

初期配置の影響についての 非階層的クラスター分析分割最適化法間の比較*

—人工データによる検討—

辻本 英夫¹⁾

I 問題

探索的なデータ解析法の1つとして多くの分野でもちいられているクラスター分析には、周知のように、数多くの手法が含まれる。実際のデータにクラスター分析を適用しようとする場合、この多様な手法群の中から、そのデータに合った手法を選択しなければならないが、そのためには、あらかじめ各手法の特性や実際的な有効性を検討し明らかにしておく必要がある。われわれは、このような目的に資するために、多様なクラスタリング手法の中から、非階層的クラスター分析の分割最適化法(以下、分割型とする)と呼ばれる手法のいくつかをとりあげ、それらの手法をデータに適用する際に問題となるいくつかの点について検討を試みた(辻本・大島, 1980)。

分割型クラスタリング手法は、クラスター数が固定か可変かによって2分されるが、われわれがとりあげた手法は、Forgy法、Jancey法、MacQueen's K-means法、convergent K-means法といういずれもクラスター数固定の分割型クラスタリング手法である。これらの手法の基本的なアルゴリズムは、まず初期配置を与え—クラスターを仮に代表する点(初期代表点)を直接与えるか、全データ=ユニットを群分けして(初期分割して)その重心を初期代表点とする—、以下、最も近い代表点へのデータ=ユニットの割り当て・クラスターの構成・代表点の更新を、あらかじめ設定された基準に達するまで反復するというものであるが、各手法のアルゴリズムにはそれぞれ異なる工夫がなされている。

われわれが検討を試みたのは、主として初期配置の違いが結果として得られるクラスターに及ぼす影響についてであり、その主な結果は次の通りである。①初期配置

の違いは、結果として得られるクラスターにかなりの影響を及ぼす。したがって、どのような初期配置が望ましいクラスターをもたらすか、あるいは、どのような方法で初期配置を決定すれば良いかということが、重要な問題となる。初期配置の方法を幾種類か比較した結果では、前もって行なった階層的クラスター分析で得られたクラスターを初期分割とした場合に、どのデータについても比較的好ましいクラスター—すなわち、クラスター内でのバラツキがより小さく、クラスター間でのバラツキがより大きいようなクラスター—が得られた。②とりあげた4つのクラスタリング手法を比較して、MacQueen's K-means法で得られるクラスターは他の手法で得られるクラスターに比べて、クラスター内でのバラツキが小さくクラスター間でのバラツキが大きいという傾向を示した。加えて、他の手法よりも初期配置の影響に対して敏感であること、得られたクラスターがサンプル順序にも影響を受けることといった点も考え合わせると、MacQueen's K-means法は他の3つの手法よりも劣ると結論づけられる。

しかしながら、これらの知見はわずか3つのデータの分析から得られたものであり、また、それらのデータがデータの種々のタイプを代表するものであるとも言えないので、これらの知見をすべてのデータに一般化することは早急すぎるであろう。したがって、クラスター数固定の分割型クラスタリング手法の実際的な有効性を明らかにするためには、さらに多くの種々のタイプのデータについて同種の分析をくりかえし、資料を蓄積していく必要がある。また、もう1つの方法として、望ましい初期配置の特性の分析とそのような初期配置を与える方法の検討を行なうこと、同時に、それぞれのクラスタリング手法のアルゴリズムのもつ「くせ」のようなものを分析し、初期配置の影響がより少ないようにアルゴリズムの改良を行なうことも重要であろう。

本研究は、後者の観点に立って、クラスター数固定の分割型クラスタリング手法における初期配置の問題の検討を試みたものである。

望ましい初期配置に関しては、われわれの行なった検討(辻本・大島, 1980)では、前もって行なった階層的

* Comparison between 4 partitioning-optimization techniques of nonhierarchical clustering methods in terms of the influence of initial configurations.

