

論文数統計による研究水準の国際比較
----対応分析法を使ったデータ内部構造析出の試み----

○孫 媛
根岸正光

Cross-national Comparison of Publication Output in Engineering:
An Exploratory Study based on Correspondence Analysis

Yuan Sun
Masamitsu Negishi

ABSTRACT

The number of scientific publication in six fields of engineering has been surveyed by searching COMPENDEX database. In this study, we quantitatively analyzes the research trend of seven countries, e.g. Japan and USA, in each field over the period 1976-1994 using correspondence analysis. The results show that the percentage of some fields, such as industrial engineering and electrical engineering, is increasing while that of mining engineering is decreasing. Also, it is revealed that Japan and France continue to increase their shares in most fields.

1. はじめに

1976年から1994年の各年における日米欧の7か国の収録論文数を、物理、化学、工学、医学の27分野別に調査した^[1]。ここでは、そのうちCOMPENDEXデータベースに収録された工学分野(「土木」「金属」「機械」「電気」「化学」「管理」の6分野)の論文数のみを対象として、年度の経過に伴う、分野別・国別の論文数あるいは論文数比率の変化を分析する。

2. データと分析

2.1. グラフの視察

各分野の論文数・比率の推移 6つの分野ごとに、7か国合計あるいは世界合計の論文数推移をあらわしたグラフ(図1)からは、つぎのような傾向を読み取ることができる。6分野を通じて、80,81年頃に、それまで漸増していた論文数が少し減る。その後増加して80年代中頃にちょっとしたピークがあった後、92年に比較的大きな落ち込みがある。そして93年以降ふたたび増加傾向に転じる。分野別に見ると、「管理」の論文数の増加がとくに目立つ。そのほか、「電気」「化学」の論文数も数が増えている。一方、「金属」の論文は80年代中頃の方がむしろ多く、ここ数年は停滞気味である。(「機械」「土木」はその中間)。

76年から94年にかけての(7か国計における)分野別比率の推移を直接的に示すのが図2である。図2を見ると、「管理」の比率が大きく増え(20%から31%へ)、「電気」の伸びも大きく、「土木」「金属」「機械」の比率が減っていることを確認できる。

つぎに、7か国ごとに分野比率の変化を見る(図3)と、どの国でも、「管理」の比率が増え、「金属」の比率が減る傾向が見て取れる。日本では「化学」「電気」の比率も増えており、その分「機械」「金属」「土木」の減少率が大きい。USAは「管理」「電気」の比率が増し、「化学」の変化は小さ

い。「土木」「機械」の比率も漸減傾向にある。カナダ、ロシアなどでは「電気」の比率も上昇している。

(注)分野比率は「全分野」ではなく6分野計を100%として計算している。

各国の論文数・比率の推移 7か国別に、全分野あるいは6分野計の論文数をあらわしたグラフ(図4)からは、7か国とも80年代中頃に論文数が減り、93年以降の伸びが大きいという傾向を読み取ることができる。国別では、ロシアの乱高下が目立つ。

76年から94年にかけての各国の(全分野に対する)論文数シェアをグラフ化した図5をみると、日本のシェアが8%から15%へと約2倍になっているほか、フランスのシェアも着実に伸びていること、USAは絶対的な論文数の増加に関わらず、シェアはむしろ落としている(53%から50%へ)こと、ドイツとロシアのシェアも減り、カナダと英国が横這いであることなども読み取ることができる。

さらに、各分野内での7か国のシェアとその変化に注目すると、すべての分野において日本の比率の上昇していることがわかる。とくに「化学」で日本(とフランス)の比率が大きく増し、USAの比率はかなり減っていること、「管理」に関して、93年以降、日本の比率が上がっており、USAは論文の絶対数の急増に関わらず、比率はむしろいくらか減っていることなども読み取れる。

2.2.分割表の分析

グラフの視察より、年度の経過とともに、分野別・国別の比率が変化している様子が見えてきた。つぎに、3元データのうち、2元ずつを分割表にまとめて、これを分析する。分析の対象となる分割表はつぎの通りである。

