

# 点事象集積検出を応用した 産業共集積形態分析の提案

井上 亮<sup>1</sup>・志賀 康平<sup>2</sup>

<sup>1</sup>正会員 東北大学准教授 大学院情報科学研究科人間社会情報科学専攻  
(〒980-8579 仙台市青葉区荒巻字青葉6-6-06)  
E-mail: rinoue@plan.civil.tohoku.ac.jp

<sup>2</sup>正会員 西日本旅客鉄道

関連性のある複数の産業が、同地域に集まって立地する産業共集積現象に関して、産業立地データを用いて共集積する産業の組み合わせを発見し、その地域を把握できる分析手法は提案されていない。本研究は、産業立地データを用いて、FDR制御法を適用した産業毎の集積地域検出、および、集積地域の共起関係に対する頻出パターンマイニングによる発見的な共集積産業の組み合わせの抽出とその組み合わせのモンテカルロ・シミュレーションによる有意性検定を通して、産業共集積を実証的に分析するアプローチを提案する。経済センサスデータを用いて提案手法の適用可能性を検証し、共集積産業の組み合わせや共集積地域の抽出が可能であることが確認された。

**Key Words :** industrial coagglomeration, cluster detection, false discovery rate, frequent pattern mining

## 1. はじめに

生産費用や輸送費の削減、高度技能労働者や中間財の集中などの誘因により、産業が地理的に集まり立地する集積現象や、関連する複数産業が集まり立地する共集積現象の発生は、世界各地で観察されている。この産業集積・共集積現象には強い関心が寄せられ、その発生メカニズムの解明に向け、地理学や空間経済学などの様々な学問分野で理論・実証の両面から研究が行われてきた。

共集積現象の分析を通して、共集積産業の組み合わせ（以下では、共集積形態と表記する）を抽出し、共集積形態の立地地域が把握できると、地域計画や企業誘致政策の立案に活かせる可能性があると期待される。

さて、産業立地データを用いた産業集積・共集積の実証研究に注目すると、Ellison・Glaeser<sup>1)</sup>やDuranton・Overman<sup>2)</sup>が提案した集積・共集積指標に基づく分析が行われている。しかし、これらの指標は、1産業の集積、あるいは、2産業の共集積の分析に焦点を当てたもので、3産業以上からなる共集積の分析には限界を有する。また、既存指標を活用して共集積形態を発見することは難しく、また、共集積地域を把握することはできない。

そこで本研究は、産業共集積形態を発見し、共集積地域を把握する新たな統計的分析手法を提案し、産業立地データへの適用を通してその有効性を確認する。

## 2. 産業集積・共集積分析の既存指標

Ellisonら<sup>3)</sup>は一産業の集積や複数産業の共集積分析を行う指標として、EG指標<sup>1)</sup>とDO指標<sup>2)</sup>を紹介している。

EG指標<sup>1)</sup>は、市町村など地域単位に集計された雇用統計を用い、産業の集積・共集積を評価する。

産業  $i$  の集積評価では、 $s_{mi}$  を産業  $i$  の全地域従業者数に対する地域  $m$  の従業者数割合、 $x_m$  を全産業の全地域従業者数に対する地域  $m$  の従業者数割合、 $H_i$  を産業  $i$  のHerfindahl指標とすると、EG指標は式(1)で定義される。

$$\gamma_i \equiv \frac{\sum_{m=1}^M (s_{mi} - x_m)^2 / \left(1 - \sum_{m=1}^M x_m^2\right) - H_i}{1 - H_i} \quad (1)$$

一産業の集積を評価するこの指標は、複数産業の共集積を評価する指標へと拡張されている。 $w_i$  を産業集団  $I$  中の産業  $i$  の従業者比率、 $s_m$  を産業集団  $I$  の地域  $m$  における従業者比率、 $\hat{\gamma}_i$  を産業  $i$  のEG指標の推定値、 $H$  を産業集団  $I$  のHerfindahl指標  $H = \sum_i w_i^2 H_i$  とすると、産業集団  $I$  の共集積を評価するEG指標は式(2)で定義される。

$$\gamma^c \equiv \frac{\sum_{m=1}^M (s_m - x_m)^2 / \left(1 - \sum_{m=1}^M x_m^2\right) - H - \sum_{i=1}^I \hat{\gamma}_i w_i^2 (1 - H_i)}{1 - \sum_{i=1}^I w_i^2} \quad (2)$$

