

点事象データから地域を観る — 犯罪発生から産業の集積まで —

東北大学 情報科学研究科
井上 亮

Web公開に際し、一部の図を削除・改変しています

詳細な位置・時刻を記録したデータの入手可能性拡大 ³

▶統計情報の公開・活用の推進

[空間] 市区町村単位・町丁目単位…
1kmメッシュ・500mメッシュ…

[時間] 5年間隔・1年間隔・四半期…

▶行政情報の公開・活用の推進

[空間] 座標

[時間] 年単位・月単位・日単位・時間単位・時刻

「地域の今」を

詳細な空間・時間解像度で分析できる可能性

本日は【点事象データに基づく地域分析】を紹介

詳細な位置・時刻を記録したデータの入手可能性拡大 ²

▶統計情報の公開・活用の推進

統計情報の有効活用を目指す施策の進展

新統計法(平成19年改正)・e-Stat・

公的統計の二次的利用制度(オーダーメイド集計)...

▶行政情報の公開・活用の推進

「オープンデータ・オープンガバメント」施策の進展

例 <https://www.data.gov/>

<https://data.cityofnewyork.us/>

点事象データとは? ⁴

事象(イベント)の発生位置や時刻を記録したデータ

(例) 事件・事故の発生地点・時刻

伝染性疾病の患者の居住地・発症日

店舗の出店・退去した地点・年月日

不動産取引が行われた物件・取引日

SNSにコメントや写真を掲載した地点・時刻

ある時刻の人の滞在場所 ⇒ 昼間人口・夜間人口など

ある日の事業所や店舗の立地場所 ⇒ 産業立地

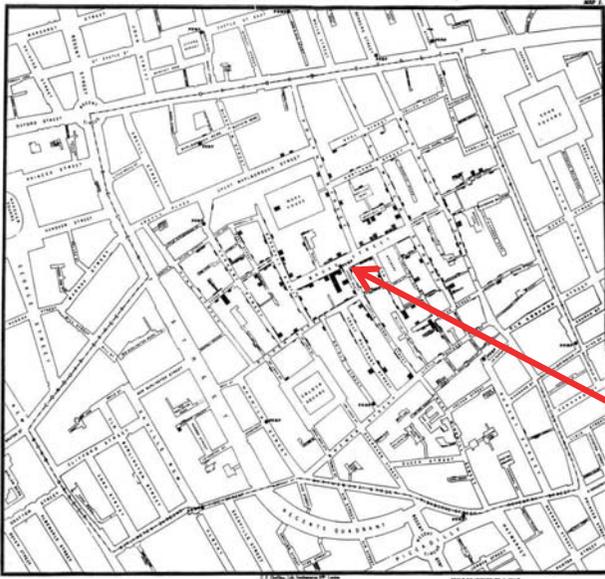
点事象データの分析

◆ 点分布パターン分析 (集積←ランダム→分散)

◆ 点事象が集積する空間領域・時期の検出

「世界最古の空間解析」伝染病罹患者の集積検出 5

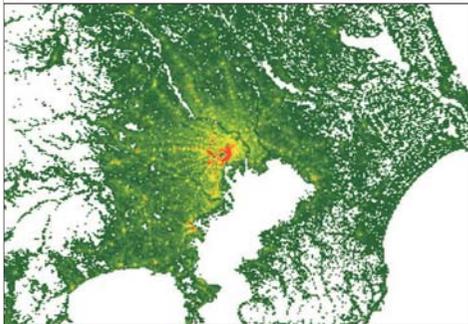
Dr. John Snowによる“1854年 ロンドンでのコレラ*流行の集積発見”
*) コッホによるコレラ菌の発見は1884年



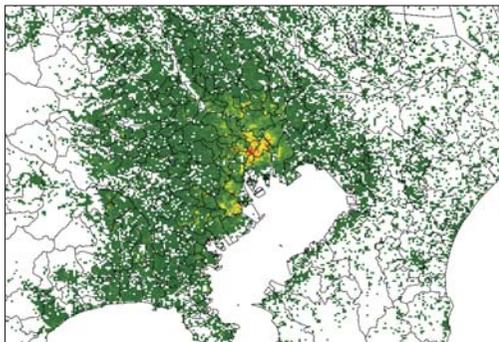
図出典
Wikipedia

産業立地の集積検出 7

⇒ 産業立地の実態把握 産業誘致施策の検討?



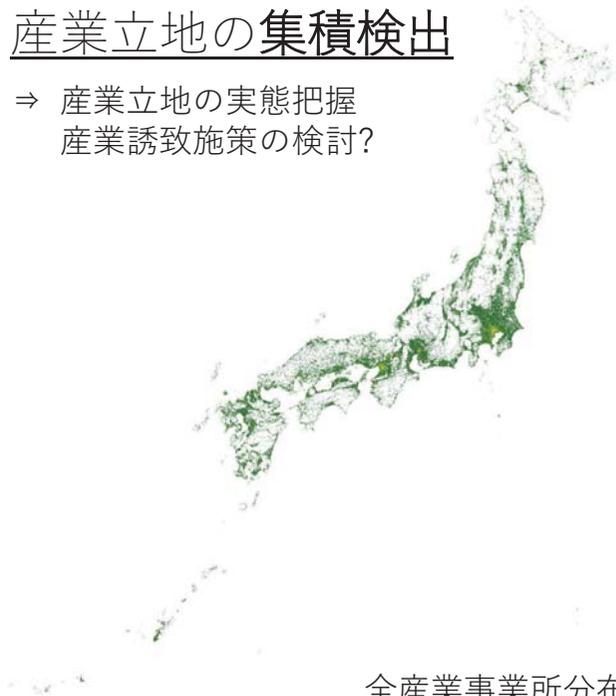
全産業事業所分布



金属製品製造業

産業立地の集積検出 6

⇒ 産業立地の実態把握
産業誘致施策の検討?



