

引用統計 Citation Statistics

京都大学数理解析研究所
国立情報学研究所国際学術情報流通基盤整備事業SPARC JAPAN
共催 連続セミナー

京都大学数理解析研究所研究集会
「紀要の電子化と周辺の話」

2008年9月2日

小田忠雄

<http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>

Joint Committee on Quantitative Assessment of Research

Citation Statistics

A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)

Corrected version,
6/12/08

Robert Adler, John Ewing (Chair), Peter Taylor

6/11/2008

引用統計

定量的研究評価に関する
国際数学連合・応用数理国際評議会・数理統計学会
合同委員会報告
R.アドラー, J.ユーイング(委員長), P.テラー

日本数学会
日本学術会議数理科学委員会
合同による和訳

予定:
日本数学会「数学通信」に要旨
日本数学会HPに全文
日本学術会議数理科学委員会
の記録として学術会議HPにリンク

以下で
内容を紹介

John Ewing HP <http://www.ams.org/ewing/>

アメリカ数学会専務理事



Ewing Shared Material



Citation Statistics

A report from the IMU in cooperation with ICIAM and IMS. On the use and misuse of citation data and statistics in assessing research. (418K)

*If you are interested in republishing this report, permission will be granted by writing to secretary@mathunion.org. A package of files, text and graphics, can be found [here](#).

**For additional articles on this subject, see the special theme volume of [Ethics in Science and Environmental Politics](#).

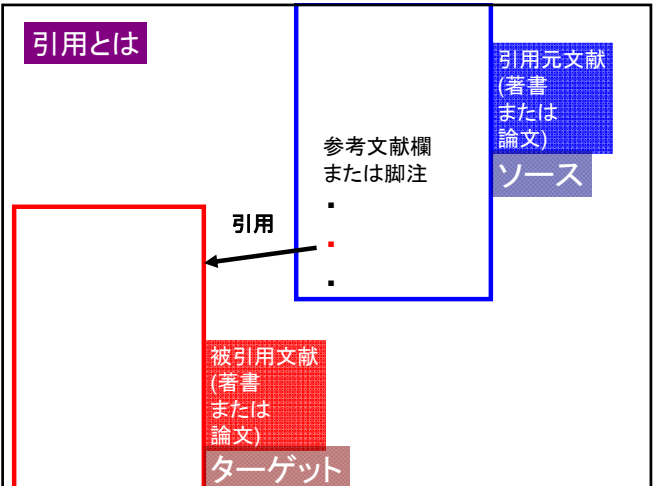
Where are journals headed? (v 6.0)

Why we should worry about the author-pay model (25K)

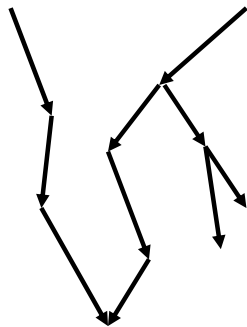
(Also in the *Notice of the AMS*, v. 55, no. 3, March 2008.

<http://www.ams.org/notices/200803/tx080300381p.pdf>).

引用とは



論文間の引用関係がデータベース化
されていれば極めて便利



引用の連鎖を辿っての
関連研究の調査

誰が引用しているかを
逆に辿って
その後の展開も把握

収録DB次第で
分野を跨った
学際的な情報検索

E. Garfield

科学情報研究所 (Institute for Scientific Information)創設
トムソン・サイエンティフィック社 (Thomson Scientific)
現在の名称は、トムソン・ロイター社 (Thomson Reuters)

窪田輝蔵著
「科学を計る---ガーフィールドとインパクト・ファクター」
株式会社 インターメディカル, 1996年,
ISBN4-900828-02-5 C1040 P2000E

約9,000の
「データ収録学術誌」
に掲載された全論文
および
それらが引用している
すべての文献
のDB

DB収録時の作業は自動化

引用元文献の供する
被引用文献に関する情報は
必ずしも万全ではない

- 筆頭著者のみ (XX, et al)とする分野が多い
- 著者の姓とイニシャルのみ
(実際の著者の同定が困難)
- 表題を挙げない分野が多い
- 書誌情報の間違い
- 書誌情報の略記

研究者が引用の連鎖を辿る目的で使用する限り、
あまり問題にならない。

PT Journal
AU Wu, SY
TI Equivariant holomorphic inequalities II: Torus and non-Abelian group actions **引用元論文**
SO JOURNAL OF DIFFERENTIAL GEOMETRY
NR 28
CR ATIYAH MF, 1968, ANN MATH, V87, P451
ATIYAH MF, 1967, ANN MATH, V86, P374
FULTON W, 1993, INTRO TOXIC VARIETIE
GROSSBERG M, 1994, DUKE MATH J, V76, P23
GUILLEMIN V, 1990, ADV MATH, V82, P160
MATHAI V, 1997, J DIFFER GEOM, V46, P78
被引用文献
略
ODA T, 1988, CONVEX BODIES ALGEBR
ODA T, 1991, GEOMETRY TOXIC VARIE, P407
略
PRATO E, 1994, COMPOS MATH, V94, P113
BP 401
EP 429
PG 29
JI J. Differ. Geom.
PY 1999
PD MAR
VL 51

