# 時系列空間データの探索的解析手法

貞広 幸雄

本論文では、地点ごと、あるいは地区ごとに記録された時系列空間データの新たな解析手法を提案する.気温、気圧、風速などの気象データ、植生、土壌、土地被覆などの地理データ、人口、就業者、通勤・通学などの社会データなど、同一地点や同一地区で継続的に記録される空間データは数多い.本手法は、同種のデータを地点・地区間で比較することにより、相互の局所的類似性を抽出、類似データをまとめることでその空間分布特性を検討する.この手法をユタ州ソルトレーク郡の人口分布変動パターンの分析に適用し、その有効性を検証する.

キーワード:時系列空間データ,探索的解析手法,局所的類似性,グラフ

#### 1. はじめに

近年,空間データの整備が進み,時系列データの利 用可能性が大きく広がってきている.気温,気圧,風 速などの気象データ,植生,土壌,土地被覆,土地利 用などの地理データ,国勢調査,商業統計調査,住宅・ 土地統計調査などの社会経済データなど,さまざまな 時系列データが空間データとして整備されており,イ ンターネットで容易に入手可能なものも多い.公的機 関だけでなく,民間企業より提供される空間データに ついても,それを蓄積して時空間分析に活用できる場 合も少なくない.

この種のデータの分析は,通常,その視覚化から始 める.時点ごとの空間データ,および,地点ごとの時 系列データを視覚的に分析することで,顕著なパター ンを抽出し,研究仮説を構築する.しかしながら,こ うした視覚的分析は,データ量が大きい場合には必ず しも効率的ではなく,特に,近年見られる膨大な時空 間データを扱うのは事実上,不可能である.

そこで本稿では、このような時系列空間データを探 索的に解析する、新たな手法の提案を行う、詳細なレ ビューは他稿 [1] に譲るが、この手法は、既存手法と 比較して柔軟性が高く、また、計算効率も良いという 利点を持つ.ここでは手法の提案と共に、データの適 用事例を示すことで、その有効性を確認する.

#### 2. 解析手法

いま, M 地点において期間  $[T_s, T_F]$  にわたって時

さだひろ ゆきお 東京大学 空間情報科学研究センター 〒 113-8656 東京都文京区本郷 7-3-1 系列データが得られているものとする. 地点 i における時系列データを  $Z_i$  と表記し,時刻 t での値を関数  $f_i(t)$  で表す.

#### 2.1 前処理

 $Z_i$ の<u>近隣</u> $N_i$ を、 $Z_i$ を垂直方向にbだけ正負両方 向に移動して得られる領域として定める(図 1). そ して、すべての近隣の重ね合わせから生ずる各小領 域をポリゴンと呼び、ポリゴンすべての集合を  $\Lambda =$  $\{P_1, P_2, \ldots, P_K\}$ と表す.ポリゴン  $P \in \Lambda$ の始点 および終点をそれぞれ  $t_S(P), t_E(P)$ と記す.ポリゴ ンの部分集合  $Q \subseteq \Lambda$ の始点および終点をそれぞれ  $t_S(Q) = \min_{P \subseteq Q} t_S(P), t_E(Q) = \max_{P \subseteq Q} t_E(P),$ Qの長さを

#### $l(Q) = t_E(Q) - t_S(Q)$

とする. 近隣 $N_i \varepsilon$ , ポリゴンの集合 $\vartheta_i$  として表し, そ の集合を  $\vartheta = \{\vartheta_1, \vartheta_2, \dots, \vartheta_M\}$  と書く.

#### 2.2 類似データの検出

次に,時系列データのなかから,類似したものを検 出する.類似した時系列データの近隣は,一般に広範 囲にわたって重なる.したがって,多くの近隣が重な



**18** (18) Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

る領域を見つけることで、類似した時系列データを検 出できる.ここではそのような、多数の近隣の重なる 領域を<u>核</u>と呼び、その集合を $\Omega = \{C_1, C_2, ..., C_N\}$ と表す. $C_i$ を完全に内包する近隣の集合と、それらの 集合をそれぞれ $\Gamma_i, \Gamma = \{\Gamma_1, \Gamma_2, ..., \Gamma_N\}$ と書き、核、 それらを内包する近隣、近隣の元の時系列データは、そ れぞれ相互に関連づけられると言う.

