

# 数値的地域区分法:NTAP

統計数理研究所 大 隅 昇  
日本アイ・ピー・エム 渋谷 政 昭

(1977年8月 受付)

## Numerical Techniques for Areal Partition: NTAP

Noboru Ohsumi

(The Institute of Statistical Mathematics)

Masaaki Sibuya

(IBM Japan)

By NTAP (Numerical Techniques for Areal Partition) we mean the procedures for data analysis of remotely sensed measurements, that is, multivariate observations at each mesh area within a region.

In this report we discuss the scope and objectives of NTAP, possible procedures and algorithms for the basic stage of NTAP, and the design policy of an experimental subroutine package NTAP-I.

Specially considered are 1) the way to specify some small areas with typical features, and how to aid the user to select them, 2) the procedure for partitioning into more connected and not scattered subregions.

To evaluate the performance of the algorithms and the package, artificial data similar to actual one and of specified characteristics are generated. The results of processing these data verify the achievement of NTAP-I and suggest the way of its better usage.

### 〈目 次〉

1. ま え が き
2. NTAP の目標
3. NTAP-I システムの機能と構成
  - 3.1 デ ー タ
  - 3.2 区分の方略
  - 3.3 画素の分類
  - 3.4 システムの構成
4. 人工データによる実験
  - 4.1 人工データの作り方
  - 4.2 人工データによる実験
    - 4.2.1 実験 (その1)
    - 4.2.2 実験 (その2)
    - 4.2.3 ISODATA による自動分類
    - 4.2.4 細分類について
    - 4.2.5 プログラムと計算効率
5. む す び

### 1. ま え が き

宇宙開発の一環として、高空より地表(地上・水上)の諸様相を観察・測定する「リモート・

センシング」の技術が発展しつつある。これは、航空機または人工衛星より、諸周波数帯の可視・可視外光線、熱線、電磁波のセンサーを用いて、地表からの反射・輻射を「撮影」し、広い領域の植相、汚染、土地利用、地質、などの状況を効率よく得ようとするものである。森林、活火山、砂漠、極寒地、海面など人間が近接し難い地域の調査、広大な領域の全体像の把握に適している。

リモート・センシング技術は、計測、画像処理、データ処理、データ解析、目標の実体と測定データ間の関係に関する知見、などから構成される<sup>5),6),16),19),20)</sup>。本論文ではこのデータ解析の部分における1つの主要課題である「対象地域をいくつかの類似の領域に区分する」という諸応用に汎用の解析法を扱う。ただし原データについて計測上の偏りの補正は、すでに行なわれていることを前提として、ここでは議論しないが、当然これは重要であり、目的によってはデータ解析と絡ませて扱うべきである<sup>16),20)</sup>。

リモート・センシングでは、2次元平面（といっても高低を扱うことがあるが）を格子状に細分・量子化した「画素(pixel)」ごとに、いくつかの周波数帯(channel, spectrum)の反射・輻射波の強度を測定する。つまり、基本データは、非常に多くの2次元格子の各点で観測される数次元の変量特性値から成っている。

平面上の座標を多変量特性値の一部とみなせる場合もあるだろうが、多くの場合に、位置は特殊な情報であろう。データ量はもちろん、対象領域の広さと測定の詳しさによるが、実用的な絵を扱うには少くとも200×300程度、つまり数万画素が必要である。

さらに各次元の特性値は数ビットであり次元が数個あるから、データ量は少な目に見積って数十万バイトから百万バイトとなる。なお既に述べたように、用いるデータは各種の補正を加えた結果のものである、ということも場合によっては問題となる。

対象領域を類似領域に区分することは、本質的には判別分析、クラスター分析<sup>10),11),18)</sup>の仕事であるが、上に述べたデータの基本特徴のために改めて議論するわけである。ここで扱う手法はリモート・センシング以外でも、いわゆるメッシュ・データの計量地理学的な分析、農業用土質調査に基づく土地区分、金属面の顕微鏡写真による分析、などに適用できよう。そこで、「画素に対応する多変量特性値による地域区分のためのデータ解析技法」を「数値的地域区分法(Numerical Techniques for Areal Partition: NTAP)」と名付けることを提案する。

この論文は、第2章でNTAPにおいて考察すべき目標を議論し、第3章で目標の第一段階を達成するためのソフトウェア、NTAP-Iの設計方針とアルゴリズムを示す。第4章ではこのようなソフトウェアの性能を確かめるとともに、NTAPの基本性質を明らかにするために人工データによる実験結果を報告する。

リモート・センシング技術は多方面で利用され、その目的に応じて最適技法、数値解析、統計的推測、データ縮約、シミュレーション、などの手法を適用しなければならない<sup>2)~4),16)</sup>。NTAPはむしろ、詳細な解析の前段階として、データの潜在的な特性を大づかみに把握する道具である。そのために、解析結果を利用者に理解しやすい形で表現することと、詳細解析との連結を計ることが必要となる。

なお、NTAPの目標設定の過程では、奥野忠一、田栗正章、岩坪秀一、矢島敬二、の諸氏に参加していただいて討論した。人工データによる実験の提案とデータ生成は渋谷が行ない、ソフトウェア作成の大部分は大隅が行なった。

## 2. NTAPの目標

対象とする地域の夾雑さと処理の特性により、地域区分の作業目標を次の4つの難易段階に

分け、それぞれについての仕事の内容を考える。

(1) 航空機からのデータにより、広い平野部の耕地を農作物の種類や生育によって分類したり、森林を樹木の種類により分類するような場合、分類結果の各領域がある程度大きく、しかもその中がかなり一様になると期待できる場合である。

(2) 人工衛星からのデータにより、都市、耕地、森林などに大まかに分類したいような場合、(1)と同様に、各領域はある程度大きい、その中はやや夾雑である。

(3) 道路、河川、活断層のような線状の対象物を領域から選び出したり、濃厚汚染地域を発見したりする場合、つまり、特定の様相をもった小目標を領域から選出する場合である。

(4) 均一な部分領域がほとんどないが、混合状況によって地域を区分したい場合で、都市部の航空写真で、緑地、住宅地、高層ビル街、道路などに区分したり、金属組織の顕微鏡写真を成分別に区分したりする場合である。混合状況の違いは、肌理(きめ、texture)という言葉で呼ばれている。

本稿では、当面の課題として(1)、(2)を取り上げる。(3)の場合には対象の幾何学的形状を仮定したり、周辺より際立たせる特性を指定する必要がある、汎用化がやや困難である。

(4)の場合は重要であり、より多くの手間を必要とする、困難ではあるが興味ある課題である。しかしながら、パターン認識の側面がより重要となり、やはり一般的手法で成功するとは思えない。

地域区分の仕事は、境界の線引きと、区分結果の表示のための「色分け」につきるとは限らない。すべての地点をどこかの区分に割当ててのではなく曖昧な点は曖昧のまま残して利用者の判断に委ねることが考えられる。市街地、森林、農地、などと大別した上で農地だけを作物によりさらに細分したい場合が考えられる。また農地だけを、湿度のような1次元の特性値の大小によって濃淡図に塗りたいこともある。

区分の判断の仕方により NTAP のアルゴリズムは2つに大別される。それは、対象地域内に典型小領域を事前に定義できるかどうかの違いによる。つまり、第1は何等かの事前の知識によって、地域内のいくつかの小部分について、これらはA地域、これらはB地域、……、と指定できる場合である。この場合各種の地域は、なるべく散らばせて広く取れるに越したことはない。典型小領域が与えられたならば、この中の画素のデータについて多変量解析を行ない、それを利用して残りの画素について分類を行なう。その詳細は次章で述べるが、要点の1つは、NTAP の特徴である座標データを積極的に利用するかどうか、利用するならば、どのようにするかである。

次に事前に信頼できる情報がない場合には、典型小領域を選ぶようなことはできずデータだけから分類を行なうことになる。この場合を一般に自動分類と呼んでいるがクラスター分析に他ならない。自動分類の手法としてはいろいろあるが、ここでは、大量の多変量データを効率よく分類できること、他のリモート・センシングの解析で広く利用されていること、などの理由から、ISODATA (Iterative Self-Organizing Data Analysis Techniques-A) を採用した<sup>9),12),13),17),22)</sup>。大量のデータを扱えるという点では、他にクラスター内分散あるいはクラスター内平方和の最小化をはかる方法、たとえば、k-means 法、Ward 法などが考えられるが、ISODATA はこうした基本的な手法の諸機能を包括した上に、クラスター内のデータ数(クラスター・サイズ)の制限、クラスターの個数の制御機能、クラスターの併合・分離の反復機能などを備えている点に特徴がある。ISODATA の弱点として、事前に与えるパラメータが多いために、計算結果の判断・解釈がやや難しいということが挙げられる。こうした不備を補うために、実際のプログラム作成にあたって、パラメータの変更が容易であること、同一データの反復利用が可能であること、結果を図・表で表示すること、などの工夫を要する。

### 3. NTAP-I システムの機能と構成

1976年春より稼働しているソフトウェア・システム，NTAP-Iの機能と構成を，3.1データ，3.2区分の方略，3.3画素の分類，3.4システムの構成，に分けて説明する。

#### 3.1 データ

まず対象とするデータはMSS (Multispectral Scanner) により記録され，適当に画像処理を経た離散データである。これは長方形領域を区切った長方形格子の上に置かれた多変量特性値であり，3次元配列のデータであるが，記録順に1次元配列に展開した逐次ファイルとしたものを原データとして用いる。

MSS各チャネルのデータは7ビットで区間 $[0, 255]$ の整数値であり，上記の配列の要素は1バイト整数の配列である。たとえば人工衛星ERTS (Earth Resources Technology Satellite) の場合，12の周波数帯での測定が行われているが，それぞれの利用のために有効なのは3~6個のチャネルである。

前述のように，1回の画像処理の対象とする画素数は，処理の目的と複雑さ，機械の能力に依存するが，ブラウン管にやや粗い画面を表示するのに必要な $240 \times 320 = 76800$ 画素が少な目の標準的な大きさであろう。NTAP-Iでも，この大きさを1回の計算処理の標準的な単位としている。