本研究のデータ分析のための計算は、名古屋大学大型計算機センターの FACOM M-200 によった。

1) 大阪市立大学

クラスター分析によって得られるクラスター——すなわち、データの中に潜在すると仮定されるクラスター——の重心ないしは重心に近いデータ=ユニットが、そのような初期配置（初期代表点）として示されたことはすでに述べた。このことは、最近シミュレーションにより、幾つかのクラスタリング手法の妥当性の検討を行なった Milligan (1980) の研究においても、同様に認められている。このような初期配置が望ましいことは、仮定されるクラスターと真のクラスターとのズレが小さいという条件つきで、直観的にも支持されよう。Milligan (1980) の研究では、さらに、真のクラスターから各クラスターにつき1つずつのデータ=ユニットをランダムに選んで初期代表点とする初期配置も、また、比較的望ましい初期配置であることが示されている。

一方、初期配置の影響についての4種の分割型クラスタリング手法間の比較については、われわれの検討からは、MacQueen's K-means 法が他の手法よりも初期配置の影響に敏感であるという傾向のみがうかがえた(辻本・大島, 1980)。

本研究では、これらの点の確認を含めて、望ましい初期配置及び望ましいクラスタリング=アルゴリズムの特性についての手掛りを得るために、簡単な人工データをもちいて、クラスター数固定の分割型クラスタリング手法間の初期配置の影響という点からの比較を試みたので、その結果について報告する。

II 方法

1. データ及びクラスター

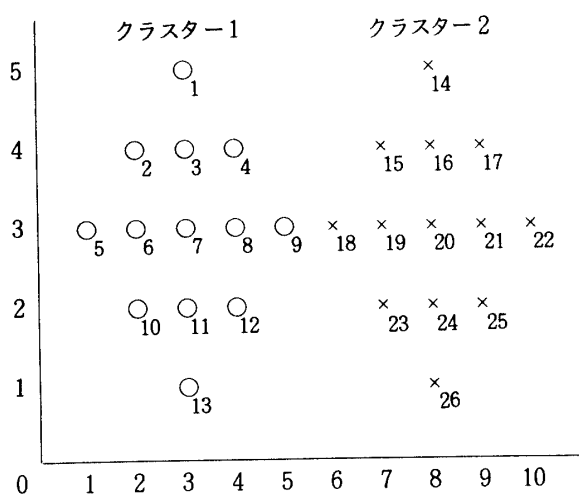


図1 データ1

図1・2に示す人工データ、データ1と2をもちいる。両データとも、26データ=ユニット、2変数からなる2次元データであり、2つの分離したクラスターを構成する。各クラスターは同数のデータ=ユニットから成る。データ1と2はデータの入力順序のみが異なり、データ2の入力順序はランダムに決定された。図に示した1~26の数字が、その入力順序を表わす。入力順序のみが異なる2つのデータをもちいるのは、MacQueen's K-means 法とconvergent K-means 法がデータの入力順序に影響を受けるためである。

2. クラスタリング手法

辻本・大島(1980)と同じく、非階層的クラスター分析の中のクラスター数固定の分割型手法である、Forgy 法、Jancey 法、MacQueen's K-means 法、convergent K-means 法の4手法をもちいる。これらのうち、ある手法同士の間には、アルゴリズム上の共通点がある。MacQueen's K-means 法以外の3つの手法は共通して、クラスター成員が変化しなくなるまで代表点の更新とデータ=ユニットの割り当てをくりかえし行なう。また、代表点の更新の仕方について、Jancey 法以外の3つの手法は共通した仕方をとる。さらに、代表点の更新の時期では、Forgy 法とJancey 法が全データ=ユニットが割り当てられた後に更新するのに対して、MacQueen's K-means 法とconvergent K-means 法は各データ=ユニットが割り当てられるごとに更新を行なう。

