

# 厳格な幾何形状の制約を置かない 産業集積の検出

氏家 晃仁<sup>1</sup>・福本 潤也<sup>2</sup>

<sup>1</sup>学生会員 東北大学大学院情報科学研究科 博士後期課程 (〒980-8579仙台市青葉区荒巻字青葉6-6-06)

E-mail:khujiie@plan.civil.tohoku.ac.jp

<sup>2</sup>正会員 東北大学大学院准教授 情報科学研究科 (〒980-8579仙台市青葉区荒巻字青葉6-6-06)

E-mail:fukumoto@plan.civil.tohoku.ac.jp

産業集積は多数の企業が空間的に集中して立地する現象である。産業集積の実態把握を行い、各産業の立地特性等を明らかにすることは、産業集積に関する理論研究で得られた知見を実証的に確認したり、産業集積に着目した政策を議論する上で有益である。

本研究では、集積領域を構成する地理的単位の連結性の仮定を緩めたモデルベースクラスタリングに基づく集積領域検出手法を用いて、首都圏から産業集積を検出する。検出された産業集積の空間的分布と交通施設や用途地域等の都市計画上の規制といった立地環境との関係に着目することで、産業の立地特性の把握を試みる。

**Key Words :** *industrial agglomeration, cluster detection, shape of cluster, spatial pattern*

## 1. はじめに

産業集積は多数の企業が空間的に集中して立地する現象である。産業集積は経済活動の空間的分布に関して観察される最も顕著な特徴の1つであり、世界中で観察される。企業が集中して立地することで生じる正の外部性は集積の経済と呼ばれる。産業集積は集積の経済によって形成・維持され、立地企業の競争力を高めると一般に考えられている。このため、産業集積に対する政策的関心は高い。地域活性化や国際競争力強化を目的として、産業集積の形成を促す多くの政策が国や地方自治体によって推進されている。

産業集積の実態について実証的知見を蓄積させることは実務・学術の両面から意義がある。産業毎の集積の程度や立地特性を把握することで、国や自治体が産業集積の形成を促進する上で有益な知見を提供できる。学術的にも、集積の経済を生み出す根源を解明したり、産業集積が形成されるメカニズムについて得られた知見を確認する上で産業集積の実態把握が有益である。

これまで産業集積の実態把握を行う研究は多数行われてきた。産業集積の程度を測る指標がEllison and Glaeser<sup>1)</sup>を始めとする研究によって考案されている。集積の程度を測る指標を用いる多くの実証研究を通じて、産業集積は世界中に存在することが明らかになった。しかし、分

析対象の空間データの持つ情報をスカラーの指標や折れ線グラフ等のグラフに要約する手法では、産業集積の地理空間上の分布を詳細に把握できない。Guillain and Gallo<sup>2)</sup>や井上・糟谷<sup>3)</sup>は町丁目等の詳細な地理的単位で集計されたデータを用いて一都市内の産業集積の分布を明らかにしている。しかし、交通施設や用途地域等の都市計画上の規制といった立地環境と都市内の産業集積の分布の関係は明らかにしていない。立地環境と産業集積の分布の関係を明らかにすることは、産業の立地特性や集積の経済の根源を明らかにする上で有益であると考えられる。そこで、本研究では、詳細な空間データから産業集積が生じている集積領域を検出し、立地環境と産業集積の空間的分布の関係を明らかにすることを目的とする。以下では、地理的単位毎に観測されたデータを利用することを想定して議論を進める。

分析対象範囲には複数の産業集積が存在すると考えられる。複数の集積領域を検出できる集積領域検出手法が必要である。複数の集積領域を検出できる集積領域検出手法はMori and Smith<sup>4,5)</sup>や井上・糟谷が提案している。空間データから集積領域を検出する際、集積領域の幾何形状の制約の仕方が問題となる。集積領域の幾何形状を制限しなければ、空間的にまとまりのある集積領域を検出できない。一方で、幾何形状の仮定を強くすると、集積領域検出結果が仮定に引きずられる恐れがある。Mori

and Smithと井上・糟谷の手法は、両者とも集積領域を構成する地理的単位の連結性を仮定している。本研究では交通施設等の立地環境と都市内の産業集積の空間的分布の関係を明らかにするため、市区町村より詳細な地理的単位で集計された空間データの利用を想定している。詳細な空間データから集積領域を検出する場合、集積領域を構成する地理的単位間の連結性の仮定は妥当でないと考えられる。そこで本研究では、集積領域を構成する地理的単位の連結性の仮定を緩めたモデルベースクラスタリングに基づく集積領域検出手法を適用する。モデルベースクラスタリングは画像処理やパターン認識の分野で広く用いられる。

本稿の構成は以下の通りである。2. では、産業集積の実態把握を行った研究を概説する。3. では、本稿で用いる産業集積検出手法を解説する。4. では、首都圏を分析対象とし、2分の1地域メッシュで観測された事業所数データを用いるケーススタディを示す。5. では、研究成果について総括する。

## 2. 既存研究

産業集積の実態把握を行う実証研究は大きく3つのアプローチに分けられる。

1つ目は、観測された空間データからスカラーの指標を算出し、分析範囲全体に対して集積の程度等を評価するアプローチである。代表的な指標として、Ellison and Glaeserが提案した産業毎の空間的な集中の程度を表す指標がある。この指標は市区町村等の地理的単位毎に観測された各産業の事業所数と従業員数のデータから算出される。Ellison and Glaeserは関連がある複数の産業が近接して立地する現象である共集積の程度を表す指標も提案している。

