

EigenFactor™ (アイゲンファクター) の数学的基礎

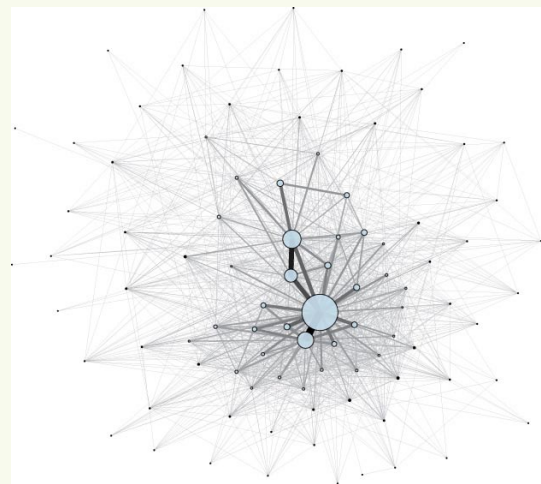
中西印刷株式会社
初版 2009/03/11

概要	1
研究モデル	2
数学的解説	3
技術的論文	4
日本語訳注	4
日本語訳の解説	5
アイゲンファクターとは	別 PDF

このページは EigenFactor.org (<http://www.eigenfactor.org/methods.htm>) のページを掲載者の許可をえて中西印刷株式会社英語チームにより翻訳、転載するものです。アイゲンファクターの実用上の意味については[こちら](#)をごらん下さい。

概要

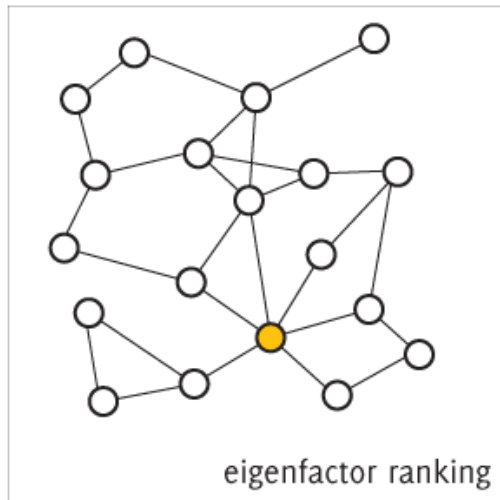
学術文献は文献目録や脚注の相互引用によって学術論文の巨大なネットワークを形成している。このネットワークのシステムは、一人一人の研究者がどの論文が重要であり、どの論文が自分の業績に関連しているかの幾多の判断を反映している。それ故、このネットワーク構造の中には、個々の雑誌 (journal) の相対的な影響力についての情報、さらに、学術の専門分野間の関連性のパターンについての情報という宝が埋もれている。アイゲンファクターの目的はこの学術相互関係や雑誌の相対的影響から情報を引き出す方法を開発することである。



アイゲンファクターはネットワーク理論からの方法を借りて、グーグルのページランクアルゴリズムがウェブページの影響力をランク付けするように、雑誌の影響力をランク付けする。このアプローチでは、他の影響力のある雑誌が、しばしば特定の雑誌を引用すればその雑誌は影響力があると考えられる。残念ながら固有ベクトル中心性 (Eigenvector centrality) の方法として知られるこの反復ランキングの枠組みは「ぶらさがり結節 (dangling nodes)」や「ぶらさがり団塊 (dangling clusters)」に、影響されやすい。これら「ぶらさがり結節」や「ぶらさがり団塊」はネットワークの他の部分に、たとえあるとしても、めったにリンクしない結節 (nodes) あるいは結節の塊のことだ。

これらの問題を打開し、かつ雑誌の引用データにある特定の特殊性にうまく対処するために、基となる固有ベクトル中心性アルゴリズムを修正したのがアイゲンファクターアルゴリズムである。

