

みんなくりポジトリ

国立民族学博物館学術情報リポジトリ National Museum of Ethnology

Introduction : Materials and Method : Computer Analyses

メタデータ	言語: jpn 出版者: 公開日: 2010-02-26 キーワード (Ja): キーワード (En): 作成者: 杉田, 繁治 メールアドレス: 所属:
URL	https://doi.org/10.15021/00003677

Ⅲ. コンピュータによる文化クラスターの分析

杉 田 繁 治*

- | | |
|-------------------|------------------|
| 1. 人文科学とコンピュータ | 3. 文化クラスター分析システム |
| 2. 計量分析とマッピングシステム | 4. 因子分析 |

1. 人文科学とコンピュータ

人文科学におけるコンピュータ利用はまだそれほど活発ではない。最近ではワープロとしての利用はかなり普及しているが、コンピュータによる情報処理を具体的な研究に適用した例は少ない。それはコンピュータの利用方法が文科系の人にとってむずかしいということだけではなく、実際問題としてコンピュータが活かせる例が見つげにくいということもあった。それは人文系の研究がどちらかといえば、計測された大量のデータから何かを引き出すというよりは、少しの事例からそれを生み出した背景を推測するということが重要であるという問題が多いからであろう。

それは情報処理というよりは分散している情報をつめて、そこから意味のある情報を引きだすための道具としての役割を求めているのである。つまり知的生産のための道具ということである。したがって人文科学の分野におけるコンピュータ利用の基本はデータベースとしての活用であろう。

あるいは生のままの形のデータからはよく分からない潜在情報をうまく引きだすために、それをいろいろに加工して利用者がデータの持つ意味をよりよく理解できるような形で表現する工夫が重要である。情報の可視化の問題が大きい。

従来のデータベースは主として文字で記述された、いわば書誌的情報のみを扱っていた。それに対して出来るだけ一次資料に近い形の情報に接近できるようにすることが大切である。歴史、文学、民俗などの研究にしてもテキスト全文をコンピュータに入力し、ある単語がどのようなコンテキストで使われているかを即座に検索することはコンピュータの得意とするところである。それを人手でやっていたのでは何年もか

* 国立民族学博物館第5研究部

かるであろう。

また人文科学で扱うデータは文字・数字、画像、音響など多岐にわたっており、しかもそれらが相互に関連して活用できることが望ましい。例えばインドネシアにおける民族音楽について知りたいとしよう。本や論文として記述されていることもあろう。それだけでは十分でなく、実際に演奏したり唄われたりしている場面のスライドや映画、ビデオなどの映像資料があれば理解はさらに深まる。その音響的特性の分析も必要であろうし、楽器があればその「もの」にも直接触れることも重要である。またその音楽に関連する儀礼や逸話など様々なことがらを調べたいであろう。マルチメディアのデータベースを構築することが必要である。

このようなデータベースでは自然言語による検索と、それをうまく活用するためのシソーラスが重要になる。資料の記述データと検索の時の入力表現が必ずしも一致しないからである。特にマルチメディアデータベースでは、異なる種類の情報に対して体系的な分類をするということは不可能に近い。図書の分類と標本資料の分類、スライドのインデックスなどを同一のカテゴリーで行なうことはできない。むしろ分類ということを放棄し、それぞれの資料が持つ性質を、自然言語によって自由に記述するというシステムのほうが、誰にでも分かりやすくまたデータ作製の観点からも望ましいのではなからうか。

人文科学におけるコンピュータ利用を一步進めるためには、エンドユーザーの立場からデータベースがどのようになっているのがよいかを追求することが必要である。ユーザーにコンピュータ操作の知識がなくても、ガイドメニューに従っていけば予備知識なしで利用できるようなインターフェイスの設計に工夫がいる。さらに異なった情報を同時に表示することが必要であり、分かりやすい出力の方法を考えることも大切である。

既存のシステムしかないものとあきらめて辛抱するのではなく、積極的に人文系の研究者にも使いやすいシステムとは何かを考え、それを要求していくことが情報処理の分野に技術革新を起させるきっかけになるのではなからうか。

2. 計量分析とマッピングシステム

コンピュータ利用というと、すぐに計量的分析ということが頭に浮かぶであろう。集積されたデータに対して統計処理をほどこし、データ全体のもつ意味をマクロにつかもうとする。統計的処理はその手順は大変明確であるが、個々のデータが持つ意味が

