数理メディア情報学レポート(2)

情報科学類二年 江畑 拓哉 (201611350)

1 Google の PageRank について下記の問いに答えなさい。

One of the reasons why Google is such an effective search engine is the PageRankˆ(tv) 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 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.
Imagine surfing the Web, going from page to page by randomly choosing an outgoing link from one page to get to the next. So, a certain fraction of the time, simply choose a random page from anywhere on the Web. This theoretical random walk of the Web is a *Markov chain or Markov process* . The limiting probability that a dedicated random surfer visits any particular page is its PageRank. A page has high rank if it has links to and from other pages with high rank.

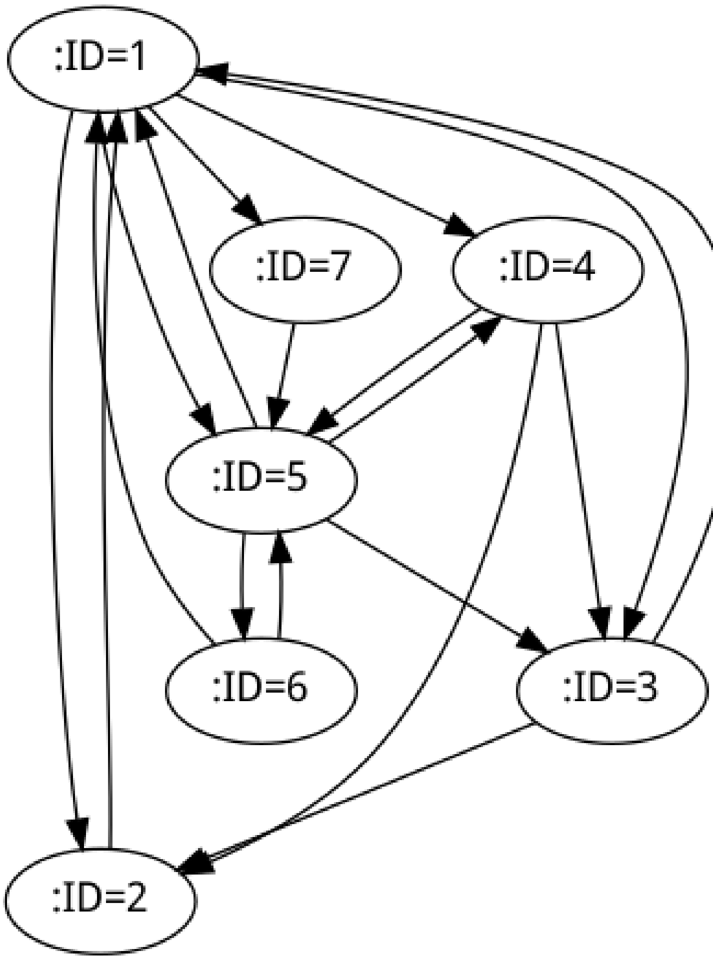
1.1 上の文章を訳しなさい

Google があのように素晴らしい検索エンジンである理由の一つに、 PageRank アルゴリズムがあります。それは Google の 創業者である Larry Page と Sergey Brin が Stanford 大学 を卒業する際に開発したアルゴリズムで、 PageRank は Web のリンク構造に寄ってのみ決定づけられます。つまりそれは Web 上のいかなる実際のコンテンツや個人のクエリにはかかわりません。そして特定のあるクエリによって、 Google は それに適合する Web 上のページを見つけ、 PageRank に順位付けられたそれらのページを列挙します。

Web サーフィンを想像してみたとき、あるページからページへの外へ向かうリンクをランダムに選択してページ間を移動しています。 つまり、ある程度の場合では Web 上の任意の場所からランダムなページを選択しているに過ぎないのです。この理論的な Web 上の散策は、 Markov chain of Markov process と呼ばれています。専用のランダムなサーファーが特定のページへとアクセスする確率を制限するものが PageRank です。高い順位のついた他のページから、あるいはそれらへのリンクのあるページには、高い順位がつけられます。

