

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 3 月 31 日現在

機関番号：11301

研究種目：若手研究(B)

研究期間：2010~2012

課題番号：22760383

研究課題名（和文）地域分析手法の精緻化に向けた空間統計解析理論の融合

研究課題名（英文）Integration of spatial statistical methods for regional analysis in a data rich environment

研究代表者

井上 亮 (INOUE RYO)

東北大学・大学院情報科学研究科・准教授

研究者番号：60401303

研究成果の概要（和文）：本研究では、空間統計学・計量地理学による空間分析手法を活用・統合した地域分析手法の開発を行い、適用可能性を検討した。まず、地点属性の時空間内挿に関して新たなモデルを提案し、既存手法より高い内挿精度を得ることを確認した。しかし、膨大な計算時間を要するため実行可能性に乏しいことが明らかとなった。また、時空間上の点事象集積検出について、柔軟な形状の集積領域を検出可能な手法を提案し、実行可能性を確認した。

研究成果の概要（英文）：This study integrated spatial analysis methods by spatial statistics and quantitative geography, proposed new regional analysis methods, and checked their applicability. First, a new model for spatio-temporal interpolation of point attributes was proposed. It was confirmed that the new model could acquire higher interpolation precision than the previous models; however, it revealed that it required considerable calculation time and lacked in practicability. Second, a new method for detecting spatio-temporal clusters of point events which have flexible shape was proposed, and its applicability was confirmed by the experiment.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	800,000	240,000	1,040,000
2011年度	500,000	150,000	650,000
2012年度	600,000	180,000	780,000
年度			
年度			
総計	1,900,000	570,000	2,470,000

研究分野：工学

科研費の分科・細目：土木工学 土木計画学・交通工学

キーワード：地域分析，空間統計，計量地理，時空間情報，内挿，集積検出

1. 研究開始当初の背景

近年、政府による統計情報の公開・流通促進へ向けた取り組みや、空間情報観測技術の発展による大量の高解像度情報の普及に伴い、国土・地域・都市を対象にする分析者は、比較的簡単に多くの空間情報を入手することが可能になってきた。例えば、国や地方自治体によってこれまでに整備されてきた統

計情報は、国土・地域・都市の過去・現在の実態を記録した貴重な情報であり、国民共有の財産であるとの認識が昨今急速に広まり、その有効活用を目指した統計情報の整備と公開が政府の施策として進められている。また、空間情報観測技術の発展は、高解像度センサーを搭載した人工衛星や、GPS・ICタグによる位置計測を利用したブローブ、地上に張り

巡らされたセンサーネットワークなど高時間・空間解像度の情報観測を可能にし、日々膨大な空間情報が取得・蓄積され、利用可能な環境が整備されてきている。また、現在では、空間位置だけではなく時刻も記録した時空間情報も利用できる環境が整いつつある。

これら質・量共に充実しつつある時空間情報を、いかに国土・地域・都市分析へと活用し、より詳細な現状把握、更には将来像構築へ活用していくかは、土木計画にとって重要であろう。しかし残念ながら、これまでのところ、時空間情報の充実に比べてその解析理論・手法は十分に進展しているとは言えない。

現在では時空間情報の統計解析理論・手法は、空間情報学と呼ばれる学際領域が形成され、研究が進められている。しかし、既存理論・手法のルーツは、空間統計学、計量地理学、計算幾何学など広い学問領域に及んでおり、またかつてはそれぞれ個別に発展してきたため、理論の前提である仮定や適用対象のデータ構造などは千差万別である。

そのため、様々な分析手法を駆使して地域分析を行うには、それらの前提条件を整理・融合し、新たな分析手法を構築することが必要である。また、空間分析手法を時空間情報への適用に向けて拡張することも求められている。

2. 研究の目的

本研究では、空間統計学・計量地理学・空間疫学において研究されてきた空間解析理論の融合を図り、不動産価格情報を例題に時空間情報を利用した地域分析への適用可能性を検証する