EG指標には簡略式も提案されており、一産業の集積

を評価するEG指標は、

$$\gamma_i = \frac{\sum_{m=1}^M (s_{mi} - x_m)^2}{1 - \sum_{m=1}^M x_m^2} \quad (3)$$

二産業の共集積を評価するEG指標は

$$\gamma_{ij}^c = \frac{\sum_{m=1}^M (s_{mi} - x_m)(s_{mj} - x_m)}{1 - \sum_{m=1}^M x_m^2} \quad (4)$$

と表される。

Duranton・Overman<sup>2)</sup>は、EG指標は地域毎に集計された雇用統計を使用するため、近接するが異なる地域内に立地する事業所の組み合わせが集積・共集積の評価に反映されないという問題を指摘している。これはいわゆる可変単位地区問題(MAUP)であるが、この回避を目的に、産業の事業所地点データを使用して事業所間距離に基づいて集積性を評価するDO指標を提案している。

事業所*ij*間距離を*d<sub>ij</sub>*、総事業所数を*n*、バンド幅*h*を与えた任意のカーネル関数を*f( )*と表すと、DO指標はまず式(5)の密度関数を推定する。

$$\hat{K}_A^{Ent}(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^n f\left(\frac{d-d_{ij}}{h}\right) \quad (5)$$

モンテカルロ・シミュレーションを通して無作為分布による密度関数の信頼区間と式(5)の密度関数推定結果を比較し、集積性を評価する。また、2産業の共集積分析に拡張した分析も提案しており、式(6)で産業*i・j*の事業所分布の密度関数を推定する。

$$\hat{K}_{ij}^{c,Ent}(d) = \frac{1}{n_i n_j h} \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} f\left(\frac{d-d_{rs}}{h}\right) \quad (6)$$

なお、DO指標では、事業所位置に加えて各事業所の従業者数を考慮した分析も提案されている。

このDO指標の集積・共分析のアプローチは、点事象の分布傾向を分析する方法として1976年に提案されたRipley<sup>4)</sup>の*K*関数や、その拡張として提案されたcross-*K*関数<sup>5)</sup>と類似性がある。分析対象地域の面積を*A*、点事象密度推定値をそれぞれ $\hat{\lambda}(=n/A)$ 、 $\hat{\lambda}_i(=n_i/A)$ 、 $\hat{\lambda}_j(=n_j/A)$ 、括弧内の条件を満たせば1、満たさない場合は0を返す指示関数を*I( )*と表すと、1種類の点事象分布傾向を分析する*K*関数は式(7)、2種類(*i, j*)の点事象分布傾向を分析するcross-*K*関数は式(8)で表される。

$$\hat{K}(d) = \frac{1}{n\hat{\lambda}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n I(d_{ij} \leq d) \quad (7)$$

$$\hat{K}_{ij}(d) = \frac{1}{\hat{\lambda}_i \hat{\lambda}_j A} \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} I(d_{rs} \leq d) \quad (8)$$

DO指標はカーネル関数形状やバンド幅設定に推定結果が依存するのに対し、*K*関数は設定が必要なパラメー

タがなく、また、無作為な点分布の下での期待値が明らかか( $E[\hat{K}(d)] = \pi d^2$ )という優れた性質を持つ。しかし、

これまでのところ、*K*関数やcross-*K*関数は産業集積・共集積分析には利用されていないようである。

さて、EG・DO指標は共に産業集積・共集積分析に使用されてきたが、これらの分析には限界を有する。

第一に、両指標は、産業分布の全体傾向として集積・共集積が生じているかを分析できるが、集積・共集積している地域を明らかにすることはできない。

第二に、両指標は、3産業以上の共集積の分析に限界を有している。

まず、DO指標を拡張し、事業所間距離を基に3産業以上の共集積を分析することを考えよう。この分析には、ある産業の事業所とその他の産業の事業所との近接性を、距離に類する基準で指標化しなければならない。しかし、3産業以上の事業所間の近接性を評価するには、各事業所から最近隣の異産業事業所までの最大距離、それらを含む凸包の面積や最小包含円の半径などを用いる必要がある。これらは、2産業の共集積を分析する際に用いる異産業の事業所間距離に比べて複雑な計算が要求される。

また、そもそもDO指標は、1産業の集積を分析する場合でも計算負荷が大きい分析手法であり、多産業の共集積分析ではその欠点により際立つ。1産業の集積分析の場合、全事業所間の距離計算を行い指標値を算出した後、無作為事業所分布下での指標値分布を推定するモンテカルロシミュレーションを繰り返さなければならない。多産業の共集積分析では異産業事業所の組み合わせが指数関数的に増加するため、DO指標の拡張による分析は実行可能性が乏しい

