様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ５ 月 １９ 日現在

機関番号：１２６０１
研究種目：基盤研究(B)（一般）
研究期間：2019～2021
課題番号：１９Ｈ０２３７５

研究課題名（和文）汎用的空間・時空間点分布分析の枠組み構築

研究課題名（英文）Framework for analyzing spatial and spatiotemporal distributions of points

研究代表者
貞広　幸雄（Sadahiro, Yukio）

東京大学・大学院情報学環・学際情報学府・教授

研究者番号：１０２４０７２２

交付決定額（研究期間全体）：（直接経費）　　13,300,000円

研究成果の概要（和文）：本研究では，点クラスター以外の多様な空間・時空間パターンを記述・評価・抽出する新たな枠組みの開発を目指した．枠組みの満たすべき要件として，1）広域的及び局所的パターンに適用可能，2）時空間次元での可視化が可能，3）統計的検定が実用時間内で可能，という3点が挙げられる．可能性のある既存分析手法として，Moran's IやGetis's Gなどの空間的自己相関係数と，一連のスキャン統計量を検討し，それらの理論的改良を行った．そしてその有効性を検証するために，NTTタウンページの提供する渋谷区内の商業施設分布，及び，スマートフォンの移動データから抽出した，渋谷地域の商圏の時間的変動に適用した．

研究成果の学術的意義や社会的意義
本研究では，点クラスター以外の多様な空間・時空間パターンを扱っている．例えば点パターンは，社会工学だけではなく，地理学，疫学，犯罪学，統計学，経済学など，非常に多様な学問領域で研究が行われている．ネットワークパターンは，交通工学，生態学，都市工学などで取り上げられており，移動軌跡パターンは，土木工学，生態学，公衆衛生学などでの応用がある．本研究で開発した手法は，上記のような極めて広範な関連領域の研究者が，自身の研究領域において利用することができる．本研究成果の第一の意義はこの点にある．

研究成果の概要（英文）：This research aimed at the development of a general framework of spatial and spatiotemporal analysis. The framework needs to permit 1) global and local pattern analysis, 2) spatiotemporal visualization, and 3) statistical tests that finish within a reasonable time. We extended spatial autocorrelation measures and scan statistics to fulfil the above requirements, and applied them to the analysis of the commercial facilities in Shibuya-ku and the trajectory data of mobile phone users in the same area.

研究分野：空間解析

キーワード：空間パターン　時空間パターン

１．研究開始当初の背景

既存の点分布分析手法の多くは，点の集塊性の評価と点クラスターの検出を目的としている．点分布には，点クラスター以外にも様々な点パターンが存在するが，それらの分析手法は未だ不十分であり，パターン評価や抽出が困難である．

２．研究の目的

そこで本研究では，多様な点パターンに対応した汎用的空間・時空間分析枠組みを構築し，それを用いて具体的な分析手法を開発する．多様な点パターンの研究が進まない理由の一つは，様々な点パターンに対応可能な汎用的分析枠組みが存在しないことである．点パターンに応じてその都度分析手法を開発するのでは，研究の進行が大きく阻害される．そこで本研究では，様々なパターンに適用可能な，汎用的枠組みの開発を行う．

３．研究の方法

汎用的枠組みを構築した後，それを用いて具体的な点パターン分析手法を開発する．さらにそれを実データに適用することで，手法の妥当性を検証する．

４．研究成果

本研究課題において中心的な役割を果たす分析手法について，以下に概説する．

4.1 空間スキャン統計量

比率は，日常的によく用いられる数値指標である．高齢化率，大学進学率，建蔽率，原価率，自給率等々，比率は様々な現象を定量的に評価する便利かつ身近な指標である．

他方，比率は母集団の数に大きく左右される．例えば，労働力人口 20 人中 1 人の失業者という地域と，100 万人中 5 万人の失業者という地域では，失業率はいずれも 5%である．しかし前者の地域では，1 人が偶然，失業状態にあると解釈することも可能であるが，後者はそのような偶然とは考え難い．このような点を加味した分析方法として，確率的現象を仮定した尤度比を利用する方法がある．空間解析においてこの方法を用いた著名な手法の一つが Kulldorff and Nagarwalla (1995)及び Kulldorff (1997)の提案した空間スキャン統計量である．空間スキャン統計量は，点分布における集塊（クラスター）を検出する手法であり，疫学や犯罪学においてよく用いられる．