表1に、各クラスタリング手法のアルゴリズムを簡単に示す。詳しくはAnderberg(1973)参照。

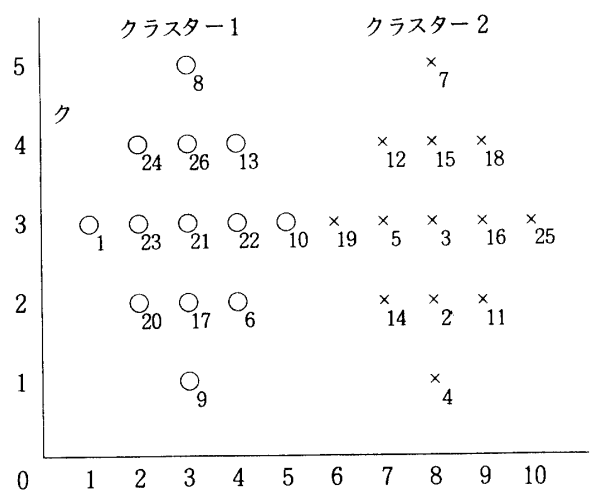


図2 データ2

表1 各クラスタリング手法のアルゴリズム

	Forgy 法	Jancey 法	MacQueen's K-means 法	convergent K-means 法
ステップ1	初期代表点が与えられる。	初期代表点が与えられる。	データの最初の k † 個のデータ=ユニットを初期代表点とする。	初期代表点を与え、データ=ユニットを最も近い代表点に割り当てて各群の重心を計算し、代表点とする。
ステップ2	代表点を固定したまま、データ=ユニットを最も近い代表点に割り当てる。	代表点を固定したまま、データ=ユニットを最も近い代表点に割り当てる。	残りのデータ=ユニットを最も近い代表点に順次割り当て、そのたびごとに重心を計算して代表点とする。	データ=ユニットを順次最も近い代表点に割り当て、割り当てによって成員が変わる場合に、その群の重心を再計算して代表点とする。
ステップ3	各群の重心を計算して新しい代表点とする。	各群の重心Cを計算し、計算前の代表点S'とCから、 $S=2C-S'$ という式で新しい代表点Sを求める。	代表点を固定したまま、再度データ=ユニットを最も近い代表点に割り当てて、最終的なクラスタとする。	各群の成員が変化しなくなるまで、ステップ2をくり返す。
ステップ4	各群の成員が変化しなくなるまで、ステップ2と3をくり返す。	各群の成員が変化しなくなるまで、ステップ2と3をくり返す。		

† k はクラスタ数を表わす。

3. 初期配置

初期配置の与え方は、前述のように、初期代表点を与えるか初期分割するかであるが、ここではすべて初期代表点を与える。

条件A：2つの初期代表点が共に同一のクラスタに属する場合を、条件Aとする。データ1の場合具体的には、図1に示されたクラスタ1の成員のうち、重心に位置するデータ=ユニット7*と周辺に位置するデータ=ユニット1・5・9・13を選ぶ。そして、これらのうちから2つをとるすべての組み合わせのうち、クラスタ1と2の重心を結んでできる直線に線対称な位置関係となるデータ=ユニットの組み合わせの一方を除外して、残りの組み合わせをすべて代表点としてもちいる。データ2の場合も、データ1とまったく同じ初期代表点**をもちいる。それらを表2に示す。

条件B：2つの初期代表点が各々異なるクラスタに

* 説明の便宜上、入力順序番号をデータ=ユニット番号とする。

** 入力順序をデータ=ユニット番号とした関係上、データ=ユニット番号も異なる。

属する場合を、条件Bとする。データ1・2とも、条件Aで選んだ5つのデータ=ユニットに加えて、同じ基準でクラスタ2からも5つのデータ=ユニットを選ぶ。これらのうちから2つをとるすべての組み合わせのうちから、まず、その2つが同一のクラスタに属する組み合わせを除外し、次に、2つのクラスタ重心を結ぶ直線に線対称な、あるいは、その直線とクラスタ間の境界線との交点(5.5, 3.0)に点対称な位置関係になるデータ=ユニットの組み合わせの一方を除外して、残りの組み合わせをすべてこの条件での初期代表点としてもちいる。それらを表2に示す。

