

空間分析におけるクラスター分析と自己組織化マップ手法の比較*

A Comparative Study of the Cluster Analysis and the Self-Organizing Map 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が、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。その理由の一つとして、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。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、比較的新しい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化マップ手法(SOM: Self-Organizing Map)の比較分析により、それぞれの手法の特徴および空間分析への有効性を考察することを目的とする。

2. 社会地域について

「社会地域」(social area)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地域であり、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、高級住宅街や職人の町などはその明白な例であるが、多くの場合は単純な観察では分かりにくい。

人種が多く、社会格差の激しいアメリカでは、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)の四つとしてとらえた¹⁾。1950年代頃には、Shevky and

Williams (1949)、Bell (1953)とShevky and Bell (1955)はロサンゼルスや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²⁾³⁾⁴⁾、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も行われるようになった。

日本においては、居住する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」の研究は欧米の「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⁵⁾⁶⁾。

手法論から見れば、Shevky and Bellの時代は、社会変動の理論から社会地域が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・統合を行った上で地域を分類する主成分分析・因子分析が多く用いられている⁷⁾。クラスター分析はこれらの統計分析において、変数を統合する道具として使われるが、個体を直接クラスタリングする研究は少ない。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、実際の応用は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が挙げられる。次に多変量解析に基づく従来の社会地域分析手法は、基本的にベクトル空間に扱いにくい情報を切り捨てるため、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。また、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はたいへん労力が要するため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比較的新しい自己組織化マップ手法(SOM)を用いて社会地域分析を行い、それぞれの手法の特徴および空間分析における有効性を考察する。結果の可視化には地理情報システムを用いる。なお、ここでいう「有効性」は、「一定の面積

*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、

TEL:0977-78-1052、E-mail:yanli@apu.ac.jp)

があり、大多数の地域と異なる社会的特徴をもつ地域（すなわち社会地域）を析出できる」という意味で用いる。

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（人口約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。なお、それぞれの段落末の数字は後述の図表に用いる変数の番号である。

(a) 総人口、世帯数、人口密度 (01-03)

(b) 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。(女性：04-24、男性：26-46)

(c) 外国人女性・男性の数 (女性：25、男性：47)

(d) 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。(持ち家：48-53、借家：54-59、間借り：60-65)

(e) 居住期間別の割合 男女別にそれぞれ1年未満、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。

(女性：66-71、男性：72-77)

(f) 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。(78-86)

なお、別府市は180個の国勢調査基本単位があるが、居住者がほとんどいない山間部の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、計算の際には、各変数値について標準化（平均が0、標準偏差が1）を行った。また、標準化された変数値のことをZ値と呼ぶ。

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化マップ (SOM) を用いる。以下では、これらの分析手法について簡単にまとめる。

なお、記述においては、データセットをベクトル X 、個体（すなわち観察単位、ここでは調査区）の数を N 、変数の数を p とし、個体 i のベクトル $(\xi_{i1}, \xi_{i2}, \dots, \xi_{ip})$ を

x_i とする。ここで、 $i=1, 2, \dots, N$ とする。

また、クラスターの重心（すなわち平均ベクトル）を $\bar{x}^c = (\bar{\xi}_1^c, \bar{\xi}_2^c, \dots, \bar{\xi}_p^c)$ と表わし、データセット全体 X の平均ベクトルを $\bar{x} = (\bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_p)$ を表わす。ただし、 $\bar{\xi}_i^c$ はクラスター c に属する変数 i の平均値であり、 $\bar{\xi}_i$ は全個体の変数 i の平均値である。なお、先述のように本研究では各変数について平均が0、分散が1になるように標準化したので、 $\bar{\xi}_i$ は0である。

(1) クラスター分析

クラスター分析とは、 N 個の個体について、それぞれ p 個の変数値が与えられたとき、近いデータをもつ個体は同一クラスターに、離れたデータをもつ個体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。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的な手法と、特定のクラスター数に分類する非階層的な手法とがある^{8) 9)}。

a) 階層的クラスタリング、Ward法

階層的な手法としては、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させる方法で、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。あらかじめ決める個体/クラスターの対に類似度あるいは非類似度（距離）を基準に、次に示す手順によってクラスターが形成される。