全体の中に埋もれてしまって重要な点をみすごしてしまうこともある。コンピュータによるデータ処理としては統計的性質のみではなく、個々の質的情報も研究者の必要に応じて見られるようにすることが重要であろう。表面的には見えない、あるいは個々のデータだけからは見えない全体像を異なった角度から見えるようにすることである。手作業では大変であるが、コンピュータをうまく活用することによってそれを可能にするのである。計量データでも単に数字の羅列でははっきりしないが、グラフを描くと一目瞭然ということがある。棒グラフがよいか、円グラフがよいか、あるいは折線グラフか、その他の表現方法を取るかは問題によって選択し工夫しなければならないが、データの表現方法によってデータの全体的性質が良く見えるようになることが多い。グラフ化は単純な操作であるが、さらにデータを加工することによってまた別の性質が浮び上がってくるということがある。

本クラスター分析においてそのグラフに相当するものはマッピングであろう。どの民族がどのような文化要素を所有しているかを地図の上に表示することによってその全体的な様子が直視できる。その分布の状態を見て民族間の関係や要素間関係などについて研究者が考えるための資料を提供する。なにか特殊な分布をしているとすればその理由は何かを考える。いろいろな文化要素の組合せについて、すぐにその分布状態が見えるというのがよい。ある仮説を持ちながら分布を見る。それによってさらに次の仮説を立てる。予想に反している場合には別の仮説を立てる。

このようにして対話的な方法でさまざまなケースについてすぐに結果が出るシステムはコンピュータでないとできない。数百の要素について予めいろいろな組合せの分布図を作っておくことは不可能に近い。コンピュータによる検索表示機能がなければできない問題であろう。またこのデータを使って単純な表示だけではなく、民族間の関係の深さや、要素間関係の深さなどを調べることも出来る。