いま地域 S において，ある疾病の罹患者が空間的に集中している地域を検出したいものとする．S には N 人の住民が居住しており，それらは点オブジェクトの集合$\{P_1, P_2, ..., P_N\}$として表される．S に含まれる領域 R 内の住民と罹患者の人数をそれぞれ $n(R)$，$a(R)$と表記する．

罹患者の集中が疑われる地域として，位置ベクトル $\mathbf{x}$ に中心を置く半径 r の円 $Z(\mathbf{x}, r)$を想定する．疾病への罹患を確率現象と捉え，円 $Z(\mathbf{x}, r)$内と $S\text{-}Z(\mathbf{x}, r)$における罹患率をそれぞれ p_0，q_0 とする．このとき，罹患現象の尤度関数は

$$\begin{aligned} & L\left(Z\left(\mathbf{x},r\right),p_0,q_0\right) \\ & = {}_{n\left(Z\left(\mathbf{x},r\right)\right)}C_{c\left(Z\left(\mathbf{x},r\right)\right)}\, p_0^{a\left(Z\left(\mathbf{x},r\right)\right)}\left(1-p_0\right)^{n\left(Z\left(\mathbf{x},r\right)\right)-a\left(Z\left(\mathbf{x},r\right)\right)} \\ & \times {}_{n\left(S-Z\left(\mathbf{x},r\right)\right)}C_{a\left(S-Z\left(\mathbf{x},r\right)\right)}\, q_0^{a\left(S-Z\left(\mathbf{x},r\right)\right)}\left(1-q_0\right)^{n\left(S-Z\left(\mathbf{x},r\right)\right)-a\left(S-Z\left(\mathbf{x},r\right)\right)} \end{aligned} \tag{1}$$

である．ここで，罹患率が円 $Z(\mathbf{x}, r)$と領域 $S\text{-}Z(\mathbf{x}, r)$で同一である，即ち，$p_0=q_0$ とする帰無仮説と，$p_0 \neq q_0$ とする対立仮説を考える．それぞれの仮説における最大尤度は

$$\begin{aligned} L_0\left(Z\left(\mathbf{x},r\right)\right) &= \max_{p_0=q_0} L\left(Z\left(\mathbf{x},r\right),p_0,q_0\right) \\ &= {}_{n\left(Z\left(\mathbf{x},r\right)\right)}C_{a\left(Z\left(\mathbf{x},r\right)\right)}\ {}_{n\left(S-Z\left(\mathbf{x},r\right)\right)}C_{a\left(S-Z\left(\mathbf{x},r\right)\right)} \times \left(\frac{a\left(S\right)}{n\left(S\right)}\right)^{a\left(S\right)}\left(1-\frac{a\left(S\right)}{n\left(S\right)}\right)^{n\left(S\right)-a\left(S\right)} \end{aligned} \tag{2}$$

及び

$$L_1\left(Z\left(\mathbf{x},r\right)\right)=\max_{p_0\neq q_0}L\left(Z\left(\mathbf{x},r\right),p_0,q_0\right)$$
$$={}_{n\left(Z\left(\mathbf{x},r\right)\right)}C_{a\left(Z\left(\mathbf{x},r\right)\right)}\left(\frac{a\left(Z\left(\mathbf{x},r\right)\right)}{n\left(Z\left(\mathbf{x},r\right)\right)}\right)^{a\left(Z\left(\mathbf{x},r\right)\right)}\times\left(1-\frac{a\left(Z\left(\mathbf{x},r\right)\right)}{n\left(Z\left(\mathbf{x},r\right)\right)}\right)^{n\left(Z\left(\mathbf{x},r\right)\right)-a\left(Z\left(\mathbf{x},r\right)\right)},$$
$$\times{}_{n\left(S-Z\left(\mathbf{x},r\right)\right)}C_{a\left(S-Z\left(\mathbf{x},r\right)\right)}\left(\frac{a\left(S-Z\left(\mathbf{x},r\right)\right)}{n\left(S-Z\left(\mathbf{x},r\right)\right)}\right)^{a\left(S-Z\left(\mathbf{x},r\right)\right)}\times\left(1-\frac{a\left(S-Z\left(\mathbf{x},r\right)\right)}{n\left(S-Z\left(\mathbf{x},r\right)\right)}\right)^{n\left(S-Z\left(\mathbf{x},r\right)\right)-a\left(S-Z\left(\mathbf{x},r\right)\right)}$$