(1)7か国について、年度×分野の論文数分割表(19×6);(2)6分野について、年度×国の論文数分割表(19×7);(3)19年度について、分野×国の論文数分割表(6×7)

2.2.1. Cramer's V

まず、行変数と列変数の関連の度合いを検討する。分割表の行と列に配置された2変数間の関連をCramer's Vによってあらわした。3通りの分割表について、SPSSのクロス集計表プロシジャを使って計算したCramer's Vを表1に示す。

表 1.1: 年度と分野の関連係数(国ごとに)

国	日本	USA	カナダ	英国	フランス	ドイツ	ロシア
Cramer's V	0.068	0.055	0.059	0.058	0.064	0.069	0.071

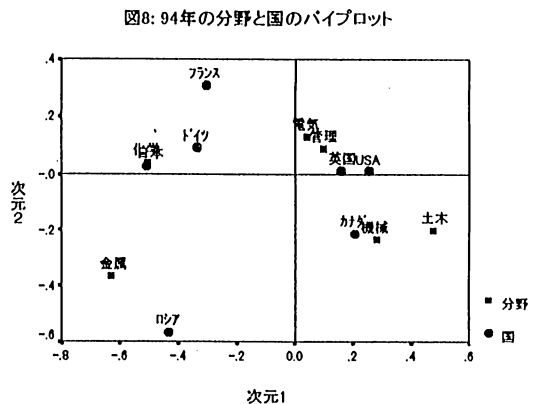
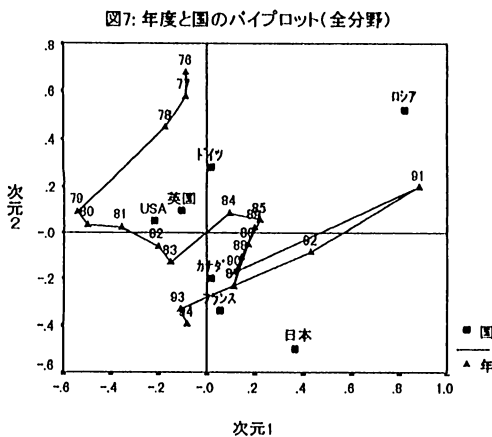
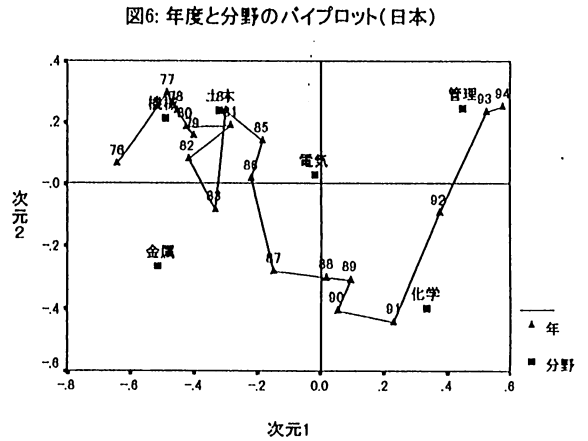
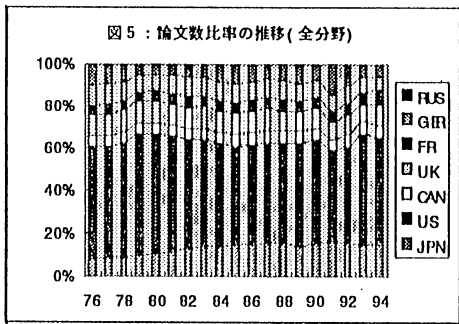
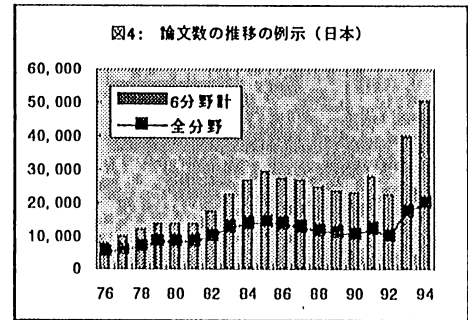
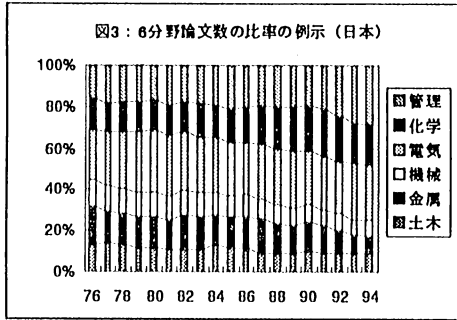
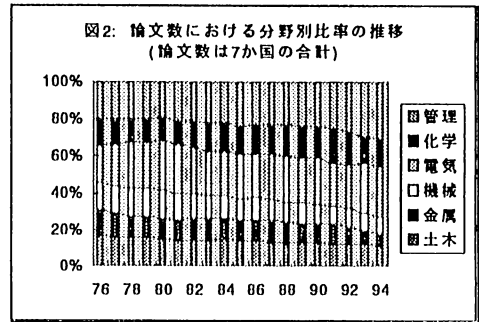
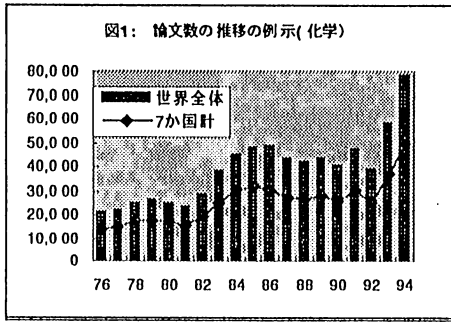
表 1.2: 年度と国の関連係数(分野ごとに)

