

# 空間分析におけるクラスター分析と自己組織写像手法の比較\*

## A Comparative Study of Cluster Analysis and the Self-Organizing Map Algorithm in Spatial Analysis\*

李 燕\*\*

By Yan LI\*\*

### 1. はじめに

都市や地域の社会的・空間的特性を分析することは都市計画を含む多くの分野において必要不可欠である。しかし、従来ではデータの整備や分析手法において不十分な点があるため、詳細なデータを用いて都市や地域全体を対象とした空間分析研究は必ずしも多いとは言えない。しかし近年、情報化による空間データの整備、地理情報システムやデータマイニング技術の発展により、このような研究が可能になりつつある。

本研究は、国勢調査の丁字詳細データを用いた大分県別府市の社会地域分析(Social area analysis)を通じて、従来の統計分析手法の一つであるクラスター分析と、近年開発された教師無し人工ニューラルネットワークの一種である自己組織写像手法 (SOM: Self-Organizing Map) の比較分析により、それぞれの方法の特徴および空間分析への適応性を考察することを目的とする。

### 2. 社会地域分析について

「社会地域」とは、都市において、異なる社会特性をもつ居住者による住み分けによって形成される空間的構造であり、居住地域 (Residential area) とも呼ばれる。スラムやゲットー、チャイナタウンなどがその分かりやすい例であるが、高級住宅街や職人の町なども社会地域の一つである。

人種が多く、社会格差の激しい欧米では、社会地域は普遍的な現象として存在することが多くの研究で確認されている。1920年代のシカゴ学派は、シカゴ市の社会空間構造が同心円状であると分析し、その背後にある社会過程を競争 (competition)、闘争 (struggle)、応化 (accommodation)、同化 (assimilation) の四つに分けてとらえた<sup>1)</sup>。1950年

代においては、Shevky and Bellはサンフランシスコを例として、アメリカにおける居住分離は「社会地位」、「アーバンニズム」(ライフスタイルや家族構成など) および「民族」(エスニック・ステータス) の3つの要因に影響されると指摘し<sup>2)</sup>、その後、社会地域分析が欧米諸国そしてその他の国において行われるようになった。

日本においては、外国人が少ないため、人種的・民族的住み分けについては大都市以外の研究は報告されていないが、人口密度から見た「都市空間構造」や住宅の質や環境を用いた「居住地域」の研究が欧米の「社会地域分析」とほぼ同時代、すなわち1950年代から1970年代を中心に、多くなされている<sup>3)</sup>。

手法論から見れば、Shevky and Bellの当時は、社会変動の理論から社会地域形成される要因を仮定し実証するという演繹的な方法を取っていたが、その後は、多変量解析を用いた帰納的な手法が主となった。その中でも、変数の選択・統合を行った上で分類する主成分分析・因子分析が多く用いられている<sup>4)</sup>。クラスター分析は直接用いることができるので便利だと考えられるが、それを用いた研究は意外と少ない。手法が多く、設定によって結果が異なる場合がある上、変数が多いと意味のあるクラスターが析出されにくいのが理由ではないかと思われる。

社会地域の分析は、都市計画の実務においても、非常に重要であるが、実際の応用が必ずしも多くはない。その原因として、まずは必要なデータの整備が不十分であったことである。次に線形的な多変量解析に基づく従来の社会地域分析手法は、多様な現実社会を捉えにくい場合が多いことが考えられる。もう一つの原因は、実際の都市の社会地域を扱う際、仮に社会地域の特徴が統計的に分かっても、その空間分布を可視化することがたいへん労力が要るので、躊躇されるのではないかと思われる。

本研究は、伝統的なクラスター分析と近年開発された自己組織写像手法 (SOM) を用いて社会地域分析を行い、それぞれの手法の有効性を考察する。結果の可視化には地理情報システムを用いることにした。

### 3. 研究対象エリアおよびデータセット

本研究はケーススタディとして、大分県別府市

\*キーワード: 社会地域、SOM、クラスター分析、GIS

\*\*正員、工博、立命館アジア太平洋大学アジア太平洋学部  
(大分県別府市十文字原1-1、  
TEL0977-78-1052、FAX0977-78-1052)

(人口12万人)を対象に、2000年に行われた国勢調査の小地域集計から、次の86個の変数を抽出した。

(1) 総人口、世帯数、人口密度。

(2) 5歳間隔の男女別人口の割合。男女それぞれ0歳から99歳まで5歳間隔の20グループおよび100歳以上の1グループに分けられ、トータル42個の変数になる。

(3) 外国人男性・女性の数

(4) 住宅形態別の割合。一戸建て、長屋、1-2階建ての共同住宅、3-5階建ての共同住宅、6-10階建ての共同住宅、11階建て以上の共同住宅の6種類あるが、それぞれ持家、借家、間借りに区別されているので、計18個の変数になる。