#### 3.2 区分の方略

利用者の広汎な分析目的，あるいは解析対象目標の様相の複雑性を考慮し，NTAP-Iでは，次のように区分の方略を考える。

〔方略 (I)〕 典型小領域を選出できる場合

比較的一様あるいは等質とみられる領域，すなわち典型領域，の代表として典型小領域を，典型ごとに複数箇所指定・抽出できるとする。この場合，さらに次の2つに分けて考える。

- (a) 典型小領域の統計解析から得た情報をもとに，すべての画素をもっとも類似した典型領域にわりあてる。そして，区分 $\{A, B, C, \dots\}$ を作る。
- (b) (a)と同様であるが，まず $\{A, B, C, \dots\}$ と大まかに区分した後に，どれか特定の領域，かりにAとする，に注目して，これを $\{A_1, A_2, A_3, \dots\}$ と詳しく区分する。この前者を「大分類」，後者を「細分類」と名づける。

〔方略 (II)〕 均一な典型小領域の選出が困難であるような場合には，典型小領域に相当する領域を作り出すことを考える。

- (a) 利用者にとって興味ある特徴を含むと思われる複数個の領域を指定する。この領域のデータをすべて合わせたデータを作り，そのデータを用いて自動分類の方法によって選出した部分についての仮りの地域区分を行う。この結果を観察して典型小領域の指定を行ない，あとは〔方略 (I)〕に従って，全領域を区分する。
- (b) 全領域を自動的に区分する。得られた分類結果を参考に典型小領域の選出を行う。

#### 3.3 画素の分類

典型小領域のデータの統計解析から得た情報にもとづいて全画素の分類を行う。分類の基本は最近隣方式 (Nearest Neighbor Rule) による。そしてNTAP-Iの特徴は，画素の座標情報を取り入れるか否かの選択を許し，それにより分類に使用する距離と分類の方式を変える点にある。

(1) 分類に使用する距離について

ある画素のデータを $z = (u, v; x_1, x_2, \dots, x_p)$ で表わす。ここで， $(u, v)$ は座標を示し，

$\mathbf{x}=(x_1, x_2, \dots, x_p)$  は  $(u, v)$  上に与えられる  $p$  次元の多変量特性値である。このとき分類の距離として次の2種類を考える。

いま、特性値の空間で、ある  $\mathbf{x}$  からある典型  $k$  までの距離を  $d_k$  とすると、2つの距離、マハラノビス距離、 $l_1$  ノルムは次のように書ける。

マハラノビス距離

$$d_k = (\mathbf{x} - \mathbf{m}_k)' W_k^{-1} (\mathbf{x} - \mathbf{m}_k) \quad (3.1)$$

ここで、 $\begin{cases} k = A, B, C, D, \dots \text{ (典型)} \\ \mathbf{m}_k = \text{典型 } k \text{ の特性値の平均ベクトル} \\ W_k = \text{典型 } k \text{ の特性値の標本分散・共分散行列, } W_k^{-1} \text{ は } W_k \text{ の逆行列} \end{cases}$   
 規準化した  $l_1$  ノルム

特性ごとの標準偏差で規準化した  $l_1$  ノルムを考える。

$$d_k = \sum_{i=1}^p |x_i - m_{ki}| / \sigma_{ki}$$

ここで、 $x_i$  は  $\mathbf{x}$  の第  $i$  成分

$m_{ki}$  は  $\mathbf{m}_k$  (典型  $k$  の平均特性値ベクトル) の第  $i$  成分

$\sigma_{ki}$  は  $\sigma_k$  (典型  $k$  の特性値の標準偏差ベクトル) の第  $i$  成分

(2) 画素の分類方式

画素の分類は次の手順により行う。

(手順1) 判定の閾値  $\delta$  を与える。続いて、 $d^* = \min_k d_k$  となる典型  $k$  を探し、

$d^* \leq \delta$  のとき、 $\mathbf{x}$  は典型  $k$  に所属と判定、

$d^* > \delta$  のとき、使用した距離が上記2つのいずれであるかにより、判定は次の2通りに分かれる。

a) マハラノビス距離のとき、 $\mathbf{x}$  は「未分類」と判定する。

b)  $l_1$  ノルムのとき、 $\mathbf{x}$  の画素の座標  $(u, v)$  を特性値として加えたうえで、次の再分類の手順により、いずれの典型に所属するかを判定する。

a) の場合には、いずれの典型にも所属しない画素が未分類として残る (あるいは未分類という典型に所属すると考えてもよい)。一方、b) の場合には、 $\mathbf{x}$  は必ずいずれかの典型に所属が確定する。

(手順2) 再分類の方法

$$d^{**} = \min_j \{ \min_k (d_k + \Delta_{kj}) \} \quad (3.2)$$

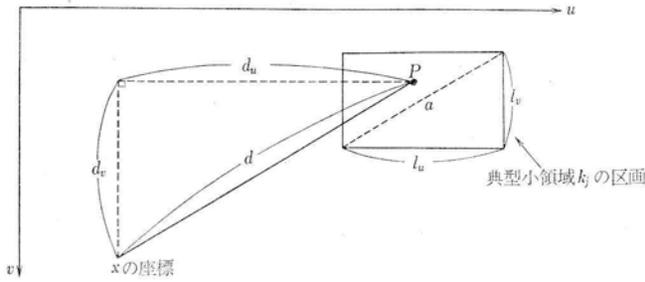
ここで、 $\begin{cases} k = A, B, C, \dots \text{ (典型)} \\ k_j = \text{典型 } k \text{ の第 } j \text{ 番小領域} \end{cases}$

となる最小の  $d^{**}$  を探し、 $\mathbf{x}$  はそのときの典型  $k$  に所属すると判定する。ここで  $\Delta_{kj}$  は座標情報を取り入れるための補正項で、次の2つが用意されている。

$$\Delta_{kj}^{(1)} = \frac{d_u}{l_u} + \frac{d_v}{l_v} \quad (3.3)$$

$$\Delta_{kj}^{(2)} = \frac{d}{a} = \frac{d}{\sqrt{l_u^2 + l_v^2}} \quad (3.3)'$$

式各記号は (図1) に示す通りである。図中の典型小領域の代表点  $P$  は、小領域  $k_j$  の区画内に適当に置く。とくに指定がない場合には、領域の重心が標準値として自動的に算出・指定される。また、 $\Delta_{kj}^{(1)}$ 、 $\Delta_{kj}^{(2)}$  のいずれを使用するかは利用者が任意に指定する。



(\*) 代表点 \$P\$ は領域 \$l\_j\$ 内に適当に指定する。指定しなければ重心が自動的に選択される。

図 1

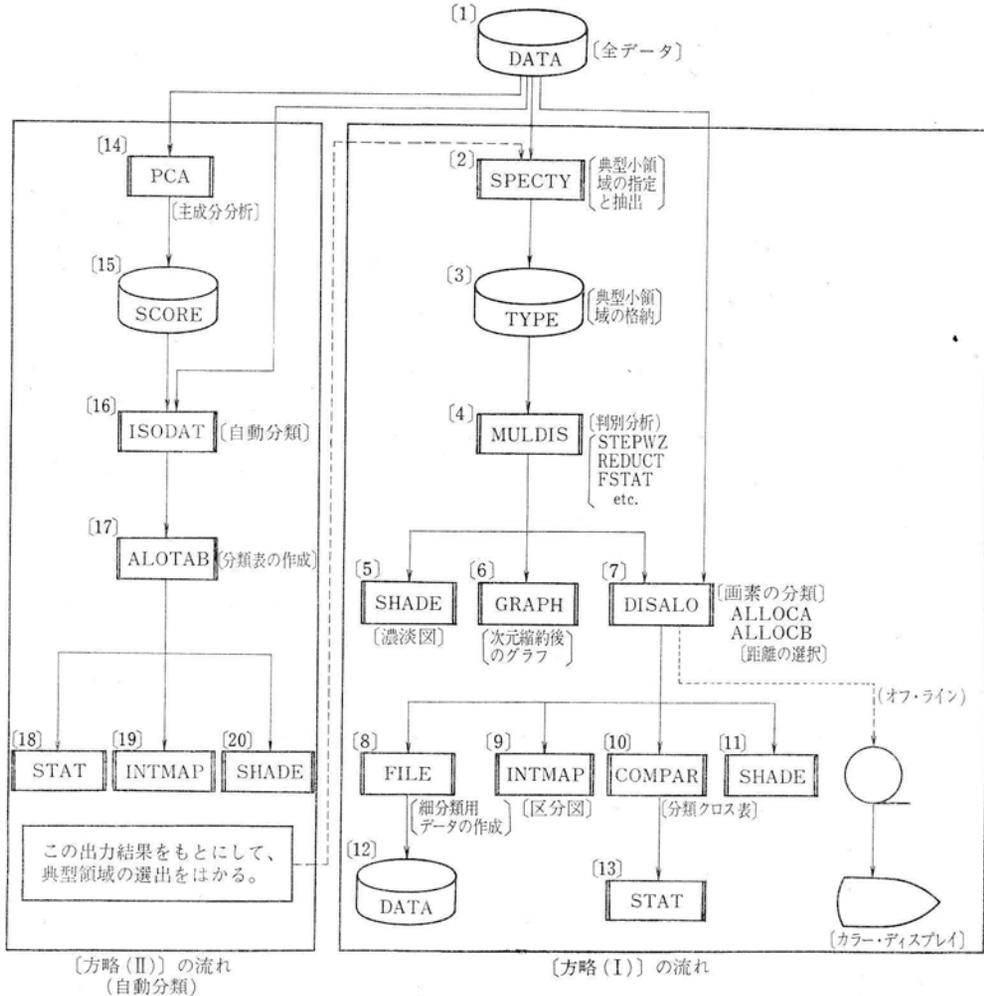


図2 NTAP 流れ図

### 3.4 システムの構成

(図2)は NTAP-I のシステム構成を示す流れ図である。この流れ図に沿って、機能の概要を説明する。なお、流れ図の各単位に [ ] で箱番号をつけた。箱番号 [2]~[13] は前述の [方略 (I)] に、[14]~[20] は [方略 (II)] に、それぞれ対応している。