① 初期状態として、 N 個の個体それぞれが、1つのクラスターを形成しているものとする。したがって、クラスターの数 K は $K=N$ とする。

② K 個のクラスターの中で最も類似度の大きい（距離の小さい）対を求め、それを1つのクラスターに融合する。 K を $K-1$ として、 $K>1$ ならば手順③へ進み、そうでなければ終了する。

③ 新しく作られたクラスターと他のクラスターとの類似度を計算し、手順②へ戻る。

上の手順はクラスター数を逐次減少する方法であるが、全体を1つのクラスターから出発し、クラスターを徐々に増加する方法もある。なお、③でクラスター C_i とクラスター C_j が融合されたとき $(C_i \cup C_j)$ 、他のクラスター C_k との非類似度は次のように計算される。

$$d(C_i \cup C_j, C_k) = \alpha_i d(C_i, C_k) + \alpha_j d(C_j, C_k) + \beta d(C_i, C_j) + \gamma |d(C_i, C_k) + d(C_j, C_k)| \quad (1)$$

この式において、パラメータ α 、 β 、 γ の与え方によって、最短距離法、最長距離法、メディアン法、群平均法、ウォード (Ward) 法、重心法など、種々のクラスター分析の手法が得られる。その中でも最も広く使われているWard法においては、パラメータは次の式(2)のよ

うになっている。

$$\alpha_i = \frac{l_i + l_k}{l_i + l_j + l_k}, \quad \beta = \frac{-l_k}{l_i + l_j + l_k}, \quad \gamma = 0 \quad (2)$$

ただし、 l_i はクラスター C_i に含まれる個体数である。この非類似度の定義から、Ward 法は、各個体から、その個体を含むクラスターの重心までの平方距離の総和（すなわち偏差平方和）の増分を最小化する方法であることが分かる。これについて詳しくは文献9を参考されたい。

b) 非階層的クラスタリング、K-means 法

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。さまざまなアルゴリズムが提案されているが、K-means法（K-平均法）は、非階層的クラスタリングとしてよく用いられる手法である。その基本的な考え方は次のようである。

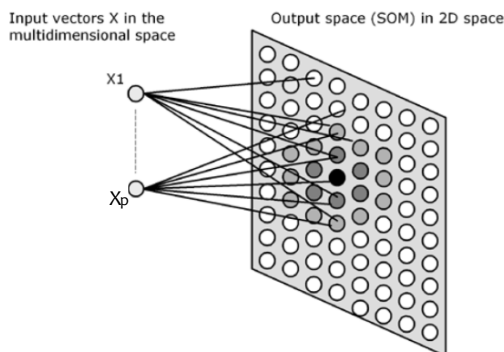
- ①初期条件として、K個数のシード点、すなわち、クラスターを形成する核となる個体を与える。
- ②逐次個体とシード点との距離（通常はユークリッド距離）を計算し、個体を最も近いシード点に分類する。
- ③ある収束条件を設定し、それが収束するまで、シード点の変更を繰り返す。

なお、収束条件としては、通常Ward法で用いられた、クラスター内偏差平方和で表わされる。

(2) 自己組織化マップ (SOM) ¹⁰⁾

「SOM」は、ニューラルネットワークの1つで、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によって提案され¹¹⁾、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、商業分野などでの実用例が報告されるようになった。

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次元などに定義することができる。各ノードには、参照ベクトルと呼ばれる p 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¹²⁾

ここで、 i 番目のノードの参照ベクトルを

$m_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{ip})$ とする。参照ベクトルはすべて長さ1の単位ベクトルに規格化される。一方、SOMに与えられる入力ベクトル $x_n = (\xi_{n1}, \xi_{n2}, \dots, \xi_{np})$ も参照ベクトルと同じく p 次元の単位ベクトルである。

入力 x_n が与えられると、 x_n に最も近い（例えば、ユークリッド距離の最も小さい）参照ベクトルをもつノードが勝者となる。ここで、ノード i とする。入力 x_n 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。

$$m_i := \frac{m_i + h_{ci} x_n}{|m_i + h_{ci} x_n|} \quad (3)$$

ここで、 $:=$ は更新を表している。この学習はすべての参照ベクトルに適用されるが、その変更の大きさは近傍関数 h_{ci} で決まる。 h_{ci} は

$$h_{ci} = \eta \exp \left\{ -\frac{|r_i - r_c|^2}{2\sigma^2} \right\} \quad (4)$$