研究のモデル



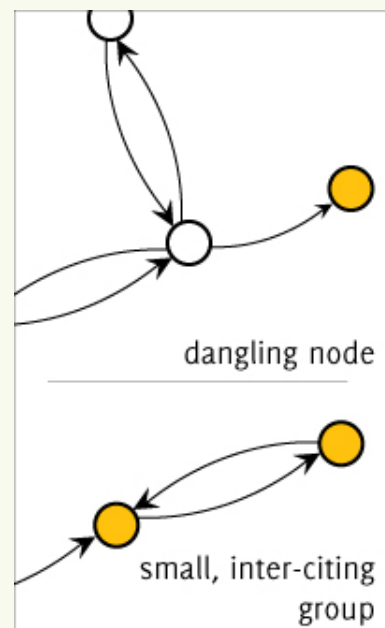
ある雑誌のアイゲンファクターによる数値とは、図書館の利用者がその雑誌に費やす時間の割合の推定値である。アイゲンファクターアルゴリズムは、研究という単純なモデルに対応している。そのモデルの中では読者は一連の引用に従って雑誌から雑誌へと移動する。ある研究者が図書館へ行って、ランダムにある論文を選んだと考えてみよう。その論文を読んだ後、研究者はその論文の引用文献からまたランダムにひとつの論文を選んで読む。そしてまたその論文から引用された雑誌に進み、そこでランダムに選ばれた論文を読み、そしてまた次の雑誌に移るためにある引用文献を選ぶ。研究者は際限なくこ

れを繰り返す。

ここで、それぞれの雑誌に費やされる総時間が、学術引用のネットワーク内にあるその雑誌の重要性の指標と考えられる。更に、実際の研究者が引用文献連鎖によって充分多い数の論文を探し出しているのなら、さきに仮定したランダムに選ぶ研究者がそれぞれの雑誌で費やす総時間は、実際の研究者がそれぞれの雑誌に費やす総時間の推定値となりうる。実際にこの実験を実行することはできないとしても、数学をもちいることで、この方法をシュミレーションすることは可能である。

どれだけ頻繁に雑誌が使われそうかという直接推定した数値を示す以上に、このアプローチは多くの利点がある。先に述べたように、Nature や Cell からの引用が、読者層の狭い三流雑誌からの引用と比較して、高く評価されているように、アイゲンファクターのランキングシステムは、引用した雑誌の威信の違いをも明らかにする。

アイゲンファクターの数値は異なった研究分野の引用パターンの相違も調整している。モデルとした上の仮定研究者の例を考えるとそれがなぜなのかがわかる。雑誌が10冊も他の雑誌を引用しても、100冊引用しても、研究者はリンクの一つだけを辿る。これは規格化された投票方式のようなもので、全投票で持って1回の投票と考えると、10回の投票では1/10の影響力を持つが100回の投票では1/100の影響力しか持たない。いずれにしても、選ぶということは影響力のある一票となりえるのだ。



数学的解説

2006年のアイゲンファクター評価値を算出するには、以下のように行う。トムソンサイエンティフィックのJCRデータセットから、ISIにリンクされた科学及び社会科学の雑誌7,611誌に対し、相互引用行列 Z を抽出する。ここで、行列要素 Z_{ij} は、雑誌*i*に2001年-2005年の間に掲載された論文が、2006年に雑誌*j*に掲載された論文から引用された数である。^{*1}同一雑誌中の自己引用を無視するため、 Z の対角要素 Z_{ij} ($i=j$)はゼロとする。

まず Z を規格化した行列 H を構築する。規格化は列ごとの合計、すなわち各雑誌*j*から引用された総数によって行う。

$$H_{ij} = \frac{Z_{ij}}{\sum_k Z_{kj}}$$

次に、論文ベクトル a を計算する。^{*2}ここに a の要素 a_i は、考慮している5年間にわたり雑誌*i*に掲載された論文の数を、同じ5年間にどれかの雑誌に掲載された論文の総数で割ったものである。行列 H にリストされた雑誌のいくつかは「ぶらさがり結節」、すなわち、自分自身以外を引用しない雑誌であろう。その要素がすべてゼロであるような H の列は、「ぶらさがり結節」である。 H におけるこのようなすべての列をベクトル a で置き換えて、修正された行列 H' を作る。作り方から、これは確率過程行列である。 H' は「研究のモデル」で上述した科学文献上の酔歩散策（ランダムウォーク）に対応する。これから、新しい確率過程行列 P を作る。

