

Q状態ポッツモデルを用いた産業集積の検出

氏家 晃仁¹・福本 潤也²

¹学生会員 東北大学大学院情報科学研究科 博士後期課程 (〒980-8577仙台市青葉区片平2-1-1)

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

²正会員 東北大学大学院准教授 情報科学研究科 (〒980-8577仙台市青葉区片平2-1-1)

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

産業集積では経済活動で関係する多数の企業が空間的に近接して立地する。産業集積の実態把握を行う手法を確立できれば、産業集積に着目した政策を議論する際の有益な基礎資料を提供できる。

本研究では、形状に制約を置かずに産業集積を検出する手法を提案する。具体的には、Q状態ポッツモデルとベイズ統計のアプローチを組み合わせた産業集積検出手法を提案する。ケーススタディを通じて、本手法により産業集積形状に制約を置くことなく、産業集積を検出できることが明らかになった。飛び地のある産業集積や凸性を満たさない形状の産業集積が検出できることを確認した。

Key Words : Industrial Agglomeration, Clustering, Bayesian Estimation, Q-state Potts Model

1. はじめに

多数の企業が空間的に近接して立地する産業集積は世界中で観察される。工業団地・都市・工業地帯等の存在が示す通り、産業集積は様々なスケールで広く存在している。理論と実証の両面から産業集積の形成要因や分布について研究が行われてきた。一般に産業集積は産業競争力を高める効果を持つと考えられており、政策的関心も高い。地域活性化や国際競争力強化のため、産業集積を促す多数の政策が推進されている。

産業集積の促進策を検討する上で、各産業の分布の特徴や、産業集積を構成する産業の種類等の把握が有効である。産業集積の実態把握の手法の確立は、産業政策の議論を行う上で有益な基礎資料を提供できる。

Mori and Smithは産業集積の数・範囲・分布を把握する手法を提案した¹⁾。Mori and Smithの手法では、一度に複数の集積を検出できる。しかし、産業集積の形状に旅行距離ネットワーク上の凸性と連続性の仮定を置いている。そのため、飛び地のある産業集積や凸性を満たさない形状の産業集積が検出できないという課題がある。

本研究では飛び地のある産業集積や凸性を満たさない産業集積を検出する手法を開発する。具体的には、Q状態ポッツモデルとベイズ統計のアプローチを組み合わせた定式化を行う。市区町村単位の事業所密度データを用いたケーススタディを通じて、本手法の有用性を検証する。

2. 既存研究

Mori and Smithは空間スキャン統計の集積検出手法を拡張した産業集積検出手法を提案した。空間スキャン統計はもともと疫病の高リスク地域を統計学的に厳密に検出するために提案された手法である。空間スキャン統計では円形等の集積候補領域を設定し、統計的に最も有意に集積が見られる領域を検出する²⁾。この手法には大きく2つの課題がある。第一に、一度に1つしか集積を検出できない。同手法を繰り返し適用すれば複数の集積を検出できるが、検定の多重性の問題が生じる。第二に、集積候補領域と同じ形状の集積しか検出できない。

Mori and Smithは、一度に複数の集積を検出でき、集積形状の制約を緩めた手法を提案した。地理的単位 (e.g. 市区町村) から成る離散的な地理空間を対象に、産業集積を事業所密度が均一な凸状の連続した地域と考えた。産業集積の集合であるクラスタスキームと事業所分布との間に確率的関係を仮定し、事業所分布に最も適合するクラスタスキームを結合法で探索する。

Mori and Smithの手法には大きく2つの課題がある。第一に、大規模データの分析に向かない。その理由は、結合法でクラスタスキームを更新する際に、全結合候補について凸性と連続性を満たす範囲を求める必要があり、計算量が多くなるからである。

第二に、産業集積形状が凸性と連続性を満たさなければならぬという制約がある。Mori and Smithの確率モデ

ルでは、ある地理的単位に事業所が立地する確率が地理的単位の経済面積に比例する多項分布に従う。同モデルでは、地理的な要素を考慮していない。空間的に近接した地域を産業集積として検出するには、凸性と連続性の仮定を置く必要がある。同仮定により、Mori and Smithの手法は町丁目単位のデータ等の詳細地理情報への適用が難しいと考えられる。産業集積形状の凸性と連続性は詳細地理情報に用いる場合、一般に成り立たないと考えられるからである。

3. 提案手法

本研究では、産業集積形状の制約を取り払い、任意の形状の産業集積を検出できる手法を提案する。具体的には、ベイズ統計のアプローチを採用し、事前確率にQ状態ポッツモデルを用いることにより、任意の形状の産業集積の検出を試みる。Mori and Smithは産業集積形状に凸性と連続性の仮定を置くことで、遠方の地理的単位が同一の産業集積を形成することがないようにしている。一方、本研究では形状の仮定を取り払う代わりに、Q状態ポッツモデルに遠方の地理的単位が同じ産業クラスタになりにくい効果（以下、分散力と呼ぶ）を取り入れることで、遠方の地理的単位が同一の産業集積を形成しないようにする。