分野	土木	金属	機械	電気	化学	管理
Cramer's V	0.059	0.064	0.063	0.046	0.068	0.054

表 1.3: 分野と国の関連係数(いくつかの年度について)

年	76	80	84	88	92	94
Cramer's V	0.064	0.056	0.050	0.065	0.057	0.048

いずれの2元間にも、弱いながらも何らかの関連がある(「2変数が独立である」という検定仮説は棄却される)。つまり、年度の経過とともに、論文数の7か国比率あるいは分野比率は若干変わってきており、また国による分野比率も変化しているといえる。



2.2.2.対応分析

つぎに、2 元間の関連について、さらに検討するために、対応分析(correspondence analysis)をおこなった。対応分析は、カテゴリー形式の項目データの相互関連を分析するための有力な手法である。カテゴリー変数どうしの分割表データに対応分析を適用すれば、2 変数間の関係を、「行と列のうち少ないカテゴリー数-1」個の解で説明することができる(解は、「行と列の変数のカテゴリーに適切なスコア」が割り当てられると同時に、そのスコアが割り当てられたときの「2 変数間の相関」という形で与えられる)。したがって、たとえば「年度×分野」の分割表(19×6)については、5 個の解が求められる。しかし、2 変数間の関係が単純なものであれば、1 個ないし 2 個の解(成分)のみで、関係をじゅうぶんに説明することができる。本データへの適用に当たっても、(とくに一方の変数が年度である場合は、)3 つ以上の成分を導出しても解釈が困難なので、原則として第 2 成分までを用いた。対応分析の計算とグラフ化には、SPSS のコレスポンデンス分析プロシジャを利用した(行と列のスコアは「正準」オプションを選択)。なお、「分割表の検定で用いられるカイ 2 乗(χ^2)」、「Cramer's V」、「対応分析により求められる相関(固有値 λ)」の間には密接な関連がある。