3. 文化クラスター分析システム

このプロジェクト全体の流れは、図1に示すとおりになっている。コンピュータが関係するのは点線で囲まれた部分である。各々のブロックについて簡単に説明を加える。

1) ワークシートのデータをコンピュータへ入力

普通コンピュータへのデータ入力はキーボードを使って入力することが多いが、今回の場合は入力項目が各文化要素に対して「有無」のみであるから、出来るだけ直接

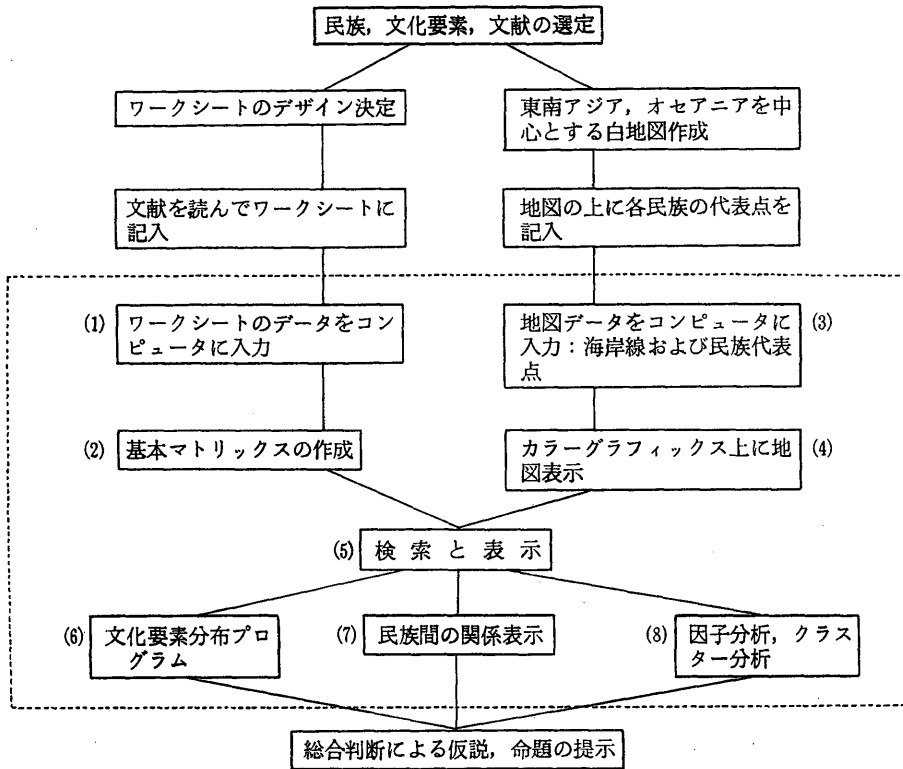


図1 クラスター分析の流れ図

的に、しかも誰にでもわかりやすい方法として、タブレットを用いる方法を開発した。タブレットにはカーソル（又はスタイラスペン）とよぶ指示器がついており、所定の場所に手で動かし、ボタンを押せばその座標値がコンピュータに入る。どの領域が何に相当するかはコンピュータで判定することができる。これを用いるとタブレット上にワークシートを置き、文化要素番号はタブレット上に記入された0～9の数字を指示することにより入力し、あとはシート上で文化要素の欄に有、無の記入があればそれをカーソル（又はペン）で、有の場合は1回、無の場合は2回、同一の場所で押さえるのみでよい。入力した結果は画面に表示され確認ができる。1冊の本に関するワークシートに対し入力時間は、15分～20分位である。

2) 基本マトリックスの作成

入力されたデータをもとにして、縦方向に民族名、横方向に文化要素を配置し、その交点に有無を記入したマトリックスを作成する必要がある。ここで問題となること

は、一つの民族に対して複数の文献から得られたデータがあるため、それをどのように扱うかである。つまりある民族に対しA文献では文化要素Mが「無」となっているのにB文献では「有」となっているケースも出現するからである。ここでは現在の状況、および過去の状況に対して次の基準でデータの集計を行なった。

すなわち、矛盾する記入があった場合はFという記号で表現する。基本マトリックスには各民族、各文化要素にたいして、現在の状況、過去の状況の二つの状態を記入している。ここに入り得る記号は、現在の場合：○（有）、×（無）、F（矛盾）、△（空白）、過去の場合：P（有）、Q（無）、F（矛盾）、△（空白）、である。空白は無記入の場合である。

3) 地図データのコンピュータへの入力

今回の場合、東南アジアからオセアニアまで広い領域にまたがるため、1枚の地図では表現がむづかしいので、2枚にわけることにした。一つは東南アジアを中心とし、他はオセアニアを中心とするものである。コンピュータ入力のための原図は縦横62 cm×87 cmの大きさであり、海岸線のみで表現された白地図である。大型タブレット上に地図をはり、1本の連結した線毎に番号をつけておき、カーソル（またはペン）でなぞっていく方式で入力した。同時に各民族の生活領域を代表する地点を点で入力した。東南アジアとオセアニアの地図の両方に含まれている地域の民族に対しては、どちらにも代表点が入れている。

4) 地図の表示

分布地図は、14インチのカラーグラフィック装置上に示される。画面上で見ただけでよいのか、結果をプリントアウトするかによって個々の表示を変えるようにしている。即ちこの装置のプリンターは白黒のみであるから、地図の陸地をカラーでぬりつぶすとプリントした場合に検索結果の点が判明しなくなる。そこで印刷する場合には海岸線だけの図形を用い、表示のみの場合は、陸と海を別々の色でぬりつぶし、検索結果の点を別のカラーで表示したものをを用いている。

5) 検索と表示

基本的な操作としては、文化要素の番号を入力し、それを有する民族の位置に印をつけた地図を出力することである。ここで問題となることは、2)の項で作成した基本マトリックスの表には、現在、および過去のデータが共存しており、また各種の状態を示す記号が入っている。それからどのようにして有の判定をするかが問題である。

解釈の仕方はいく通りかあろうし、それらをすべて試みることも可能であるが、今回は、下記の基準で判定することにした。つまり現在ははっきりと「有」となっているか、過去に「有」であった場合のみ「有」とする。現在は「無」となっているも過去に「有」であれば「有」とすることについては疑問もあろうが、この表の判定基準をいろいろに変えて試みることも、コンピュータでは容易であり、これによりクラスター分析の結果がどのように変化するかを見ることも面白い問題である。

6) 文化要素の分布

クラスター分析のための基本的なデータを表示するプログラムを開発した。これは文化要素番号を入力することにより、それがどこに分布するかを一覧するためのものである。現在の機能はいくつかの文化要素を同時に入力し、(a) それを共有する民族のみの位置を示すもの、(b) 各文化要素毎に記号を変えてそれを有する民族を表示することも簡単である。図2にサンプルを示す。