(1) 本手法の枠組み

N 個の地理的単位から成る地理空間を考え、地理的単位 i のデータを f_i で表す。全ての地理的単位のデータを $f = (f_1, \dots, f_N)^T$ で表す。分析範囲には K 個の産業集積と産業集積以外の領域（以下、非集積領域）1つの合計 $(K+1)$ 個の領域が存在すると仮定する。非集積領域の領域番号を0番とする。各地理的単位は $(K+1)$ 個の領域の1つに必ず所属する。地理的単位 i が所属する産業集積を a_i で表し、 $a = (a_1, \dots, a_N)^T$ と定義する。 a をクラスタ分割と呼ぶ。データ f は次のプロセスを経て生成されると仮定する。

- 1) あるクラスタ分割 a が事前確率 $\Pr(a)$ に従って生起する。
- 2) 生起したクラスタ分割 a の下で、尤度 $\Pr(f | a)$ に従ってデータ f が生成される。

クラスタ分割はベイズ推定の枠組みで原因と呼ばれる。観測データ f に対して、原因となりうるクラスタ分割は多数考えられる。ベイズ推定では、各クラスタ分割が原因であるデータ f についての条件付き確率（事後確率）を求める。本手法では事後確率が最大のクラスタ分割を

産業集積検出結果とする。産業集積検出問題は式(1)で定義される。

$$\max_a \Pr(a | f) \propto \Pr(f | a) \Pr(a) \quad (1)$$

式(1)は組み合わせ最適化問題であり、大域的最適解を求めることは一般に難しい。本研究では式(1)の問題を式(2)の事後周辺確率最大化問題に置き換え、近似的にクラスタ分割を求める。

$$\max_{k \in \{0, 1, \dots, K\}} \Pr(a_i = k | f) \quad \forall i \quad (2)$$

(2) 定式化

a) Q状態ポッツモデルによる事前確率

本研究ではクラスタ分割を通じて産業集積を検出するが、検出される産業集積は近接した地理的単位によって構成されることが望ましい。そこで、本研究ではQ状態ポッツモデル³⁾に基づく式(3)で定義する事前確率を用いる。

$$\begin{aligned} \Pr(a | \alpha, \alpha') \\ = \frac{1}{Z_a} \prod_{ij} \exp \left(\frac{1}{2} \alpha (w_{ij} - \bar{w}) \delta(a_i = a_j \neq 0) + \frac{1}{2} \alpha' w_{ij} \delta(a_i = a_j = 0) \right) \end{aligned} \quad (3)$$

w_{ij} は地理的単位 i, j 間の空間重み係数、 \bar{w} は空間重み係数の平均値、 α, α' はパラメータ、 δ は括弧内の条件式が真の場合に1、偽の場合に0になる関数である。

指数の中の第1項は地理的単位が同じ産業集積に所属する場合の作用を表し、第2項は地理的単位が共に非集積領域に所属する場合の作用を表している。第1項について、地理的単位 i, j が近接していれば、 $w_{ij} - \bar{w} > 0$ となり、地理的単位 i, j が同じ産業集積に所属する場合に事前確率が大きくなる。一方、地理的単位 i, j が近接しておらず、 $w_{ij} - \bar{w} < 0$ となる場合は、同じ産業集積に所属しない場合に事前確率が大きくなる。

b) 空間相互作用を考慮した尤度

Mori and Smithは産業集積内の事業所密度を均一と考えて定式化した。実際のデータでは観測値が均一になるとは限らない。観測値の不均一さに対応する尤度として、式(4)の空間相互作用を考慮する正規分布を考える。

$$\begin{aligned} \Pr(f | a, \lambda, \mu, \sigma) \\ = \frac{1}{Z_f} \prod_i \exp \left\{ -\frac{(f_i - \mu)^2}{2\sigma^2} \right\} \prod_{ij} \exp \left\{ \frac{1}{2} \lambda w_{ij} (f_i - \mu)(f_j - \mu) \delta(a_i = a_j) \right\} \end{aligned} \quad (4)$$

λ, μ, σ はパラメータである。近接していて、観測値が平均値より大きい地理的単位が1つの産業集積になる場合に尤度が大きくなる。式(4)の相互作用項は i について固定して変形すると、式(5)のように変形できる。

$$\exp \left\{ \frac{1}{2} \lambda (f_i - \mu) \sum_j w_{ij} (f_j - \mu) \delta(a_i = a_j) \right\} \quad (5)$$

Anselinが提案したLocal Moran's I^4 は次式で定義され、

$$I_i = (f_i - \mu) \sum_j w_{ij} (f_j - \mu) \quad (6)$$

式(5)の指数の中身と式(6)が同じ構造を持っていることが分かる。観測値が平均値よりも大きく、近接した地理的単位を検出しようとする定式化であると言える。

(3) 提案モデルの解法

提案モデルは次の事後確率最大化法により解く。

- 1) パラメータの初期値と産業集積の数 K を与える。
- 2) パラメータを与えてクラスタ分割問題を解く。

$$\max_{k \in \{0, 1, \dots, K\}} \Pr(a_i = k | f, \alpha, \alpha', \lambda, \mu, \sigma) \quad \forall i \quad (7)$$