3. 研究の方法

GWRK と空間スキャン統計の 2 分析手法に関して、理論の検討と、不動産価格情報を用いた適用可能性の検討を行った。以下に、それぞれの研究について記す。

(1) GWRK

GWRK は、空間統計学手法の普遍クリギング(UK)と計量地理学手法の地理的加重回帰(GWR)を融合した分析手法である(例えば、Fotheringham et al., 2002)。本研究は、GWRK の理論を精査し時空間分析への拡張を図り、公示地価を用いて地域分析への適用可能性を検証した。

① 時空間 GWRK の構築

a) GWR

GWR は、ある点の近傍に重み付けした回帰モデルを用い、点固有のモデルを推定する手法である。ここで、点 i の被説明変数 y_i を内挿する場合を考える。まず点 i に関するモデルを設定する。空間バンド幅 δ_d ・時間バンド幅 δ_t を設定したガウス型カーネル関数に、

点 ij 間距離 d_{ij} ・時間差 t_{ij} を代入した

$$w_{ij} = \exp\left(-\left(d_{ij}/\delta_d\right)^2 - \left(t_{ij}/\delta_t\right)^2\right) \quad (1)$$

を、第 j 対角要素とする対角行列を \mathbf{W}_i とすると、点 i に関するモデルが得られる。

$$\mathbf{W}_i^{1/2}\mathbf{y} = \mathbf{W}_i^{1/2}\mathbf{X}\boldsymbol{\beta}_i + \mathbf{u}_i \quad \mathbf{u}_i \sim N(0, \sigma^2\mathbf{I}) \quad (2)$$

ただし、 \mathbf{y} は被説明変数ベクトル、 \mathbf{X} は説明変数行列、 $\boldsymbol{\beta}$ はパラメータベクトル、 \mathbf{u}_i は攪乱項ベクトル、 σ^2 は分散を表す。このとき、点 i の内挿値の最良線形不偏予測量(BLUP)は

$$\hat{y}_i = \mathbf{x}_i'(\mathbf{X}'\mathbf{W}_i\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_i\mathbf{y} \quad (3)$$

となる。なお本研究では、空間・時間のバンド幅は交互に交差検定を行い推定する。

b) UK

UK は、回帰モデルの攪乱項の空間確率場に二次定常性を仮定して、その共分散を相対位置ベクトルの関数として空間相関を構造化し、任意地点の内挿値を求める手法である。通常、二次定常性に加えて等方性も仮定し、共分散を距離の関数として構造化する。

時系列相関も扱う場合は、時間と距離に対する二次定常性を仮定する。本研究では、異方性パラメータを用いて時間軸を空間軸の一つと考える簡易的な指数型時空間共分散関数(式(5))を用いる。

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad \mathbf{u} \sim N(0, \mathbf{V}) \quad (4)$$

$$V_{ij}(h_{ij}) = \begin{cases} \theta_0 + \theta_1 & \text{if } h_{ij} = 0 \\ \theta_1 \exp(-h_{ij}^2/\theta_2^2) & \text{if } h_{ij} > 0 \end{cases} \quad (5)$$

$$\theta_0 + \theta_1 : \text{sill}, \theta_0 : \text{nugget}, \theta_2 : \text{range}, h_{ij} = \sqrt{d_{ij}^2 + \phi^2\tau_{ij}^2}$$

点 i の内挿値は、内挿点と観測点の攪乱項間共分散ベクトルを \mathbf{c} とすると式(6)で表される。

$$\hat{y}_i = \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{c}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad \boldsymbol{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (6)$$

c) GWRK

GWRK は、GWR の攪乱項に対し、UK と同様の二次定常性を仮定する手法である。内挿点 i に関するモデルは式(7)で表される。

$$\mathbf{W}_i^{1/2}\mathbf{y} = \mathbf{W}_i^{1/2}\mathbf{X}\boldsymbol{\beta}_i + \mathbf{u}_i \quad \mathbf{u}_i \sim N(0, \mathbf{V}_i) \quad (7)$$