2つ目のアプローチは、空間データから条件を変えて算出した指標を折れ線グラフ等に視覚的に表すアプローチである。Duranton and Overman<sup>6</sup>が提案したK-densityという指標やMacron and Puech<sup>7</sup>が提案したRipleyのK関数<sup>8</sup>に基づく指標をグラフにプロットする手法が該当する。Duranton and Overmanは、Ellison and Glaeserの指標値がデータの集計単位に依存する可変単位地区問題(MAUP)を抱えていると指摘し、事業所間の距離に基づくMAUPのない指標を提案した。K-densityをグラフにプロットすることで、2つの事業所間の距離の分布が把握できる。

3つ目のアプローチは、産業集積が起きている場所等を地図上に出力し、産業集積の空間的分布を把握するアプローチである。Guillain and Galloは詳細な地理的単位で集計された産業別従業員数データを用いてモラン散布図

を作成した。従業員数の空間的自己相関を可視化することで、産業毎の空間的分布を調査した。Mori and Smith<sup>4,5</sup>や井上・糟谷は複数の産業集積の数や空間範囲を明らかにできる手法を提案している。以下、Mori and Smithと井上・糟谷の手法を概説する。

Mori and Smithは集積領域検出結果である地理的単位の組合せをクラスタスキームと呼んだ。Mori and Smithは観測される空間データの生成過程が確率モデルで表されると仮定し、確率モデルを用いてクラスタスキームと空間データの関係を記述している。モデル選択基準BICを用いて最適なクラスタスキームを選択することで、集積領域を検出する。Mori and Smithの確率モデルは事業所をダーツと見立てた的当てモデルであり、集積領域の幾何形状とは独立である。集積領域の幾何形状は確率モデルとは別に仮定する。Mori and Smithは集積領域を時間距離ネットワーク上の凸包であり、かつ幾何学的な空洞がない地理的単位の集合と仮定している。

井上・糟谷は多重検定の問題に対処する方法の1つであるFDR-controlling法と空間スキャン統計<sup>9</sup>を利用する手法を提案した。空間スキャン統計は代表的な集積領域検出手法であるが、空間スキャン統計を繰り返し適用すると多重検定の問題が生じる。井上・糟谷は空間スキャン統計を繰り返し適用して求めた集積領域の候補の一部をFDR-controlling法に基づいて選択し、集積領域検出結果とした。FDR-controlling法を用いることで、空間スキャン統計を繰り返し適用しながらも検定の多重性の問題に対処している。集積候補領域の幾何形状は、地理的単位の連結性のみを仮定している。

## 3. 本研究の産業集積検出手法

既存の集積領域検出手法はいずれも集積領域を構成する地理的単位の厳格な連結性を仮定している。しかし、詳細な空間データから集積領域を検出する場合、集積領域を構成する地理的単位間の連結性の仮定は妥当でない。そこで本研究では、集積領域を構成する地理的単位の連結性の仮定を緩めたモデルベースクラスタリングに基づく集積領域検出手法を適用する。

### (1) モデルベースクラスタリング

画像処理やパターン認識分野では、画像データを同一の被写体が写された画素の集合に分割し、領域分割を行う手法が多く開発されてきた<sup>10</sup>。その1つに、モデルベースクラスタリングに基づく手法がある。

モデルベースクラスタリングでは、観測された画像データの生成過程が確率モデルで表されると仮定する。この確率モデルは、画像データと領域分割結果の関係を記

述する。確率モデルを仮定した下で、ベイズの定理に従い、事後確率が最大の領域分割を探索することで、画像データを領域分割する。観測された画像データの生成過程を表す確率モデルでは、ある領域分割が事前分布に従って実現し、実現した領域分割に従って画像データが生成されると仮定する。事前分布に従って実現した領域分割と画像データの関係を表す確率モデルをベイズ統計の枠組みで尤度と呼ぶ。

事前分布には Q 状態ポッツモデルがよく用いられる。Q 状態ポッツモデルは磁性体の物性を説明するために提案された確率モデルである。2次元の正方形格子の格子点上のスピンの Q 個の状態を取り得る場合に、距離が近いスピンの同じ状態を取る確率が高いことを表現する。格子点を画素、スピンの状態を画素が所属する領域とみなすと、領域分割の事前分布に利用できる。Q 状態ポッツモデルを事前分布に仮定することで、距離が近い画素が同じ状態をとりやすいという画像データの空間的自己相関を確率モデルに組み込むことができる。

本研究では Q 状態ポッツモデルの格子点を地理的単位、スピンの状態を地理的単位が所属する領域とみなす。これにより、集積領域を構成する地理的単位の連結性の仮定を緩めたモデルベースクラスタリングに基づく集積領域検出手法を構築できる。

## (2) 集積領域検出手法の枠組み

集積領域検出手法の枠組みを解説する。N 個の地理的単位から成る離散的な地理空間を考える。地理的単位  $i$  の観測値を  $f_i$  で表す。分析範囲全体の観測データを  $f = (f_1, \dots, f_N)^T$  で表す。分析範囲に集積領域以外の領域（以下、非集積領域）1つと K 個の集積領域が存在すると仮定する。地理的単位は (K+1) 個の領域の1つに所属する。地理的単位  $i$  が所属する領域を  $a_i$  ( $\in \{0, 1, \dots, K\}$ ) で表し、全ての地理的単位の所属領域を表すベクトルを  $a = (a_1, \dots, a_N)^T$  と定義する。a をクラスタ分割と呼ぶ。クラスタ分割が持つ情報は Mori and Smith のクラスタスキームの情報と等価である。Mori and Smith と同じく、本研究では次の確率的过程を経て観測データが生成されると仮定する。確率モデルを用いてクラスタ分割と観測される空間データの関係を記述する。