(3)

であり，これらの尤度比は

$$\lambda\left(Z\left(\mathbf{x},r\right)\right)=\frac{L_1\left(Z\left(\mathbf{x},r\right)\right)}{L_0\left(Z\left(\mathbf{x},r\right)\right)}.$$

(4)

となる．p_0 と q_0 が近い値の場合，尤度比$\lambda(Z(\mathbf{x}, r))$の値は小さくなり，円 $Z(\mathbf{x}, r)$と領域 $S\text{-}Z(\mathbf{x}, r)$とで罹患率の差異が小さいことになる．反対に尤度比$\lambda(Z(\mathbf{x}, r))$が大きな場合，$p_0$ と q_0 の差異が大きい，即ち，円 $Z(\mathbf{x}, r)$の内外で罹患率の差が大きいことになる．$p_0>q_0$ の場合には，円 $Z(\mathbf{x}, r)$における罹患率が $S\text{-}Z(\mathbf{x}, r)$と比べて高く，反対に $p_0<q_0$ の場合には円 $Z(\mathbf{x}, r)$における罹患率が相対的に低い．

帰無仮説が成り立つ，即ち

$$p_0=q_0=\frac{a(S)}{n(S)},$$

(5)

の場合，$\lambda(Z(\mathbf{x}, r))$が得られている値よりも小さくなる確率を $\Pr[Z(\mathbf{x}, r)]$とし，以下の変数を定義する．

$$\kappa\left(\mathbf{x},r\right)=\begin{cases}\Pr\left[Z\left(\mathbf{x},r\right)\right] & \text{if } p_0>q_0\\ -\Pr\left[Z\left(\mathbf{x},r\right)\right] & \text{otherwise}\end{cases},$$

(6)

変数$\kappa(\mathbf{x}, r)$は変域が-1～1 であり，その絶対値が円 $Z(\mathbf{x}, r)$と $S\text{-}Z(\mathbf{x}, r)$の罹患率の乖離度を表す．

この変数$\kappa(\mathbf{x}, r)$を用いることで，冒頭に述べた「20 人中 1 人の失業者と，100 万人中 5 万人の失業者」を区別することができる．失業者を罹患者と読み替えると，20 人中 1 人が失業する確率は，100 万人中 5 万人が失業する確率よりも遥かに高い，即ち，偶然である可能性が相対的に高い．反対に 100 万人中 5 万人が失業しているという状態は，偶然起こり得ない，即ち，何らかの構造的要因によるものであり，20 人中 1 人の失業者という事象と比べて遥かに深刻な問題である．

空間スキャン統計量では$\lambda(Z(\mathbf{x}, r))$を算出後，円 $Z(\mathbf{x}, r)$を半径を 0 から既定値 r_{MAX} まで徐々に拡大し，$p_0>q_0$ という条件の下で$\lambda(Z(\mathbf{x}, r))$が最大となる円を探索する．この円は地点 $\mathbf{x}$ の関数 $Z_{\mathrm{MAX}}(\mathbf{x})$と表記され，様々な地点における値の中で最大の尤度比を与えるものについて有意性の検定を行い，最も異質性の高い地域を抽出する．

4.2 手法の一般化と課題

空間スキャン統計量は，元来は疾病罹患者の集塊検出を目的とする手法であるが，点分布として表現される空間オブジェクトの集合であれば，広範に適用可能である．特に尤度比$\lambda(Z(\mathbf{x}, r))$に基づいた評価の枠組み自体は，その汎用性が非常に高い．即ち，ある確率的空間現象について，円領域 $Z(\mathbf{x}, r)$の内外で想定するモデルの同一性という観点から，$Z(\mathbf{x}, r)$内の空間現象の異質性を評価し，顕著に異質な地域を検出する．換言すれば，確率的現象と想定できる空間現象であれば，枠組み自体はさらに広範な対象に適用可能である．以下，具体例を述べる．

