Patterns and models that are?

valid useful unexpected understandable

Data Mining tasks.

① Descriptive methods eg. cluster

② Predictive methods eg. recommender sys

Usage Quality context streaming

scaiability 扩展

Links as votes. ① more links, Page more impor

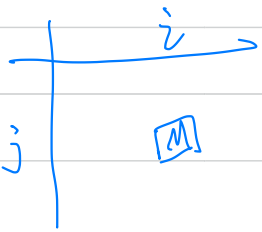
② Links from important Page count more.

in-link: 点连接至该页面 $i \to j$ $M_{ji} = \frac{1}{d_i}$

Value j = 来源Value分支

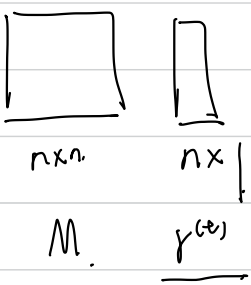
$\sum_i r_i = 1$ All Page score

$r = M \cdot r$



eigenvector：特征向量

$Ax=\lambda x$ why 其中A是M矩阵

$y^{(t+1)}$ = 

Link analysis approaches:
- Page Rank
- Topic-specitic Page Rank
- Web spam detecting algorithms

Page Rank:
- 有更多link的 page越重要. in-link like vote
- Link from 重要页面的 count more.
- 公式描述:
  - 一个 page 的 score 为它所有 in-link 权重之和 → r_j
  - 每个 link 的权重等于它 source page 权重/out-link数目.

$$r_j = \sum_{i \to j} \frac{r_i}{d_i}$$ 其中 d_i 为节点 i 的出度数目

$M_{ji} = \frac{1}{d_i}$ $M_{ji} = 0$

M 列的和为 1.

M 列随机矩阵

权重向量 r: 代表每个 page 的 score

$\sum_i r_i = 1$ ← 所有页面 score 之和为 1

M_{ji} 指从 i 出发到 j 的值是 $\frac{1}{d_i}$

目的 \ 出发	0	1	2	j
0				
i 1				
2				

$M \cdot r = r$ 故 1 是特征值

特征向量: $Ax = \lambda x$ 则 x 是特征向量, λ 是特征值

考虑到 r 向量中每个分量之和为 1 而 M 的每 column 之和为 1 故 $Mr \le 1$

迭代法: 依据 $r = Mr$ 计算

初始化: $r^{(0)} = \begin{bmatrix} \frac{1}{N} \\ \vdots \\ \frac{1}{N} \end{bmatrix}$ N 为结点数目

$r^{(t+1)} = M \cdot r^{(t)}$ until $|r^{(t+1)} - r^{(t)}| < \varepsilon$

总结 $r^{(n)} = M \cdot r^{(n-1)} = M^n r^{(0)}$

结果接近主特征向量

随机游走：$P(t+1) = M \cdot P(t)$，假设达到一个状态 $P(t+1) = M \cdot P(t) = P(t)$

则 $P(t)$ 是 stationary distribution

结论：对于满足特定条件的图形，静态分布是唯一的，无论 t=0 时初始概率为何，最终都达到静态分布.

Each time step，都执行

Page Rank的问题：是 problem

① dead ends：结点无出点 to go out　　r 很易变成形如 $\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

② spider traps：局限在某个 group.　　trap 很容易到最终的 r 为类似 $\begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$

不算 problem 哈.

Trap 的 solution：Teleports.

每次跳有2个 choice
- β 概率 follow link random
- 1-β prob: jump to random page

β 一般 0.8~0.9

Dead End 的 solution：Teleport too!

1.0 概率随机跳

sparse matrix：稀疏矩阵

使用 Teleport 之后，公式变成 $r_j = \sum_{i \to j} \beta \frac{r_i}{d_i} + (1-\beta)\frac{1}{N}$

即 $A = \beta M + (1-\beta)\left[\frac{1}{N}\right]_{N \times N}$

$r = Ar = \beta M r + \underbrace{\left[\frac{1-\beta}{N}\right]_{N \times N}}$ → since $\sum_{i=1}^{n} r_i = 1$

为常数，故 $r^{new} = \beta M \cdot r^{old}$

直接加 $\frac{1-\beta}{N}$