

# 非階層的クラスタリングによる東京大都市圏の考察<sup>1)</sup>

佐藤 仁志<sup>2)</sup>

## 1. はじめに

都市圏とは何らかの社会的・経済的な関係性を持つ一体的な地域である。東京をはじめとした巨大な都市では、単独の市区町村で経済・社会が完結することは不可能である。そのため、市区町村以外にも都市圏単位でのデータ整備の必要性は古くから訴えられているが、日本では公的な取り組みがない現状が続いている。

日本の大都市圏の多くが、都心の業務地と都市郊外の住宅地を通勤によって結びつける依存関係が最も強い。そのため、通勤構造に着目した都市圏分析も多い（例えば李・鈴木[2006]、高津・大林[2008]など）。研究者レベルで比較利用されている都市圏の定義の一つに、金本・徳岡[2002]の都市雇用圏がある。都市雇用圏は、山田・徳岡[1983]が提唱する標準大都市雇用圏の欠点の一つであった中心都市が単一である点を変更し、「都市圏は、中心都市と、それと社会的・経済的に密接な関係を有する周辺地域、すなわち郊外、によって形成されています。）」<sup>3)</sup>という考えにもとづいて定義されたものである。金本・徳岡らが提唱する都市雇用圏の考えにもとづいた研究には、既に周・大竹[2006]などをはじめとしたいくつかの成果がある。

以上のように、都市圏の実証研究は、数多くあり様々な方法が試みられている。既存の都市圏の定義は、その多くが市区町村を単位として都市圏の決定を行っている。しかし、近年では平成の大合併といわれるように町村部を中心とした合併や編入によって、行政区域の広域化が急速に進んでいる。合併等により出現した広大な行政範囲には、従来よりも様々な状況が同時に存在しうる可能性が拡大している。つまり、一つの市区町村の行政範囲の中に都心部への通勤者が多いために都心とのつながりが非常に強い地区と、逆に自市町村内や周辺の通勤者が多い地区が混在する可能性である。特に、都心からある程度の距離が離れた郊外ほど上記の可能性が高まる。この問題の解決方法の一つとしてデータの収集単位の細分化があげられる。市区町村単位より小さな統計データとして、日本では小地域統計の整備が進みつつある。国勢調査も1995年度から小地域によるデータの公開を開始しており、小地域統計を用いた研究も一般的になりつつある（例えば坂西[2007]・関根[2004]など）。

上記の問題に対応するために、佐藤[2009]では金本・徳岡[2002]の都市雇用圏を国勢調査の小地域統計を用いた形式に変更することを試みた。佐藤[2009]では、都市圏を設定する際の基準を金本・徳岡[2002]と同様に、中心都市への通勤率を10%としている。しかし、10%という

1) 本研究は「麗澤大学重点研究助成金」によって実施した研究成果である。記して感謝申し上げます。

2) 麗澤大学経済学部 hsato15@reitaku-u.ac.jp

3) 都市雇用圏 [http://www.urban.e.u-tokyo.ac.jp/UEA/uea\\_def.htm](http://www.urban.e.u-tokyo.ac.jp/UEA/uea_def.htm)

数値の基準については明確な基準があるわけではなく、過去の研究蓄積などを踏まえ経験的に設定された基準である。そこで、本研究では同一都市圏内では通勤に関する同質性を持つ部分に着目し、通勤構造に関して同質性のある地域を統計学の手法に基づいて抽出することで、都市圏の設定にできるだけ主観的・経験的な基準を排除することを試みる。具体的には、非階層的なクラスタリング手法である K-means 法または X-means 法を用いて通勤構造に関して同質性を持つ地域の特定を行い都市圏の設定を行う。

## 2. 使用データについて

都市圏の導出には、2005年度の国勢調査の常住地別通勤データの小地域統計を使用する。今回は埼玉県・千葉県・東京都・神奈川県のみを対象として分析を行うこととする<sup>4)</sup>。

## 3. 都市圏の設定方法について

本研究では、金本・徳岡[2002]が提唱する都市雇用圏の設定方法を小地域統計に適用した佐藤[2009]の順序に変更を加えた下記の順序で都市圏の設定を行う。

- (1) 中心都市の選出は、金本・徳岡[2002]と同様の方法でおこなう（市区町村単位の通勤データを利用している）。中心市区町村への通勤率の基準値は10%で設定している。
  - (ア) 中心都市の算出は、金本・徳岡[2002]と同様に市区町村帯の通勤データを用いて複数ラウンドの計算を行うこととする。
  - (イ) 中心都市は「さいたま市、戸田市、千葉市、東京特別区、立川市、武蔵野市、横浜市、川崎市、厚木市」の9市区となった。
- (2) 中心都市の選定終了後、国勢調査の常住地による従業市区町村の結果を用いて、下記の項目に関する比率を市区町村単位で計算する。
  - (ア) 他県で従業する人数のうち中心都市へ通勤する者の比率
  - (イ) 県内他市区町村で従業する人数のうち中心都市へ通勤する者の比率
- (3) (2)で求めた比率を用いて小地域統計単位で中心都市への通勤率を算出する。
- (4) 中心都市への通勤率を使用して都市圏を算出する。中心都市への通勤率の同質性に着目し、都市圏の算出を行う。つまり、中心都市への通勤率に関する同質性を持つ地域をクラスタリングすることで都市圏の設定を行う。