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(F_{ij} - E_{ij})^2}{E_{ij}}$$

$$V = \sqrt{\frac{\chi^2}{N [m \ln(r, c) - 1]}}$$

$$N(\lambda_1 + \lambda_2 + \Lambda + \lambda_{\min(r,c)-1}) = \chi^2$$

年度と分野 7 か国のそれぞれと世界計について、年度×分野の論文数分割表を作り分析をした(これは、「米国における『1960 年から 1976 年まで』の『12 の専門分野』における博士号取得者数の分割表を対応分析」した Greenacre の応用(1984)とよく似ている)。第 1 成分と第 2 成分の寄与率は表2の通りである。いずれの国でも第 1 成分の寄与率が比較的大きく(いちばん小さなロシアで約 60%)、第 2 成分までとれば 90%以上の寄与率が得られている。

表2.1:年度×分野の対応分析の結果

	日本		USA		カナダ		英国		フランス	
	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2
固有値	.0176	.0038	.0128	.0015	.0134	.0032	.0146	.0015	.0138	.0050
寄与率	.762	.165	.831	.100	.758	.181	.859	.087	.667	.240
累積寄与率	.762	.928	.831	.931	.758	.939	.859	.946	.667	.907

	ドイツ		ロシア	
	成分 1	成分 2	成分 1	成分 2
固有値	.0196	.0026	.0153	.0081
寄与率	.820	.108	.601	.320
累積寄与率	.820	.928	.601	.920

Greenacre に倣い、行(年)と列(分野)に与えられた、第 1 成分と第 2 成分のスコアを同一座標上

にプロットした(図 6)。これらの図より、第 2 成分の解釈は難しいが、第 1 成分は、ほぼ年度の経過に相当すると思われる。どの国についても、かつての「金属」から「管理」へと、比率の高い分野が変わってきている様子、「化学」が 91 年以降比率を落としていることなどを読み取ることができる。

年度と国 これも、6 つの分野および全分野について、年度と国のスコアのバイプロットを描いた(図 7)。第 1 成分、第 2 成分の寄与率は、分野によって少し違うが、第 2 成分までの累積寄与率は、おおむね 90% 近くにはなっている。「年度と分野」に比べると、全般的に第 1 成分の寄与率が小さく、年度の経過が第 1・2 成分に混じり合っている、あるいは、第 2 成分が年度の経過であるような印象を受ける。第 2 成分の負(図では下)の方向へゆくほど「最近」であることを表すとすれば、日本が他の国にもまして、比率を増やしている傾向が読み取れることになる。

表2.2:年度×国の対応分析の結果

	土木		金属		機械		電気		化学		管理	
	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2
固有値	.0147	.0039	.0137	.0098	.0142	.0055	.0063	.0049	.0188	.0083	.0103	.0064
寄与率	.708	.186	.551	.395	.603	.231	.496	.496	.673	.296	.585	.359
累積寄与率	.708	.894	.551	.947	.603	.834	.392	.888	.673	.968	.585	.944

国と分野 76,80,84,88,92,94 年の 6 つの年度および 76 年から 94 年までの論文数を合計した上での分析をおこなった(図 8)。94 年についてみると、第 1 成分の寄与率が 91.5%、第 2 成分までの累積寄与率が 97.4% となっており、この 2 つの成分でほとんどの変動を説明できることがわかる。第 1 成分の方向は、アングロサクソン系の国(USA,英国,カナダ)とそれ以外の国(日本,ドイツ,フランス,ロシア)とを分けるようである。またアングロサクソン系の国は、「土木」「管理」「機械」などとの分野と関連し、第 2 グループの国は「電気」「化学」と近いこと、とくに、日本は「化学」ととても近いことなどがわかる。ただし、ロシアと「金属」は、他の国・分野から大きく離れている。

表2.3:国×分野の対応分析の結果

	76		80		84		88		92		94	
	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2	成分 1	成分 2
固有値	.0184	.0011	.0120	.0024	.0078	.0028	.0137	.0062	.0121	.004	.0105	.001
寄与率	.889	.052	.761	.150	.621	.225	.637	.289	.731	.237	.915	.059
累積寄与率	.889	.941	.761	.911	.621	.845	.637	.926	.731	.968	.915	.974