全産業事業所分布

平成21年経済センサス基礎調査 地域メッシュ統計 2分の1地域メッシュ

「点事象データから地域を観る」 8

「点事象の集積検出を通して、地域の特徴把握を目指す」

近年、開発してきた分析手法について、紹介する。

本日のメニュー

1. 点事象集積検出の基礎
2. ネットワーク・時空間上の点事象集積手法の提案とその応用 (飲食店集積と犯罪集積)
3. 複数点事象の集積パターンの分析手法の提案とその応用 (産業共集積現象)

1. 点事象集積検出の基礎

点事象集積検出の基礎

点事象の分析手法

➤ **分布パターンの識別** 代表的手法: K関数法

集積 ← 無作為分布 → 分散

➤ **集積領域の検出** 代表的手法: 空間スキャン統計
集積パターンを示す点分布に対して
その**集積**している(時)空間領域を検出する分析手法
=「点分布密度が、周囲よりも高い」

集積領域の検出

【前提となる仮定】

分析対象領域 G では、ポアソン点過程に従い点が分布

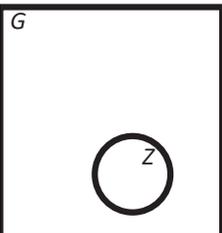
【部分領域 Z が集積か否か】

領域 Z 内の点過程の母数(単位面積当たり平均個数・点密度) λ_Z が
領域 Z 外 Z^c の点過程の母数 λ_{Z^c} より大きいかな？を判定

対立仮説 $\lambda_Z > \lambda_{Z^c}$

帰無仮説 $\lambda_Z = \lambda_{Z^c}$ を立て仮説検定すればよい？

Geographical Analysis Machine by Openshaw (1987)



複数の部分領域に対して検定すると
「多重検定」問題が発生！

多重検定(比較)問題とは？

統計的検定を複数回繰り返すと、
偽陽性 (有意ではない対立仮説を誤って採択) の
可能性が増加する問題

有意水準5%の検定は、**帰無仮説の下で**、
観測された値が生じる**確率が5%以下**なら

「**帰無仮説を棄却、対立仮説を採択**」

⇒ 真の状態が「**帰無仮説** ○・**対立仮説** ×」の場合にも
偽陽性(第一種の過誤)の検定結果を返す可能性が5%はある

独立な検定を、有意水準5%で100回繰り返すと...

各検定が偽陽性の結果を返す可能性が5%あるので
すべての検定結果が正しい可能性は

$(1 - 0.05)^{100} = 0.00592$ 0.6%以下しかない！

対処法1) 空間スキャン統計 (Kulldorff, 1997)

13

「**多重検定問題**」のため、**多数**の部分領域 Z に対して集積の有無を検定する操作を繰り返してはいけない!!

そこで、**最も集積性の強い部分領域 Z_{MLC}** を探した後にその領域に関してのみ検定を行う方法が提案

対立仮説と帰無仮説の尤度比 $LR(Z)$

$$LR(Z) = \begin{cases} \frac{\left(\frac{\#(Z)}{\text{size}(Z)}\right)^{\#(Z)} \left(\frac{\#(Z^c)}{\text{size}(Z^c)}\right)^{\#(Z^c)}}{\left(\frac{\#(G)}{\text{size}(G)}\right)^{\#(G)}} & \text{if } \frac{\#(Z)}{\text{size}(Z)} > \frac{\#(Z^c)}{\text{size}(Z^c)} \\ 1 & \text{otherwise} \end{cases}$$

を基準に、**尤度比最大**となる Z を探索 $\Rightarrow Z_{MLC}$

Z_{MLC} の有意性を、 G 内に無作為に発生させた点分布から得られる最大尤度比の分布を用いて評価。

\Rightarrow **一領域の集積のみを検定**することで**多重検定問題を回避!**

(2. ネットワーク上・時空間上の集積検出に利用)

対処法2) False Discovery Rate (FDR) 制御法の応用¹⁴

de Castro & Singer (2006), Brunson & Charlton (2011)

FDR (偽陽性率)とは?

m 本の仮説を検定した時に「有意と判定した仮説数 R 」に占める「偽陽性の仮説数 V 」の割合の期待値。

$$FDR = E\left(\frac{V}{R}\right) \quad (\text{ただし } FDR=0 \text{ if } R=0)$$

	帰無仮説を棄却 (対立仮説を有意と判定)	帰無仮説を保留 (対立仮説を有意ではないと判定)	計
帰無仮説真 <small>偽陽性 (第1種の過誤)</small>	V	U	m_0
対立仮説真	S	T <small>偽陰性 (第2種の過誤)</small>	$m - m_0$
計	R	$m - R$	m

FDR制御法

FDRを事前に設定した一定水準 α 以下に抑えて仮説検定を行う方法。

多数の検定を行うと偽陽性の発生数 V を増えるが、**一定程度、偽陽性の検定結果が含まれることを許容**するアプローチ。
[従来の多重検定への対処法 (FWER)では、 $Prob(V < 1) < \alpha$ (有意水準)]

対処法2) False Discovery Rate (FDR) 制御法の応用¹⁵

FDR制御法 (Benjamini & Hochberg, 1995)

1. 検定する m 個の仮説 (m 地域の集積の有無)について p 値を計算する。
2. p 値の昇順に並べ替え, $p_{(1)}, \dots, p_{(m)}$ と表す。
3. 有意水準 α に対して, $p_{(k)} \leq \frac{k}{m} \alpha$ となる最大の k を探す
4. $p_{(1)}, \dots, p_{(k)}$ に対応した帰無仮説を棄却
 $\Rightarrow k$ 本の仮説を採択
 (k 地域で集積があると判定)

(3. 産業集積の分析に利用)

16

2. ネットワーク上・時空間上の点事象集積手法の提案とその応用

— 空間スキャン統計に基づく分析 —

空間スキャン統計に基づく集積検出 17

地域単位の点数集計データから、集積地域を検出。
分析の最小単位は**地域**

1. 「単純な形状」

任意の地点から一定距離以内の地域を集積領域の候補と設定

2. 「複雑な形状」

隣接地域の集合を集積領域の候補と設定

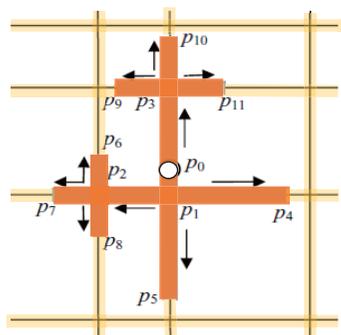
空間上の集積領域検出 18

図出典 : Duczmal et al. (2006)

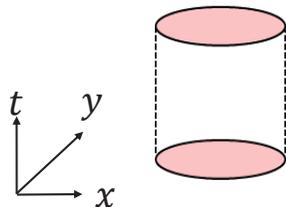
単純な形状 Kulldorff (1997)など	柔軟な形状 Duczmal et al. (2007)など
<p>【特徴】</p> <ul style="list-style-type: none"> ある地点から一定距離内の領域に点集積する現象の分析を想定 円形などのコンパクトな領域を検出 <p>【課題】</p> <ul style="list-style-type: none"> 点分布が少ない地域も含んだ集積を検出する可能性 	<p>【特徴】</p> <ul style="list-style-type: none"> 隣接地域を結合した領域を検出 <p>【課題】</p> <ul style="list-style-type: none"> 点分布が少ない地域を含む広範囲を覆う領域を検出する可能性 複雑な形状の領域を検出し、結果の解釈が困難な場合あり
<p>形状複雑度を制御 Duczmal et al.(2006)など</p> <ul style="list-style-type: none"> 提案した領域の形状複雑度指標を基に、検出領域の形状を調整 単純・柔軟な形状の集積領域検出手法を折衷した分析が可能 	