- 1) あるクラスタ分割が事前分布  $p(a | \beta)$  に従って実現する。β は事前分布のパラメータの集合である。
- 2) 実現したクラスタ分割の下で尤度  $p(f | a, \gamma)$  に従って空間データが生成される。γ は尤度のパラメータの集合である。

本研究では、確率モデルを仮定した上で、モデル選択基

準ICLを用いてクラスタ分割を選択し、集積領域を検出する。ICLはBiemacki<sup>12)</sup>が提案したモデルベースクラスタリングにおける代表的なモデル選択基準であり、式(1)で定義される。

$$ICL(K) = \int p(f, a | \theta_K) \pi(\theta_K) d\theta_K \quad (1)$$

$\theta_K = \beta \cup \gamma$  とおく。Biemackiは式(1)を式(2)で近似した。

$$ICL(K) \approx \log p(f, \hat{a} | \hat{\theta}_K) - \frac{v_K}{2} \log N \quad (2)$$

$\hat{\theta}_K$  はパラメータの最尤推定値、 $\hat{a}$  は  $\hat{\theta}_K$  の下で事後確率が最大のクラスタ分割（式(3)）、 $v_K$  はパラメータの総数である。

$$\hat{a} = \arg \max_a p(a | f, \hat{\theta}_K) \propto p(f, a | \hat{\theta}_K) \pi(\hat{\theta}_K) \quad (3)$$

Mori and Smithが用いたBICは観測データをよく説明するモデルを選択するため、クラスタ分割が過学習を起こす恐れがある。一方、ICLは最も妥当なクラスタリング結果を選択することを目指す指標である。ICLはBICよりもクラスタの数が少ないモデルを選択する傾向がある。ICLは安定的に適切なクラスタの数を選択できる指標として知られており、実証分析で広く用いられる。そこで、本研究ではICLを用いてクラスタ分割を選択する。本研究の集積領域検出問題を式(4)で定式化する。ICLに基づくクラスタ分割の選択を通じて、内生的に集積領域の数を決定できる。パラメータは全て無情報事前分布に従うと仮定する。

$$\begin{aligned} (\hat{a}, \hat{\beta}, \hat{\gamma}) &= \arg \max_{a, \beta, \gamma} ICL(a, \beta, \gamma) \\ &= \arg \max_{a, \beta, \gamma} \ln p(f | a, \gamma) p(a | \beta) - \frac{v_K}{2} \ln N \end{aligned} \quad (4)$$

## (3) 確率モデルの定式化

空間データの生成過程を表す確率モデルが満たすべき要件としては以下の3つを考える。第一の要件は、観測値の小さい地理的単位と観測値の大きい地理的単位が別々の領域に所属する場合に高い確率を示すことである。第二の要件は、距離が近い地理的単位が集積領域を構成する場合に高い確率を示すことである。第三の要件は、集積領域の幾何形状の制約を満たさない場合にゼロに近い確率を示すことである。本研究では第一の要件には多変量正規分布を尤度に用いることで、第二と第三の要件にはQ状態ポッツモデルを事前分布に用いることで対応する。

### a) 尤度の定式化

尤度は式(5)の観測値間の独立性を仮定した多変量正規分布で定義する。

$$p(f | a, \mu, \sigma) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(f_i - \mu_{a_i})^2}{2\sigma^2} \right\} \quad (5)$$

$\mu$  ( $=\{\mu_0, \dots, \mu_k\}$ ) は正規分布の平均値,  $\sigma$  は分散を表すパラメータである ( $\gamma = \mu \cup \sigma$ ). 式(5)の尤度は観測値が類似する地理的単位が同じ領域に所属するほど大きな値を取る. 観測値の小さい地理的単位と観測値の大きい地理的単位が別々の領域に所属する場合に高い確率を示す. 式(4)の定式化の下では, 尤度のパラメータ推定値は次式(6)となる.

$$\begin{cases} \hat{\mu}_k = \sum_{i \in C_k} f_i / N_k \\ \hat{\sigma} = \sqrt{\sum_i (f_i - \mu_{a_i})^2 / N} \end{cases} \quad (6)$$

$N_k$  は領域  $k$  に所属する地理的単位の数,  $C_k$  は領域  $k$  に所属する地理的単位の集合である.

## b) 事前分布の定式化

クラスタ分割の事前分布は画像処理分野でよく用いられる式(7)のQ状態ポッツモデルを援用して定式化する.

$$p(a | \alpha) = \frac{1}{Z(\alpha)} \exp(-H(a, \alpha)) \quad (7)$$

$\alpha$  はパラメータ ( $\alpha \geq 0$ ),  $H$  はポッツモデルのエネルギー関数 (式(8)),  $Z$  は規格化定数である.

$$H(a, \alpha) = - \sum_{k \in \{0, \dots, K\}} \sum_{(i, j) \in N_{C_k}} \frac{1}{2} \alpha w_{ij} \quad (8)$$

$w_{ij}$  は地理的単位  $(i, j)$  間の空間重み係数,  $N_{C_k}$  は領域  $k$  に所属し, かつ隣接する地理的単位のペアの集合である. ポッツモデルのエネルギー関数値が小さいクラスタ分割ほど, 事前確率は大きくなる. 式(7)のポッツモデルは距離が近い地理的単位が集積領域を構成する場合に高い確率を示す. しかし, エネルギー関数値は常にゼロ以下であり, 任意のクラスタ分割が一定の確率で実現すると仮定することを意味する. 式(7)の事前分布を用いた場合, 遠く離れた地理的単位からなる集積領域等, 集積領域として不適切なクラスタ分割が出力される恐れがある. そこで, 本研究ではポッツモデルのエネルギー関数に集積領域の幾何形状の制約を表すペナルティを加えることで, 確率モデルの中で集積領域の幾何形状を制限する. ペナルティが大きくエネルギー関数値が大きいクラスタ分割は事前分布で実現する確率がほぼゼロになる.

