

# 多項離散選択モデルへの固有ベクトル空間フィルタリングの適用

## 吉田崇紘・堤盛人

### Eigenvector-based Spatial Filtering Approach to a Multinomial Logit Model

#### Takahiro YOSHIDA and Morito TSUTSUMI

**Abstract:** It is vital to deal appropriately with spatial dependence, that is, when using regression models dealing with spatial data. Recent studies on spatial econometrics have suggested many methods to consider spatial dependence in quantitative data. However, these methods, termed spatial discrete choice models, typically require a computationally burdensome, iterative calculation for parameter estimation. In addition, spatial “multinomial” discrete choice models are still being finalized, and these methods are not well established. Hence, the present study applies another approach, namely eigenvector-based spatial filtering, in which spatial dependence can be considered only by introducing eigenvectors of a modified spatial weight matrix as explanatory variables, to a multinomial logit model and we compare its predictive accuracies to that of the conventional model.

**Keywords:** 空間依存性 (spatial dependence), 空間多項離散選択モデル (spatial discrete choice model), 固有ベクトル空間フィルタリング (eigenvector-based spatial filtering)

#### 1. はじめに

近年の空間計量経済学における重要なトピックのひとつに、空間依存性を考慮した質的データのモデリングがある。Pinkse and Slade (2010) は、空間依存性を空間ラグパラメータによって直接構造化する空間ラグモデルや空間エラーモデルを離散選択モデルに拡張した“空間離散選択モデル”が、空間依存性の取り扱いに理論的側面からの検証が必要であると述べている。Smirnov (2010) では、空間離散選択モデルが実用面で抱える問題として、パラメータ推定に繰り返し計算を要するため、高い計算負荷が掛かることを挙げている。また、Smirnov (2010) は、特に空間“多項”離散選択モデルにおける理論、及び推定法は、ま

だ十分に議論されていないとしている。

一方、近年、空間依存性を考慮する別のアプローチとして、空間統計学の代表的な手法の一つである空間フィルタリング、中でも、たとえば Griffith (2003) によって提案がなされた固有ベクトル空間フィルタリング (ESF) 法を用いた質的データのモデリングが注目されている。ESF 法は、説明変数に隣接性を表す行列から抽出した固有ベクトルを追加するだけで空間依存性が考慮できるため、実用性を長所にもつ。また、ESF 法において固有ベクトルで表現される空間依存性は、除外変数の代理変数として解釈可能であるため、空間依存性の取り扱いが明白である。このような利点から、ESF 法の適用は近年広がりを見せており、さまざまなモデルに追加した際に生じる ESF 法の問題や利点の検証など、現在はその知見を蓄えることが求められている。特に、これまでに ESF 法を多項離散選択モデルに適用した例は、筆

---

吉田崇紘： 〒305-8573 茨城県つくば市天王台 1-1-1

筑波大学大学院 システム情報工学研究科

Tel.: 029-853-5572

E-mail: yoshida.takahiro@sk.tsukuba.ac.jp

者らの知る限り存在しない。

以上のような背景から、本研究では、空間依存性を考慮した質的データのモデリングにおいて、空間計量経済学で議論されている空間離散選択モデルとは別のアプローチとして、ESF 法の多項離散選択モデルへの適用を提案するとともに、適用に当たっての問題について考察する。

## 2. 多項ロジットモデルへの ESF 法の適用

### 2.1 固有ベクトルの算出

空間依存性の代表的な検定統計量として、Morans'  $I$  統計量 ( $MI$ ) が広く知られており、

$$MI = \frac{n}{\mathbf{1}^T \mathbf{C} \mathbf{1}} \frac{\mathbf{y}^T (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) \mathbf{C} (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) \mathbf{y}}{\mathbf{y}^T (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) \mathbf{y}} \quad (1)$$

と定義される(式(1))。ここで、 $n$  はサンプル数、 $\mathbf{1}$  は 1 からなる  $n \times 1$  のベクトル、 $\mathbf{y}$  は空間依存性を検定する変数の  $n \times 1$  のベクトル、 $\mathbf{I}$  は  $n \times n$  の単位行列、 $\mathbf{C}$  は  $n \times n$  の近接行列(なお、本研究では隣接行列)である。

ESF 法の固有ベクトルは、 $MI$  の近接性を表す行列である  $(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) \mathbf{C} (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)$  の固有ベクトルであり、サンプルの数だけ算出される(式(2))。

$$\{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_n\} = \text{eigenvector}\{(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n) \mathbf{C} (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)\} \quad (2)$$