## 4. K-means 法

K-means 法は、クラスタリングの一種である。クラスタリングには、樹形図（dendrogram）に基づいた分類を行う階層的手法と K-means 法などの非階層的手法がある。階層的手法は、様々なタイプのデータを用いることができ、そのデータ間の類似性の構造を樹形図として表すことが出来るため、結合関係を細かく議論できる点が長所である。しかし、本研究では

---

4) 金本・徳岡[2002]でも示されているように東京特別区を中心とする大都市雇用圏は、1都3県以外にも茨城県・栃木県の一部にも存在している。今回は用意したデータの制約上、やむを得ず1都3県で分析を行うこととした。小地域統計については、佐藤[2009]や(財)統計情報研究開発センターなどを参照。  
(財)統計情報研究開発センター：<http://www.sinfonica.or.jp/>

表1 金本・徳岡 [2002] の都市雇用圏の設定基準

都市圏設定基準	都市雇用圏 (Urban Employment Area)
区 分	大都市雇用圏 (Metropolitan Employment Area) : 中心市町村の DID 人口が 5 万以上 小都市雇用圏 (Micropolitan Employment Area) : 中心市町村の DID 人口が 1 万以上 5 万未満
都 市 圏	DID 人口が 1 万以上の市町村を含む。
中心都市	以下の条件のいずれかを満たす市町村を中心都市とする。複数存在する場合には、それらの集合を中心とする。 (1) DID 人口が 1 万以上の市町村で、他都市の郊外でない。 (2) 郊外市町村の条件を満たすが、(a)従業常住人口比が 1 以上で、(b)DID 人口が中心市町村の 3 分の 1 以上か、あるいは 10 万以上である。
郊 外	中心都市への通勤率が (a) 10% 以上のものを (1 次) 郊外市町村とし、 (b) 郊外市町村への通勤率が 10% を超え、しかも通勤率がそれ以上の他の市町村が存在しない場合には、その市町村を 2 次以下の郊外市町村とする。 ただし、 (1) 相互に通勤率が 10% 以上である市町村ペアの場合には、通勤率が大きい方を小さい方の郊外とする。 (2) 中心都市が複数の市町村から構成される場合には、それらの市町村全体への通勤率が 10% 以上の市町村を郊外とする。 (3) 通勤率が 10% を超える中心都市が 2 つ以上存在する場合には、通勤率が最大の中心都市の郊外とする。 (4) 中心都市及び郊外市町村への通勤率がそれぞれ 10% を超える場合には、最大の通勤率のものの郊外とする。

出典：金本・徳岡 [2002] 「日本の都市圏設定基準」

対象地域内のデータ数は 20000 件を超えるため、樹形図を用いたデータ分類は非常に困難である。

そこで、多数のデータの分類 (クラスタリング) に適した K-means 法を用いて分類を行うこととする。K-means 法は、あらかじめクラスタ数を設定している点が階層的手法と大きく異なる点である。K-means 法によるクラスタリングの手順は、下記の通りである。

- (1) それぞれのデータに対してランダムにクラスタに所属させる。
- (2) (1) で設定したクラスタごとに重心を求める。
- (3) それぞれのデータに対して、(2) で求めた各クラスタの重心<sup>5)</sup>までの距離を計算<sup>6)</sup>する。
- (4) 各データを重心までの距離が最も近いクラスタへ所属を変更する。
- (5) (4) で割り当てたクラスタの結果に基づいて(2)以降の手順を計算した結果に変化がなければクラスタリングを終了する。変化がある場合には、(2)以降の手順を繰り返す。

金本・徳岡 [2002] や佐藤 [2009] の都市圏の形状を踏まえ、本研究では K-means 法によるクラスタ数は都市圏内と都市圏外の 2 地域と考えてクラスタリングを行うこととする<sup>7)</sup>。

## 5. K-means 法による都市圏の抽出

前述(3)および(4)の手順に基づいて、「さいたま市、戸田市、千葉市、東京特別区、立川市、

5) K-means 法の重心は各クラスタに含まれるデータの平均値に等しい。

6) 距離の計算は、ユークリッド距離で計算している。例えば、2次元空間の 2 点  $x(x_1, x_2), y(y_1, y_2)$  のユークリッド距離  $d$  は  $d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$  で求めることができる。

7) K-means 法および後述の X-means 法によるクラスタリングは、データマイニングのオープンソースとして開発されている「Weka(ver.3.6.3)」を用いた。