ネットワーク上・時空間上の単純な形状の集積検出 19

【ネットワーク】



【時空間】



任意地点から一定のネットワーク距離以内の領域

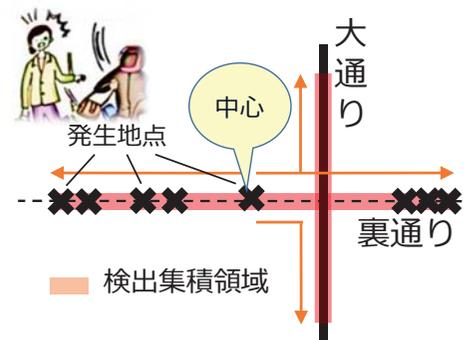
Shiode (2011), Shiode & Shiode (2012)

同じ空間領域で一定期間継続する点事象の集積地域・期間
Kulldorff (2001)など

ネットワーク上・時空間上の単純な形状の集積検出 20

【ネットワーク】

(例) ひったくりが裏通りだけで発生する現象だったら

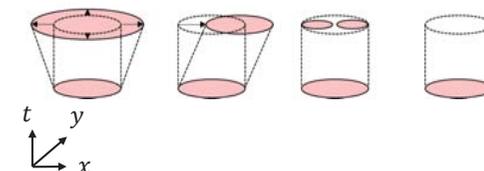


多発地点から近いだけで大通りも集積として検出

【時空間】

時間経過につれて集積空間領域が変化する事象

(例) 空気感染する病気 防犯対策実施前後の犯罪発生



空間領域が変化しない集積検出では対応できない

ネットワーク/時空間上で柔軟な形状の集積領域検出の必要性

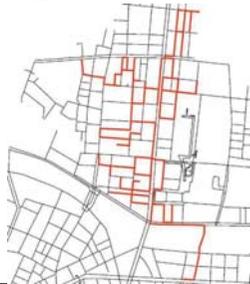
【アプローチ】

1. ネットワーク上/時空間上の分析対象領域を離散的な領域に分割（空間分析の「地域」に相当）し、隣接領域を結合して集積領域の候補を生成
2. ネットワーク上・時空間上の領域形状の複雑度を指標化
3. 形状複雑度指標に関する制約下で、集積領域を探索

【ネットワーク】

リンクを分析の最小単位に設定

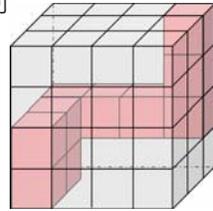
接続リンクを繋ぎ、集積領域の候補とする



【時空間】

離散的な領域「ボクセル」に分割

隣接ボクセルを繋ぎ、集積領域の候補とする



形状指標

【ネットワーク】

$$K_n(Z) = \frac{\max_{n \in N(\text{Compact}(Z))} nd(n, \text{Center}(\text{Compact}(Z)))}{\max_{m \in N(Z)} nd(m, \text{Center}(Z))}$$

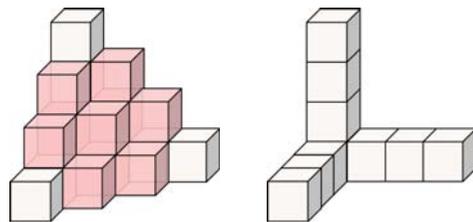


$K_n(Z) = 1$

$K_n(Z) = 0.30$

【時空間】

$$K(Z) = \frac{(\text{領域}Z\text{内のボクセル数})}{(\text{領域}Z\text{の凸包内のボクセル数})}$$



$K(Z) = 1$

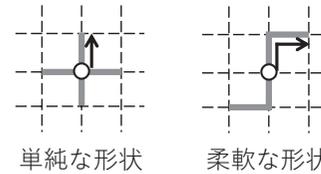
$K(Z) = 0.5$

【アプローチ】

1. ネットワーク上/時空間上の分析対象領域を離散的な領域に分割（空間分析の「地域」に相当）し、隣接領域を結合して集積領域の候補を生成
2. ネットワーク上・時空間上の領域形状の複雑度を指標化
3. 形状複雑度指標に関する制約下で、集積領域を探索

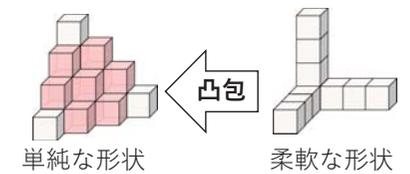
形状複雑度指標: 単純(コンパクト)な形状との対比

【ネットワーク】



「中心点から最遠点までの距離の比」

【時空間】



「凸包との体積の比」

【アプローチ】

1. ネットワーク上/時空間上の分析対象領域を離散的な領域に分割（空間分析の「地域」に相当）し、隣接領域を結合して集積領域の候補を生成
2. ネットワーク上・時空間上の領域形状の複雑度を指標化
3. 形状複雑度指標に関する制約下で、集積領域を探索

形状複雑度指標の制約下で尤度比を最大化するリンク/ボクセルの組合せを求める組合せ最適化問題として定式化.

遺伝的アルゴリズムを用いて尤度比最大の集積候補領域を探索

ネットワーク上の集積領域検出 適用例 25

<分析対象地域>

仙台市青葉区東部(東北自動車道の東側)