また、内挿点の BLUP は式(8)で求まる。

$$\hat{y}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}}_i + \mathbf{c}_i'\mathbf{V}_i^{-1}\mathbf{W}_i^{1/2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_i) \quad (8)$$

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}'\mathbf{W}_i^{1/2}\mathbf{V}_i^{-1}\mathbf{W}_i^{1/2}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}_i^{1/2}\mathbf{V}_i^{-1}\mathbf{W}_i^{1/2}\mathbf{y}$$

なお Harris et al. (2010)では、GWRK の手順を「(i) 内挿点に関する GWR のパラメータを推定・内挿値を算出」「(ii) GWR の残差を、通常クリギングを用いて内挿」「(iii) GWR による内挿値と残差の内挿値を加算」し、GWRK による内挿値としている。しかしこの方法は、GWR では攪乱項の共分散がないことを仮定(式(2))しているにも関わらず、その残差に二次定常性を仮定するという不合理な推定を行っており、この手法には問題がある。

一方、本研究で提案する式(8)のモデルにも不合理な仮定が存在する。内挿点から離れた地点には小さい重みが与えられるため、通常、

残差も小さな値となる。すなわち、内挿点からの距離に応じて分散や共分散は小さくなるはずだが、攪乱項には二次定常性を仮定しており不合理である。なお、同様の問題は、GWR のモデルにも存在する。式(2)は攪乱項の分散均一を仮定しているが、重みにより、内挿点から離れた地点では攪乱項の分散は必然的に小さくなることが予想され、両者の設定は不合理である。

このように、GWR は理論的に不整合な仮定を設定しており、GWR を基に構築された GWRK にも問題があることが確認された。

② 公示地価を用いた適用可能性の検討

GWR・UK・GWRK の内挿精度を検証する。なお、2000～2012 年の東京都区部住居系用途地域内の地積 150 m²未満の情報を用いる。

被説明変数は地価の対数値とし、説明変数は、地積(m²)・最寄り駅から都内主要駅(東京・新宿・池袋・渋谷・上野)迄の鉄道所要時間(分)・最寄り駅迄の距離(m)・前面道路幅員(m)・前面道路方位ダミー(南東・南・南西:1, それ以外:0)・住宅ローン金利(%)を用いる。

a) 異方性パラメータ・バンド幅の設定

異方性パラメータは、通常最小二乗法の残差を利用し同一時点・同一地点のデータによる距離・時間差に関する経験セミバリオグラムを作成し、それぞれに線形の関数を当てはめ分散が等しくなる距離と時間差の比を求めて設定した。本研究では、1年=420m と換算する異方性パラメータを設定した。

バンド幅は、GWR では交差検定を用いて推定する。しかし、GWRK では、交差検定では膨大な計算を要するため、本研究では、空間は 500～1700m の 100m 間隔、時間は 1～10 年の 1 年間隔で格子点探索を行い、推定した。

b) 内挿精度の検証

内挿精度は、Leave-one-out cross validation による平均二乗平方根誤差(RMSE)で評価する。計算には R の gstat を使用した。

GWR・UK・全期間共通のバンド幅による GWRK・各年の最適バンド幅による GWRK の内挿精度を表-1・図-1 に示す。

推定結果より、全モデルで、対象期間の両端と、2006～2010 年に精度が悪化することが確認された。まず、対象期間の端部では、内挿に利用可能な近隣の観測点が少なくなるため内挿精度が低下すると考えられる。また、2006～2010 年は、地価変動が大きい時期に当たる。いずれの手法も、直前・直後の時期に強く影響されるため、価格上昇の開始期や、下落に転じる時期、下落が沈静化する時期に、内挿精度が大きく低下している。