Weka ホームページ <http://www.cs.waikato.ac.nz/~ml/index.html>

図1 K-means法による都市圏

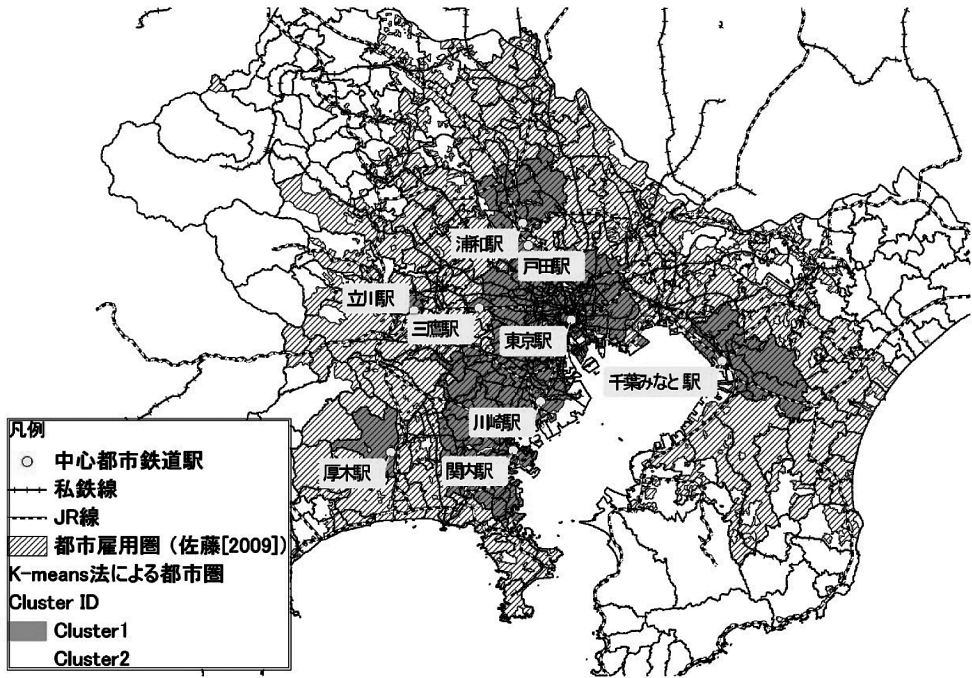


表2 K-means法による中心都市への通勤率に関する平均値（上段）と標準偏差（下段）

	都区部	さいたま市	戸田市	千葉市	立川市	武蔵野市	横浜市	川崎市	厚木市
全体	41.73 34.42	2.77 9.81	0.27 2.47	2.73 9.86	0.52 2.93	0.29 1.74	6.28 16.95	2.04 8.42	0.71 4.83
クラスター ID1	89.98 8.31	5.73 16.80	0.66 4.45	4.78 16.62	0.60 4.79	0.44 2.98	16.61 27.46	5.53 14.65	1.22 8.26
クラスター ID2	20.82 14.90	1.49 3.20	0.10 0.33	1.85 4.13	0.48 1.54	0.23 0.70	1.81 4.36	0.52 0.97	0.48 1.91

武蔵野市、横浜市、川崎市、厚木市」の9市区が中心都市として抽出された。そこで、各地区から中心地区への通勤率を算出し、このデータに用いてK-means法を行った結果が図1である。図1を見ると、クラスターのIDが1の地域が東京の大都市圏と考えることができるが、この地域は中心都市としてあげた「さいたま市、戸田市、千葉市、東京特別区、立川市、武蔵野市、横浜市、川崎市、厚木市」の9市区の面積にきわめて近く、各種の都市圏研究の結果や経験や実感と大きくかけ離れている。また、比較対象として提示している佐藤[2009]の大都市圏と比べてもかなり狭い範囲のみを大都市圏としてクラスタリングを行っている。K-means法は対象となるデータを必ずいずれかのクラスターに所属させる方法であるため、クラスター数を2と設定した場合には中心都市との関係がきわめて強い地域のみが都市圏として抽出されていると思われる。これは、表2で示される各クラスターの平均値の違いにも表れている。つまり、中

心都市への通勤率のデータのみを用いて都市圏とそれ以外の地域のようにクラスタ数を2として分割するのでは、クラスタ数の設定が不十分である可能性が示唆されている。そこで、対象となる地域のクラスタ数をより適切な数に設定することで都市圏設定の最適化を試みる。

## 6. X-means 法による都市圏の抽出

Pelleg, Moore[2000]によって提案された X-means 法は、K-means 法を拡張したものである。K-means 法では、何らかの仮定に基づいてクラスタ数の設定を行い、事前に設定したクラスタ数に合致するクラスタリングを行うものであった。しかし、X-means法ではクラスタ数をある一定の範囲で与え、その名から後述する情報量基準に従って最も適切なクラスタ数を探することができるため、K-means 法よりもクラスタリングの精度が上がると考えられている方法である。

X-means 法におけるクラスタリングでは、再帰的に K-means 法を行い、クラスタ分割の前後でベイズ情報基準量 (Bayesian information criterion, BIC) を計算する。そして、クラスタ分割の前後の BIC の値を比較しクラスタ分割が妥当と判断されるまでクラスタ分割を続けることによって適切なクラスタ数を求めようとする考え方である。X-means 法によるクラスタリングの手順は下記の通りである。

- (1) 与えられたデータに対して  $k=2$  で K-means法によるクラスタリングを行い、作成されたクラスタを  $C_1, C_2$  とする。
- (2)  $i=1, 2$  に対して、次の手順(3)~(8)を繰り返す。
- (3) クラスタ  $C_i$  に対して K-means 法を用いて 2 つのクラスタ  $C_i^1, C_i^2$  に分割する。
- (4) クラスタ  $C_i$  に含まれるデータに対して  $p$  変量正規分布を仮定し、そのときの BIC を計算し、その値を  $B$  とする。
- (5)  $C_i^1, C_i^2$  に分割したデータに対しても、同様に  $p$  変量正規分布を仮定し BIC を計算し、その結果を  $B'$  とする。
- (6)  $B > B'$  ならば二分分割したモデルを好ましいと判断し、 $C_i \leftarrow C_i^1$  とする。 $C_i^2$  はスタックに積む。手順(3)へ。
- (7)  $B \leq B'$  ならば二分分割しないモデルを好ましいと判断し、 $C_i^1$  に対する二分分割を停止する。スタックからデータを取り出し手順(3)へ。スタックが空の場合には手順(8)へ。
- (8)  $C_i$  における二分分割が終了。
- (9) クラスタリング終了。

