

点分布における点の生成と消滅の分析

貞広幸雄

A New Method for Analyzing the Appearance and Disappearance of Points

Yukio SADAHIRO

Abstract: This paper proposes a new method for analyzing the appearance and disappearance of points. The method helps analysts capturing the overall picture and regional variation of event pattern and detecting significant local patterns. Four measures are defined that indicate the intensity of spatial and temporal patterns of events. The measures are visualized as grid maps. Statistical test evaluates the significance of the measures to extract the regions of significant patterns. The proposed method is applied to the analysis of shops and restaurants in Shibuya, Tokyo.

Keywords: 点分布 (point distribution), 生成 (appearance), 消滅 (disappearance)

1. はじめに

本論文では、点分布における点の生成と消滅を分析する新たな手法を提案する。点の生成と消滅は、植物の生死、建物の建築と解体、店舗の開閉店などを表す時空間現象である。こうした現象を扱う手法はこれまで多数提案されているが、本論文では既存手法の問題点を解決し、点の生成と消滅の空間パターンと時間パターンを分析する新たな手法を提示する（レビュー及び手法の詳細は Sadahiro (2019) 参照のこと）。

2. 手法の概略

本節では手法の概略を説明する。いま、領域 S_0 において、期間 T の間に N 個の点が存在したものとする。点の一部は全期間を通じて存在し、その他は一部期間のみ存在している。領域 S_0 内に中心 \mathbf{x} 、半径 r の円 $Z_0(\mathbf{x}, r)$ を取り、円柱 Z を $Z_0(\mathbf{x}, r) \otimes T$ 、円柱 S を $S_0 \otimes T$ とそれぞれ定義する。

2.1 空間パターン

まず、生成と消滅の頻度を図化し、それらが特に高い或いは低い地域を抽出する方法を提案する。生成と消滅は同様の手続きで扱うことが可能であるため、以下の説明は生成についてのみ行う。

いま、 Z 及び $S-Z$ において、生成が確率的に発生するモデルを考える。それぞれの発生確率を p_0 , q_0 と表記し、その値は実際のデータから最尤法によって推定する。推定された p_0 が q_0 よりも大きければ、 Z では生成が相対的に頻繁であり、 p_0 が小さければ、生成が少ないとということになる。

次に、 Z と $S-Z$ で生成確率が異なるかどうかを考える。 $p_0=q_0$ 及び $p_0 \neq q_0$ をそれぞれ帰無仮説、対立仮説とし、最尤法によって導出した尤度関数の最大値をそれぞれ $L_0(Z)$, $L_1(Z)$ と書く。このとき対数尤度比は

$$\lambda_A(Z) = \log L_1(Z) - \log L_0(Z) \quad (1)$$

で与えられる。そして、空間スキャン統計と同様に、各地点 \mathbf{x} において Z の半径を徐々に拡大し、予め定めた最大値までの間で $\lambda_A(Z)$ が最大となる円を抽出する。そのときの $\lambda_A(Z)$ を生成頻度指標

貞広幸雄

東京大学大学院情報学環

E-mail: sada@csis.u-tokyo.ac.jp

$\alpha_A(\mathbf{x})$ と呼ぶ.

$$\alpha_A(\mathbf{x}) = \max_{r \leq r_{\max}} \lambda_A(Z) \quad (2)$$

$p_0 > q_0$ のとき, Z では生成の頻度が相対的に高く, $p_0 < q_0$ のときには頻度が相対的に低いことになる. 以下, これらを {A+, A-} と表記する (消滅の場合は {D+, D-} と書く). この 2 つを異なる色相で区別し, $\alpha_A(\mathbf{x})$ を明度或いは彩度で表現することで, 生成の相対的頻度を図化し, その空間パターンを把握することが可能となる.

次に, 相対的頻度が他と比べて統計的に有意に異なる地域を抽出する. まず S_0 に格子網を重ね, 各格子点で $\alpha_A(\mathbf{x})$ を計算し, 降順に並べる. 次に, 大きな値の $\alpha_A(\mathbf{x})$ に対応する円から順に, 互いに重ならないものを予め定めた個数になるまで選択する. そして, FDR(False Discovery Rate)を制御することで, 多重検定の問題を回避しながら統計的に有意な円を抽出する.

2.2 時間パターン

次に, 生成と消滅の頻度の時間的変化, 特に, その増減傾向を分析する. いま, 期間 T を M の等間隔区間 $\{T_1, T_2, \dots, T_M\}$ に分割し, Z_i と S_i をそれぞれ $Z_0(\mathbf{x}, r) \otimes T_i$ と $S_0 \otimes T_i$ により定義される円柱, Z_i と S_i における生成の発生確率 p_i 及び q_i とする. 全ての i について $p_i = kq_i$ を帰無仮説, ある i について $p_i \neq kq_i$ を対立仮説とし, それぞれの妥当性を比較する.

