

## 線形代数—ページランカー

Copyright ©2006 by Shigeto R. Nishitani

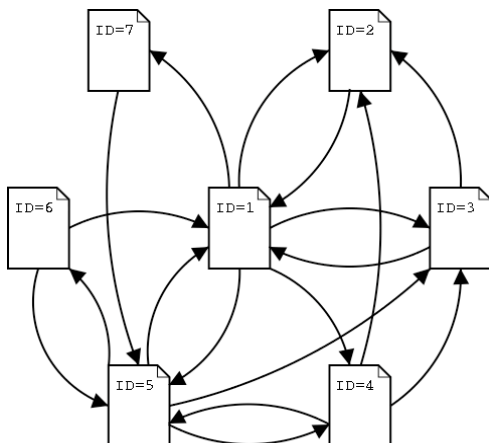
Googleのpage rankは非常に単純な仮定

「多くの良質なページからリンクされているページはやはり良質なページである」から成り立っている。

詳しい解説は<http://www.kusastro.kyoto-u.ac.jp/~baba/wais/pagerank.html>にある。

### ▼ ホームページのリンク

[つぎのようなリンクが張られたページを考える。



### ▼ 計算手順

1 リンクを再現する隣接行列を作る。ページに番号をつけて、その間が結ばれているi-j要素を1, そうでない要素を0とする。

上の例では,

1 2 3 4 5 6 7

```
-----  
0 1 1 1 1 0 1  
1 0 0 0 0 0 0  
1 1 0 0 0 0 0  
0 1 1 0 1 0 0  
1 0 1 1 0 1 0  
1 0 0 0 1 0 0  
0 0 0 0 1 0 0
```

となる。

2 行列を転置する。これはページランカーが「どれだけリンクしているか」ではなく、「どれだけリンクされているか」を評価するためである。

3 それぞれの列ベクトルの総和が1となるように規格化して推移確率行列をつくる。

4 初期ベクトルは、すべての要素が同じ値で、足して1になるように用意する。

5 初期ベクトルに何度か(例えば10回、あるいは収束するまで)推移確率行列を掛ける。この操作は、行列の最大固有値に属する固有ベクトルを見つけることに相当する。

6 得られたベクトルの各要素が対応するページの得点とみなせ、得点順にランクが高くなる。

### ▼ 演習

1 うえの手順にしたがって、ページランクをもとめよ。

2 初期ベクトルとして

$$v_0[1]=100.0, v_0[2]=0.0, \dots (\text{後は} 0)$$

として行列のかけ算に伴う各要素の変化を観察し、前問と比較せよ。遷移行列の意味を記せ。

3 行列の固有値・固有ベクトルをもとめ、最大固有値に対応する固有ベクトルを取り出せ。前問までに得られた結果と比較し、一致していることを確かめよ。ただし、固有ベクトルの大きさは任意であることに気をつけよ。