- 3) クラスタ分割を与えてパラメータ推定問題を解く

$$\max_{\alpha, \alpha', \lambda, \mu, \sigma} \Pr(\alpha, \alpha', \lambda, \mu, \sigma | f, a) \quad (8)$$

- 4) 2)-3)をクラスタ分割が収束するまで繰り返す

a) クラスタ分割問題

式(7)の目的関数である事後周辺確率は確率伝搬法⁵⁾で求める。確率伝搬法は大規模確率モデルの近似アルゴリズムの1つであり、大規模な確率モデルから効率よく周辺確率を計算できる。確率伝搬法を用いることで、提案モデルのように複雑な確率モデルに対しても周辺確率を計算できる。

尚、クラスタ分割問題を解く際、所属する産業集積を固定した地理的単位（以下、コア）を設定する。クラスタ分割問題では産業集積を番号で区別するが、番号自体は意味を持たない。コア等を設定しなければ、次に説明する産業集積番号の対称性の問題が生じる。例えば、宮城県に仙台を中心とする産業集積がある場合、仙台の産業集積にどの番号を割り当てても、事後確率は変わらない。事後周辺確率最大化でクラスタ分割問題を解くには、産業集積番号で産業集積を区別できる必要がある。そこで、産業集積に番号を固定したコアを設定し、番号で産業集積を区別できるようにする。コアに設定された地理的単位は所属領域がそれぞれの産業集積に固定されているため、コアの数と検出される産業集積の数は一致する。

b) パラメータ推定問題

式(8)のパラメータ推定問題は $\Pr(\alpha, \alpha', \lambda, \mu, \sigma)$ が無情報事前分布に従うと仮定すると、産業集積の数

$$\begin{cases} \max_{\lambda, \mu, \sigma} \Pr(f | a, \lambda, \mu, \sigma) \\ \max_{\alpha, \alpha'} \Pr(a | \alpha, \alpha') \end{cases} \quad (9)$$

と等価になる。式(9)の一階条件式は式(10)である。

$$\begin{cases} \frac{\partial \Pr(a | \beta)}{\partial \alpha} = \sum_{ij} (w_{ij} - \bar{w}) \delta(a_i = a_j \neq g) \\ - \sum_{ij} (w_{ij} - \bar{w}) \sum_{k \neq g} \Pr(z_i = z_j = k | \beta) = 0 \\ \frac{\partial \Pr(a | \beta)}{\partial \alpha'} = \sum_{ij} w_{ij} \delta(a_i = a_j = g) \\ - \sum_{ij} w_{ij} \Pr(z_i = z_j = g | \beta) = 0 \\ \frac{\partial \Pr(f | a, \theta)}{\partial \lambda} = \sum_{ij} w_{ij} \delta(a_i = a_j) (f_j - \mu) (f_i - \mu) \\ - \sum_{ij} w_{ij} \delta(a_i = a_j) \Sigma_{ij} = 0 \\ \frac{\partial \Pr(f | a, \theta)}{\partial \mu} = \lambda \sum_{ij} w_{ij} \delta(a_i, a_j) \frac{(f_i + f_j - 2\mu)}{2} \\ - \sum_i \frac{f_i}{\sigma^2} + \frac{N\mu}{\sigma^2} = 0 \\ \frac{\partial \Pr(f | a, \theta)}{\partial \sigma} = \sum_i \left\{ (f_i - \mu)^2 - \Sigma_{ii} \right\} = 0 \end{cases} \quad (10)$$

Σ_{ij} は式(4)の正規分布の分散共分散行列の第 (i, j) 成分である。 $\beta = (\alpha, \alpha')$, $\theta = (\lambda, \mu, \sigma)$ とおいた。本手法では式(10)の連立方程式を数値的に解くことでパラメータを推定する。具体的には、数値計算言語MATLABで提供されている信頼領域法⁶⁾によるアルゴリズムを利用する。パラメータ推定の手順は次の通りである。ただし、 $x = (\alpha, \alpha', \lambda, \mu, \sigma)$,

$$F(x) = \begin{cases} F_\alpha = \frac{\partial \Pr(a | \beta)}{\partial \alpha} \\ F_{\alpha'} = \frac{\partial \Pr(a | \beta)}{\partial \alpha'} \\ F_\lambda = \frac{\partial \Pr(f | a, \theta)}{\partial \lambda} \\ F_\mu = \frac{\partial \Pr(f | a, \theta)}{\partial \mu} \\ F_\sigma = \frac{\partial \Pr(f | a, \theta)}{\partial \sigma} \end{cases} \quad (11)$$

とおく。

- 1) $t=0$ とし、パラメータの初期値 x_t を与える。
- 2) 終了条件を満たすまで3)-6)を繰り返す。
- 3) $F(x_t)$ を計算する。
- 4) $\nabla F(x_t)$ を差分近似で求める。
- 5) 信頼領域法の部分問題を解き、 x_{t+1} を決定する。
- 6) $t=t+1$ とする。

