

日本語学会誌論文間での相互引用統計に基づく学会間類縁性の測定と その応用可能性

○孫 媛、根岸正光
国立情報学研究所

Measuring the Relationship among Japanese Academic Societies with the Cross Citations Found in Their Papers, and Its Practicability

Yuan Sun, Masamitsu Negishi

This study is based on the citation data extracted from the Citation Database for Japanese Papers (CJP), produced by the National Institute of Informatics (NII), Institute of Electronics, Information & Communications Engineers (IEICE) and other 13 academic societies which have strong citing or cited relations with each other have been chosen for the analysis. Converting scheme of the raw citation counts into dissimilarity data, and statistical analyses by multidimensional scaling method to measure relationship among the societies are described. The potentiality of bibliometric research on the CJP database for practical applications will also be discussed.

1. はじめに

計量書誌学的分析では米国ISI社のデータベースを用いることが多いが、日本の研究活動を正当に評価するために、われわれは、国立情報学研究所で作成している、わが国の学会誌論文を対象とした「引用文献索引データベース」(CJP)を用いた研究を行っている。¹⁾本報告では、CJPから採取した引用統計により、電子情報通信学会をはじめ、互いに引用・被引用関係で結合される14の学会間の「距離」を様々な方法で算出し、多次元尺度法、クラスタ分析によって、学会間の関連性・類縁性を分析し、その結果の応用可能性について論じる。

2. データと分析

CJPデータベースを用いて、電気関連5学会、すなわち、電気学会、電子情報通信学会、映像情報メディア学会、情報処理学会、計測自動制御学会の学会誌を中心とする、論文相互間の引用回数を調査し、これら5学会と強い引用・被引用関係をもつ9学会を抽出した。²⁾それら14学会の会誌・論文誌における1998年、99年の発表論文が2000年の論文から引用された回数を表1に示す。たとえば、電子情報通信学会の雑誌論文は電気学会の雑誌論文を30回引用したのに対して、逆方向の引用は61回である。また、データ行列の対角要素は各学会雑誌の自己引用回数を表す。

本研究では、表1のデータに多次元尺度法(multidimensional scaling: MDS)を適用し、14学会を2次元ユークリッド空間に布置し、学会間の関係を幾何学的・視覚的に把握することを試みた。ただし、MDSを適用するには、表1のデータにいくつかの変換を施す必要がある。

まず、表1のデータは非対称であるが、MDSの代表的なアルゴリズムの1つであり、今回の分析に用いたALSCALは、対称行列に適用される。そこで、学会のペアごとに、引用回数と被引用回数を両者の平均値で置換して、対称行列に変換した。また、引用・被引用回数が多い学会どうしは類似性が高いとみるのが自然であるが、MDSで扱うデータは非類似度(距離)でなければならない。したがって、何らかの方法で原データを非類似性データに変換する必要がある。この変換法には多様なものが考えられるが、ここでは3種の変換法とそれらを用いた結果を示す。

表1 電気電子関連14学会の雑誌の相互引用回数
(1998年と1999年の発表論文の2000年論文における引用回数)

引用元の学会 (雑誌)	引用先 (被引用) の学会 (雑誌)													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	IEICE	IEEJ	ITE	IPSJ	SICE	JSAP	ASJ	MSJ	JSPE	JSME	ISCIE	RSJ	JSAI	JAMIT
1 電子情報通信学会	824	30	23	123	10	84	57	13	1	3	2	4	17	35
2 電気学会	61	585	4	14	30	31	2	21	14	5	10	7	1	1
3 映像情報メディア学会	127	7	97	23	1	34	1	19			2	1	6	
4 情報処理学会	73	3	5	197		1	1			1		2	12	
5 計測自動制御学会	15	32	3	6	86				6	8	21	14	2	
6 応用物理学会	42	13		1	1	897		24	1					4
7 日本音響学会	27	10		16	3	32	474		1	4			5	
8 日本応用磁気学会	7	54				18		224						
9 精密工学会	4	6	3	3	7	18			130	25		5		
10 日本機械学会	4	6		4	20				25	78	5	29		
11 システム制御情報学会	19	16	2	4	34		1		1	19	45	21	3	
12 日本ロボット学会	7	5	2	1	12				1	14	4	79	3	
13 人工知能学会	8	1		32	6				1	1	2	1	22	
14 日本医用画像工学会	6			3		1								19