<点事象>

居酒屋立地点 492店舗

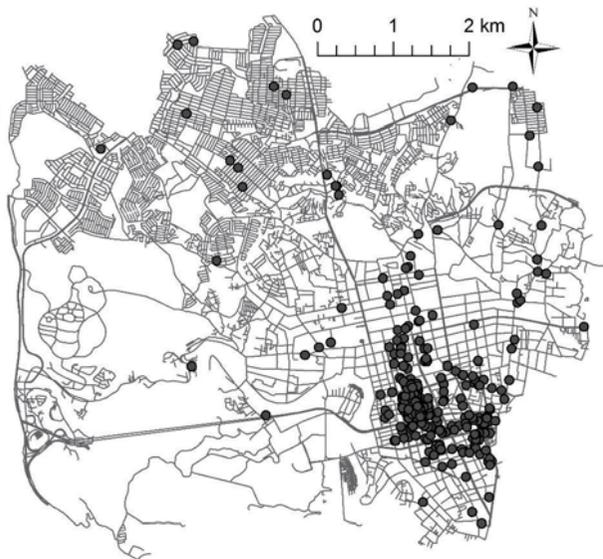
座標付き電話帳
「テレポイント®Pack!」
2010年9月版

<道路ネットワーク>

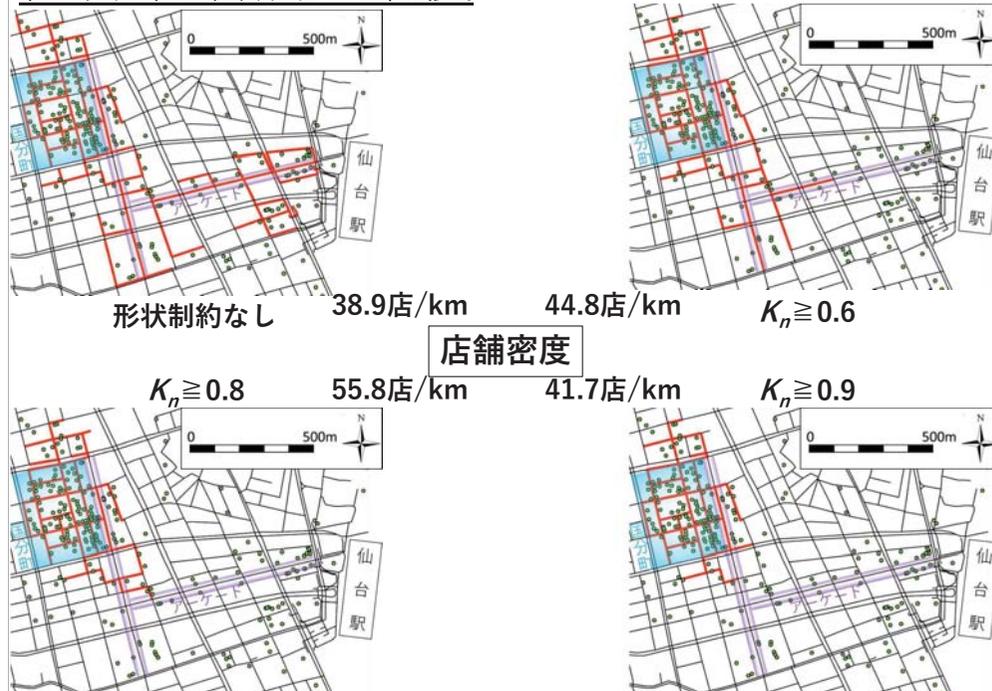
総延長647km

7,953リンク

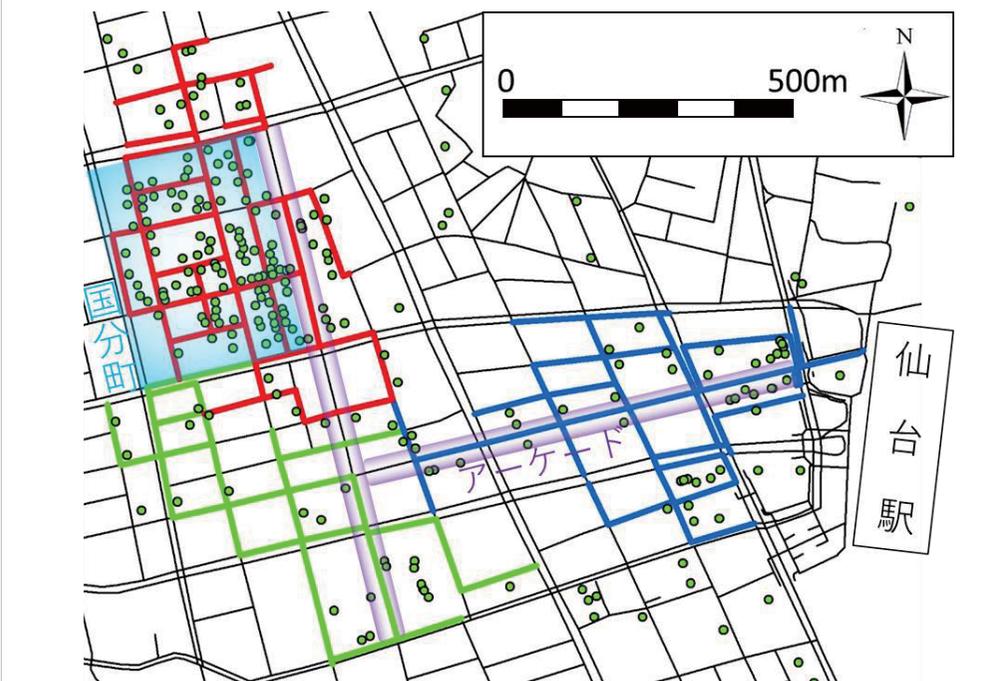
数値地図25000



仙台市 居酒屋集積 26



複数回・強い形状制約での集積検出例 27



時空間上の集積領域検出 適用例 28

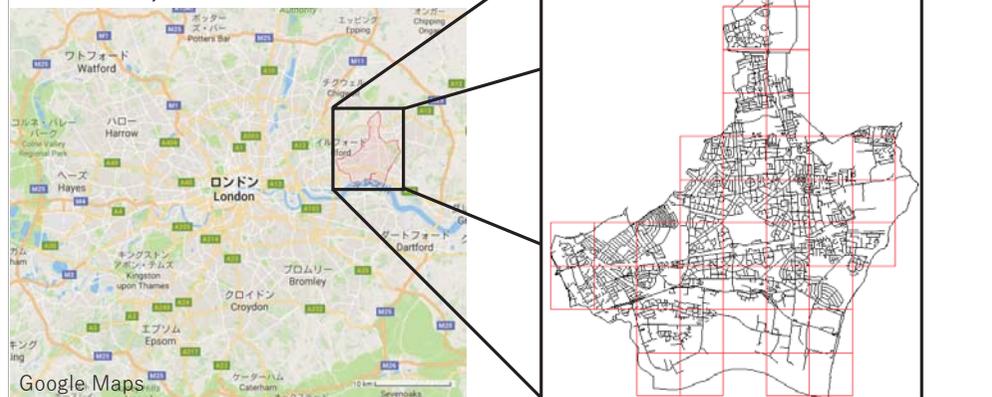
侵入盗(空き巣)発生 by ロンドン警視庁 (Metropolitan Police)

<地域> バーキング・アンド・ダゲナム・ロンドン特別区

<期間> 2014年1月1日~2015年12月31日 (1ヶ月単位)

<ボクセル設定> 1km×1km×1ヶ月 ※ボクセル内の対象領域が小さいものは除く
(1ヶ月あたり41個, 24ヶ月で全984個)

<件数> 3,362件

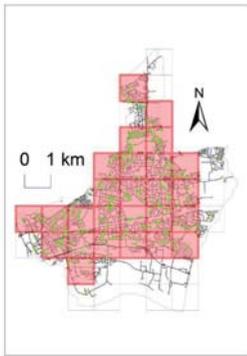


侵入盗発生の時空間集積

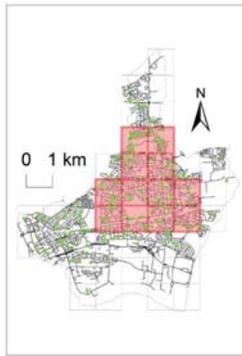
29

□ 集積領域(形状制約なし)
 □ 集積領域(形状指標 ≥ 0.8)