$F(x)$ の計算で必要になる周辺確率は確率伝搬法で近似計算する。差分近似を前進差分で行う場合、4)で $F(x)$ をパラメータ数だけ計算する。 x の更新一轮には、6回 $F(x)$ を評価するため、確率伝搬法を6回適用する必要がある。

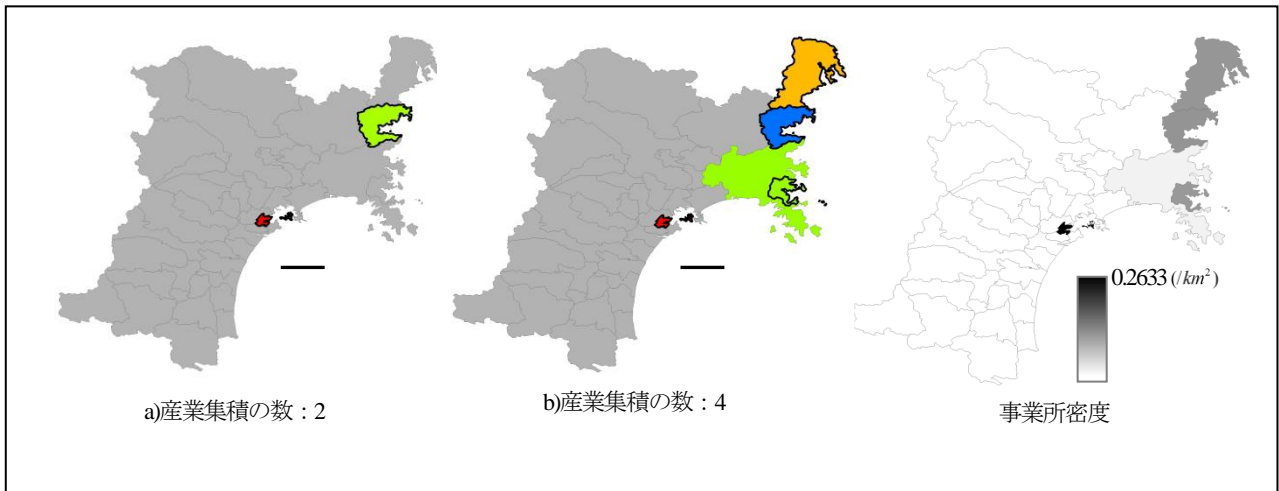


図-1 海面漁業の産業集積検出結果（分析範囲：宮城県）

c) 確率伝搬法を用いないパラメータ推定方法の検討

分析範囲を全国とする場合、確率伝搬法を1回適用するだけで2時間以上の時間を要する。本手法を実用的な手法とするには、確率伝搬法を用いずにパラメータを推定する必要がある。

式(9)のパラメータ推定問題を確率伝搬法を用いずに解く方法として、Grelaudらが提案したABCアルゴリズム⁷⁾を用いる方法が考えられる。ABCアルゴリズムではクラスタ分割が与件である場合にポッツモデルのパラメータを推定できる。式(9)のパラメータ推定問題ではクラスタ分割は与件であるため、ABCアルゴリズムが適用できると考えられる。式(10)の一階条件で周辺確率の計算が求められるのは α, α' の一階条件のみであるから、 α, α' をABCアルゴリズムで推定できれば、確率伝搬法を用いずにパラメータを推定できる。

d) 隠れマルコフモデルの解法を提案した既存手法

本研究のモデルに類似した隠れマルコフモデルからパラメータを推定する手法に、補助変数を用いるMCMC⁹⁾を用いたCucala and Marin¹⁰⁾の手法やCeleuxらが提案したEMアルゴリズムに平均場近似を組み合わせた手法¹¹⁾、変分ベイズ法を用いたMcGroryら¹²⁾の手法がある。MCMCを用いるCucala and Marinの手法は、パラメータの一階条件を計算する必要はなく、パラメータと同時にクラスタ分割を決定できる。ただし、推定するパラメータの事前分布を仮定する必要がある。平均場近似を用いるCeleuxらの手法では、クラスタ分割問題を式(12)で示した通り、地理的単位毎の問題に近似する。

$$a_i = \arg \max_{a_i} \Pr(f_i | a_i) \Pr(a_i | \tilde{a}_{-i}) \quad \forall i \quad (12)$$

\tilde{a} は平均場近似によって求めたクラスタ分割を表し、 \tilde{a}_{-i} は*i*以外の地理的単位の所属領域を表す。本研究の提案モデルをCeleuxらの手法で解くには、式(13)のように f_i 同士が独立でなければならない。

$$\Pr(f | a) = \prod_i \Pr(f_i | a_i) \quad (13)$$

しかし、本研究の提案モデルでは f_i 同士は独立ではない。本研究の提案モデルをCeleuxらの手法で解くには、式(4)の尤度の確率モデルを近似する必要が生じる。

4. ケーススタディ

(1) 使用データ

事業所数データとして、平成24年経済センサス-活動調査による市区町村単位のデータを用いる。事業所数を可住地面積で割り、事業所密度データとした。可住地面積とは、総面積から林野面積と主要湖沼面積を差し引いた面積である。

