

引用分析

浅井 勇夫

1. はじめに

記録された情報群の中から、目に見えない有用な情報を抽出できれば、知識は増大し、人間の頭脳労働は軽減する。このような夢を実現するために、計量情報学では、主に蔵書や学术论文を計数し、客観的でオートマチックな方法を用いて、特定の情報群に関する全体的な特性や構造を明らかにしてきた。ここでは、情報を抽出するためのデータとして文献の末尾にある引用文献を用いる分析手法を概略し、従来の Citation 分析と異なる Referation 分析と、それを実現するパソコン用ソフトについて述べる。

2. 引用文献データ

引用文献は、論文を読む研究者に非常に重要な情報を提供する。論文に関係の深い先人の研究論文は列挙され、分野内の論文の位置づけが明確になる。したがって、引用文献は文献間の関係を表わす貴重なデータであり、重要な文献属性である。

現在のオンライン用文献データベースは、残念ながら、この引用文献データが入力されていない。人間の知的作業を通して、書誌事項、抄録、キーワード、分類コードなどを付与したデータベースが作られる。そして、書誌項目やキーワードなどをキーとして、必要な文献群を検索する。引用文献を抄録誌の中にリストしたり、データベ

ス化するのが非常に困難なために、現在のような形態になったものと思われる。

ここで、引用分析の世界で扱う引用文献データについて説明する。ある文献Aが文献Bを引用すれば、文献Bは引用文献である。文献の末尾にリストされるのは、このBの集合である。このAとBの関係は誰でも知っている。しかし、引用分析では、逆転の発想というか、AとBの関係の逆も用いる。すなわち、文献Bは文献Aに引用されたとみなす。入力は1回であるが、その逆も使用するので、データは倍増し、効果を発揮する。

文献は2種類の文献群をもつ。1つは引用文献群 (References) であり、もう1つは被引用文献群 (Citations) である、われわれが簡単に知ることができるのは引用文献群である。それは年月が経過しても増えることはなく、むしろ退化していくように感じられる。しかし、引用文献群を転置して得られる被引用文献群は、文献がどのように利用され発展していったかを示し、文献の価値に応じて増大し、いかにも生命力をもっているように思われる。

従来、引用文献群や被引用文献群を別々に扱っていたが、最近、それらを統一した Referations が使われるようになってきた。図1は、その概念を示すものである。Referations の定義は、ある文献に関する引用文献群、被引用文献群、そして文献自体を1つにまとめたものである。

Citations や Referations の概念がいくら良くても、データの構築が不可能であれば実用化でき

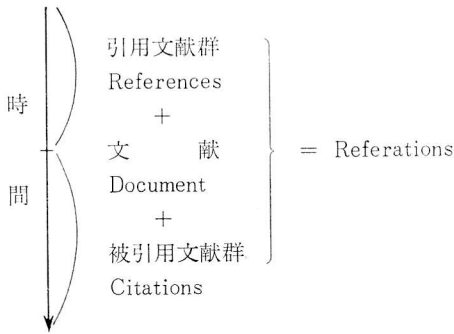


図1 Referationsの定義

ない。この分野の貢献者であるGarfieldは、1955年に、Citation Index概念を提案した[1]。1963年には、613種類の科学技術雑誌に掲載された140万種類の引用文献データをコンピュータ処理し、Science Citation Indexを冊子体で発行した[2]。現在、ISI社のSCI、SSCIデータベースは、DIALOGでオンライン検索可能である。

図2は、SCIやSSCIの入力書式の例を示す。引用文献は、第1著者、出典の省略形、巻・頁・発行年で構成してあり、文献にリストされたすべての引用文献を入力する。これは、大変な労力と記憶容量が必要である。一般のデータベースと違って、抄録やキーワードなどはないが、非常に速報性があり、データベースの価値は高い。