変換法1：最小非類似値が0になるような定数から引用・被引用回数（平均）を引く。この変換による非類似性データを間隔尺度と見なして、SPSSのALSCALプロシジャにより求めた2次元空間布置を図1に示す。これを見ると、学会同士が環状になっており、情報処理、電子情報通信、応用物理、電気、機械系（精密工学、計測自動制御、機械）、システム制御、ロボット、人工知能という順序を経て、情報処理へ戻る。これは、学会間のコミュニケーション・ネットワークの形成状況を示しているように解釈できる。ここで電気学会が物理系と機械系の中間に位置するのが興味深い。

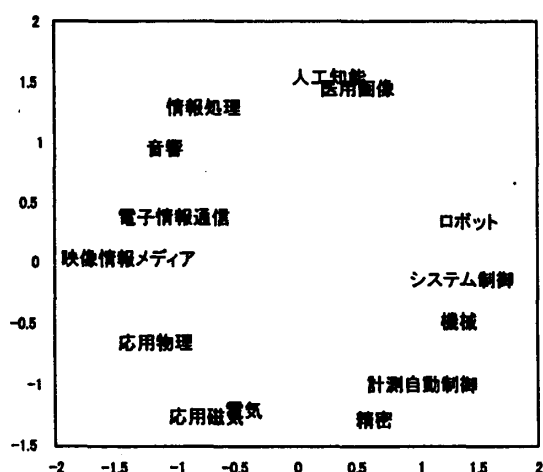


図1. 変換法1による14学会のMDS布置

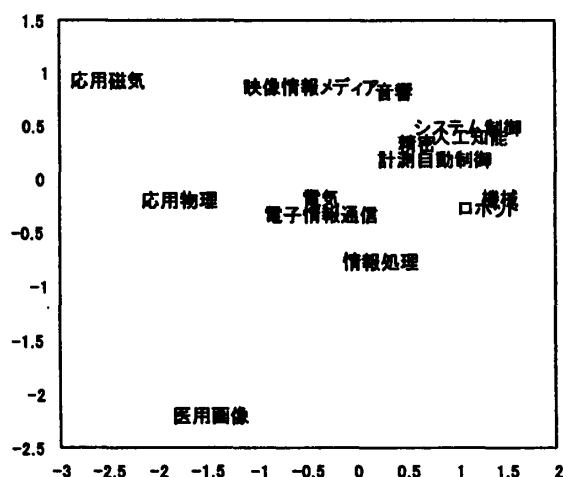


図2. 変換法2による14学会のMDS布置

変換法2：引用・被引用回数（平均）の逆数を非類似度とする。ただし、引用・被引用回数（平均）が0の箇所については、逆数を求める前に小さな値として0.01に置き換えておく。この変換による非類似性データを比率尺度と見なしてALSCALプロシジャによって分析した結果を図2に示す。データとモデルの適合度を示す指標（ストレス）は0.69であり、1の方法による0.33と比べると適合度は劣るが、解釈しやすい布置が得られている。すなわち、電子情報通信、電

気、情報処理の各学会は相対的に近くに布置されて、近縁の学会と見なせる。また、ロボット、機械、システム制御情報等の学会が、別のクラスタを形成している。また、応用物理と応用磁気学会が図の左側にあり、左下から右上に伸びる軸を引くと、これは“物理－電気電子－機械”の次元と解釈することも可能であろう。