条件A'：条件Aで初期代表点としてもちいるデータ=ユニットの各対について、その入力順序を変えた場合を条件A'とする(表2参照)。これは、2つの初期代表点から等距離に位置するデータ=ユニットが存在する場合、そのデータ=ユニットは最初に入力した初期代表点のクラスタに割り当てられるため、初期代表点の入力順序によって最初のデータ分割が異なってくるためである。条件Bの場合には、どの初期代表点についても、そのようなデータ=ユニットは存在しないので、条件Aに対する

表2 初期代表点

条件	初期代表点
A-1	7, 9
2	7, 5
3	7, 1
4	9, 5
5	9, 1
6	5, 1
7	1, 13
B-1	7, 20
2	7, 18
3	7, 22
4	7, 14
5	9, 18
6	9, 22
7	9, 14
8	5, 22
9	5, 14
10	1, 14
11	1, 26
A'-1	9, 7
2	5, 7
3	1, 7
4	5, 9
5	1, 9
6	1, 5
7	13, 1

条件A'のような、条件Bに対する条件B'は要しない。

4. 距離測度・クラスター概念・プログラム

辻本・大島(1980)と同じ距離測度・クラスター概念・コンピュータープログラムをもちいる。

III 結果

表3に、データ1における条件AとBの場合のクラスターリングの結果を示す。表中の値はクラスター内平方和*

* クラスター内平方和は次式によった；

$$\sum_{k=1}^m \left(\frac{1}{n_k} \sum_{\alpha, \beta \in C_k} d_{\alpha\beta}^2 \right)$$

ここで、 α と β はクラスター C_k の成員、 $d_{\alpha\beta}$ は α と β の距離、 n_k はクラスター C_k の成員数、 m はクラスター数。

を表わす。また、「P」という記号はあらかじめ設定されていたクラスターが完全に復元されたことを表わし、その場合のクラスター内平方和は56.000である。以下、表4、表5においても、表中の値・記号の意味は同様である。

条件Bの場合、すなわち、あらかじめ設定された2つのクラスターからそれぞれ1つずつ初期代表点を選んだ場合には、どのクラスターリング手法をもちいた場合にも、すべての条件(B-1~B-11)で設定されていたクラスターが完全に復元された。

それに対して条件Aの場合、すなわち、あらかじめ設定された2つのクラスターの片方から2つの初期代表点を選んだ場合には、どのクラスターリング手法をもちいたかによって少々結果が異なる。Forgy法あるいはJancey法をもちいた場合には、条件A-3とA-7を除いた残りのすべての条件で、設定されていたクラスターが完全に復元された。convergent K-means法の場合も、Forgy法、Jancey法の場合と結果はほぼ同じで、条件A-7のみを除くすべての条件において、設定されていたクラスターが完全に復元された。この3つの手法に比べて、MacQueen's K-means法の場合には結果は良くなく、設定されていたクラスターが完全に復元されたのは、わずかに条件A-3の場合だけであった。

次に、データ1における条件A'の場合の結果を表4に示す。どのクラスターリング手法においても、結果は条件Aとまったく同一であるかほぼ同じであり、初期代表点の入力順序による影響はあまりみられなかった。Forgy法、Jancey法、convergent K-means法の3手法については、条件A'-3とA'-7以外のすべての条件において、設定されていたクラスターが完全に復元されるという結果が示された。この3手法とは対照的に、MacQueen's K-means法の場合には、A'-1からA'-7までの条件A'でのどの条件においても、設定されていたクラスターが復元されるという結果は得られなかった。

データ2についての結果は、表5に示した通りである。Forgy法とJancey法では、すべてのデータ=ユニットが割り当てられるまで代表点は固定されたままであり、データの入力順序には結果は影響されないので、MacQueen's K-means法とconvergent K-means法のみを、データ2に適用した。

MacQueen's K-means法の場合には、条件Bについては、どの条件においても設定されているクラスターが完全に復元されるという、データ1と同一の結果が得られ、条件AとA'に関しては、14の条件のうち約3分の2の条件において設定されていたクラスターが完全に復元されるという、データ1の場合よりも良い結果が得ら