本研究では2種類のペナルティを用いることを提案する. 第一のペナルティは, 遠く距離を隔てた地理的単位が同じ集積領域に所属することを防ぐためのペナルティである (以下, ペナルティAと呼ぶ). 集積領域の連結性を仮定しない場合, 非常に離れた地理的単位が同じ集積領域に所属する可能性が生じる. 空間的にまとまりのある集積領域を検出するには, 遠く距離を隔てた地理的単位が同じ集積領域に所属することを防ぐ必要がある. 第二のペナルティは, 多数の集積領域が入り乱れたクラスタ分割の検出を防ぐためのペナルティである (以下, ペナルティBと呼ぶ). 正規分布を仮定する式(5)の尤度

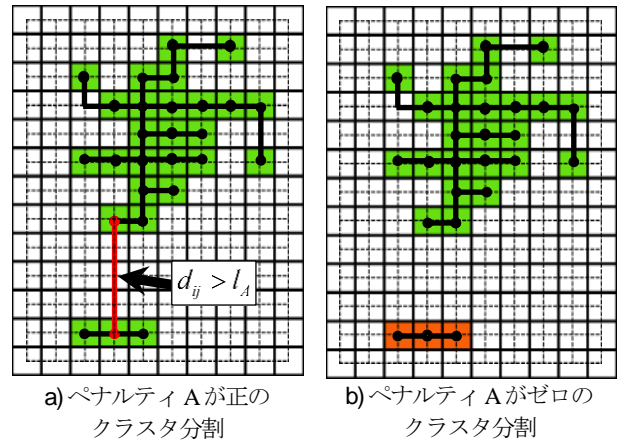


図-1 ペナルティ A

は, 観測値の大きさ毎に地理的単位が別れて所属すると最も大きな値になる. しかし, 尤度が最大になるクラスタ分割は, 多数の集積領域が入り乱れ, 集積領域検出結果として妥当でないと考えられる. 多数の集積領域が入り乱れたクラスタ分割の検出を防ぐ必要がある.

本研究ではペナルティAを仮想的な時間距離ネットワークを利用して定式化する. 社会経済活動を観測した空間データからの集積領域検出を考える場合, 地理的単位間の距離は交通ネットワーク上の時間距離と考えるのが妥当である. しかし, 現実の交通ネットワークを確率モデルに組み込むのは容易ではない. そこで, 仮想的な時間距離ネットワーク  $G$  を仮定して地理的単位間の距離を定義する. グラフ  $G$  は地理的単位をノード, 隣接する地理的単位間にリンクを張る無向重み付きグラフである. リンクの重みは地理的単位間の時間距離である.

グラフ  $G$  を利用すると, ペナルティAは式(9)で定式化できる.

$$P_A = \sum_{k \in \{1, \dots, K\}} \sum_{(i, j) \in L_{C_k}} (d_{ij} \delta(d_{ij} > l_A)) \quad (9)$$

$P_A$  はペナルティAの値,  $l_A$  はパラメータ,  $d_{ij}$  は地理的単位  $(i, j)$  間のグラフ  $G$  上の最短距離,  $L_{C_k}$  は集積領域  $k$  に所属する地理的単位をノード,  $d_{ij}$  をリンクの重みとするグラフの最小全域木  $T_k$  を成すリンクの集合,  $\delta(d_{ij} > l_A)$  は  $d_{ij} > l_A$  の場合に1, それ以外は0を取るデルタ関数である. 図-1にペナルティAが正のクラスタ分割とゼロのクラスタ分割を例示した. 実線の格子で区切られた長方形が地理的単位を表す. 地理的単位は長方形の辺を共有する場合に隣接すると考える. 図-1の破線が時間距離ネットワーク  $G$  を表す. 隣接地理的単位間の時間距離は一定とおく. 同じ色の地理的単位は同じ集積領域に所属することを意味する. 白色の地理的単位は非集積領域に所属する.  $l_A$  を隣接地理的単位間の時間距離の3倍とおく. 図-1 a)のクラスタ分割は, 集積領域の最小全域木  $T$  が赤線で示した  $d_{ij} > l_A$  であるリンクを持つ

ためペナルティAが正となる。図-1 b)のクラスタ分割は、全ての集積領域の最小全域木も  $d_{ij} > l_A$  であるリンクを持たないためペナルティAがゼロとなる。ペナルティAを式(9)で定式化することで、遠く距離を隔てた地理的単位が同じ集積領域に所属することを防ぐことができる。

ペナルティBは各集積領域の地理空間上の凸包に基づいて式(10)で定式化できる。

$$P_B = \sum_{k \in \{1, \dots, K\}} \sum_{k' \in \{1, \dots, K\} \setminus k} s_{k,k'} \quad (10)$$

$s_{k,k'}$  は集積領域  $k$  と  $k'$  の凸包が重複する部分の地理的単位の面積の和である。式(10)は 2 つの集積領域の凸包が重複する場合に、複数の集積領域が入り乱れていると判定し、ペナルティ B を課す。図-2 にペナルティ B が正のクラスタ分割とゼロのクラスタ分割を例示した。図-2 の見方は図-1 と同じである。図-2 a)のクラスタ分割では、緑色の集積領域の凸包が黄色の集積領域を含むため、ペナルティ B が正となる。緑色の集積領域と黄色の集積領域が 1 つになり、図-2 b)のクラスタ分割になると、ペナルティ B はゼロとなる。ペナルティ B を式(10)で定式化することで、多数の集積領域が入り乱れたクラスタ分割の検出を防ぐことができる。

