最大固有値解法の性能評価

森, 淳史 / MORI, Junji

(出版者 / Publisher)
法政大学大学院理工学・工学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 理工学・工学研究科編 / 法政大学大学院紀要. 理工学・工学研究科編
(巻 / Volume)
55
(開始ページ / Start Page)
1
(終了ページ / End Page)
4
(発行年 / Year)
2014-03-24
(URL)
https://doi.org/10.15002/00010382

法政大学大学院工学研究科紀要　Vol.55(2014 年 3 月)　法政大学

最大固有値解法の性能評価

Performance Evaluation for Maximum eigenvalue solution

森　淳史
Junji Mori
指導教員　李　磊

法政大学大学院工学研究科情報電子工学専攻修士課程

Eigenvalue problem of the matrix is used in a wide range of fields between the analysis of correlation problem, large-scale computer simulation and big data of the vibration system. There are PageRank is a technique used in calculating importance of the Web pages and such popularity as one of its use. Matrix used in PageRank is a non-negative matrix, implements the maximum eigenvalue solution corresponding to it, compared with the conventional method. ***Key Words** : eigenvalue, PageRank, non-negative matrix*

１．はじめに

行列の固有値問題は振動系の解析や大規模コンピュータシミュレーション, ビッグデータにおけるデータ相関関係の解析など幅広い分野で用いられている. その利用法の一つとして Web ページの重要度, 人気度などの算出の際に用いられる手法である PageRank がある．PageRank で用いられる行列は非負行列であり, それに対応した最大固有値解法を実装し, 従来手法と比較した.

２．評価と順位

ネット検索の結果として表示される Web ページの順位は Web ページの宣伝効果に多大な影響を与えるため,Web ページ作成者の関心を集めている.また,対象の一対比較結果から対象の順位を決めたいという場面は多くある.例えば,Web ページ相互のリンクによる参照関係からページの重要度あるいは人気度を算出し,それに応じた順位付けをする場面や,チーム間での対戦の得点と失点が得られた時の全チームの順位付けの場面などであるこれらの一対比較の結果は

$$a_{ij} := \begin{cases} 1 & \text{ページ}j\text{から}i\text{へリンクあり} \\ 0 & \text{ページ}j\text{から}i\text{へリンクなし} \end{cases}$$

で決まる隣接行列 $A := \lfloor a_{ij} \rfloor_{i,j=1,\dots,n}$ や

$$b_{ij} := (\text{チーム}i\text{のチーム}j\text{からの得点}) - (\text{チーム}i\text{のチーム}j\text{への失点})$$

である得失点差行列 $B = \lfloor b_{ij} \rfloor_{i,j=1,\dots,n}$ で与えられる.

また, 対象に対して複数の個人(あるいは複数の基準)によって異なる順序が提案された時, それらを一つの順序にまとめる際には

$$c_{ij} := (i\text{を}j\text{よりも上位とした人数}) - (j\text{を}i\text{よりも上位とした人数})$$

である行列 $C := \lfloor c_{ij} \rfloor_{i,j=1,\dots,n}$ が基準になる.

３．固有ベクトルによる評価の算出

ページ人気度を算出する方法である PageRank は固有ベクトル法である. ページ間のリンクの有無を表現した(2. 1)は通常非常に疎な行列である. PageRank は A の第 i 行をその行和 $\sum_{j \in N} a_{ij}$ で除した行列

$$H := \left[\frac{a_{ij}}{\sum_{k \in N} a_{ik}} \right]_{(i,j) \in N^2}$$

の左主固有ベクトルを人気度として与える. この H も A と同様に非常に疎な行列である. H は行和 1 の確率行列である.

ネットサーファーがリンク先のページへ移動する選択行動は推移確率行列 H のマルコフ連鎖に従うとすると, ネットサーファーが訪れるページの定常確率 p は H の左主固有ベクトルである. PageRank ではページの人気度をこの定常確率 p で与えている.

４． PageRank

PageRank は「多くの良質なページからリンクされているページは,やはり良質である」という再帰的な関係をもとにすべてのページの重要度を判定したものである.PageRank が高くなるポイントは大きく分けると次の 3 つである

・被リンク数
・おすすめ度の高いページからのリンクかどうか
・リンク元ページでのリンク数

５． フロベニウスの定理

第 2 章で例示した固有ベクトル法の入力行列はいずれも非負行列である.非負行列における正の固有ベクトルの存在と一意性に関する性質がフロベニウスの定理として知られている.

