

ビブリオメトリックスを活用した 研究評価の現状と展望

Bibliometrics as a tool for research evaluation

国立情報学研究所
National Institute of Informatics

孫 媛
Sun, Yuan

SPARC Japan 2012年度 第1回

2012年5月25日

研究評価とは

- 研究活動に関わる何らかの意思決定を行うために、評価対象の価値を判断する作業
- 評価対象
 - 国／大学・研究機関／研究者個人
 - 研究プロジェクト／研究に関わる政策
 - 論文／雑誌

何のための研究評価

- 研究評価の本来の目的
 - 優れた研究の探索
 - 研究の学問的質の維持
- 1980年代に始まり, 90年代さらに加速したトレンド
 - 科学政策と連動: 研究優先度の定義・研究資金の配分
 - 典型例:
 - 大学ランキング(機関)
 - 研究費配分(プロジェクト, 個人, 政策)
 - 昇進・給与(個人)

研究評価手法の変化

- 1970年代までのスタンダードな手法：
ピアレビュー（peer review）
 - 評価対象と研究分野を同じくする専門家が判断を下す同業者評価
 - “Process of subjecting an author’s scholarly work, research or ideas to the scrutiny of others who are experts in the same field” (wiki)
 - 科学コミュニティにおいて様々なレベルで行われる
 - 論文の査読, 研究者への学術賞の授与, 資金配分を行うプロジェクトの選定, 人事採用, 昇進...

研究評価手法の変化

- 1980年代以降, 客観的評価指標への要望が高まる
 - 研究活動の拡大と国の研究予算増加の鈍化
 - 評価による選択的資金配分
 - プロジェクト間, 分野間比較の必要性
 - 専門内でのピアレビューには限界
 - 説明責任(accountability)の強調
 - 研究活動への公的資金投入の正当性が問われる
 - 効果は投入資金に見合っているか(パフォーマンス指標)
 - ピアレビューの問題点
 - 評価者の主観性・能力による意識・無意識的バイアス
 - 多大の時間・労力・費用を要する
- ビブリオメトリックス指標による研究評価が増え, ビブリオメトリックスが評価ツールとして定着していく

ビブリオメトリックスとは

- 図書などの情報伝達媒体(communication media)に対する数学と統計的方法の適用 (Alan Pritchard , 1969)
 - 文献データの統計分析手法だが, 科学コミュニケーションの解明・測定にも用いられ, Scientometricsと同義語で使われることが多い(科学文献データの統計分析は, ビブリオメトリックスという言葉の出現以前から)
- 研究対象は情報伝達媒体
 - 学術雑誌論文, 図書, モノグラフ, レポート, 学位論文等。最近では, 特許やウェブデータ等も
- 分析データは書誌情報
 - 雑誌・論文, 著者, 著者の所属機関, 引用文献, 出版年, 主題分野分類等

ビブリオメトリックスの略史

- 概ね50年の歴史
 - ‘bibliometric’という語は1969年に生まれる (それ以前は‘science of science’)
 - 先駆者として, Cole, Lotka, Bradford, Zipf など
- Cole & Eales (1917)
 - “The history of comparative anatomy: Part I. – a statistical analysis of the literature”, *Science Progress*, 11, 578-596.
- Garfield
 - 1960年代に, SCI (Science Citation Index) を創設
 - Library information tool
 - 引用統計の調査が可能となり, ビブリオメトリックスの応用研究の可能性が飛躍的に拡大

ビブリオメトリックスの略史

- Derek de Solla Price
 - Analysis of exponential growth of science (1956)
 - Little Science, Big Science (1963)
 - 科学コミュニケーションの分析を通じて、ビブリオメトリックスの系統的手法を提示。論文数・引用数を用いる現在の研究評価手法の基礎を築く
- Henry Small
 - 共引用分析
- Francis Narin, CHI
 - ビブリオメトリックス指標(論文数・引用数)を、はじめて研究評価・科学政策に適用。ピアレビュー測度との相関を示す