■ 形状指標 0.67
 ■ ボクセル数 524個
 ■ 事象数 2,985個
 ■ 点密度 (点/ボクセル)
 領域内(赤) 5.70
 領域外(色なし) 0.82



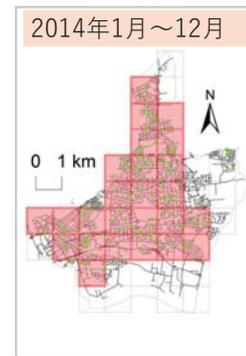
■ 形状指標 0.81
 ■ ボクセル数 229個
 ■ 事象数 1,534個
 ■ 点密度 (点/ボクセル)
 領域内(赤) 6.69
 領域外(色なし) 2.42



侵入盗発生の時空間集積

30

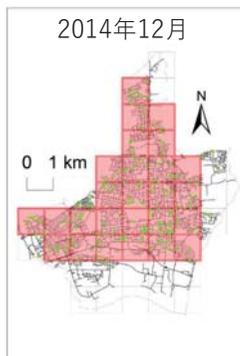
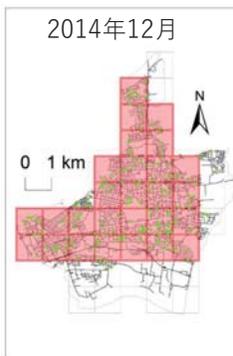
□ 集積領域(形状制約なし)
 □ 集積領域(形状指標 ≥ 0.8)



侵入盗発生の時空間集積

31

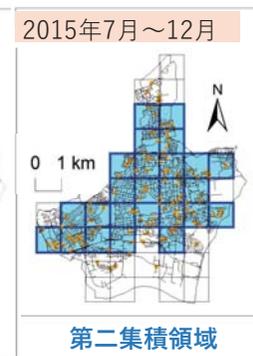
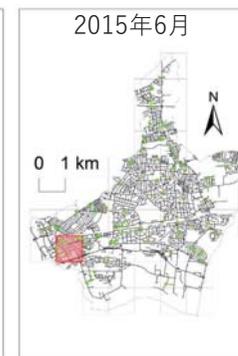
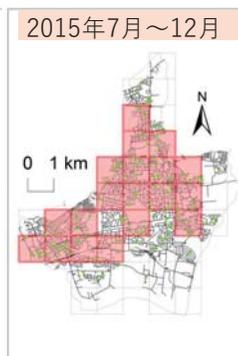
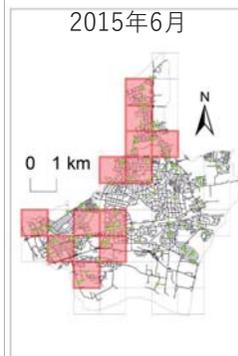
□ 集積領域(形状制約なし)
 □ 集積領域(形状指標 ≥ 0.8)



侵入盗発生の時空間集積

32

□ 集積領域(形状制約なし)
 □ 集積領域(形状指標 ≥ 0.8)

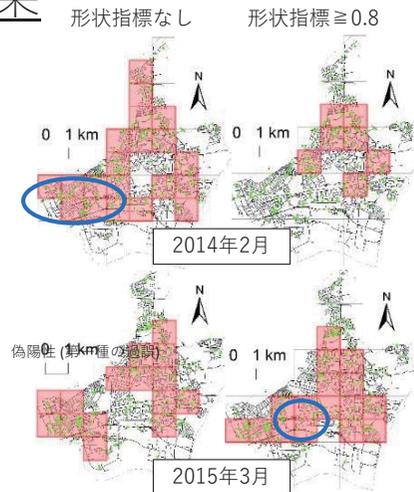


形状制約の設定と検出結果

33

- 形状制約を設定しないと、**点事象数が少ない地域を包含した**大きな領域を
集積として検出する可能性

- 形状制約を設定すると、**点事象数が少ない地域を包含した**
コンパクトな形状の領域を
集積として検出する可能性



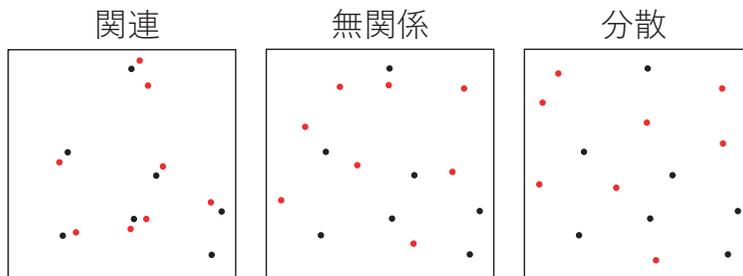
どちらの検出でも、点事象数が少ない領域を
集積として検出する可能性あり

→ 形状制約の強さを変えた分析を通して、
検出結果を比較し注意深く解釈する必要性

多種類の点事象の空間分布分析

35

▶ 分布パターンの識別 代表的手法: Cross K関数法



三種類以上の点事象分布の分析は？

▶ 多種類の点事象が関連して集積する領域の検出？

複数種類の点事象分布の関連性に関心がある研究分野
「産業の集積・共集積現象の分析」

3. 複数点事象の集積パターンの 分析手法の提案

34

— FDR制御法に基づく分析 —

産業の地理的集積現象

36

【誘因】

- 生産費用・輸送費の削減
- 企業間の取引費用の削減
- 高度技能労働者・中間財の集中

...

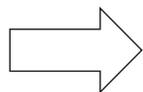
集積現象の発生メカニズムを解明する
様々な研究が行われている。



「関連性がある複数の産業が同じ地域に立地する現象」

発生メカニズムの解明に向けて

- ・ 理論モデルに基づく分析
- ・ 産業立地データを用いた実態把握・実証分析に関する研究が数多く行われている。



地域計画・企業誘致政策の立案など

ここでは、産業立地データを用いた共集積現象の実態把握に着目

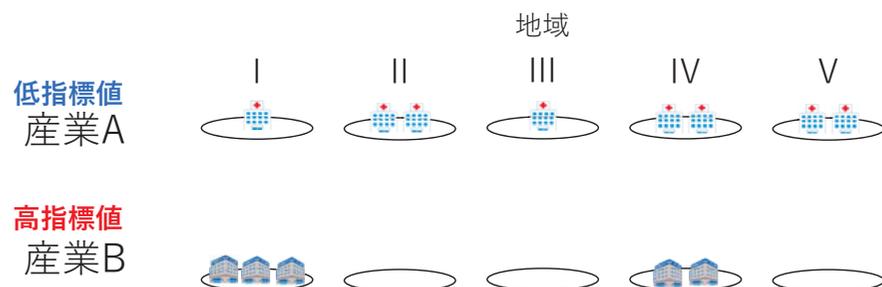
EG指標 (一産業 i の集積)