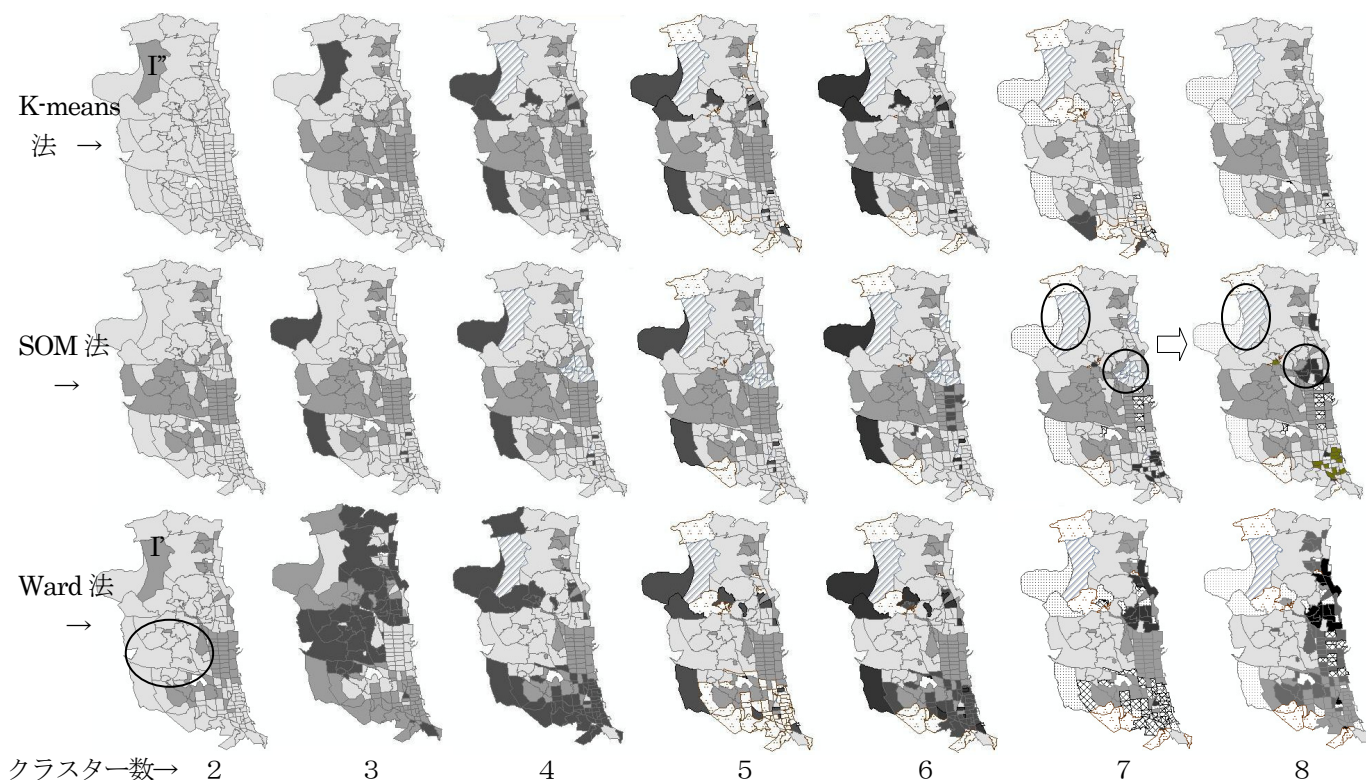
で与えられている。ここで r_i は i 番目のノードの2次元配列上での位置を表している。 η は学習の強さを表す1より小さな正の定数である。また、式(4)の中の絶対値はノード配列上のユークリッド距離である。 σ は近傍の広がりを決める正の定数である。

入力 x_n 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 x_n が入力されたときや x_n に類似する入力となされたときには、このときの勝者がまた勝者になって学習を起こ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。このように、 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 p 次元の入力ベクトルのパターンが、2次元の出力スペースに現れる。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。なお、個体のクラスター分けが必要な場合は出力スペースに現れる写像を直接観察することや、従来のクラスター分析などが応用される。

(3) 本研究の分析手法と諸設定

本研究は、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ー分析とWard法を用いた階層的クラスター分析の2つ、ニューラルネットワークの手法としてSOM法を用いることにする。これらの手法を比較



