論文 / 著書情報
Article / Book Information

論題(和文)	主成分析を応用した科学的貢献度評価指標 "o-index"
Title(English)	Scientific Evaluation Index “ o-index ” using Principal Component Analysis
著者(和文)	大槻明, 川村雅義
Authors(English)	Akira Otsuki, Masayoshi Kawamura
掲載誌(和文)	Deim2013予稿集
Citation(English)	Proceedings of Deim2013
Vol, no, pages	, ,
発行日 / Pub. date	2013, 3

主成分分析を応用した科学的貢献度評価指標"o-index "

大槻 明† 川村 雅義‡

†東京工業大学〒152-8550 東京都目黒区大岡山 2-12-1
‡東京大学 〒113-8656 東京都文京区弥生 2-11-16
E-mail: †cecil2005@hotmail.co.jp, ‡kawamura.masa@nifty.com

あらまし 従来の代表的な科学的貢献度指数として，*h*-Index，*g*-Index，A-Index 及び R-Index などが存在するが，これらは過去に公刊された論文に基づく評価であるため，経験の長い科学者，または共同科学者が多い科学者ほど値が大きくなる傾向がある．また，優れた若手科学者を発掘する目的には適さない．ゆえに，本研究では，主成分分析の観測値として，研究領域の成長度と被引用論文年度の分散度合を考慮したページランクアルゴリズムによる，科学者の重要度評価値を算出し．そして，これら２つの観測値を基に主成分分析を行うことで新たな合成変数（科学者の科学的貢献度評価指数）を算出する手法について提案する．

キーワード 科学的貢献度指数, Big Data, Data Mining, Citation analysis, Database

Scientific Evaluation Index “o-index” using Principal Component Analysis

Akira OTSUKI† and Masayoshi KAWAMURA‡

†Tokyo Institute of Technology 2-12-1 Ookayama, Meguto-ku, Tokyo, 152-8550 Japan
‡The University of Tokyo 2-11-16 Yayoi, Bunkyo-Ku, Tokyo, 113-8656 Japan
E-mail: †cecil2005@hotmail.co.jp, ‡kawamura.masa@nifty.com

Abstract As typical scientific contribution indexes, such as h-Index, g-Index, A-Index and R-Index, have been conventionally assessed based on literatures published in the past, these values tend to be higher in case of well-experienced scientists or those who have larger number of colleagues. In addition, these are inappropriate for discovering excellent young scientists. Therefore, calculating importance evaluation values of scientists as an observation value of principal component analysis based on PageRank algorithm in consideration of degree of growth in area of study and degree of variance of years of cited literatures, we propose a method for calculating new synthetic variables (scientific contribution estimated index for scientist) by conducting principal component analysis based on these two observation values in the study.

Keyword Scientific Contribution Index,Big Data, Data Mining, Citation analysis, Database

1. はじめに

科学者の科学的貢献度の評価は，研究の内容や研究領域特有の慣習など，様々な要素が絡んでくるため極めて難しい.科学的貢献度を分析する初期の段階では，研究の成果である論文が他の論文に引用された数（被引用数）を規準にすることが一般的であった.しかし，被引用数を基準とした評価では，質の評価を行には限界がある．例えば，50 の引用を受ける論文 1 編と，10 しか引用を受けない論文 5 編が等価になる問題が挙げられる．この問題に対し，Hirsch, J. E [1]は，Web of Science の Times Cited（被引用数）を元に h-Index を考案した.この指標は，当該科学者の論文の量（論文数）と論文の質（被引用数）とを同時に 1 つの数値で表すことができるという利点を持つ．また，Egghe [2]が考案した g-Index は，h-index と似た指数で，上位 g 番目までの論文の被引用数の総和が g^2 以上となる最大の g の値を指標とするものである．その他にも，A-index[3]や R-index[4]などが研究されてきた．これらの指標により，従来の被引用数を基準とした評価における，質の分析が行いにくかったという課題をある程度克服することが可能となった．しかし，上述した Index は，過去に公刊された論文に基づく評価であるため，経験の長い科学者，または共同科学者が多い科学者ほど値が大きくなる傾向がある．また，研究領域の成長度について考慮がなされておらず，例えば，優れた若手科学者を発掘する目的には適さない．