2.2.3.対数線形分析

ここまででは、3 元の情報を 2 元ごとに切り取って分析してきたが、最後に 3 元の分割表データに対して、対数線形分析をおこなった。3 元分割表に対する対数線形分析のモデルにおいては、分割表の各セルの度数が

$$\log_e F_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$$

という「飽和モデル」で完全に説明されるが、期待度数と観測度数の食い違いを小さく押さえたまま、交互作用パラメータを減らすことができるならば、それがより良いモデルとして採用されることになる。対数線形モデルのうち、もっとも一般的な「階層モデル」に基づく計算を、Statistica の対数線形モデル分析プロシジャを用いておこなった。検定の結果、2 次の交互作用まですべてを含

む「飽和モデル」が採用された(仮説 $u_{ijk}=0$ が棄却される)。しかし、N(全論文数)がきわめて大きいことを考えると、この仮説が棄却されるのは、むしろ自然であるともいえる。そこで、交互作用項を減らしたモデルについても期待度数をもとめ、それらと観測度数の関係グラフを視察した。その結果、3つの1次交互作用をすべて含んだモデルはもちろん、3つのうちどれか2つの交互作用項だけを含むモデルでも、観測度数と期待度数の食い違いは小さいことがわかった(しかし、1つの1次交互作用項だけでは、明らかに系統的で大きな逸脱が生じた)。つまり、本データは「年度・分野」と「年度・国」あるいは「年度・分野」と「分野・国」のような2つの1次交互作用を想定したモデルとよく適合する。2次交互作用を想定しなくてもよい($u_{ijk}=0$)ということは、「変数 1,2 間の交互作用が変数 3 の全水準において同一であると仮定することに等しい(Everitt 邦訳書 p88)」。たとえば、「年度」「分野」「国」をそれぞれ、変数 1,2,3 とみなして、具体的に言い換えるならば、このことは『「年度と分野」の間には交互作用があるが、その交互作用はどの『国』においても等しい』ことを意味している。7か国のそれぞれに関する対応分析の結果が、どれもよく似た傾向を示していたことも、2次の交互作用を認めなくても良いという結論と一貫性を持つ。以上のことから、3元データを3つの2元分割表に分けて分析したことは、一応支持されるといえるだろう。

3.まとめ

「年度と分野」、「年度と国」、あるいは「国と分野」の間の連関は、Cramer's Vによると、それほど大きなものではない。しかし、年度の経過に伴って分野別比率・国別比率が変化していること、国によって分野比率に差があることは間違いない。グラフの視察から、分野においては(国を問わず)「金属」などの比率が減り、「管理」などが増えていること、国においては(分野を問わず)日本やフランスの比率が上昇して、USAなどが比率を落としていることが見て取れるが、分割表に対する対応分析の結果も、そうした結論を裏付けるものであった。対応分析はまた、グラフだけでは見落とされる変数間の関連を示している。今回の報告では、3元同時の分析はじゅうぶんにおこなわれていない。しかし、線形対数分析の結果などから、本データが持つ情報は、2元分割表を分析することで、ほぼ引き出すことができることも示唆された。

参考文献

- [1] 学術論文数の国際比較調査—理学・工学・医学分野の学術論文数の動向 1976-1993, 学術情報センター 1996.
- [2] Everitt, B.S. 山内光哉(監訳) 1980 質的データの解析. 新曜社.
- [3] 西里静彦 1982 質的データの数量化. 朝倉書店.
- [4] 大隅昇・L.ルバール・A.モリノウ・K.M.ワーウィック・馬場康維 1994 記述的多変量解析法. 日科技連.
- [5] 柳井晴夫 1994 多変量データ解析 理論と応用. 朝倉書店.

著者所属:

学術情報センター研究開発部

(R&D department, National Center for Science Information Systems)