#### (1) [1]: DATA

計算対象となる全データが格納されたファイルである。

#### (2) [2], [3]: SPECTY, TYPE

典型および典型小領域の指定・抽出を行う。区分対象とする領域の種々の事前情報 (たとえば、地表対照物 (ground truth) の情報、画像情報など) にもとづいて、代表的な、またはほぼ等質とみられる領域を指定する。指定領域の形状は長方形とし、指定領域の重なりは許されない。また1典型につき複数個の小領域を指定できる。抽出は、1つの小領域ごとに、その典型の番号と、長方形の四隅の座標を指定するだけでよい。指定によって全データの格納ファイル (DATA) から必要個所のデータを取り出し、新たに典型小領域だけのデータからなるファイル、TYPE を作成する。同時に、次の処理に必要なパラメータ類 (典型および典型小領域ごとのデータ数、典型小領域の代表点  $P$  の座標など) をカードまたはファイルに出力する。

#### (3) [4], [5], [6]: MULDIS, SHADE, GRAPH

ファイル TYPE に格納された典型のデータを使って、判別分析を中心とする統計解析を行う。計算内容として、

- i) 典型、典型小領域ごとの初等統計量 (平均ベクトル、標準偏差、分散・共分散行列、相関行列など) の算出。
- ii) 次元縮約を伴う判別分析。
- iii) 変数選択による特性値の吟味。
- iv) 判別得点の図示。
- v) 各種検定 (等平均、等分散の検定など)。

などが挙げられる。この他、通常の判別分析の機能はほとんど備えている。サブプログラム MULDIS は文献 [14] を参考に手直したものを作成、使用している。

計算結果のうち、図的表示が可能なものについては、濃淡図 (たとえば図 6, 8, 9) や、判別得点図などにより表示する。SHADE, GRAPH はこの機能をはたすサブプログラムである。

#### (4) [7]: DISALO

前述の 3.3 画素の分類の分類方式に従って全画素を分類する。NTAP-I の最も重要なサブプログラムの1つである。ALLOCA はマハラノビス距離、ALLOCB は、 $t_1$  ノルムを使用する場合のサブプログラムである。DISALO により分類した結果は、一時ファイルに格納し、これをもとに以下のサブプログラムで、分類の状況、結果などを表示する。

#### (5) [8]: FILE

「細分類」の指定を初めに与えておくと、この作業に必要なデータを作成する。利用者が分類対象として取り上げた典型のうち、とくに関心があって細分類の対象としたい典型、かりに  $A$  とする、の指定を行うと、全データの分類後、典型  $A$  に所属すると判定されたデータの全情報を細分類ファイル DATA (箱番号 [12]) に格納する。これを改めて箱番号 [1] の DATA と同じように扱って、[2] 以下の分析を行えば細分類の結果が得られる。

#### (6) [9], [10], [11]: INTMAP, COMPAR, SHADE

観察を容易にするため、分類結果はできる限り図・表により出力表示する。サブプログラム INTMAP は分類結果の地域区分図を作成する (たとえば図 12)。これは、各画素がどの典型と判定されたかという名札情報の区分図である。一方、サブプログラム SHADE は [5] と同

様に濃淡図を出力する。このとき、区分領域の特徴を見易くするために、典型ごとに濃淡図を分離・表示することも可能である。サブプログラム COMPAR は指定した典型領域と分類後の区分領域の関係を分類クロス表として作成、表示する。利用者は、この表をもとに、分類結果の典型ごとの画素の構成比率、未分類の状況、座標情報を考慮したデータの割合などを一覽できる。

(7) [13]: STAT

分類により得られた区分領域について、典型別、典型小領域別、座標吟味の有無別などの属性によりデータを層別した上で各種統計量を算出する。

以上の [2]~[13] は [方略 (I)] に対応する諸機能である。一方、次に述べる [14]~[20] は [方略 (II)] に対処する自動分類の機能である。

(8) [14], [15]: PCA, SCORE

現在稼働中の NTAP-I では、全次元 (3~6 次元) の特性値をすべて取り入れて、全データ (76800 画素) を自動分類することは難しい。そこでまず、主成分分析により次元縮約をはかった上で、求めた主成分得点を新たなデータとして自動分類を行う。このとき指定により任意の 2 主成分が自動分類に利用可能である。サブプログラム PCA で主成分分析の計算を行い、指定された主成分得点はファイル SCORE に格納される。

(9) [16]: ISODAT

ISODATA によるクラスタリングを行う自動分類の中核となるサブプログラムである。主成分分析の得点を利用せずに直接原データを対象に分類を行う場合には、関心のある領域を中心に、ある区域を指定・抽出する。指定箇所は 1 箇所である必要はなく、領域の重なりがなければ何箇所選んでもよい。ただし、画素の総数は、パラメータの与え方にもよるが、6 次元データの場合を例にとると、現用大型計算機の能力では 3 万個程度に限定される。こうして指定された領域のデータを、まとめて 1 つのデータ・セットと考えて、これを自動分類する。

(10) [17]: ALOTAB

抽出箇所として指定した領域に、仮の典型番号をあらかじめ与えておくと、これと、作られたクラスターとの関係をクロス分類表として出力表示する。指定がない場合、単にクラスターの情報だけを表示する。

(11) [18], [19], [20]: STAT, INTMAP, SHADE

作り出した仮の典型、すなわちクラスター、の性質を数量的に把握するため、クラスターごとの統計量を算出する。さらに、実際の区分を観察するため、区分図、濃淡図を作成する。

(12) 反復機能について

ISODATA に限らず、クラスター分析の欠点として、初期パラメータの選び方に明確な規準がないこと、1 回の分類結果だけで判断することは危険で試行錯誤により何回かの分類を繰り返す必要があること、などが挙げられる。この欠点を補うために、計算の初めに複数組のパラメータ群をまとめて与えると、[16]~[20] の計算手続きを、同じデータ・セットに対し、指定した組数だけ繰り返し計算し、結果を順次出力する。利用者はこの結果を比較、観察して、適当なものを選べばよい。

以上が NTAP-I の主な機能であるが、これが意図したように働くかどうか、またプログラムが正しく機能するかどうかの検証を行うためには、構造の知れた、しかもなるべく現実のデータに近い人工データを作成し、これを使用した実験が必要となる。

4. 人工データによる実験

4.1 人工データの作り方

現実のデータに似せた人工データを生成するためには、まず、現実のものが各画素の上に置かれた多変量特性値であることを考えに入れることが必要となる。また、区域の特徴として現実には、明確な境界をもった区域が存在する場合、境界あるいは区域と呼べるものがあいまいな場合、などが考えられるので、この点も考慮に入れたい。

このため、a) 明確な境界をもつ区域が複数個あり、しかも各区域の特性値が適当にばらついている場合、b) 区域があいまいなうえに、特性値がばらついている場合、を想定し、2種類の人工データを生成した。前者を「ブロック・データ」、後者を「ファジィ・データ」と名づける。それぞれ、生成方法は次の通りである。

まず3次元特性値からなる4つの区域 A, B, C, D を考え、4つの区域のうちのすべての2区域間の誤分類の確率をある程度任意に与えて、対応する区域の平均ベクトルの配置を作る。ある程度任意とことわったのは、平均ベクトル間に距離の公理を保つ必要があるためである。たとえば、区域 A と B の間の誤分類の確率を  $P(B|A) = P(A|B) = 0.19$  と与え、各区域が分散が1の正規分布と仮定すると、

$$2(1 - \Phi(x)) = 0.19,$$

$$\text{ここで、} \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

から、AB 間の距離、 $2x \approx 2.62$  を得る、A, C 以外の4つの対に対してもそれぞれ誤分類の確率を与え同じようにして、それらの間の距離を計算する (図3)。次に  $\overline{BD}$  (x 軸) を軸として、三角形 ABD を回転させ、 $\overline{AC} = r$  が、指定した誤分類の確率に対応する距離となるようにする。いま (図4) のように A の座標が  $(-1.6, y, z)$  の位置にあるとして、図中の記号を使うと、次の関係がある。

$$\begin{cases} a^2 + (b+y)^2 + z^2 = r^2 \\ y^2 + z^2 = c^2 \end{cases}$$

これを y について解くと、

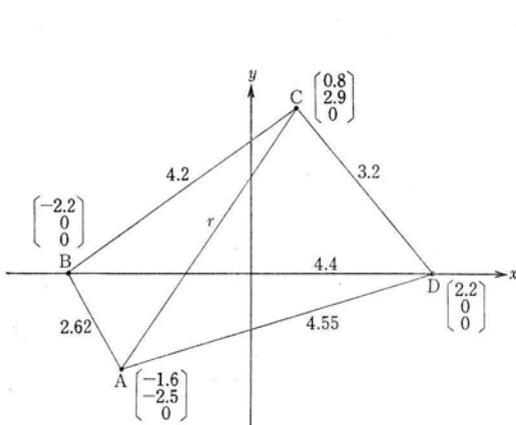


図 3

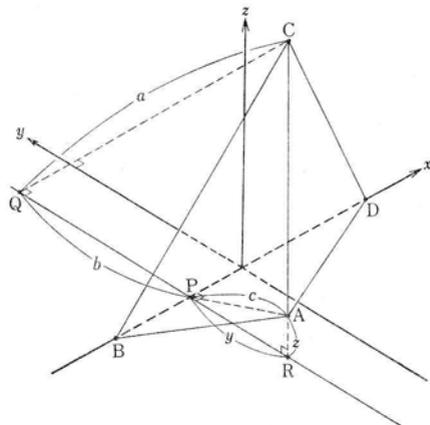


図 4

$$y = \{r^2 - (a^2 + b^2 + c^2)\} / 2b$$

となる。  $a=2.4$ ,  $b=2.9$  のとき,  $c=2.5$  となるので,  $P(A|C)=P(C|A)=0.01$  と与えて,  $r=5.2$  を得る。これらの値を上式に代入して,  $y=1.1$ , したがって  $z=2.3$  を得る。よって各区域の平均ベクトルとして, 次の値を得る。

$$\mu_A = \begin{bmatrix} -1.6 \\ -1.1 \\ 2.3 \end{bmatrix}, \quad \mu_B = \begin{bmatrix} -2.2 \\ 0.0 \\ 0.0 \end{bmatrix}, \quad \mu_C = \begin{bmatrix} 0.8 \\ 2.9 \\ 0.0 \end{bmatrix}, \quad \mu_D = \begin{bmatrix} 2.2 \\ 0.0 \\ 0.0 \end{bmatrix} \quad (4.1)$$