上記の方法で、計算した結果が図 2 である。X-means 法によって対象地域は、122 のクラスタに分割された。クラスタ ID は、下記に記すように基本的には地域ごとに集積する傾向が見られる。

- クラスタ ID が 1 ~ 50 : 主に中心都市の 9 市区周辺地域に集積
- クラスタ ID が 51 ~ 60 : 主に分析対象地域の最郊外に集積
- クラスタ ID が 61 ~ 80 : 主に横浜市に集積
- クラスタ ID が 81 ~ 90 : 主にさいたま市に集積
- クラスタ ID が 91 ~ 100 : 主にさいたま市に集積
- クラスタ ID が 101 ~ 105 : 主に川崎市に集積
- クラスタ ID が 106 ~ 115 : 主に厚木に集積

図2 X-means 法による都市圏

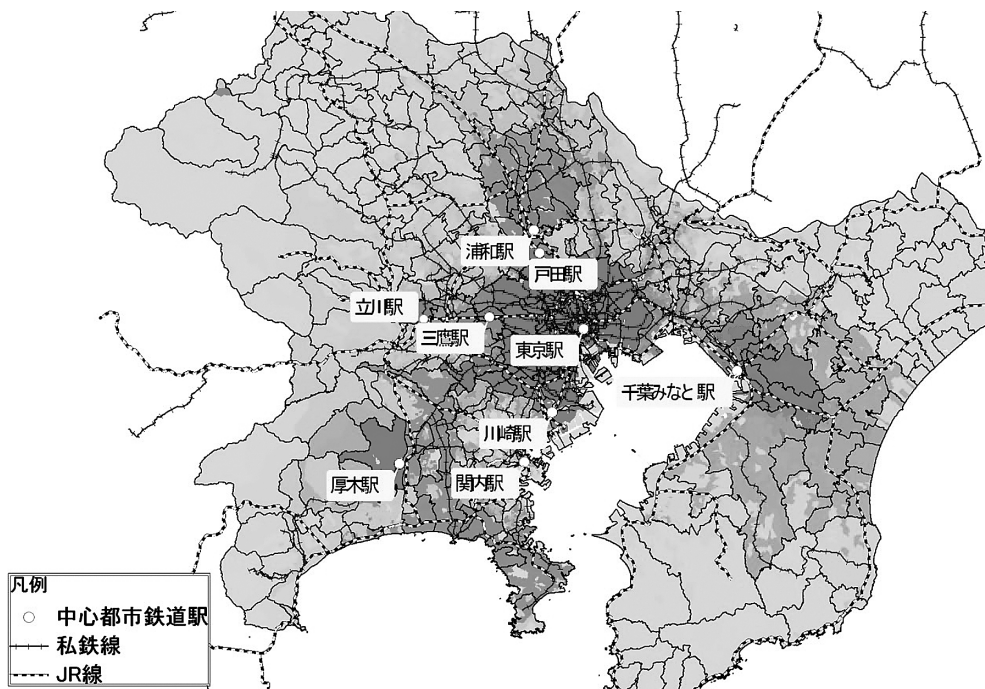


表3 Global Moran's I を用いた距離帯別のクラスターの地理的集積度の比較

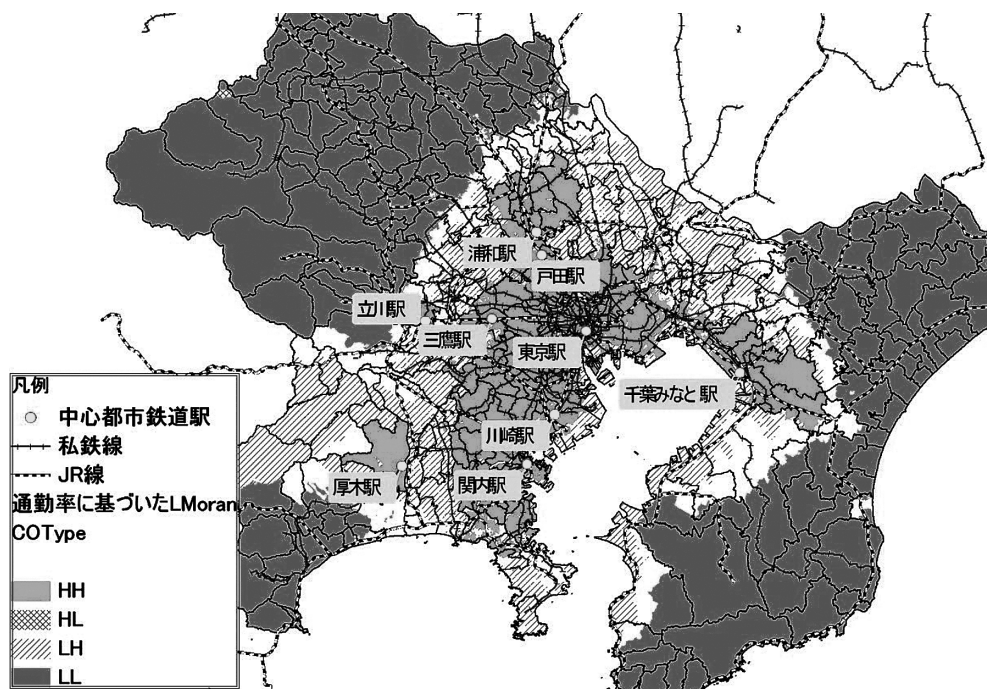
	都区部内	都区部から0~10km	都区部から10~20km
Global Moran's I	0.17	0.51	0.66
E(I)	$-3.0 \times 10^{-4}$	$-2.0 \times 10^{-4}$	$-2.0 \times 10^{-4}$
Var(I)	$1.6 \times 10^{-5}$	$5.0 \times 10^{-6}$	$4.8 \times 10^{-5}$
Z score	44.02 **	232.04 **	94.97 **
	都区部から20~30km	都区部から30~40km	
Global Moran's I	0.83	0.70	
E(I)	$-2.1 \times 10^{-4}$	$-3.9 \times 10^{-4}$	
Var(I)	$4.3 \times 10^{-5}$	$2.5 \times 10^{-5}$	
Z score	125.80 **	140.10 **	