ゆえに，本研究では，「ある科学者が所属するクラスタ（研究領域）の成長度」，及び「被引用論文年度の分散度合を考慮したページランクアルゴリズムによる，科学者の重要度評価」を主成分分析の観測値として算出し，これらの値を基に主成分分析を行うことで新たな合成変数（科学者評価指数）を算出する．この指数の事を o-index（otsuki-Index）呼ぶ．o-index により，現在及び今後成長していくことが期待されるクラスタ

（研究領域）や，また，その中の中心的な役割を果たす科学者の科学的貢献度を評価できる新たな指数が実現できることが期待される．

2. 先行研究

計量書誌学（Bibliometrics）とは，論文などの書誌を構成する要素を計量的に研究する学問である．学問の歴史としては浅く，WILLIAM W.[5]によれば，文献の統計解析(Cole and Eales, 1917) がこの分野の最初の業績とされている．そして，1955 年に Eugene Garfield[6]が自然科学・社会科学分野の学術雑誌を対象として，その雑誌の影響度を測る指標としてインパクトファクター (impact factor)を考案した．また，科学者の科学的貢献度を示す指標としては，これまでに h-Index，g-Index，A-index 及び R-index などが提案されている．2.1 節以降で，これらの科学的貢献度指標について概観する．

2.1 h-index

カリフォルニア大学サンディエゴ校の Jorge Hirsch 博士が 2005 年に提唱した指標である．表 1 に示すとおり，ある科学者の *h*-Index は，その科学者が発表した論文のうち，被引用数が h 以上であるものが h 以上あることを満たすような指標である．

表1. Example of *h*-Index

科学者	科学者の業績， （）内は被引用数	*h*-index
A	論文 A(9)，論文 B(7)，論文 C(5)， 論文 D(4)，論文 E(4)	5
B	論文 A(35)，論文 B(9)，論文 C(5)， 論文 D(3)，論文 E(1)	3

2.2 *g*-index

また，*g*-Index も *h*-index と似た指数で，下記(1)に示すとおり，上位 g 番目までの論文の被引用数の総和が，g^2 以上となる最大の g の値を指標とするものである．

$$g \leq \frac{1}{g}\sum_{i \leq g} c_i \qquad (1)$$

2.3 *A*-index

A-index は，Jin BH が考案した指標で，下記(2)のとおり，*h*-index=*h* としたとき，被引用数（順位）が *h* 番目までの論文の平均被引用数を求めてそれを指標とするものである．

$$A = \frac{1}{h}\sum_{j=1}^{h} cit_j \qquad (2)$$

2.4 *R*-index

R-index も Jin BH が考案した指標で，*h*-index=*h* としたとき，被引用数 h 番目までの論文の被引用総数の平方根を求めてそれを指標とするものである．

$$R = \sqrt{\sum_{j=1}^{h} cit_j} \qquad (3)$$

2.5 従来の科学的貢献度指数の問題

前節までにおいて，代表的な科学的貢献度指数についていくつか概観した．これらの指標により，従来の被引用数を基準とした評価における，質の分析が行いにくかったという課題をある程度克服することが可能となった．しかし，これらの指数は過去に公刊された論文に基づく評価であるため，経験の長い科学者，または共同科学者が多い科学者ほど値が大きくなる傾向がある．具体的には，*h*-Index では，被引用回数の上位の論文は評価に反映されないため，計算式上，例えば 1 位の論文が 10000 回引用されてようが，被引用数 10 番目の論文の被引用回数が 10 であれば *h*-Index は 10 となる．ゆえに，論文発表本数が少ない優れた若手科学者は指数が低く算出される傾向がある．

また，概観した全ての指数共通に言えることは，評価対象の科学者における，当該研究領域の成長度について考慮がなされていないということである．つまり，いくら引用数が多い論文であったとしても，分析行う現時点において，その研究領域が衰退しているような場合には，過去には科学的貢献度の高い研究領域であったかもしれないが，現時点ではその役目を終えている可能性が高いと考えられる．ゆえに，科学的貢献度を計るうえでは，その研究領域が成長しているのか．それとも既に衰退しているのか，と言った点も考慮することが質的な評価という側面では極めて重要であると考えるが，従来の指標ではこの点を考慮していない．