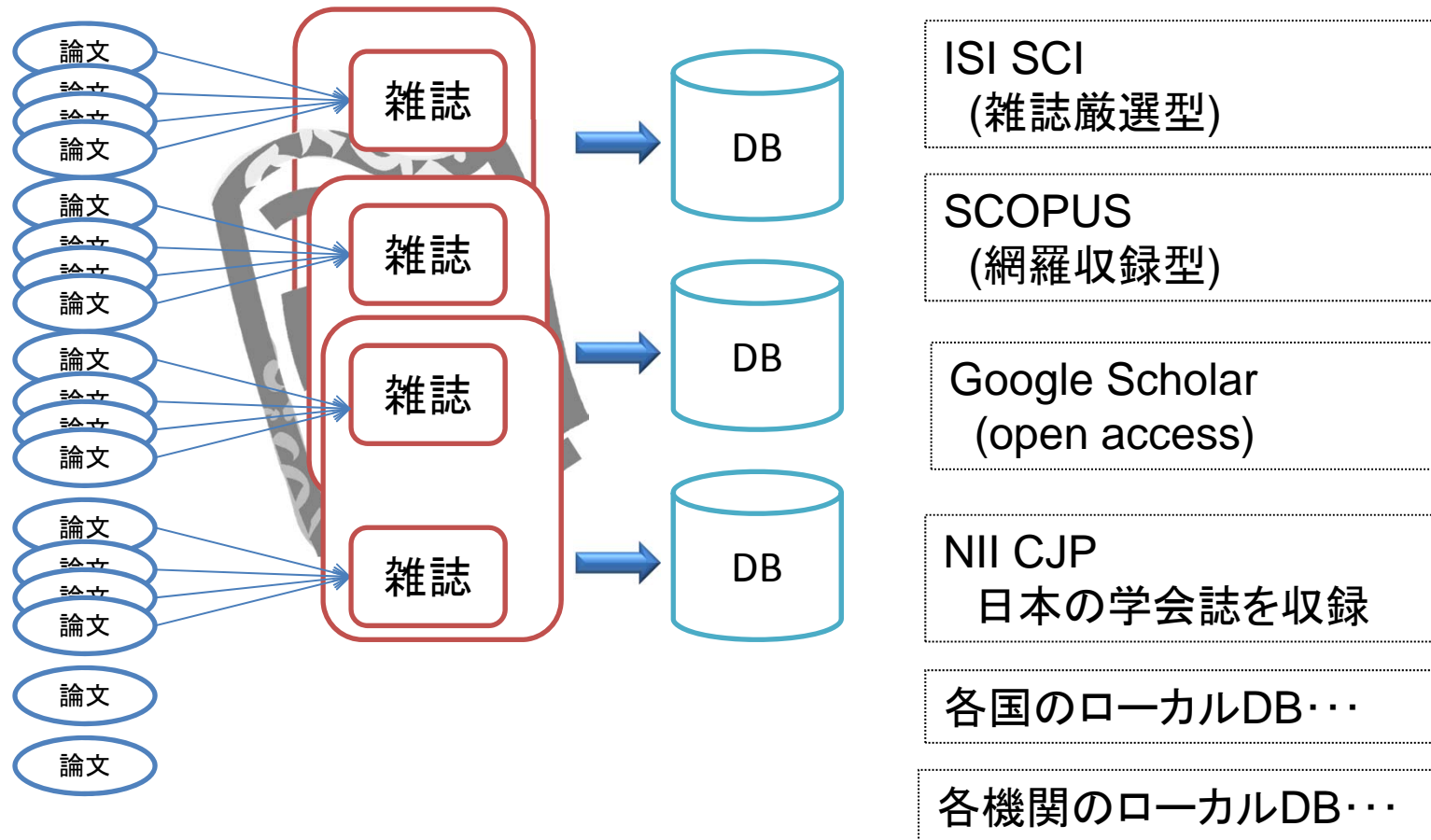
ビブリオメトリックスの略史

評価・政策に活用してビブリオメトリックス研究を行う研究グループ

- アメリカ、**CHI Research** (現在はThe Patent Board) (Narin)
- イギリス、**SPRU** (Science and Technology Policy Research),
Sussex大学
- オランダ、**CWTS** (The Centre for Science and Technology
Studies), Leiden大学
- オーストラリア、**CPI** (Center for Policy Innovation) ANU大学
- スペイン、**CINDOC** (The Centre for Scientific Information and
Documentation (Centro de Información y Documentación
Científica-CINDOC))
- ハンガリー科学院図書館、**ISSRU** (Information Science and
Scientometrics Research Unit)
- 80年代中頃、**OTA** (Office of Technology Assessment) は、ビブリオメ
トリックス指標をピアレビューと併用することを提唱
 - 主観的手法が研究評価を支配していた時代の終わり
(OTA, Research Funding as an Investment: Can We Measure the Returns?-A Technical Memorandum.1986)

