

土地利用データを対象とした 多項ロジットモデルへの 固有ベクトル空間フィルタリングの適用

吉田 崇紘¹・堤 盛人²

¹学生非会員 筑波大学大学院 システム情報工学研究科 (〒305-8573 茨城県つくば市天王台1-1-1)

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

²正会員 筑波大学教授 システム情報系社会工学域 (〒305-8573 茨城県つくば市天王台1-1-1)

E-mail: tsutsumi@sk.tsukuba.ac.jp

空間データを扱う回帰モデルでは、空間データのもつ性質のひとつである空間的依存性に対処する必要がある。近年、空間計量経済学と空間統計学において、質的データに対し空間的依存性を考慮する手法が提案されている。一般に、空間計量経済学のアプローチは、推定に繰り返し計算が必要で計算負荷が高いという欠点を持つ場合が多い。一方、空間統計学のアプローチ、中でも固有ベクトル空間フィルタリング (ESF) は、説明変数に空間情報を元に抽出された変数を追加するだけで空間的依存性を考慮でき、標準的な統計パッケージへの実装が容易であるという実用的利点をもっている。本研究では、土地利用データを対象として、ESFを多項ロジットモデルに適用し、質的データにおける空間的依存性を考慮した手法の実用性と推定精度について、ESFの有効性を確認する。

Key Words : *spatial effect, eigenvector spatial filtering, land use data, multinomial logit model*

1. はじめに

位置情報をもつデータである空間データを扱う回帰モデルにおいては、空間データのもつ一般的性質である空間的影響 (spatial effect)、すなわち空間的依存性 (spatial dependence) と空間的異質性 (spatial heterogeneity) に対処することが重要である (Anselin, 1988¹⁾。

近年、空間計量経済学と空間統計学の分野において、質的データに対する空間的影響を考慮する方法の提案がなされてきている。両分野の手法を比較すると、空間計量経済学における手法は、推定に繰り返し計算が必要で計算負荷が高いという欠点をもつ場合が多い (たとえば、Chakir and Parent, 2009²⁾; Wang et al., 2012³⁾)。一方、空間統計学における手法、中でも空間フィルタリング・アプローチの固有ベクトル空間フィルタリング (eigenvector-based spatial filtering, ESF) は、説明変数に空間情報を抽出した変数を追加するだけであり、標準的な統計パッケージへの実装が容易であるという実用的な利点をもっている (Griffith and Paelinck, 2011⁴⁾)。

また、空間計量経済学のモデルが近接性を表す変数として距離を用いるのに対し、ESFはMoran's I 統計量の近

接行列の固有ベクトルを用いて空間的相関をモデル化する。このようなアプローチを用いることで、距離だけでは説明することのできない地域毎の空間構造も表現できる (村上, 2012⁵⁾)。

上に挙げた長所からESFは適用は広がりつつあり、現在はその利点の検証や、さまざまなモデルに追加した際の固有ベクトルの解釈の仕方など、現在はその知見を蓄積することが求められている。特に、多項離散選択モデルに適用した例は筆者らの知る限りその様な研究は見当たらず、固有ベクトルと推定されるそのパラメータの線形和の選択肢間の差異の解釈は不透明なままである。したがって本研究では、ESFを多項ロジットモデルに適用し、空間的依存性を考慮有無による推定精度やモデルの当てはまりについて検証することに加え、各選択肢ごとに算出される固有ベクトルとそのパラメータの線形和は、どのように解釈されるのか考察する。

以下では、まず、第2章でESFについて説明し、次に、第3章では、茨城県の土地利用データを対象とした実証分析を行う。最後に、第4章で分析の結果を要約し、多項ロジットモデルにおける固有ベクトルとそのパラメータの線形和の解釈について考察を行う。

2. 固有ベクトル空間フィルタリング (ESF)