<データ> 地域単位に集計された雇用統計

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

s_{mi} : 産業 i の全従業者のうち、地域 m の従業者割合
 x_m : 全産業従業者のうち、地域 m の従業者割合
 H_i : 産業 i の企業単位のHerfindahl指標 (独占度指標)

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



• Ellison and Glaeser (1997)

地域単位の雇用統計データを利用
 各地域・各産業従業者割合に基づく指標
 ⇒ EG指標

• Duranton and Overman (2005)

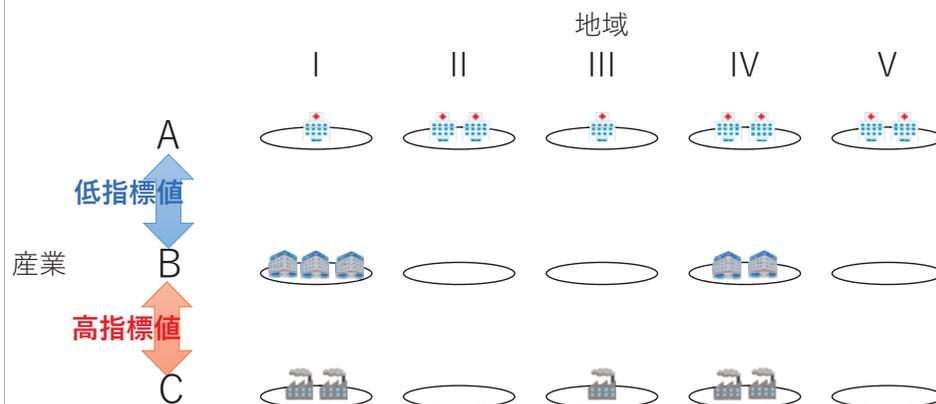
事業所立地点・従業者数データを利用
 事業所立地点間の距離に基づく指標
 ⇒ DO指標

EG指標 (二産業 ij 間の共集積)

<データ> 地域単位に集計された雇用統計

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

s_{mi} : 全国の産業 i の従業者のうち、地域 m の従業者割合
 x_m : 全産業従業者のうち、地域 m の従業者割合



EG指標 (多産業の組み合わせ I 内の共集積) ⁴¹

<データ> 地域単位に集計された雇用統計

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

w_i : 産業の組み合わせ I の従業者に占める産業 i の従業者数
 s_m : 地域 m における組み合わせ I の従業者割合
 H : 組み合わせ I の企業単位のHerfindahl指標
 $s_m = \sum_i w_i s_{mi}$
 $H = \sum_i w_i^2 H_i$

多産業の組み合わせを事前に設定 ⇒ 共集積の程度を評価可能
 しかし、共集積する多産業の組み合わせを効率的に見出す方法は提案なし。
 産業の組み合わせは大量にあるので、分析は困難。

Duranton & Overman (2005) は、EG指標は可変単位地区問題 (MAUP) を有すると批判。
 ・ 地域区分を変えると指標値が変わる
 ・ 同一地域に立地していない産業の組み合わせは考慮に入れない
 ⇒ 隣接地域に立地していても無視される



産業立地地点間の距離に基づく指標を提案

DO指標 (一産業 i の集積) ⁴²

<データ> 各産業事業所の立地位置・従業者数

各事業所の位置のみを考慮する場合

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

n : 全事業所数
 d_{ij} : 事業所 ij 間の距離
 f : バンド幅 h のガウス型カーネル密度関数

K関数 (Ripley, 1976) との類似性あり

各事業所の位置+従業者を考慮する場合

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

$e(i)$: 事業所 i の従業者数

DO指標 (二産業 ij 間の共集積) ⁴³

<データ> 各産業事業所の立地位置・従業者数

各事業所の位置のみを考慮する場合

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

n_i : 産業 i の事業所数
 d_{rs} : 事業所 rs 間のユークリッド距離
 f : バンド幅 h のガウス型カーネル密度関数

クロスK関数との類似性あり

各事業所の位置+従業者を考慮する場合

$$\hat{K}_{ij}^{Emp}(d) = \frac{1}{h \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} e(r)e(s)} \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} e(r)e(s) f\left(\frac{d-d_{rs}}{h}\right)$$

$e(i)$: 事業所 i の従業者数

既存の共集積指標のまとめ ⁴⁴

	Ellison & Glaser (1997)	Duranton & Overman (2005)
2産業間の共集積指標	地域の従業者割合 (簡略式) $\gamma_{ij}^c = \frac{\sum_{m=1}^M (s_m - x_m)(s_{mj} - x_m)}{1 - \sum_{m=1}^M x_m^2}$	事業所立地点 $\hat{K}_{ij}^{cEm}(d) = \frac{1}{n_i n_j h} \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} f\left(\frac{d-d_{rs}}{h}\right)$ 事業所立地点+従業者数 $\hat{K}_{ij}^{Emp}(d) = \frac{1}{h \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} e(r)e(s)} \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} e(r)e(s) f\left(\frac{d-d_{rs}}{h}\right)$
3産業以上の共集積指標	$\gamma^c \equiv \frac{\sum_{m=1}^M (s_m - x_m)^2 / \left(1 - \sum_{m=1}^M x_m^2\right) - H - \sum_{i=1}^I \gamma_i w_i^2 (1 - H_i)}{1 - \sum_{i=1}^I w_i^2}$	なし
長所	<ul style="list-style-type: none"> ➤ 地域単位の統計から算出可能 ➤ 3産業以上の共集積も分析可能 	<ul style="list-style-type: none"> ➤ 事業所間距離に基づく分析で、可変単位地区問題なし
短所	<ul style="list-style-type: none"> ➤ 可変単位地区問題あり ➤ 3産業以上からなる産業共集積を効率的に探索するアルゴリズムが提案されていない 	<ul style="list-style-type: none"> ➤ 事業所立地点データが必要 ➤ バンド幅設定に結果が依存 ➤ 3産業以上から構成される共集積の分析は困難
	<ul style="list-style-type: none"> ➤ 全体的な共集積の程度を示すのみ、共集積が生じている地域を示すことはできない 	

3産業以上からなる共集積を分析することは困難

- 3点以上の近接性の評価は？
- 3つ以上の異なる点過程から得られた指標値の統計的性質は？

(また、産業 i, j それぞれの空間分布は考慮されていない)