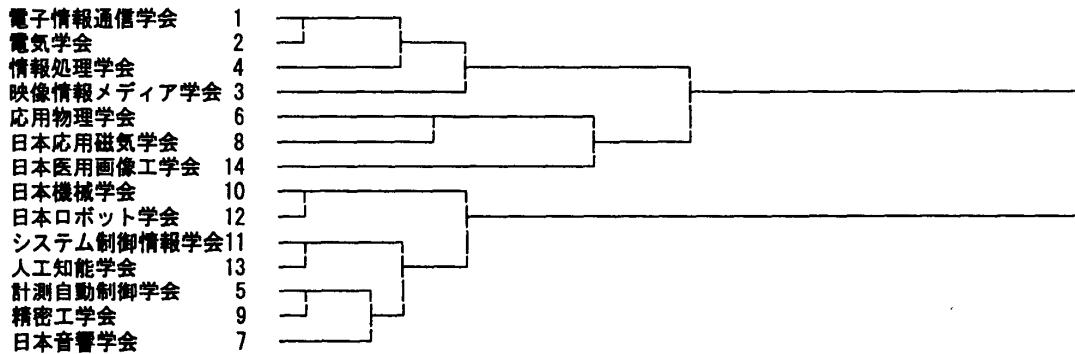


図3. 図2の座標値による学会のクラスタ形成を示すデンドログラム

この図における学会クラスタをさらに明確に示すために、図2の座標値に対してWard法によるクラスタ分析をおこなった。結果のデンドログラム（図3）を見ると、電気電子情報関連学会群（電子情報通信、電気、情報処理、映像情報メディアの4学会）、機械系学会群（機械、ロボット、システム制御情報、人工知能、計測自動制御、精密、音響の7学会）、物理系学会群（応用物理、応用磁気、医用画像工学の3学会）という3つのクラスタが形成されることがわかる。

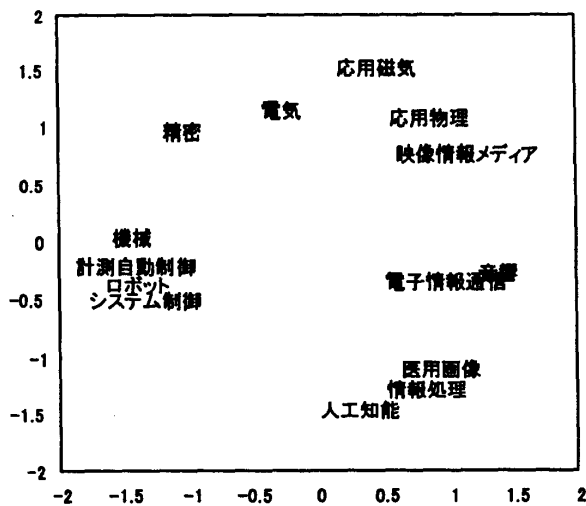


図4. 変換法3-1による14学会のMDS布置

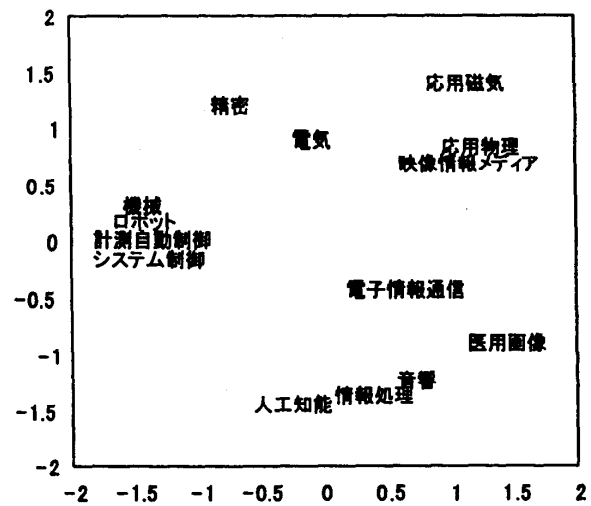


図5. 変換法3-2による14学会のMDS布置

変換法3-1: 学会規模の影響を取り除くための基準化をおこなう。基準化の方法も種々考えられるが、引用回数合計が学会規模を表すと見なし、14学会各々について、引用・被引用回数（平均）を引用回数の合計（自己引用は除く）で割った上で、最小非類似値が0になるような定数から引くことにより、非類似性データを求めた。この変換による結果を図4に示す。この布置は、学会どうしが環状に列なる点で図1と似ているが、基準化をおこなったことで学会間の遠近関係が強調されている。