空間的依存性の代表的な検定統計量として、Moran's I が用いられることが多い。Moran's I は、Griffith(2000)⁹ に倣うと、

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

と定義される。ここで、 n はサンプル数、 \mathbf{y} は空間的相関を検定したい変数の $n \times 1$ のベクトル、 \mathbf{I} は $n \times n$ の単位行列、 $\mathbf{1}$ は 1 からなる $n \times 1$ のベクトル、 \mathbf{C} は $n \times n$ の近接行列、 $(\mathbf{I}-\mathbf{1}\mathbf{1}'/n)$ は射影行列である。以下では、簡単のため、 \mathbf{C} は隣接行列として扱う。

ESF は、Moran's I の近接性を表す $(\mathbf{I}-\mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}'/n)$ の固有ベクトルを用いて、空間的依存性を考慮する手法である。 $(\mathbf{I}-\mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I}-\mathbf{1}\mathbf{1}'/n)$ の最大固有値をもつ固有ベクトル \mathbf{E}_1 は、Moran's I で説明可能な最も大域的な空間パターンを表し、 n 番目に大きな固有値をもつ固有ベクトル \mathbf{E}_n は、 $n-1$ 番目までのすべての固有ベクトルと直交かつ無相関であり、また、Moran's I で説明可能な n 番目に大域的な空間パターンを表す (Griffith, 2007⁷; 村上, 2012⁹)。

多項ロジットモデルにおけるESFの基本モデルは、

$$\mathbf{U}_j = \mathbf{X}\boldsymbol{\beta}_j + \mathbf{E}_r\boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}_j \quad (2)$$

で与えられる。ここで、 \mathbf{U}_j は選択肢 j を選んだ場合の効用の $n \times 1$ のベクトル、 \mathbf{X} は $n \times k$ の説明変数行列、 \mathbf{E}_r は近接行列から算出された固有ベクトルを要素にもつ $n \times r$ の行列、 $\boldsymbol{\beta}_j$ は $k \times 1$ 、 $\boldsymbol{\gamma}_j$ は $r \times 1$ のパラメータベクトルである。 $\boldsymbol{\varepsilon}_j$ は $n \times 1$ の誤差ベクトルであり、要素同士が互いに独立で、累積分布関数がタイプ I の極値分布 (対数ワイブル分布) に従う。

式(2)の \mathbf{E}_r 中の r 個の固有ベクトルは、予めMoran's I が最大の固有値をもつ固有ベクトルの 0.25 倍以上の固有ベクトルを抽出し、次にステップワイズ法などを用いてそれらの中から選定することで決定できる (Griffith, 2003⁸; 村上, 2012⁹)。

一般に空間計量経済学と空間統計学におけるモデルは、空間的依存性を考慮した結果として、共分散が距離の関数で与えられる。ゆえに、線形モデルにおいては、通常最小二乗法が望ましい推定量を得るための仮定のひとつである共分散が 0 の仮定を満たさないため、通常最小二乗法が適用できない。一方で、ESF は、分布形を仮定したモデルにおいても、固有ベクトルを説明変数に追加することで空間的依存性を考慮することが可能である。

3. ESFを用いた土地利用区分の推定

(1) 用いるデータの概要

本研究では、3 次メッシュで集計された 2006 年度の茨城県の土地利用区分を対象に、多項ロジットモデルに ESF を適用し、空間パターンを表す固有ベクトルの有無により推定精度、モデルの当てはまりがどの程度向上するのかを検証する。

被説明変数に国土数値情報ダウンロードサービスで公開されている、2006 年度の土地利用 3 次メッシュデータを用いる (図-1)。このデータは、3 次メッシュ (概ね 1km メッシュ) 毎に、各土地利用区分の面積を整備したものであるが、本研究では、以下に示すようにデータを集計し、各メッシュ内で最も占有面積の大きい土地利用区分を質的データとして与える。土地利用区分の内、荒地、幹線交通用地、その他の用地、ゴルフ場を「その他の用地」、また河川地及び湖沼、海浜、海水域を「水域」として集計し、「田」、「その他の農用地」、「森林」、「建物用地」はデータ加工せずに用いる。各メッシュにカテゴリーを当てるため、当該メッシュの中で最大の土地利用区分のカテゴリーをメッシュに与える。当該メッシュ内で土地利用区分の占有面積が等しくなる場合は、分類不能として分析からは除外した。また、大佛・倍田(1996)において、推定・予測の際には、土地利用が変化しないと考えられる安定地点について考慮することにより推定精度が向上することが示されていることから、当該メッシュ内において「水域」が最大の占有面積をもつメッシュは、不変地点として分析からは除外した。