既存の産業共集積分析指標では分析ができない/難しい

- 共集積する産業の組み合わせ（産業共集積形態）の発見
- 産業共集積が起きている地域の表示

を行う手法を提案

接近法

1. 産業別の集積地域検出

- ✓ False Discovery Rate (FDR)制御法

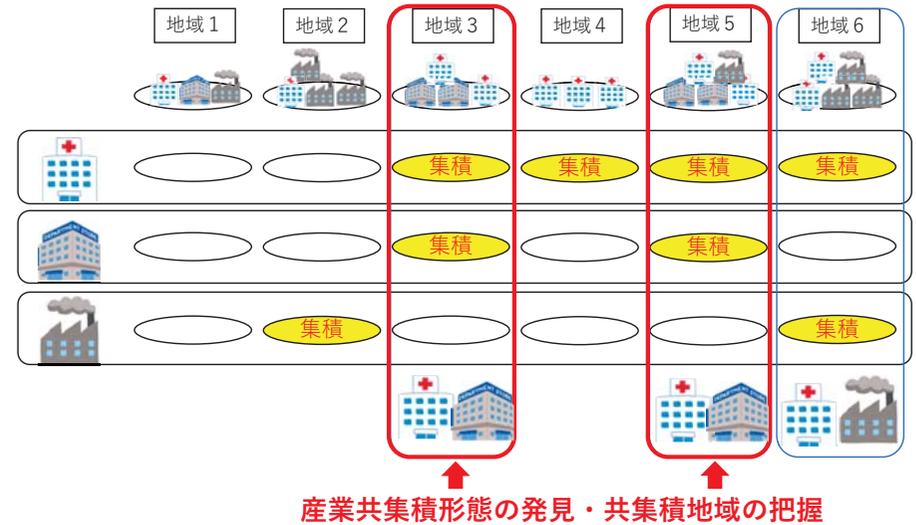
2. 産業共集積形態の発見・検定

- ✓ 頻出パターンマイニング
- ✓ Monte Carlo simulationによる有意性検定



[接近法1] 産業別の集積地域検出

[接近法2] 産業共集積形態の発見・検定

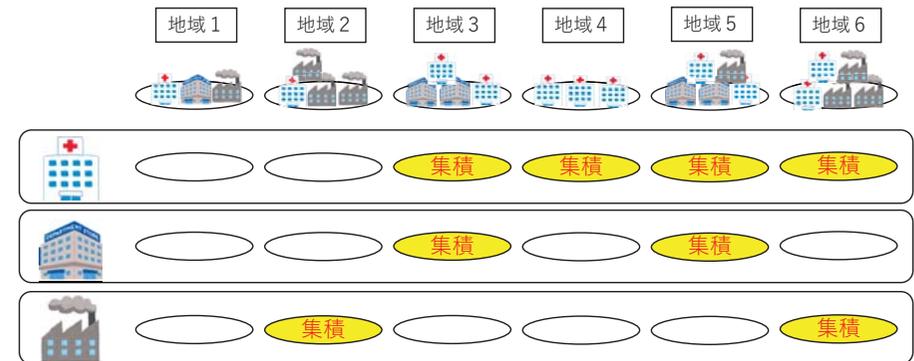


EG指標と提案手法の比較

	本研究の接近法	EG指標
データ	地域単位の産業・雇用統計	
分析の特徴	<ul style="list-style-type: none"> ➢ FDRに基づく統計的検定により産業立地（雇用）データを各産業の集積地域データに縮約 ➢ データマイニング手法により産業共集積形態を効率的に発見可能 ➢ 各産業共集積形態に対し共集積地域を示すことが可能 ➢ シミュレーションを通して共集積の有意性を評価可能 	<ul style="list-style-type: none"> ➢ 各産業の集積地域だけでなく全地域の従業者数分布が共集積指標に反映 ➢ 共集積を評価するためには産業共集積形態を事前に設定する必要あり ➢ 共集積地域は不明 ➢ 指標の統計的性質について議論されている

[接近法1] 産業別の集積地域検出

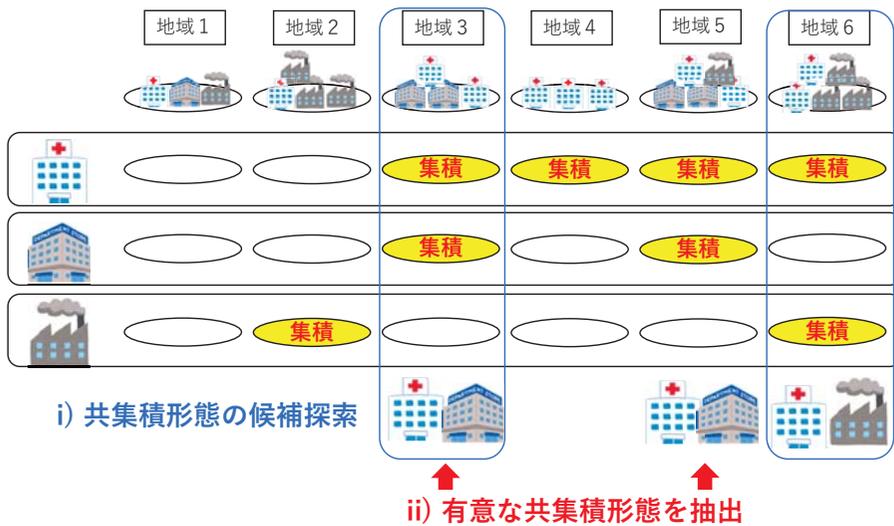
各産業が各地域で集積しているかを検定すると多重検定問題を生じる
⇒ FDR制御法を用いて分析



[接近法2] 産業共集積形態の発見・検定 49

i) 共集積形態の候補探索 頻出パターンマイニング

ii) 有意な共集積形態を抽出 Monte Carlo simulation



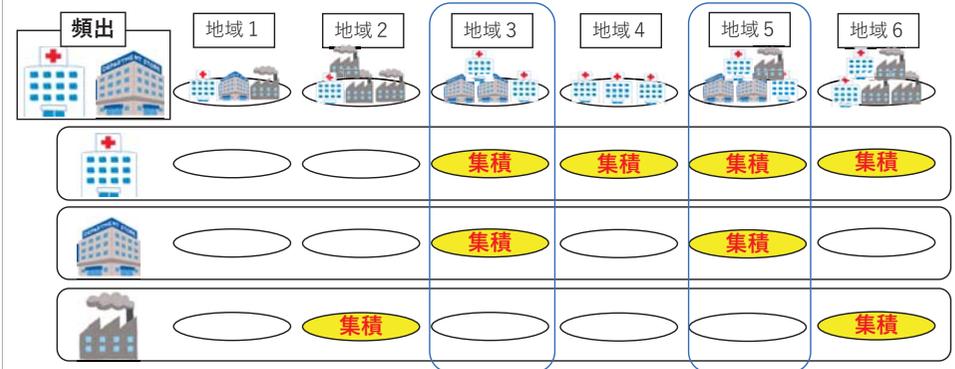
[接近法2] i) 共集積形態の候補探索 50

頻出パターンマイニング (Agrawal & Srikant, 1994)