5年程前に WoSの
Cited Referencesで
ODA, T
を検索した一例

ODA T, 1978, TAT I FUNDAMENTAL RE, V58
ODA T, 1978, TATA I FUND RES LECT, V58
ODA T, 1978, TATA I FUNDAMENTAL R, V57
ODA T, 1978, TATA I FUND RES, V58
ODA T, 1978, TATA I FUND RES, V58
ODA T, 1978, TATA I LECTURE NOTES
ODA T, 1978, 58 TAT I FUND RES
ODA T, 1978, TIFR LECTURE NOTES, V11
ODA T, 1978, LECT TOR EMB APPL TA, P58
ODA T, 1978, LECTURES TORUS EMBED, V58
ODA T, 1978, TORUS EMBEDDINGS APP

Tata研究所での
講義録
の引用のされ方

ODA T, 1988, ERGEB MATH GRENZGEB, V15
ODA T, 1988, CONVEX BODIES ALGEBR, P15
ODA T, 1987, ERGEBNISSE MATH IHRE, V15

Springer社刊
単行本
の引用のされ方

TR社(TS社, ISI)の本来の対象である
学術誌掲載論文ではないため、
DBへの収録の際の簡略化は不統一

ODA T, 1973, P INT C MANIFOLDS RE, P373
ODA T, 1975, ALMOST HOMOGENEOUS A, P373
ODA T, 1975, MANIFOLDS TOKYO, P373
MIYAKE K, 1973, MANIFOLDS TOKYO, P373
MIYAKE K, 1973, ALMOST HOMOGENEOUS A
MIYAKE K, 1975, HOMOGENEOUS ALGEBRAI, P373

国際会議で発表した
三宅克哉氏との
共著論文
の引用のされ方

ODA T, 1979, T AMS, V253, P1
SESHADRI CS, 1979, T AM MATH SOC, V253, P1

Seshadriとの共著論文

KATZ NM, 1968, J MATH KYOTO U, V8, P199

Katzとの共著論文

ODA T, 1977, MATH ANN, V231, P97
ODA T, 1981, P JAPAN ACAD A, V57, P415
ODA T, 1982, PERIODS HILBERT MODU

織田孝幸氏の論文

データ収録誌に掲載された論文である限り、同定は可能

研究評価

(透明性, 説明責任, 資金の有効活用)

嘆かわしい趨勢(世界的)

- 質の評価の困難さから
- ピア・レビューは非効率, 主観的とみなし
- 計測可能な数字に置き換え,
「簡単かつ客観的」に
- 分野を跨る比較も必要
- 引用データに基づく統計を活用
文献計量学(ビブリオメトリック)

評価とランク付け

学術雑誌のランク付け **インパクト・ファクター**

学術雑誌の比較に関しても不十分
僅かな情報のみを活用
「体重のみで人の健康を判断する」
ようなもの

誤用(言語道断)

以下で詳説

論文
研究者
研究機関
学術分野
のランク付け

インパクト・ファクター

トムソン・サイエンティフィック社の「データ収録誌」(9,000誌以上)

TS社DB データ収録誌掲載の全論文の書誌と、その論文が引用する全文献の(省略形)書誌

TSの「データ収録誌」の2008年インパクト・ファクター

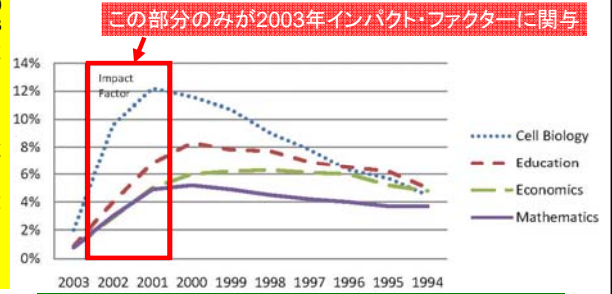
$$= \frac{\text{2006年, 2007年にJ誌に掲載された論文が2008年のデータ収録誌掲載論文で引用された回数}}{\text{2006年, 2007年にJ誌に掲載された論文数}}$$

2年の期間が妥当か否かは分野により異なる

掲載論文数の少ない学術誌では変動大(ターゲット年, ソース年)