一方、EG指標も、3産業以上の共集積分析への拡張には限界がある。EG指標では、事前に設定した産業組み合わせについて、式(2)を用いて分析する枠組みが提示されているが、共集積する産業の組み合わせを産業立地データから発見する方法については検討されていない。多くの産業から構成される共集積を分析する場合には、考え得る産業の組み合わせは膨大となるため、効率的な探索法を考案しないと実用的な計算時間で結果を得ることはできない。

以上のように、産業共集積の既往分析指標には、分析上の限界が存在する。

### 3. 産業共集積分析手法の提案

本章では、産業立地データから、産業共集積形態の発見・共集積地域の特定を行う接近法を提案する。なお、本研究では、EG指標のように、地域単位に事業所数や

従業者数が集計されたデータを用いた分析を前提とする。

提案の接近法は、下記の2段階で構成されている。

まず第1段階では、産業毎に集積地域を検出する。通常、産業集積は複数の地域で生じるため、集積地域を発見しようとする、多数の地域に関して集積の有無を検定する操作を繰り返さなければならない。しかし、同種の検定を繰り返すと、真の帰無仮説を誤って棄却する第一種の過誤の確率が増加する、多重検定問題が生じる。

この問題への対処法として False Discovery Rate (FDR) 制御法<sup>6)</sup>が提案されており、点事象の地理的集積検出への応用<sup>7,8)</sup>もなされている。本研究では分析の第1段階として、FDR制御法を用いた各産業の集積地域抽出を行い、各産業・各地域の事業所数や従業者数など産業立地データを、集積の有無を表すバイナリの情報に縮約する。

次に第2段階では、第1段階で得られた各産業の集積地域検出結果を利用して、共集積する産業の組み合わせ(以後、共集積形態と記す)を分析する。まず、産業集積が同地域に共起する状態を共集積と定義し、データマイニングを通して多くの地域で見られる共集積形態を抽出する。その後、各地域・各産業の集積発生確率を設定し、各産業が各地域に無相関に集積するとの条件の下で、共集積が生じる地域数分布を推定するモンテカルロ・シミュレーションを行い、抽出された共集積形態が観測される地域数との比較を通してその有意性を検定する。

各段階の分析手順の詳細について以下に記す。

### (1) 各産業の集積地域の検出

ここでは、単位面積当たり事業所数を基準にした集積地域検出を例に説明する。各地域・各産業の事業所数はポアソン分布に従うと仮定した上で、他の地域よりも事業所数が有意に多い地域を探す分析を行う。

分析領域内の部分領域  $Z$  の面積を  $a_Z$ 、事業所数を  $n_Z$ 、領域  $Z$  外の面積を  $a_{Z^c}$ 、事業所数を  $n_{Z^c}$ 、領域  $Z$  内外の単位面積当たり事業所数をそれぞれ  $\lambda_Z, \lambda_{Z^c}$  とすると  $n_Z \sim \text{Poisson}(a_Z \lambda_Z)$ 、 $n_{Z^c} \sim \text{Poisson}(a_{Z^c} \lambda_{Z^c})$  と表せる。

領域  $Z$  が集積地域かを検定するため、対立仮説  $H_1: \lambda_Z > \lambda_{Z^c}$ 、帰無仮説  $H_0: \lambda_Z = \lambda_{Z^c}$  を立てる。今、分析対象領域内の全事業所数の観測値  $N$  が得られたとしよう。この条件下で帰無仮説が真ならば、領域  $Z$  内の事業所数  $n_Z$  の確率分布は二項分布に従う。

$$n_Z \sim \text{Bi} \left( N, \frac{a_Z}{a_Z + a_{Z^c}} \right) \quad (9)$$

ここで領域  $Z$  の事業所数観測値が  $n$  なら、帰無仮説の  $p$  値は式(10)となる。

$$p = \sum_{i=n}^N \binom{N}{i} \left( \frac{a_Z}{a_Z + a_{Z^c}} \right)^i \left( \frac{a_{Z^c}}{a_Z + a_{Z^c}} \right)^{N-i} \quad (10)$$

表-1 産業集積検出結果の例

地域	集積産業
I	A, C,
II	B, C, D
III	A, B, C
IV	$\phi$
V	A, B, C, D
VI	A, B

本研究は FDR 制御法による検定を通して、多重検定問題を回避した集積地域検出を行う。FDR とは、棄却した帰無仮説総数に対する誤って棄却した真の帰無仮説数の割合の期待値である。FDR を事前に定めた有意水準以下に抑えて検定する FDR 制御法が提案されており、その一つの B-H 法<sup>9)</sup>の手順は以下の通りである。