ここで、 $\text{eigenvector}\{\bullet\}$  は、行列  $\bullet$  の固有ベクトルを取り出す演算子である。

### 2.2 モデルに投入する固有ベクトルの選出

Griffith (2003) は、substantial な空間依存性を考慮するには、式(2)で算出された固有ベクトルの中から、最大の固有値をもつ固有ベクトルの  $MI$  の 0.25 倍以上 ( $MI_{Max*0.25}$  基準) の  $MI$  をもつ固有ベクトルを説明変数の候補にすべきとしている。しかし、単純に説明変数を過大に増やすことは過剰適合 (over-fitting) を引き起こすためモデル特定上好ましくない。したがって、自由度調整済み決定係数の最大化や、残差の  $MI$  の最小化に基づいたステップワイズ法などによる変数選択(村上、

2012) を行うとともに、候補となる固有ベクトルの数も調整する必要がある。筆者らの従前の研究では、後述する提案モデルにおいて、 $MI_{Max*0.25}$  基準を満たす固有ベクトルを候補に AIC 最小化基準に基づく変数選択を行った結果、的中率が 100% となった。そこで本研究では、最大固有値をもつ固有ベクトルの  $MI$  の 0.50 倍以上 ( $MI_{Max*0.50}$  基準) の  $MI$  をもつ固有ベクトルを変数選択の候補とした。なお、Griffith (2003) は、 $MI_{Max*0.50}$  基準では、weak な空間依存性を考慮できないが、moderate な空間依存性は考慮可能としている。

### 2.3 ESF 法を適用した多項ロジットモデル

空間依存性を ESF 法によって考慮する多項ロジットモデルの基本モデルは、

$$p_{ij} = \Pr(y_i = j) = \frac{\exp(\mathbf{X}_i^T \boldsymbol{\beta}_j + \mathbf{E}_i^T \boldsymbol{\gamma}_j)}{\sum_{j=0}^J \exp(\mathbf{X}_i^T \boldsymbol{\beta}_j + \mathbf{E}_i^T \boldsymbol{\gamma}_j)} \quad (3)$$

$$i = 1, 2, \dots, n$$

$$j = 0, 1, \dots, J$$

$$\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0 = \mathbf{0}$$

と表現できる。ここで、 $y$  は  $n \times 1$  の多値の被説明変数ベクトル、 $i$  は地点、 $j$  は選択肢 ( $j=0$  は参照選択肢)、 $p_{ij}$  は  $y_i$  が  $j$  を選択する確率、 $\mathbf{X}_i = (1 X_{2i} \dots X_{ki})$  は  $n \times k$  の説明変数行列、 $\mathbf{E}_i = (E_{1i} E_{2i} \dots E_{mi})$  は  $n \times m$  の固有ベクトルの行列、 $\boldsymbol{\beta}_j$  及び  $\boldsymbol{\gamma}_j$  は  $j$  を選択したときのそれぞれ  $k \times 1$ 、 $m \times 1$  の係数ベクトルである。尤度関数  $L$  は

$$L = \prod_{i=1}^n \prod_{j=0}^J p_{ij}^{d_{ij}}, \quad d_{ij} = \begin{cases} 1, & y_i = j \\ 0, & y_i \neq j \end{cases} \quad (4)$$

となるため、パラメータは、上式(4)に対して両辺に対数をとった対数尤度関数に最尤法を行うことで得られる。ESF 法を適用した多項ロジットモデルは、説明変数に式(2)で算出した固有ベクトルを追加するだけであるため、式(3)から  $\mathbf{E}_i^T \boldsymbol{\gamma}_j$  項を除いた式は、通常の多項ロジットモデルと同一である。

### 3. ケーススタディ：土地利用推定

#### 3.1 分析の概要

本章では、式(3)で表現される提案モデル(spatial モデル)と通常の多項ロジットモデル(Conventional モデル)の推定精度を比較する。

分析の対象には、土地利用データを用いる。土地利用データは、各カテゴリー間、または、各カテゴリーの選択確率に空間依存性が存在すると考えられるため、代表的な質的空間データとして、空間多項離散選択モデルとその推定法の開発、発展の実証データに利用されてきた(たとえば、Chakir and Parent, 2009)。そこで本研究においても、既往研究に倣い、土地利用データを対象に分析を行う。

#### 3.2 使用データ

表-1に使用する変数の概要を示す。被説明変数に関して次に補足する。H21年度の土地利用3次メッシュデータは、各メッシュに12のカテゴリーそれぞれの面積が格納されているが、本研究では、各メッシュ内で最大の土地利用を代表値として、5つのカテゴリー(Paddy Field, Other Agri. Land, Forest, Land for Building, Other land)に該当したメッシュのみを抽出し、被説明変数とした(図-1)。