(5) 居住期間別の割合。男女別にそれぞれ1年以下、1-4年、5-9年、10-19年、20年以上、出生時からの6区分が設けられているので、全体では、12個の変数になる。

(6) 主な家計収入別の世帯の割合。賃金・給料のみ、恩給・年金のみ、賃金・給料+その他の収入、恩給・年金+その他の収入、仕送り、農業、農業以外の事業収入、内職とその他の9種類に分けられる。

なお、別府市は180個の国勢調査基本単位があるが、人口のほとんどない山間部の17個を除いたため、データセットには変数が86個、調査区数が163になる。なお、計算に使われる際には、各変数について標準化(平均が0、標準偏差が1)を行った。

#### 4. 分析手法について

本研究は、上述の別府市のデータについて、伝統的な統計分析の一つであるクラスター分析と、ニューラルネットワークの手法である自己組織写像 (SOM) を用いて社会地域分析を行い、各手法の有効性について検討する。以下では、まずこの二種類の手法について簡単に紹介する。

##### 4.1 クラスター分析

クラスター分析とは、N個の観察単位について、それぞれK個の変数値が与えられたとき、近いデータをもつ観察単位は同一クラスター(区分)に、離れたデータをもつ観察単位は別のクラスターになるよう区分けを定める分析である<sup>6)</sup>。さまざまな手法が提案されているが、大きく分けるとデータの分類が階層的になされる階層型手法と、特定のクラスター数に分類する非階層的手法とがある。

階層的手法としては、クラスター数をあらかじめ特定せず、クラスター数を逐次増加あるいは逐次減少する方法で、分類体系(デンドログラム/樹形図)が得

られることが特徴である。一方、非階層的な手法はあらかじめ指定したクラスター数に最適な区分けをする方法である。いずれの手法においても、クラスター間の距離の定義やクラスターを集約することによって生じる情報量のロス、クラスター代表点の与え方、逐次増加や減少の方法、基礎データの処理方法(変数の取捨、標準化)などによって、分類結果が異なることはしばしば起きる。

##### 4.2 自己組織写像 (SOM)

「SOM」は、大脳皮質の感覚野に見られる機能地図つまり特徴地図が、知覚経験に基づくある種の学習によって形成されていくありさまを説明する数理モデルである。1980年代Kohonenによる提案され<sup>7)</sup>、日本においては、1990年代後半から、経済学、土木工学、情報工学、商業分野などの実用例が報告されるようになった<sup>8)</sup>。

基本的なSOMは2次元に配列されたノード(神経細胞)からなっている(図1の出力スペース)。この出力スペースは必要に応じて1次元や3次元などに行うことができる。各ノードには、参照ベクトルと呼ばれるn次元ベクトルが初期値として与えられている。

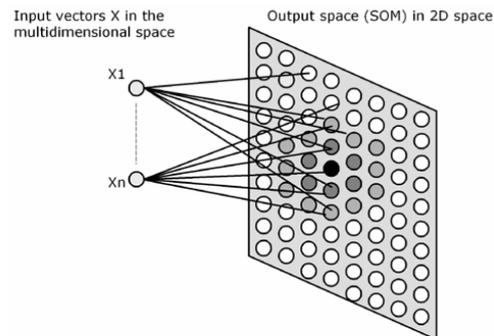


図1 SOMの基本概念<sup>9)</sup>

同じくn次元の入力ベクトルXから入力x1を与えられると、x1に最も合致する参照ベクトルをもつノードが「勝者」となり、そのノードおよびその周囲のノードがこの入力を学習する。ここで、入力x1に対する「学習」は、入力に合致する方向へすべてのノードの参照ベクトルを変更することでなされるが、ノード配列上のユークリッド距離に影響されるため、主として勝者の周囲で起こる。その結果、後で再びx1が入力されたときやx1に類似する入力が入力されたときには、このときの勝者がまた勝者になって学習を起しやすい傾向が生じる。同様に、ほかの入力が与えられたときもそれぞれの勝者を決め、学習を起こす。このようなXからの入力、勝者としてのノードの決定、周囲ノードの学習が繰り返行われ、その結果、もと

もとn次元の入力ベクトルのパターンが、2次元の出力空間に現れる。このように、SOMは高次元のデータや、ベクトル空間上にないデータを、2次元の平面上など、より低次元で容易に観察できる空間に写像する（次元削減する）ことでデータの可視化を実現する。データセットの次元を減らし、線形的な統計手法で見つからない知識を発見できるとされている。