念のため, 上の値から平均間距離, 誤分類確率を求めると, (表1)-(2)となる。(表1)-(1)の初めに与えた数値と若干異なるが, これは数値の丸めによるものである。

表1 エグデータの平均間の距離と誤分類の確率

(1)

区 域	A	B	C
B	2.62 (0.19)		
C	5.20 (0.01)	4.20 (0.04)	
D	4.55 (0.02)	4.40 (0.03)	3.20 (0.11)

左側: 平均間距離 右側 ( ) 内: 誤分類確率

(2)

区 域	A	B	C
B	2.62 (0.1902)		
C	5.20 (0.0093)	4.17 (0.0366)	
D	4.58 (0.0220)	4.40 (0.0278)	3.22 (0.1074)

左側: 平均間距離 右側 ( ) 内: 誤分類確率

(\*) (1) は初めに目安として定めた数値

(2) は実際に出来上がったデータからのもの

なお, 4分類(4区域)以下で, 4次元以上のデータを考える場合には, ゼロ成分を上記の平均ベクトルに追加するだけでよい。ここで, 各特性間に適度の相関を与えたほうがより現実的であるから, 次のように適当な線形変換を行う。

$$\mathbf{x}_i = L(\boldsymbol{\mu}_i + \mathbf{e}) \quad (i = A, B, C, D) \quad (4.2)$$

ここで,  $\mathbf{e}$  は独立な単位正規乱数  $N(\mathbf{0}, I)$  である。

( $\mathbf{0}$  はゼロベクトル,  $I$  は単位行列)。

いまここでは, 分散・共分散行列は4つの典型に共通で,

$$V = LL' = \begin{bmatrix} 1.44 & 0.72 & 0.36 \\ 0.72 & 1.00 & 0.50 \\ 0.36 & 0.50 & 1.94 \end{bmatrix}$$

とした。

したがって, 相関係数行列  $R$  は,

$$R = \begin{bmatrix} 1.000 & 0.600 & 0.215 \\ 0.600 & 1.000 & 0.359 \\ 0.215 & 0.359 & 1.000 \end{bmatrix}$$

となり, 特性値の間に, やや相関の高いものから, それほど相関がないものまでを適当に配置したことになる。

ブロック・データの作り方

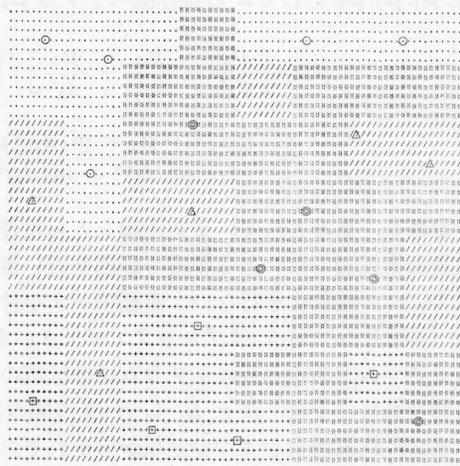
上の関係を利用して、まず明確な境界をもった区域を想定したブロック・データを生成する。そのため、対象の画面を4つの区域に対応させて適当に区分し、ブロック状の領域を作る。各区域の配置は(図5)のように与え、区域ごとに画素を、(表2)のようにわりあてる。

次に式(4.2)により求めた  $x$  を実際の7ビットのデータにあわせて、 $[0, 255]$  内の整数となるように変換する。

$$y_i = [10x_i + 128] \quad (i = A, B, C, D) \quad (4.3)$$

(ここで  $[ ]$  は整数部を示す)

(4.3) により生成したデータを、それぞれ4つの区域に対応させる。そして、各画素の



(\*) 記号は次の典型に対応する

$$1 \cdot = A \quad | \cdot = B \quad + \cdot = C \quad * \cdot = D$$

図5 ブロックの配置とその代表画素の位置

表2 ブロックに配置された画素数と構成比率の表

区域	画素数		構成比率 (%)
	データ (I)	データ (II)	
A	660	13,200	39.06
B	840	16,800	
C	720	14,400	60.94
D	1,620	32,400	
計	3,840	76,800	100.00

座標ベクトル  $(u, v)$  を  $y_i$  につけ加えた上で5次元のベクトル  $(u, v; y_i)$  とし、これを対象のデータとする。なお、画素数が3840個と76800個の2種類のデータ、(I)、(II)を用意した。後者のデータ(II)は前者のデータ(I)の1画素を縦を5倍、横を4倍することで得られる。また整数化後の平均ベクトルは、およそ次のようになる。

$$\mu_A^* = \begin{bmatrix} 108 \\ 109 \\ 148 \end{bmatrix}, \quad \mu_B^* = \begin{bmatrix} 101 \\ 114 \\ 121 \end{bmatrix}, \quad \mu_C^* = \begin{bmatrix} 137 \\ 156 \\ 142 \end{bmatrix}, \quad \mu_D^* = \begin{bmatrix} 154 \\ 141 \\ 134 \end{bmatrix} \quad (4.4)$$

フuzzy・データの作り方

ブロック・データの区配置を利用して、区域の境界をより一層あいまい化する。まず、各区域の代表点(「代表画素」と名づける)を各区域ごとに複数個選ぶ。そして同一領域内の代表画素には、その対応する区域の平均ベクトルを与える。代表画素以外のすべての画素に対しては、代表画素間の特性値が滑らかに変化するように、代表画素の特性値の重みつき平均をとる。つまり、ある画素から最も近い3個の代表画素を選び、選んだ画素の上に置かれた特性値ベクトルを、その画素から3つの代表画素までの平方距離に反比例する重みをつけて平均する。いま、ある画素の特性値  $x$  からもっとも近い3個の代表画素をそれぞれ  $i, j, k$  とし、これらの各点からその画素までの距離および、代表画素の平均ベクトルを、それぞれ、 $d_i, d_j,$

$d_k; \mu_i, \mu_j, \mu_k$  として  $x$  の特性値ベクトル  $\mu_x$  を,

$$\begin{aligned} \mu_x &= \left[ \frac{\mu_i}{d_i^2} + \frac{\mu_j}{d_j^2} + \frac{\mu_k}{d_k^2} \right] / \left[ \frac{1}{d_i^2} + \frac{1}{d_j^2} + \frac{1}{d_k^2} \right] \\ &= \frac{\{(d_j d_k)^2 \mu_i + (d_k d_i)^2 \mu_j + (d_i d_j)^2 \mu_k\}}{(d_j d_k)^2 + (d_k d_i)^2 + (d_i d_j)^2} \end{aligned} \quad (4.5)$$

で定める。重みとして、平方距離をとった理由は、代表画素の間を滑らかに変化させ、しかも代表画素近辺では、ある程度の範囲を近い値に保つためである。この  $\mu_x$  に  $N(\mathbf{0}, V)$  を加え、式(4.3)の整数化を行ない、座標データを付け加えてファジィ・データとする。

(図6)(a)~(c)はブロック・データの各次元の濃淡図である。また(図7)は各次元のヒストグラムである。ヒストグラムに各区域の平均のおよその位置を示したが生成の狙いどおり各特性値だけでは、AとB、CとDの間の識別をつけることは難かしい。

また、代表画素を4つの区域からそれぞれ5個取って(図5)のように布置する。これを使わず、式(4.5)により境界をあいま化すると(図8)のように、あいま化された区域がえられる。これに乱数を加えてできたファジィ・データが(図9)である。また(図10)は(図8)、(図11)は(図9)に、それぞれ対応するヒストグラムである。これをみると、いくつの区域があると考えられるかの判定はかなり困難である。

## 4.2 人工データによる実験

### 4.2.1 実験(その1)

まず、画素数が3840個のデータ(I)を用い、典型小領域の配置の仕方、抽出する領域の大きさ、閾値 $\delta$ の与え方、などが分類結果におよぼす影響を調べた。

初めに典型小領域の抽出個所をいろいろ変えたパターンをいくつか用意する(このパターンを「ウィンドウ」と名づける)。一例として(図12)に4つのウィンドウとそれを用いた場合の、[方略(I)]による区分図を示した。ブロック・データ、ファジィ・データに対して用意した、それぞれ2つのウィンドウの特徴と分類に使った距離、閾値の関係は(表3)に挙げた。また各ウィンドウの典型小領域の位置、抽出画素数は(表4)のように与えた。

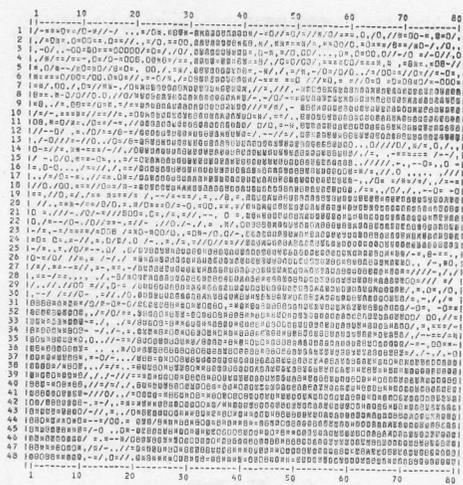
こうした実験を繰り返した結果、次の知見がえられた。

(1) 抽出する領域の大きさは、適当な大きさがあれば十分で、とくに大きくなる必要はない。むしろ、広い領域を指定すると、特性値の変動が大きくなることがある。経験的には、画素の数にして1抽出個所当り数十~百程度、1典型当り数百程度あれば十分である。

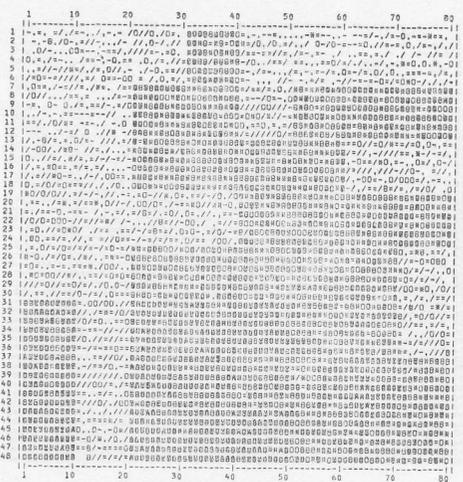
(2) むしろ、抽出の位置が問題で、できる限り等質に近いと思われる部分を指定するとよい。ウィンドウとしてB-2やF-2を用いた例にみられるように(図12)、小領域の位置によっては分類結果が夾雑になる。したがって、抽出の位置がまずいと分類結果から区域の特徴、差異を見わけることが困難になる。