市区町村間の距離データには、市区町村の役場・役所間の直線距離を用いる。役場・役所の経緯度データからHubenyの距離計算式により距離データを作成する。市区町村の役場・役所間の距離が d_{ij} の市区町村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。

(2) 適用結果

a) 海面漁業の産業集積検出結果

海面漁業に対して複数の分析範囲で産業集積検出を行った結果を示す。図-1は宮城県を分析範囲とした結果、図-2は全国を分析範囲とした結果である。コアに設定した市区町村の輪郭線を太線にした。色が同じ市区町村は同じ産業集積に所属することを表す。コアのない灰色の領域が非集積領域と解釈できる。図中の直線は提案モデルの分散力が作用する最短の距離を示す。

図-1より次の2点が確認できる：1)事業所密度と比較して有意な産業集積が検出できる、2)設定する産業集積の数を変えても産業集積を検出できるが、産業集積の数

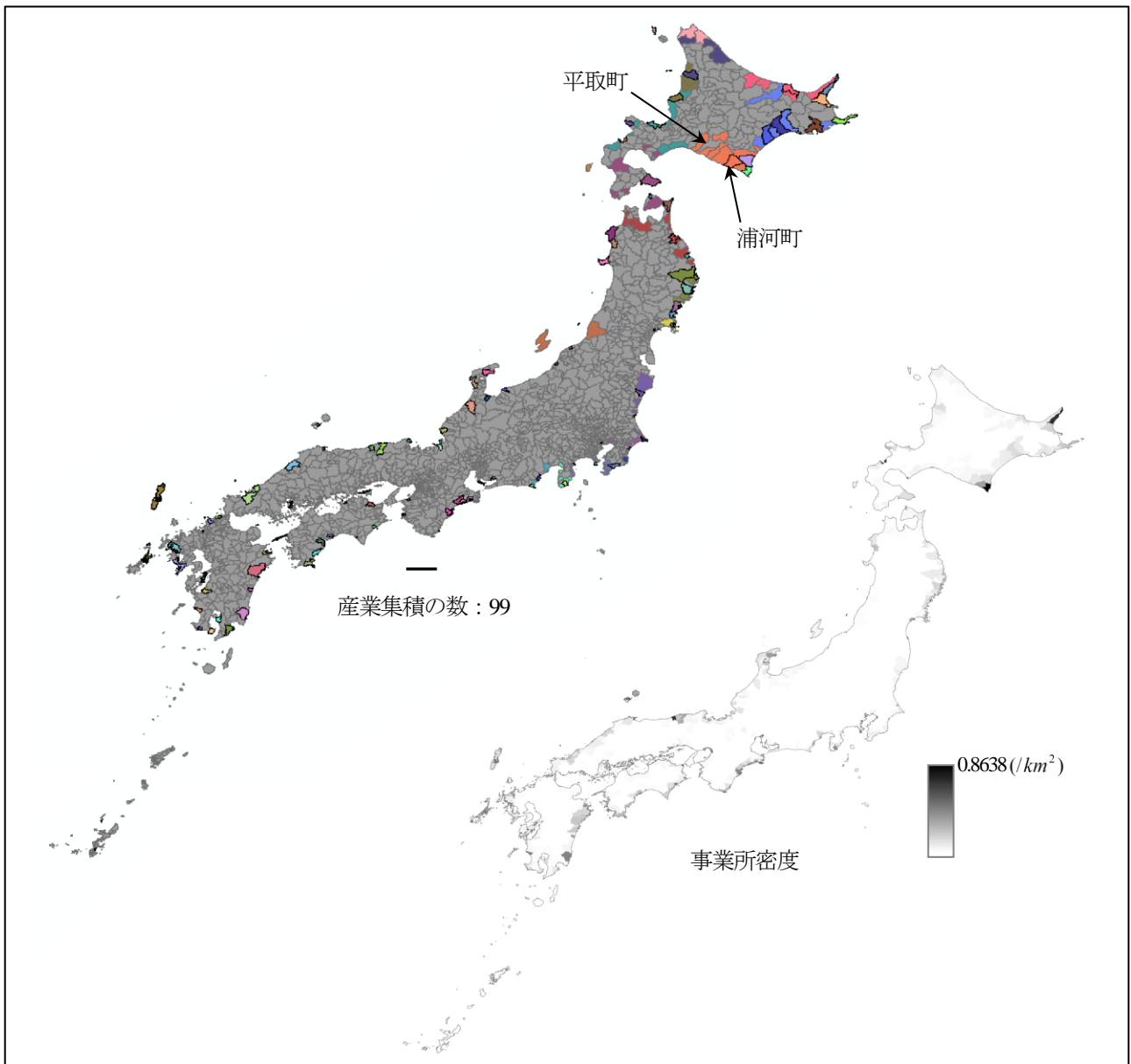


図-2 海面漁業の産業集積検出結果（分析範囲：全国）

によって産業集積検出結果が異なる。産業集積の数はクラスタ分割問題の制約条件の1つであり、設定する産業集積の数によって産業集積検出結果が変化する。産業集積の数の決定方法については、5.で解説する。