モデルを最尤法によって推定し, 得られた尤度関数の最大値をそれぞれ $L^T_0(Z)$, $L^T_1(Z)$ とおくと, 対数尤度比は

$$\lambda^T_A(Z) = \log L^T_1(Z) - \log L^T_0(Z) \quad (3)$$

で与えられる.

2.1 節と同様, 各地点 \mathbf{x} において Z の半径を徐々に拡大し, 予め定めた最大値までの間で $\lambda^T_A(Z)$ が最大となる円を抽出, $\alpha^T_A(Z)$ を以下のように定める.

$$\alpha^T_A(\mathbf{x}) = \max_{r \leq r_{\max}} \lambda^T_A(Z) \quad (4)$$

Z と $S-Z$ において生成頻度の変化を比較するために, 以下の 2 つの変数を定義する.

$$k_i = \frac{p_i}{q_i} \quad (5)$$

$$\kappa_i = \frac{k_i}{\frac{1}{M} \sum_i k_i} \quad (6)$$

$\kappa_i > 1$ のとき, Z_i では S_i-Z_i よりも生成の頻度が高いことを示す. 従って各 κ_i を 1 と比較することで, Z における生成の時系列パターンを評価できる. 例えば $M=2$ のとき, $\kappa_1 < 1$ かつ $\kappa_2 > 1$ か, $\kappa_1 > 1$ かつ $\kappa_2 < 1$ の 2 通りが有り得る. 前者は Z において生成が相対的に増加, 後者は反対に減少していることを表し, 以下 {A_I, A_D} と書く (消滅の場合は {D_I, D_D}). 2.1 節と同様, 増減を 2 つの異なる色相で区別し, $\alpha^T_A(\mathbf{x})$ を明度或いは彩度で表現することで, 生成確率の時間パターンを図化することができる. 統計的に有意なパターンの抽出も 2.1 節と同様の方法で行う.

3. 適用例

2 節で提案した手法の有効性を検証するために, 本節では NTT タウンページの電話帳データを利用し, 2001~2016 年の渋谷区における商業施設分布の分析を行う. 紙面の都合上, ここでは衣料品店と飲食店に関する分析結果のみを示す. 図 1 及び図 2 はそれぞれ 2016 年における衣料品店と飲食店の分布である.

まず初めに, 店舗の開閉店の頻度を図化する. 図 3 は衣料品店の $\alpha_A(\mathbf{x})$ と $\alpha_D(\mathbf{x})$ の分布を表す. 紫色が濃い地域では開閉店が頻繁であり, 店舗間競争が厳しいことを表す. 開閉店は渋谷駅北や代官山駅周辺で最も頻繁であり, 原宿駅東や恵比寿駅南がそれに続く. 濃い緑色は開閉店が少ないことを表しており, 笹塚・恵比寿・広尾駅周辺, 渋谷駅南などでは衣料品店間の競争がそれほど厳し

くないと考えられる。図3は図1と類似しており、一般的に、店舗が多い地域ほど競争が厳しく、少ない地域ほど競争がないことができる。渋谷駅周辺では、北と南で開閉店の頻度が大きく異なる。前者では若者向けの様々な種類の衣料品店が多く競争も厳しいが、後者では古着やアメカジの店が長く存続していることがその要因と考えられる。