ビブリオメトリックス指標の 研究評価への利用

研究評価データ



データの属性: 著者, 所属機関, 国, 年, 分野, 雑誌, DB

ビブリオメトリックス指標の 研究評価への利用

- 論文数
 - 研究の生産性, 活発さを表す
 - グレードの高い雑誌への掲載は, (ある程度)「質」も保証
- 被引用数
 - 研究の質(インパクト・影響度)を表す
- 基本指標(論文数, 被引用数)をもとに多様な指標:
 - 雑誌評価指標: JIF; SNIP; SJR
 - h-index, ...
 - Crown Indicator, ...
- 評価対象
 - 国／雑誌／大学・機関／個人

- スウェーデンの医学校Karolinska Institutetは、自らの研究を細かく評価し、研究戦略を策定するために、2006年の初めにKarolinska Institutet Bibliometricsプロジェクトを発足させた。
 - 種々の指標をまとめた非常に便利なマニュアルやHandbookをインターネット上で公開している。
 - <http://ki.se/ki/jsp/polopoly.jsp?l=en&d=1610&a=17742>



Karolinska
Institutet

Bibliometric indicators

- definitions and usage at Karolinska Institutet

さまざまな研究評価指標：
定義・短長所・使い方について

標準化指標

有名なオランダのライデン大学のCWTS(Centre for Science and Technology Studies)が開発した、論文あたりの引用数を分野(同タイプ・同発表年)で基準化した「クラウン指標」。「クラウン指標」を含めて多重指標による「ライデン・ランキング」といわれる欧州と世界の大学ランキングを発表、インターネットで公開。

特定の一年間において、その雑誌に掲載された論文、平均何回引用されているかを示す

本来は、特定の研究分野における雑誌の影響度を測る指標だが、実際には、研究者評価のために個別論文の評価指標として利用されることが多い

Indicators

Publications	論文数	5
Number of publications		5
Number of ISI publications		5
Number of publications in top journals		6
CEST world share of publications		7
CEST degree of specialization		7
CEST relative citation index		8
Citations	引用数	10
Number of citations		10
Citations per publication		10
CWTS field normalized citation score (crown indicator)		11
Field normalized citation score		13
Total field normalized citation score		14
Logarithm-based citation z-score		15
Top 5%		16
CEST relative impact index		
CEST normalized mean impact		
CWTS journal normalized citation		
Journal normalized citation score		
Journal packet citation score		
h-index		
Uncitedness		
Self citedness		
Cooperation		26
Co-authoring	構造的指標	26
Journals	可視化	27
ISI journal impact factor	雑誌評価	27
Normalized journal impact		27
Journal to field impact score		29
Citation reference values		31
Field citation reference value		31
Top 5% citation reference value		31
Journal citation reference value		32

研究者の生涯業績を示す指標として、物理学者Hirschが2005年に提案。ある研究者が過去に発表した論文のうち、h編それぞれについて、少なくともh回以上引用されたとき、その研究者のh指数はh。大学評価にも利用

基本指標1: 論文数

- 研究の生産性, 活発さを表す指標として用いられる。最大の長所は, 単純でわかりやすいこと
- 論文数だけで「質」を判断するのはおかしいという認識から, 引用指標の併記が求められることも
 - 学術雑誌, 特に「質」が高いとされる雑誌に掲載されるまでに, 何人ものピアレビューを経ているはずであり, 論文数はある程度「質」を保証するという考え方もある。
- とくに国・機関評価の場合, 以下の指標の利用も
 - 研究者1人あたりの論文数
 - 高引用論文数; 良い雑誌に掲載の論文数
 - 加重; 重み付けなし; (共著の場合のカウント法)