被引用曲線

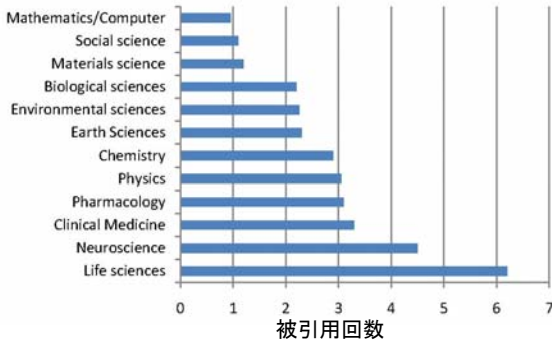
2003年論文による被引用パーセント



4分野で2003年出版の学術誌に掲載された論文が引用した論文の発表年

数学での引用の約90%は、この2年の期間外に発生している(MRDB)

分野毎の論文当たり被引用回数平均



分野毎に引用の文化が異なる。
分野を跨る比較でインパクト・ファクターは無意味

学術誌の評価に際しては、インパクト・ファクターは単なる一方法
他の情報による補足が不可欠

注意が必要

異分野の学術誌の比較
学術誌のタイプ
(掲載論文数, 言語, 書評掲載, レター)
年(ソース)毎の変動
データ収録誌でない学術誌の存在
計測期間(ターゲット)

実は、数学内部でも、統計分野での引用文化は他と異なる

インパクト・ファクターの言語道断な誤用

掲載学術誌のインパクト・ファクターが論文や著者の評価比較に拡大解釈されている

例:
大学内での研究評価に際し、論文掲載誌のインパクト・ファクターによるウエイト付け

権威ある学術誌に掲載された論文の質は、一般的に高い。
しかし、インパクト・ファクターが高いことと権威あることとの関連は?

Mathematical Reviews DBによる実験

TS社の数学関係データ収録誌は約400誌
MRDBでは1,200誌以上を対象、その内で、コアとなる800誌以上は掲載論文全てを対象。
1940年以降に刊行された数学文献、被引用文献の厳密な同定

コア誌に2000年以降に掲載された論文の引用文献300万件以上のデータ収録

引用の
50%は過去10年間の論文、
25%はその前の10年間、
12.5%は更にその前の10年間

MRDBによる例

米数学会 Proceedings誌(2001—2004) 被引用回数分布

短い論文
2,381論文, 約15,000ページ

2005年発行学術誌による引用
4年型「インパクト・ファクター」0.434

ソース年
ターゲット年
短かすぎる

米数学会 Transactions誌(2001—2004) 被引用回数分布

長い論文
1,165論文, 25,000ページ

2005年発行学術誌による引用
4年型「インパクト・ファクター」0.846

無作為抽出P誌掲載論文が,
無作為抽出T誌掲載論文以上の
被引用回数を持つ確率
62%

被引用回数分布は歪んでおり
いわゆる「市乗法則」に従う

研究者の被引用状況によるランク付け

研究者 S 氏の h 指数

$$= \max \left\{ n \mid \begin{array}{l} \text{S氏の発表論文の内,} \\ \text{被引用回数が} n \text{ 以上の論文が} n \text{ 編ある} \end{array} \right\}$$

発表論文数と被引用回数分布を単一数値で表す試み
Hirsch (2006)

(他にも, m 指数, g 指数等)

10回引用されている論文10編の研究者と,
10回以上引用されている論文10編と, 9回引用されている論文90編を持つ研究者の比較?

10回引用されている論文10編の研究者と,
100回引用されている論文を10編のみ持つ研究者との比較?

Anne-Wil Harzing (Melbourne大学教授 International Management) 作
Publish or Perish <http://www.harzing.com/> から入手可能

h指数, g指数 自動計算

Google Scholar から取得したデータ

ある実在する研究者の例(84編の論文)

被引用回数が
15以上の論文
15編

トップ15論文の被引用状況

h指数 15 だけでは,
このような実態
まで把握できない

根源的問題: 「引用」の意味

被引用回数が多いのは良い論文か?

「引用」の意味や解釈は極めて主観的

- 知的恩義を負っていることの謝辞
- 著名な著者の名声を借りる(「後光」効果)
- 引用する側と引用される側の関係
- 引用する方の論文の結果とは直接関係ないが, 学術的会話の進行のために引用
- 単一文献で, いくつかの結果を参照できる利便性 (それ以前の論文に「抹消効果」を及ぼす虞)
- 不備な結果への警告
- 貢献否定

統計の賢明な使用

使用するデータは適切か

TS社の場合, 「データ収録誌」に限られている

h指数の場合, Google Scholar 文献同定の困難

統計分析, 不確実性, 調整, 信頼性

統計モデルの設定, 結果の不確実性,
交絡因子(学術分野, 掲載論文のタイプ, 実験・理論等)
を考慮した調整, ランク付けの信頼性

解釈とインパクト

結果が「質」を示す量として妥当か
ランクを上げるために行う行動変更が及ぼす影響