\*\* 1%の有意水準で帰無仮説を棄却する。

● クラスタ ID が116~122：主に都区部に集積

図2では、 $k=2$ で行ったK-means法の図1に比べてかなり細分化が進んでいる。特に東京都区部から横浜にかけての地域や、中心都市周辺の地域のクラスターが細分化されている。また、図2では判別しにくい部分もあるが、地域によってクラスターの空間的な集積状況に違いが生じている。そこで、東京都区部の外縁部から10kmごとの距離帯別にGlobal Moran's Iを算出しクラスターIDの空間的な集積度の違いを確認した。表3をみるとGlobal Moran's Iは全て正の値で有意であるため、クラスターIDの空間的な集積(偏り)が全ての距離帯で生じていることが示されている。また、都区部に近い場所よりも郊外の方がGlobal Moran's Iの値が大きいため、

図3 中心都市への通勤率の分布 (Anselin's Local Moran's I)



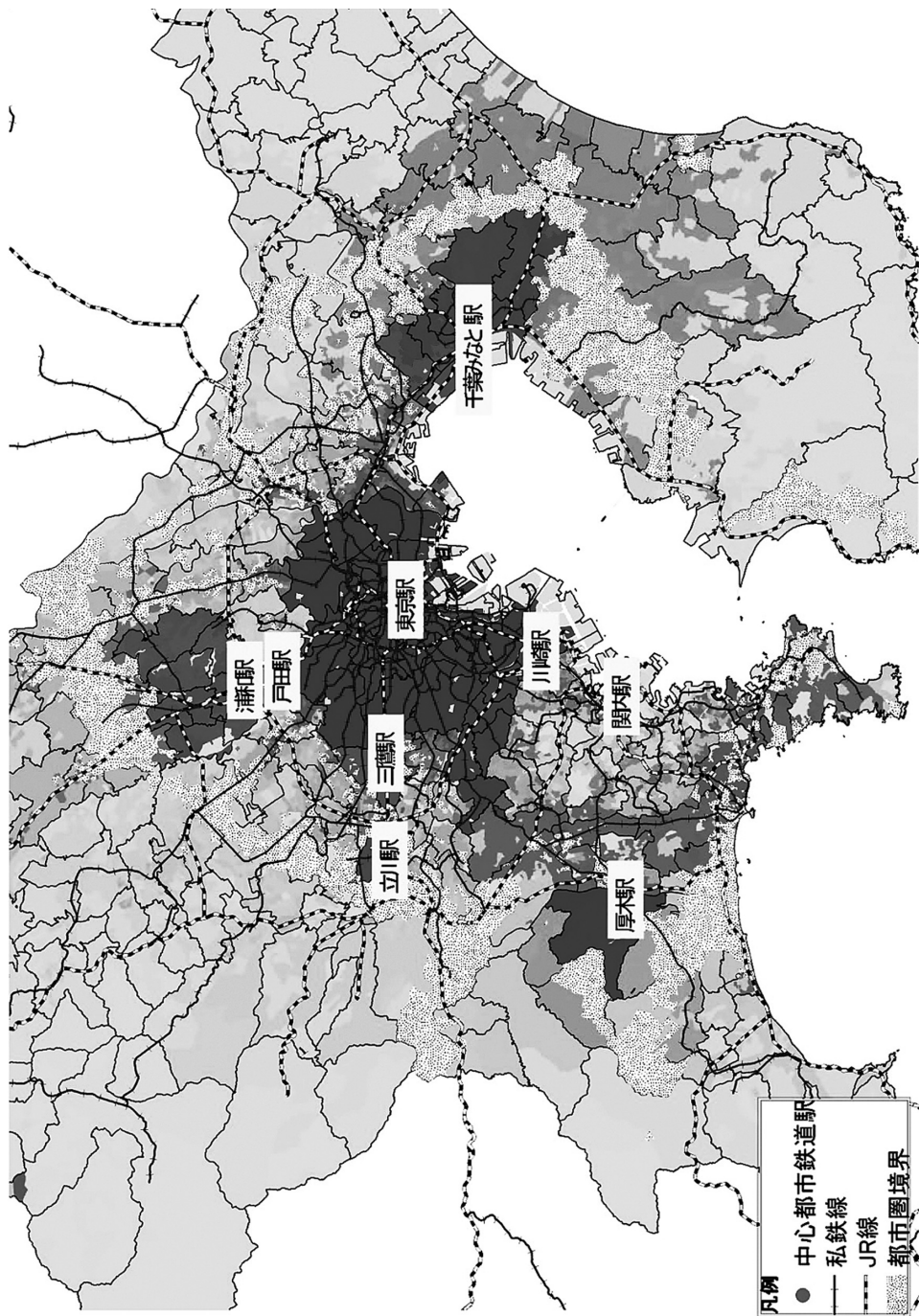
郊外ほどクラスタIDの空間的な集積（偏り）が生じている。これは、郊外ほど鉄道網が疎らになるため、鉄道駅周辺とその他地域でクラスタが異なるためであると思われる。以上のことから、クラスタIDが1~50の地区あたりで都市圏境界を設定することが妥当であると思われる。しかし、図2で示すX-means法を用いたクラスタリングでは、対象地域が細分化されすぎたため境界の設定が困難になっている。

