小地域集計データによる地域クラスタリングにおける幾何学的制約 とパラメータ推定の安定性の関係について

阪田知彦・吉川 徹

A study on the relations between stability of parameter estimation and geometrical restriction
in regional clustering based on small regional unit data

Tomohiko SAKATA and Tohru YOSHIKAWA

Abstract: This study is a basic research towards a proposal of regional clustering methods for analyzing regional data which is totaled per geometrical irregular shapes. To this end, this paper analyzed the relations between stability of distribution parameter estimation and geometrical restriction in regional clustering in the case of numbers of business use buildings by floorage in the center of Tokyo Wards. First, we proposed two methods of regional clustering with Akaike Information Criterion; one considers the maximum length of the common edge of two polygons; the other considers their maximum area. Second, based on the two methods and the usual method, we composed the suitable polygons from 402 units within CHUO-KU using the number of buildings by floorage scaling estimated by the Tokyo Metropolitan Government GIS data. Third, we compared the stability of parameter estimation of the three unit patterns.

Keywords: 幾何学的制約 (Geometrical Restriction), 地域クラスタリング (Regional Clustering), 調査区単位 (Enumeration District Number Units), 赤池情報量規準 (Akaike Information Criterion), 延べ床面積 (Floorage), 東京都心部 (Center of Tokyo Wards)

1. はじめに

本研究は、幾何学的に不整形な基礎集計単位で集計された地域データを解析するための地域クラスタリング手法の提案に向けた基礎的研究である。特に本研究は、延べ床面積規模別棟数を基に事業所系建物群の空間的特性を把握する際の適切な集計単位について、集計単位の幾何学的性状と集計単位内の分布パラメータ推定に関する安定性との関係を実証

的に分析したものである¹⁾。

近年、各種の小地域統計データの整備が進んだことや、計算機の高性能化に伴い、小地域単位での都市計画分析は従来よりも手軽に行えるようになってきた。この中には地理情報システム技術を活用し、建物や敷地などの個別の事物のデータから様々な集計単位でのデータを生成し、都市の様態を詳細に分析しようという試みも多い。例えば、阪田・吉川(2001a)では、東京都都市計画地理情報システム(以下、東京都GIS)建物データより推定した建物ごとの延べ床面積を1/4地域メッシュ(約250×250m)単位で集計して延べ床面積規模別棟数を求め、東京都心部における事業所系建物分布について

阪田：〒305-0802 茨城県つくば市立原1

独立行政法人建築研究所 住宅・都市研究グループ
Department of Housing and Urban Planning, Independent
Administrative Institutions Building Research Institute,
1 Tachihara Tsukuba, Ibaraki-pref. 305-0802 Japan
Tel. 0298-79-0698 or 64-3805
E-mail:sakata@kenken.go.jp

図1 基礎集計単位の大きさと街区・建物との位相関係

分析している。また、国勢調査基本単位区集計や事業所・企業統計調査調査区集計など、国などが公表する統計調査集計でも集計単位が小さい小地域統計データが利用できるようになってきている。ただし、こうした小さな集計単位によって指標を観察する場合、大きな集計単位で見るとよりも偶然性に左右されやすくなる。こうした指標の過度な詳細性による不安定さを回避するには、集計単位を適切にグルーピングするプロセス（いわゆる地域クラスタリング）が必要となる。

こうした視点からの先行研究としては前掲の阪田・吉川（2001a）の他に、木島・渡辺（1972）、青木ほか（1986）、玉川（1987）、樋口・玉川（1988）、中谷（1996）、吉川（1999）、大佛・中山（2001）、伊藤（2002）などがあげられる。ただし、これら先行研究で議論の対象としているのは、集計単位内の「分布のパラメータ推定の安定性」という観点のみであり、かつ中谷（1996）などの少数例を除いて、多くの研究は幾何学的に整形なメッシュを基にした検討である。そこで本研究では、小地域データを分析するためには、次の2つの検討が必要であると考えた。

（1）不整形な基礎集計単位のクラスタリングによって不整形な集計単位を構成する手法の検討

任意の地域における空間的特性をなるべく良く捉えられるように地域クラスターを構成するには、可能な限り細密な基礎集計単位を基にクラスタリングを行うことが望ましい。メッシュシステムでは、こうした操作が簡単にできる。しかしメッシュシ