資 料

表3 データ1の条件A, Bにおけるクラスタリングの結果

条 件	初期代表点	クラスタリング手法			
		Forgy	Jancey	MacQueen's K-means	convergent K-means
A-1	7, 9	P	P	60.024	P
2	7, 5	P	P	71.758	P
3	7, 1	200.444	200.444	P	P
4	9, 5	P	P	60.024	P
5	9, 1	P	P	60.024	P
6	5, 1	P	P	60.024	P
7	1, 13	200.444	200.444	71.758	200.444
B-1	7, 20	P	P	P	P
2	7, 18	P	P	P	P
3	7, 22	P	P	P	P
4	7, 14	P	P	P	P
5	9, 18	P	P	P	P
6	9, 22	P	P	P	P
7	9, 14	P	P	P	P
8	5, 22	P	P	P	P
9	5, 14	P	P	P	P
10	1, 14	P	P	P	P
11	1, 26	P	P	P	P

† 表中の値はクラスター内平方和を示す。「P」は設定されたクラスターが完全に復元されたことを示し、その場合のクラスター内平方和は56.000である。

表4 データ1の条件A'におけるクラスタリングの結果

条 件	初期代表点	クラスタリング手法			
		Forgy	Jancey	MacQueen's K-means	convergent K-means
A'-1	9, 7	P	P	60.024	P
2	5, 7	P	P	71.758	P
3	1, 7	200.444	200.444	60.024	200.444
4	5, 9	P	P	60.024	P
5	1, 9	P	P	60.024	P
6	1, 5	P	P	71.758	P
7	13, 1	200.444	200.444	71.758	200.444

† 表中の値はクラスター内平方和を示す。「P」は設定されたクラスターが完全に復元されたことを示し、その場合のクラスター内平方和は56.000である。

れた。convergent K-means 法の場合には、データ1の場合とまったく同じ結果が得られた。すなわち、条件A-7, A'-3, A'-7を除く条件A, B, A'における残

りのすべての条件の下で、設定されていたクラスターが完全に復元されるという結果が示された。データ2の場合においても、データ1の場合と同様、

表5 データ2におけるクラスタリングの結果

条 件	初期 代表点	クラスタリング手法	
		MacQueen's K-means	convergent K-meant
A-1	7, 9	P	P
2	7, 5	60.024	P
3	7, 1	P	P
4	9, 5	P	P
5	9, 1	P	P
6	5, 1	60.024	P
7	1, 13	60.024	200.444
B-1	7, 20	P	P
2	7, 18	P	P
3	7, 22	P	P
4	7, 14	P	P
5	9, 18	P	P
6	9, 22	P	P
7	9, 14	P	P
8	5, 22	P	P
9	5, 14	P	P
10	1, 14	P	P
11	1, 26	P	P
A'-1	9, 7	P	P
2	5, 7	60.024	P
3	1, 7	P	200.444
4	5, 9	P	P
5	1, 9	P	P
6	1, 5	60.024	P
7	13, 1	P	200.444

† 表中の値はクラスター内平方和を示す。「P」は設定されたクラスターが完全に復元されたことを示し、その場合のクラスター内平方和は56.000である。

初期代表点の入力順序による影響はあまりみられなかった。また、データの入力順序の影響については、convergent K-means 法では、辻本・大島(1980)の結果と異なり影響がみられなかった。これは、もちいたデータの構造があまりにも単純であったためだと思われる。

以上の結果は、次の2点にまとめられる。

第1に、2つのデータ及び4つのクラスタリング手法のどの場合においても、設定された2つのクラスターの各々から1つずつデータ=ユニットを選んで初期代表点とする条件Bの場合には、すべて、設定されていたクラ

スターが完全に復元されるという結果が得られた。この結果は、真のクラスターからそれぞれ1つずつランダムに選ばれたデータ=ユニットが初期代表点として望ましいものである、という Milligan (1980) の得た結果と一致する。