## 7. 通勤率の地理的集積度からみた都市圏境界

前述のX-means法によるクラスタリングの結果だけでは、クラスタ数が細分化されすぎたため都市圏の境界を定めることが難しい。都市圏は様々な項目に関して同質性を持つため、今回の研究で使用している中心都市への通勤率に関しても、同一の都市圏に属しているならば都市圏内ではある程度同じ傾向を持つはずである。従って、都市圏の境界となるような地域では周辺と比較して異なった性質を持つはずである。そこで、Anselin's Local Moran's Iによって周辺地域と比較した特異性を持つ地域を明らかにすることを試みる。1つのデータに基づいてAnselin's Local Moran's Iが計算されるため、これまで9市区の中心都市への通勤率を別々に扱ってきたが、ここではそれらをまとめて中心都市への通勤率として扱うことにする。図3は中心都市への通勤率を基準としたAnselin's Local Moran's Iを求め、ローカルな空間的自己相関が有意な地区をタイプ別に地図上に表示した。具体的には、Anselin's Local Moran's Iが5%の有意水準で統計的に有意な地域を下記の4つのタイプに分けて塗り分けを行っている。

- (1) 周辺を含め中心都市への通勤率が高い地区 (HH)

图4 都市圏境界





- (2) 周辺を含め中心都市への通勤率が低い地区 (LL)
- (3) 周辺よりも通勤率が高い地区 (HL)<sup>8)</sup>
- (4) 周辺よりも通勤率が低い地区 (LH)

図3をみると、都心から順にHH→LH→LLと変化している。LLで示される地区は当該地区も周辺地区も中心都市への通勤率が低いことを示しているため、この地区は明らかに都市圏外であろう。同様の判断基準を適用すると、HH地区は都市圏内になる。また、LH地区に隣接してHH地区が存在しているため、LH地区は隣接するHH地区の影響を受けていることが考えられる。

そして、LHとLLで示される地区の間にAnselin's Local Moran's Iの値が統計的に有意ではなく、4タイプのいずれにも評価できない地区が広がっている。この地域は、図2でクラスタIDが1~50となる「主に中心都市の9市区周辺地域に集積」の外縁部となるような地域にほぼ一致している。以上のことから、LHとLLで示される地区の間に広がるAnselin's Local Moran's Iの値が統計的に有意でない空間が、都市圏内から都市圏外への遷移地帯の役割を果たしていると判断し、この地域を都市圏境界と考える。

都市圏境界と判断した地域とX-means法によるクラスタリングの結果を重ね合わせた地図が、図4である。図4をみると、埼玉・東京・神奈川では都市圏境界と判断した空間が、同心円上に明確に連続して存在している。しかし、千葉県の北西部から埼玉県との県境にかけての地域では、都市圏境界が非連続な場合や境界が二重になる場合があり、他地域と比べると境界が明確になっていない。また、これらの地域ではJR常磐線沿線や東武伊勢崎線のような都心と郊外を結ぶ放射状の鉄道の沿線地域の一部にも都市境界が明瞭に現れている点も他地域と異なる点である。

## 8. 結 論

本研究では、東京大都市圏を事例として小地域統計を用い、各地区からの都市圏の中心への通勤率に基づいた都市圏の設定を試みた。その結果を下記に記す。

- 従来の都市圏の定義では、都市圏の設定を行う者の経験などによって判断されることが多かった判断基準に対して、本研究では同質性を持つ地域を統計的指標に基づいて抽出し、それによって都市圏の境界を定めた。
- 「さいたま市、戸田市、千葉市、東京特別区、立川市、武蔵野市、横浜市、川崎市、厚木市」の9市区を中心都市と定義し、小地域統計の各地区からそれぞれの中心都市への通勤率をデータとしてクラスタ数を2でK-means法を実施した。しかし、この方法では中心都市とその他の都市でクラスタ分割されただけの状態にほぼ等しい状態になった。
- クラスタの分割数を最適化することで、上記の問題を解決し適切な都市圏の設定を行うことを考えた。そこで、BICにより望ましいクラスタ数を設定することができるX-means法を用いてクラスタリングを行った。その結果、対象地域は122のクラスタに分割され、分割されたクラスタは空間的に集積し偏りのある結果となった。また、中心都市への通勤率に基づいたクラスタリングの結果には、全ての地域において強い正の空間的自己相関があった。特に、郊外地域では都心に近い地域に比べて、中心都市への通勤を行う地区とそ