テムに基づいて基礎集計単位を構成した場合、メッシュを細かくしていくと、そのメッシュに内包される建物や敷地などの事物の数が減ると同時に、その内包のされ方が歪（いびつ）になること（図1）がある。どれくらい歪になるかについての定量的分析は続稿に譲るとして、この現象を回避する代替案として、基礎集計単位を街区データや事業所・企業統計調査調査区データに代表される不整形な領域とすることが考えられる。ただし、不整形な基礎集計単位から得られる地域クラスターは不整形なものになる。

（2）クラスタリング後の集計単位の形状をなるべく単純なものにする手法の検討

クラスタリング後の集計単位内の分布パラメータは、その空間的特性に応じた最適化が図られており、空間統計学的にはクラスタリング前と比較してより容易に結果を解釈することが可能になっているはずである。ただし気を付けなければならないのは、クラスタリング後の集計単位形状が複雑であると、集計結果を解釈することが逆に困難になることである。中谷（1996）の提案したクラスタリング手法では、集計結果の解釈しやすさを考慮して飛び地のクラスタリングを避けている。この考え方を推し進めれば、飛び地になっても過度に形状が複雑な集計単位は、結果の解釈がやはり困難であるので、避けるべきであると考えられる。すなわち本来、「結果が読みやすい集計単位を構成する」ことを目的に、分布パラメータの過度な詳細性を解消するためにクラスタリングを施したにもかかわらず、共通の分布パラメータを持つ領域を示す集計単位の形状そのものが複雑になるという、別の意味での過度な詳細性が発生するために、結果の解釈が逆に困難になりかねない。特にこの現象は、不整形な基礎集計単位を基にした地域クラスタリングでは頻出する恐れがある。これではクラスタリングを施す意義自体が薄れてしまう。したがって、不整形な領域に対するクラスタリングでは、集計単位内の分布パラメータ推定の安定化と、統合後の集計単位形状の単純化がともに図れてこそ、本当の意味で「結果が読みやすい集計単位を構成できた」といえる。この観点を

加味した先行研究は管見では少ない。したがって、上述の現象が実際の市街地でどのように起こりうるのかを分析することから始める必要がある。

これらの課題に対して本稿では、小地域統計データの分析への適用も勘案して、街区と同様の規模の基礎集計単位を用いた、適切な集計単位を構成するためのクラスタリングに関する基礎的分析を実際の市街地データを用いて行うことにする。その視点としては、従来からの集計単位内の分布パラメータ推定の安定化に加えて、クラスタリングの過程で幾何学的制約を課した場合における、集計単位形状の単純さとクラスタリングの程度のトレードオフの状況に着目して考察することにする。

2. 幾何学的制約を課しつつ適切な集計単位を構成する手法

ここでは地域クラスタリング手法に幾何学的制約を課することによって、上記の目的を満たす手法を探索する。

2.1 地域クラスタリング手法の理論

地域クラスタリングの手法と一口に言っても、その原理はいくつか考えられる。本研究では、それらの中でも簡明なモデルによってクラスタリングが可能な赤池情報量規準（以下、AIC）による方法（中谷，1996；吉川，1999；阪田・吉川，2001a；大佛・中山，2001）を使用することにする。特に本稿では吉川（1999），阪田・吉川（2001a）と同様の多項分布に従うデータを念頭におき、延べ床面積規模別棟数データを例として分析を進める。

ここで、隣接している2つの集計単位（以下、調査区境域データがポリゴンであることより、単にポリゴンと呼ぶ）を考える。今、ポリゴン*i*内に含まれる建物が、ランダムかつ独立に延べ床面積規模のカテゴリー*j*のどれに含まれるかが、多項分布に従うものと仮定する。この2つのポリゴンをそれぞれ独立の多項分布モデルとして考えた場合と、2つのポリゴンを統合して1つの多項分布モデルとして考えた場合とを設定し、両方のモデルのAICを計算すれば、この2つの集計単位を統合すべきかが判断出

