

産業の共集積形態分析の試行

井上 亮・志賀 康平

A Trial for Industry Coagglomeration Pattern Analysis

Ryo INOUE and Kohei SHIGA

Abstract: The geographic concentration of industries has attracted researchers' interest, and many theories have been proposed to explain the causes of the phenomenon. However, empirical researches which analyze the coagglomeration of industries are scarce, as is noted by Ellison et al. (2010). This study proposes an analysis method which extracts the coagglomeration patterns of industries from the small-area statistics on economic activities. The proposed method is composed of two procedures: the spatial cluster detection of each industry by false discovery controlling method and the frequent pattern mining of the combination of detected industrial clusters. The proposed method is applied to the 500-meter grid data of the 2012 Economic Census for Business Activity in Japan to check the applicability and validity of the proposed analysis method.

Keywords: 共集積 (coagglomeration), 集積検出 (cluster detection), FDR (false discovery rate), 頻出パターンマイニング (frequent pattern mining)

1. はじめに

関連性の強い複数の産業が地理的に集積して立地する現象は、地理学や経済学など様々な分野で永らく注目されており、そのメカニズムを説明する多くの理論が地理学や空間経済学において提案されている。しかし、Ellison *et al.* (2010) が指摘するように、産業の共集積形態を実証的に分析する研究はほとんど行われていない。

本研究は「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) を用いて、産業の共集積形態を小地域単位で分析することを試みる。

本研究ではまず、False Discovery Rate (FDR) に

基づく統計的検定を通して産業別の集積メッシュを検出した上で、得られた産業集積の共起関係に対して頻出パターンマイニングを適用し、複数産業の共集積形態の抽出を試みる。

なお本研究では、あるメッシュの産業集積を、「基準① 単位面積当たりの産業別事業所数」または「基準② 単位面積当たりの産業別従業者数」で判定する。

2. では、各産業の集積地域検出方法である FDR-controlling 法を、3. では産業の集積形態の抽出に用いる頻出パターンマイニングを概説し、4. で経済センサスデータへの適用結果を示す。

2. FDR-controlling 法による産業別集積地域の検出

集積検出に当たり、事業所や従業者の空間分布を点事象分布と考え、単位面積当たりの産業別事

井上 亮 〒980-8577 仙台市青葉区片平 2-1-1

東北大学 大学院情報科学研究科

Phone: 022-217-6368

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

業所数や産業別従業者数がポアソン分布に従うと仮定して分析する。

さて、点事象の集積検出方法として空間スキャン統計(例えばKulldorff (1997))が注目されており、多くの派生手法が提案されている(例えばDuczmal et al. 2006)。空間スキャン統計は、点事象が従う確率分布を仮定した上で、ある地域の点集積を尤度比で評価する。分析対象領域を走査して尤度比最大となる集積地域の候補を検索し、その候補の集積性をモンテカルロシミュレーションで検定する。しかし「仮説検定を複数回繰り返すと第一種の過誤が生じる確率が増加する」多重検定問題への対応から、空間スキャン統計は同時に一つの集積地域しか検出できない。Mori and Smith (2009)は、空間スキャン統計の枠組みを拡張し、複数集積地域を同時に検出できる手法を提案したが、集積候補地域群の検索およびその検定に多大な時間を要する場合があることが報告されており、実用性に課題を抱えている。

一方、Brunsdon and Charlton (2011)は、多重検定問題への接近法であるFDR controlling法の一つであるB-H法(Benjamini and Hochberg, 1995)を点事象集積検出に応用することを提案している。FDR-Controlling法は、「棄却された帰無仮説群に含まれる、第一種の過誤を有する帰無仮説の割合の期待値(FDR)」を基準に検定する。多重検定問題への伝統的対応法であるBonferroni法などでは、棄却された帰無仮説群の中に第一種の過誤を有する帰無仮説が一つでも含まれる割合(Familywise error rate)が、有意水準以下になるよう検定を行うため、多数の仮説を検定する場合、過度に保守的な検定となる。一方、FDR-Controlling法は、棄却された帰無仮説の中に第一種の過誤を有するものが一定の割合以下含まれることを許容するため、Familywise error rateを基準にする検定法よりも検出力が高い。

Brunsdon and Charlton (2011)は、英国センサスの小地域 (LSOA: 人口1,000~3,000人の地域)単位

のデータを用いて家庭内介護を行っている人口の割合の高い地域を抽出する適用と、シミュレーションデータを用いた実験を通して、FDR-Controlling法による集積検出の有効性を確認している。この方法は、分析対象の空間単位が小さくなると、検出力が低下することが懸念されるが、本研究ではBrunsdon and Charlton (2011)と同様、B-H法による集積検出方法を適用する。