次に、図-2より次の2点に分かる：1)飛び地のある産業集積も設定され、柔軟な形状の産業集積が検出できる、2)全国を分析範囲として産業集積を検出できる。ただし、分析範囲に産業集積検出結果が依存する。

1点目の具体例として、北海道浦河郡浦河町をコアとする産業集積が挙げられる。浦河町をコアとする産業集積は、事業所密度がゼロの平取町を含んでおらず、凸性を満たさない形状である。2点目について、分析範囲を全国とした場合と宮城県とした場合の2つの場合で、宮城県に検出される産業集積を図-3に示した。図-3のa)とb)では、宮城県内に設定されたコアの数と市区町村は同

一である。しかし、宮城県を分析範囲とした図-3のa)では、石巻市は南三陸町をコアとする産業集積に所属しているのに対して、全国を分析範囲とした図-3のb)では、塩釜市をコアとする産業集積に所属している。この原因は2点考えられる。第一に、提案モデルの分散力が作用する最短の距離が分析範囲に依存することが原因として考えられる。宮城県を分析範囲とする場合は、分散力は約16kmより離れた市区町村間に作用する。全国を分析範囲とする場合は、約75kmより離れた市区町村間に作用する。塩釜市と石巻市の距離は約28kmであるため、分析範囲が宮城県の場合は塩釜市と石巻市は同じ産業集積に所属しにくい、分析範囲が全国の場合には塩釜市と石巻市は同じ産業集積に所属しやすくなる。第二の原因は、式(4)の尤度の確率モデルにあると考えられる。事業所密度が μ より高い地理的単位が事業所密度がよ