来る。このとき、この2つのモデルのAICの差であるDAICは、自由度を*C*とすると、

$$DAIC = (-2) \sum_{j=1}^C \left(\sum_{i=i_1, i_2} n(i, j) \right) \log \left(\frac{\sum_{i=i_1, i_2} n(i, j)}{\sum_{i=i_1, i_2} \sum_{j=1}^C n(i, j)} \right) - (-2) \sum_{i=i_1, i_2} \sum_{j=1}^C n(i, j) \log \left(\frac{n(i, j)}{\sum_{j=1}^C n(i, j)} \right) - 2(C-1)$$

・・・式 (1)

であたえられ、DAICが負であれば、これらのポリゴンを統合してもよいと判断出来る。

上記の方法では、ある対象地域内のそれぞれの集計単位のもつAICの改善量が最大となる組合せ（最小AICモデル；MAICEと呼ばれる）を解くこと（AIC最小化問題）になる。これを厳密に捉えれば対象地域内でのMAICEを求めるには、対象地域内の集計単位全ての組合せについて解くことになり、計算量が爆発的に大きくなり計算不能に陥る。

そこで代替案として、中谷（1996）と同様にAICをもとに隣接する2つの調査区ポリゴンについてのDAICを計算し、これを基に逐次的に統合していく計算プロセスを採用することにした。この方法では、必ずしもMAICEが得られるとは限らないが、近似的に求めたMAICE（以下、本研究では特に断らない限り、MAICEとはこの近似MAICEのことを指す）によって適切な集計単位を構成していく。

図2 最大接辺長条件・最大接面積条件

2.2 統合するポリゴンのペアの候補を選択する際の幾何学的制約

上記の手法の問題点として、統合に関して何らの幾何学的制約が課せられていないために、結果と

して得られる集計単位が単純な形になる保証がないことがあげられる。そこで本研究では、統合するポリゴンのペアの候補を選択する際に、一定の幾何学的制約条件を課すことにする。

まず大前提として、ある対象地域内にあるすべてのポリゴンから隣り合う（接する）ポリゴンを選び出すことは前述の通りである。選び出す方法は次のようにした。メッシュシステムであれば、自分のメッシュの位置関係より、隣り合う上下左右のメッシュの位置関係をメッシュコードより算出するのは比較的簡単である。しかし、本分析では形状が不整形なポリゴンから選び出す必要があるので隣り合うポリゴンを抽出するのは格段に難しくなる。そこで、GISの空間解析機能により自らのポリゴンに薄いバッファを生成させ、このバッファと交差するポリゴンを選択候補とする。その中から、1ペアを選び出す制約条件として、

条件a：選択候補の中で接している辺の長さが最大のもの（最大接辺長条件）

条件b：選択候補の中で接しているポリゴンの面積と自分のポリゴンの面積の合計が最大のもの（最大接面積条件）

条件c：選択候補の中で自分のポリゴンと統合した際のAICの差が最小であるもの（MAICE条件）の3つを設定した。

このうち、条件a、bはどのような接し方をしているかを問題とする幾何学的制約を付けたものである（図2）。これらは、ポリゴン同士の幾何学的な連続性を最大化することによって、統合後のポリゴンがなるべく単純な形状になることを目指している。

一方で条件cは比較対照のため、幾何学的制約を付けていない計算条件である。これは、中谷（1996）のポアソンモデルに基づくクラスタリング手法を多項分布モデルで再定義した吉川（1999）、阪田・吉川（2001a）と同様のものである。

2.3 計算手順

前節の手法に関して、具体的な計算の流れを図3に示す。まず、調査区境域ポリゴンより条件a、b、cのうち何れかひとつの制約条件を考慮しつつ隣接

している2つのポリゴンを選択する。次に、このポリゴン間でDAICを計算する。これを全ポリゴン分繰返し、これらのうちDAICが最小であるポリゴンの組合せを選び出し、お互いを統合する。そして、この統合したポリゴンを含めて、再度DAICの計算を全ポリゴンについて繰返す。この一連の計算が $DAIC > 0$ になったら計算終了とし、この時点での集計単位を考察に用いる。