次に、Referationに関しては、著者がパソコン用の入力処理システムを開発した[3]。引用文献データは、文献ファイルに登録した文献の文献番号だけで構成し、画面上に表示した著者索引の支援により簡単に入力できる。引用ファイルを転置して、Referationファイルをつくるが、それは文献番号で構成する。番号から著者、標題、出典などの書誌項目が即時に呼びだせ、書誌項目全

- BRADFORD SC (DOCUMENTATION, 1949)
- SMART SC (AM ED RES J, V18, P399, 1981)
- SUMMERS EG (READING RES Q, V19, P102, 1983)
- PRICE DO (KEY PAPERS INFORMATI, P195, 1980)
- WEINSTOCK M (ENCY LIBRARY INFORM, V5, P14, 1971)

図2 SCI, SSCIの引用文献データの入力書式

部を入力したのと同じ効果をもつため、豊富な分析が可能になる。

なお、引用文献データの利用と批判は、Smithが解説している[4]。CitationsはReferencesと同義に使われるが、引用分析の分野では、上記のように被引用文献として用いる。Garfieldが、この新しい概念を新語で表現していたならば、用語上の混乱は避けられたものと思われる。

最近、論文の参照文献の書き方に関する新しい基準が、この特集で書いておられる仲本氏等により作られた[5]。参照文献の書き方が標準化されれば、引用分析の研究者だけでなく、文献情報の流通面で大いに役立つ。

3. 引用分析

引用分析は科学者や研究者の研究活動の分析に使われる。最近の研究は、データベース化されたSCI, SSCIのCitationデータを用いるものが多い。Garfieldが書いた著書[6]は、この分野での代表的なものである。日本語の文献として、坂本[7]、上田・緑川[8]が参考になる。以下、データ別に分析方法を説明する。

3.1 Reference Analysis

簡単な例で説明すれば、日本OR学会論文誌を過去5年位収集し、論文にあるすべての引用文献の著者名、雑誌名、年齢、分野別の頻度統計をとれば、引用パターンの解剖学的所見が得られる。数年前に、経営工学関係の洋雑誌AIIEの引用文献のコンピュータ処理を試みたが、引用文献の入力や識別は大変困難であった。

この分野ではPriceの業績が顕著で、邦訳の「リトル・サイエンス ビッグ・サイエンス」がある[9]。単に頻度統計をとるのではなく、文献間の関連性を分析する方法が、Kesslerにより1963年に提案された[10]。文献Aと文献Bに共通にある引用文献を計数し、書誌結合(Bibliographic Coupling)と名づける。これをベースにいろいろな測度を定義し、類似した文献をグループ化する。

基になる文献群が最近のものであれば、研究開発動向を識別でき、また検索にも適用できる。

Reference Analysisは、コンピュータ化することが困難なため、次のCitation Analysisの普及により、あまり使われていない。しかし、最新の文献群を分析するには、CitationデータよりもReferenceデータのほうがよい結果を得る。

3.2 Citation Analysis

SCI, SSCI データベースは、科学界に大きな影響を与えている。ある分野では、Citationカウントを評価に用いる。論文、著者、所属機関、雑誌、国など、カウント・データのaggregateの仕方により、いろいろな評価が可能である。有名なものに、Journal Citation Reportsがある。

以前に、SSCI を用いて経営工学関係の洋雑誌の分析を行なった[11]。11種類の雑誌間の関連性を調べたが、JORSA, ORQ, MS, AIIEの4誌が互いに強く関連し合っていることが定量的に確認された。

Citationデータを評価に使う場合には、その特性に注意する必要がある。ある論文が発行されてから引用した論文が発行されるまでには、ある程度の年月がかかる。たとえば、流通、入手、精読、論文作成、投稿、査読、印刷、そして出版の過程を経る。このため、発行後、数年間はカウントされない。これらの現象をよく理解したうえで、評価に使うべきである。

1973年、Smallは文献間の関連性をはかる測度として、共引用(Co-citation)を提案した[12]。書誌結合のReferencesを使う代わりにCitationsを用いるものである。SCI, SSCIはコンピュータ処理が可能のため、非常に多くの研究がなされている。単に分類するだけでなく、科学の発展過程をマッピングしたりする。これらに興味のある方は、文献[6, 7, 8]を参照していただきたい。