核の検出は、以下の手順による(図 2).まず、最も 多くの近隣の重なるポリゴン  $P_{11}$ を選択し、核を表すポ リゴン集合 $\Theta$ の元とする、次に、順にその隣接領域、例 えば  $\{P_{21}, P_{22}\}$ ,  $\{P_{31}, P_{32}, P_{33}\}$ ,  $\{P_{41}, P_{42}, P_{43}, P_{44}\}$ という具合に、選択領域を拡大する、ただし、拡大が 水平方向に進行するように、以下のような基準で隣接 領域を選択する.

最初に、図 2(b) にあるように、すべてのポリゴン を $x = t_s(\Theta)$  (図中の太線) と $x = t_E(\Theta)$  で切断し、  $P_{11}$  に隣接するポリゴン ( $P_{21}$ ' と  $P_{22}$ ") を考える. そ れらの元のポリゴンの長いほう  $P_{21}$  を集合  $\Theta$  に加え、  $\Theta$  のすべての元を内包する近隣の集合を  $\Psi$  とする. 次 に、すべてのポリゴンを  $x = t_s(\Theta)$  (図中の太点線) と $x = t_E(\Theta)$  で切断し、 $\Theta$  の元である  $P_{11}$  あるいは  $P_{21}$  に隣接するポリゴンを考える. それらの元のポリ ゴンの中で最も長いものを集合  $\Theta$  に加え、 $\Psi$  のうち、  $\Theta$  の元を一つでも内包しないものを除去する. この操 作を、核  $\Theta$  が既定値  $L_{min}$  よりも長くなるまで続け、 その結果得られる  $\Theta$  が  $C_1$ 、 $\Psi$  が  $\Gamma_1$  となる.

以上のように  $C_1$  が得られた後,切断したポリゴンは すべて元のとおりに接合し,また, $C_1$  に含まれるポリ ゴンを, $\Gamma_1$  に含まれるすべての近隣を構成するポリゴ ン集合から除去する.そして,再度,同様の手順を繰り 返し,最も多くの近隣の重なるポリゴンの選択段階に おいて,重なるポリゴンの個数が既定値  $\alpha$  よりも小さ くなった段階で終了する.その結果,核および近隣の 集合  $\Omega = \{C_1, C_2, ..., C_N\} \ge \Gamma = \{\Gamma_1, \Gamma_2, ..., \Gamma_N\}$ を得る.

同一の核を共有する時系列データは、少なくともそ の核の範囲では、互いに局所的に類似していると言う ことができる.従来の手法が、時系列データ相互の全 体的類似性を評価するのに対し、本手法は、核という 概念を通じて局所的類似性に基づく評価を行うという 特徴を有しており、時系列データのより柔軟な解析が 可能である.

上記手順では、あらかじめ 3 つのパラメータ b,  $L_{\min}, \alpha$  を定める必要がある.この種のパラメータは、 探索的分析の場合、できるだけ多くの値を試みて結果





 図 2 図 1 右上部分における核の検出手順.
(a) 核の拡大順序. P<sub>11</sub>, P<sub>21</sub>, P<sub>31</sub>, P<sub>41</sub>, の順に核 が拡大する. P<sub>41</sub> が加わった段階で l(Θ)>L<sub>min</sub> とな り, 核 C<sub>1</sub> の検出が完了する. (b) ポリゴンの切断.

を比較することが望ましい.しかし一般には,そのような網羅的な試行は現実的に困難であり,ある程度,選 択の指針が必要である.