まずは，量的データを属性として持つ点分布に関する分析である．例えば，地域 S 居住者の身長の分布を考える．分布は概ね正規分布に従うものと仮定し，$Z(\mathbf{x}, r)$の内外でそれぞれ身長が正規分布 $N(\mu_0, \sigma_0^2)$，$N(\mu_1, \sigma_1^2)$に従うものとする．分布自体の差異を評価する場合は，$\mu_0\neq\mu_1$ かつ$\sigma_0\neq\sigma_1$ と，$\mu_0=\mu_1$ かつ$\sigma_0=\sigma_1$ という 2 つの異なる仮定の上でそれぞれ分布を推定し，尤度比$\lambda(Z(\mathbf{x}, r))$を算出，可視化すれば良い．分布の平均値の差異を評価するには，$\mu_0\neq\mu_1$ かつ$\sigma_0=\sigma_1$ と，$\mu_0=\mu_1$ かつ$\sigma_0=\sigma_1$ という仮定に基づいた分析を行う．分散の差異であれば，$\mu_0=\mu_1$ かつ$\sigma_0\neq\sigma_1$ と，$\mu_0=\mu_1$ かつ$\sigma_0=\sigma_1$ とすれば良い．

次に，点分布で表されるバス停の日平均利用者数を利用圏内の高齢者数で説明する，即ち，交

通弱者である高齢者が多いほど，バス利用者も多いという単回帰モデルを考える．地域によっては高齢者の多くが自家用車を利用しているためにバス利用者が少なかったり，反対に自動車利用率が低くバス利用者が多いなどの差異が見られる可能性がある．このような地域的差異の可視化には，以下の方法が考えられる．

いま，i 番目のバス停の位置ベクトルと利用者数をそれぞれ $\mathbf{y}_i$，b_i，バス停利用圏を半径 r の円とし，その内部の高齢者数を e_i とおく．このとき，以下の単回帰モデルを考える．

$$b\left(\mathbf{y}_i\right)=\begin{cases}\alpha_{10}e_i+\beta & \text{if } \mathbf{y}_i\subset Z\left(\mathbf{x},r\right)\\ \alpha_{11}e_i+\beta & \text{otherwise}\end{cases} \tag{7}$$

このモデルを $\alpha_{10}=\alpha_{11}$ 及び $\alpha_{10}\neq\alpha_{11}$ という二つの条件下で推定し，$\lambda(Z(\mathbf{x}, r))$ を算出する．この値は $Z(\mathbf{x}, r)$ 内における高齢者数の影響の大小を表し，$\alpha_{10}>\alpha_{11}$ の場合には $Z(\mathbf{x}, r)$ 内部の方が高齢者数の影響が強く，$\alpha_{10}<\alpha_{11}$ の場合には影響が弱いことになる．

上記の単回帰モデルは，重回帰モデルにも拡張できる．交通弱者である高齢者が多く，かつ，道路の渋滞率が高いほど，バス利用者も多いと考えると，i 番目のバス停の面する道路の混雑率を c_i として，

$$b\left(\mathbf{y}_i\right)=\begin{cases}\alpha_{10}e_i+\alpha_{20}c_i+\beta & \text{if } \mathbf{y}_i\subset Z\left(\mathbf{x},r\right)\\ \alpha_{11}e_i+\alpha_{21}c_i+\beta & \text{otherwise}\end{cases} \tag{8}$$

というモデルを考えることができる．このモデルを，帰無仮説は $\alpha_{10}=\alpha_{11}$ かつ $\alpha_{20}=\alpha_{21}$，対立仮説を $\alpha_{10}\neq\alpha_{11}$ かつ $\alpha_{20}=\alpha_{21}$ と，$\alpha_{10}=\alpha_{11}$ かつ $\alpha_{20}\neq\alpha_{21}$ という 2 通りを想定してそれぞれ推定し，2 つの尤度比 $\lambda(Z(\mathbf{x}, r))$ を算出する．α_{10} と α_{11} の大小関係については，単回帰モデルの場合と同様に，α_{20} と α_{21} については，$\alpha_{20}>\alpha_{21}$ の場合には $Z(\mathbf{x}, r)$ 内部の方が道路の渋滞率の影響が強く，$\alpha_{20}<\alpha_{21}$ の場合には影響が弱いと解釈できる．高齢者数と道路の渋滞率の影響をそれぞれ $\lambda(Z(\mathbf{x}, r))$ に基づいて異なる色相によって可視化すると，高齢者数と渋滞率の影響がいずれも強い地域，いずれも弱い地域，一方のみ強い地域などを視覚的に把握できる．

回帰モデルの例は，通常の回帰分析における誤差分布の可視化でもある程度代替可能である．しかし回帰分析では，可視化の際に標本数の大小を考慮しないため，母集団の数への依存性には対応できない．