図3 計算手順

図4 分析対象地域

3. 適用

3.1 対象地域の概要と計算結果

では実際の規模データに適用する。ここでは、東京都GIS建物データより推定した建物単位での延べ床面積（阪田・吉川，2001a；阪田，2002）を基に、0-50、50-150、150-500、500-1500、1500-3000、3000m²以上の6段階の延べ床面積レンジ別で集計した規模別事業所系棟数データを用いる。また、分析対象地域は東京都中央区内の402調査区（図4）²⁾とする。これに基づき計算を行った結果が表1である。

3.2 統合計算結果の幾何学的な特徴

2.2で設定した3つの制約条件によって構成した集計単位領域と、その領域内の延べ床面積規模別棟数分布を図5～7に示す。まず、その領域の幾何学的特徴を定量的に分析する。吉川ほか(1986)は、図8のような不整形な領域に対する形状係数(Haggett, et.al, 1977)で定量化して形の特徴を把握しているので、本研究でも形状係数を適用する。

ここで、 S , L , M をそれぞれ統合後ポリゴンの面積、周長、長径(図8)を示すものとする、次の4つの指標 $C_1 \sim C_4$ が導かれる。

$$C_1 = S / L \quad \dots \text{式(2)}$$

$$C_2 = S / 2M \quad \dots \text{式(3)}$$

$$C_3 = 4\pi S / L^2 \quad \dots \text{式(4)}$$

$$C_4 = 4S / \pi M^2 \quad \dots \text{式(5)}$$

これらの指標のうち本分析の分析目的に合うものとして、ここでは統合後のポリゴン形状の単純さを示す C_3 と C_4 を観察する。この2つの指標についてそれぞれの統合結果ごとに算出した結果が表2、その指標の分布を示したのが図9である。

図5 統合計算結果(最大接辺長条件)

図6 統合計算結果(最大接面積条件)

図7 統合計算結果(MAICE条件)

まず、 C_3 について見る。この指標自体は、1に近いほど形状として単純であるといえ、平均・中央値ともに、最大接辺長>最大接面積>MAICEの順である。したがって、この3つの条件のうち、最大接

辺長条件を設定することにより集計単位形状の単純化が図れることがいえる。また、指標の分布の幅を示す分散や四分位数間範囲を見ると、最大接辺長<最大接面積<MAICEの順であることから、もっとも単純な集計単位形状を得ることが出来るのは最大接辺長条件の場合であることがわかる。

次に、 C_4 について見る。この指標は、巨視的な視点から1に近いほど形状として円形(矩形)に近いことを示す指標である。先程の C_3 の挙動と違い、今回設定した3つの計算条件の間では決定的な差異は無いが、この場合でも最大接辺長>最大接面積>MAICEの順でより正方形に近い(団子状の)集計単位形状を構成出来ることがわかる。

さらに、上述の定量的解釈を補完するため、領域の特徴について定性的・即地的な考察を行う(図5~7)。一見して、制約条件による統合後の集計単位

位の広がり方に違いが見られる。最大接面積条件(図6)やMAICE(図7)の場合では、対象地域の中央部に大きく形状が複雑な集計単位が広がっているが、もっとも単純な集計単位を構成できた最大接辺長条件(図5)の場合では、比較するとコンパクトな集計単位が広がっている。すなわち、最大接辺長条件によって形状が単純であるのに加えて、規模も揃った集計単位が構成されている。以上の定性的考察は表1の面積の平均や標準偏差からも裏付けられる。

また、広幅員道路と集計単位の位置関係についても、制約条件によって違いが見られる。最大接面積条件やMAICE条件では、広幅員道路と無関係な形状の集計単位が少なからず見られる。それに比べて最大接辺長条件では、広幅員道路の両側が均等に統合されたものや、広幅員道路に囲まれた単純な形態のものが多く、東京に関する都市計画研究では、

図8 面積・周長・長径

表1 計算結果

	統合前	計算条件		
		最大接辺長	最大接面積	MAICE
ポリゴン数	402	120	105	86
計算時間※ (時間:分:秒)	—	9:02:02	7:33:32	14:38:26
平均面積(㎡)	4528.9	15176.9	17345.0	21177.0
平均周長(m)	288.4	499.0	634.3	909.4
面積標準偏差(㎡)	2996.5	16084.5	46325.8	18173.6
周長標準偏差(m)	109.1	285.3	1153.5	638.3

