20)} = (0.051706, 0.073681, 0.057414, 0.348701, 0.199903, 0.268594)^{\mathrm{T}}$$
$$\mathbf{x}^{(25)} = (0.051705, 0.073679, 0.057412, 0.348704, 0.199904, 0.268596)^{\mathrm{T}}$$

幂法

- 完全正、列随机矩阵幂法的收敛性 $\boxed{\mathbf{1}^{\mathrm{T}}\mathbf{P}=\mathbf{1}^{\mathrm{T}}}$ $\boxed{\mathbf{P}=\left(p_{ij}\right)_{n\times n},\ p_{ij}>0}$
  - 记 $\mathbf{V}$ 为满足 $\mathbf{1}^{\mathrm{T}}\mathbf{v}=0$ 的 n 维列向量 $\mathbf{v}=(v_1,\cdots,v_n)^{\mathrm{T}}$ 全体组成的集合。记 $\|\mathbf{v}\|_1=\sum_{i=1}^{n}|v_i|$
  - 对任意的 $\mathbf{v}\in\mathbf{V}$，$\mathbf{w}=\mathbf{P}\mathbf{v}\in\mathbf{V}$，且 $\|\mathbf{w}\|_1\le c\|\mathbf{v}\|_1$，其中 $c<1$
    - 由 $\mathbf{1}^{\mathrm{T}}\mathbf{w}=\mathbf{1}^{\mathrm{T}}\mathbf{P}\mathbf{v}=\mathbf{1}^{\mathrm{T}}\mathbf{v}=0$ ，$\mathbf{w}\in\mathbf{V}$
    - 若 $\mathbf{w}=\mathbf{0}$，$\|\mathbf{w}\|_1\le c\|\mathbf{v}\|_1$ 显然成立。若 $\mathbf{w}\ne\mathbf{0}$，记 $e_i=\operatorname{sgn} w_i$ ，则 $c=\max\limits_{j}\left|\sum_{i=1}^{n}p_{ij}e_i\right|<1$
    - $\|\mathbf{w}\|_1=\sum_{i=1}^{n}|w_i|=\sum_{i=1}^{n}e_iw_i=\sum_{i=1}^{n}e_i\left(\sum_{j=1}^{n}p_{ij}v_j\right)=\sum_{j=1}^{n}v_j\left(\sum_{i=1}^{n}p_{ij}e_i\right)\le\sum_{j=1}^{n}|v_j|\left|\sum_{i=1}^{n}p_{ij}e_i\right|\le c\sum_{j=1}^{n}|v_j|=c\|\mathbf{v}\|_1$
  - $\lim\limits_{k\to\infty}\mathbf{x}^{(k)}=\mathbf{X}$ ，其中 $\mathbf{X}$ 为属于特征值 1 的特征向量 $\boxed{\mathbf{P}\mathbf{X}=\mathbf{X}}$
    - 记 $\mathbf{v}_0=\mathbf{x}^{(0)}-\mathbf{X}\in\mathbf{V}$，$\mathbf{x}^{(k)}=\mathbf{P}^k\mathbf{x}^{(0)}=\mathbf{P}^k(\mathbf{X}+\mathbf{v}_0)=\mathbf{P}^k\mathbf{X}+\mathbf{P}^k\mathbf{v}_0=\mathbf{X}+\mathbf{P}^k\mathbf{v}_0$
    - 由 $\left\|\mathbf{P}^k\mathbf{v}_0\right\|_1\le c^k\|\mathbf{v}_0\|_1$ ，$\lim\limits_{k\to\infty}\mathbf{x}^{(k)}=\mathbf{X}$

随机浏览

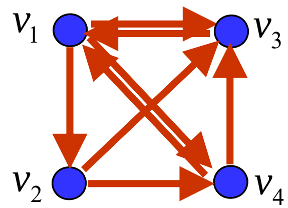
数学建模 *MATH T*

- 随机浏览（Random Surfer）
  - 按以下模式浏览互联网的网页
    - 有时从当前网页的链接中随机打开一个网页
    - 有时键入网址新建一个网页

$5:1 \quad \frac{5}{6} \approx \alpha \quad \bar{\bar{\mathbf{P}}} = \alpha \bar{\mathbf{P}} + (1-\alpha)\frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}$