本研究ではペナルティAとペナルティBをQ状態ポッツモデルのエネルギー関数に加えることで、集積領域の幾何形状を制約する。式(11)でクラスタ分割の事前分布を定義する。

$$p(a | \alpha, b_A, b_B) = \frac{1}{Z(\alpha, b_A, b_B)} \exp(-H(a, \alpha, b_A, b_B)) \quad (11)$$

$\alpha$  と  $b_A$ ,  $b_B$  はパラメータ,  $H$  はポッツモデルのエネルギー関数である (式(12)) .

$$H(a, \alpha, b_A, b_B) = - \sum_{k \in \{0, 1, \dots, K\}} \sum_{(i,j) \in L_k} \frac{1}{2} \alpha w_{ij} + b_A P_A + b_B P_B \quad (12)$$

ポッツモデルのエネルギー関数の第一項 (以下, 集積力項) の目的は, 距離が近い地理的単位が同じ領域に所属するクラスタ分割の実現確率を高くすることである。集積力項は集積領域及び非集積領域毎に分解できる。ペナルティAの定式化で用いた最小全域木  $T$  を利用して各領域の集積力項を定義する。

尚, 式(11)の事前分布のパラメータは本稿では分析者が設定する。パラメータを固定することで, 厳密な計算が現実的に不可能なポッツモデルの規格化定数  $Z$  の計算を回避する。事前分布のパラメータの推定は今後の課題である。

#### (4) モデルの解法

式(4)で定式化される集積領域検出問題は地理的単位の組合せを求める組合せ最適化問題であり, 大域的最適解を求めることは困難である。そこで本研究では, 分析

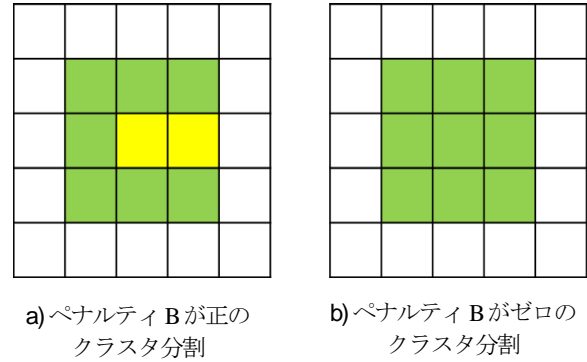


図-2 ペナルティ B

範囲を複数の分析範囲に分割し, 近似的に式(4)の問題を解く。全体の分析範囲を分割して得られる分析範囲を小分析範囲と呼ぶ。分析範囲を複数の小分析範囲に分割して近似的に式(4)の問題を解く手順は次の通りである。

- 1) 全体の分析範囲を複数の小分析範囲に分割する。
- 2) 小分析範囲毎に式(4)の問題を解き, それぞれクラスタ分割を求める。
- 3) 小分析範囲毎に得られたクラスタ分割をd)とe)の探索を行って統合し, 全体の分析範囲に対する集積領域検出結果とする。

小分析範囲毎にクラスタ分割を求める際は, 組合せ最適化問題に対する代表的な近似解法である遺伝的アルゴリズムを用いる。クラスタ分割を個体の遺伝子情報とし, ICLで個体の適合度を評価する。遺伝的アルゴリズムの終了条件は, 最適なクラスタ分割の適合度と個体群全体の適合度の平均値の差の絶対値が5回連続で5以下になることと設定した。本研究の遺伝的アルゴリズムの操作は, 局所探索と突然変異, 適合度の高い個体の複製から成る。以下1(1)a)からc)でそれぞれの役割を解説する。

##### a) 局所探索

式(4)の組合せ最適化問題は局所最適解を求めることは困難と考えられるが, 定式化の特徴から適合度が高い個体の特徴を予想できる。式(5)の尤度は観測値の小さい地理的単位と観測値の大きい地理的単位が別々の領域に所属する場合に大きくなる。式(11)の事前分布は2つのペナルティがゼロで距離が近い地理的単位が同じ領域に所属する場合に大きくなる。観測値が大きく距離が近い地理的単位で1つの集積領域が構成され, 観測値が小さい地理的単位で非集積領域が形成されるクラスタ分割の個体の適合度が高くなると予想できる。本研究では少ない世代交代回数で適合度の高い個体を生み出し, 計算時間を短くするため, 確定的な局所探索を行う。具体的には, 適合度が上昇しなくなるまで集積領域を拡大する

局所探索，集積領域を縮小する局所探索，2つの集積領域を1つに統合する局所探索を行う。

#### b) 突然変異

a)の確定的な局所探索だけでは，大域的最適解を求められない可能性が高い。そこで，局所最適解から個体を脱出させるため，適合度の改悪を許容する確率的な操作を突然変異として実行する。具体的には，確率的な集積領域の拡大と縮小，統合，分割，新設，消去，複数の地理的単位の所属領域の変更を行う。

#### c) 適合度の高い個体の複製

通常の遺伝的アルゴリズムでは，交差で適合度の高い個体と類似した個体を作成し，局所最適解の探索を行う。一方，本研究のアルゴリズムでは，a)の局所探索で確定的に局所最適解の探索を行っており，改めて局所最適解の探索を行う必要はないと考えられる。そこで，本研究では適合度の高い個体を複製することで，適合度の計算を回避しつつ，適合度の高い個体を作成する。本研究の定式化で最も計算に時間を要する部分は，各クラスタ分割の集積領域毎に最小全域木 $T$ を求める部分である。適合度を計算する回数が少ないほど，計算を高速化できる。