以上から，科学的貢献度の質的評価という側面から，多分に改善の余地はあると考える．ゆえに，本研究では，この質的評価の向上を目指し，次章に本研究のコンセプトを提案する．

3. 提案手法

前章の課題を解決するために，本研究では，下記の①～②を主成分分析の観測値として算出し，これらの値を基に主成分分析を行い，新たな合成変数（科学者評価指数）を算出する．この指数の事を *o*-index（otsuki-Index）呼ぶ．なお，本研究の主な分析対象は，学術論文や特許である．

① ある科学者が所属するクラスタ（研究領域）の成長度
② 被引用論文年度の分散度合を考慮したページランクアルゴリズムによる，科学者の重要

度評価値

上記①は，同じ引用数でも，その研究領域が成長傾向にあるのか，それとも衰退傾向にあるのか，といったクラスタ（研究領域）の成長度を測るためのものである．このように，研究領域の成長度を考慮していることから，現在及び将来にわたり発展していく研究領域かどうかを評価するこができる．また，②は，ある科学者の被引用文献の年度の分散値を算出し，その値を PageRank アルゴリズム[7]に応用することにより，その科学者自体の重要度の評価を行うものである．具体的には，ある科学者の被引用文献の年度の分散値を算出し，その値を PageRank アルゴリズムに応用することにより,その科学者自体の重要度を評価している．インパクトファクターは，引用数の合計値で評価しているが，引用件数が同じでも,「一時期に大量に引用された」場合や,「長期間少しずつ引用されている」場合などが考えられるため，従来の引用分析だけでは，それぞれの重要度を計算する事が難しい．ゆえに，本研究では，上述のそれぞれの「場合」に対し，論文をノード，引用をエッジとする有向グラフと考え，各ノードに発表年数を持たせたうえで，あるノードに入るエッジの元ノードの発表年数の分散を調べることでそれぞれの重要度の計算を試みる．

以上の２つの観測値を基に主成分分析を行うことにより，研究領域の評価など，従来の指数よりもより質的な評価が行える新たな科学者の科学的貢献度指数の実現が期待される．

3.2 クラスタ成長度の算出

o-index の１つ目の観測値として，クラスタ成長度を算出する．本研究で取り扱うクラスタは，ランダムネットワークを礎にしている．ランダムネットワークとは，1960 年頃に Paul Erdös 及び Alféd Rényi [8-10]が提唱した各ノード間にエッジがランダムに存在するネットワークのことである．ランダムネットワークを作成する手順について述べる．総ノード数を N，各エッジが存在する確率を p とする．まず初めにノードを N 個用意する．この場合，最大で下記(2)の本数のエッジが存在しうる．

$$ {}_{N}C_{2} = \frac{N(N-1)}{2} \tag{2}$$

それらのエッジを等しい確率 p でつくる．このようにして作った，確率 p のランダムネットワークには平均で下記(3)の本数のエッジがある．

$$ p\,{}_{N}C_{2} = p\frac{N(N-1)}{2} \tag{3}$$

また，ノード 1 個あたりの平均のエッジ数<k>は下記(4)のとおりとなる．

$$ \langle k \rangle = pN \tag{4}$$

このように N と p を与える事によりランダムネットワークを実現する．ランダムネットワークでは，ノード間の全エッジは等確率で存在しており，基本的にクラスタ化していない．ランダムネットワークにおいては，ネットワーク中の任意の 2 つのノードがリンクされる確率は p なので，ランダムネットワークでのクラスタ係数は下記(5)のとおりとなる．本研究では，このクラスタ係数の成長度を主成分分析の観測値（*ClusterGrowth*）として使用する．