1. 検定する  $m$  個の帰無仮説を  $p$  値の昇順に並べ替え、それぞれの帰無仮説の  $p$  値を  $p_{(1)}, \dots, p_{(m)}$  と表す。
2. 有意水準  $\alpha$  に対し、 $p_{(k)} \leq \alpha k/m$  となる最大の  $k$  を探す。
3.  $p_{(1)}, \dots, p_{(k)}$  に対応した帰無仮説を棄却する。

本研究では、この B-H 法を用いて、各産業・各地域の集積検出を行う。

### (2) 産業共集積形態の抽出

産業別の集積地域検出結果を基に、同地域に集積する産業の組み合わせである産業共集積形態を抽出する。

#### a) 共集積形態候補の抽出

まず、データマイニング手法の一つである頻出パターンマイニングを用いて共集積形態の候補を抽出する。頻出パターンマイニングとは、顧客の購買履歴から同時に頻りに購入される商品の組み合わせを発見する方法として Agrawal・Srikant<sup>9)</sup>が最初に提案した分析手法である。この分析では、ある組み合わせが頻出であるかは、全データ数に対する出現数の比である「支持度」で評価する。分析を行う際にはまず最小支持度と呼ばれるパラメータを設定し、データ数と最小支持度の積で与えられる閾値よりも出現数が多い組み合わせを列挙する。

ここで、頻出パターンマイニングを用いた共集積形態の候補抽出について簡単な例を用いて示す。4産業 (A, B, C, D) の集積地域検出結果を地域毎に整理し、表-1を得たとしよう。最小支持度を50%と設定し、3地域以上で共集積する産業の組み合わせを共集積形態の候補として列挙することを目指す。この例では {A}, {B}, {C}, {A, B}, {A, C}, {B, C}, {A, B, C} が抽出され、これらから他の組み合わせの部分集合ではない飽和集合を選ぶと {A, B, C} が得られる。なお本研究では、高速な頻出パターンマイニングアルゴリズムとして知られる FP-Growth アルゴリズム<sup>10)</sup>を用いる。

頻出パターンマイニングによる分析では、通常、全ての組み合わせに対して共通の閾値を設定する。しかし、共通の閾値設定に基づき分析すると、限られた地域に集積し集積地域数が閾値よりも少ない産業については、共集積産業形態を発見できない。表-1の例で3地域以上で共集積する産業の組み合わせを探すと、2地域のみで集積する産業Dについては、共集積形態が発見できない。しかし、産業Dの集積地域には必ず産業B・Cが集積しており、この形態が有意である可能性はある。

マーケティングへの応用を目指し顧客の購買履歴を分析する場合は、ある商品の組み合わせが同時に購入される回数の多寡は重要な情報の一つで、購入回数に対する閾値の設定は合理的である。しかし、本研究の分析では共集積地域数の多寡よりも、ある産業の集積地域数に占める、その産業を含む産業共集積形態の地域数の割合が重要であろう。そこで、ある産業の組み合わせを共集積形態の候補として抽出するか評価する場合は、その形態内の最も集積地域数が少ない産業の集積地域数を基準に設定した閾値で判断することとする。ただし、全データ件数に対する支持度で設定した共通の閾値を外すと、数件しか観測されない例外的な形態も抽出されてしまうため、共通の閾値を足切り用の基準として設定する。

#### b) 共集積形態候補の検定

次に、得られた共集積形態候補の有意性を検定する。本研究は、産業毎に集積地域が無作為に決まるとの仮定の下で生じる共集積の地域数分布をモンテカルロ・シミュレーションで推定した上で、産業立地データから得られた共集積地域数と比較して検定する。

ただし、シミュレーションを通じて集積地域数分布を作成する場合、サービス業をはじめとする多くの産業が人口が多い都市に立地する実態を踏まえた分析を行わないと、共集積地域数の分布を過小に推定し、結果として、形態候補の有意性を過度に評価することに繋がる。そこで、各産業の各地域における集積立地確率は、(1)で得られた集積産業数に比例すると仮定して設定した上で、分布推定・有意性検定を行う。