フロベニウスの定理とは

1. 非負の正方行列が既約であれば,その絶対値最大の固有値は正で,かつ実数である.
2. 非負の正方行列が既約であれば絶対値最大の固有値は A の固有方程式の単純根である.
3. 非負の正方行列が既約であれば,絶対値最大の固有値に対応する固有ベクトルの成分はすべて正であるかまたはすべて負である.

非負行列の A が既約であるとは,行と列に対して同じ置換をどのように施しても変換後の行列がブロック上三角行列

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

とならないことである.ただし A_{11}, A_{22} は正方行列である. 既約でない行列 A を可約という.

ノルム 1 の正で単一の固有ベクトルをペロンベクトルと呼び,行列 A の固有値に関わらず,非負の固有ベクトル $(= p)$ は, p とそのスカラー倍しかない.

６． 行列の固有値問題

$n \times n$ の行列 A においてあるスカラー量 λ が 0 ベクトルでない n 次元列ベクトル x とともに

$$Ax = \lambda x$$

を満たすとき, λ を A の固有値, x を A の λ に対する固有ベクトルという. 固有値, 固有ベクトルの計算は行列 A によって表されている線形システムなどの性質を明らかにするのに役立つ.

７． べき乗法

べき乗法は次数の大きな行列の最大固有値を求める際に有効な手法である.そのアルゴリズムを以下に示す.

do

$\upsilon = Ax^{(k)}$

$\lambda^{(k)} = \left(x^{(k)}, \upsilon\right)$

$x^{(k+1)} = \dfrac{\upsilon}{\|\upsilon\|_2}$

k^{++}

$while\left(\|\upsilon\|_2^2 - \left|\lambda^{(k)}\right|^2 < \alpha\right)$

END

８． H 行列

定義 $\mathrm{A} = \{a_{ij}\}, A \in R^{n\times n}$ とする.正の対角行列

$$\mathrm{D} = \mathrm{diag}\{d_i | d_i > 0 \quad i = 1,2,\ldots,n\}$$

が存在するとき,この行列に対して

$$|a_{ii}| d_i > \sum_{j\neq i} |a_{ij}| d_j \quad (i = 1,2,\ldots,n)$$

の条件を満たすとき,A を H 行列(一般化優対角行列)という.

９． 手法１

この手法は入力するベクトルが非負行列の場合にのみ可能な手法である.PageRank の問題において用いられる行列は非負行列であるため,この手法が適用可能である.アルゴリズムを以下に示す.

$M^{(0)} = I - D^{-1}A$

$for(k = 0,1,\ldots)$

$for(i = 0,1,\ldots,n-1)$

$for(j = 0,1,\ldots,n-1)$

$if\ (i \neq j) \quad S_i^{(k)} = S_i^{(k)} + \left|m_{i,j}\right|$

$end \quad for$

$end \quad for$

$for(i = 0,1,\ldots,n-1)$

$d_i^{(k)} = \dfrac{S_i^{(k)} + \varepsilon}{\left|m_{i,j}\right| + \varepsilon}$

$end \quad for$

$if\ (d_i^{(k)} \max - d^{(k)}_i \min < \alpha) \quad exit$

$$D^{(k)} = \begin{pmatrix} d_0^{(k)} & & & 0 \\ & d_1^{(k)} & & \\ & & \ddots & \\ 0 & & & d_{n-1}^{(k)} \end{pmatrix}$$

$$M^{(k+1)} = M^{(k)} D^{(k)}$$

$end \quad for$

この手法は本来,行列がH行列かどうかを判別するものであり,その際,対角行列 D のすべての要素が1よりも小さくなった場合にH行列と判定する.そのため,早い段階で切り上げていたが,何度も反復させるとある値に収束する.この値が最大固有値かどうかを確認する.

１０．数値実験

1つ目は参考文献[5]より行列(10.1)

$$A = \begin{pmatrix} 16 & -1 & 1 & 2 \\ 2 & 12 & 1 & -1 \\ 1 & 3 & -24 & 2 \\ 4 & -2 & 1 & 20 \end{pmatrix} \qquad (10.1)$$

をべき乗法を用いて絶対値最大固有値,対応する固有ベクトル,反復回数を求めるこの際,初期ベクトルを,収束判定条件を1.0×10^{-8}とする.