  - 从任一网页开始，充分长时间后，访问各网页的概率即为网页重要度
- 极限概率
  - 记事件 $\{X_m = j\}$ 为时刻 m 访问网页 j，则 $P\{X_m = i \mid X_{m-1} = j\} = p_{ij}$
  - 若 $P\{X_{m-1} = j\} = x_j, j = 1,\cdots,n$，则 $P\{X_m = i\} = \sum_{j=1}^{n} P\{X_m = i \mid X_{m-1} = j\} P\{X_{m-1} = j\} = \sum_{j=1}^{n} p_{ij} x_j$
  - 记 $\mathbf{x}^{(m)} = (P\{X_m = 1\}, P\{X_m = 2\}, \cdots, P\{X_m = n\})^{\mathrm{T}}$ 为时刻 m 浏览各网页的概率，

$$\mathbf{x}^{(m)} = \bar{\bar{\mathbf{P}}}\mathbf{x}^{(m-1)}$$



$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{pmatrix}$$

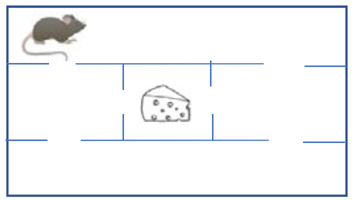
随机过程

数学
建模
MATH T

- 随机过程（stochastic process）
  - 描述随机现象随时间推移而演化的一类数学模型
  - 一族随机变量 $\{X(t), t \in T\}$，其中 t 是参数，T 为参数集
    - T 为整数集的随机过程称为随机序列
- Markov过程
  - 在已知目前的状态（现在）的条件下，它未来的演变（将来）不依赖于它以往的演变（过去）
  - 随机序列$\{X_n, n = 0, 1, 2, \cdots\}$，$X_n$ 只能取有限个或可数个数值
    - $P\{X_{n+1} = i_{n+1}\}$ 只与 X_n 有关，而与 $X_{n-1}, X_{n-2}, \cdots, X_0$ 无关
    - 对任意 $n \geq 0$和一列状态 $i_0, i_1, \cdots, i_{n-1}, i_n, i_{n+1}$，

$$P\{X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1}, X_n = i_n\} = P\{X_{n+1} = i_{n+1} \mid X_n = i_n\}$$

Andrey Andreyevich Markov
（1856 –1922）
俄罗斯数学家



PageRank模型

数学
建模
MATH T

- PageRank模型
  - 数学基石
  - 成功经验
    - 数学理论与现实应用的相互适应、完美结合
    - 建模与求解全过程体现简明性与实效性
    - Google搜索引擎与Google公司的运作
  - 思想源流
    - 学术论文的引用与文献计量学（bibliometrics）

1906	Markov	Markov theory
1910s	Perron, Frobenius	Perron–Frobenius theorem
1929	von Mises, Pollaczek-Geiringer	Power method

Jon Michael Kleinberg (1971-)
美国康奈尔大学计算机科学系、信息科学系教授，**2006**年**Nevanlinna**奖得主

Kleinberg JM. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5): 604-632, 1999

Papers are judged not only on their original thinking, but also on the number of papers they cite, the number of papers that subsequently cite them back, and the perceived importance of each citation.

——Battelle J. The Birth of Google. *Wired*, 2005.

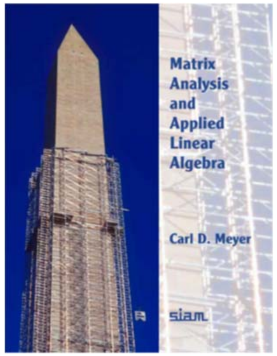
- 体育排名

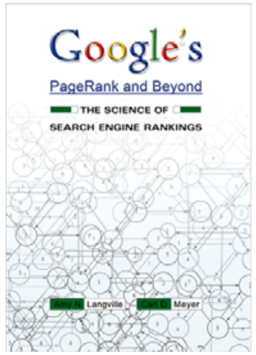
Keener JP. The Perron-Frobenius theorem and the ranking of football teams. *SIAM Review*, 35, 80-93, 1993.

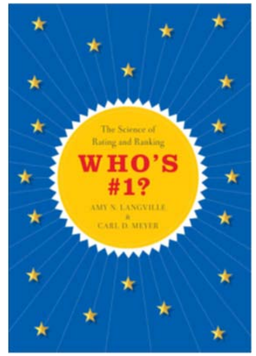
PageRank模型

- PageRank模型
  - In addition to saying something useful, the Perron–Frobenius theory is elegant. It is a testament to the fact that beautiful mathematics eventually tends to be useful, and useful mathematics eventually tends to be beautiful
  - The now seminal 1998 paper used Markov chains, our favorite mathematical tool, to rank webpages. Two little known graduate students had used Markov chains to improve search engine rankings and their method was so successful that it became the foundation for their fledgling company, which was soon to become the extremely well-known search engine Google
  - Netflix places such importance on their ability to accurately rank movies for their users that in 2007 they hosted a competition. Their Netflix Prize awarded $1 million to the individual or group that improved their recommendation by 10%

Gleich DF. PageRank beyond the Web. *SIAM Review*, 2015, 57(3): 321-363.







Meyer CD, *Matrix Analysis and Applied Linear Algebra*, SIAM, 2000.

Langville AN, Meyer CD. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2011.

Langville AN, Meyer CD. *Who's # 1?: The Science of Rating and Ranking*. Princeton University Press, 2012.


谢 谢