

空間的に連続なデータのための固有ベクトル空間フィルタリングの拡張

村上大輔

Development of eigenvector spatial filtering for spatially continuous data

Daisuke MURAKAMI

Abstract Eigenvector spatial filtering (ESF) is a relatively new technique that considers spatial autocorrelation. It is easy to implement and extend, thus expanding its scope of application in the future. However, ESF is restrictive in that it cannot consider continuity of space. In this study, we extend ESF so as to consider the continuity of space. To confirm the effectiveness of our method, our method, linear regression model, and kriging, a geostatistical method, are compared using a case study.

Keywords: 固有ベクトル空間フィルタリング (Eigenvector spatial filtering) , 空間的相関 (Spatial autocorrelation) , 空間内挿 (Spatial interpolation) , 通常最小二乗法 (Ordinary least squares)

1. はじめに

空間的相関 (空間的に近接したデータは類似した傾向を示すという空間データの基本的性質) を考慮することのできる比較的新しい手法に固有ベクトル空間フィルタリング (ESF : Eigenvector Spatial Filtering ; Griffith, 1996) がある。

ESF の最大の長所は簡便性である (Griffith, 2003). ESF の基本モデルは線形回帰モデルと同一であり, 標準的な統計パッケージを用いて実装できる. また, 一般化線形モデル, 混合効果モデルといった線形回帰モデルに基づいたモデルが容易に援用できる.

一方で, ESF は, 隣接の有無に基づいて空間的相関をモデル化しなければならならず, 空間上の任意地点で観測可能なデータの空間的な連続性が考慮できない. 従って, ESF は, 空間統計学を中心に研究の盛んな空間的に連続なデータ補間 (例: 気温分布図の作成) への適用もまた, 一般にできない. ま

た, 固有ベクトルを算出するための計算負荷が大きいこともまた, ESF の短所である (Griffith, 2003).

以上を踏まえ, 本研究では, 計算負荷に配慮しながら ESF を連続空間に拡張する.

2. 固有ベクトル空間フィルタリング

空間的相関の検定統計量として, モラン I 統計量 (1)式が知られている.

$I_{moran} = \mathbf{y}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{y} / \mathbf{y}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{y}$ (1)

n はデータ数, \mathbf{y} は空間的相関を計測する変数のベクトル, \mathbf{I} は $n \times n$ の単位行列, $\mathbf{1}$ は 1 を要素を持つ $n \times 1$ のベクトル, \mathbf{C} は $n \times n$ の隣接行列である.

ESF は, モラン I 統計量の中で用いられる隣接行列 $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ \mathbf{C} $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ の固有ベクトル \mathbf{E}_i を用いて空間的相関を考慮する.

$$\{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_n\} = evec[(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)] \quad (2)$$

ここで, $evec[\quad]$ は $[\quad]$ 内の行列の固有ベクトルの演算子であり, \mathbf{E}_n は $n-1$ 番目までの全ての固有ベクトル $\{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_{n-1}\}$ と無相関, かつモラン I 統計量で説明される最も大域的な空間パターンを表す. ここ

で、 \mathbf{C} が隣接行列であることは、各固有ベクトルが無相関となるための必要条件である(Griffith, 2003).

ESFの基本モデルは(3)式で与えられる.

$$y_s = \sum_k x_{s,k} \beta_k + \sum_l E_{s,l} \gamma_l + \varepsilon_s \quad \varepsilon_s \sim N(0, \sigma^2) \quad (3)$$

ここで、 s は地点である。 k は説明変数、 l は固有ベクトルの添え字であり、 y_s 、 $x_{s,k}$ 、 ε_s は、地点 s の、被説明変数、説明変数 k 、攪乱項をそれぞれ表し、 $E_{s,l}$ は(2)式より与えられる固有ベクトル l の第 s 要素である。 β_k 、 γ_l 、 σ^2 はパラメータを表す。モデルに投入する固有ベクトルは、自由度調整済み決定係数の最大化や、残差のモラン I 統計量の最小化等に基づいて選択される(例えれば、Griffith, 2006).