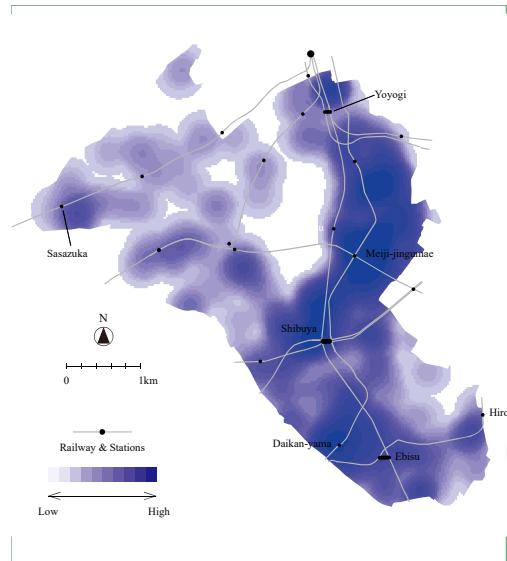


図1 2016年における衣料品店の分布

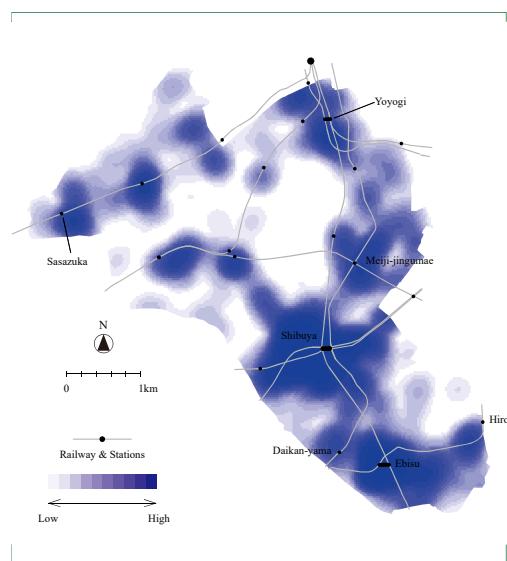


図2 2016年における飲食店の分布

図4は飲食店の $\alpha_A(x)$ と $\alpha_D(x)$ の分布を表す。衣料品店とは異なり、図4は必ずしも図2とは類似

していない。例えば渋谷駅周辺には飲食店が多いが、開閉店はそれほど頻繁ではない。渋谷駅近隣には老舗の飲食店が多く、昔からの顧客が根強く支持していることがその要因と思われる。恵比寿駅周辺にも飲食店は多いが、図4を見ると閉店が少なく開店が多い、即ち、飲食店が増加している。この地域では商業集積が西に拡大しつつあることが背景にあるものと考えられる。

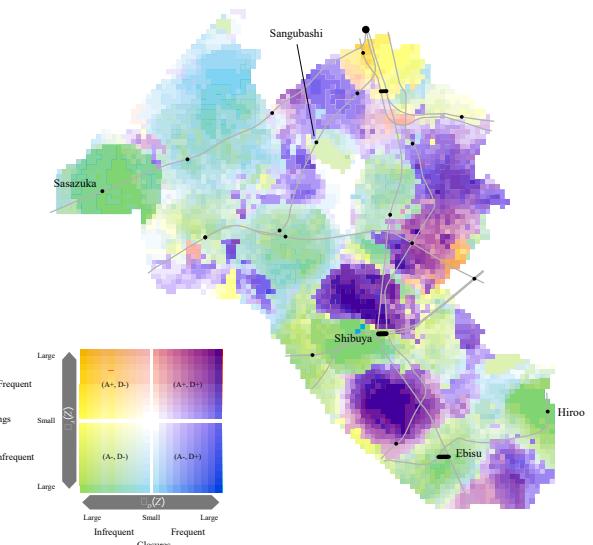


図3 衣料品店の $\alpha_A(x)$ と $\alpha_D(x)$ の分布

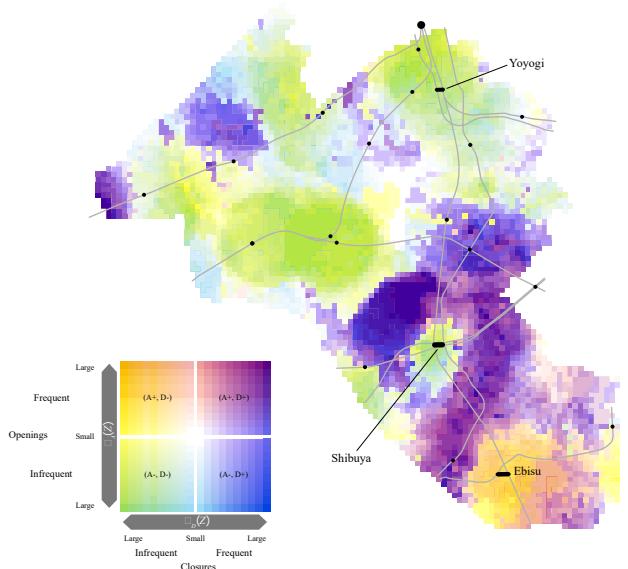


図4 飲食店の $\alpha_A(x)$ と $\alpha_D(x)$ の分布

図5は開閉店が統計的に有意に多い或いは少な

い地域を示している。概ね図3・4と合致しているが、統計的検定によって $\alpha_A(\mathbf{x})$ や $\alpha_D(\mathbf{x})$ に対する評価がより細分化されることがわかる。例えば図3では渋谷、恵比寿、広尾駅周辺は同様の色に見えるが、図5aでは渋谷駅周辺が1・3位、広尾駅周辺が5位、恵比寿駅周辺は11位以下となっており、その差異が明確である。