上記はいずれも，尤度比 $\lambda(Z(\mathbf{x}, r))$ に基づく分析例の一部であり，この枠組みはより広範な適用可能性を有している．空間パターンの可視化手法，特に，標本数を加味した探索的分析手法として有望である．今後さらに多くの適用例を重ねることが望ましい．

反面，尤度比に基づく枠組みでは，尤度比の計算自体が困難な場合がある．上記例では，パラメータの最尤推定量が陽に表されるため，尤度比の計算は容易である．しかし一般には，最尤推定量が陽に表されず，数値計算に頼らざるを得ない．幸い近年は，maxLik（Henningsen and Toomet (2011)）や maxlogL（Gutiérrez and Hernández (2020)）などパッケージが提供されているので，それらを利用した効率的な尤度比の計算が求められる．

また，異質地域の検出には，帰無仮説下での尤度比の確率分布が必要である．その導出にはモンテカルロシミュレーションを用いることが多いが，計算量の大きさが課題となる．この点については，今後，より効率的なアルゴリズムの開発等が望まれる．

———参考文献———

Gutiérrez JM, Hernández F (2020) Maxlogl: A general computational procedure for maximum likelihood estimation in r.

Henningsen A, Toomet O (2011) Maxlik: A package for maximum likelihood estimation in r. *Computational Statistics* 26: 443-458

Kulldorff M (1997) A spatial scan statistic. *Communications in Statistics - Theory and Methods* 26: 1481-1496

Kulldorff M, Nagarwalla N (1995) Spatial disease clusters: Detection and inference. *Stat Med* 14: 799-810

Sadahiro Y (2019a) Analysis of the appearance and disappearance of point objects over time. *International Journal of Geographical Information Science* 33: 215-239

Sadahiro Y (2019b) Statistical analysis of spatial segregation of points. *Computers, Environment and Urban Systems* 76: 123-138

5．主な発表論文等

〔雑誌論文〕　計5件（うち査読付論文　5件／うち国際共著　1件／うちオープンアクセス　1件）

1．著者名	4．巻
Yukio Sadahiro	25
2．論文標題	5．発行年
A method for analyzing the expansion and shrinkage of point pattern	2021年
3．雑誌名	6．最初と最後の頁
Transactions in GIS	534-550
掲載論文のDOI（デジタルオブジェクト識別子）	査読の有無
なし	有
オープンアクセス	国際共著
オープンアクセスではない、又はオープンアクセスが困難	－

1．著者名	4．巻
Yukio Sadahiro	58
2．論文標題	5．発行年
A method for analyzing the daily variation in the spatial pattern of market area	2021年
3．雑誌名	6．最初と最後の頁
Journal of Retailing and Consumer Services	-
掲載論文のDOI（デジタルオブジェクト識別子）	査読の有無
なし	有
オープンアクセス	国際共著
オープンアクセスではない、又はオープンアクセスが困難	－

1．著者名	4．巻
Wang S and Y Sadahiro	15
2．論文標題	5．発行年
A fine-grained evaluation of social vulnerability to COVID-19, healthcare access and built environmental disparities in Tokyo, Japan	2022年
3．雑誌名	6．最初と最後の頁
International Journal of Eigital Earth	2006-2027
掲載論文のDOI（デジタルオブジェクト識別子）	査読の有無
なし	有
オープンアクセス	国際共著
オープンアクセスとしている（また、その予定である）	該当する

1．著者名	4．巻
Sadahiro Y	24
2．論文標題	5．発行年
A method for evaluating the degree of spatial and temporal avoidance in spatial point patterns	2022年
3．雑誌名	6．最初と最後の頁
Journal of Geographical Systems	241-260
掲載論文のDOI（デジタルオブジェクト識別子）	査読の有無
なし	有
オープンアクセス	国際共著
オープンアクセスではない、又はオープンアクセスが困難	－

１．著者名 Sadahiro Y	４．巻 -
２．論文標題 Event pattern analysis: Spatial clustering of sequential events and temporal change of events over time	５．発行年 2023年
３．雑誌名 Transactions in GIS	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計0件

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	山田　育穂 (Yamada Ikuho) (00594756)	東京大学・空間情報科学研究センター・教授 (12601)	
研究分担者	奥貫　圭一 (Okunuki Keiichi) (90272369)	群馬大学・社会情報学部・教授 (12301)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関