#### d) 観測値の低い集積領域の消去

分析範囲を複数の小分析範囲に分割する場合，小分析範囲毎に非集積領域に所属する地理的単位の観測値の平均値 $\mu_0$ や分散パラメータ $\sigma$ が異なる。集積領域として検出されるために必要な観測値の大きさに小分析範囲間で差が生じる。その結果，分析範囲全体で評価すると，非集積領域とした方がICLが大きくなる集積領域が現れる場合がある。そこで，小分析範囲毎に得られたクラスタ分割を連結させる際に，各集積領域について分析範囲全体でICLを評価し，分析範囲全体で見ると観測値が低い集積領域を非集積領域にする操作を行う。

#### e) 小分析範囲の境界部に位置する集積領域の統合

分析範囲を分割する際の境界を跨ぐ集積領域は，複数の小分析範囲で別々の集積領域として検出される。しかし，これらの集積領域は分析範囲を分割する際の境界を跨いでいなければ，1つの集積領域として検出されていたと考えられる。そこで，ペナルティ $A$ が作用しない小分析範囲の境界からの距離が $l_A$ 以下の地理的単位を含む集積領域のペアについて，統合した場合のICLの方が大きければ集積領域を統合する操作を行う。

## 4. ケーススタディ

3. で解説した集積領域検出手法を2分の1地域メッシュで観測された事業所数データに適用し，産業集積の空間的分布を把握する。

### (1) ケーススタディの条件

#### a) 使用データ及び事前分布のパラメータ

平成21年経済センサス・基礎調査で観測された2分の1地域メッシュの86の産業中分類の事業所数を用いる。事業所数データをメッシュの面積で割った事業所密度データに本研究の集積領域検出手法を適用する。本研究の集積領域検出手法は市区町村や町丁目等の面積に差がある地理的単位で観測された空間データへの適用も想定しているため，面積で基準化した事業所密度データを用いる。2分の1地域メッシュの面積は，独立行政法人統計センターが運用管理するウェブサイト (<http://www.e-stat.go.jp/SG1/estat/toukeiChiri.do?method=init>) から入手した2分の1地域メッシュのポリゴンデータからArcGISで算出する。2分の1地域メッシュ間の時間距離は，メッシュの重心間の直線距離とする。メッシュの重心もArcGISで求める。重心間の時間距離が $d_{ij}$ のメッシュのペアに対して空間重み係数 $w_{ij}$ を $w_{ij} = 1/d_{ij}^2$ と設定する。ポリゴンの辺を共有しているメッシュ同士を隣接しているとみなす。

事前分布のパラメータは $\alpha = 5$ ， $b_A = b_B = 10^6$ ， $l_A = 2(km)$ と設定する。

#### b) 分析範囲と小分析範囲の設定

東京都を中心とする62,800メッシュを分析範囲とし，図-3に赤枠で示した165の小分析範囲に分析範囲を分割して集積領域検出手法を適用した。

### (2) 適用結果

本研究では86の産業中分類の事業所密度データから産業集積を検出した。計算は開始から3日で終了した。本稿では，第三次産業の例として情報サービス業，第二次産業の例として，化学工業の産業集積検出結果を示す。