(3) したがって、典型小領域はなるべく小さく、しかも1典型当り複数個をとり、広い範囲にばらまくとよい。B-1、F-1とB-2、F-2の区分例を較べると、この傾向がはっきり観察できる。

とくにファジィ・データのように区域の境界があいまいであっても、典型小領域を小さめに、また数多くの場所にばらまくと、領域の区分特徴をかなり把握できる。さらに、どの小領域をどの典型としてわりあてたらよいか判断に迷うときでも、各小領域を1つ1つの典型と考えて(つまり1典型当り1小領域として)一旦分類し、分類クロス表を観察し、重複の多い小領域を併合して、典型と考え直し、再吟味するという方策も考えられる。



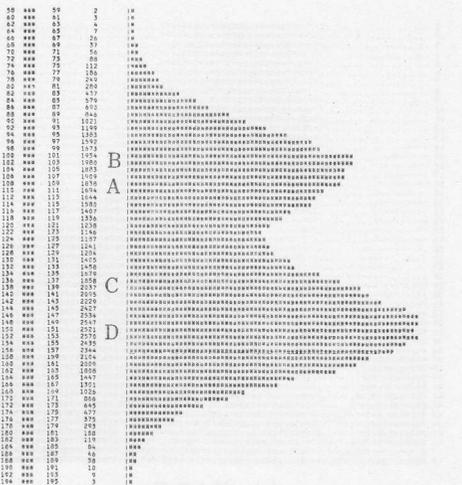
(a) 第1次元



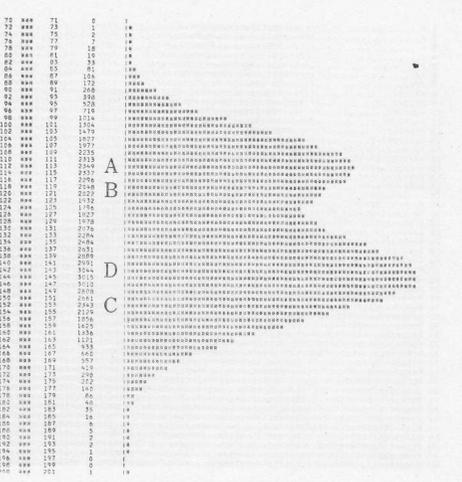
(b) 第2次元



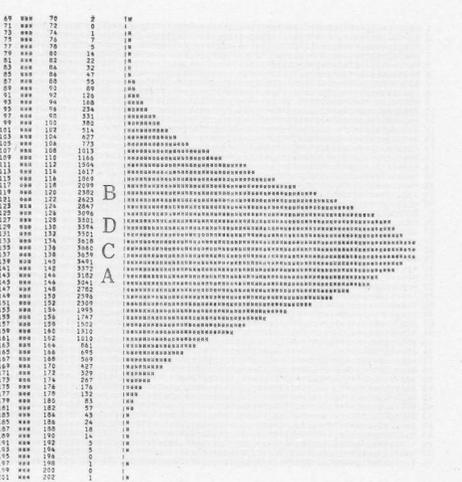
(c) 第3次元



(a) 第1次元



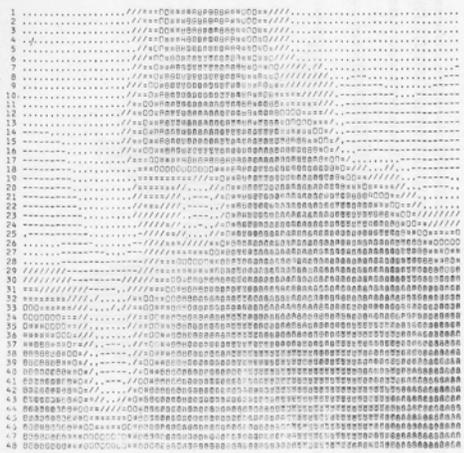
(b) 第2次元



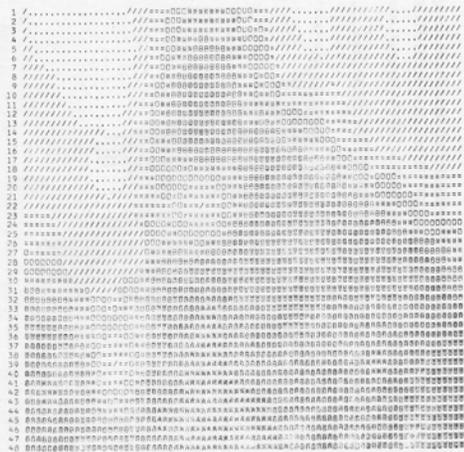
(c) 第3次元

図6 ブロックデータの濃淡図 (正規乱数を加えたあと)

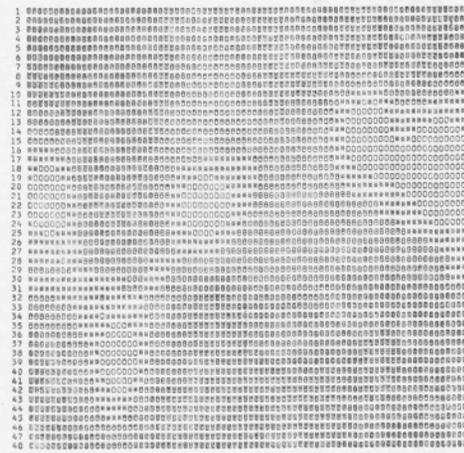
図7 ブロックデータのヒストグラム (76800個のとき)



(a) 第1次元



(b) 第2次元

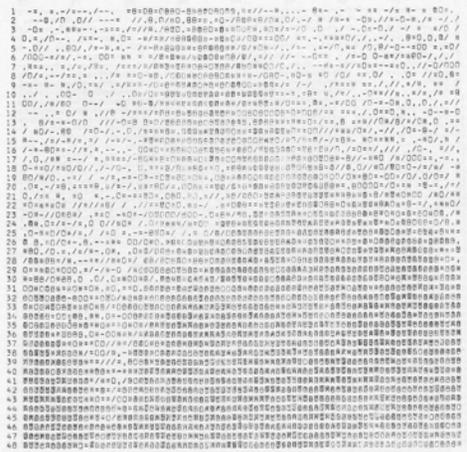


(c) 第3次元

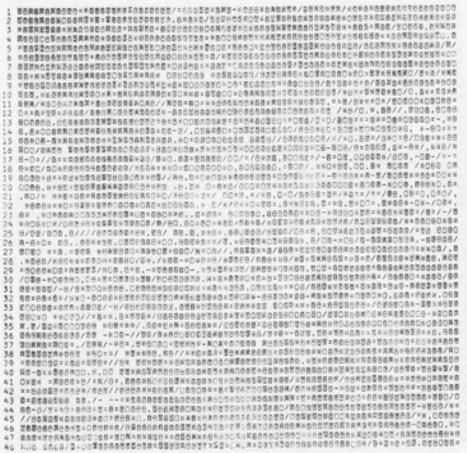
図8 ファジィデータの濃淡図 (正規乱数を加える前)



(a) 第1次元

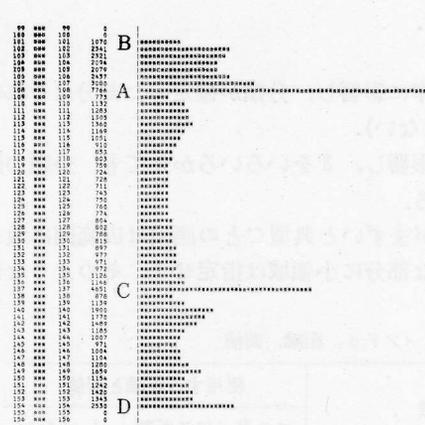


(b) 第2次元

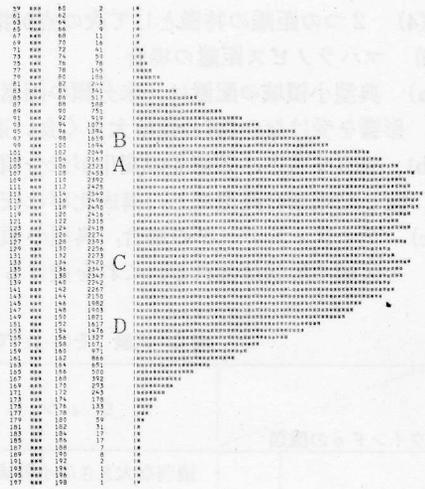


(c) 第3次元

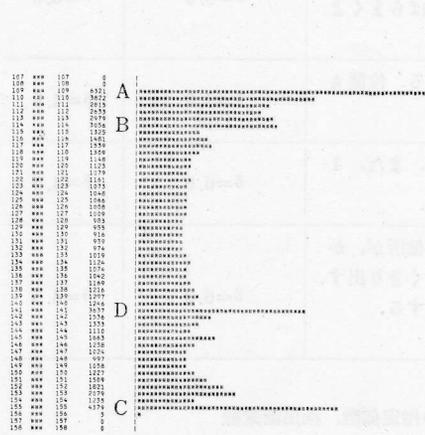
図9 ファジィデータの濃淡図 (正規乱数を加えた後)



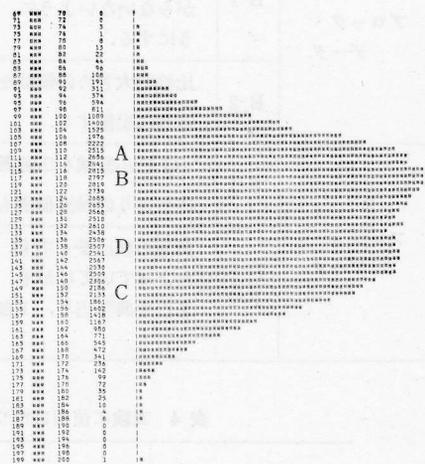
(a) 第1次元



(a)



(b) 第2次元



(b)



(c) 第3次元



(c)

図10 ファジィ・データのヒストグラム (76800個のとき)

図11 ファジィ・データのヒストグラム (正規乱数を加えた後)

(4) 2つの距離の特徴として次の点が指摘できる.

i) マハラノビス距離の場合

a) 典型小領域の配置は、未分類の画素の出現比率に影響し、分類が確定した部分はそれ程影響を受けない(誤分類は大きく食い違うことはない).