1.2 上記の PageRank を求める計算方法について具体的に詳述しなさい。

具体的に以下の図を用いて説明する。



この有向グラフに基づいて以下のルールで行列を作成する。

- ID 番号の数　× ID 番号の数の零行列を作る
- 各行、各列にそれぞれの ID 番号を割り振る
- i 行に対応する ID から j 列に対応する ID に向かってリンクがなされているとき、i 行 j 列の値を 1 と置く

- 最終的に各行で、1 の数を足してその数で行全体の 1 の要素をすべて割る

(最後の試行は現在いる位置からそれぞれのリンク先に飛ぶ確率が均一であるという仮定に基づいている)
これにより、以下のような行列が求まる

	1	2	3	4	5	6	7
1	0	1/5	1/5	1/5	1/5	0	1/5
2	1	0	0	0	0	0	0
3	1/2	1/2	0	0	0	0	0
4	0	1/3	1/3	0	1/3	0	0
5	1/4	0	1/4	1/4	0	1/4	0
6	1/2	0	0	0	1/2	0	0
7	0	0	0	0	1	0	0

この行列は確率行列 P となり、 $M = P^T$ である M が Google の推移行列である。M の固有値を λ 、固有ベクトルを V とする。
ここで M 、 λ 、V の値を示しておくと、

$$M = \begin{bmatrix} 0.00000 & 1.00000 & 0.50000 & 0.00000 & 0.25000 & 0.50000 & 0.00000 \\ 0.20000 & 0.00000 & 0.50000 & 0.33333 & 0.00000 & 0.00000 & 0.00000 \\ 0.20000 & 0.00000 & 0.00000 & 0.33333 & 0.25000 & 0.00000 & 0.00000 \\ 0.20000 & 0.00000 & 0.00000 & 0.00000 & 0.25000 & 0.00000 & 0.00000 \\ 0.20000 & 0.00000 & 0.00000 & 0.33333 & 0.00000 & 0.50000 & 1.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.25000 & 0.00000 & 0.00000 \\ 0.20000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \end{bmatrix} \tag{1}$$

$$\lambda = \begin{bmatrix} 1.00000 + 0.00000i \\ -0.44433 + 0.23415i \\ -0.44433 - 0.23415i \\ 0.02731 + 0.31430i \\ 0.02731 - 0.31430i \\ -0.16595 + 0.00000i \\ -0.00000 + 0.00000i \end{bmatrix} \tag{2}$$

$$V = \begin{bmatrix} A & B \end{bmatrix} \tag{3}$$

$$A = \begin{bmatrix} 0.69946 + 0.00000i & 0.63140 + 0.00000i & 0.63140 - 0.00000i & -0.56600 + 0.00000i \\ 0.38286 + 0.00000i & -0.28715 + 0.15402i & -0.28715 - 0.15402i & -0.26420 + 0.05040i \\ 0.32396 + 0.00000i & -0.07422 - 0.10512i & -0.07422 + 0.10512i & 0.10267 - 0.14787i \\ 0.24297 + 0.00000i & 0.00707 - 0.24933i & 0.00707 + 0.24933i & 0.11643 - 0.02319i \\ 0.41231 + 0.00000i & -0.28417 + 0.44976i & -0.28417 - 0.44976i & 0.49468 + 0.14385i \\ 0.10308 + 0.00000i & 0.22951 - 0.13211i & 0.22951 + 0.13211i & 0.14749 - 0.38066i \\ 0.13989 + 0.00000i & -0.22243 - 0.11722i & -0.22243 + 0.11722i & -0.03106 + 0.35747i \end{bmatrix}$$

$$B = \begin{bmatrix} -0.56600 - 0.00000i & 0.32958 + 0.00000i & -0.00000 + 0.00000i \\ -0.26420 - 0.05040i & -0.14584 + 0.00000i & -0.40825 + 0.00000i \\ 0.10267 + 0.14787i & -0.24608 + 0.00000i & 0.00000 + 0.00000i \\ 0.11643 + 0.02319i & 0.24398 + 0.00000i & 0.00000 + 0.00000i \\ 0.49468 - 0.14385i & -0.42562 + 0.00000i & 0.00000 + 0.00000i \\ 0.14749 + 0.38066i & 0.64118 + 0.00000i & 0.81650 + 0.00000i \\ -0.03106 - 0.35747i & -0.39720 + 0.00000i & -0.40825 + 0.00000i \end{bmatrix}$$