図ー2 各手法における、地域が2 クラスタから8 クラスタまで析出されていく過程

対象とするのは、いずれも代表的な手法で、お互いにアルゴリズムは異なるものの、共通な点もあるからである。

アルゴリズムについて、上でも述べたが、K-means法は与えられたクラスター数にベストな分けをするのに対して、Ward法では、クラスター数を逐次増加や減少の手法でクラスターの樹形図を得る。一方、SOM法は高次元のデータを低次元の出力空間に学習させながら写像し、出力空間で分けを行う。

クラスター分析においては、クラスター間の距離の定義、クラスター代表点の与え方やクラスターを集約することによって生じる情報量ロスの定義、逐次増加や減少などのクラスタリングのプロセスによって手法が多種多様である。しかし、K-means法とWard法は、クラスター間の距離の定義（ユークリッド距離）、クラスター代表点の与え方（重心）やクラスターを集約することによって生じる情報量ロスの定義（偏差平方和）などにおいて共通している。

一方、SOM法は出力スペースの写像をクラスタリングする際、従来の手法を応用することができるが、本研究で用いたソフトウェアViscovery SOMineにおいては、Ward法を改良した方法、SOM-Ward法と呼ばれるものが用いられている。アルゴリズムは未公開であるが、少なからずWard法との共通点もあると思われる。

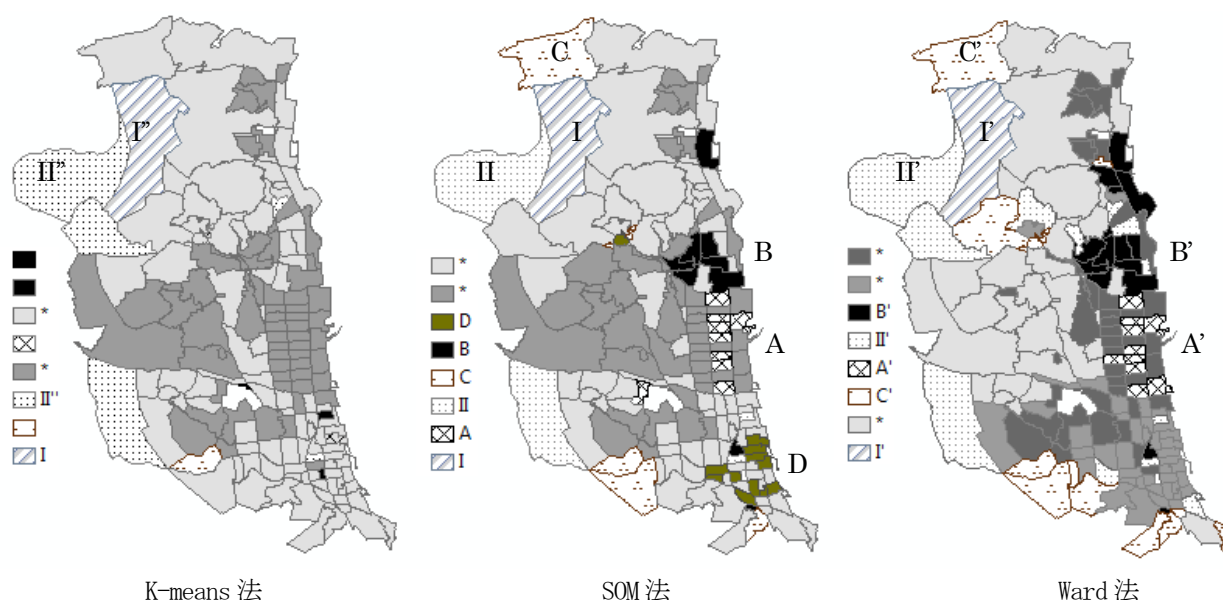
なお、クラスター分析はSPSS 13.1を用いて計算を行

った。K-means法においては、シード点をあらかじめ与える方法もあるが、3手法とも予備情報がない状態で比較したいので、シード点を与えていない。また、K-means法はデータの入力順序から影響を受けやすいとされているので、複数回において順序を変えて計算したが、異なる結果は見られなかった。Ward法においては、類似性（非類似性）について、内積や相関係数など複数の方法を選択することができるが、K-means法と同じくユークリッド距離で計算した。