第2に、条件Bでは4つのクラスタリング手法の間に結果の違いがみられなかったのに対して、設定された2つのクラスターのうちの1つから初期代表点とするデータ=ユニットを2つとも選ぶという条件AとA'の場合には、MacQueen's K-means 法と他の3手法の間に、ク

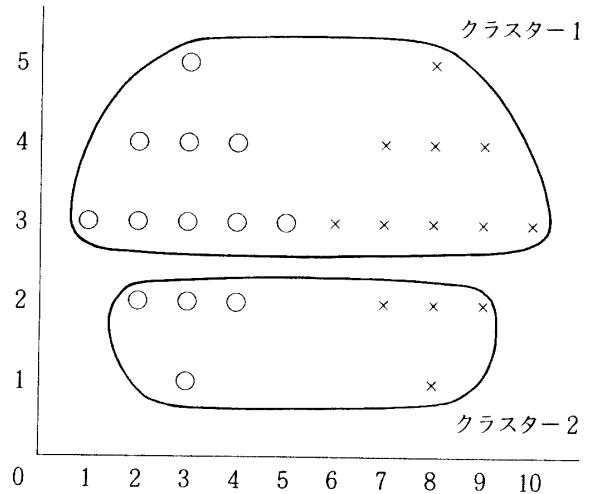


図3 MacQueen's K-means 法以外の手法によるクラスタリングの結果(1)
(クラスター内平方和 = 200.444)

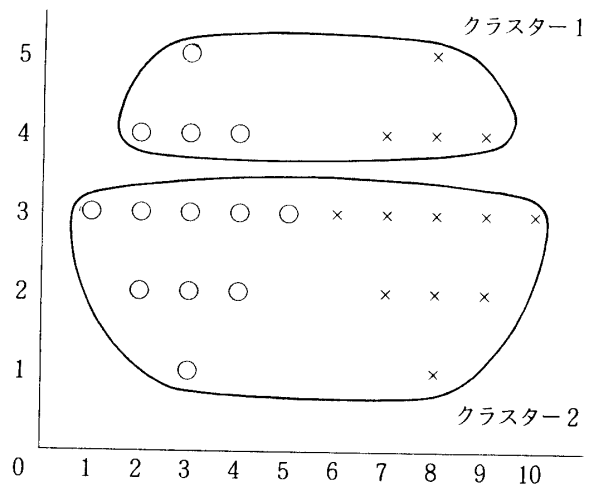


図4 MacQueen's K-means 法以外の手法によるクラスタリングの結果(2)
(クラスター内平方和 = 200.444)

ラスタリング結果に違いがみられた。Forgy法, Jancey法, convergent K-means法では, いくつかの条件を除いてほとんどの場合, 条件Bと同様, 設定されていたクラスターが完全に復元されるという結果が得られた。設定されていたクラスターが復元されなかった場合というのは, いずれも, 設定された2つのクラスターの重心を結ぶ直線と直交する直線上に2つの初期代表点が配置された場合であった。ただ, convergent K-means法については, そのような初期代表点をもちいた場合でも, 初期代表点の入力順序によっては, 設定されていたクラ

スターが完全に復元される場合がありうるということが示された。他方, MacQueen's K-means法の場合には, データの入力順序によって違いはあるものの, 概してこれらの条件の下では, 他の3手法に比べて結果は良くなく, 約3分の1の条件においてしか, 設定されていたクラスターが完全に復元されるという結果が示されなかった。4つのクラスタリング手法の優劣に関するここでの結果は, 辻本・大島(1980)の結果と一致する。

しかしながら, 設定されていたクラスターが復元されなかった場合の結果について, MacQueen's K-means法と他の3手法とを比較してみると, 他の3手法によって得られたクラスター(図3ないしは図4)が設定されたクラスターとかなり異なるのに対して, MacQueen's K-means法では設定されていたクラスターがほぼ復元されていることがわかる(図5, 図6)。これは興味深い結果であり, この点に十分留意する必要がある。