b) 換言すると、閾値 $\delta$ の変化が未分類の画素に影響し、 $\delta$ をいろいろかえても、分類の確定した画素の典型ごとの構成比率は安定している.

c) ファジィ・データの場合、典型小領域の設定がまずいと典型ごとの画素は広範囲に散在し区域の形状は乱れる. これを避けるには夾雑な部分に小領域は指定せず、そのような部

表3 実験(その1)で使用したウィンドウ, 距離, 閾値

ウィンドウの種類		ウィンドウの特徴	使用する距離と閾値	
			マハラノビス距離	$I_1$ ノルム
ブロック・データ	B-1	適当な大きさの小領域を1典型当り複数個とする. 配置の場所はなるべく区域の境界にまたがらないように, また広域にばらまくようにする.	$\delta=9.0$	$\delta=3.0$
	B-2	比較的大きめの領域を数少なくとる. 位置も適当に配置する.	$\delta=6.0$	$\delta=4.5$
ファジィ・データ	F-1	小さい小領域を広範囲にばらまく. また, 1典型当りの抽出個所も数多くとる.	$\delta=6.0$	$\delta=4.5$
	F-2	典型の数は多くても, 領域の抽出個所が, かたよっていても, 領域を大きくきり出す. また1典型当り, 1個所だけ抽出する.	$\delta=6.0$	$\delta=4.5$

表4 実験に使用したウィンドウの指定個数, 抽出画素数

ウィンドウの種類		1	2	3	4	5	計
実験(その1)の場合	B-1	42	64	24	64		339
		22 (64)	72 (136)	39 (63)	12 (76)		
	B-2	80	160	75	144		459
	F-1	16	16	16	16		320
16 16 (80)		16 16 (80)	16 16 (80)	16 16 (80)			
F-2	133	144	128	171	120	696	
実験(その2)の場合		187	208	143	187		2698
		189	208	198	78		
		143	143	272	273		
		102	187	78	102		
		(621)	(746)	(691)	(640)		

分は境界部の未分類区域として残し、座標情報も含めて判定すればよい（たとえば F-1）.  
ii)  $l_1$  ノルムの場合

- a) 適度な座標情報の導入は区域を明確にする。つまり、小領域の位置が多少まづい場合でも、 $\delta$  を座標情報が適当に働くように与えると、ウィンドウの不都合を補って、ある程度妥当な区分結果を示す（座標吟味の画素の割合が 5~15% 程度）。
- b) 全く座標情報を考慮に入れぬ場合は、小領域の位置の影響を敏感に受けやすい。

この数値実験では各領域における特性値の分散・共分散行列を等しくしている。それは、最初に説明したように誤分類の確率を事前に指定しなかったからである。相関行列は同じで分散を変えた場合、相関行列も変えた場合、さらに正規性を仮定しない場合をも実験する必要がある。こうした実験は何通りか試みたが、場合ごとに様相が変わり、明確な規則性はないのであえてここでは取り上げなかった。ファジィの場合の実験では領域をあいまいにすることにより、領域を夾雑にし、その結果として分散・共分散も一様でなくなっている。

表 5 分類後の典型ごとの構成比率の表  
マハラノビス距離を使用した場合

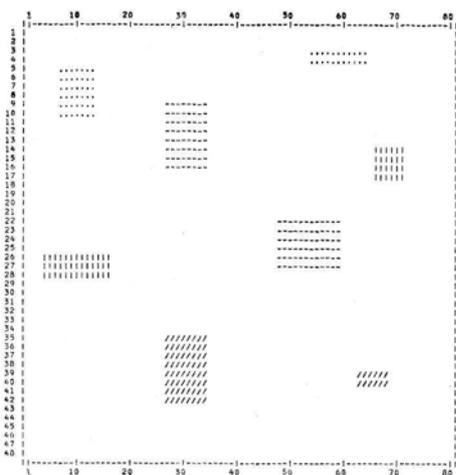
ブロック・データのとき				ファジィ・データのとき		
$\delta$	3.0	7.0	9.0	$\delta$	3.0	9.0
典型				典型		
A	11.42 (17.77)	16.26 (17.64)	17.02 (17.62)	A	14.71 (25.41)	23.30 (24.09)
B	14.03 (21.83)	19.95 (21.64)	20.97 (21.71)	B	14.05 (24.27)	25.41 (26.28)
C	11.91 (18.53)	17.57 (19.06)	18.45 (19.10)	C	14.28 (24.66)	23.08 (23.87)
D	26.91 (41.87)	38.40 (41.66)	40.16 (41.57)	D	14.85 (25.65)	24.91 (25.76)
未分類	35.73	7.84	3.40	未分類	42.10	3.30

(注) ( ) 内の数値は未分類を除く A~D を 100 としたものの構成比率を示す。

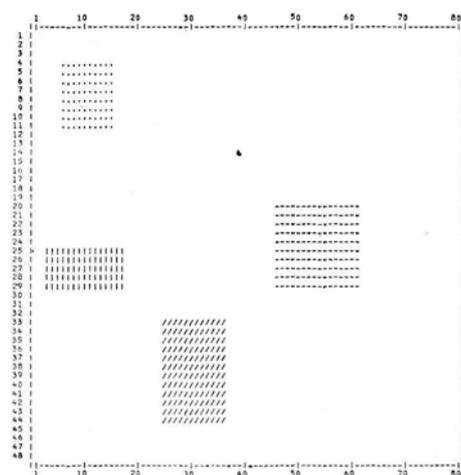
$l_1$  ノルムを使用した場合

ブロック・データのとき			ファジィ・データのとき			
$\delta$	1.5	4.5	$\delta$	1.5	3.0	4.5
典型			典型			
A	5.31 12.90 (18.20)	16.63 1.00 (17.63)	A	6.12 17.18 (23.30)	16.87 6.16 (23.03)	21.92 1.33 (23.26)
B	6.56 15.30 (21.86)	20.61 1.27 (21.87)	B	5.63 21.04 (26.68)	17.81 9.18 (26.99)	25.74 1.17 (26.91)
C	5.65 15.11 (20.76)	18.44 1.28 (19.72)	C	5.52 16.28 (21.80)	16.07 5.72 (21.79)	21.15 0.82 (21.96)
D	12.66 26.51 (39.17)	38.93 1.85 (40.78)	D	6.33 21.89 (28.22)	18.87 9.32 (28.19)	26.45 1.42 (27.87)
計	30.18 69.82	94.60 5.40	計	23.60 76.40	69.62 30.38	95.26 4.74

(注) { 上段が座標情報を取り入れずに判定された比率  
下段が座標情報を取り入れて判定された比率  
( ) 内は両者の和、つまり確定した典型の画素数の比率



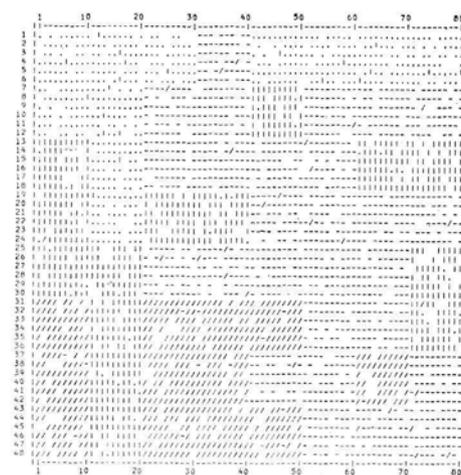
ウィンドウ (B-1)



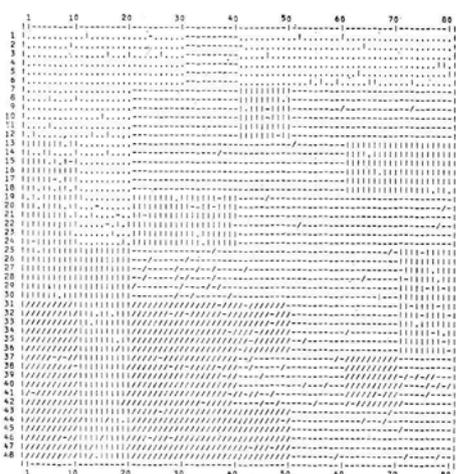
ウィンドウ (B-2)



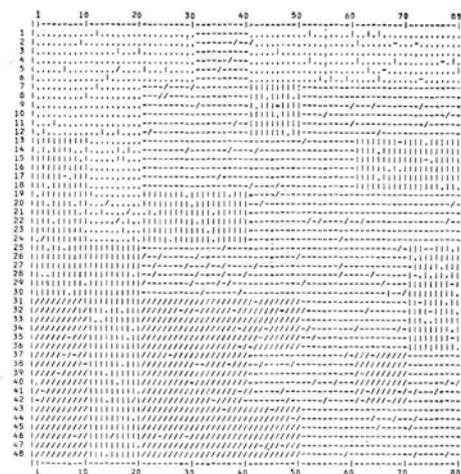
(マハラノビス距離;  $\delta=9.0$ )



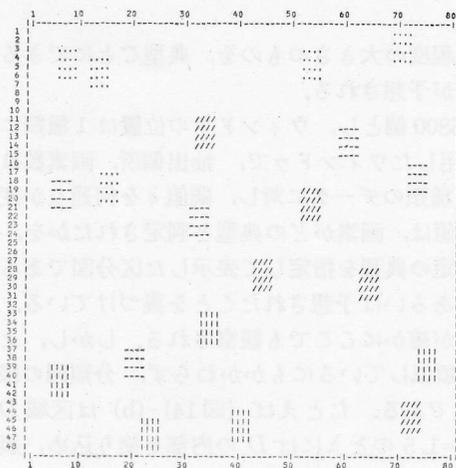
(マハラノビス距離;  $\delta=6.0$ )



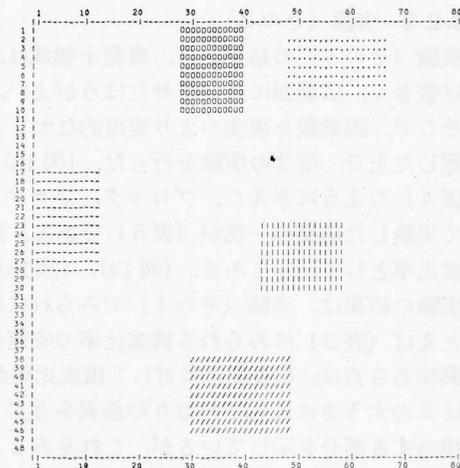
( $I_1$  ノルム;  $\delta=3.0$ )