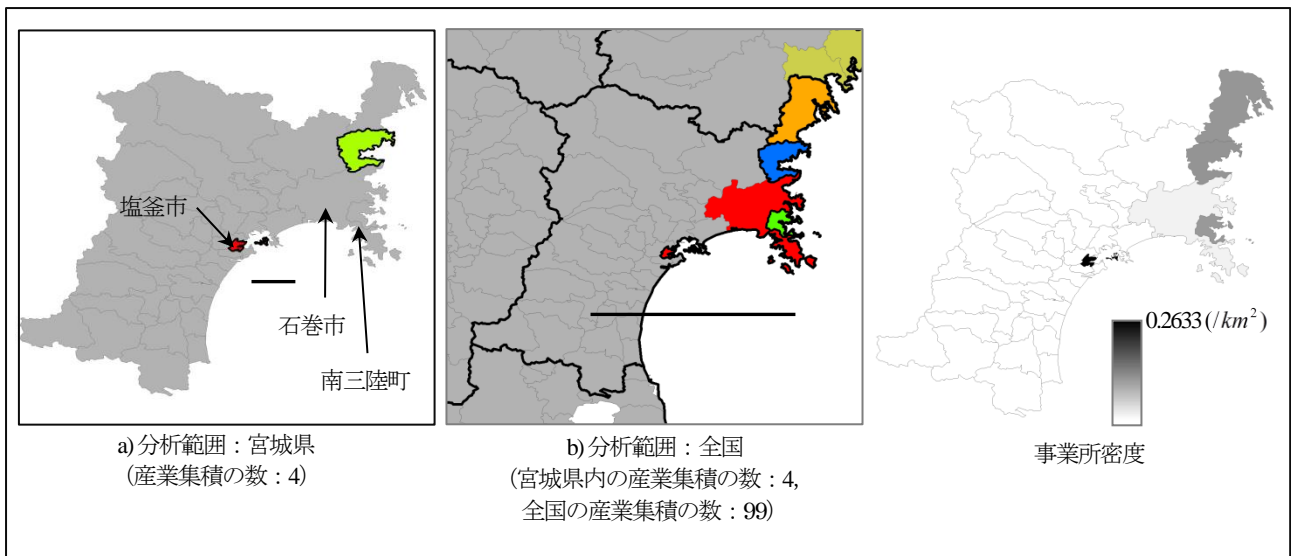


図-3 分析範囲の変化と産業集積検出結果（海面漁業）

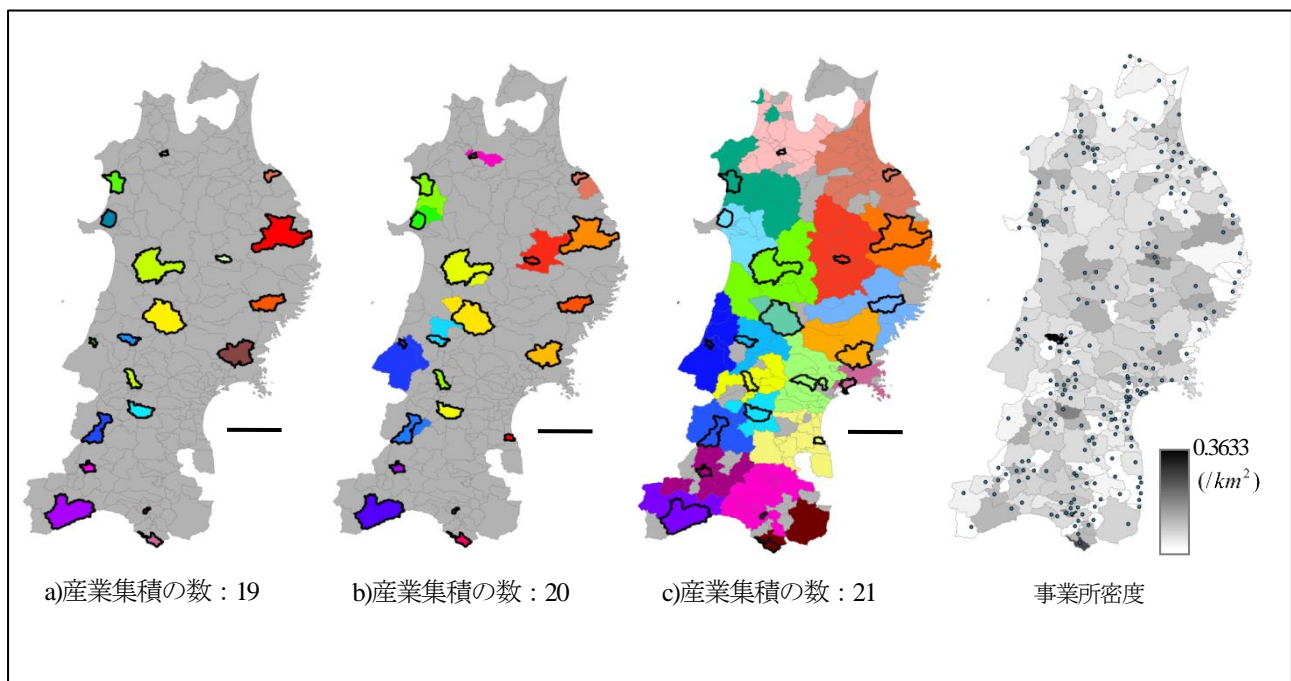


図-4 耕種農業の産業集積検出結果（分析範囲：東北6県）

り高い地理的単位と同じ産業集積に所属するほど式(4)の尤度は大きくなる。石巻市の周辺にはコアが複数あるが、塩釜市は他のコアと比べて事業所密度が高い。塩釜市と同じ産業集積に所属する方が尤度が大きくなるため、石巻市は全国を分析範囲とする場合に塩釜市をコアとする産業集積に所属する結果となったと考えられる。

b) 耕種農業の産業集積検出結果

図-4は分析範囲を東北6県とし、耕種農業について産業集積を検出した結果である。事業所密度を示した図中の点は市区町村の代表地点の位置を示す。パラメータ推定結果を表-1に示した。

表-1 耕種農業パラメータ推定結果

産業集積の数	19	20	21
α	19.94	3.819	4.032×10^{-2}
α'	3185	5.556	0
λ	28.94	1.398	1.554
μ	5.246×10^{-2}	5.243×10^{-2}	5.242×10^{-2}
σ	2.601×10^{-3}	2.601×10^{-3}	2.601×10^{-3}

図-4から次の2点が分かる：1)海面漁業と同様に柔軟な形状の産業集積を検出できる、2)産業集積の数によって産業集積に所属する市区町村の数が大きく異なる。2点目について、産業集積の数が19の場合はコアに設定し

た市区町村のみ産業集積に所属するのに対し、産業集積の数が21の場合は大半の市区町村が産業集積に所属する結果となった。この原因はパラメータ α, α' の推定値にあると考えられる。 α は市区町村の産業集積への所属のしやすさを表し、 α' は非集積領域への所属のしやすさを表す。産業集積の数が19の場合、 α は α' と比べて格段に小さく、市区町村は非集積領域に所属しやすい。そのため、コアの市区町村のみ産業集積に所属した。産業集積の数が21の場合は $\alpha'=0$ であり、 α が α' より相対的に大きいと、大半の市区町村が産業集積に所属する結果となったと考えられる。

5. おわりに

本研究では任意の形状の産業集積を検出できる手法を提案した。Q状態ポッツモデルとベイズ統計のアプローチを組み合わせた定式化を行うことで、形状に制約を置かず産業集積を検出する手法を提案した。

市区町村単位の事業所密度データを用いたケーススタディを通じて、次の2点が明らかになった：1)事業所密度と比較して有意な産業集積を検出できる、2)柔軟な形状の産業集積を検出できる。2点目に関して、飛び地のある産業集積や凸性を満たさない産業集積も検出できることが確認できた。