$$ C_{rand} = p = \frac{\langle k \rangle}{N} \tag{5}$$

3.3 ページランクアルゴリズムを応用した科学者の重み付け値の算出

o-index の２つ目の観測値として，科学者の重要度の重み付けを算出する．筆者ら[11]は，Newman 法などにより同定（クラスタリング）された各領域（クラスタ）から，主要論文（ノード）をダイナミックに抽出する手法を提案した．具体的には，論文をノード，引用をエッジとする有向グラフと考え，各ノードに発表年数を持たせたうえで，あるノードに入るエッジの元ノードの発表年数の分散を調べることでそれぞれの分散値（Variance）を算出する．そして，この Variance をページランクアルゴリズムに適応する．Variance を適応しないページランクアルゴリズムでは，各論文に「流れ込む」引用の得点の総和と，各論文から「流れ出す」引用の得点の総和が等しくなるようにして，その総和をその論文の得点と考え，この得点が高いほどその論文は重要であると考える．しかし,筆者らの手法では,この各論文に「流れ出す」引用の得点計算に Variance の値を適応することにより,従来のアルゴリズムでは,「流れ出す」引用が複数個あった場合，得点は均等に割り振られていたが，Variance の値が高いものにより多く「流れ出す」と考え計算する．これにより，被引用年次の分散度を反映した形で主要論文の重要度を算出できる．本研究では，この重要度を主成分分析の観測値の一つ（*PagerankOtsuki*）として使用する．

3.4 主成分分析を応用した科学的貢献度評価指標”o-index”

前述の２つの観測値を基に主成分分析を行い，科学的貢献度指数を算出する．この指数の事を *o*-index と呼ぶ．主成分分析は，潜在変数を従属変数として２変量以上の独立変数でそれを説明しようとする回帰分析モデルである．つまり，主成分分析は２つ以上の変数から,新たな１つの合成変数を作り出す手法といえる．*o*-index では，先ず，前述の２つの観測値を年度毎に算出し，表.1 に示すようなデータフレームを用意する．

表.1　観測値のデータフレームの例

No	ClusterGrowth	PagerankOtsuki
1	1319	0.0226901155865
2	1205	0.023357562845
3	1057	0.0252106135308
4	914	0.0272025998314
5	727	0.0274367637258

そして表.1 の全データを観測値として主成分分析を行う．合成変数を説明する各変数の重み（weight）のことを，主成分負荷量というが，これは回帰分析における偏回帰係数に当たるものである．これにしたがって,本研究における合成変数は次のように表される．

o-index = (n * ClusterGrowth) + (n * PagerankOtsuki)

n は整数を表す．また，観測値間における格差をなくすために，２つの観測値はそれぞれ標準化したうえで使用する．観測値の *o-index* への効果について説明する．例えば，ClusterGrowth 及び PagerankOtsuki がマイナスだった場合は，*o-index* に対して負の効果をもたらす．逆に，変数がプラスだった場合は，*o-index* に対して正の効果をもたらす．なお，これら２つの変数の効果量はどれも同じ程度に設定する．このようにして *o-index* を算出することにより，「クラスタ成長度」及び「ページランクアルゴリズムを応用した科学者の重み付け」を考慮した，新たな科学的貢献度指数の実現を目指す．

4 評価実験

4.1 評価実験の概要

h-index，*g*-index，*A*-index，*R*-index と *o*-index を比較検証することにより *o*-index の有用性について考察する．まず，論文データベース Scopus から，2012 年度までの論文を検索対象として，さらに，”Data Mining” をクエリとして対象論文を取得した．取得した対象論文の合計は 31,501 件である．これらの中で引用数上位 5 論文を表.2 に示す．次に，評価実験の対象科学者を選定する．表.3～表.6 は，表.2 の４人の著者が第１著者であり，さらに，論文キーワードがデータマイニングである論文をリスト化したものである.さらに,表.3～表.6 は，それぞれ *h*-index，*g*-index，*A*-index，*R*-index と *o*-index を算出した結果をまとめたものである．なお，o-index の算出は 4.2 節に述べるとおり行った．

表.2　引用数上位 5 論文（論文 DB：Scopus，クエリ：Data Mining）

	引用数	著者	論文名	出版年
1	3649	Agrawal,R., et.al	Mining association rules between sets of items in large databases	2009
2	1594	Fawcett,T., et.al	An introduction to ROC analysis	2006
3	1340	Zimmermann, P. et.al	GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox	2004
4	1305	Agrawal, R., et.al	Mining sequential patterns	2005
5	749	Foster, I. et.al	Grid services for distributed system integration	2002