共集積形態候補の有意性検定手順は下記の通りである。

まず、産業毎に各地域の集積確率を設定する。その際、各地域の集積確率が地域面積および集積産業数に比例し、その和が産業の集積地域総数と一致するように調整する。ただし、確率が1を越える場合は、他の地域の集積確率を増やして調整する。

次に、産業毎・地域毎に乱数を発生させ、産業集積の有無に関する模擬データを作成し、各共集積形態候補の地域数を計数する。この過程を事前に定めた回数繰り返して、共集積形態候補の地域数分布を推定する。

最後に、共集積形態候補の地域数をシミュレーションで推定した地域数分布と比較し、事前に定めた有意水準

を基準に検定する。なお、この過程も同一の集積地域模擬データを用いた多重検定であるため、FDR制御法を用いて検定する。(1)で検定する仮説は、各地域・各産業で独立な仮説だと考えられるのに対し、(2)で検定する仮説は、共通の産業を含む共集積形態候補に関する仮説であるため、独立な仮説とは言えない。そこで、関連のある仮説群に関してFDRを基準に検定を行う方法である、Benjamini・Yekutieli<sup>1)</sup>の手法を用いて検定する。

なお、本研究では、ある共集積形態候補の有意性が棄却された場合、その形態候補を部分集合に含む形態候補も有意ではないと判断する。例えば、表-1から得られた共集積形態候補の検定によって、形態候補{A, C}が棄却されたとする。本研究では、この結果を受けて{A, C}を部分集合に含む{A, B, C}も棄却し、{A, C}、{B, C}のみを有意な共集積形態と評価する。

#### (3) 提案手法の特徴

提案手法は、産業立地データを産業集積の有無に関するデータに圧縮した上で、大量のデータから効率的に情報を抽出するデータマイニング手法を適用することにより、同地域に集積する産業の組み合わせである共集積形態を分析することを可能としている。その結果、提案手法は、多数の産業区分に基づく産業立地データから発見的に共集積形態を探索することも可能になっており、EG指標に対する優位性を備えている。また、EG・DO指標では把握できない、共集積形態が集積する地域を示すことができる点も大きな利点である。

更に、有意性検定過程で、多くの産業が都市などの一部地域に集積して立地するという全産業の立地形態を踏まえた上で、より顕著な共集積が生じているかを分析することができる。DO指標では考慮できない産業立地の偏りを踏まえた分析ができる利点を備えている。

一方で、DO指標が回避している、地域設定に分析結果が依存するMAUPには対応できない。また、EG指標では、同地域に共に多くの事業所が立地する状況に加えて、同地域に共に少ない事業所しか立地しない状況も考慮して分析されるのに対し、提案手法では、前者のみを考慮した分析となる。そのため、EG指標と提案手法が分析している共集積の定義には、多少違いがあることには注意が必要である。

## 4. 提案手法による共集積形態の分析

### (1) 使用データ

提案手法の適用可能性を検証するため、「平成21年経済センサス-基礎調査に関する地域メッシュ統計」の2分の1地域メッシュ(約500m四方)を利用し、単位面積当

たりの産業別事業所数を基準に共集積形態抽出を行う。本データは、平成19年11月改訂の日本標準産業分類の定義に基づき、農業・漁業・分類不能産業を除いた86中分類についてデータを記録している。本分析ではこの86中分類の産業分類に従い、共集積形態分析を行う。

なお、調査時点の全国の総メッシュ数は1,515,129だが、本データは1事業所以上が立地する336,646メッシュのみを記録している。本データはゼロ切断データであるため、その効果を補正して事業所密度を推定した<sup>12,13)</sup>。

## (2) 分析結果と考察

### a) 産業分類別の集積地域検出

FDR0.01以下の基準で産業分類別の集積地域を検出した。26,184メッシュで少なくとも1産業が集積しているとされ、全国の1.7%、事業所立地メッシュの7.8%に相当する。表-2に集積メッシュ数上位の10産業を示し、図-1・2に「60 その他の小売業」「24 金属製品製造業」の東京近郊の集積地域を示す。なお、図上の実線は鉄道路線を表す。その他の小売業が、都心や鉄道駅周辺に集積しているのに対し、金属製品製造業が、墨田・葛飾区などの下町や大田区、神奈川県川崎市、埼玉県川口市に集積する様子が観察され、概ね実感に即した結果が得られた。