SCIやSSCIデータベースは、Citation Indexが強調されすぎている。そのため、Referenceデータを使う分析が、ほとんど行なわれていない。

せっかく苦勞して入力したデータも、半分だけしか利用されていない。今後、文献の識別が自動的に行なえるソフトの開発が望まれる。

3.3 Referation Analysis

文献の属性をReferenceデータやCitationデータだけで表わすのは、あまりよい結果を生まない。古い文献にはReferencesがなく、新しい文献にはCitationsがない。文献間の関連性を求める書誌結合は比較的新しい文献同士に、共引用は古い文献同士に有効である。上記のReferationsを使えば、文献間の発行年による差異を取り除くことができる。図3は、研究用に収集した1160編の文献からなる、パーソナル・データベースにおけるReferationsの例を示す。たとえば、1番目のPriceの論文は、Referationsが113編、Citationsが102編、Referencesが10編あることを示す。

図4は、2つの文献(10860, 10137)の関連度数を表わす。6個の書誌結合(10070~10081の実線)や共引用(11081)を含むとともに、ReferenceとCitationとの新しい結合関係(10525と10119)や直接引用(10860と10137)を含んでいる。これをもとに分析すれば、前述の場合よりも、分析精度は向上し、すべての文献同士の関連性に適用できる。

このようなReferationデータを使えば、新しい文献検索が可能になる[13]。文献をキーとして(1)文献のReferationsを検索、(2)文献に関連性の高い n 編の文献を検索、などができる。さらに検索した文献群に関連性の高い文献群、キーワード群、著者群、そして出典群を求めることができる[14]。得られた情報は一種の専門情報であり、

No	: FDocN	= Yr	Cdr	Cit	Ref	Author
1	: 30120	= 63	113	102	10	PRICE, D. J. DE.
2	: 10121	= 65	98	88	9	PRICE, D. J. DE.
3	: 00009	= 48	76	75	0	BRADFORD, S. C.
4	: 10056	= 72	105	74	30	GARFIELD, E.
5	: 10011	= 69	53	47	6	BROOKES, B. C.

図3 Referationsの例

Cit.	1: 11078=84	ETO, H.			
	2: 41110=83	SCHUBERT, A.			
	3: 11081=83	SUMMERS, E. G.	-----	1: 11091=83	SUMMERS, E. G.
	4: 10978=81	KANTOR, P. B.			
Doc.	5: 10860=81	ASAI, I.	=====	2: 10860=81	ASAI, I.
	6: 10597=78	PRAUNLICH, P.		3: 10808=78	BOICESCU, V.
	7: 10607=78	BROOKES, B. D.			
	8: 10605=78	DROTT, M. C.			
Ref.	9: 10592=77	BOOKSTEIN, A.		4: 10567=77	NARIN, E. F.
	10: 10586=77	LEIMKUHLE, F.			
	11: 10527=76	PRICE, D. J. DE.			
	12: 10525=76	HASPERS, J. H.	-----	5: 10525=76	HASPERS, J. H.
	13: 10119=75	POPE, A.	-----	6: 10119=75	POPE, A.
	14: 10090=73	LAWANI, S. M.			
	15: 10137=73	SARACEVIC, T.	=====	7: 10137=73	SARACEVIC, T.
	16: 10156=72	WILKINSON, E. A.		8: 10144=71	TAYLOR, L. J.
	17: 10285=71	NARANAN, S.		9: 10136=71	SARACEVIC, T.
	18: 10070=70	GOFFMAN, W.	-----	10: 10070=70	GOFFMAN, W.
	19: 10011=69	BROOKES, B. C.	-----	11: 10011=69	BROOKES, B. C.
	20: 10069=69	GOFFMAN, W.	-----	12: 10069=69	GOFFMAN, W.
	21: 20166=69	BROOKES, B. C.			
	22: 10260=69	FAIRTHORNE, R.	-----	13: 10260=69	FAIRTHORNE, R.
	23: 10093=67	LEIMKUHLE, F.	-----	14: 10010=68	BROOKES, B. C.
	24: 20319=67	GROOS, O. V.		15: 10093=67	LEIMKUHLE, F.
	25: 10035=62	COLE, P. F.		16: 10296=67	GOFFMAN, W.
	26: 10081=60	KENDALL, M. G.	-----	17: 10066=66	GILCHRIST, A.
	27: 10705=55	SIMON, H. A.	-----	18: 10081=60	KENDALL, M. G.
	28: 10008=34	BRADFORD, S. C.		19: 30261=49	ZIPF, G. K.
				20: 00009=48	BRADFORD, S. C.