一方、SOM法の計算においては、出力スペースの定義（ノードの数と配列など）、近傍関数（式(4)）のパラメータ η 、 σ の与え方と学習回数設定によって、出力スペースの写像がかなり違ってくことはよく知られているが、本研究で使ったツールはこれらの問題を解消するような独自の工夫をされているので¹²⁾、今回はデフォルト設定のまま計算を行った。

5. 社会地域分析における3手法の比較

図ー2は、K-means法、SOM法およびWard法の3つの手法において、地域が2 クラスタから8 クラスタまで析出されていく過程を示している。2 クラスタ（クラスター数が2）において、K-means法は後ほどの段階で他の手法によっても析出されるクラスターIだけを割



図ー3 クラスターの空間分布

り出した。一方、SOM法は縦に長い対象地域を大きく中部とその両端（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、中西部の地域（図中の左下の○で示した部分）でSOM法と異なる。SOM法とWard法の結果は階層的なので、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ず、クラスターの境界も毎回少し異なり、7クラスターのように、Ward法に類似しているケースもあるが、中西部の分け方についてはおおむねSOM法に類似している。

クラスターを細分化していく過程を見ると、SOM法においては、全地域を2クラスターに分けた後、7クラスターまで、薄く塗りつぶされた部分を5個のクラスター、灰色の部分で2個のクラスターに分けた。8クラスターにおいては、7クラスターまで生成したクラスターの一つをさらに細分化した（右の○で示された部分）。一方、Ward法では、3クラスターの段階で、これまでの2クラスターの中の一つをさらに二分化し全地域を三分化した後、それぞれ2つや3つの地域に分けて8クラスターまで至る。以下では、社会地域の数としても妥当で、両手法における樹形図レベルも同じである8クラスターに分けた場合の結果を用いて比較を行う。

（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クラスターにおいて（図ー3）、3つの手法は共通してクラスターIとクラスターIIを析出した。Iはひとつの面積の広い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法は最も多くの調査区を含んでいる。

このほか、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に類似する地域は析出されていない。一方、K-means法においては中部と両端の構造はSOM法に類似するが、いくつかの小さな調査区だけ析出された。

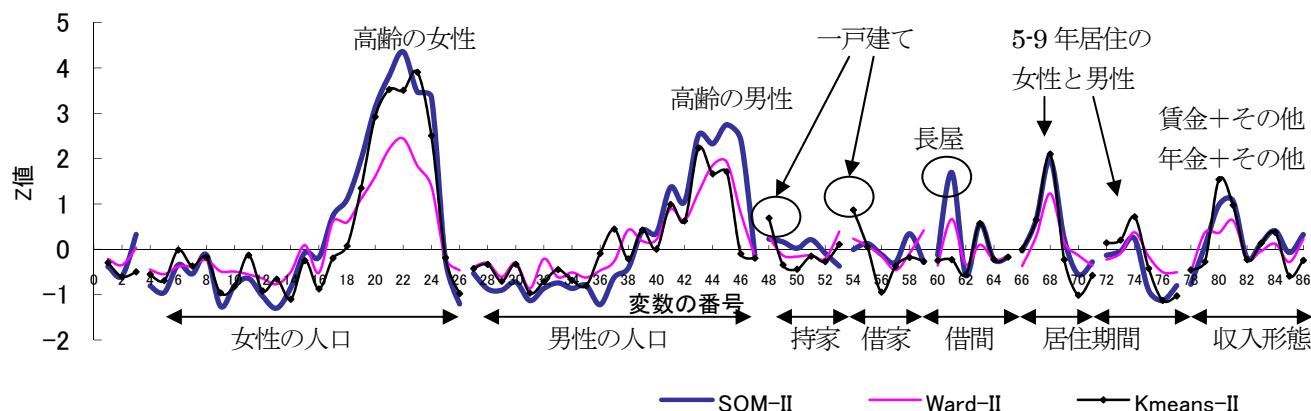
社会地域は複数の調査区、あるいは一定の面積を有する地域であるとする、社会地域となりうるクラスターの数（図ー3の凡例ではクラスター名か「*」が付いているクラスター）は、K-means法は4つだけであるのに対して、SOM法とWard法は共に8つになっている。

以上のように、3手法のうち、SOM法とWard法はより類似した結果が得られている。一方、K-means法では社会地域として考えられない面積の小さい個別の調査区が析出されやすいという特徴が見受けられる。