しかし、集積地域検出結果には問題も存在する。集積メッシュ数第7位の「49 郵便業（信書便事業を含む）」は、10メッシュに2事業所、3,672メッシュに1事業所がある、極めて分散して立地する産業分類である。ゼロ切断データを考慮した事業所密度推定の結果、0.022事業所/km<sup>2</sup>、すなわち1メッシュ当たり約0.0056事業所と極めて低い密度推定値が得られたため、事業所が立地する全てのメッシュが集積と判定された。他の産業分類では著しく低い事業所密度推定結果は得られず、同様の問題は生じていないが、1地域当たりの事業所が少ない産業に関する集積地域検出に検討が必要であることが確認された。

### b) 産業共集積形態の抽出

1産業以上が集積するメッシュ数に対し最小支持度を1% (262メッシュ)、形態内の集積地域数最小の産業のメッシュ数に対し最小支持度を5%と設定し、両者を満たす共集積形態候補を抽出した。54,304形態が候補として抽出され、その中で1,865形態が飽和集合である。

次に、999回のシミュレーションから共集積地域数分布を作成し、FDR 0.01を基準に共集積形態候補の有意性を検定した。1,177形態が棄却され、棄却された形態を部分集合として含む42,601形態を除いた結果、11,703形態が有意となり、そのうち824形態が飽和集合である。

抽出された共集積形態のうち、10産業からなる構成産業数が最大の13形態を表-3に示し、最多312地域を検出した共集積形態の地域を図-3に示す。都心に加えて横浜、千葉、大宮、八王子など、JRの主要駅周辺に多く共集積

地域が分布していることが確認される。

棄却された形態のうち、メッシュ数が最多の形態は、「69 不動産賃貸業・管理業」「76 飲食店」の組み合わせで、3,086メッシュで共集積する。しかし、両産業とも集積地域数が多く、シミュレーションによる1%タイル値がメッシュ数が3,299となるため、棄却された。

38産業分類は共集積形態に含まれるが、残りの48産業分類は共集積が確認されなかった。そのうち、36産業分類は集積地域数が共通の閾値を下回っている。なお、共通の閾値を小さくすると、これらの産業分類についても共集積形態候補の抽出は可能である。例えば、1産業以上が集積するメッシュ数に対する最小支持度を本分析の半分の0.5%に設定すると、131メッシュ以上で観測される共集積形態を抽出できる。頻出データマイニングによって、標準的な性能の計算機で1分程度の計算時間で、1,610,573形態、そのうち9,417形態が飽和集合である候補が抽出できた。しかし、シミュレーションを用いた有意性検定には多大な時間を要するため、この設定に対して共集積形態の分析を行うことは難しい。

表-4に示す12産業分類は、集積メッシュ数が共通の閾値を超えているが、共集積形態が発見されなかった。そのうち3産業分類は、頻出パターンマイニングでは候補が抽出されたが、検定で棄却された。棄却された共集積形態の中には、前述の「53 郵便業（信書便事業を含む）」も含まれている。産業分類別の集積地域検出で多くの集積メッシュが検出され、頻出パターンマイニングでも他の産業分類と共集積しているメッシュが多く発見されたが、有意性検定により共集積形態を構成する産業として抽出されることを防ぐことできた。

これらの分析結果は、データマイニング手法による共集積形態候補の抽出だけでは、偶然、同じ地域に集積が生じる産業の組み合わせを除外することができないことを示しており、提案手法にモンテカルロ・シミュレーションによる検定を導入する効果が確認された。

表-2 集積メッシュ数上位10産業

産業分類 (中分類)	集積メッシュ数
60 その他の小売業	7,569
76 飲食店	7,563
78 洗濯・理容・美容・浴場業	7,431
58 飲食料点小売業	7,116
69 不動産賃貸業・管理業	5,452
83 医療業	4,423
49 郵便業 (信書便事業を含む)	3,682
7 職別工事業 (設備工事業を除く)	2,545
57 織物・衣服・身の回り品小売業	2,545
6 総合工事業	2,441