$$P = aH' + (1 - \alpha)ae^T$$

ここに、 e^T はその要素がすべて1の行ベクトル（すなわち(1,1,...,1)）である。^{*3}したがって、 ae^T はすべての列が同じ a であるような行列である。^{*4} P のこの項は、確率 $1 - \alpha$ で文献を追跡し、雑誌に掲載された論文の数に比例する重みをもって無作為に選んだ雑誌に移行する過程に対応している。

グーグルのように、 $\alpha = 0.85$ とする。 P において各雑誌で費やされた時間の割合に対応する P の主固有ベクトルによって、ベクトル π^* を定義する。^{*5}これらの割合は、雑誌影響力の重みとして役立つ。

アイゲンファクターの評価値 EF は、次のように定義される。^{*6}

$$EF = 100 \frac{H\pi^*}{\sum_i [H\pi^*]_i}$$

各雑誌*i*に対する論文影響力評価値 AI_i は、雑誌の論文ごとの引用影響力に関する指標である。論文影響力評価値は次のように計算される。

$$AI_i = 0.01 \frac{EF_i}{a_i}$$

ここに、 EF_i は雑誌*i*に対するアイゲンファクター評価値、 a_i は規格化された論文ベクトルの*i*番目の要素である。

技術的論文

これ以上のこの方法に関する詳しい情報は PDF で提供される。Pseudocode も PDF で提供している。また Mathematica による完全なソースコードも PDF として提供する。

Eigenfactor.org によって雑誌をランクづけするのに使われた修正された固有ベクトル中心性のアルゴリズムは学術出版の影響を定量化するために使われてきた反復的方法の 30 年にわたる伝統を拡張したものである。もっとも重要な先駆的業績は以下の文献 [4-9] にある。

1. D. J. de Solla Price	Networks of Scientific Papers Science 169: 510-515 (1965) [PDF]
2. S. Brin and L. Page	The Anatomy of a Large-Scale Hypertextual Web Search Engine WWW7 / Computer Networks 30 (1-7): 107-117 (1998) [PDF]
3. P. Bonacich	Factoring and weighting approaches to clique identification Journal of Mathematical Sociology, 2: 113-120. (1972)
4. G. Pinski and F. Narin	Citation Influence for Journal Aggregates of Scientific Publications: Theory, with Application to the Literature of Physics Information Processing and Management 12: 297-326, 1976
5. S. J. Liebowitz and J. P. Palmer	Assessing the relative impacts of economics journals Journal of Economic Literature 22: 77-88, 1984 [JSTOR]
6. P. Kalaitzidakis and T. Stegnos and T. P. Mamuneas	Rankings of academic journals and institutions in economics Journal of the European Economic Association 1: 1346-1366, 2003
7. I. Palacios-Huerta and O. Volij	The measurement of intellectual influence Econometrica 72: 963-977, 2004 [PDF]
8. Y. K. Kodrzycki and P. Yu	New Approaches To Ranking Economics Journals B. E. Journal of Economic Analysis and Policy, 5(1), Article 24, August 2006 [B.E. Press]
9. J. Bollen and M. A. Rodriguez and H. Van de Sompel	Journal Status Scientometrics, 69(3), 669-687, December 2006 [PDF]

日本語訳注

*1 「行列」は数学の術語で、数または文字（要素という）を方形に配したものである。第 1 添字は行番号、第 2 添字は列番号を表す。ただし、「行」とは横並び列、「列」とは縦並び列である。2 つの行列の「積」が定義される。すなわち行列の積は、第 1（左側）の行列の行の各要素に第 2（右側）の行列の列の各要素をかけて足し合わせたもの（もちろん両者の個数が一致する場合のみ可能）を、今考えた行と今考えた列との交点の位置に配してできる行列のことである。したがって積は、かける順序に依存する。