図4 文献860と文献137との関連度数

ば、さらに多量の文献が処理できる。

ソフトは、N88-BASIC (86) MS-DOS版のコンパイラを使用し、試行錯誤をくりかえしながら、著者一人で開発した。ソートやマージなどの高速化が必要なところは、8086アセンブラを使った。プログラムの容量は、BASICの中間言語で約300KBで、マルチ・ステートメントのプログラム・リストは、1枚60行で約150枚である。

操作は、ユーザーが使いや

すいように設計されている。メニュー方式になっており、ジョブは、ファンクション・キーの選択と画面の最下行に示される簡単なコマンド入力で行う。表1は、ファンクション・キーの一覧表を示す。

入力には文献IN、引用IN、作成の3種類のジョブからなる。文献INは、文献の書誌事項を入力し、文献ファイルを作成する。種別、文献番号、著者、標題、出典、発行年からなり、各種の索引に支援されて入力する。引用INは、文献のReferenceデータを、著者索引の支援のもとに入力するジョブである。番号コードの入力だけで、Referenceデータが簡単に作成できる。なお、あらかじめ55種類の索引ファイルを作成して、検索や分析を高速化した。

出力として、リスト、検索、関連性の3種類がある。リストはデータベースの内容をいろいろな書式で出力する。[f1]の引用は、Referencesを文献ごとにリストする。[f2]は、標題のKWIC索引を出力する。[f3]は、文献番号順、標題キー、著者、出典、発行年順、標題など、6種類の索引リスト、[f4]は4種類のランクリスト、[f5]は4種類の書誌リストがある。[f6]はブラッドフ

研究者に有用なものである。

Citation Analysisで使われた各種の分析手法は、Referation Analysisにも適用できる。この新しい分析法は、非常に可能性を秘めたものである。やがて研究者は、独自のパーソナル・データベースをパソコン上に作成し、それに支援されて研究をすすめる時代がくる。

4. パソコンを使った引用分析

研究者は、高性能で低価格なパソコンなどのOA機器を利用して、研究効率をあげる必要がある。また、パソコン通信の普及で、研究者間のコミュニケーションは大きく変化する。ここでは、研究者が収集した文献のコレクションから、研究活動に役立つ情報を分析するパソコン用ソフトについて述べる[3, 14]。

使用したパソコンのハードとして、(1)NEC製のPC-9801E(主記憶640KB)、(2)カラー高解像度ディスプレイ、(3)1MBフロッピー・ディスク装置、そして、(4)プリンタである。操作性を高めるために、各種のファイルをメモリ上に展開する。テスト用に1160編の文献コレクションを処理した。主記憶の大きいワークステーションを使え

ワード分布やロトカ分布など、計量文献学に関する10種類の分布を出力する[15].