次に、GWRK の内挿精度に着目する。まず、全期間共通のバンド幅による GWRK は、2008 年を除き GWR・UK より高精度を示した。このことから、GWRK の有効性が確認できる。また、年毎にバンド幅設定を行った結果を見

表-1 内挿精度の比較

	GWR	UK	GWRK			
			全期間共通 (1300m, 3年)	年毎	距離 (m)	時間差 (年)
全期間	0.041	0.034	0.028	0.022	-	-
2000	0.053	0.042	0.037	0.036	1700	5
2001	0.039	0.023	0.018	0.013	700	5
2002	0.019	0.025	0.015	0.011	600	1
2003	0.023	0.022	0.014	0.009	600	2
2004	0.019	0.024	0.014	0.008	700	2
2005	0.020	0.025	0.015	0.010	700	2
2006	0.046	0.038	0.034	0.033	1300	6
2007	0.074	0.031	0.020	0.016	800	8
2008	0.048	0.056	0.063	0.036	800	1
2009	0.030	0.036	0.020	0.016	600	9
2010	0.058	0.032	0.023	0.023	1300	3
2011	0.017	0.024	0.015	0.007	600	2
2012	0.041	0.046	0.029	0.028	600	1

(*) GWR バンド幅: 720m, 1年

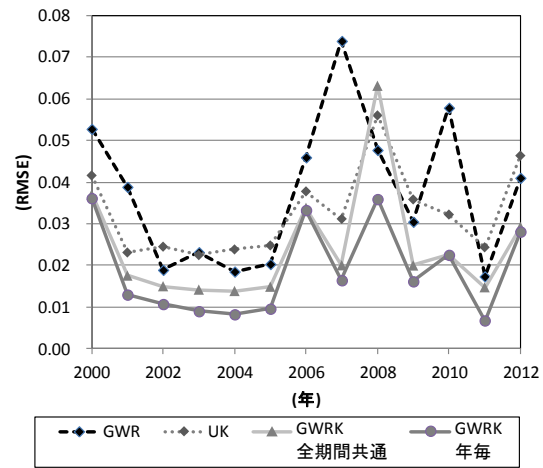


図-1 内挿精度の比較

ると、最適バンド幅は時期により異なることが確認されたが、地価の特性とバンド幅の関係に決まったパターンは認められず、交差検定以外の方法でバンド幅設定を行う方法は見いだすことができなかった。

前述の通り、GWRK のバンド幅推定には膨大な計算時間を要することを考えると、GWRK を用いた内挿手法は、内挿精度を改善できる可能性はあるものの、実用性に乏しいことが明らかとなった。

(2) 時空間スキャン統計

時空間スキャン統計は、時空間上の点事象の集積時期・地域を検出する統計的手法である(Kulldorff et al., 1998)。

① 時空間スキャン統計による集積検出

時空間スキャン統計は、事前に設定した集積候補領域に対して、対立仮説・帰無仮説の尤度比を用いて集積性を評価する。

まず、点事象の位置が 2 次元空間座標と時刻(x, y, t)で表されており、この点事象の集積領域(空間領域・時期)を検出することを考える。今、ある空間・時間領域 Z を集積の候補領域として設定したとする。

点事象の発生が特定の確率分布に従って

いるという仮定の下に、対立仮説を「候補領域 Z の内外で事象の発生確率は異なる」、帰無仮説を「候補領域 Z の内外で事象の発生確率は同一である」として、尤度比を計算する。

通常、点事象の分布には Poisson 分布を仮定し、尤度比を算出する。ここで、分析対象の全領域を G 、候補領域 Z 内の点事象発生確率を p 、候補領域 Z 外の点事象発生確率を q と表すとする。また、領域 A の大きさを $\mu(A)$ 、領域 A 内の事象数を n_A と表す。このとき、候補領域 Z の対立仮説の尤度関数は、

$$L(Z, p, q) = \frac{\exp\{-p\mu(Z) - q(\mu(G) - \mu(Z))\}}{n_G!} p^{n_Z} q^{(n_G - n_Z)} \quad (9)$$