表.3　Agrawal,R のデータマイニング系論文における各種 index 比較

o-index	h-index	g-index		A-index	R-index	Number of Citation	Name of the Papers
1.353000		1(1)	3649			3649	Mining association rules between sets of items in large databases, 1993.
1.3614909		2(4)	4954			1305	Mining sequential patterns, 1995.
1.3633497		3(9)	5688			734	Privacy-preserving data mining, 2000.
1.3735004		4(16)	6328			640	Automatic subspace clustering of high dimensional data for data mining applications, 1998.
1.2927639		5(25)	6813			485	Database mining: A performance

							perspective, 1993.
1.3522393		6(36)	7211			398	Parallel mining of association rules, 1996.
1.3732959		7(49)	7282			71	Automatic subspace clustering of high dimensional data, 2005.
1.2776390	**8**	**8**(64)	7313	**914.125**	**85.52**	31	Securing electronic health records without impeding the flow of information, 2007.
1.343410							

表.4　Fawcett,T.のデータマイニング系論文における各種 index 比較

o-index	h-index	g-index		A-index	R-index	Number of Citation	Name of the Papers
1.2445739		1(1)	1613			1613	An introduction to ROC analysis, 2006.
1.3730406		2(4)	2062			449	Adaptive fraud detection, 1995.
1.2475221		3(9)	2112			50	Using rule sets to maximize ROC performance, 2001.
1.1460667	**4**	**4**(16)	2125	**531.25**	**46.10**	13	PRIE: A system for generating rulelists to maximize ROC performance, 2008.
1.252801							

表.5　Zimmermann, P.のデータマイニング系論文における各種 index 比較

o-index	h-index	g-index		A-index	R-index	Number of Citation	Name of the Papers
1.4045851	**1**	**1**(1)	1226	**1226**	**35.01**	1226	GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox

表.6　Foster, I.のデータマイニング系論文における各種 index 比較

o-index	h-index	g-index		A-index	R-index	Number of Citation	Name of the Papers
1.4032871		1(1)	749			749	Grid services for distributed system integration, 2002.
1.4000199	**2**	**2**(4)	782	**391.00**	**27.96**	33	Data integration in a bandwidth-rich world, 2003.
1.4016535							

4.2 *o*-index の算出

先ず，表.7 に例示とおり，前述の２つの観測値（ClusterGrowth と PagerankOtsuki）を年度毎に，直近 5 年分(2008～2012)を算出した． 表.7 は一例として Zimmermann, P.の観測値データフレームを示してものであるが，このようなデータフレームを対象の全論文について算出した.

表.7　Zimmermann, P.の観測値データフレーム（ClusterGrowth と PagerankOtsuki）

Year	ClusterGrowth	PagerankOtsuki
2012	1319	0.0226901155865
2011	1205	0.0233575628450
2010	1057	0.0252106135308
2009	914	0.0272025998314
2008	727	0.0274367637258

本研究では主成分分析のツールとして R を使用する．R には主成分分析を行うために prcomp()と princomp()という 2 つの関数が用意されているが，どちらも大きな違いはないため，本研究では princomp()を採用することとした．Zimmermann, P.を一例に，princomp()を使って主成分分析を行った結果を図.1 に示す．引数 cor=TRUE は，原データを標準化して解析することを指定するものである．また，summary()の引数に loadings=TRUE とすることで、主成分負荷量が出力できる.

```
> data <- read.csv("/mydata.csv",head=F)
> data2 <- princomp(data, cor=TRUE)
> data2
Call:
princomp(x = data, cor = TRUE)
Standard deviations:
```

Comp.1	Comp.2
1.4045851	0.1647446

2 variables and 9 observations.

```
> summary(data2, loadings=TRUE)
Importance of components:
```

	Comp.1	Comp.2
Standard deviation	1.4045851	0.16474458
Proportion of Variance	0.9864296	0.01357039
Cumulative Proportion	0.9864296	1.00000000

Loadings:

