中学生のための玉原数学教室　　東京大学
玉原国際セミナーハウス
講義 1 (10:30-11:40)　三竹　大寿
2018 年 10 月 13 日

「連立方程式とページランク」

1. 序

おそらく，ここにいる全ての方がインターネットを利用して，調べたい単語を検索した経験があると思います．例えば，今，スマートフォンで Google の検索エンジンを利用して,「東京大学，数理科学，玉原」というキーワードを入れたらどのようなウェブページが検索されますか？沢山のウェブページが現れると思いますが，みなさんはどのページをまず初めに開くでしょうか？(注意：Google アカウントを持たれている方は，ログアウトしてから検索してみましょう.)

実際，私が上記した単語を Google を利用して検索してみると，約 9620 件ものウェブページが検索され,

1 番目：http://tambara.ms.u-tokyo.ac.jp
2 番目：http://www.ms.u-tokyo.ac.jp/tambara/index.html
3 番目：https://www.ms.u-tokyo.ac.jp/video/lecture/tanbara-open/index.html

といった順番でウェブページのハイパーリンクが出てきました．私は，まずこの一番目のページを始めにクリックしてみました．おそらく，多くの方も同じことをしたのではないでしょうか？ここで質問です．どうして，一番目に来たページをまず開きましたか？

現在，世界中には数百億（もっとあるかも？）ものウェブページあると言われています．その中から，どのようにこのようなウェブページが検索され，どのような基準で順番に並んでいるか，皆さんは考えたことありますか？検索する側としてみたら，自分が検索した単語に対して，もっとも「重要な」または「信頼できる」ウェブページが順番に出てきて欲しいですね.

今回の講義では，特に，莫大なウェブページのデータに対して,「重要度」を決めるための数学の一つの概念で,「ページランク」という考え方について紹介します．この考え方は，1999 年に Sergey Brin と Larry Page（Google の共同創業者）らによって導入されたもので，例えば，Google の検索結果の表示順序の決定方法などに利用されています．この考え方（処理方法）は Google の商標として登録され，特許を取得されています．余談になりますが，ページランクの名前の由来は，ウェブ “ページ”と Larry “Page”をかけたものだそうです.

2. 発想

以下では,「ウェブページ」を簡単に「ページ」と呼びます．まず,「重要なページ」とは何かということを, 次の２つの考え方で定義します.

(A) 多くのページからハイパーリンク（以下では，リンクと呼びます）されているページは重要なページである.
(B) 重要なページからリンクされているページは重要なページである.

ポイント. (A) は分かりやすいですね．しかし，これだけで十分でしょうか？また，(B) はどうでしょう？一見，トートロジー（a ならば a という論理）のように聞こえますね．(B) の本意は,「よりリンク先を厳選しているページは重要である “=”リンクを乱発しているようなページは重要でない」としたいということにあります.

3. 問題のモデル化（数学の言葉で定式化）

ここでは，節 2 で述べた発想を，数学を用いて定式化してみましょう．まずは，インターネット上のページを数学の言葉を利用して表現してみましょう.

(1) 各ページを, それぞれ 1 つの「点」と考える.
(2) ある点 A で表されたページから, 他の点 B で表されたページへのリンクが存在する場合, A から B への矢印を与える.
(3) ２の操作をすべてのリンクに対して行う.

このような,「点」と「(有向な) 辺」の集合を考える数学を「グラフ理論」と呼びます．このグラフ理論上の言葉を利用して，ページランクを定義してみましょう.

その前に，グラフ理論における最低限の言葉，記号を定義しましょう.

定義 1. 頂点の集合 V と，２つの頂点を結ぶ辺の集合 E の組 $G = (V, E)$ をグラフと呼ぶ. 特に, 各辺 $e \in E$ に対して, 向きを与えたグラフを有向グラフと呼ぶ. 有向グラフの各辺 $e \in E$ に対して, その始点を $o(e) \in V$, 終点を $t(e)$ と書く．ここで，$v \in V$ に対して,

$$
\begin{aligned}
V_v &:= \{u \in V \mid o(e) = v, t(e) = u \text{ となる } e \in E \text{ が存在する}\} \\
&= \{\text{ページ } v \text{ からのリンク先のページの集合}\}, \\
|v| &:= V_v \text{の要素の数}, \\
B_v &:= \{u \in V \mid t(e) = v, o(e) = u \text{ となる } e \in E \text{ が存在する}\} \\
&= \{\text{ページ } v \text{ へリンクしているページの集合}\}.
\end{aligned}
$$

と置く.

定義 2 (ページランク)**.** 考えているウェブデータの有向グラフを $G = (V, E)$ として, $V = \{v_1, \ldots, v_n\}$ とおく．このとき, ウェブページ v_i のページランク $r(v_i)$ を

$$
r(v_i) := \sum_{v_j \in B_{v_i}} \frac{r(v_j)}{|v_j|} = B_{v_i} \text{の全ての元 } v_j \text{に対する} \frac{r(v_j)}{|v_j|} \text{の和}
$$

と定義する.

定義 2 において，各 v_i は, 一つのページを表しており，$V = \{v_1, \ldots, v_n\}$ を考えることは n 個のウェブページを考えることに対応しています.

4. ページランクの計算例

授業中に色々と考えましょう．手を動かして計算してみることが大切です.

5. 最後に

さて，超特急でページランクの基本中の基本について解説をしてきましたが，その雰囲気を掴めたでしょうか？実際に手を動かしてみることで，ページランクの定義の意味と求め方が良く分かったと思います．

今日の話に興味を持ってもっと勉強をしたいと思った方に，メッセージです．今日の話では，高々手計算で確かめられる程度の問題を実際に計算することで，ページランクを求めることができました．しかし，実際にはウェブページは数百億以上あると言われているわけですが，それでもページランクは求められるでしょうか？もう少し数学的に言えば，

- 「任意の n 個のウェブページに対して，ページランクは存在するのか？」
- 「存在した場合に，どのように求めるのか？」

といった問題になります．

この問いには，理系に進めばおそらくどの大学でも学ぶ「線形代数」を学ぶことで答えることができます（ただし，しっかりと線形代数を勉強しないと，最後まできちんと答えることは難しいと思います）．身近な話題ですので，色々と他にも素朴な疑問が出てくると思います．そのような素朴な疑問の中に，実は未知の数学の「問い」が隠れてるかもしれません．(実際，いい数学の研究は素朴な疑問から始まることが多いです.)

最後に注意ですが，実際の Google 検索の結果に表示される順序は, ページランクの順序そのものではないようです．ページランクが重要な役割を果たしていることに間違いないですが，他にも様々なことを加味することで，より実用的な検索エンジンとなるそうです．詳しくは，専門書 [1] を読んで下さい．レクチャーノート [2] は，ページランクを簡明に解説されています．必要な線形代数の知識についてもコンパクトにまとまっているので，チャレンジ精神のある方だったら中学生でも読むことは可能かもしれません．本講義の内容も参考にさせて頂きました．

References

1. A.N.Langville, C.D.Meyer, Google PageRank の数理, 共立出版, 2009.
2. 内藤久資, グラフ上のランダムウォークと Google のページランク, 2011 年度数学アゴラ. (http://www.math.nagoya-u.ac.jp/\~naito/lecture/high_school_2011/summer.pdf)