基本指標2：論文の被引用数

- 論文が発表されてから現在までに、他の論文から引用された回数
- 当該研究領域における被引用論文の貢献を示す
- 論文の被引用回数は論文の質に関するほとんど唯一の客観的データと考えられている
- 国・機関評価の場合、以下の指標の利用も
 - 研究者1人あたりの被引用数
 - 1論文あたりの被引用数
- 研究評価のための指標が、ビブリオメトリックスの領域でいろいろと開発・提案されているが、そのほとんどは論文数と被引用回数をもとに構成されている。

インパクトファクタ

- JIF (Journal Impact Factor)は、米国Thomson Reuters社 (旧ISI社)が作成した引用索引データベースの収録雑誌について、掲載された論文一編あたりの被引用回数の平均を算出したもの
 - 同社が編集発行するJCR(Journal Citation Report)を通じて毎年公表
- SCIやCurrent Contentsへの収録雑誌を選定するための定量的基準としてISI社自身が用いるために考案された
- わかりやすく便利な指標として、研究者個人、研究機関の評価指標に転用される
- しかし、IFの提唱者であるGarfield自身は、IFは雑誌の重要度を示す指標であり、研究者の業績評価の指標として利用すべきではないと繰り返し注意を喚起

h指数 (h-index)

- 研究者の生涯業績を示す指標として、物理学者Hirschが2005年に提案 (Hirsch index)
- ある研究者が発表した論文のうち、少なくともh回以上引用された論文がh編あるとき、その研究者のh指数はh
 - 被引用数10以上の論文を10編持つ研究者のh指数は10
 - 被引用数が10以上の論文が9編ならばh指数は9
- 高いh指数を得るためには、高い質の (引用度の高い) 論文を数多く発表しなければならない。
 - 量と質の両立が求められる
- 発表当初から、h指数は多くの研究者、政策立案者、メディアに注目されている。
- Web of Science, Scopus等でh指数が利用可能になってから、h指数に基づく研究評価が流行。最近過剰の利用が指摘されている。

ビブリオメトリックス指標を用いることの意義

- 客観的指標への社会的な要請
 - ピアレビューの問題点(主観性・信頼性・意識的・無意識的バイアス)を補完する
 - 説明責任に応える
 - 分野間・機関間・国間のグローバルな比較が可能に
 - Fundingと繋がるが多くなってきている
 - 世界大学ランキング・研究力ベンチマーキング

ビブリオメトリックス指標を用いる際の留意事項

- 技術的な問題

- 論文数・被引用数を正確に求めること自体，膨大な手間を要する作業
- いわゆる「名寄せ」の問題：著者の記述通りの情報がデータベースに収録されることが多いが，
 - 引用文献の表記法，雑誌名の省略法の違い，書誌情報記載の誤りなどにより，同一の論文が別のものとしてカウントされることがある
 - 姓と名のイニシャルの英文表記だけでは，著者を同定できないことがある
 - 所属機関名英文表記のゆれ・略名，組織改組などによる機関名の変化

ビブリオメトリックス指標を用いる際の留意事項

- 論文数や被引用数には種々の要因が影響
 - － 雑誌の採録範囲
 - Web of Science, Scopus, CiNii, それぞれに特徴
 - － 例えば, Negishi,E(1977)の2010/11/4時点での被引用数:
WoSでは491件, Scopusでは292件(1996年以降)
 - － 引用習慣や論文の寿命など, 研究分野間の差異
 - － 統計を取る期間
 - － 英語論文, レビュー論文, 方法的論文は引用されやすい
- 共著論文の論文数・被引用数統計
 - － 1論文あたりの共著者数・共著者機関数が増加傾向
 - － 共著論文の場合の集計法をどうするか
 - － 自己引用, 否定的な引用をどう扱うか

ビブリオメトリックス指標を用いる際の留意事項

- 各指標の特徴：固有の限界
 - － 評価対象に適合するか？
 - － 分野間の違い
 - － 引用統計の問題
 - 分布の歪み(性質)
 - 分野間の習慣の違い
 - 引用年数の問題
 - 研究の質と関係のない要因が被引用数を増やす
 - － たとえば、英語論文；レビュー論文；理論的・方法論的論文；国際共著が引用されやすい
 - 影響度の高い研究が謝辞等に記されることもある
 - あまりに先端的な研究は、優れたものでも最初は理解されず
 - 引用のタイムラグ問題