（2）共通クラスターの特徴の比較

ここでは、3手法で共通して析出されたクラスターIとII、およびSOM法とWard法で共通して析出されたクラスターA、B、Cについてデータの比較分析を行い、それぞれの手法の特徴をしてみる。

クラスターIはどの手法においても一つの調査区で構成されているが、析出される順序に注目する（図ー2）。K-means法では最初の区分け（2クラスター）の段階で析出されたのに対して、SOM法は8クラスターで析出された。Ward法もより早い段階（4クラスター）で析出さ



図ー4 各手法におけるクラスターIIの変数のZ値の比較

表ー1 SOM法とWard法の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

れた。このクラスターにあたる地域においては、大学および大学寮だけが立地しており、一般住民はほとんど居住していない。したがって、ここでも（1）で得た結果と同じく、K-means法はユニークな地域を見つけやすいと考えられる。

SOM法とWard法の共通クラスターについては、各クラスター内の平均 $\bar{x}^c$ とデータセット全体 X の平均 $\bar{x}$ を用いてクラスター間の比較を行う。表ー1では、SOMクラスターの $\bar{x}^c$ とWardクラスターのそれとの相関係数と分散を示している。なお、分散は、クラスター内平均 $\bar{x}^c$ の変量と全個体の $\bar{x}$ の変量との偏差平方和の平均なので、分散が大きければ、クラスター平均が全体平均との乖離が大きく、そのクラスターがより鮮明な特徴をもつことを示していると考えられる。

表ー1から分かるように、両手法は比較的大きな相関をもっているが、SOM法のほうが分散が大きい。すなわち、SOM法の方が地域のより鮮明な特徴を捉えていると考えられる。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していることから、SOM法の方がより特徴的で、地理的にまとまった地域を析出できたことが分かる。

K-means法はクラスターII'で他の手法と共通している

ので、このクラスターを見る。調査区はSOM法と同じく5つであるが、SOMとの相関係数は0.896、Wardとは0.892、いずれもSOM-Ward間のそれ（0.949）より低い。しかし、分散は1.138でSOM法に次ぐ値であるため、K-means法はこのクラスターに関して、他の手法と異なる特性を捉えている可能性が高い。これを見るために、この3つの手法における変数のZ値を図ー4で示す。この図から分かるように、3手法とも高齢者で、5-9年間住み、賃金や年金に加えて他の収入のある方が目立っている。K-means法の地域は主としてSOM法やWard法と住宅形態に関する変数（48-65）のZ値で異なる。SOM法とWard法では、借り間の長屋の住宅形態が目立ち、K-means法では一戸建ての持家と借家が多い。したがって、SOM法とWard法では、この地域の高齢化以外、多く分布している老人ホームや温泉療養地の特徴を捉えられているのに対して、K-means法は面積が広いので、この地域における住民全体の高齢化を捉えたと考えられる。

（3）クラスターの社会特性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を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。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ー2にまとめた。なお、Ward法とK-means法で析出されたクラスターの多くはSOM法のクラスターと類似する範囲を持っているので、この2手法のクラスターも社会的特徴をもつと考えられ、その分析は省略する。

表ー2から分かるように、SOM法では、まず地域全体を中部と北部・南部に二分化した。中部は若い年齢層、賃金収入、マンション借家、短い居住年数の特徴をもつことでその他の地域と異なる。この中部エリアはさらに6クラスター分類の際にさらに若いサラリーマンの核家

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

＊ 注：○は後に細分化されるクラスター。●は8クラスターまで残ったクラスター。

また、○や●の後の括弧の中の数字はクラスター数を表している

族のクラスターAを析出した。北部・南部のエリアからは、短い居住期間の高齢者地域（クラスターII）、高齢者農家の多い地域（クラスターC）、居住期間の長い温泉・旅館・商店街地域（クラスターD）、大学のみの地域（クラスターI）と市街地に立地する大学地域（クラスターB）を析出した。なお、クラスターの社会特徴の分析は基本的に図－4のように変数のZ値をプロットし、突出した特徴を捉える方法を用いるが、詳細については文献14を参照されたい。