ここで以下の式を導入する。
$x_0 = c_1 * v_1 + c_2 * v_2 + ... + c_n * v_n$ より、
$k\ s.t\ k \to \infty$ について $M^k * x_0 = M^k * c_1 * v_1 + M^k * c_2 * v_2 + ... + M^k * c_n * v_n$ であり、
$M^k * v_i = \lambda_i^k * v_i$ であることを重ねて

$$\begin{aligned} M^k * x_0 &= \lambda_1^k * c_1 * v_1 + \lambda_2^k * c_2 * v_2 + ... + \lambda_n^k * c_n * v_n \\ &= \lambda_1^k * (c_1 * v_1 + (\lambda_2/\lambda_1)^k * c_2 * v_2 + ... + (\lambda_n/\lambda_1)^k * c_n * v_n) \end{aligned} \tag{4}$$

であることがわかる。
ここで、 λ_1 が最大固有値であり、 $\lambda_1 = 1$ あるとき、 $(\lambda_i/\lambda_1)^k \to 0$ と近似できるため、これにより

$$\begin{aligned} M^k * x_0 &= 1^k * c_1 * v_1 \\ &= c_1 * v_1 \end{aligned} \tag{5}$$

という定常状態に至った。 $\|v_1\|_1 = 1$ となるようにしたとき、これが PageRank となる。
即ち

$$v_1 = \begin{bmatrix} 0.303514 \\ 0.166134 \\ 0.140575 \\ 0.105431 \\ 0.178914 \\ 0.044728 \\ 0.060703 \end{bmatrix} \tag{6}$$

これがこの場合の PageRank となる。

2 ハウスドルフのコンパクト図形の空間について下記の問いに答えなさい

2.1 「縮小写像」の定義と、「縮小者像の原理（不動点定理）」を記述し証明しなさい

完備な距離空間 (X, d) ($X \subseteq R^n$) について写像 $f : X \to X$ が存在して、ある μ ($0 \leq \mu < 1$) について、

$$d(f(x), f(y)) \leq \mu d(x, y) \quad (x, y \in X) \tag{7}$$

であるとき、この写像 f を縮小写像という。
そしてこの空間上に必ず不動点 p が存在することを縮小写像の原理という。
ここで不動点とは f(p) = p を満たす p を示す。
x_0 ($x0 \in X$) に $x_{n+1} = f(x_n) \quad n = 1, 2 \cdots$ を定めることで、点列 { x_n } を置く。このとき縮小写像の性質から、

$$\begin{aligned} d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq \mu d(x_n, x_{n-1}) &= \mu d(f(x_{n-1}), f(x_{n-2})) \\ &\leq \mu^2 d(x_{n-1}, x_{n-2}) \leq \cdots \leq k^n d(x_1, x_0) \end{aligned}$$

これに三角不等式を用いて、

$$\begin{aligned} d(x_m, x_n) \leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + d(x_{m-2}, x_{m-2}) + \cdots + d(x_{n+1}, x_n) \\ \leq (\mu^{m-1} + \mu^{m-2} + \cdots + \mu^n) * d(x_1, x_0) \leq \mu^n / 1 - \mu * d(x_1, x_0) \end{aligned} \tag{8}$$

ここで、 $\mu^n \to 0$ であることから、 $d(x_m, x_n) = 0 \quad (m, n \to \infty)$ であるとわかった。つまり x_n はコーシー列であり、完備な空間上の点列であることから、ある α に収束する。
ここで f が縮小写像であるのだから、

$$\alpha = \lim_{n \to \infty} x_{n+1} = \lim_{n \to \infty} f(x_n) = f(\lim_{n \to \infty} x_n) = f(\alpha) \tag{9}$$