パラメータ b は、時系列データの類似性の定義を与 えるものであり、大きな値にすると、多くの時系列デー タが類似していると見なされることになる。その結果、 近隣の重複も増加し、多くの核が検出される.他方、大 きな b は時系列データの微少な変動を隠してしまうた め、過大な値は必ずしも望ましくない、少なくとも、時 系列データの最小振れ幅

$$y_f = \min_{i \in \{1, \dots, M\}} \left\{ \max_{t \in [T_S, T_E]} f_i(t) - \min_{t \in [T_S, T_E]} f_i(t) \right\}$$

よりも十分に小さい必要がある.現実的には、上記 y<sub>f</sub> の 5-10%程度の値から始め、徐々にその値を大きくし て、十分な個数の核を検出した時点で終了するという 方法が現実的であろう.

パラメータ  $L_{\min}$  と  $\alpha$  は、いずれも検出される核の 重要性を決定づける、大きな値にすると、長く、多く の時系列データの関連づけられた核のみが抽出される。 それらは分析上は大変重要であるが、反面、ほとんど 核が抽出されないということもあり得る.したがって 分析の最初は小さな値、例えば、 $\alpha = 0.001 \times M$ およ び $L_{\min} = 0.05 \times (T_E - T_S)$ などから始め、妥当な数 の核が抽出されるようになるまで徐々に値を大きくす るのが望ましい.

なお、多数の時系列データを扱う場合、近隣の重ね 合わせの結果、膨大な数のポリゴンが生ずることがあ る.この場合、上記手順は実用時間内では終了しない という問題が発生する、それに対処するためには、あ らかじめ空間を離散化し、ラスターに基づいた核の抽 出が有効である.m×nの格子網を用いた場合、その 計算量は O(Mmn) となる(詳細は文献 [1] を参照の こと).

### 2.3 時系列データの分類

上記手順では、集合 ⊖ に含まれるポリゴンは増加す る一方、集合 Ψ に含まれる近隣は減少する.集合 Ψ からは、類似性の低い近隣が順に除去されることから、 除去の順序が類似性の高低を反映する.したがってこ の作業を、それぞれの核の検出後も続けることで、核 に関連づけられるすべての時系列データを分類するこ とができる.

例えば前述の図 2(b) の例では、 $C_1 = \{P_{11}, P_{21}, P_{31}, P_{41}\}$ であるが、この核に関連づけられる時系列 データは  $Z_1, Z_2, Z_3, Z_6$ である.これらの近隣は、  $N_1, N_2, N_6, N_3$ の順に集合  $\Psi$  から除去される.した がって、 $\{Z_2, Z_6, Z_3\}$ は  $Z_1$ よりも相互類似性が高く、  $\{Z_6, Z_3\}$ は  $Z_2$ よりも相互類似性が高いと言える.こ のことから、 $Z_1, Z_2, Z_3, Z_6$ は  $\{\{Z_1\}, \{Z_2, Z_6, Z_3\}\}$ 、  $\{\{Z_1\}, \{Z_2\}, \{Z_6, Z_3\}\}, \{\{Z_1\}, \{Z_2\}, \{Z_6\}, \{Z_3\}\}$ な どと分類することができる.しかし、元の順序に矛盾 する  $\{\{Z_1, Z_6\}, \{Z_2, Z_3\}\}$ や  $\{\{Z_1, Z_3\}, \{Z_2, Z_6\}\}$ な どの分類は認められない.

複数の核が検出された場合,時系列データが複数の 核に同時に関連づけられる可能性がある.これは,核 が時系列データの局所的類似性を示すものであるとい う性質による.ただし,このような重複分類が望まし くない場合には,手順を若干修正することで,重複を 禁止することも可能である(文献 [1]).

## 2.4 核,時系列データ,ポリゴンの相互関係の 可視化

上の手順で抽出された核,時系列データ,ポリゴンの相互関係を可視化するために,ここではグラフ構造 に基づいた,位相図(topology diagram)を応用した 方法を提案する.位相図とは,ハッセ図([6,7])の部



 図 2 における核の抽出過程を表す位相図.
(a) 集合 Θ に追加されるポリゴン群の順序, (b) 核 C<sub>1</sub> と, それに関連づけられる時系列データおよびポ リゴンの相互関係.実線は集合 Θ へのポリゴンの追 加, 点線は集合 Ψ からの時系列データの除去,太線 は集合 Θ の拡大を表す.