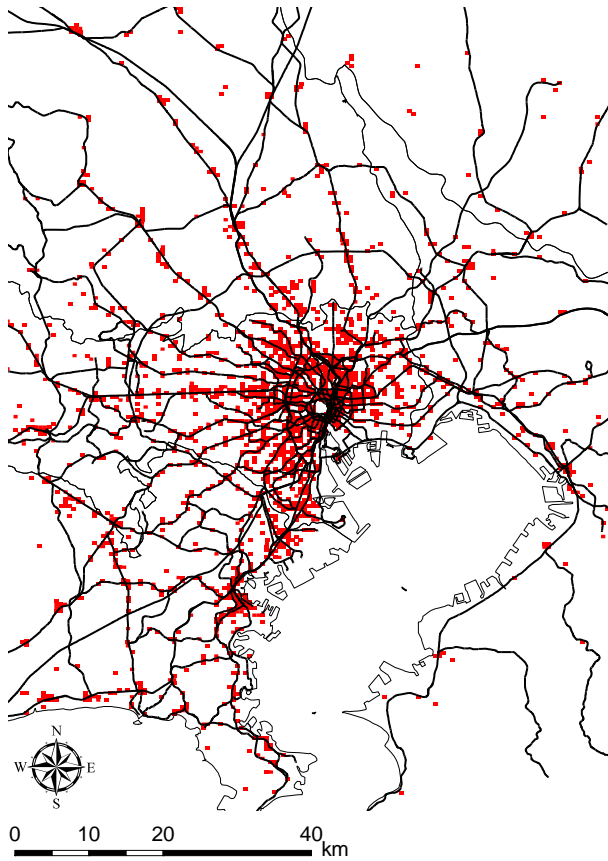


図-1 「60. その他の小売業」の集積地域

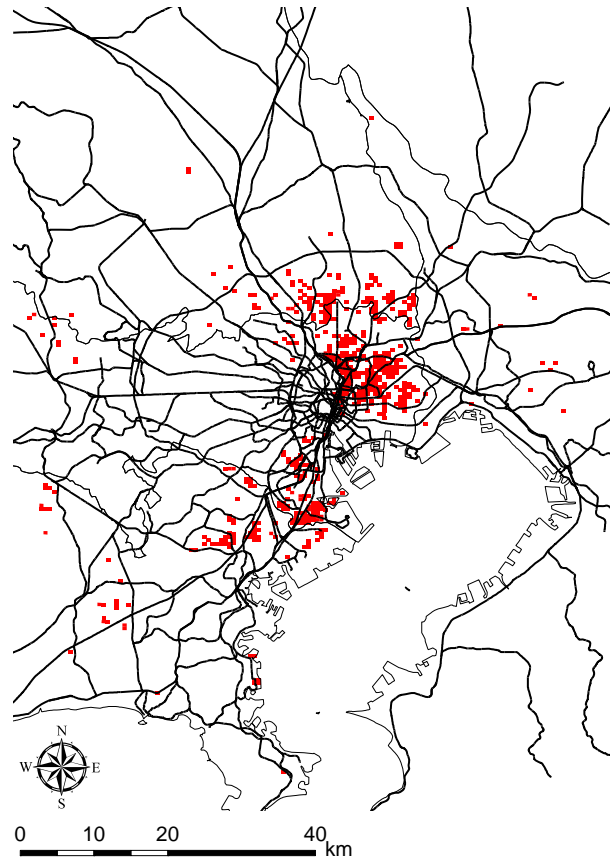


図-2 「24 金属製品製造業」の集積地域

表-3 産業数最大の共集積形態の候補

産業分類 (中分類)	共集積形態												
	1	2	3	4	5	6	7	8	9	10	11	12	13
39 情報サービス業									○				
57 織物・衣服・身の回り品小売業		○		○	○	○	○	○		○	○	○	○
58 飲食料品小売業	○	○	○	○	○	○	○	○	○	○	○	○	○
60 その他の小売業	○	○	○	○	○	○	○	○	○	○	○	○	○
68 不動産取引業	○	○	○	○		○	○	○	○	○	○		
72 専門サービス業 (他に分類されないもの)	○		○	○	○	○	○		○		○	○	○
74 技術サービス業 (他に分類されないもの)	○												
76 飲食店	○	○	○	○	○	○	○	○	○	○	○	○	○
78 洗濯・理容・美容・浴場業	○	○	○	○	○	○	○	○	○	○	○	○	○
79 その他の生活関連サービス業		○	○		○		○	○			○		○
80 娯楽業		○				○				○		○	○
82 その他の教育学習支援業	○	○	○	○	○	○	○	○	○	○		○	○
83 医療業	○	○	○	○	○	○	○	○	○	○	○	○	○
92 その他の事業サービス業	○		○	○	○			○	○	○	○	○	
共集積メッシュ数	312	301	293	282	282	279	278	227	273	269	262	261	261