次に検索では、検索の手がかりとなるデータベースの内容を表示して、利用者を支援できるようにした。13画面、22種類の情報が画面の下半分に表示される。書誌キー以外に、文献をキーとする検索も可能である。検索結果の論理演算やソート、画面やプリンタへ文献群の出力ができる。

検索で得られた文献群に関する関連性情報を作成することができる。まず、[f1]で検索式を選択してから、[f2]の文献、[f3]のキーワード、[f4]の著者、そして[f5]の出典を選択すればよい。図5は、著者情報とその検索画面例である。画面には、共引用分析で有名な著者 Small が書いた16編の文献群について、

Referations をベースに

した関連度数、関連係数、そして関連指数の高い著者名がリストされている。[A04]は関連度数の高い20人の著者群を検索したことを表わす。このように任意の文献群について、キーワード群、著者群、出典群を検索・操作・出力できる。

最新の引用分析手法を利用した研究者レベルのパーソナルな文献管理が、パソコンを用いて実現できた。そこから引き出される質の高い専門情報

表1 ファンクションキーの一覧表

	f-1	f-2	f-3	f-4	f-5	f-6	f-7	f-8	f-9	f-10
初期画面	検索	関連性			リスト	文献IN	引用IN	作成	交換	終了
1検索	引用SW	索引SW	ランクSW	操作	表示	呼出	登録	1行	印刷SW	終了
2関連性	式 00	文献	キー	著者	出典	操作	表示		印刷SW	終了
5リスト	引用	KWIC	索引	ランク	書誌	分布		1行	印刷SW	終了
6文献IN	作成	索引				呼出	登録		印刷SW	終了
7引用IN	作成	索引	リスト	削除		呼出	登録		印刷SW	終了
8作成	文献	キー	引用	分布		全部			印刷SW	終了

[2 関連性] SMALL, H./A (16)

[引用文献管理]

◎著者検索式										
[A01]	15	BIBLIOMETRIC/K<X								
[A02]	15	BIBLIOMETRIC/K<Z								
[A03]	10	(1*2)								
[A04]	20	SMALL, H./A<X								
[A05]	20	SMALL, H./A<Z								
[A06]	14	(4*5)								

	[1] 関連度数 < X			[2] 関連係数 < Y			[3] 関連指数 < Z		
No	Author	Freq	Total	Author	Coef	Freq	Author	Index	Freq
1	SMALL, H.	1444	5312	SMALL, H.	0.3438	1444	SMALL, H.	100.00	1444
2	GARFIELD, E.	937	14128	WHITE, H. D.	0.2354	298	GRIFFITH, B.	64.86	641
3	GRIFFITH, B.	641	3264	STONEHILL, J.	0.2285	154	WHITE, H. D.	56.95	298
4	PRICE, D. J. D.	355	7056	GRIFFITH, B.	0.2169	641	GARFIELD, E.	50.35	937
5	WHITE, H. D.	298	1232	MALIN, M. V.	0.1543	147	STONEHILL, J.	49.56	154
6	CAWKELL, A. E.	241	3504	MULLINS, N. C.	0.1380	147	MALIN, M. V.	38.39	147
7	NARIN, F.	233	5056	STUDER, K. E.	0.1332	71	MULLINS, N. C.	36.03	147
8	COLE, S.	196	4304	SULLIVAN, D.	0.1296	84	PRICE, D. J. D.	32.13	355
9	COLE, J. R.	177	4144	BARBONI, E.	0.1250	60	CHUBIN, D. E.	31.68	160
10	CHUBIN, D. E.	160	1376	LENOIR, T.	0.1182	52	SULLIVAN, D.	30.91	84

1 式03	2 文献	3 キー	4 著者	5 出典	6 操作	7 表示		9 印刷SW	10 終了
[1-3, 1-88 CR] 選択, 検索 [0] 結合 [,] 次 [.] 前 [P] 印刷 [=] 次F									

図5 著者情報とその検索画面例

に支援されて研究を進めれば、研究は効率化し、研究資源の有効な活用が期待できる。

5. おわりに

計量情報学関係のデータの特徴は、(1)1, 2, 3と数える、せいぜい、2~3桁のカウント・データ、(2)月・日ではなく、年単位のデータ、(3)生データのコンピュータ化は困難、などである。このよ