	Comp.1	Comp.2
V1	-0.707	-0.707
V2	0.707	-0.707

>

図.1　主成分分析を行った結果（Zimmermann, P.）

主成分分析では観測変数の数と同じ数の合成変数が作成されることになるので，適当な数だけ合成変数を採用して，それ以降の合成変数は切り捨てる必要がある．本研究では，その判断基準として標準偏差と累積寄与率を使用した．ここでいう標準偏差とは，合成変数の標準偏差のことであり，主成分分析では，作成される合成変数の分散が最大になるように重みが計算されるので，この値が大きいほど"よい合成変数である"ということがいえる．判断の基準としては，大まかに 1 以上の合成変数を採用し，それ以下の合成変数を切り捨てるというのが一般的である．一方で累積寄与率もいくつの合成変数を採用するかの判断基準として役立つ．簡単にいえば，この累積寄与率とは第 n 番目までの合成変数がデータ全体を説明している割合のことである．慣例的に 80%（0.80）という寄与率があれば，十分にデータ全体を説明していると判断される．したがって，今回の場合は，第 1 番目の合成変数 1.4045851（Comp. 1）だけでも，すでにデータ全体の 98%（0.98）を説明していることになるので，第 2 番目以降の合成変数は切り捨てればよいことになる．この主成分得点を表. 3〜表.6 の各論文毎に算出した．また，表. 3〜表.6 の o-index の最終行は，論文が複数ある場合にその平均値を算出したものである．他の指数との比較にはこの平均値を用いる．

4. 3 考察

表. 3〜表.6 に示すとおり，Zimmermann, P.や Foster, I. は論文数が少なく，逆に，Agrawal,R や Fawcett,T.は論文数が多い．h-index，g-index，A-index 及び R-index は，論文数に比例して指数が上がっていた．特に，Agrawal,R.は研究歴が長く，表. 3 に示すとおり，データマイニング関連の論文だけでも 8 つ存在する．このように，研究歴が長く論文数が多い科学者においては，*h*-index，*g*-index，*A*-index 及び *R*-index の指数は高くなるという特徴を確認することができた．これは，グラフとして表示すると違いが明確になる．図.2〜図.6 に示すとおり，*o*-index 以外は論文数の少ない Zimmermann, P.になるにつれて指数が低くなっている．ちなみに，Zimmermann, P.の A-index は高く算出されているが，A-index は平方根を求める指数であり，また，Zimmermann, P.の対象論文が 1 つであることから，その論文の引用数（1226）がそのまま指数として算出されているだけである．

表. 2 のとおり，Zimmermann, P.は 1226 の引用数を誇る論文を有する有望な科学者であるが，論文数が少ないため，h-index，g-index，A-index 及び R-index では指数が低く算出されていた．対して，o-index は，論文数に大小に影響されることなく指数が算出されていた．以上から，*o*-index は，研究領域の成長度合いや，筆者独自手法により，科学者の重要度を算出し，それらの値を考慮することにより，*h*-index，*g*-index，*A*-index，*R*-index では算出することが困難であった，研究歴が短く論文数が少ない有望な科学者（Zimmermann, P.）についても，より質的な評価が行えるようになったと考えられる．