図-3 共集積形態1の共集積地域

表-4 集積メッシュ数が多いが共集積形態がない産業分類

産業分類 (中分類)	集積メッシュ数	共集積形態がない理由
9 食料品製造業	502	候補なし
11 繊維工業	1,029	棄却
13 家具・装備品製造業	313	候補なし
21 窯業・土石製品製造業	392	候補なし
32 その他の製造業	411	候補なし
43 道路旅客運送業	559	候補なし
48 運輸に附帯するサービス業	274	候補なし
49 郵便業 (信書便事業を含む)	3,682	棄却
75 宿泊業	1,127	棄却
89 自動車整備業	361	候補なし
90 機械等修理業 (別掲を除く)	278	候補なし
94 宗教	1,171	候補なし

## 5. おわりに

本研究は、産業立地データに対する、FDR制御法による産業別の集積地域の検出、頻出パターンマイニングによる共集積産業の組み合わせの抽出、および、モンテカルロ・シミュレーションによる共集積の有意性の検定を

通して、産業共集積を分析する新たな手法を提案した。この提案手法を経済センサスデータを用いた分析に適用し、共集積産業の組み合わせを抽出し、その共集積地域を明示できることを確認した。

提案手法は、産業立地データを産業集積地域データへデータ縮約することと、大量のデータから関係性を抽出する分析手法として考案された頻出パターンマイニングを適用することによって、多くの産業からなる共集積を発見的に分析できるようになった。これは、既存のEG・DO指標では不可能な分析であり、提案手法の大きな利点だと言える。また、有意性検定を導入することにより、EG指標と同様、多くの産業が一部地域に集積して立地することを踏まえた共集積集積分析が可能である。

一方、地域単位の集計データを用いる分析手法では必然的に生じるMAUPへの対応ができないことが限界として挙げられる。

最後に、提案手法について今後検討すべき事項について述べる。まず、現状では、集積地域数の少ない産業と関連性を持って立地する産業を発見することは、検定過程に計算を要するため難しい。そのため、より効率的な有意性の検定手法を検討する必要がある。また、本研究の分析では、抽出された共集積形態の中から特徴的な形態のみに注目し、結果を紹介した。しかし、他にも大量の共集積形態が抽出されており、そのすべてを把握し解釈することは難しい。抽出された形態を基に、産業間の関係を可視化する方法の検討が必要であろう。これらの検討を今後の課題としたい。

**謝辞：**本研究の一部は、JSPS科研費 26289169の助成を受けた。また、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. 456)による成果であり、「平成21年経済センサス-基礎調査に関する地域メッシュ統計」データを利用した。

## 参考文献

- 1) Ellison, G. and Glaeser, E.: Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, Vol. 105, No. 5, pp 889-927, 1997.
- 2) Duranton, G. and Overman, H.: Testing for localization using microgeographic data. *Review of Economic Studies*, Vol. 72, pp 1077-1106, 2005.
- 3) Ellison, G., Glaeser, E.L. and Kerr, W.R.: What causes industry agglomeration? Evi-dence from coagglomeration patterns. *The American Economic Review*, Vol. 100, pp. 1195-1213, 2010.
- 4) Ripley, B.D.: The second-order analysis of stationary point processes. *Journal of Applied Probability*, Vol. 13, No. 2,

- pp. 255–266, 1976.
- 5) 例えば Cressie, N.A.C.: *Statistics for Spatial Data*. Wiley, 1993.
  - 6) Benjamini, Y. and Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, Vol. 57, No. 1, pp.289–300, 1995.
  - 7) Caldas de Castro, M. and Singer, B.: Controlling the false discovery rate: a new application to account for multiple and dependent tests in local statistics of spatial association. *Geographical Analysis*, Vol. 38, pp. 180–208, 2006.
  - 8) Brunson, C. and Charlton, M.: An assessment of the effectiveness of multiple hypothesis testing for geographical anomaly detection. *Environment and Planning B: Planning and Design*, Vol. 38, pp. 216–230, 2011.
  - 9) Agrawal, R. and Srikant, R.: Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Data Bases*, pp. 487–499. 1994.
  - 10) Han, J., Pei, J., and Mao, R.: Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, Vol. 8, pp. 53–87, 2004.
  - 11) Benjamini, Y. and Yekutieli, D.: The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, Vol. 29, pp. 491–507, 2001.
  - 12) Cohen, A.C., Jr.: Estimating the parameter in a conditional Poisson distribution. *Biometrics*, Vol. 16, No. 2, pp. 203–211, 1960.
  - 13) 岩崎 学: カウントデータの統計解析, 朝倉書店, 2010.

(2015. 4. 23 受付)

## ANALYSIS OF COAGGLOMERATION OF INDUSTRIES BASED ON POINT EVENT CLUSTER DETECTION

Ryo INOUE and Kohei SHIGA