であり、この点が不動点であることを示している。
次に、 $f(\beta) = \beta$ となるような β の存在を仮定して、

$$d(\alpha, \beta) = d(f(\alpha), f(\beta)) \leq \mu d(\alpha, \beta) \tag{10}$$

であるのだから、

$$(1-\mu)*d(a,b) \leq 0 \tag{11}$$

$1-\mu > 0$ より $d(\alpha,\beta)=0$
距離関数において、これは $\alpha=\beta$ を意味している。

2.2 ハウスドルフの距離の定義を述べ、それが距離であることを示しなさい

ある2つの空間 A, B の距離を考えたときに、A から δ_A だけ広げた空間上に B がちょうど全て含まれるとき、これを A から B への距離として、同様に B から A の距離を δ_B とすると、A, B に対するハウスドルフ空間の距離 $d_H(A,B)$ は、

$$d_H(A,B) = max(\delta_A, \delta_B) \tag{12}$$

とする。
これが距離であることを示すためには、1) 正値性、2) 対称性、3) 最短性　を示す必要がある。

1. 正値性に関しては4つの場合を考えることで証明できる。

   - A と B に包含関係がない場合
     δ_A, δ_B が正の値であることは明らかであるため、正値性は成り立つ。

   - A の一部と B の一部が重なっている場合
     δ_A, δ_B が正の値であることは明らかであるため、正値性は成り立つ。

   - A が B を含んでいる場合
     δ_A が負である可能性はあるが、 δ_B は正であることが明らかであるため、正値性は成り立つ。

   - A と B が等しい場合
     $\delta_A=0, \delta_B=0$ は明らかであるため、正値性は成り立つ。

2. 対称性に関しては、 $max(\delta_A, \delta_B) = max(\delta_B, \delta_A)$ であることから成立する。

3. 最短性に関しては、以下に詳しく示していくことにする。

二点間の距離を $dist(x,y)$ とする。その上で点 x と 図 A の隔たりを $d(x,A)$ とする。つまり、 $d(x,A) = \min_{a\in A} dist(x,A)$ と表すことができる。

ここで、 δ_A を $d(b,A)$ が最大となるものに定める。$(b \in B)$。
つまり、 $\delta_A = \max_{b\in B} d(b,A) = \max_{b\in B} \min_{a\in A} dist(b,a)$
同様にして $\delta_B = \max_{a\in A} d(a,B) = \max_{a\in A} \min_{b\in B} dist(a,b)$ を定める。
式からわかるように、 δ_A と δ_B は同じものを示しているわけではない。
わかりやすくするため、 $\delta_A = D(B,A), \delta_B = D(A,B)$ として以下の三角不等式を示す。

$$\begin{aligned} D(A,B) &\leq D(A,C) + D(C,B) \\ D(B,A) &\leq D(B,C) + D(C,A) \end{aligned} \tag{13}$$

$a \in A$ を固定して、二点間の距離に関する三角不等式を用いて、

$$\begin{aligned} d(a,B) &= \min_{b\in B} dist(a,b) \\ &\leq \min_{b\in B}(dist(a,c) + dist(c,b)) \\ &= dist(a,c) + \min_{b\in B} dist(c,b) \end{aligned} \tag{14}$$

両辺を $c \in C$ で最小化すると、

$$\begin{aligned} d(a,B) &= \min_{c\in C} dist(a,c) + \min_{c\in C} \min_{b\in B} dist(b,c) \\ &\leq \min_{c\in C} dist(a,c) + \max_{c\in C} \min_{b\in B} dist(b,c) \\ &= \min_{c\in C} dist(a,c) + D(C,B) \end{aligned} \tag{15}$$

この式は、 $\forall a \in A$ で成り立つので $a \in A$ で最大化することで、

$$\begin{aligned} \max_{a\in A} d(a,B) &= \max_{a\in A} \min_{c\in C} dist(a,c) + D(C,B) \\ &\Updownarrow \\ D(A,B) &= D(A,C) + D(C,B) \end{aligned} \tag{16}$$