IV 討 論

望ましい初期配置の特性は何かという問題に関して, 本研究では, Milligan(1980)同様, 真のクラスター(あらかじめ設定されたクラスター)の各々から1つずつ選ばれたデータ=ユニットを初期代表点とした場合に, それらのクラスターがクラスタリングの結果十分復元されうるという結果を得た。したがって, 初期代表点とするデータ=ユニットがそれぞれ, データ構造の中に潜在する異なるクラスターの成員であるならば, それらの初期代表点は望ましい初期代表点であると考えられる。もちろん, ここでもちいたデータは人工データであり, データ構造も単純で, 行なった分析も組織的ではないので, 初期代表点についてのこの条件は, 望ましい初期配置(初期代表点)としての十分条件とは断定できない。また, 同様な結果を得ているMilligan(1980)の研究では, 本研究とは研究目的が異なるため, 初期配置の問題に関しては十分組織的な分析を行なっているとは言えず, その結果からも, 初期代表点についてのこの条件が十分条件であるとは言いきれない。

逆に, この条件を望ましい初期配置の必要条件として限定することも早急すぎるように思われる。なぜならば, 本研究で, 設定されたクラスターの1つから2つのデータ=ユニットを選んで初期代表点とした場合にも, MacQueen's K-means法以外の3手法では, おおむね設定されていたクラスターが完全に復元されるという結果を示したからである。加えて, この後に述べるように, MacQueen's K-means法に少し工夫を加えたアルゴリズムによれば, 本研究でもちいたどのデータ・初期代表点の場合においても, 設定されたクラスターが完全に復元さ

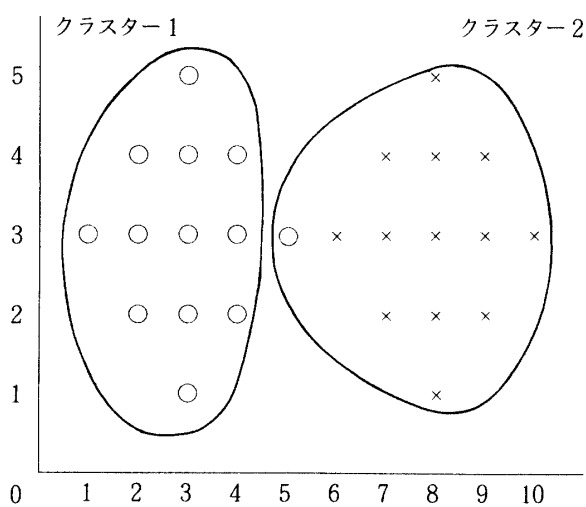


図5 MacQueen's K-means法によるクラスタリングの結果(1)
(クラスター内平方和 = 60.024)

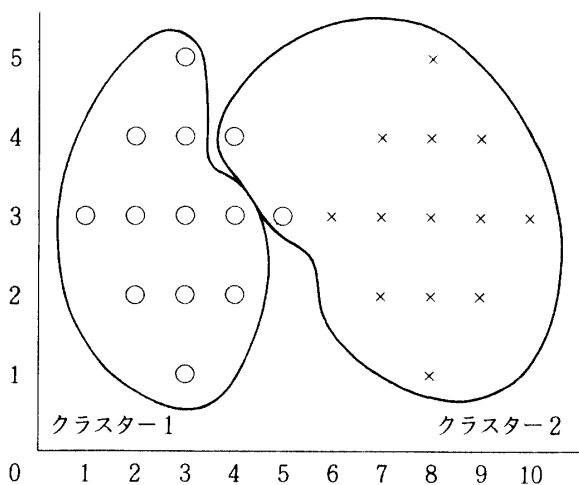


図6 MacQueen's K-means法によるクラスタリングの結果(2)
(クラスター内平方和 = 71.758)