3. 固有ベクトル空間フィルタリングの連続空間への拡張

3.1 固有値・固有ベクトルの解析解

X 軸方向に P 個、 Y 軸方向に Q 個のメッシュが敷き詰められた空間を考える(図1)。同空間上の隣接行列 \mathbf{C} を、隣接する4メッシュに等しい重み1を与えることで定義した場合、 \mathbf{C} の固有値は(4)式、固有ベクトルは(5)式でそれぞれ与えられる(Griffith, 2000).

$$\lambda_{p,q} = 2 \left[\cos\left(\frac{p\pi}{P+1}\right) + \cos\left(\frac{q\pi}{Q+1}\right) \right] \quad (4)$$

$$E_{s,p,q} = \frac{2}{\sqrt{(P+1)(Q+1)}} \sin\left(\frac{p_s p \pi}{P+1}\right) \sin\left(\frac{q_s q \pi}{Q+1}\right) \quad (5)$$

$p=\{1,2\dots P\}$, $q=\{1,2\dots Q\}$, $p_s=\{1,2\dots P\}$, $q_s=\{1,2\dots Q\}$ であり、それらの取り方に応じて \mathbf{C} の固有値と、固有ベクトルが与えられる。

ここで、(4)、(5)式は \mathbf{C} の固有値・固有ベクトルを表し、 $(\mathbf{I}-\mathbf{1}\mathbf{1}'/n) \mathbf{C} (\mathbf{I}-\mathbf{1}\mathbf{1}'/n)$ の固有値・固有ベクトルとは異なる。しかしながら、(6)式を用いることで、 \mathbf{C} の固有ベクトル $\mathbf{E}_{p,q}$ を、 $(\mathbf{I}-\mathbf{1}\mathbf{1}'/n) \mathbf{C} (\mathbf{I}-\mathbf{1}\mathbf{1}'/n)$ の固有ベクトル $\mathbf{E}_{p,q}^*$ に変換できる(Griffith, 2000).

$$\mathbf{E}_{p,q}^* \propto \begin{cases} (1/\sqrt{n})\mathbf{1} & \text{if } p=q=1 \\ k_{p,q}(\mathbf{I}-\mathbf{1}\mathbf{1}'/n)\mathbf{E}_{p,q} & \text{otherwise} \end{cases} \quad (6)$$

ここで、 \propto は比例を表す記号、 $k_{p,q}$ は $\mathbf{E}_{p,q}$ の標準偏差の逆数を表す定数である。

3.2 固有値・固有ベクトルの解析解の連続空間への拡張

本研究では、メッシュを極限まで細分化することで連続空間上を定義し、同空間上でESFを再定式化する。まず、 P, Q を極限まで大きくすることで $P^{(\infty)}$ 、 $Q^{(\infty)}$ する。 $P^{(\infty)} \times Q^{(\infty)}$ のメッシュ間について構築される隣接行列 $\mathbf{C}^{(\infty)}$ の固有値は(7)式、固有ベクトルは(8)式でそれぞれ与えられる。

$$\lim_{(P,Q) \rightarrow (\infty,\infty)} \lambda_{p,q} = \lambda_{p,q}^{(\infty)} = 2 \left[\cos\left(\frac{p}{P_\infty}\pi\right) + \cos\left(\frac{q}{Q_\infty}\pi\right) \right] \quad (7)$$

$$\lim_{(P,Q) \rightarrow (\infty,\infty)} E_{s,p,q} = E_{s,p,q}^{(\infty)} \propto \sin(p'_s p \pi) \sin(q'_s q \pi) \quad (8)$$

(8)式の $p'_s \{ = p_s / P^{(\infty)} \}$ は、図2に示す空間上の X 軸方向の総長に占める、 X 軸上の端点から地点 s の X 座標の間の長さの割合、 $q'_s \{ = q_s / Q^{(\infty)} \}$ は、 Y 軸についての同割合となる。

(6)、(8)式より、図2に示す連続空間上で定義された $(\mathbf{I}-\mathbf{1}\mathbf{1}'/n) \mathbf{C}^{(\infty)} (\mathbf{I}-\mathbf{1}\mathbf{1}'/n)$ の固有ベクトルの第 s 要素は(9)式で与えられる。