同様に、 $D(B,A) = D(B,C) + D(C,A)$ が求まり、結果としてハウスドルフの距離の最短性を示すことができた。
よって、ハウスドルフの距離は距離として成立しているといえる。

2.3 ハウスドルフの空間上の２つの縮小写像 F1 F2 があるとき、 $\Phi = F1 \cup F2$ もまた縮小写像となることを示しなさい

まず縮小写像の和についての定義を行う。
あるハウスドルフ空間 C について $(F1 \cup F2)(C)$ を行うとき、これは $F1(C) \cup F2(C)$ と

等価である。さて、 $F1 \cup F2$ が縮小写像であることを示すために

$$d_H(A \cup B, C \cup D) \leq max\{d_H(A,C), d_H(B,D)\} \tag{17}$$

を示していく。

まず、式 1 $D(X \cup Y, Z) = max\{D(X,Z), D(Y,Z)\}$ を示したい。

この証明のためには、 $d(x,Z)$ を最大化できるように $x \in X$ を、 $d(y,Z)$ を最大化できるように $y \in Y$ を考えると、つまり Z との距離が最も離れている X 内の点と Y 内の点がわかり、つまり以下の等式が成立する。

$D(X \cup Y, Z) = max\{d(x,Z), d(y,Z)\} = max\{D(X,Z), D(Y,Z)\}$

これによって式 1 は示された。

次に、式 2 $D(X, Y \cup Z) \leq min\{D(X,Y), D(X,Z)\}$ を示したい。

どのような点 $x \in X$ でも、 $d(x, Y \cup Z) \leq d(x,Y)$ と $d(x, Y \cup Z) \leq d(x,Z)$ であることは明らかであるので、以下の式が成立する。(以下の式での $x \in X$ は、$d(x, Y \cup Z)$ を最大化する x である)

$$\begin{aligned} D(X, Y \cup Z) \quad &= d(x, Y \cup Z) \\ &= min\{d(x,Y), d(x,Z)\} \\ &\leq min\{D(X,Y), D(X,Z)\} \end{aligned} \tag{18}$$

これによって式 2 は示された。

以上の事実より、元の式 $d_H(A \cup B, C \cup D) \leq max\{d_H(A,C), d_H(B,D)\}$ を示す。

$$\begin{aligned} d_H(A \cup B, C \cup D) \quad &= max\{D(A \cup B, C \cup D), D(C \cup D, A \cup B)\} \\ &= max\{D(A, C \cup D), D(B, C \cup D), \\ &\qquad D(C, A \cup B), D(D, A \cup B)\} \\ &\leq max\{D(A,C), D(B,D), D(C,A), D(D,B)\} \\ &= max\{d_H(A,C), d_H(B,D)\} \end{aligned} \tag{19}$$

よって元の式は示された。

次に空でないハウスドルフ空間 A, B について $\Phi = F1 \cup F2$ を適用すると

$$\begin{aligned} d_H(\Phi(A), \Phi(B)) \quad &= d_H(F1(A) \cup F2(A), F1(B), \cup F2(B)) \\ &\leq max\{d_H(F1(A), F1(B)), d_H(F2(A), F2(B))\} \\ &\leq max\{\mu_1 d_H(A,B), \mu_2 d_H(A,B)\} \\ &= max\{\mu_1, \mu_2\} d_H(A,B) \end{aligned} \tag{20}$$

ここで、μ_1 、μ_2 は

$$\begin{aligned} d(F1(x), F1(y)) \leq \mu_1 d(x,y) \quad (x, y \in X) \\ d(F2(x), F2(y)) \leq \mu_2 d(x,y) \quad (x, y \in X) \end{aligned} \tag{21}$$

によって定められるものとする。

すると F1,F2 が縮小写像であることから、 $0 \leq \mu_1, \mu_2 < 1$ なので、上式から Φ も縮小写像である。

3 参考文献:

Invitation to Dynamical Systems (著者: Edward R. Scheinerman)