#### 3.3 説明変数に追加する固有ベクトル

Spatial モデルの説明変数に追加する固有ベクトルの候補基準は、前述したように、 $MI_{Max*0.50}$ 基準とした。 $MI_{Max*0.50}$ 基準を満たす固有ベクトルは、448個あり、全サンプル数、及び全固有ベクトル数  $n = 5614$  の7.98%であった。Spatial モデルでは、候補となった固有ベクトルは全て追加した。

#### 3.4 推定結果の比較

Conventional モデル、Spatial モデルの推定結果から予測される土地利用図をそれぞれ図-2、図-3に示す。表-2は、被説明変数と2つのモデルの推定結果の対応を比較しており、対角項が的中した土地利用区分の数と割合を示している。全体の的中率は、それぞれ76.93%、89.19%であった。ま

表-1 使用する変数の概要

Variable	Description	Source of data
Land Use	土地利用区分 (Paddy Field: 1, Other Agri. Land: 2, Forest: 3, Land for Building: 4, Other Land: 0)	土地利用3次 メッシュ、NLNI
Dens. POP	人口密度 [単位: 人/km <sup>2</sup> ]	Census 2005
Ave. SLOPE	平均傾斜度 [単位: 度]	標高・傾斜度3次 メッシュ、NLNI
Dist. Sta.	メッシュ重心点から最寄駅までの直線距離 [単位: km]	鉄道時系列、 NLNI
GEOM2	微地形区分が山麓地: 1, otherwise: 0	
GEOM3	微地形区分が丘陵: 1, otherwise: 0	
GEOM8	微地形区分が砂礫質台地: 1, otherwise: 0	
GEOM9	微地形区分がローム台地: 1, otherwise: 0	
GEOM10	微地形区分が谷底低地: 1, otherwise: 0	
GEOM12	微地形区分が自然堤防: 1, otherwise: 0	JEGCM
GEOM13	微地形区分が後背湿地: 1, otherwise: 0	
GEOM15	微地形区分が三角州・海岸低地: 1, otherwise: 0	
GEOM16	微地形区分が砂州・砂礫州: 1, otherwise: 0	
GEOM17	微地形区分が砂丘: 1, otherwise: 0	
GEOM18	微地形区分が干拓地: 1, otherwise: 0	

NLNI: 国土数値情報。Census: 国勢調査。  
JEGCM: 日本の地形・地盤デジタルマップ(若松ら, 2005)。

た、2つのモデルの尤度比( $LR$ )検定を行った結果、 $LR = 8443.4 > \chi^2_{0.05}(448) = 498.3$ となり、5%水準で有意であった。

図-1~3、表-2より、Conventional モデルは、地理的条件から土地利用が決定されやすい、特に被説明変数が Forest であるメッシュの的中率は85.91%と良好であるが、社会経済的条件から土地利用が決定されやすい Other Land のメッシュの的中率は5.99%と良くないといえる。

一方、Spatial モデルは、サンプル数に対して7.98%分の空間依存性を示す説明変数を追加しただけで、conventional モデルと比較してモデル全体の的中率が12.26%の上昇を示している。そして、 $MI_{Max*0.50}$ 基準を満たす固有ベクトルは、Other Land に顕著である局所集積のある土地利用区分に対する各選択肢の選択確率の空間依存性の考慮を可能にし、その結果、Other Land の推定精度は81.44%の向上を示した。