となり、帰無仮説の尤度関数は

$$L_0(Z, p) = \frac{\exp\{-p\mu(G)\}}{n_G!} p^{n_G} \quad (10)$$

となるので、領域 Z の対立仮説と帰無仮説の尤度比 $\lambda(Z)$ は、領域 A の事象発生密度を $d(A)$ ($=n_A/\mu(A)$)、 $I(\cdot)$ を指示関数とすると

$$\lambda(Z) = \frac{d(Z)^{n_Z} d(\bar{Z})^{(n_G - n_Z)}}{d(G)^{n_G}} \cdot I(d(Z) > d(\bar{Z})) \quad (11)$$

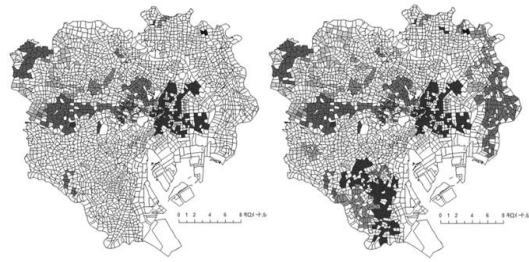
と表すことができる。

空間スキャン統計では、分析対象領域内 G に候補領域 Z を走査させ、最大尤度比を出力する候補領域(Most Likely Cluster, MLC)を検索する。検出された MLC は、対象領域内に無作為に作成した点事象分布から得られる最大尤度比分布を作成するモンテカルロシミュレーションを通して、有意性を検定する。

②候補領域の設定

上記の尤度比を用いた集積性の評価を行う手法では、まず時空間上を走査する候補領域の形を設定しなければならない。Kulldorff et al. (1998)は、円形の空間領域に時期を設定した時空間上で円筒の領域を候補領域と設定した分析を提案している。また、空間上の集積検出に限定した手法であるが、Kulldorff et al. (2006)は楕円領域を設定し、より柔軟な形状の抽出を試みている。しかし、地域分析、特に人間活動の結果である社会経済現象の分析では、ある点から一定の距離内の範囲に等方的に影響が波及する構造の伝播パターンを示すとは限らず、例えば交通ネットワークに沿った点分布の集積が発生するなど、複雑な形状の集積が生じる可能性がある。そのような複雑な形状の領域を候補領域として設定する方法として、Duczmal and Assunção (2004)は、町丁目界のように領域分割された空間単位を利用し、隣接領域の組み合わせを候補領域とする方法を提案し、その拡張を行う研究が行われている(例えば Yao et al. 2011)。

本研究は、隣接空間単位の組み合わせから作成する複雑な形状の候補領域を用いて集積検出を行う手法に着目し、時空間分析への



(a) 2003年1~6月 (b) 2007年1~6月

図-2 不動産取引集積の検出例

拡張を行う。

③新たな候補領域法の提案

空間単位に基づいた集積検出を行う場合、隣接空間単位を結合して集積候補領域を走査する必要がある。空間単位数が多くなると、隣接空間単位の組み合わせ数は指数関数的に増加するため、全ての組み合わせを走査することは不可能である。Duczmal and Assunção (2004)では、焼き鈍し法を利用した集積領域探索を提案しているが、空間単位数が増加すると、実行可能な計算時間で尤度比が大きい集積領域を探索できるか疑問である。

そこで、本研究は空間単位を階層的に結合し集積を検出するアルゴリズムを提案した。以下に提案アルゴリズムの流れを示す。

「(i) 点事象がある空間単位のみを選定し候補領域とする」「(ii) 候補領域の隣接関係を利用し、10程度の候補領域からなる候補領域集合を設定して分析対象領域を分割する」「(iii) 各候補領域集合内で、候補領域の組み合わせを列挙し尤度比を算出する」「(iv) 各候補領域集合内で最大尤度比となる組み合わせを選定し、新たな候補領域に加える。また、最大尤度比の候補領域を含む組み合わせを除き、他の候補領域集合内の候補領域と隣接する領域も新たな候補領域に加える」「(v) 候補領域数が2以上なら(ii)に戻り繰り返す。1なら候補領域を出力して終了する」