7) 民族間の関係表示

これは一つの例にすぎないが、アッサム・ビルマ地方の民族について、空間的な配置と文化要素からみた共有の割合を表示するプログラムである。まず画面上に民族の配置図が出る。そこでライトペンによってそのうちの任意の一つを指し示すと、その点を中心とした同心円が描かれ、その民族から見た、他の民族との相対距離が色の濃淡で表示される。これは単純に要素を共有する割合で評価した関係であって要素のカテゴリー別、あるいは基本マトリックスの使い方などによって、様々な角度からの距離を表示することが出来る。結果のプリントも出来るようになっている。図3に出力例を示す。

8) クラスター分析

基本マトリックスを使って民族間、文化要素間のクラスター分析をおこなった。各要素間の相関係数としては2値データにたいする標準的なフェイ係数(ϕ)を採用した。これは下に示すように、二つの事象I、IIに対する相関を取る場合に、両者において共に1となっている項目の数をaとし、Iにおいては1であるが、IIにおいては0となっている項目の数をb、その逆の場合をc、およびI、IIにおいて共に0となっている項目の数をdとする。

例えば次のデータの場合には

I	1 1 0 0 1 0 1 0 0 0 0 1 1 1 0 1 0
II	1 0 0 1 1 1 0 0 0 1 1 0 1 0 1 0 0
a=3	b=5
c=5	d=4

	II		
I		1	0
	1	a	b
	0	c	d

となる。

その時、IとIIとの相関係数は次の式で与えられる。

$$\frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

このデータを用いてクラスター分析を行なった。途中のクラスター間の距離の評価に関しては最遠隣法とかいくつかの方法がとれるようになっている。結果は樹状のグラフとして表示されるのでたいへん見やすくなっている(図4)。

4. 因子分析

民族や文化要素間に何か共通のファクターがあるかもしれないということで、基本マトリックスを使って機械的に因子分析を施してみた。使用した方法は最小残差法である。この結果を2次元的に表示したものが図5である。いろいろな軸の組合せを表示させることが出来る。この分析結果が意味を持つかどうかは専門家の解釈によるが、データの隠れた一面をあらわにするという意味では一つの武器になろう。



図3 アッサム・ビルマにおける諸民族間の距離の例示
 (図では、ビルマを中心として19民族との距離をしめしてある)



図4 クラスター分析の結果をツリーで表現したもの（一部）

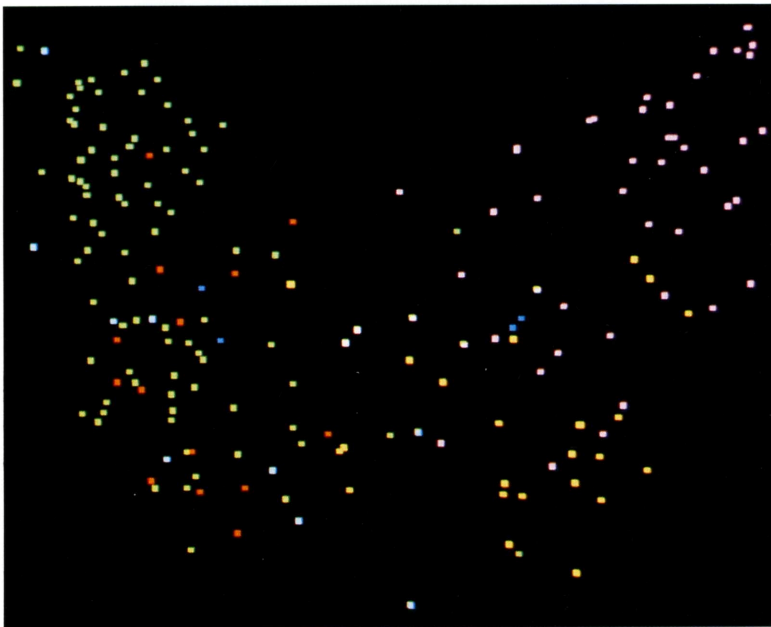


図5 因子分析の結果の表示：因子2と因子4の関係



写真1 コンピュータによる検索



写真2 ワークシートのコンピュータへの入力