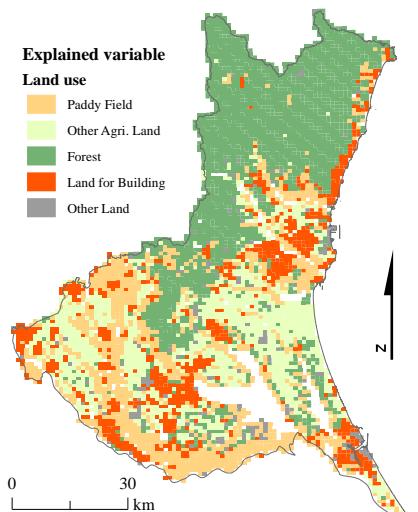


図-2 H21 年度の土地利用図

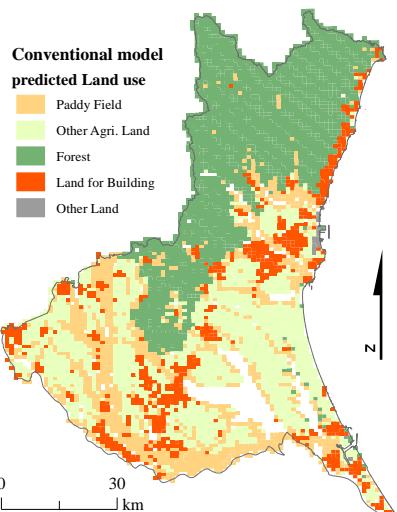


図-3 Conventional モデルの推定結果

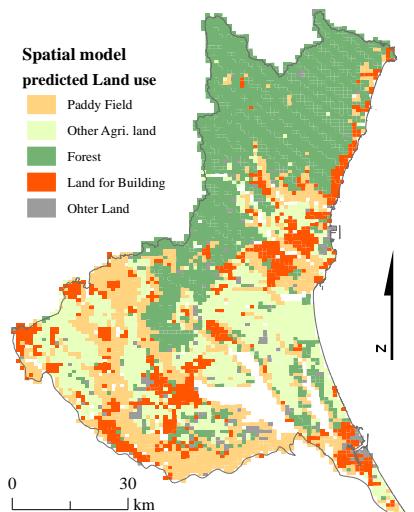


図-4 Spatial モデルの推定結果

表-2 推定結果の的中率の比較

Obs. \ Pred.	Paddy Field	Other Agri. Land	Forest	Land for Building	Other Land	Total
upper: Conventional						
lower: Spatial						
Paddy Field	901 (71.51%)	228 (18.10%)	84 (6.67%)	47 (3.73%)	0 (0.00%)	1260 (100%)
	1075 (85.32%)	86 (6.83%)	60 (4.76%)	32 (2.54%)	7 (0.56%)	1260 (100%)
Other Agri. Land	129 (11.01%)	936 (79.86%)	63 (5.38%)	44 (3.75%)	0 (0.00%)	1172 (100%)
	82 (7.00%)	1005 (85.75%)	49 (4.18%)	31 (2.65%)	5 (0.43%)	1172 (100%)
Forest	55 (2.50%)	233 (10.59%)	1890 (85.91%)	18 (0.82%)	4 (0.18%)	2200 (100%)
	35 (1.59%)	78 (3.55%)	2076 (94.36%)	8 (0.36%)	3 (0.14%)	2200 (100%)
Land for Building	56 (6.87%)	155 (19.02%)	19 (2.33%)	582 (71.41%)	3 (0.37%)	815 (100%)
	40 (4.91%)	57 (6.99%)	7 (0.86%)	705 (86.50%)	6 (0.74%)	815 (100%)
Other Land	21 (12.57%)	87 (52.10%)	42 (25.15%)	7 (4.19%)	10 (5.99%)	167 (100%)
	7 (4.19%)	3 (1.80%)	7 (4.19%)	4 (2.40%)	146 (87.43%)	167 (100%)
Total	1162	1639	2098	698	17	5614
	1239	1229	2199	780	167	5614

#### 4. まとめ

本研究では、空間依存性を考慮した質的データのモデリングとして、ESF 法を多項ロジットモデルに適用したモデルを提案し、通常の多項ロジットモデルと推定精度の比較を行った。ESF 法の多項ロジットモデルへの適用の利点として、 $MI$  が小さい固有ベクトルを説明変数に追加することで、選択肢の局所的な集積地の的中率を向上できることを示した。

#### 謝辞

本研究を進めるにあたり、小荒井衛様・仲埜貴元様（国土地理院）より、ご教示頂いた。ここに記して心より感謝申し上げます。

#### 参考文献

- 村上大輔 (2012) : 固有ベクトル空間フィルタリングの連続空間への拡張, *GIS-理論と応用*, **20** (2), 1-12.
- 若松加寿江, 松岡昌志, 杉浦正美, 久保純子, 長谷川浩一 (2005) : 「日本の地形・地盤デジタルマップ」, 東京大学出版.
- Chakir, R. and Parent, O., 2009. Determinants of land use change: A spatial multinomial probit approach. *Papers in Regional Science*, **88** (2), 327-344.
- Grifith, D. A., 2003. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*, Springer.
- Pinkse, J. and Slade, M. E., 2010. The future of spatial econometrics. *Journal of Regional Science*, **50** (1), 103-117.
- Smirnov, O. A., 2010. Modeling spatial discrete choice. *Regional Science and Urban Economics*, **40**, 292-298.