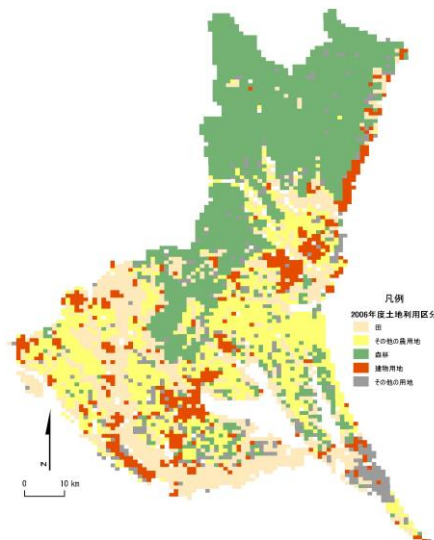


図-1 茨城県の 2006 年度土地利用 3 次メッシュ

説明変数には、地理的条件として標高・傾斜、含水域ダミーを用い、社会的条件として人口、最寄駅までの距離、東京駅までの距離を用いる。標高・傾斜は国土数値情報の2009年度標高・傾斜度次メッシュデータを用い、含水域ダミーは当該メッシュに国土数値情報土地利用3次メッシュの土地利用区分のうち河川地及び湖沼、海浜、海水域の面積を含むメッシュを1とした。人口は総務省統計局の e-stat からダウンロードできる 2005 年度の国勢調査に基づく人口総数、最寄駅までの距離と東京駅までの距離は当該メッシュの重心から各駅までの直線距離（単位：m）を用いる。なお、各駅のデータは国土数値情報の鉄道時系列データの中から 2006 年度に運用されている駅のみを用いた。

(2) 分析の結果

2013年5月7日現在、分析中につき、口頭発表時にレジュームを配布する。

4. まとめ

事前の予備的な分析結果より、空間的依存性を考慮しない場合と固有ベクトルを用いて空間的依存性を考慮する場合を比較すると、モデルの当てはまりを示すAIC、カテゴリー選択の適中率（%）ともに改善が見られた。

固有ベクトルを用いて空間的依存性を考慮するESF手法の今後の課題として、大量の固有ベクトルの変数選択の効率化、多岐の質的データに対する空間的依存性の検定手法の適用、空間的異質性への対処の3点が挙げられる。

参考文献

- 1) Anselin, L. : *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, 1988.
- 2) Chakir, R., Parent, O. : Determinants of land use changes: A spatial multinomial probit approach, *Papers in Regional Science*, Vol.88, No.2, 2009.
- 3) Wang, X., Kockelman, K., Lemp, J. : The dynamic spatial multinomial probit model: analysis of land use change using parcel-level data, *Journal of Transport Geography*, Vol.24, pp.77–88, 2012.
- 4) Griffith, D. A., Paelinck, J. H. : *Non-standard Spatial Statistics and Spatial Econometrics*, Springer, 2011.
- 5) 村上大輔：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，Vol.20, No.2, pp.1–12, 2012.
- 6) Griffith D. A. : A linear regression solution to the spatial autocorrelation problem, *Journal of Geographical Systems*, Vol.2, pp.141–156, 2000.
- 7) Griffith, D. A. : Spatial Structure and Spatial Interaction: 25 Years Later, *The Review of Regional Studies*, Vol.37, No.1, pp. 28–38, 2007.
- 8) Griffith, D. A. : *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization*, Springer, 2003.
- 9) 大佛俊泰，倍田賢一：安定地点を考慮した土地利用遷移予測モデル，GIS—理論と応用，Vol.4, No.1, pp.1–6, 1996.