ビブリオメトリックス指標を用いる際の留意事項

- データベースへの依存性(収録範囲の問題):
 - 従来からのトムソン・ロイターのWoS
 - 近年現れたエルゼビア社の競合製品Scopus
 - 最近の強いライバルGoogle Scholar
 - 無料, カバー範囲が広い(論文・会議・図書・非英語文献...)
 - 質保証がない(範囲・基準が不明確)
 - 各種の評価指標は, 各DB内で計算される
 - 評価指標の改良や新指標の提案も特定のDB中での工夫
 - ⇒ 異なるDBに基づいた指標間の比較ができない
- たとえば, 日本応用物理学会刊行の英文誌*Japanese Journal of Applied Physics, Part 2: Letters*の2000年IFは, CJP版では0.59, SCI JCR版では1.16

ビブリオメトリックス指標を用いる際の留意事項

- 分野分類法の問題：
 - 各データベースの分野分類法が異なる
 - 論文ごとではなく、掲載雑誌に対して行われている
 - ⇒ 分野ごとの評価に問題が生じる可能性
 - ⇒ データ統合(後で紹介)の際に問題が生じる可能性
- 概念・指標の再解釈問題：
- 研究評価・ベンチマーキングのツールとしてビブリオメトリックスは、従来のビブリオメトリックスに新しい見方をもたらし、概念・指標に対して再解釈を行うことが必要
- 再解釈自体に、必ずしも筋が通っていない。また解釈・利用することによって、評価対象そのものを変えてしまうこと

研究評価のこれから

いくつかの課題

① 評価データの問題

- 応用研究・人文社会科学分野への適用
- 評価データの統合
- 論文の分野分類手法

② 指標の研究・開発

- ファンディング
- 機関(大学・学部・研究グループ等)評価
- ランキングと多次元評価

③ 論文等では測れない研究活動の成果

評価データの問題

- “データ”は最重要の問題の一つ
- 現在のDB: WoS, Scopus, Google Scholar...
 - 収録方針・特徴は違うが, 共通の問題を抱えている
 - 雑誌の収録範囲(人文社会分野が少ない)
 - 文献の種別: 雑誌・会議録・図書・章...
 - 国の代表性(国ごとの収録雑誌数が違う)
 - 言語の代表性(英文雑誌が圧倒的に多い)
 - 分野分類体系(現在, すべてのDBは論文ごとではなく, 雑誌で分野分類している。しかも, DB間で用いられる分類法が異なる)

評価データの問題

- バイアスを無くし, 公平な評価, 目的に合った評価を行うには?
- 各国内でのローカル評価
 - ファンディング (performance-based funding system)
 - S&T Innovation system: 産学連携・研究者情報
 - 自国・機関の研究について現状の把握
 - 自国雑誌が網羅されていない
 - 人文社会科学・応用分野のデータが不十分
- グローバル評価
 - 国間の研究プロフィール・ネットワーク分析
 - 世界大学ランキング・ベンチマーキング
 - 評価データの偏り

ローカル評価用データの整備

- 研究評価のための統合的DB作成の試み
 - 既存のDBの拡張(CWTS, Netherlands; SOOI, Belgium)
 - 基準の違い・文献種別の割合の違いなどの問題は未解決
 - ビブリオメトリックス指標を折り込んだ新しいDBの作成 (REPP, Australia; NIFU STEP, Norway; VABB-SHW, Flanders)
 - 使える指標ができるまで時間がかかる
 - ⇒ 各国でデータ整備が進行
 - 自国の引用索引データDB
 - 日本, 中国, 韓国, 台湾, タイ, フィリピン, マレーシア...
 - ⇒ 尺度間等化方法が必要(容易ではないが)

各国の研究センター

- ビブリオメトリックスを評価に適用するためのデータ整備・方法・指標開発を目的とする研究センター等が各国で創設されている。伝統的なものとして、例えば
 - CWTS (オランダ)
 - ISSRU (ハンガリー)
 - VINITI (ロシア) (All-Union Institute for Scientific and Technical Information, a branch of the USSR Academy of Sciences, 1952)
 - NISTADS (インド) (National Institute of Science, Technology and Development Studies)