※ 計算時間の計測環境: PentiumIII500MHz, RAM:512MB, WinNT4.0, MapInfo6.0J+MapBasic

表2 形状指標の比較

形状指標 計算条件	C_3			C_4		
	最大接辺長	最大接面積	MAICE	最大接辺長	最大接面積	MAICE
平均	0.660	0.568	0.400	0.494	0.430	0.326
中央値	0.700	0.606	0.366	0.516	0.440	0.304
標準偏差	0.116	0.191	0.197	0.103	0.134	0.128
分散	0.013	0.037	0.039	0.011	0.018	0.016
最小	0.306	0.046	0.091	0.198	0.109	0.066
最大	0.807	0.807	0.770	0.645	0.628	0.614
合計	79.226	59.681	34.404	59.334	45.099	28.078
ポリゴン数	120	105	86	120	105	86

図9 形状指標の分布

市街地の空間的特性を把握する上で、広幅員道路沿道の「ガワ」と、それに囲まれた「アン」という二重構造が重要な視点であるとされている(権ほか, 1997など)。対象地域においてもこの構造の存在が指摘できる(阪田・吉川, 2001b)。このことから、広幅員道路に関して集計単位の統合のされ方が単純な方が結果の解釈が容易になる。今回設定した制約条件では、最大接辺長条件の結果がこうした要求に答えやすいといえる。

以上より、幾何学的な単純性を求めるための幾何学的制約条件としては、最大接辺長条件が有効であることがわかった。

3.3 計算過程の観察

以上の幾何学的性状の分析と比較するために、情報量統計学の観点からの考察が必要である。

そこで、今回の3つの計算条件それぞれのDAICの変化(図10)を見ていくことにしよう。まず、全ポリゴンのペアの中で最小のDAICを常に統合条件としていたので概ね試行が進むにつれてDAICが大きくなるが、時折DAICが小さくなることがある。これは次の事情による。直前の試行で統合候補になった2つのポリゴンの周囲には、より短い接辺長やより小さな合計面積で接しているために統合候補にならな

かったポリゴンが存在する。こうしたポリゴンが、直前の統合によって新たに生成されたポリゴンとの間で制約条件を満たし、かつ全ポリゴンのうちで最小のDAICポリゴンのペアになる場合に、上記のようにDAICが再び小さくなることがある。また、計算の後半でこのような挙動が多数見られるのは、統合によってポリゴン数が少なくなるのと共に、こうした統合回数の多いポリゴンが幾何学的制約条件を満たす確率が高まるという2つの要因が複合的に働いているためであると考えられる。さらにこの挙動は、最大接辺長>最大接面積>MAICEの順で起こ

図10 DAICの挙動

図11 AIC改善量の比較

りやすいこともわかる。

ここから想定出来るのは、幾何学的な制約を優先させたために、AICの改善（パラメータ推定の安定性）を妨げているのではないかという仮説である。そこで、試行回数とAICの改善量を示したのが図11である。AICの最大改善量（試行の終局値）を見るかぎり、この対象地域全体としてもっともパラメータ推定から見て安定した集計単位が得られる条件は幾何学的制約条件を付けないMAICEであり、最大接辺長条件は最大接面積条件よりもパラメータ推定の安定性としては不利となる幾何学的制約であることがわかる。

4. まとめと展望

4.1 まとめ

本稿では、不整形で小さい集計単位内の多項分布データを基にクラスタリングする際の幾何学的制約とパラメータ推定の安定性の関係についての基礎的分析を行ってきた。特に、接するポリゴン群から1組のポリゴンのペアを選び出す際に2種類の幾何学的な制約を付けた場合、その計算後の形状はより単純になるがパラメータ推定の安定性の観点からは不利になるトレードオフの状況を具体的・定量的に観察することができた。

また、統合後の領域に対しての定性的・即地的考察では、広幅員道路との関係に着目し、幾何学的制約条件による違いを明らかにし、その都市空間構造の把握における意味を示した。このことから、単に集計単位内における分布パラメータの安定性だけに着目するのでは分からない都市の様相を把握するために、他の制約条件を加味した地域クラスタリングが有用であることが示唆される。本研究での実証分析によって、幾何学的制約条件がその制約条件として有効であるもののひとつであることが示された。