6. 終わりに

本稿では、国勢調査のデータを用いて、伝統的なK-means法およびWard法と、比較的新しい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化マップ手法による社会地域分析を行い、この3つの手法の比較を行った。「一定の面積があり、大多数の地域と異なる社会的特徴をもつ地域（すなわち社会地域）を析出できる」という意味では、SOM法が最も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、より多くの調査区が析出されるため、特徴が明確でない傾向がある。また、SOM法には析出されたがWard法では析出できなかった地域もある。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、一定の面積を必要とする社会地域分析には適用しにくいと思われる。

なお、以上の結果は、一地域のデータセットで試し

た結果であり、分析においてパラメータを変えるなど感度分析的なことも行っていないので、一般的な結論としてはさらなる研究が必要である。また、従来の多変量解析を用いた社会地域分析においては、まず変数について主成分分析を行い、主成分の得点を用いてさらに個体のクラスター分析を行うという複合的な手法を取る場合が多いので、SOM法とこの複合手法との比較も有意義であると思われるが、これは今後の課題とする。

最後に、SOM法は伝統的な多変量解析手法とでは本質的に相違することを強調したい。多変量解析手法は、細かいオプションの違いによって解が異なる場合はあるが、原則的には同一のデータからは同一の分析結果・解釈が導かれる。一方、SOM法は同一のデータから同一の分析結果・解釈が導かれるとは限らない¹⁵⁾。本研究で用いたツールはこれらの問題を解決したとされているが、詳細なアルゴリズムやパラメータが公開されずにブラックボックスであるという問題も残る。しかし、SOM法は従来の手法で捨てられる情報も捉えることができるので、探索的な分析手法として非常に有用であると思われる。今後、SOM法で得られたクラスターをK-means法のシード点として与えることでより特徴の鮮明な社会地域を得られるかどうかについても検討してみたい。

謝辞

本研究は科研費19510165の助成を受けたものである。

参考文献

- 1) Park, R. E. and Burgess, E. W.: Introduction to the Science of Sociology, University of Chicago Press, Chicago, 1928.
- 2) Shevky, E. and Williams, M.: The Social Areas of Los Angeles. University of California Press, Los Angeles, 1949.
- 3) Bell, W.: The social areas of the San Francisco Bay Region, American Sociological Review Vol.18-1, pp.39-47, 1953.
- 4) Shevky, E. and Bell, W.: Social Area Analysis. Stanford University Press, Stanford, 1955.
- 5) 木内信蔵: 都市地理学原理, 古今書院, 1979.
- 6) 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, pp.73-78, 1980.
- 7) Economic Geography, Vol. 47, Supplement: Comparative Factorial Ecology, 1971.
- 8) 上田尚一: クラスター分析, 朝倉書店, 2003.
- 9) 佐藤義治: 多変量データの分類-判別分析・クラスター分析-, 朝倉書店, 2009.
- 10) 徳高平蔵, 藤村喜久郎, 山川烈: 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.
- 11) Kohonen, T.: Self-organised formation of topographically correct feature maps. Biological Cybernetics, Vol.43, pp.59-69, 1982.
- 12) Koual, E. L. and Kraakl, M. J.: Geovisualization to support the exploration of large health and demographic survey data, International Journal of Health Geographics Vol.3, p.12, 2004.
- 13) マインドウェア総研: データマイニング用SOM, <http://www.mindware-jp.com/basic/faq2.html>, 2009. 7. 14に閲覧。
- 14) 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, No. 5-2(CD-ROM), 2006.
- 15) 豊田秀樹: データマイニング入門, 東京図書, 2008.

空間分析におけるクラスター分析と自己組織化マップ手法の比較*

李 燕**

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が、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十分とは言えない。その原因の一つとして、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。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、比較的新しい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化マップ手法(SOM: Self-Organizing Map)の比較分析により、それぞれの手法の特徴および空間分析への有効性を考察した。

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

By Yan LI**

Understanding a city's or an area's socio-spatial attributes is of fundamental importance for many policy-making activities including city planning. However, the socio-spatial analysis of real cities has never been fully explored. The reason is multiple. One of them is that the uses of linear correlation methods often result in patterns incommensurate with the real world. This research compares the conventional cluster analysis methods with a non-linear non-parametric correlation method of Self Organizing Map (SOM) techniques in a case study of Beppu City, Oita, Japan. It demonstrates that the SOM method can deal with geographical distribution matters better than the cluster analysis methods.