2つ目は参考文献[4]より行列(10.2)

$$A = \begin{pmatrix} 0 & 1 & 1/2 & 0 & 1/4 & 1/2 & 0 \\ 1/5 & 0 & 1/2 & 1/3 & 0 & 0 & 0 \\ 1/5 & 0 & 0 & 1/3 & 1/4 & 0 & 0 \\ 1/5 & 0 & 0 & 0 & 1/4 & 0 & 0 \\ 1/5 & 0 & 0 & 1/3 & 0 & 1/2 & 1 \\ 0 & 0 & 0 & 0 & 1/4 & 0 & 0 \\ 1/5 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad (10.2)$$

をべき乗法と手法 1 を用いて最大固有値を求める.べき乗法の初期ベクトルを$x = (1.0,...,1.0)$とし,収束判定条件を$\alpha = 1.0\times10^{-6}$とする.手法 1 は$\varepsilon = 1.0\times10^{-4}$,収束判定条件を$\alpha = 1.0\times10^{-6}$とする.

3つ目は0から5までの値を要素に持つ500×500のランダム行列にべき乗法と手法 1 を適用する.条件は2つ目と同じとする.

１１．実験結果

行列(10.1)をべき乗法で解いた結果を表11.1,11.2に示す.

表 11.1 行列(10.1)の実装結果 1

	絶対値最大固有値	反復回数
参考文献[5]	−24.148522	120
実装結果	−24.148522	120

表 11.2 行列(10.1)の実装結果 2

固有ベクトル	参考文献[5]	実装結果
1	−0.024477	−0.024477
2	−0.026883	−0.026883
3	0.999105	0.999105
4	−0.021629	−0.021629

参考文献と等しい結果が出たことからプログラムが確からしいことを確認できた.

次に行列(10.2)をべき乗法,手法 1 で解いた結果を表11.3,11.4に示す.

表 11.3 行列(10.2)の実装結果 1

	最大固有値	反復回数	計算時間[s]
べき乗法	0.999872	12	0.000000
手法 1	1.000000	23	0.000000

表 11.4 行列(10.2)の実装結果 2

固有ベクトル	参考文献[4]	べき乗法	手法1の最後のs
1	0.69946	0.69933	2.12449
2	0.38286	0.38295	1.16288
3	0.32396	0.32400	0.98397
4	0.24297	0.24299	0.73740
5	0.41231	0.41239	1.25234
6	0.10308	0.10303	0.31308
7	0.13989	0.13996	0.42489

行列(10.2)の最大固有値は 1 でありべき乗法,手法 1 ともに限りなく解に近づいた.若干手法 1 の方が近い答えが出た.また手法 1 の中に最大固有値に対する固有ベクトルがないか調べてみたがそれらしいものは見当たらなかった.

最後にランダム行列の結果を表 11.5 に示す.

表 11.5 ランダム行列の実装結果

	最大 固有値	反復回数	計算時間 [s]
べき乗法	1252.05	5	0.005
手法 1	1249.56	6	0.026

大規模な行列を解いたとき, 計算時間に明らかな差が見られた.

１２． まとめ

手法 1 は最大固有値を計算できることが確認できた. しかし, 固有値に対応する固有ベクトルも探してみたがこちらは見つからなかった. 固有ベクトルは逆反復法のような近似固有値から近似固有ベクトルを求める手法を組み合わせて求められる可能性があるので今後の課題とする. 反復回数, 計算時間からべき乗法と比較して良い値が出ないことを明らかにした.

１３． 謝辞

本研究を遂行するにあたり,多大なるご指導を頂きました法政大学教授李磊教授に厚くお礼を申し上げます.また,本研究室の大学院生の方々にも感謝いたします.

参考文献

1) 関谷　和之, 山本　芳嗣 “ランキングを求める数理的方法”　電子情報通信学会　Vol.９４,No.９,２０１１

2) 拓植 覚,獅々堀 正幹,北 研二“Non-negative Matrix Factorization を用いた情報検索”情報学基礎６１－１,自然言語処理１４２－１，２００１

3) 訳者,岩野　和生,黒川　利明,黒川　洋 “Google PageRank の数理”　２００９

4) Google の秘密　-　PageRank　徹底解説 http://homepage2.nifty.com/baba_hajime/wais/pagerank.html

5) 皆本晃弥 “C 言語による数値計算入門”サイエンス社