分集合を修正したものであり,文献 [2,3,4,5] でその 具体例が提案されている.ハッセ図では、ノードは核, 近隣で示される時系列データ、ポリゴンを表す.近隣が ポリゴンを内包する時系列データは、そのポリゴンと 直接あるいは間接的にエッジによって結合される.縦 軸は各空間オブジェクトの長さを表す.

位相図は一意に定まるわけではない. 典型的な位相 図は, それぞれの核ごとに, その抽出過程をそのまま グラフとして図化したものである. 図 3(a) は, 核  $C_1$ の抽出におけるポリゴンの追加順序を表す. ポリゴン  $\{P_1, P_2, P_3, P_4, P_5, P_6, P_7\}$ はこの順に集合 $\Theta$ に追加さ れ, 並行して時系列データ $\{Z_4, Z_8, Z_7, Z_1, Z_2, Z_3, Z_6\}$ の近隣が集合  $\Psi$  から除去される. 図 3(b) において, 前者は実線, 後者は点線, 集合 $\Theta$ の拡大は太線でそれ ぞれ表される.

位相図は、核、時系列データ、ポリゴンの位相関係 を理解するのに有用なだけではなく、前述した時系列 データの分類にも利用可能である.図 3(b)の位相図に おいて、エッジ  $\{E_1, E_2, E_3, E_4, E_5, E_6\}$ のうちいくつ かを切断すると、時系列データもいくつかの部分グラフ に分割されるが、それがそのまま自然な分類を定める. 例えば  $E_3$ を切断すると、時系列データは  $\{Z_4, Z_7, Z_8\}$  と  $\{Z_1, Z_2, Z_3, Z_6\}$  という 2 つに分類される.

## 3. 実証分析

前節までに提案した方法を,ここではユタ州ソルト レーク郡の昼夜間人口分布変動パターンの分析に適用す る.人口分布データは,543のTAZ (Transportation Analysis Zone) ごとの人口変動を,住民の居住地と就 業地データに基づいて推計したものであり,アメリカ 交通省によってWeb上で公開されている[8].人口変 動の詳細な推計は,日本でもNTTドコモのモバイル 空間統計など,さまざまな方法で行われるようになっ てきており,この種のデータの分析は今後さらに進む ものと思われる.

図4はソルトレーク郡の午後0時と午前0時の人口 分布である.昼間は中心市街地のほかに、ソルトレー ク空港とユタ大学に人口が集中していることがわかる. 543のデータのうち、半数以上は人口が夜間に増加、





**A** 4 フルドレーク部の重複间八 (a) 午後時, (b) 午前時.

昼間に減少という,住居地域の変化傾向を示している. 反対に,昼間人口が夜間人口よりも多い場合としては, 図 5 の例が挙げられる(図 5(a)中,縦軸は各 TAZ の人口を表す.以下,図 6(a),7(a),8(a),8(b)も同 様).これらは, b=0.004の場合に検出された核に関 連づけられた時系列データの一部(黒線)と,それら の観測された TAZ の分布であり,13の核と,それら に関連づけられた404の時系列データを示す.図 5(b) より,これらの地域はソルトレーク郡の中心市街地と 空港付近に多いことがわかる.

図 6 は, b =0.006 の場合に検出された核に関連づけ られた時系列データと、それらの観測された TAZ の 分布である. 黒線の地域では、昼夜間人口比が他地域 ほど極端ではなく、住宅地と就業地のいずれも含む地 域であることが示唆される. 実際、図 6(b) にはユタ 大学が含まれているが、大学内には学生向けの寮や教 職員向けの宿舎なども整備されており、昼間人口と比 べて夜間人口はそれほど大きく減少するわけではない.