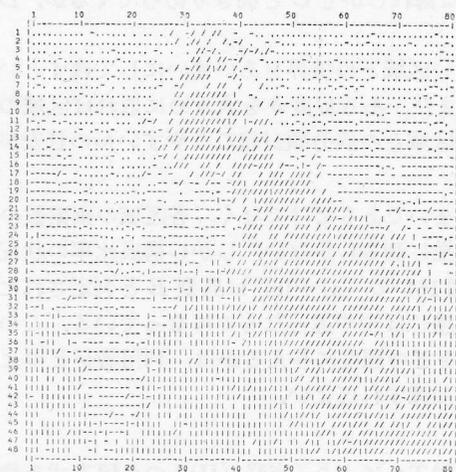
( $I_1$  ノルム;  $\delta=4.5$ )



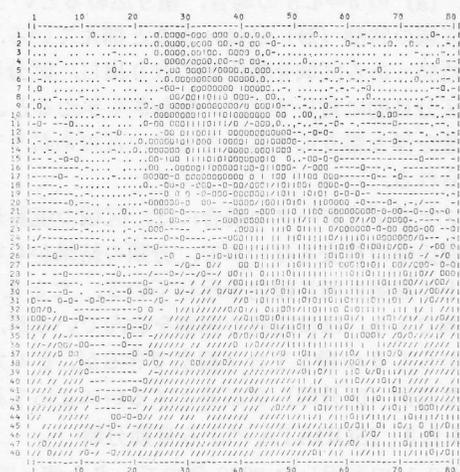
ウィンドウ (F-1)



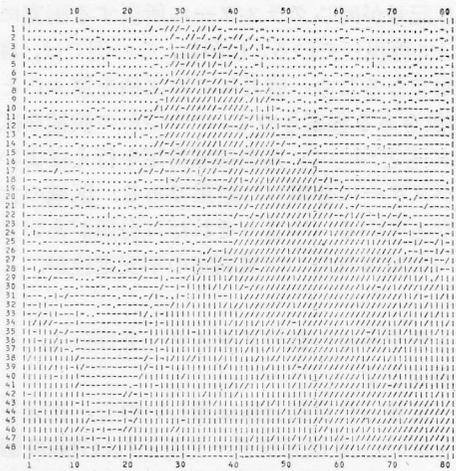
ウィンドウ (F-2)



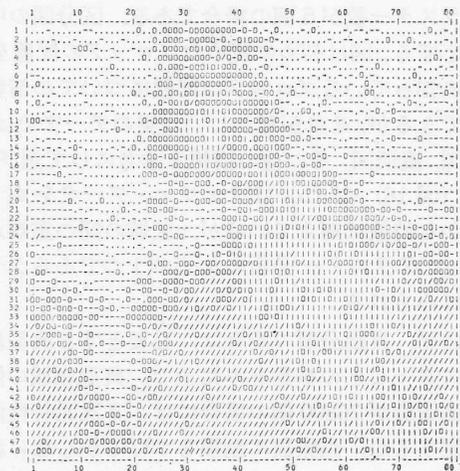
(マハラノビス距離;  $\delta=6.0$ )



(マハラノビス距離;  $\delta=6.0$ )



( $I_1$  ノルム;  $\delta=4.5$ )



( $I_1$  ノルム;  $\delta=4.5$ )

図 12 (つづき)

#### 4.2.2 実験 (その2)

実験 (その1) の結果から, 典型小領域は, ある程度の大きさのものを, 典型ごとに行き渡るだけ数多く, 広範囲に散在させたほうがよい, ことが予想される。

そこで, 画素数を現実のより実用的な大きさの76800個とし, ウィンドウの位置は1種類に固定した上で, 第2の実験を行った。(図13)が使用したウィンドウで, 抽出個所, 画素数は(表4)のように与えた。ブロック, ファジィの2種類のデータに対し, 閾値 $\delta$ を何通りか変えて実験した結果の一部が(表5)である。表の数値は, 画素がどの典型と判定されたかを, 構成比率として示してある。(図14), (図15)は特定の典型を指定して表示した区分図である。

実験の結果は, 実験 (その1) でみられた傾向, あるいは予想されたことを裏づけている。たとえば(表2)にみられる構成比率の安定化傾向が確かにここでも観察される。しかし, より興味ある点は,  $\delta$ の変化に対して構成比率が安定類似しているにもかかわらず, 分類図の様子は $\delta$ の大きさによってかなりの差異を生じることである。たとえば(図14)-(b)は区域Dに相当する部分を示しているが, これをみると,  $\delta=1.5$ のときにはDの内部を塗り込み, 誤分類の起こりにくい, つまり, 区域AまたはBと接している境界部を明確に捕えている。一方, (a)の $\delta=4.5$ の場合, 塗り込めるといふ現象は薄れしかもCと混ざりあってくるが, Dの区域の特徴(形)ははっきりと捕えている。

結局, ある区域を集中的に取り出し, 区分しにくい区域との境界をあいまいにして残すという区分目的に対しては,  $\delta$ を小さく与えるとよい。また, ある程度 $\delta$ を大きめに取れば区域の特徴を浮き彫りにするとともに, どのあたりでその区域との誤分類が起こりやすいかの目安がえられる。

ファジィ・データの場合には, 区域の境界があいまいであるだけでなく, 各区域の画素数の構成比率の目安もないので, 分類の結果から挙動を推し測るしかない。しかし構成比率の状況は上の説明と同じ傾向を示している。たとえば(図15)の(a)~(c)は典型Bと判定された区域であり, ほぼ同数の画素を含んでいるが, 図の感じはやや違ったものになっている。この例からも $\delta$ の働き具合を観察できるが,  $\delta$ の大きさをどの程度に与えるかは, 利用者がどのような図柄(区分図)を必要とするかによって適当に加減すればよい。

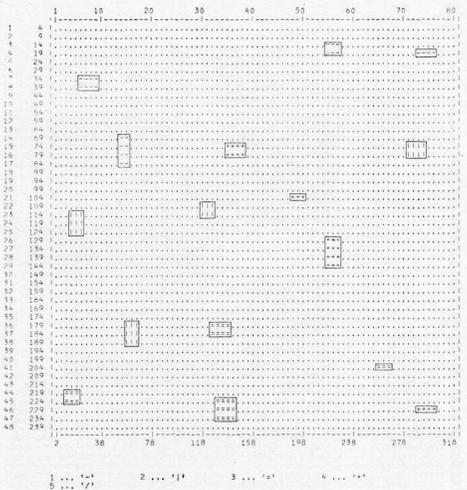
#### 4.2.3 ISODATAによる自動分類

4.2.1で利用した3840個のデータを使ってISODATAにより「方略(II)」の分類方式を検討した。(図16)は分類結果の一例である。(a)はブロック・データの分類結果である。この例はクラスター数が5個の場合である。区分図の中の同じ記号が同じクラスターに所属する画素に対応する。図は初めに定めたブロックの配置をきれいに再現している。しかし, この例では, 事前に知っているブロックの配置についての情報を利用できるため, 容易にこの図柄を選ぶことができるが, それが難しい現実のデータでは, 典型数のある程度きめて, いくつかの候補を区分図, 濃淡図, その他の情報を参考にして選び, これにもとづいて, 典型領域指定を行えばよい。

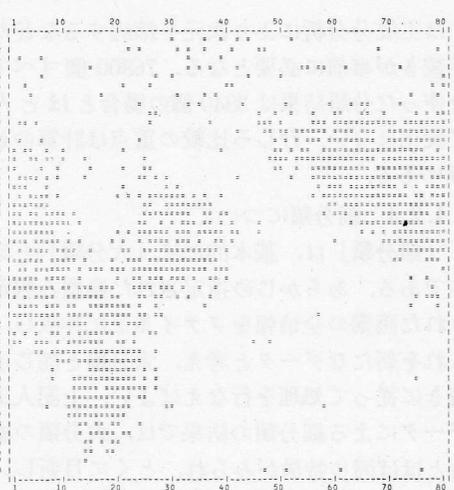
(図16)の(b)はファジィ・データを4つのクラスターに分類した結果である。区域の大まかな特徴は捕えているが, 区域の間の重複が多くなっている(つまりそれだけ分類があいまい)。クラスター分析の欠点の1つであるクラスターの重複部分の所属判定の不安定性が, ここにみられる。

結局, 自動分類の場合は, 必要とする典型数よりもやや多めの数のクラスターを作って, クラスターごとの挙動をみた上で, 典型領域の候補を選ぶとよい。

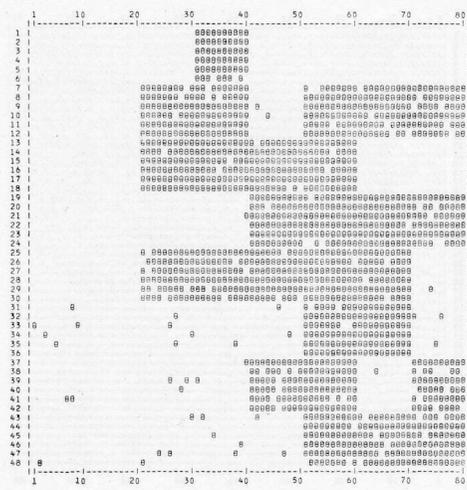
画素数が76800個の場合には, 3次元特性値をすべて使った自動分類は難しいので, 次元を限定するか, 画素数に制限をつけるかしなければならぬ。2次元のデータであれば76800個すべての分類が可能であり, 次元がそれ以上の場合には, 特定の次元の特性値を指定する, ある



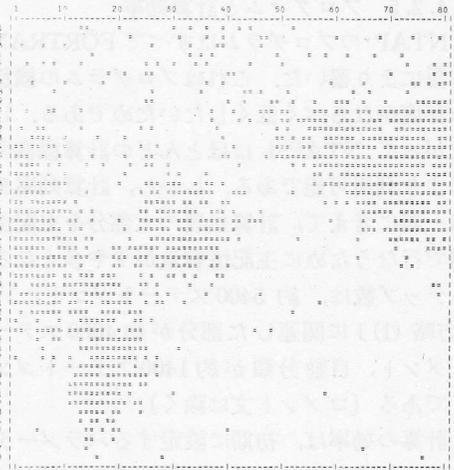
(\*) 同じ記号が同じ典型に対応する。  
図13 実験(その2)の典型指定箇所