*2 ベクトルはその要素を縦に並べたものである。したがって、それは列番号が一つしかない行列ともみなせる。

*3 肩の T は転置 (transposed) を表す。すなわち、縦並びを横並びに変えたものが行ベクトルである。

*4 脚注 1 で説明した行列の積の定義から、列ベクトルかける行ベクトルは各要素の積で構成される行列になる。

*5 脚注1で説明した行列の積の定義により、ベクトルに P を左からかけると、ベクトルになる。 $P\pi = \lambda\pi$ となるような、(全要素がゼロのベクトル以外の) ベクトル π を P の固有 (アイゲン) ベクトルという。ただし、 λ は数 (一般に複素数) で、 P の固有値と呼ばれる。 π の全要素を未知数と見ると、固有ベクトルの方程式は、 λ をパラメータとして含む n 元連立斉次1次方程式 (今の例では、 $n = 7611$) である。これが、ゼロベクトル以外の解を持つためには、係数の作る行列式がゼロにならなければならない。これは λ の n 次代数方程式で、その解として固有値が決まる。絶対値が最大の固有値に対応する固有ベクトルを主固有ベクトルという。

*6 脚注1で説明した行列の積の定義により、行列とベクトルの積はベクトルである。この式はベクトルの形で書いてある。

日本語訳の解説

本件アイゲンファクターの構成は、何度読んでもなかなかわかりにくいと思います。以下、推量ですが、理解の参考にはなるとおもいます。

アイゲンファクターは図書館にいつまでも座って、論文の引用文献の中から無作為にその掲載雑誌を次々選んでいく人が費やす時間の長さの比でもって、雑誌の重要度を計ろうという主旨の話です。これを相互引用行列に基づいた確率過程とみなして考えてみます。

相互引用行列をかける (行列の積の意味で) ということは、上の EigenFactor ranking の図の雑誌を表す丸印から別の雑誌を表す隣の丸印に確率的に移行することです。これを繰り返すと、いろいろな移行のパターンが確率的に考えられることとなります。それぞれの丸印が確率的にどれだけ通過されるかで、その雑誌で費やされる相対的な時間の長さを計ることになります。

すべての雑誌が互いに十分よく引用し合っているならば、規格化された相互引用行列 H を次々かけていくと、各雑誌で費やす時間が正しく反映され、重要度が推定できると思われれます。しかし、もし他の雑誌の論文を決して引用をしない雑誌「ぶらさがり結節」があると、いったんそこへ入ったら行列の積の性質上、出られなくなってしまいます。この困難を避けるために、強制的に論文ベクトル a をそこに挿入して、ゲームを続けられるようにするのが、 H を H' に置き換えるということにあたります。しかし、これはあまりにも粗っぽいやり方で、「ぶらさがり結節」とそれに非常に類似の状況との取扱いに差があり過ぎることになります。それを幾分でも緩和するのが、 H' を P で置き換えるという操作です。

さて、任意のベクトルに P を何回でもかけていくと、次第にある特定のベクトルに近づいていきます (力学系の用語で「アトラクター」と呼ばれるもの)。それはどういうベクトルかという、 P をかけても変わらない (ただし規格化して考えているので、大きさは問題にしない) ベクトルのはずです。これを数学で P の固有ベクトルと呼んでいます (訳注5参照)。もちろん、固有ベクトルがすべてアトラクターになるわけではありません。ですが、何らかの条件のもとで、絶対値が最大の固有値は実数で、それに属する固有ベクトル π^* はアトラクターになることがいえるのではないかと推測されます。

最後に、この π^* に本来の H をかけ、規格化したものでもってアイゲンファクターを定義します。もうゲームは終了しているわけですから、「ぶらさがり結節」があっても問題にはならないということになります。