図7は、b=0.006の場合に、いずれの核にも関連づけられなかった時系列データである。昼間人口の増加と、夜間人口の減少が他地域と比べて緩やかであり、かつ、遅めの時間帯であることが図より読み取れる。こ





**図5** b = 0.004 の場合に検出された核に関連づけられた 時系列データ(a, 濃灰色)と、それらの観測された TAZ の分布(b, 灰色).



**図6** b = 0.006 の場合に検出された核に関連づけられた 時系列データ(a, 濃灰色)と、それらの観測された TAZ の分布(b, 灰色).



(a)



**図7** b = 0.006 の場合にいずれの核にも関連づけられな かった時系列データ(a, 濃灰色)と、それらの観測 された TAZ の分布(b, 灰色).

れらは図 7(b) より,主として郊外部に位置するショッ ピングモールやアウトレットストア,レストランなど, 就業時間が通常の業務地域と比べて遅い地域であるこ とがわかる.

図 8(a) は、 b =0.006 の場合に検出された核である. その多くは、人口変動の安定する昼間および夜間に集 まっている. 図中、黒色で示されている核に関連づけ



**図8** b = 0.006 の場合に検出された核と、その一つ(黒 色)に関連づけられた時系列データ.



図 9 図 8(a) の黒色で示されている核に関連づけられた時系列データ.
(a) 位相図,(b) 点線で囲まれた時系列データに対応する TAZ.

られた時系列データが図 8(b) である. この核に関す る位相図の上半分が図 9(a) であり, 点線で囲まれてい る TAZ と他の間に明確な差異が見いだされる. 前者 は図 9(b) に示される地域であり, これらはいずれも 郡の中心部に位置している. 図 8(b) を見ると, これ らの地域での人口変動は図中の他地域よりも緩やかで

**22** (22) Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

オペレーションズ・リサーチ

あり,通勤行動が徐々に行われていることを示唆して いる.都心部では居住費が高いことから,これらの地 域へは比較的遠方からの通勤が行われているのではな いかと思われる.

#### 4. 結論

本論文では、時系列空間データを探索的に解析する、 新たな手法を提案した.本手法の特徴の一つは、局所 的な類似性に基づいて時系列空間データを分類する点 であり、部分的にしか似ていないデータであっても、そ れらを相互に結びつけて一つの集合とし、その空間分 布を論ずることが可能となる。

時系列空間データの整備に伴い,このような探索的 分析手法に対する需要はますます高まってきている。 後さらに,新たな手法の開発が望まれる。

#### 参考文献

[1] Y. Sadahiro and T. Kobayashi, Exploratory analysis of spatially distributed time series data: Detection of similarities, clustering and visualization of mutual relations. Discussion Paper Series, 108, Department of Urban Engineering, University of Tokyo, 2012 (available from

http://ua.t.u-tokyo.ac.jp/pub/due-dp/108.pdf).

- [2] Y. Sadahiro, Analysis of the spatial relations among point distributions on a discrete space. International Journal of Geographical Information Science, 24, 997–1014, 2010.
- [3] Y. Sadahiro, Analysis of the relations among spatial tessellations. *Journal of Geographical Systems*, 13, 373–391, 2011.
- [4] Y. Sadahiro, Spatial relations among polygons: an exploratory analysis. Geographical Analysis, to appear (available from

http://ua.t.u-tokyo.ac.jp/pub/due-dp/102.pdf).

- [5] Y. Sadahiro, R. Lay, and T. Kobayashi, Trajectories of moving objects on a network: detection of similarities, visualization of relations, and classification of trajectories. Transactions in GIS, to appear.
- [6] G. Birkhoff, Lattice Theory (3rd Ed.), American Mathematical Society, 1979.
- [7] B. A. Davey and H. A. Priestley, Introduction to Lattice and Order, Cambridge University Press, 2002
- [8] U.S. Department of Transportation, Census Transportation Planning Products, 2000
- (http://www.fhwa.dot.gov/ctpp/).