各国の研究センター

- 比較的新しいところとして, 例えば
 - REPP (オーストラリア) (Research Evaluation and Policy Project)
 - NIFU STEP (ノルウェー) (The Nordic Institute for Studies in Innovation, Research and Education)
 - SOOI (ベルギー) (K.U. Leuven – Steunpunt O&O Indicatoren)
 - KAWAX (チリ) (Chilean Observatory of Science, Technology and Innovation)
 - WISELAB (中国)
 - iFQ (ドイツ) (The Institute for Research Information and Quality Assurance (iFQ – Institut für Forschungsinformation und Qualitätssicherung))
 - Science-Metrix (カナダ)
 - Evaluametrics (イギリス)
 - Evidence (イギリス)

グローバル評価用データの問題

- グローバル評価
 - 研究プロフィール, 研究ネットワーク
 - 世界大学ランキング
 - 用いられているデータベース
 - WoS, Scopus
 - Google Scholar
 - ウェブデータ
 - データのバイアス・DB間結果比較の難しさ
 - 大学ランキングの結果
 - 論文の被引用数・h-index, JIFの結果
- ⇒ 使用し結果を解釈する際に, 認識していることが重要

新しい指標の研究開発

- 機関(大学・学部・研究グループ)・個人評価
- ファンディング
- ランキングと多次元評価
 - 指標の多様性
 - 多次元評価・ベンチマーキング
- 代替(Alternative)指標の開発
 - 研究発表形態の変化
 - OAジャーナル, e-printアーカイブ, 個人のホームページ
 - 従来引用にはタイムラグ問題

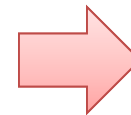
代替指標研究開発の必要性

- 学術情報を巡る環境の激変
 - 電子ジャーナル・雑誌論文のオープンアクセス(OA)化・機関リポジトリ・研究者自身による論文のウェブ公開など
- 現在の評価対象・評価指標が変わっていく可能性
 - OA化により被引用数が影響を受ける?
 - 機関リポジトリなどにより評価対象が広がる?
 - ⇒ トップジャーナル英文論文のみならず、国内学会誌論文、さらに雑誌に掲載されない論文・調査など、研究者の幅広い研究活動を評価対象とすることが可能になる
- 新しい評価指標(代替指標)の研究開発
 - ダウンロード数
 - リンク数
 - ウェブ上でのvisibility

研究評価は、どこへ向かう？

- 研究評価の対象・指標の意味の見直し
 - 研究には多次元的な側面がある。そもそも何を評価対象にするか。それをどう評価するのか
 - 既存のデータにとらわれず。
DBに収録されていない ≠ 価値がない
⇒ 多次元データ・多指標評価を積極的に行うべき
- 期待できること
 - 雑誌の電子化および「名寄せ」処理技術の進展
- 予測不能なこと
 - 研究の形態自体が今後大きく変化する可能性がある。それに伴い、評価指標も変わらざるを得ない？

- 研究評価は、科学政策に資するのか？
 - 評価のコストに見合う成果は得られたのか？
 - evidence-based policyがよかったevidenceがあるのか？
- 良いか悪いかは別にして、要請がある以上、研究評価は行われる
 - ビブリオメトリックス研究者としては、最善を尽くして新しい要請や環境に合わせて、新しい技術・指標の開発・改良に努力
 - しかし、追いつかない現状がある
 - 指標の限界・注意点を十分に認識することが重要
 - 評価の目的に合った指標を選ぶ
 - 複数の指標の利用も検討する
 - 各指標を正しく利用する
 - 結果を正しく解釈する



図書館員の
出番？

- モデルとして, 教育評価
 - 小論文⇒客観的テスト(多肢選択)⇒適応型客観テスト+その他⇒+認知診断テスト(学習の補助)
 - ピアレビュー中心 ⇒ 指標に依拠 ⇒ 多様な方法(評判・ピアレビュー・調査...)⇒+研究・機関意思決定の支援(ベンチマーキング, データ・ツールの提供)