本研究は、上述の別府市のデータについて、伝統的な統計分析手法としてK-means法を用いた非階層的クラスタ分析とWard法を用いた階層的クラスタ分析の2つ、ニューラルネットワークの手法としてSOM法を用いて社会地域分析を行い、各手法が社会地域分析における有効性を検討することにする。ここでK-means法を用いたのは、これが非階層的な手法として最もよく使われているからである。また、Ward法を用いたのは、SOMのアルゴリズムにおいてもWard法に類似した手法（SOM-Ward法）が使われているからである。なお、距離計算については三手法同様、ユークリッド距離を使う。クラスタ分析はSPSS 13.1を、SOMはViscovery SOMine 4.0を用いて計算を行った。

## 5. 社会地域分析における三手法の比較

図2は、地域を2クラスターに分けた場合の結果である。この図から分かるように、K-means法は後ほど他の手法によっても析出されるクラスターIだけを割り出した。一方、SOM法は縦長い対象地域を大きく中部と両端（北部と南部）に分けた。それに対して、Ward法はクラスターIを析出したことと、円で示した部分（以降、中西部と呼ぶ）の分け方でSOM法と異なる。なお、SOM法とWard法は階層的クラスタリング手法なので、3クラスター以降、この基本的な分け方に基づいて細分化していく。K-means法は分類体系が形成されないが、結果を見ると、この中西部の分け方については基本的にSOM法に類似している。紙面の関係で、以下では、地域を8クラスターに分けた場合の各手法による結果を用いて比較を行う。

### 5.1 クラスターの類似性と社会地域になりうるクラスターの数

8クラスターの場合（図3）は、2クラスターの場合と比べると、SOM法は中部地域からクラスターA、両端の地域からクラスターB、C、Dを析出したのに対して、Ward法においては、SOM法のクラスターA、B、Cに位置が類似するA'、B'、C'が析出され

たが、クラスターDは析出されていない。一方、K-means法において中部と両端の構造はSOM法に類似するが、新たに析出されたのはいくつかの小さな調査区だけであった（面積が小さいので、図3では確認しにくい）。

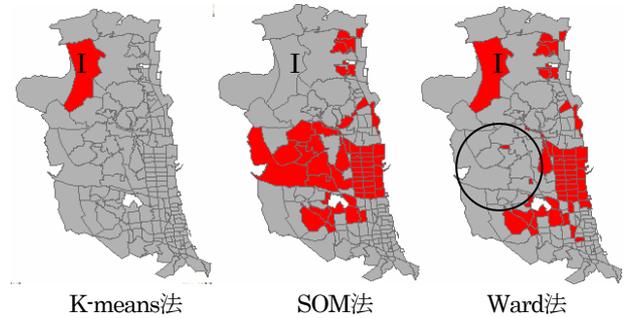


図2 2クラスター分けの地域分布

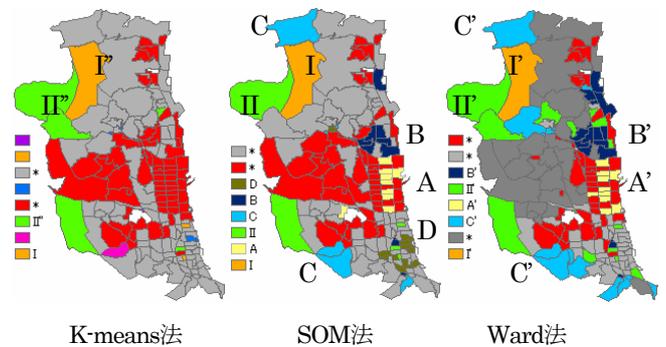


図3 8クラスターの地域分布

3つの手法は共通的にクラスターIとクラスターIIを析出した。IIはひとつの調査区で構成されるので、3つの手法とも同じ範囲である。IIは共に西部の辺縁の調査区をメインとするが、範囲はそれぞれ異なり、Ward法のクラスターII'はより多くの調査区が含まれている（SOM法は5、Ward法は14個の調査区）。

社会地域は複数の調査区、あるいは一定の面積を有する地域であるとする、社会地域となりうるクラスターの数（図3の凡例ではクラスター名か「\*」が付いているクラスター）は、K-means法は4つだけであるのに対して、SOM法とWard法は共に8つになっている。

以上のように、三手法の中、SOM法とWard法はより類似した結果が得られているが、社会地域になりうるクラスターの数を見ると、K-means法では個別の面積の小さい調査区が析出されやすいので、社会地域分析においては、有効な方法とは言えないようである。