れうることを示すことができるからである。

したがって、初期代表点がそれぞれ異なる真のクラスターの成員であるという条件は、ここでは、望ましい初期配置としての必要条件であるとも十分条件であるとも言えない。しかしながら、現在の研究段階においては、このような初期代表点をもちいることが最も実用的であろう。実際の分析においては、真のクラスターは未知であるので、何らかの方法で真のクラスターにできる限り類似したクラスターを仮定する必要がある。その上で、この仮定されたクラスターのそれぞれから1つずつデータ・ユニットを選んで初期代表点とするか、あるいは仮定されたクラスターに基づいて初期分割することになる。その具体的方法としては、Milligan & Sokol (1980) や辻本・大島 (1980) が提唱するように、階層的クラスター分析と組み合わせてクラスタリングを行なうことが望ましいと思われる。

初期配置の影響がより少ないクラスタリング=アルゴリズムに関しては、本研究では1つの注目すべき結果が得られた。MacQueen's K-means法が他の3手法よりも初期配置の影響に敏感であることは、先回の報告(辻本・大島, 1980)で述べた結果の1つであり、本研究の結果においても認められた。しかしながら、ここでの分析結果について検討してみると、その原因は、単に、クラスター成員が変化しなくなるまで代表点の更新とデータ・ユニットの割り当てを反復するか否かという点にあるように思われる。すなわち、MacQueen's K-means法のアルゴリズムでは、他のクラスタリング手法と異なり、この種の反復計算が行なわれないという点があるが、その原因であると考えられる。今回行なった分析で、MacQueen's K-means法をもちいた場合、設定されたクラスターを復元できなかった場合に生じたクラスターは、図5と図6に示したどちらかである。どちらの場合についても、それらのクラスターの重心を代表点として、再度代表点を固定したまま全データ・ユニットの再割り当てを行なえば、明らかに設定されていたクラスターが完全に復元される。実際、図5、図6に示されたクラスターは共に、Forgy法によって設定されていたクラスターが完全に復元された場合のうちのある場合において、そのクラスタリング過程の途中結果として見出すことができるのである。図5に示されたクラスターについては、同様に、Jancey法、convergent K-means法によるクラスタリング過程にも見出される。したがって、MacQueen's K-means法に、クラスター成員が変化しなくなるまで

代表点の更新とデータ・ユニットの割り当てを反復するというアルゴリズムをつけ加えれば、少なくともここでもちいたデータ・初期代表点についてはどんな場合でも、設定されたクラスターは完全に復元される。ここでの結果からは、このよう改良されたアルゴリズムは、初期配置の影響がより少ないと言えよう。もちろん、この点についても、単純な構造しかもたないデータをもちいたからこのような結果がでたのではないか、という批判はあてはまる。先の望ましい初期配置の特性という問題も含めて、今後、組織的なシミュレーション研究、あるいは実際のデータへの適用という形での検討が必要であろう。

V 要約

簡単な人工データをもちいて、非階層的クラスター分析のクラスター数固定の分割最適化法(分割型クラスタリング手法)と呼ばれる、Forgy法、Jancey法、MacQueen's K-means法、convergent K-means法の4つのクラスタリング手法について、望ましい初期配置の特性及び初期配置の影響の少ないクラスタリング=アルゴリズムの特性に関する手掛りを得るために、初期配置の影響という点からの比較を試みた。望ましい初期配置の特性については明確な知見は得られなかったが、初期配置の影響のより少ないクラスタリング=アルゴリズムについては、MacQueen's K-means法に代表点の更新及びデータ・ユニットの割り当ての反復計算をつけ加えたものが、そのようなアルゴリズムとして示唆された。

文 献

- Anderberg, M. R. 1973 *Cluster analysis for applications*. Academic Press.
- Milligan, G. W. 1980 An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, **45**, 325-342.
- Milligan, G. W. & Sokol, L. M. 1980 A two-stage clustering algorithm with robust recovery characteristics. *Educational & Psychological Measurement*, **40**, 755-759.
- 辻本英夫・大島将人 1980 非階層的クラスター分析の分割最適化法における初期配置の影響 名古屋大学教育学部紀要—教育心理学科一, **27**, 135-145.

(1981年7月31日 受稿)