#### a) 情報サービス業の産業集積検出結果

図-4に情報サービス業について産業集積を検出した結果を示す。図-4a)は集積領域検出の対象である事業所

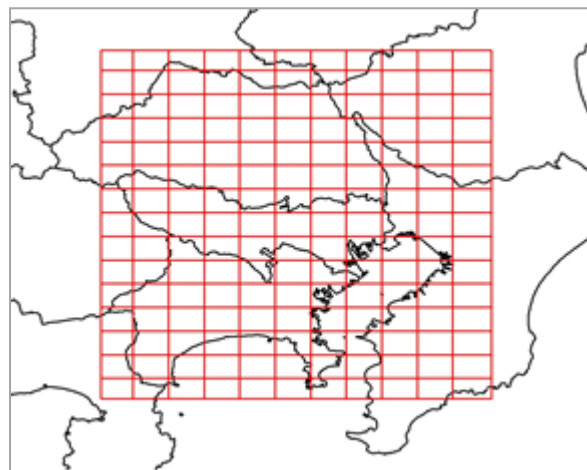


図-3 分析範囲と小分析範囲



図-4 情報サービス業の産業集積検出結果

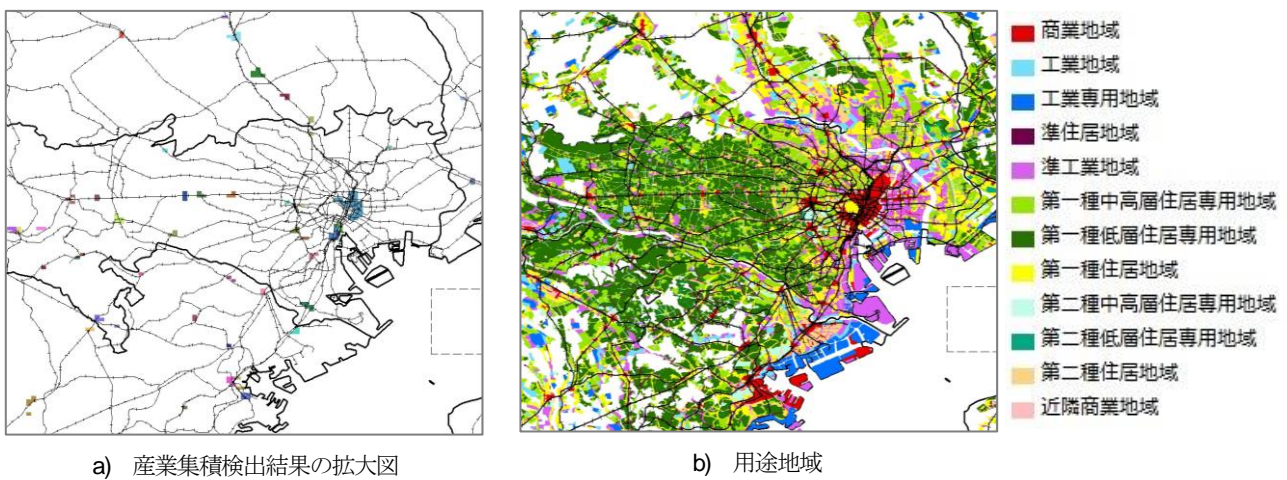


図-5 情報サービス業の産業集積検出結果の拡大図

密度データである。黒の濃いメッシュほど事業所密度が高いことを意味する。図-4b)が集積領域検出結果である。同じ色のメッシュは同じ集積領域に所属することを意味する。白色のメッシュは非集積領域に所属する。図-4c)は小分析範囲毎に求めたクラスタ分割を連結させただけのクラスタ分割である。図-4a)より、情報サービス業の事業所は東京の都心部に多く立地していることが読み取れる。図-4c)より、小分析範囲毎に求めたクラスタ分割では、分析範囲全体で見ると有意ではない産業集積が多数検出されることが分かる。図-4b)より、観測値が低く有意ではない産業集積は、3.3(4)d)で説明した観測値が低い集積領域を非集積領域にする操作で取り除かれていることが確認できる。一方で、図-4a)で赤丸で囲んで示した範囲は事業所密度が高いにも関わらず、集積領域として検出されていない。この原因は、小分析範囲毎に得られたクラスタ分割を統合する際に、小分析範囲毎の分析で非集積領域に割り当てられた地理的単位を集積領域にする操作を行っていない点にある。小分析範囲毎に得られたクラスタ分割を統合する際に、分析範囲全体で見

て観測値が高い地理的単位を集積領域にすることを検討する操作を加える必要がある。

図-5 a)に図-4 b)の東京都区部を含む範囲を拡大して示した。図-5 b)は図-5 a)と同じ範囲の平成23年度時点の用途地域を示した図である。図-5には鉄道路線網も併せて示した。図-5より、多くの場合、情報サービス業の事業所は鉄道路線沿いの商業地域に立地することが分かる。図-5 a)で集積領域になっているメッシュのほとんどが図-5 b)で赤色の商業地域になっている。

#### b) 化学工業の産業集積検出結果

図-6に化学工業について産業集積を検出した結果を示した。図の見方は図-4と同じである。図-6より、次の2点分かる。第一に、化学工業の事業所は情報サービス業と比べて広い範囲に立地しており、東京都区部以外にも多くの産業集積が検出されている。第二に、情報サービス業の場合と同様に、分析範囲全体で見ると観測値が高いにも関わらず、非集積領域に割り当てられている地理的単位が東京都区部に多く存在する。化学工業について

