

Google ページランクの数理

相転移 P (@phasetr)*

2014-07-17

1 概要

Google の検索ではその精度を上げるために 100 を越える要素を考えていることが知られている。最近は少しずつトレンドが変わりつつあるが、このうち Google 自身が「ネット上の民主主義」として強調しているページランクについて議論する¹。インターネット上の Web ページをノード (vertex) と思い、ネットワークをグラフとして抽象化して線型代数の問題に叩き落とす。検索結果の即時性・計算機資源の有限性など現実的な制約も多いが、数値計算に関する理論的なバックアップまで含め、知識としては教養レベルで解決できる。何から何まで教養レベルで簡単に解決すると思われても困るが、線型代数が支配する領域の広さと高い攻撃力を実感してほしい。

数学としては Perron-Frobenius の定理を目標に議論を進める。最後に純数学的な問題として、確率論から見ると離散時間の Markov 連鎖とエルゴード問題に関係していること [?, 7 節], 理論物理では Hubbard モデルの強磁性相転移に関係していること [?], Perron-Frobenius の無限次元版が汎関数積分 (経路積分) で酷使されていること [?] など簡単に紹介したい。ニコニコ動画で動画 [?] も作っていて、物理との関係や小ネタもいろいろ説明しているのでそちらも参考にしてほしい。またニコニコ動画の動画は [?] を元としている。

次のような内容・順番で講演する予定だ。多少変わる可能性はあるが参考にしてほしい。

1. ページランクの基本思想
2. 問題の直観的定式化
3. 問題の数学的定式化
4. べき乗法とそれに関する工学的説明
5. ハイパーリンク行列と問題: 具体例とともに
6. Google 行列の導入
7. Perron-Frobenius の定理による解決
8. 数学・物理的な応用

2 対象

対象は限定しない。純粋な予備知識としては教養の線型代数で十分。その辺の数学科学部生と戦える程度で線型代数を武器として扱える力量があれば問題ない。今回具体的に出てくる中で一番難しいのは行列と線型写像の区別だ。今回、本当に完全に数学的な意味での線型写像は出てこないが、その本質に近い形で話を展開する。上記区別がついていないという方もこれで勉強し直すくらいの気持ちで来て頂きたい。

*phasetr@gmail.com, <http://phasetr.com>, <http://phasetr.com/members/formadd/>

¹最近リンクによる SEO の限界という話が出ていて、ページランクの重要性も落ちてきているようだ。ただ、歴史的に Google が重要視していたという事実は変わらない。そこで自然言語処理を使って内容をきちんと見るようにしていこうとしているようだ。これはこれで統計学を背景にしている数学との関係も深い。