4.2 展望

今後の展望としては、次の事項があげられる。まず、本稿で取り上げたような不整形な基礎集計単位に対する幾何学的単純さと集計単位内のパラメータ推定の安定性を同時に最適化するトレードオフ問

題の定式化が必要である。これには前述のように、基礎集計単位の幾何学的特性の実証的把握や、幾何学的安定性の理論モデルの構築が必要となる。さらに、メッシュ状の基礎集計単位に対する同様の手法の検討も考えられる。これにはまず、メッシュの細密化に伴う、個別的データ（例えば、建物データや敷地・土地利用データなど）の内包関係の定量的分析が必要である。また、分析の目的に応じた制約条件の選択手法についての検討も必要である。これらは課題としたい。

【注釈】

- 1) 本稿は、阪田（2002）、阪田・吉川（2001b, 2002）を再構成したものである。
- 2) 本研究では、東京都心部でも事業所系建物が集積している代表的な地域であり、かつ事業所・企業統計調査調査区が、阪田・吉川（2001b）での1/4地域メッシュよりも細密に設定されている地域を対象地域とした。

【参考文献】

- 青木義次・大佛俊泰・尚 炯鐘（1986）都市メッシュデータ解析におけるメッシュサイズの効果，日本都市計画学会学術研究論文集，**21**，247-252。
- 伊藤香織（2002）時空間を特徴づける領域分割の最適化に関する研究，日本建築学会計画系論文集，**556**，341-348。
- 大佛俊泰・中山大樹（2001）空間クラスターと情報量損失を考慮した空間データの視覚化，地理情報システム学会講演論文集，**10**，159-162。
- 木島安史・渡辺仁史（1972）アーバンダイメンション その2・重複集計によるメッシュアナリシス，日本建築学会論文集，**198**，37-42。
- 権 昌希・玉川英則・市古太郎（1997）ポリゴンデータを用いた大都市幹線道路周辺市街地の空間構造分析，地理情報システム学会講演論文集，**6**，267-270。
- 阪田知彦（2002）『東京都区部における事業所系建物の延べ床面積規模分布の把握とその推定精度に関する研究』，東京都立大学大学院工学研究科博士論文。
- 阪田知彦・吉川 徹（2001a）GIS建物データとメッシュシステムに立脚した東京都区部における延べ床面積規模別事業所系建築物の地理的分布の分析，日本建築学会計画系論文集，**545**，189-196。
- 阪田知彦・吉川 徹（2001b）調査区単位内の延べ面積規模別棟数から見た東京都心部の事業所系建物群の空間的特性，日本建築学会大会学術講演梗概集，F-1，

241-242.

阪田知彦・吉川 徹 (2002) 多項分布特性を持つ調査区単位データをクラスタリングする際の幾何学的制約と情報量の関係, 日本建築学会大会学術講演梗概集, F-1, 659-660.

玉川英則 (1987) 土地利用比率の同質性からみた最適メッシュ規模に関する考察, 日本都市計画学会学術研究論文集, 22, 229-234.

中谷友樹 (1996) 死亡率地図における空間単位の情報量統計学的評価ー地理情報システムによる疾病地図解析システムの構築に向けてー, GISー理論と応用, 4 (1), 53-60.

樋口忠彦・玉川英則・アグス ブディ プルノモ イサク

(1988) 連続変量の最適メッシュ区分についてーメソナルマップを用いてのケーススタディーー, 日本都市計画学会学術研究論文集, 23, 37-42.

吉川 徹 (1999) 東京都多摩地区南部の鉄道沿線の通勤時間分布の地域的差異の分析, 日本建築学会大会学術講演梗概集, F-1, 81-82.

吉川 徹・坂内正夫・大沢祐・峰村治実 (1986) ベクトル化処理による土地利用図のデータ化とその定量的分析のための手法に関する一考察, 日本不動産学会秋季大会学術講演集, 31-34.

Haggett, P, Cliff, A.D. & Frey, A. (1977) *Locational Analysis in Human Geography (Second Edition)*, London: Edward Arnold Ltd.