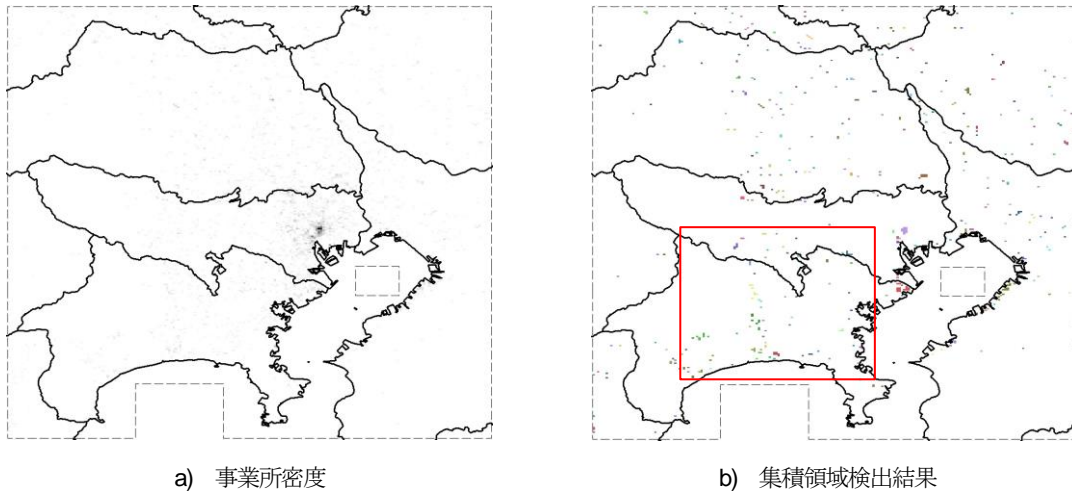


図-6 化学工業の産業集積検出結果

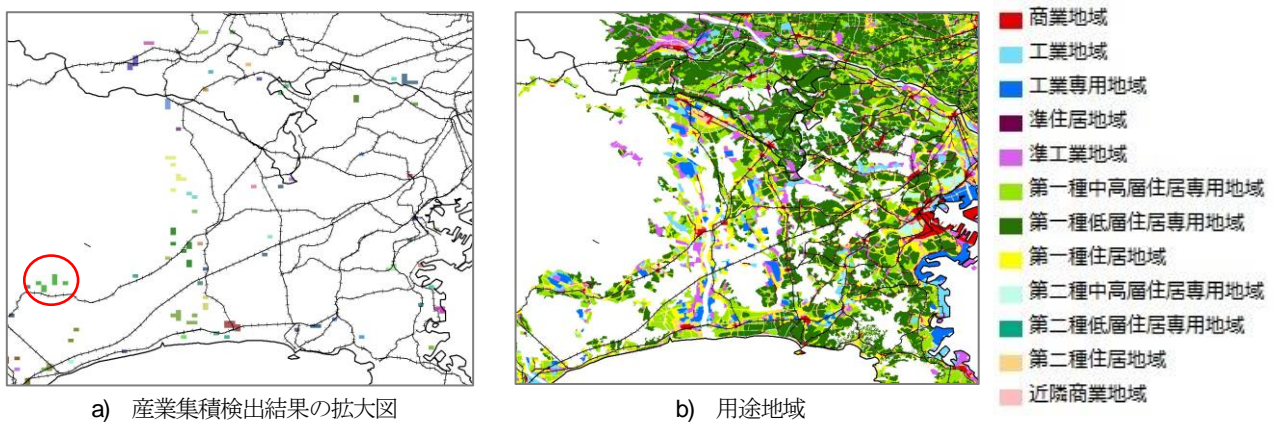


図-7 化学工業の産業集積検出結果の拡大図

の結果からも、小分析範囲毎に得られたクラスタ分割を統合する際に、分析範囲全体で見て観測値が高い地理的単位を集積領域にすることを検討する操作を加える必要があることが分かる。

図-7は図-6b)で赤枠で囲んだ範囲を拡大した図である。図-7b)は図-7a)と同じ範囲の用途地域を示した図である。図-7には鉄道路線網も併せて示した。図-7より化学工業の事業所は沿岸部の工業地域・工業専用地域だけでなく、内陸に点在する工業地域・工業専用地域にも多く立地することが読み取れる。図-7a)で集積領域になっているメッシュと対応する場所が図-7b)で工業地域・工業専用地域になっているケースが多く観察される。また、図-7a)で赤丸で囲んだ集積領域が連結しないメッシュから構成されていることも確認できる。本研究で用いた集積領域検出手法により、連結しない地理的単位から構成される集積領域を検出できる。

## 5. おわりに

本研究では、集積領域を構成する地理的単位の連結性の仮定を緩めた集積領域検出手法を用いて、首都圏から産業集積を検出した。本手法により、所属する地理的単位が連結していない産業集積を検出できることを確認した。また、鉄道路線網や用途地域と検出された産業集積の分布を比較し、立地環境と都市内の産業集積の空間的分布が関連することを示唆する事例を示した。

本研究の今後の課題として、次の3点が挙げられる。第一に、本稿の分析は断片的であるため、より網羅的に産業集積の空間的分布と立地環境の関係を明らかにする必要がある。第二に、小分析範囲毎に得られたクラスタ分割を統合する際に、分析範囲全体で見て観測値が高い地理的単位を集積領域にすることを検討する操作を加える必要がある。第三に、集積領域検出結果が集積領域検出手法で用いたポッツモデルのパラメータに依存するため、ポッツモデルのパラメータをデータから推定できるようにする必要がある。



謝辞：本研究は東京大学空間情報科学研究センターの空間データ利用を伴う共同研究（No. 582）の成果であり，平成21年経済センサス・基礎調査地域メッシュ統計のデータを利用した。

#### 参考文献

- 1) Ellison, G. and Glaeser, E. L. : Geographic concentration in US Manufacturing industries: A dartboard approach, *Journal of Political Economy*, *Journal of Political Economy*, Vol. 105, No.5, pp.889-927, 1997.
- 2) Guillain, R. and Gallo, J. L. : Agglomeration and dispersion of economic activities in and around Paris: an exploratory data analysis, *Environment and Planning B: Planning and Design*, Vol. 37, pp.961-981, 2010.
- 3) 井上 亮, 糟谷志帆 : 複数の柔軟な形状を持つ点事象集積地域の検出法の開発, *土木計画学研究・概要集*, Vol. 49, 2014.
- 4) Mori, T. and Smith, T. E. : An industrial agglomeration approach to central place and city size regularities, *Journal of Regional Science*, Vol. 51, No. 4, 2011, pp.694-731, 2011.
- 5) Mori, T. and Smith, T. E. : A probabilistic modeling approach to the detection of industrial agglomerations, *Journal of Economic Geography*, pp. 1-42, 2013.
- 6) Duranton, G. and Overman, H. G. : Testing for localization using micro-geographic data, *Review of Economic Studies*, Vol. 72, Issue 4, pp. 1077-1106, 2005.
- 7) Macron, E. and Puech, F. : Evaluating the geographic concentration of industries using distance-based methods, *Journal of Economic Geography*, Vol. 3, pp.409-428, 2003.
- 8) Ripley, B. D. : Modelling spatial patterns, *Journal of Royal Statistical Society B*, Vol. 39, pp.172-212, 1997.
- 9) Kulldorff, M. and Nagarwalla, N. : Spatial disease clusters – detection and influence, *Statistic in Medicine*, Vol.14, pp. 799-810, 1995.
- 10) Cheng, H. D., Jiang, X. H., Sun, Y. and Wang, J. : Color image segmentation: advances and prospects, *Pattern Recognition*, Vol. 34, pp.2259-2281, 2001.
- 11) Cucala, L. and Marin, J.-M. : Bayesian inference on a mixture model with spatial dependence, *Journal of Computational and Graphical Statistics*, Vol. 22, Issue 3, pp.584-597, 2013.
- 12) Biernacki, C., Celeux, G. and Govaert, G. : Assessing a mixture model for clustering with the integrated completed likelihood, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 7, pp.719-725, 2000.

(2015. 4. 24受付)