変換法3-2: また、合計引用回数で割った後、さらに、平均値0、標準偏差1になるように標準化を施した結果が図5である。この図では、左下から右上に伸びる“機械－電気電子－物理”の

軸と、それに垂直な“電気－電子情報通信－情報処理”の軸を見てとることができよう。

3. 考察

(1) 統計手法的課題

本研究では、学会間相互の引用・被引用回数が学会どうしの類似度を表すと考え、これを種々の方法で非類似度に変換して、多次元尺度法による2次元布置を試みた。これにより、学会間の関連性が地図上の位置関係のように表現され、関連性を吟味するための手がかりが得られる。

引用・被引用データに多次元尺度法を適用する際のもっとも重要な点は、引用・被引用回数（類似度）を非類似度（距離）に変換する方法であろう。この変換法には論理的に唯一絶対といえるものがなく、本研究でも妥当と思われる方法をいくつか提案し、分析をおこなった。一般によく用いられるのは、最小非類似度（もっとも近い箇所）が0になるような定数から類似度を減ずる方法（前記変換法1）であるが、今回のデータでは各学会がほぼ等間隔で環状に列ぶ布置が得られた。このメリハリのなさは、全体の引用・被引用回数の中で、学会内自己引用の割合が非常に大きいために生じるものと思われる。実際、それぞれの学会内で自己引用による基準化をおこなった場合（変換法3）、学会間の遠近に格差が現れた。今回のデータに関してもっとも解釈の容易な布置は、引用・被引用回数の逆数をとる方法であった（変換法2）。この布置では、電気電子情報関連学会群、機械系学会群、物理系学会群という3つのクラスターの形成がみられた。

今回の分析に関しては、非類似度への変換法に恣意的要素のあることや、モデルの適合度、非対称データの対称形への変換など、統計理論的観点からの問題が残り、これら理論面での妥当性の向上が課題である。

(2) 実際的应用に向けて

ところで、明治初頭以来のわが国の近代学会の設立数に関する推移統計をみると、経済循環との相関も観察され、今次の長期不況化において、このところ学会の創設は停滞している³⁾。その中で、バブル期の後遺症ともいえる小規模学会乱立の現状を憂う声にも強いものがある。グローバル化の時代に勝ち抜くには、国際発信力の強化が不可欠であるが、そのためには、学会の合併統合もさりながら、むしろテーマ毎の学会間共同事業の機動的な設定、推進が有効と考えられる。

本報告に示した分析方法は、こうした共同活動の企画、検討、説得の場において、これに「客観的」基礎を与え得るものと期待できる。そのためには、上述の統計手法的課題を処理しつつ、より幅広く、しかも細部に立ち入った分析が必要であり、今後CJPデータベースの充実を図るとともに、この種のビブリオメトリックス的分析の応用可能性を追求してゆきたい。

【参考文献】

- 1) 根岸正光, 西澤正己, 磯谷峰夫, 養毛堅一郎, 浅野正一郎, 電子情報通信学会の評価 — 学会誌の引用数分析を中心に, 電子情報通信学会誌, 82(5), pp479-484, 1999.
- 2) M. Negishi, Y. Sun, K. Shigi, *Citation database for Japanese papers: a new bibliometric tool for Japanese academic society*, Proceedings of the 9th International Conference on Scientometrics and Informetrics, August 25-29, 2003, Beijing, China, 10p. (to appear)
- 3) 今野省造, 日本の学会設立年を辿る, SCIENTIA((財)日本学会事務センター), No. 18, p12, 2002.