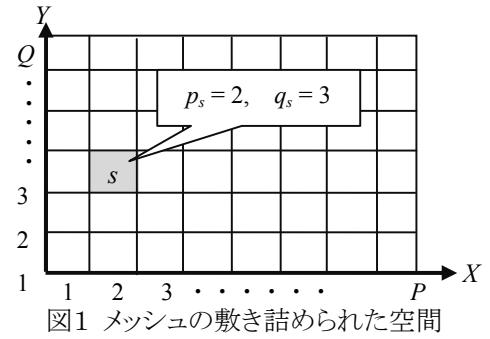


図1 メッシュの敷き詰められた空間

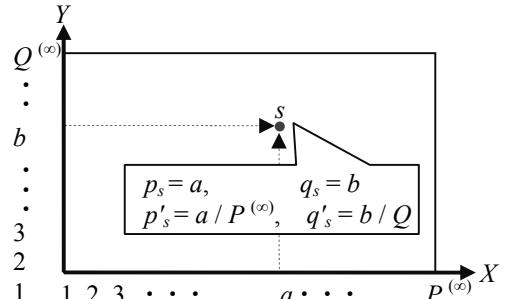


図2 メッシュの極限まで細分化された空間

$$E_{s,p,q}^{(\infty)*} \propto \begin{cases} (1/\sqrt{\infty})1 & \text{if } p = q = 1 \\ \sin(p_s' p \pi) \sin(q_s' q \pi) & \text{otherwise} \end{cases} \quad (9)$$

(9)式より得られる固有ベクトルは、長方形空間内（図2）の空間パターンを表し、各ベクトルは長方形空間内で直交、かつ無相関である。

3.3 モデルの構築

(9)式で与えられる非第一固有ベクトルをESFの基本式(3)式に代入することで、(10)が与えられる。

$$y_s = \sum_k x_{s,k} \beta_k + \sum_p \sum_q \sin(p_s' p \pi) \sin(q_s' q \pi) \gamma_{p,q} + \varepsilon_s$$

$$\varepsilon_s \sim N(0, \sigma^2) \quad (10)$$

従来のESFと同様、(10)式は通常線形回帰モデルと同一であり、実装性・拡張性に富む。また、固有ベクトルの算出を、解析解を用いて行うことを前提としており、計算量は小さい。固有ベクトルの選択には、(i)自由度調整済み決定係数やF値、(ii)モランI統計量といった基準に基づいたステップワイズ法が適用できる。

以降では、以上で提案した手法を固有ベクトル連続空間フィルタリング（ECSF：Eigenvector Continuous Space Filtering）と呼ぶこととする。

4. 提案手法の実データへの適用

4.1 分析の概要

ECSF、線形回帰モデル（Non-Spatial）、及びKrigingを地価関数の推定、及び地価の補間に適用する。

本実証では、2007年度住宅地公示地価データ、及び都道府県地価調査データを使用する。対象範囲は東京都及び埼玉、千葉、神奈川の3県及び茨城県南部であり（図3）、対象範囲内のデータ（地点）数は9,245である。

被説明変数は公示地価（万円/m²）の自然対数、説明変数は表1に示す各変数である。以上に加え、ECSFについては大きな固有値を持つ6,000の固有ベクトルを抽出し、説明変数の候補とした。

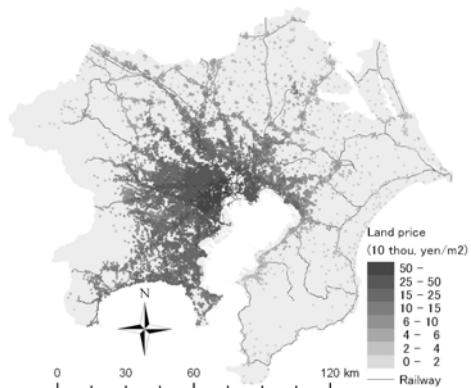


図3 地価の実測値

表1 説明変数の概要

説明変数	出典・年度	概要
東京距離	Yahoo! 路線情報 (2009)	最寄り駅から東京、新宿、池袋、渋谷、品川川のいずれかの駅までの最短所要時間（分）
駅距離	国土数値情報 (2007)	最寄り駅までの距離（km）
市街化区域		市街化区域であれば1を示すダミー変数
田	国土数値情報 (2006)	所属する3次メッシュ（約1km ² ）あたりの面積（m ² / km ² ）
その他農地		
森林		
荒地		
交通用地		
その他用地		
河川		
海浜		
海水域		
ゴルフ場		

4.2 パラメータ推定結果

ステップワイズ法とVIF診断に基づいた変数選択をECSFに適用した結果、707の固有ベクトルが有意であるとの推定結果を得た。

固有ベクトルを除く各説明変数のパラメータ推定結果を表2に示す。表中で有意となった各推定値の符号は直観と整合的である。一方で、Non-Spatialのみで有意となった各変数は、空間的相関を無視したことにより生じたバイアスである可能性がある。