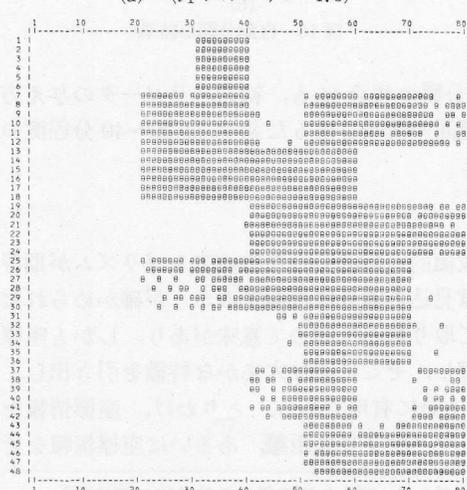
(a) (マハラノビス距離;  $\delta=9.0$ )



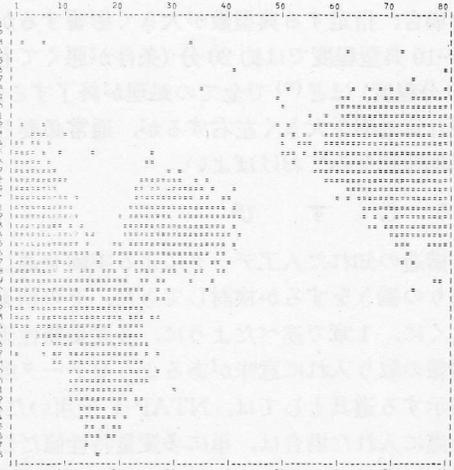
(a) ( $l_1$  ノルム;  $\delta=4.5$ )



(b) ( $l_1$  ノルム;  $\delta=4.5$ )



(b) ( $l_1$  ノルム;  $\delta=1.5$ )



(c) ( $l_1$  ノルム;  $\delta=1.5$ )

図14 ブロック・データの区分図(典型Dの場合)

図15 ファジィ・データの区分図(典型Bの場合)

いは主成分分析により次元を縮約するなどの手続きが事前に必要となる. 76800 個すべてを使った分類結果は 3840 個の場合とほとんど変化はない.むしろ比較の重点は計算の効率にある.

4.2.4 細分類について

「細分類」は, 基本的には「大分類」と同じである. あらかじめ指定された典型と判定された画素の全情報をファイル上に作成し, これを新たなデータと考え, 大分類と同じ手続きに従って処理を行えばよい. 上記人工データによる細分類の結果では, 大分類の場合とほぼ同じ特徴がみられ, とくに目新しい傾向はなかった.

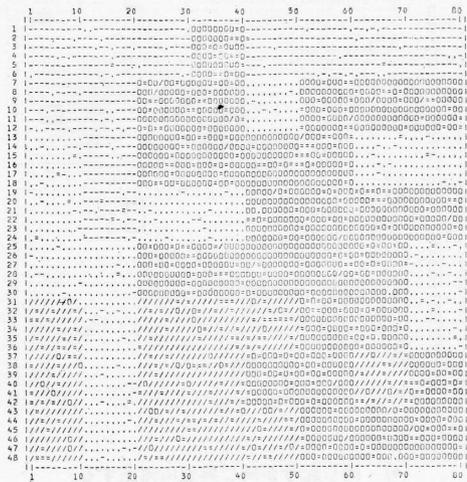
4.2.5 プログラムと計算効率

NTAP のプログラムはすべて FORTAN 言語により書いた. これはプログラムの機種依存度をなるべく低くしたいためである. したがって NTAP-I はほとんどの計算機機種により利用可能である. しかし, 計算効率を第一義に考えて, 計算処理の大部分を主記憶内で行なうために主記憶容量がやや多くなる. ステップ数は, 約 5400 ステートメント, 内 [方略 (I)] に関連した部分が約 4000 ステートメント, 自動分類が約 1400 ステートメントである (コメント文は除く).

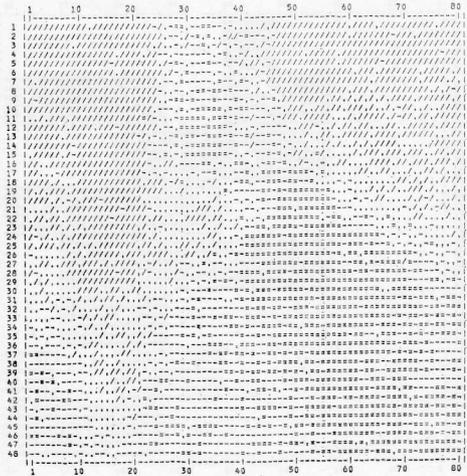
計算の効率は, 初期に設定するパラメータによって若干異なるが, 予想していた以上に良い結果を得ている. [方略 (I)] による分類の場合, 指定する典型数が大きく影響するが, 4~10 典型程度では約 20 分 (条件が悪くても 40 分程度) ほど (\*) で全ての処理が終了する. 自動分類の場合にも, 初期パラメータの与え方が計算効率を大きく左右するが, 通常必要とする範囲の情報を得るためには 30~40 分程度の処理時間をみておけばよい.

5. む す び

構造の知れた人工データによる実験を通して, 数値的地域区分のためのアルゴリズムが期待通りの働きをするか検討してきた. その結果, ほぼ見込み通りに機能することが確かめられた. とくに, 1章で述べたように, 多変量特性値として取り扱って初めて意味があり, しかも座標情報の取り入れに意味があるようなデータ解析に際し, そこに潜む大まかな特徴を引き出し, 表示する道具としては, NTAP-I で用いた手法は十分に有用である. とりわけ, 座標情報を考慮に入れた場合は, 単に多変量特性値だけによるマハラノビス距離, あるいは座標情報を考



(a)



(b)

図 16 自動分類の結果

[(\*) 統計数理研究所に設置された HITAC-8700 で, シングルジョブとして取扱った時の総処理時間]

えに入れない。ノルムだけによる分類にくらべて、分類目的によっては予想以上の効力を発揮することが観察される。また、実験を通して、典型小領域の閾値の与え方が分類結果に与える影響の大まかな特徴も知ることができた。もちろん、これは性質が比較的素直な人工データを使った上でのことではあるが、NTAP-I の機能を確認するには充分であろう。

現実のデータに NTAP-I を適用したとき、はたして見込み通りの機能を発揮するか、若干の疑問が残らぬでもない。とくに、画像にテクスチャ（肌理）のパターンが多く含まれるような場合、典型ごとの変動の大きさがまちまちであるような場合、などに遭遇したとき、座標情報を考慮した区分法がどのように働くか、また細分類などの機能が区分にどのように影響するか、などの問題が起こる。

これらを吟味するため、実際のデータの解析をも進めているが、結果の一部から、NTAP-I が実用上十分に役立つことがわかってきている。こうした現実のデータ解析については別の機会に報告する。

なお、プログラム開発と実験に際して、統計数理研究所の桂康一氏および日本大学（当時）の松下久君お二人の多大のご協力をえた。また有益な助言をいただいた論文審査員に深謝いたします。

## 参 考 文 献

- [1] 奥野忠一他 (1971), 多変量解析法, 日科技連出版社.
- [2] キャノンイメージ編集室編 (1976), Remote Sensing.
- [3] 坂井利之 (1975), 情報の探検, 岩波新書.
- [4] ブルーシー・アンド・グリーンランド財団編 (1975), 日本国土海洋総合学術診断 報告書.
- [5] 和達清夫他 (1976), リモートセンシング, 朝倉書店.
- [6] 朝倉堅五 (1975), リモートセンシングの可能性, 数理科学, Vol.13, No.12 pp.62-72, サイエンス社.
- [7] 計測と制御 (小特集リモートセンシング) (1976), 15, No.7, 社団法人計測制御学会.
- [8] 渋谷政昭, 大隅 昇 (1977), 数値的地域区分法: NTAP, 第3回リモートセンシング・シンポジウム
- [9] Ball, G.H. and Hall, D.J. (1965) ISODATA, A novel method of data analysis and pattern classification, *Tech. Rep.* Stanford Research Institute.
- [10] Batchelor, B.G. (1974) *Practical Approach to Pattern Classification*, Plenum Press.
- [11] Dasarathy, B.V. (1975) An innovative clustering technique for unsupervised learning in the context of remotely sensed earth resources data analysis, *Int. J. System Sci.*, 6, 23-32.
- [12] Dubes, R.C. (1970) *Information Compression, Structure Analysis and Decision making with a Correlation Matrix*, Michigan State Univ.
- [13] Duran, B.S. and Odell, P.L. (1974) *Cluster Analysis: A Survey*, Springer-Verlag.
- [14] Eisenbeis, R.A. and Avery, R.B. (1972) *Discriminant Analysis and Classification Procedures-Theory and Applications*, Heath, D.C. and Co.
- [15] Fu, K.S. (ed.) (1977) *Syntactic Pattern Recognition, Applications*, Springer-Verlag.
- [16] Haralick, R.M. (1976) Automatic Remote Sensor Image Processing, in *Digital Picture Analysis*, Rosenfeld, A. (ed.), Springer-Verlag.
- [17] Hall, D.J. and Khanna, D.V. (1976) The ISODATA method computation for the relative perception of similarities and differences in complex and real data in *Statistical Methods for Digital computer, III of Math. Methods for Digital Computers*, Ralston, K.A., Wilf, H.S. (eds.), John Wiley.
- [18] Ingram, D.S., Actkinson, A.L. (1973) The applicability and effectiveness of cluster Analysis, *NASA, JSC Internal Note No. 73-FM-53*.
- [19] Odell, P.L. (1976) Remote sensing, in special issue of *Communication in Statistics*, A5.
- [20] Rosenfeld, A. (1973) Progress in picture processing, *Computing Surveys*, 5.
- [21] Su, M.Y. (1972) Unsupervised classification of multispectral sensing data, *NASA, NAS 8-27364*.
- [22] Tou, J.T. and Gonzalez, R.C. (1974) *Pattern Recognition Principles*, Addison Wesley.