8) HLとなった地区は秩父市の一地区のみである。これは、当該地区の人口が極端に少ないことによって生じた結果であり、外れ値と考えることが妥当であろう。従って、HLになる地区は実質的には存在していないと思われる。

れ以外の場所へ通勤を行う地区が空間的に分離していることが、Global Moran's I の値から明らかにされた。

- 小地域統計の各地区から中心都市への通勤率に対して Anselin's Local Moran's I を算出し、中心都市に対する通勤率によってローカルな空間的自己相関を明らかにした。その結果、大半の地区において周辺地区の通勤率との空間的自己相関を持っていることが明らかになった。しかし、郊外ではローカルな空間的自己相関が統計的に有意ではない地域が同心円状に存在していることが明らかになった。このような地域では、中心都市への通勤の比重が高い地域と低い地域が混在している。そこで、この地域を都市圏境界と判断した。都市圏境界は帯状の空間として存在し、またその幅も一様ではない。特に、千葉県北西部から埼玉県との県境に渡る地域では、都市境界が明瞭に表れなかった。これらの地域の都市境界が他地域と異なった理由は、都心への鉄道網が他に比べて少ないことや、住宅開発の時期が遅かったことが影響していると思われる。

(麗澤大学准教授)

#### 付論1 X-means 法における BIC の計算について

ベイズ情報基準量 (Bayesian information criterion, BIC) は下記の式で計算する。クラスター  $C_i$  に含まれるデータ  $x_i$  を  $p$  変量正規分布と仮定し式 (A-1) で表す。

$$f(\theta_i; x) = (2\pi)^{-\frac{p}{2}} |V_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)' V_i^{-1} (x - \mu_i) \right\} \quad (\text{A-1})$$

このとき BIC は下記の式 (A-2) で計算する。

$$BIC = -2 \log L(\hat{\theta}_i; x) + q \log n_i \quad (\text{A-2})$$

- $p$  : データ次元数
- $\hat{\theta}_i = [\hat{\mu}_i, \hat{V}_i]$  :  $p$  変量正規分布の最尤推定値
- $\mu_i$  :  $p$  次の平均値ベクトル
- $V_i$  :  $p \times p$  の分散・共分散行列
- $q$  : パラメータ  $\hat{\theta}_i$  の次元数であり。
- $n_i$  : クラスター  $C_i$  に含まれるデータ数
- $L(\cdot)$  : 尤度関数を表しており  $L(\cdot) = \prod f(\cdot)$

#### 付論2 Global Moran's I について

Global Moran's I は、空間的自己相関を示す最も代表的な指標である (Moran(1948, 1950))。Global Moran's I は下記の式 (A-3) で計算することができる。

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{X})(x_j - \bar{X})}{W \sum_{i=1}^n (x_i - \bar{X})^2} \quad (\text{A-3})$$

$n$  はデータ総数、 $w_{ij}$  は地点  $i$  と地点  $j$  の結合関係を表すウェイト値、 $x_i$  は地点  $i$  の観測値、 $\bar{X}$  は観測値  $x$  の平均値である。ウェイト値  $w_{ij}$  は地点  $i$  と地点  $j$  が接するときは 1、接しないときは 0 の値をとる。また、 $W$  はウェイト値の合計であり、下記の式 (A-4) で計算される。

$$W = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \quad (\text{ただし } i \neq j) \quad (\text{A-4})$$

Global Moran's I の検定は、式 (A-5) を用いて検定を行うことができる。

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}} \quad (\text{A-5})$$

ただし、 $E[I]$  は  $I$  の期待値、 $V[I]$  は  $I$  の分散を示し、それぞれ下記の式 (A-6)、(A-7) で定義される。

$$E[I] = \frac{-1}{n-1} \quad (\text{A-6})$$

$$V[I] = E[I^2] - E[I]^2 \quad (\text{A-7})$$

そして、Global Moran's I の値は 1 に近ければ集積度合いが高く、-1 に近ければ空間的に分散している。また、 $z_I$  スコアに基づいて算出された  $p$  値によって、Global Moran's I の値に対する統計的な有意性の検定を行うことができる。

### 付論 3 Anselin's Local Moran's I について

Anselin's Local Moran's I は下記の式 (A-8) で計算することができる (Anselin(1995))。

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{ij} (x_j - \bar{X}) \quad (\text{A-8})$$

$n$  はデータ総数、 $w_{ij}$  は地点  $i$  と地点  $j$  の結合関係を表すウェイト値、 $x_i$  は地点  $i$  の観測値、 $\bar{X}$  は観測値  $x$  の平均値である。ウェイト値  $w_{ij}$  は地点  $i$  と地点  $j$  が接するときは 1、接しないときは 0 の値をとる。

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n w_{ij}}{n-1} - \bar{X}^2 \quad (\text{A-9})$$

Anselin's Local Moran's I の検定は、式 (A-10) を用いて検定を行うことができる。

$$z_{II} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \quad (\text{A-10})$$