4.3 補間結果

各モデルの予測精度をRMSEに基づいて検証し

表2 パラメータ推定結果（***は1%，**は5%，*は10%水準で有意であることを表す）

説明変数	Non-Spatial		Kriging		ECSF	
	推定値	t値	推定値	t値	推定値	t値
定数項	12.7	597 *** ¹⁾	11.3	548 **	11.3	746 ***
東京距離	-1.92×10 ²	-101 ***	-4.61×10 ³	-13.4 ***	-6.34×10 ³	-40.6 ***
駅距離	-6.01×10 ²	-23.1 ***	-1.06×10 ¹	-33.8 ***	-9.65×10 ²	-62.1 ***
市街化区域	4.55×10 ⁻¹	29.1 ***	5.91×10 ⁻¹	59.2 ***	6.00×10 ⁻¹	81.0 ***
田	-1.07×10 ⁻⁶	-33.9 ***	-2.92×10 ⁻⁷	-14.0 ***	-2.19×10 ⁻⁷	-14.2 ***
その他農地	-7.55×10 ⁻⁷	-23.4 ***	-2.34×10 ⁻⁷	-10.3 ***	-2.39×10 ⁻⁷	-14.5 ***
森林	-4.21×10 ⁻⁷	-13.2 ***	-2.70×10 ⁻⁷	-11.9 ***	-3.02×10 ⁻⁷	-19.3 ***
荒地	-9.32×10 ⁻⁷	-7.15 ***				
交通用地	5.90×10 ⁻⁷	3.64 ***			2.54×10 ⁻⁷	3.56 ***
その他用地	-4.87×10 ⁻⁷	-11.9 ***				
河川	-4.30×10 ⁻⁷	-8.92 ***	-2.48×10 ⁻⁷	-8.70 ***	-2.31×10 ⁻⁷	-10.7 ***
海浜	1.89×10 ⁻⁶	2.89 ***				
ゴルフ場	-3.33×10 ⁻⁷	-3.49 ***			-2.20×10 ⁻⁷	-5.32 ***
Partial-sill			2.15×10 ⁻¹			
Nugget			3.05×10 ⁻²			
Range			80.2			

た(表3). この表より, ECSF と Kriging の RMSE は Non-Spatial の半分程度であり, 空間的相関を考慮することによる補間精度の向上が確認できる. また, ECSF の補間精度は Kriging よりもわずかに上回っており, 補間手法としての ECSF の有用性が確認された. 最後に, Kriging と ECSF の両補間結果を図4, 5 に示す. 郊外部において, 両補間結果に一定の差異が確認できる.

参考文献

- Griffith, D. A. (1996) Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying georeferenced data. *The Canadian Geographer*, **40**, pp.351–367
- Griffith, D. A. (2000) Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra & Its Applications*, **321**, pp.95–112.
- Griffith, D. A. (2003) *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin: Springer-Verlag.
- Griffith, D. A. (2006) Hidden negative spatial autocorrelation, *Journal of Geographical Systems*, **8** (4), pp.335–355

表3 RMSEの算出結果(万円/m²の対数値)

	Non-Spatial	Kriging	ECSF
RMSE	3.79×10^{-1}	1.67×10^{-1}	1.56×10^{-1}

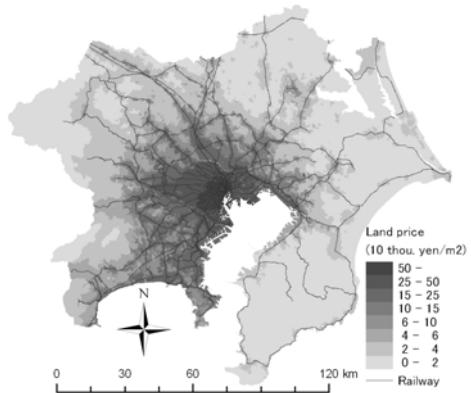


図4 Kriging の補間結果

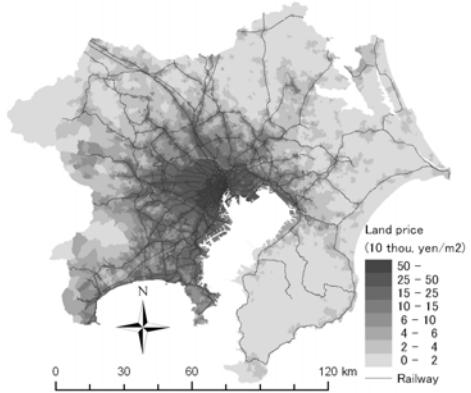


図5 ECSF の補間結果