FP-growth algorithm (Han et al., 2000) を使用

[接近法1] で得られた産業別集積地域の共起関係を用いて頻出パターンマイニングで共集積候補を探索する。

共集積地域数に閾値を設定し、閾値を満たす共集積形態を候補として抽出する。



[接近法2] ii) 有意な共集積形態の抽出 51

Monte Carlo simulation

各形態候補について、シミュレーションとデータから得られた共集積地域数を比較し、有意な共集積形態を抽出する。

多くの産業が都市に立地することを考慮するため、各産業の各地域の集積発生確率は、地域の集積産業数に比例するよう設定。

<集積発生確率の設定>

		地域1	地域2	地域3	地域4	地域5	地域6
集積地域数	集積産業数	0	1	2	1	2	2
4		0	1/2	1	1/2	1	1
2		0	1/4	1/2	1/4	1/2	1/2

提案手法の適用 使用データ 52

「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

2分の1 地域メッシュ (約500m四方)

産業別事業所数データ

全国のメッシュ総数: 1,515,129 (湖沼を除く)

事業所のあるメッシュ数: 336,646 (ゼロ切断データ)

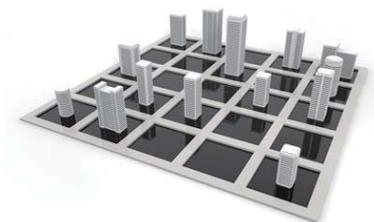
事業所数: 6,009,389[件]

(事業内容等が不詳の事業所及び農林漁業を除く)

産業分類:

日本標準産業分類(H19)

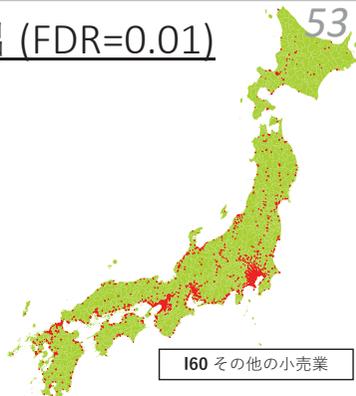
中分類 (87産業分類)



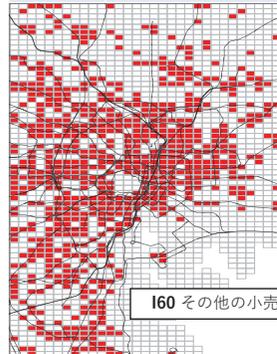
[接近法1] 産業別の集積地域検出 (FDR=0.01)

53

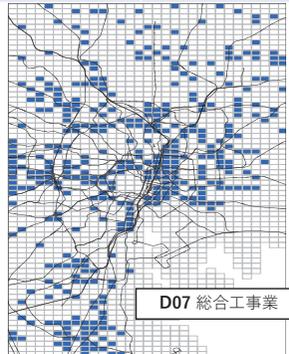
産業中分類	集積地域数上位10産業	集積地域数
60	その他の小売業	7,569
76	飲食店	7,563
78	洗濯・理容・美容・浴場業	7,431
58	飲食料点小売業	7,116
69	不動産賃貸業・管理業	5,452
83	医療業	4,423
49	郵便業(信書便事業を含む)	3,682
07	職別工事業(設備工事業を除く)	2,545
57	織物・衣服・身の回り品小売業	2,545
06	総合工事業	2,441



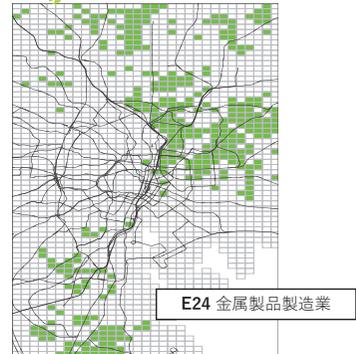
160 その他の小売業



160 その他の小売業



D07 総合工事業



E24 金属製品製造業

産業共集積検出に関して

55

[提案手法の利点]

- 産業別の集積地域検出を通じた産業立地データの圧縮により共集積形態を発見的に探索可能。
- FDR制御法を用いた集積検出により
 - 多重検定問題を回避
 - 多地域から各産業の集積地域を短時間に検出
- 頻出データマイニング手法を用いることにより多地域・多産業のデータから効率的に共集積形態候補を抽出可能。
- シミュレーションを通して、共集積形態の有意性を評価可能。
- 既往の産業共集積分析方法では把握できない共集積地域を示すことが可能。

[課題]

- 大量の産業組み合わせが抽出されるため、結果の解釈が困難。
⇒ 得られた産業間の関係を要約・可視化する手法が不可欠。
- EG指標と同様に、可変単位地域問題(MAUP)を抱える。
近隣メッシュの立地も考慮した分析が可能か検討の必要あり。

[接近法2] 産業共集積形態の発見・検定

54

共集積形態候補の探索

閾値: min(1産業以上の集積存在地域の1% (262メッシュ), 共集積形態内最小集積地域数の5%)

有意な共集積形態を抽出

(999回試行・FDR=0.01・部分集合が有意でない形態は棄却)

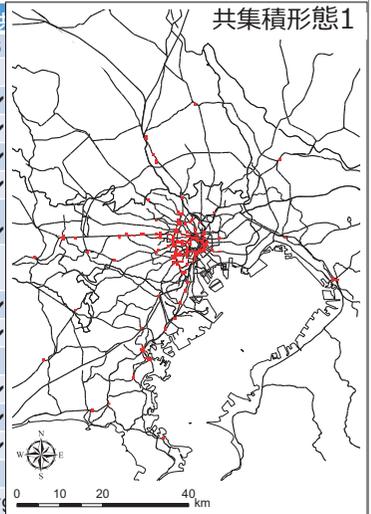
候補数: 54,304

うち 他の部分集合ではない形態数: 1,865

有意な形態数: 11,703

うち 他の部分集合ではない形態数: 824

産業分類 (中分類)	1	2	3	4	5	6
39 情報サービス業						
57 織物・衣服・身の回り品小売業		✓		✓	✓	✓
58 飲食料点小売業	✓	✓	✓	✓	✓	✓
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



共集積形態1

「点事象データから地域を観る」

56

「点事象の集積検出を通して、地域の特徴把握を目指す」

1. 空間スキャン統計に基づく

ネットワーク・時空間上の点事象集積手法の提案とその応用 (飲食店集積と犯罪集積)

2. FDR制御法に基づく

複数点事象の集積パターンの分析手法の提案とその応用 (産業共集積現象)

今後、オープンデータ・IoT (Internet of Things)の進展により多くの地理空間情報・データの流通が期待。データを活用した地域モニタリングを考えていきたい。