④提案手法の適用

提案手法を不動産取引価格情報の取引地点データに対して適用し、取引の集積地域・時期の検出を試行した(図-2)。

本提案手法により、不動産取引が活発であった時期や地域を、実行可能な計算時間で抽出することができることを確認した。

4. 研究成果

本研究では、地点属性データを利用した時空間内挿と、点事象分布データを利用した集積検出に関して、新たな分析手法を提案した。

現在、時空間情報として大きく分けて、地点属性データ・点事象データ・空間/時間集計データの3種類のデータが提供されている。本研究で提案した点事象集積検出を活用すると、点事象の集積を考慮した地点属性デー

タや空間/時間集計データの分析が可能になると考えられる。本研究の成果により、多種の時空間情報を活用した地域分析に向けて、新たな一つの可能性を提示することができたと考えている。

参考文献

- Duczmal, L. and Assunção, R. 2004. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistic and Data Analysis*, **45**: 269–286.
- Fotheringham, A.S. Brunsdon, C., and Charlton, M.E. 2002. *Geographically Weighted Regression: The analysis of spatially varying relationships*. John Wiley & Sons Ltd.
- Harris, P., Fotheringham, A.S., Crespo, R., and Charlton, M. 2010. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Mathematical Geosciences*, **42**: 657–680.
- Kulldorff, M., Athas, W., Feuer, E., Miller, B., and Key, C. 1998. Evaluating cluster alarms -A space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American Journal of Public Health*, **88**(9): 1377–1380.
- Kulldorff, M. 1997. A spatial scan statistic, *Communication Statistic Theory and Method*, **26**(6): 1481–1496.
- Yao, Z., Tang, J., and Zhan, F.B. 2011. Detection of arbitrarily-shaped clusters using a neighbor-expanding approach: A case study on murine typhus in South Texas. *International Journal of Health Geographics*. **10**: 23.

5. 主な発表論文等

[雑誌論文] (計4件)

- ① 糟谷志帆・井上 亮. 不動産取引価格情報に基づく小地域単位の市況変動の抽出. 土木計画学研究・講演集, 45, CD-ROM. 2012.
- ② 川松 祐太・井上 亮. 時空間における地価形成要因影響度の変遷・差異を考慮した普遍クリギングによる地価内挿の実証的検討. 土木計画学研究・講演集, 45, CD-ROM. 2012.
- ③ 井上 亮・渡邊 拓也. 不動産取引事例情報に基づく取引位置・価格水準の時空間集積の抽出. 土木計画学研究発表会・講演集, 44, CD-ROM. 2011.
- ④ 井上 亮・中西 航・杉浦 綾子・中野 拓・米山重昭. 取引価格と公的地価指標の比較を通じた地価情報提供の検討. 地理情報システム学会講演論文集, 19, CD-ROM, 2010.

[学会発表] (計2件)

- ① Ryo Inoue・Wataru Nakanishi・Ayako Sugiyura・Taku Nakano・Shigeaki Yoneyama.

Publication of comparison information between interpolated appraised land prices and transaction land prices. *GeoViz 2011*, 2011年3月10日, ハンブルグ, ドイツ.

- ② 井上 亮・杉浦 綾子. 不動産の取引価格と公的地価指標の比較による情報提供法の検討—Webサービス「井上・TAREA地価情報提供システム」の開発—. 日本不動産学会 平成22年度秋期全国大会, 2010年11月28日, 東京.

6. 研究組織

(1) 研究代表者

井上 亮 (INOUE RYO)

東北大学・大学院情報科学研究科・准教授

研究者番号：60401303