ここで、基準①による分析を例に手順を説明する。メッシュ i の事業所数を n_i 、面積を a_i 、全国の事業所数を N 、面積を A とする。今、メッシュ i の単位面積当たりの産業別事業所数の推定値は $\hat{\theta}_i = n_i/a_i$ 、全国の単位面積当たりの産業別事業所数の推定値は $\hat{\theta} = N/A$ である。点事象集積検出を行う際に検証すべき対立仮説は $\hat{\theta}_i > \hat{\theta}$ 、帰無仮説は $\hat{\theta}_i = \hat{\theta}$ となる。ポアソン分布の母数の推定値の差を検定する場合、二項分布が用いられる(岩崎, 2010)。すなわち、帰無仮説のもとでメッシュ i の事業所数分布が得られる確率 p_i は、次式で表せる。

$$p_i = {}_N C_{n_i} (a_i/A)^{n_i} \{(A-a_i)/A\}^{N-n_i} \quad (1)$$

さて、B-H法は、無作為に分布する真の帰無仮説の p 値は0~1の間で一様分布するとの原理を基に、FDRに基づく検定手順を提案している。 m 個の帰無仮説を p 値の昇順 ($H(1), \dots, H(m)$) に並べた上で、事前に定めた有意水準 α に対して

$$P_{(k)} < \frac{k}{m} \alpha \quad (5)$$

を満たす最大の k を求め、 k 番目までの帰無仮説を棄却すれば、FDRを α 以下に抑えられることを示している。本研究では、式(1)で得られた各メッシュの p 値に対してB-H法を適用し、各産業の集積の有無を検定する。

ただし、経済センサスデータは、事業所が一つ以上存在するメッシュだけを記録するため、ゼロ切断データである。一方、全国土を覆うメッシュを考え、事業所数0のメッシュを大量に含むデータを作成すると、山間部や湖沼など事業所が立地

できない場所を含むため、ゼロ過剰データとなる。本研究では、ゼロ切断データとして扱い分析する。

3. 頻出パターンマイニングによる産業共集積形態の抽出

集積検出の結果から得られた各メッシュにおける集積産業の組み合わせに対して、頻出パターンマイニングを行い、同じメッシュに集積して立地する産業の組み合わせを抽出する。なお、本研究では、FPGrowth 法 (Han *et al.*, 2004)を用いる。

4. 産業共集積の分析

日本標準産業分類の中分類に従って記録された事業所数・従業者数データを用いて産業集積地域の検出・共集積形態の分析を行う。なお、平成 21 年経済センサスの 2 分の 1 地域メッシュは全 336,646 メッシュ、表-1 に示す 87 中分類に関するデータが記録されている。

4.1 産業集積地域の検出

各産業中分類に対し、FDR1%以下で集積地域検出を行った。一例として、基準①により最多の 6,746 メッシュが集積地域として検出された、[M76 飲食店]の集積地域検出結果について、東京近郊の拡大図を図-1 に示す。鉄道路線に沿い、集積が検出されていることが分かる。なお、基準②

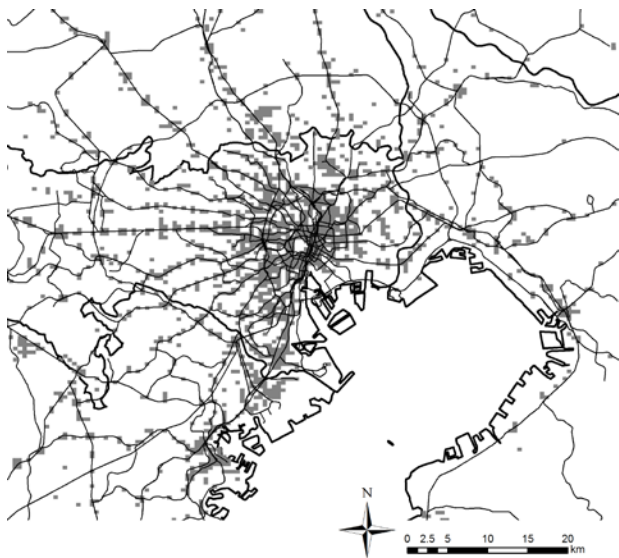


図-1 単位面積あたり事業所数に基づく飲食店集積

表-1 平成 21 年経済センサス 基礎調査

地域メッシュ統計の産業中分類の数

大分類*1	中分類
C 鉱業, 採石業, 砂利採取業	1
D 建設業	3
E 製造業	24
F 電気・ガス・熱供給・水道業	4
G 情報通信業	5
H 運輸業, 郵便業	8
I 卸売業, 小売業*2	11
J 金融業, 保険業*3	2
K 不動産業, 物品賃貸業	3
L 学術研究, 専門・技術サービス業	4
M 宿泊業, 飲食サービス業	3
N 生活関連サービス業, 娯楽業	3
O 教育, 学習支援業	2
P 医療, 福祉	3
Q 複合サービス事業	2
R サービス業(他に分類されないもの)*4	8
S 公務(他に分類されるものを除く)	1