ただし、 $E[I_i]$  は  $I_i$  の期待値、 $V[I_i]$  は  $I_i$  の分散を示し、それぞれ下記の式 (A-11)、(A-12) で定義される。

$$E[I_i] = -\frac{\sum_{j=1, j \neq i}^n w_{ij}}{n-1} \quad (\text{A-11})$$

$$V[I_i] = E[I_i^2] - E[I_i]^2 \quad (\text{A-12})$$

そして、Anselin's Local Moran's I の値が正の場合は対象地区の値は近隣地区と近い値を持つことを示し、逆に Anselin's Local Moran's I の値が負の場合は対象地区の値は近隣地区と異なる値を持つことを示している。また、 $z_{II}$  スコアに基づいて算出された  $p$  値によって、Anselin's Local Moran's I の値に対する統計的な有意性の検定を行うことができる。

### 参考文献

- AGUILERA, A., "Growth in commuting distances in French polycentric metropolitan areas : Paris Lyon and Marseille", *Urban Studies*, 42, pp.1537-1547, 2005.
- Anselin, L., "Local indicators of spatial association-LISA", *Geographical Analysis*, 27, pp.93-115, 1995.
- Fotheringham, A.S., Rogerson, P.A., "The SAGE handbook of Spatial Analysis", Sage, 2008.
- MCMILLEN, D. P., "The number of subcenters in large urban areas", *Journal of Urban Economics*, 53, pp.321-

- 338, 2003.
- Moran, P.A., "The interpretation of statistical maps", *Biometrika*, 35, pp.255-260, 1948.
- Moran, P.A., "A test of for the serial dependence of residuals", *Biometrika*, 37, pp.178-181, 1950.
- PELLEG, D., Moore, A., "Accelerating Exact k-means Algorithm with Geometric Reasoning", *KDD-99*, pp.277-281, 1999.
- PELLEG, D., Moore, A., "X-means: Extending K-means with Efficient Estimation of the Number of Clusters", *ICML-2000*, pp.727-734, 2000.
- WITTEN, I. H., Frank, E., "DATA MINING: Practical Machine Learning Tools and Techniques, Second Edition", Morgan Kaufmann, 2005.
- 石岡恒憲「クラスター数を自動決定する k-means アルゴリズムの拡張について」, 『応用統計学』, 応用統計学会, 2000, 29巻3号, 141-149.
- 金本良嗣・徳岡一幸「日本の都市圏設定基準」, 『応用地域学研究』, 応用地域学会, 2002, 7巻, 1-15.
- 小松拓磨・中川雅之「マイクロデータを用いた居住地選択モデルによる郊外居住の実証分析」, 『都市住宅学』都市住宅学会, 2007, 57巻, 52-60.
- 坂西明子「東京大都市圏の産業別従業者と人口の郊外化—GISを用いた小地域分析（フォーラム論文 特集 都市経済の視角）」, 『都市研究』, 近畿都市学会, 2006, 5巻, 29-37.
- 佐藤仁志「小地域統計を用いた都市雇用圏の設定基準の拡張」, 『麗澤経済研究』, 麗澤大学経済学会, 2009, 17巻1号, 19-28.
- 周燕飛・大竹文雄「都市雇用圏からみた失業率の地域的構造」, 『応用地域学研究』, 応用地域学会, 2006, 11巻, 1-12.
- 鈴木勉「東京大都市圏における職住割当の最適化に関する実証的研究」, 『都市計画 別冊都市計画論文集』, 日本都市計画学会, 1992, 337-342.
- 関根智子「国勢調査小地域統計の空間分析（特集 GISによる統計の利用）」, 『統計』, 日本統計協会, 2004, 55巻8号, 20-26.
- 高津定弘・大林千一「都市・雇用データによる都市機能指標と圏域設定の基礎整備」[JILPT 資料シリーズ], 労働政策研究・研修機構, 2008, No.44.
- 谷口守・阿部宏史・松中亮治他「DVD 電話帳を用いた認識に基づく地域範囲に関する研究—東京都市圏を対象とした検討から」, 『都市計画論文集』, 日本都市計画学会, 2004, 39巻, 56-61.
- 元田浩・津本周作・山口高平・沼尾正行「データマイニングの基礎」, オーム社, 2006.
- 李召熙・鈴木勉「1965-2000年における大都市圏の通勤距離の変化動向及びその要因に関する研究」, 『都市計画 別冊 都市計画論文集』, 日本都市計画学会, 2006, 217-222.
- 山田浩之・徳岡一幸「わが国における標準大都市雇用圏：定義と適用—戦後の日本における大都市圏の分析(2) —」, 『経済論叢』, 京都大学, 1983, 132巻3・4号, 145-173.

## Summary

### Analysis of Tokyo metropolitan area by non-hierarchical clustering

Hitoshi Sato

The purpose of this study is to identify the structure of the Tokyo metropolitan area by using small area statistics. A Metropolitan area is the sphere that has a strong relationship between downtown and suburbs. Also, there is some kindred area within the same Metropolitan area. We tried to clear the structure of the Tokyo metropolitan area by non-hierarchical clustering. Specifically, we tried to classify the Tokyo metropolitan area according to the K-means method and the X-means method by using the data of the commuting rate.

As a result, we obtained the following results:

1. We were not able to set an appropriate metropolitan area by the k-means method when the number of clusters was assumed to be two. However, because the classified kind of number was a little, we were not able to do an appropriate classification by this method.

2. In the x-means method, we classified the metropolitan area into 122 kinds of clusters. We combined the results of this result and Anselin 's Local Moran' s I and clarified the structure of the Tokyo metropolitan area.

（受付 平成22年11月18日）  
（校了 平成23年1月31日）