本研究の課題として次の3点が挙げられる：1)分析範囲に検出される産業集積の空間的規模が影響を受ける、2)モデルのパラメータ推定に日単位の時間を要する、3)産業集積の数の選択を行っていない。1点目の課題の原因は2ある。第一に、分散力が作用する距離が分析範囲に依存する定式化になっている。第二に、事業所密度が平均 μ より高い地理的単位がより事業所密度が高い地理的単位をコアとする産業集積に所属しやすい定式化になっている。1点目の課題を克服するには定式化の修正が必要である。2点目の課題について、最も時間を要しているのは、パラメータの一階条件式に現れる周辺確率の計算である。周辺確率は効率的なアルゴリズムである確率伝搬法を用いて求めているが、多数回確率伝搬法を適用する必要があるため日単位の時間がかかる。パラメータ推定を速めるために別のアプローチを検討する必要がある。産業集積の数の選択基準としては、 $BIC^{13)}$ や $ICL^{14)}$ 等の利用が考えられる。本研究のモデルの場合、 BIC は式(14)で定義され、実行が困難なクラスタ分割についての総和の計算が求められる。

$$BIC = \ln \sum_a \Pr(f, a | \hat{\lambda}, \hat{\mu}, \hat{\sigma}) - \frac{5}{2} \ln N \quad (14)$$

このような問題に対して BIC を近似した指標がForbes and Peyrardによって提案されている¹⁵⁾。Biernackiらは、混

合モデルを用いたクラスタリングにおいてクラスタ数を選択する指標として式(15)で定義される ICL を提案した。

$$ICL = \log \Pr(f | \hat{a}, \hat{\lambda}, \hat{\mu}, \hat{\sigma}) - \frac{5}{2} \log N + \log \Gamma\left(\frac{K}{2}\right) + \sum_{k=1}^K \log \Gamma\left(\hat{n}_k + \frac{1}{2}\right) - K \log \Gamma\left(\frac{1}{2}\right) - \log \Gamma\left(n + \frac{K}{2}\right) \quad (15)$$

\hat{n}_k はクラスタ k に所属する地理的単位の数である。式(4)の正規分布 $\Pr(f | a)$ の正規化定数は、分散共分散行列から計算できる。 ICL は容易に計算が可能な指標である。本研究では今後 BIC や ICL 等の指標を基に産業集積の数の選択を行う予定である。

参考文献

- 1) Mori, T. and Smith, T. E. : A probabilistic modeling approach to the detection of industrial agglomerations, *KIER Discussion Paper*, 2010.
- 2) Kulldorff, M. and Nagarwalla, N. : Spatial disease clusters: detection and inference, *Statistics in Medicine*, Vol.14, pp.799-810, 1995.
- 3) Wu, F. Y. : The potts model, *Rev. Mod. Phys.* Vol. 54, pp.235-268, 1982.
- 4) Anselin, L. : Local indicators of spatial association-LISA, *Geographical Analysis*, Vol. 27, pp.93-115, 1995.
- 5) Pearl, J. : Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, *Morgan Kaufmann Publishers Inc.*, 1988.
- 6) Conn, A. R., Gould, N. I. M. and Toint, Ph. L. : Trust Region Methods, *SIAM*, 2000.
- 7) Grelaud, A., Robert, C., Marin, J.-M., Rodolphe, F. and Taly, J.-F. : ABC likelihood-free methods for model choice in Gibbs random fields, *Bayesian Anal.* Vol. 4, No. 2, pp.317-336, 2009.
- 8) Møller, J., Pettitt, A. N., Reeves, R. and Berthelesen, K. K. : An efficient markov chain monte carlo method for distributions with intractable normalising constants, *Biometrika*, Vol. 93, Issue. 2, pp.451-458, 2006.
- 9) Murray, I., Ghahramani, Z. and MacKay, D. J. C. : MCMC for doubly-intractable distributions, *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, pp.359-366, 2006.
- 10) 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.
- 11) Celeux, G., Forbes, F. and Peyrard, N. : EM procedures using mean field-like approximations for markov model-based image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 36, Issue 1, pp.131-144, 2003.
- 12) McGrory, C. A., Titterton, D. M., Reeves, R. and Pettitt, A. N. : Variational bayes for estimating the parameters of a hidden potts model, *Stat. Comput.* Vol. 19, Issue 3, pp.329-340, 2009.

- 13) Schwarz, G. : Estimating the dimension of a model, *The Annals of Statistics*, Vol. 6, pp.461-464, 1978.
- 14) Biernacki, C., Celeux, G. and Govaert, G. : Assessing a mixture model for clustering with the integrated classification likelihood, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 22, Issue 7, pp.719-725, 2000.
- 15) Forbes, F. and Peyrard, N. : Hidden markov random field model selection criteria based on mean field-like approximations, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 25, Issue 9, pp. 1089-1101, 2003.

(2014. 4. 25 受付)

Q-STATE POTTS MODEL APPROACH TO THE DETECTION OF INDUSTRIAL AGGLOMERATION

Akihito UJIIE and Junya FUKUMOTO

Industrial agglomeration is widespread phenomenon across the countries. An industrial cluster is a region that has concentration of business institutions. In this study, we propose a new method of detecting industrial clusters based on Bayesian statistics. Its feature is that it can detect industrial clusters without any constraint on form of clusters. The result of case study shows that our approach can detect various-form industrial clusters such as cluster that has an enclave or cluster that doesn't have convexity.