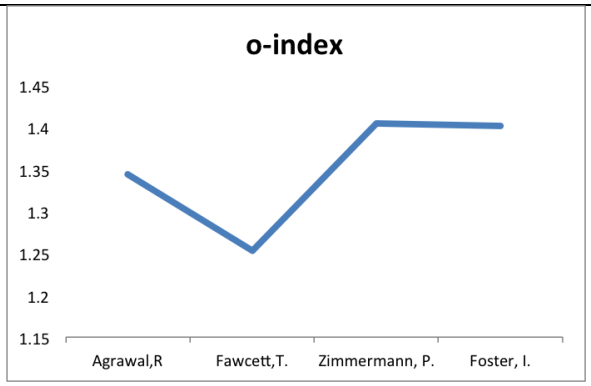

図.2　*o*-index

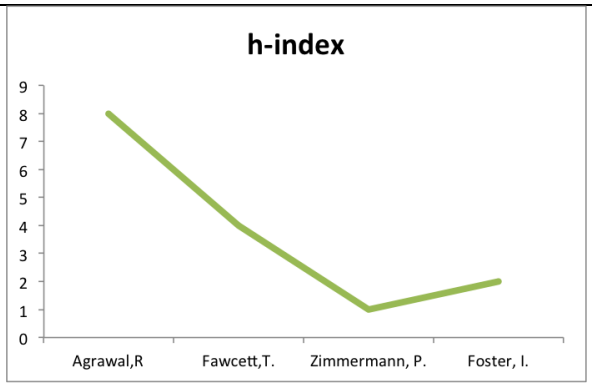

図.3　*h*-index

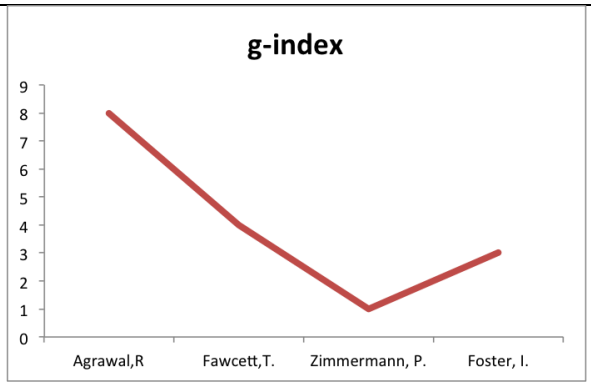

図.4　*g*-index

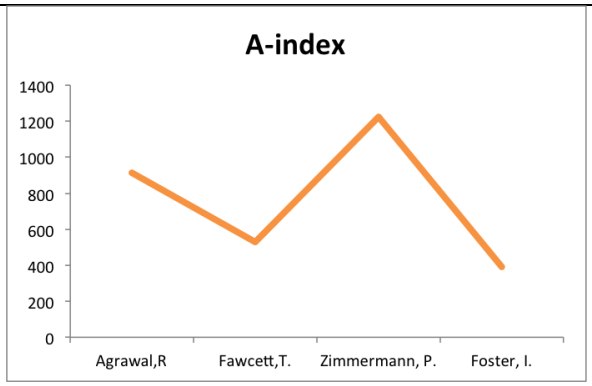

図.5　*g*-index

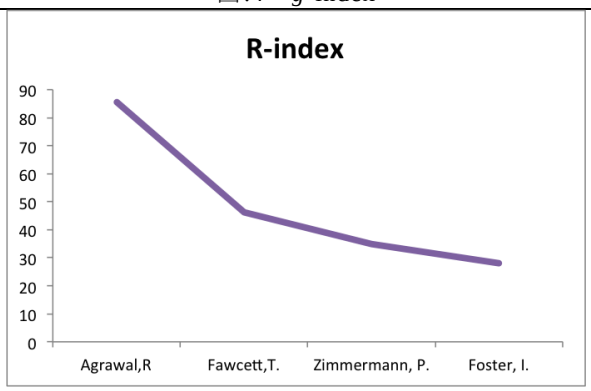

図.6　*g*-index

より質的な評価が行えるようになった原因について，さらに考察する．図.7 は，Agrawal, R.のデータマイニング論文（8 つ）におけるクラスタサイズの成長度推移を比較したものである．また，図.7 では分かり辛いため，図.8 に，「Mining association rules between sets of items in large databases, 1993.（以下，Agrawal, R. 1993:Database)」だけのクラスタサイズの成長度推移を示した．表.2,3 に示したとおり，Agrawal, R.1993:Database は，3649 もの被引用数を誇る論文である．これは，本評価実験で扱う論文の中では 1 番の被引用数である．しかし，図.8 に示したとおり，この論文の研究領域は 2010 年度をピークに衰退傾向にあるため，結果として表.3 に示したとおり，o-index では低いスコアが算出されていた．以上に述べたとおり，o-index は，研究領域の衰退傾向も評価することができるため，論文数に比例して指数が上がっていた h-index, g-index, A-index 及び R-index では行えなかった，質的な評価がより行えるようになったものと考えられる．