うな貧弱なデータ構造であるため、ORの研究者を満足させるようなモデル化は、あまりなされていない。さらに、得られる情報が実際に利用可能かどうかという実務面からの厳しい評価がある。

オンライン情報検索を使えば、研究に関する情報は簡単に収集可能である。2~3のデータベースを使用して、研究者別、大学別、あるいは会社別の業績を調べたり、さらに、計量的な情報分析により、ある種の専門情報を作り出すことが可能である。情報面で、非常にオープンな社会が到来しつつあることが感じられる。

情報は無料であるという慣習があるため、日本では、情報の価値はあまり認められず、ましてコストのかかる情報分析は存在しえなかった。しかし、欧米のデータベースに簡単にアクセスできるようになると、欧米流の情報料を払わなければならないが、情報コストや情報価値の認識は必然的に高くなる。今後、情報群であるデータベースの計量的情報分析が発展するものと思われる。

参 照 文 献

- [1] Garfield, E. : Citation Indexes for Science. *Science*. Vol. 122, No. 7 (1955), 108-111
- [2] Weinstock, M. : Citation Indexes. *Encyclopedia of Library and Information Science*, Vol. 5, 1971, 16-40
- [3] 浅井勇夫: パソコンによる引用文献データベースの開発. 第21回情報科学技術研究会発表論文集, 日本科学技術情報センター, 1984, 21-31
- [4] Smith, L. C. : Citation Analysis. *Library Trends*, Vol. 30, No. 1(1981), 83-106
- [5] 仲本秀四郎他: 科学技術情報流通技術基準—参照文献の書き方. 第21回情報科学技術研究会発表論文集, 日本科学技術情報センター, 1984, 155-161
- [6] Garfield, E. : *Citation Indexing—Its Theory and Application in Science, Technology, and Humanities*. John Wiley & Sons, 1979, New York
- [7] 坂本徹朗: 引用分析. 情報検索, 雄山閣, 1976, 第6章, 119-143
- [8] 上田修一, 緑川信之: 科学技術情報—利用者の立場から. *ドクメンテーション研究*, Vol. 34, No. 4 (1984), 175-185
- [9] デレック・プライス著, 島尾永康訳: *リトル・サイエンス ビッグ・サイエンス*. 創元社, 1970
- [10] Kessler, M. M. : Bibliographic Coupling Between Scientific Papers. *American Documentation*, Vol. 14, No. 1 (1963), 10-25
- [11] 西山徹, 浅井勇夫: SSCIによる経営工学分野に属する洋雑誌の分析. 日本経営工学会秋期研究発表会予稿集, 1980, 195-196
- [12] Small, H. : Co-citation in the Scientific Literature—A New Measure of the Relationship Between Two Documents. *Journal of the American Society for Information Science*. Vol. 24, No. 4, 1973, 265-269
- [13] 浅井勇夫: Referationによる引用検索. 情報学シンポジウム講演要旨集, 1985, 31-32
- [14] 浅井勇夫: Referationを用いた特定専門分野情報の分析. 第22回情報科学技術研究会発表論文集, 日本科学技術情報センター, 1985
- [15] 浅井勇夫: 計量文学で用いられるパソコンソフト. 第20回情報科学技術研究会発表論文集, 日本科学技術情報センター, 1983, 167-171

次 号 予 告

特集 経営財務とOR

情報および期待と証券価格 仁科 一彦
 会計情報と経営財務 伊藤 邦雄
 企業評価と

CAPM(資本資産評価モデル) 青山 護
 権利, 契約の財務理論と情報 久保田敬一
 経営財務とリスク・マネジメント 若杉 敬明

解 説

ポートフォリオ理論: その発展と意義 榎原 茂樹

研究レポート

株式投資と残差分析 斎藤 進

事例研究

債務格付けの実際 黒沢 義孝
 債券先物の実際 横山 直樹

連載 パソコンによるOR(4)

プロジェクト・スケジューリング 古林 隆

紹介・SSOR 大鑄 史男・塩出 省吾