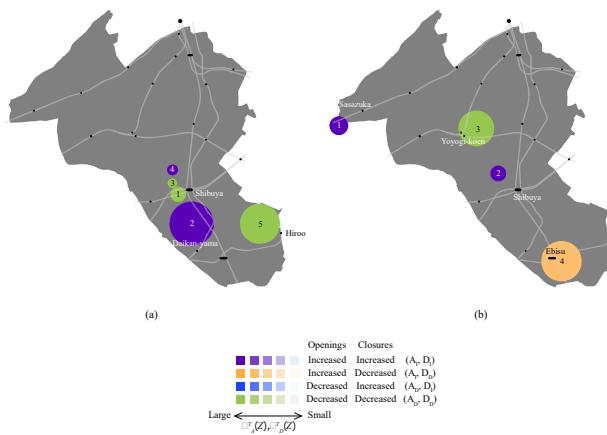


図5 開閉店の統計的に有意に多い或いは少ない地域。

(a) 衣料品店, (b) 飲食店

次に、全期間を2つに分け、前期から後期への開閉店頻度の変化を見る。図6は衣料品店の $\alpha_A^T(\mathbf{x})$ と $\alpha_D^T(\mathbf{x})$ の分布である。渋谷駅周辺では開店が減少し、閉店が増加している。一方代官山駅周辺では開閉店とも減少しており、店舗間競争が落ち着いてきているものと思われる。代々木駅南では開店が増加し、閉店が減少している。図3と合わせると、この地域では今後、店舗間競争が厳しくなることが予想される。

図7は飲食店の $\alpha_A^T(\mathbf{x})$ と $\alpha_D^T(\mathbf{x})$ の分布である。渋谷駅周辺では開閉店とも減少しており、図4で見られる店舗間競争は落ち着いてきているものと考えられる。明治神宮前駅周辺では、閉店が減少、開店が増加し、全体的に店舗が増加している。これは、当該地域で飲食店街全体が拡大していることによる。

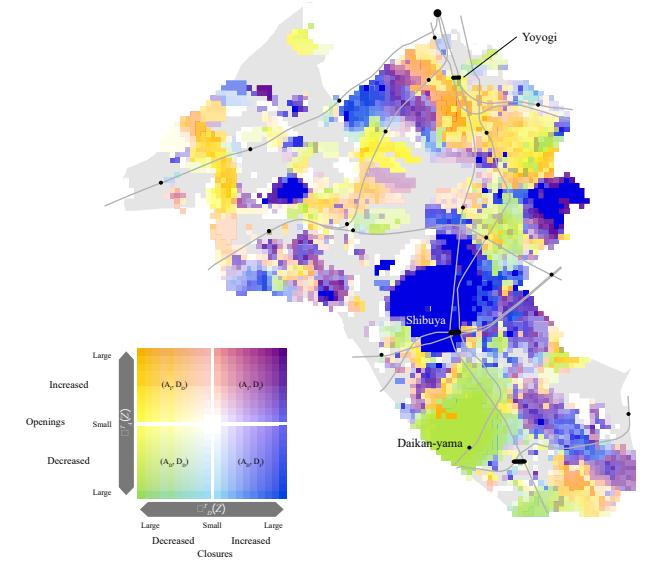


図6 衣料品店の $\alpha_A^T(\mathbf{x})$ と $\alpha_D^T(\mathbf{x})$ の分布

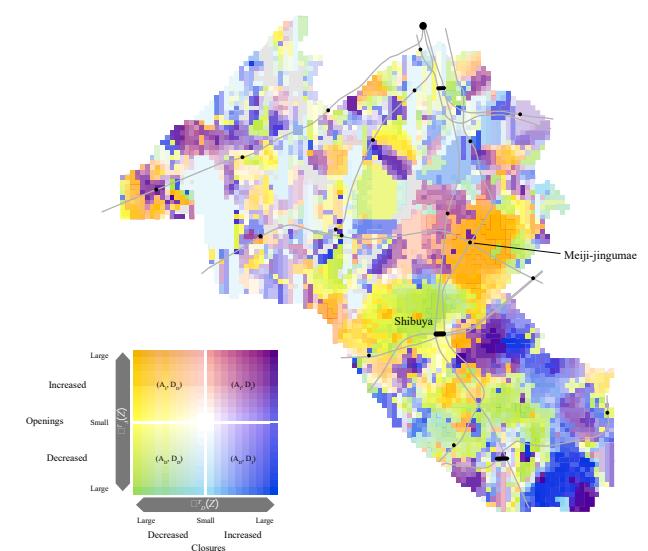


図7 飲食店の $\alpha_A^T(\mathbf{x})$ と $\alpha_D^T(\mathbf{x})$ の分布

謝辞

本研究ではNTTタウンページ(株)からデータをお借りしました。ここに記して謝意を表します。

参考文献

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