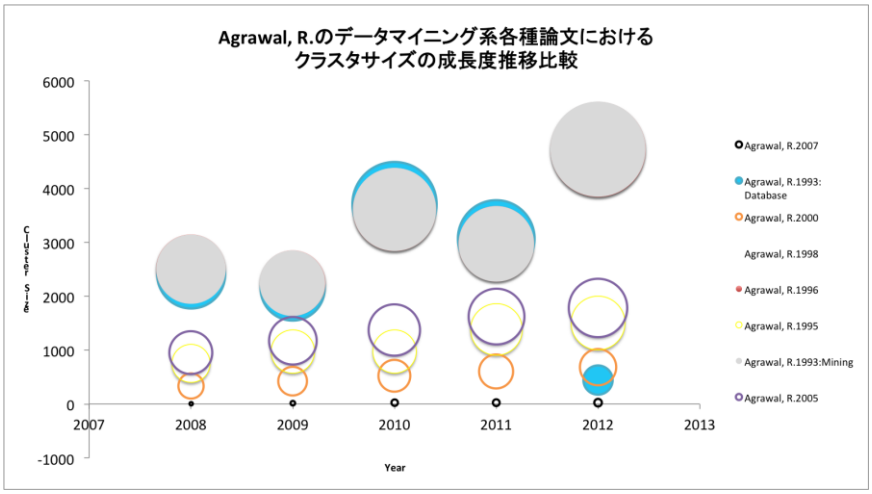

図.7　**Agrawal, R.のデータマイニング系各種論文におけるクラスタサイズの成長度推移比較**

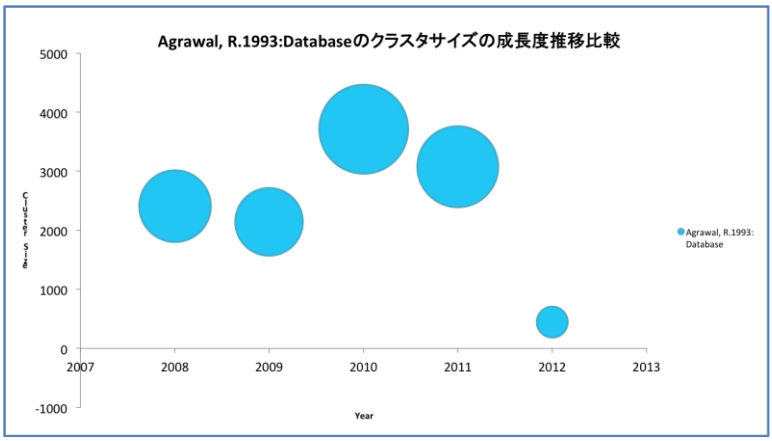

図.8　**Agrawal, R.1993:Database のクラスタサイズの成長度推移比較**

5 むすび

本研究では，従来の代表的な科学的貢献度指数を概観し，それらの指数が経験の長い科学者，または共同科学者が多い科学者ほど値が大きくなるという特徴を指摘した．そして，評価実験においてこの特徴を確認した．本研究では，この課題を解決するために，*o*-Index を提案した．*o*-Index は，研究領域の成長度と被引用論文年度の分散度合を考慮したページランクアルゴリズムによる，科学者の重要度評価値を主成分分析の観測値として算出する．評価実験の結果，*o*-index は，*h*-index，*g*-index，*A*-index，*R*-index では算出することが困難であった，研究歴が短く論文数が少ない有望な科学者についても質的な評価が行えていた．今後は，*o*-index をより効率的に算出できる手法について研究していきたい．

参　考　文　献

[1] Hirsch, J.E. (2007). Does the h index have predictive power?, Proceedings of the National Academy of Sciences of the United States of America 104 (49) , pp. 19193-19198.

[2] Egghe, L. (2006). Theory and practise of the g-index, Scientometrics 69 (1), pp. 131-152.

[3] Jin, B. H. (2006). h-Index: An evaluation indicator proposed by scientist. Science Focus, 1(1), 8–9. (In Chinese)

[4] Jin BH, Liang LM, Rousseau R, Egghe L (2007). The R- and AR-indices: Complementing the h-index. Chinese Science Bulletin 52(6):855-863.

[5] WILLIAM W. HOOD, CONCEPCIÓN S. WILSON (2001). The literature of bibliometrics, scientometrics, and informetrics.

[6] Garfield E. (1955). Citation indexes to science: a new dimension in documen-tation through association of ideas, Science 122(3159): 108-11.

[7] Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd (1998). The Pag-eRank Citation Ranking: Bringing Order to the Web.

[8] Erdös, P., & Rényi, A. (1959). On random graphs, Publicationes Mathematicae Debrecen, 6, 290–297.

[9] Erdös, P., & Rényi, A. (1960). On the evolution of random graphs. Magyar Tud. Akad. Mat. Kut. Int. Kzl., 5, 17–61.

[10] Erdös, P., & Rényi, A. (1961). On the strength of connectedness of a random graph. Acta Math., Acad. Sci. Hungar., 12, 261–267.

[11] Akira Otsuki and Ayumi Kawakami (2012). Academic Landscape Based on Network Analysis Considering Analysis of Variation in the Years of Lucubration Publishing, New Research on Knowledge Management Models and Methods, InTec Open Science, Chapter 17, pp.371-378.