*1 大分類[A農業, 林業][B漁業][T分類不能の産業]は対象外

*2 [I61無店舗小売業]は対象外

*3 [J64貸金業, クレジットカード業等非預金信用機関][J65金融商品取引業, 商品先物取引業][66 補助的金融業等][J67保険業(保険媒介代理業, 保険サービス業を含む)]は対象外

*4 [R95その他のサービス業]は対象外

[M76 飲食店]では 13,080 メッシュが抽出された。また、基準②で最多のメッシュが集積地域として検出された産業は、[I58 飲食料品小売業]で 18,497 メッシュである。

4.2 産業共集積形態の抽出

頻出パターンマイニングを用いて、産業共集積形態の抽出を行った。なお、最小サポートは1%と設定し、一産業以上の集積が検出されたメッシ

ユ(基準①: 62,327メッシュ, 基準②: 268,229メッシュ)の1%以上で生起した産業の共集積形態を抽出している。

基準①で検出された最も産業数が多い組み合わせは、11産業から構成されている。その中で地域数が最も多い形態は、[I57織物・衣服・身の回り品小売業][I58飲食料品小売業][I60その他の小売業][K68不動産取引業][K69不動産賃貸業・管理業][M76飲食店][N78洗濯・理容・美容・浴場業][N79その他の生活関連サービス業][N80娯楽業][O82その他の教育, 学習支援業][P83医療業]で構成、211メッシュで各産業が集積している。これらは小売業中心・繁華街に立地する産業の組み合わせである。これらと産業構成が異なる組み合わせの例は、6産業・214メッシュで構成される[G39情報サービス業][I53建築材料, 鉱物・金属材料等卸売業][I54機械器具卸売業][I55その他の卸売業][R92その他の事業サービス業][R93政治・経済・文化団体]や、3産業・304メッシュの[D06総合工事業][D07職別工事業(設備工事業を除く)][D08設備工事業]などが検出された。

一方、基準②では、最大8産業から構成される産業共集積が抽出され、地域数最多(1025メッシュ)の組み合わせは[I58飲食料品小売業][I60その他の小売業][K68不動産取引業][K69不動産賃貸業・管理業][M76飲食店][N78洗濯・理容・美容・浴場業][O82その他の教育, 学習支援業][P83医療業]で、同じく小売業中心・繁華街型の共集積が検出された。その他には、建設業関連・卸売業関連など複数の形態の集積地域が抽出された。

5. おわりに

本研究では、経済センサスの小地域集計データから、産業別集積地域検出と集積地域組み合わせの分析を通じた産業共集積形態の分析の試行を行った。提案したアプローチにより様々な共集積を抽出できることは確認できたが、まだその意味解釈を行うに至っていない。結果の解釈を通して、

この分析の有効性を評価することは今後の課題としたい。

なお、本研究では、ある産業の従業者数比率が全国平均より有意に高い地域の検出も行い、特定の産業に特化した地域を抽出した。FDRによる検出は分析の可能性を示したが、集積検出の定義から、複数産業が同地域に集積することが起こりにくいため、共集積検出には適していなかった分析とはならなかった。これら異なる集積定義による共集積の発見も検討課題である。

謝辞

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 年経済センサス地域メッシュ統計を利用した。

参考文献

- 岩崎学(2010):「カウントデータの統計解析」, 朝倉書店.
- Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, **57**(1), 289–300.
- Brunsdon, C. and Charlton, M., 2011. An assessment of the effectiveness of multiple hypothesis testing for geographical anomaly detection. *Environment and Planning B: Planning and Design*, **38**, 216–230.
- Duczmal, L., Kulldorff, M., and Huang, L., 2006. Evaluation of spatial scan statistics for irregularly shaped clusters. *Journal of Computational and Graphical Statistics*, **15**(2), 1–15.
- Ellison, G., Glaeser, E. L., and Kerr, W. R., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, **100**, 1195–1213.
- Han, J., Pei, J., and Mao, R. 2004. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, **8**, 53–87.
- Mori, T., Smith, T. E., 2009. *A probabilistic modeling approach to the detection of industrial agglomerations*. Discussion paper No. 682, Institute of Economic Research, Kyoto University.