### 5.2 共通クラスターの比較

ここでは、3方法で共通として析出されたクラスターIとII、およびSOM法とWard法で共通として析出

表1 SOM法による析出プロセス(2から8クラスターまで)およびクラスターの主な社会特徴

クラスター数	2 →	3 →	4 →	5 →	6 →	7 →	8
中部 西部	若い年齢層、賃 金収入、マンシ ョン借家、短い 居住年数				●さらに若いサラリー マンの核家族のエリア (A)		●残りのエリア
北部 南部	その他	●さらに高齢者(農家)が多い 周辺エリア(II)	(後、細分化:仕送りの若者が多い地域) →	●短い居住期間の高齢者 (老人ホームの多い地域)(C)		●別府大学エリア(B) ●APUエリア(I)	●居住期間の長い自営業 (温泉、旅館、商店街な ど)(D) ●残りのエリア

注:●は8クラスターまで残ったクラスターである。

表2 共通クラスターの統計分析

クラスター	調査区の数		相関係数 SOM-Ward	分散	
	SOM	Ward		SOM	Ward
A, A'	9	10	0.975	0.943	0.82
B, B'	9	14	0.973	0.381	0.232
C, C'	6	14	0.905	0.967	0.525
II, II'	5	14	0.949	1.539	0.501

注:K-means法について、クラスターII”の調査区の数5、相  
関係数:SOMとは0.896、Wardとは0.892、分散1.138

されたクラスターA,B,Cについてデータの比較分析を  
行い、それぞれの方法の特徴をしてみる(表2)。こ  
の表から分かるように、SOM法とWard法は比較的  
に大きな相関をもっているが、SOM法のほうが調査  
区の数少ないのに分散が高い(すなわち平均値から  
離れている)。また、特にクラスターCのように、  
Ward法は地理的に分散した数多くの調査区をひとつ  
のクラスターとして析出されてしまったことから、  
SOM法はより特徴的で、地理的にまとまった地域を  
析出できると言えよう。

### 5.3 クラスターの特徴からみた手法の有効性

析出されたクラスターは、社会地域としての特徴が  
捉えられるかどうかという視点も、手法の有効性を判  
断する上で必要である。紙面の関係で、詳細について  
は文献10を参照されたいが、ここでは、SOM法の8  
クラスターまで析出された地域の特徴とプロセスを表  
1にまとめた。各段階において析出されたクラスター  
は明確な社会的特徴があると考えてよいであろう。

## 6. 終わりに

本稿では、国勢調査のデータを用いて、SOM法お  
よび伝統的なK-means法とWard法による社会地域分  
析を行った。析出されたクラスターの数と範囲、地理

的接近性、統計的特徴および社会的特徴からみて、  
SOM法が優れていると考えられる。Ward法はSOM  
法に近い社会地域が析出できたが、敏感度が高くない  
と見られる。一方、K-means法は特徴的な個別調査  
区が析出されやすいので、社会地域分析には適用しに  
くいと思われる。

強調すべきは、本研究で使用した変数は、人口の社  
会特性を描写するもので、取捨は行われておらず、デ  
ータの統計処理も標準化のみであったということであ  
る。伝統的な統計分析においてはデータの取捨・加工、  
手法の選択などが必要であり、主観的な要因を排除し  
にくい面が大きいことは否定できない。このことから  
みても、SOM法は便利で有効な手法であると言えよ  
う。また、本研究で用いた変数においては、お互いの  
位置関係を表すものは使用していない。にもかかわらず、  
空間的に連続的な社会地域の存在がGISのマッピ  
ングによって確認できたことも興味深い。日本の都市  
においても、欧米のように明白でないかもしれないが、  
社会地域というものが確実に存在していると言えよう。

### 参考文献

- 1) Park, R. E. and Burgess, E. W. (1928): Introduction to the Science of Sociology, Chicago: University of Chicago Press.
- 2) Shevky, E. and Bell, W. (1955): Social Area Analysis. Stanford, CA: Stanford University Press.
- 3) 木内信蔵:都市地理学原理、古今書院、1979。
- 4) 川上秀光:地方中心都市における密度構造の変容、都市計画論文集、1980、pp.73-78。
- 5) Economic Geography, Vol. 47, Supplement: Comparative Factorial Ecology, Jun., 1971
- 6) 上田尚一:クラスター分析、朝倉書店、2003。
- 7) Kohonen, T. (1982). Self-organised formation of topographically correct feature maps. Biological Cybernetics, 43:59-69.
- 8) 徳高平蔵、藤村喜久郎、山川烈監修:自己組織化マップ応用事例集:SOMによる可視化情報処理、海文堂出版、2002.10。
- 9) Koua1, E. L. and Kraak1, M. J. (2004): Geovisualization to support the exploration of large health and demographic survey data, International Journal of Health Geographics 2004, 3:12
- 10) 李 燕、シャンムガナタン スパナ: SOMとGISを用いた社会地域分析の試み、都市計画報告集No5-2(CD-ROM)、2006.8