

階層的な隣接地域結合操作に基づく点事象の時空間集積検出の提案

糟谷 志帆・井上 亮

A New Spatio-Temporal Cluster Detection Method by Searching Hierarchical Combination of Adjacent Area Units

Shiho KASUYA and Ryo INOUE

Abstract: 点事象が集積する空間や時空間の領域を抽出する方法については、空間スキャン統計に基づく手法が数多く提案されている。その多くは、ある地点を中心とする波及構造を仮定し、円形に近い空間領域を集積として仮定する。しかし社会経済現象では、交通網を通じた影響の波及や行政区単位の地域属性の相違による波及速度の差異等により、等方的に影響が波及する構造を仮定することは必ずしも妥当ではない。そこで本研究では、より柔軟な形状の集積を、階層的な隣接地域結合操作を用いて高速に検出する手法を提案し、不動産取引地点・時点データを利用して時空間集積検出への適用可能性を確認する。

Keywords: スキャン統計 (scan statistic), 点事象 (point event data), 時空間集積検出 (spatio-temporal cluster detection)

1. はじめに

近年、政府による統計情報・行政情報の公開施策の進展や、測位・地図利用技術など地理情報技術の一般社会への普及により、詳細な位置情報を持った大量の地理情報を利用した地域分析が可能な環境が整いつつある。また、現在流通する地理情報には、詳細な時刻情報が記録されているものが数多く存在しており、時空間情報を活用した分析手法開発の必要性が高まっている。

本論文では、時空間情報を持つ点事象の分析に注目する。時空間上の点事象分布の分析方法として、時空間スキャン統計(Kulldorff *et al.*, 2005)が提案されている。これは、点事象の空間集積領域を検出する空間スキャン統計(Kulldorff and Nagarwalla, 1995; Kulldorff, 1997)を時空間に拡張したもので、集積地域・期間を検出する方法で、点分布の地域的差異や時間的变化など地域の特徴を知ることができる分析手法である。

さて、通常、空間・時空間スキャン統計では、まず集積領域の候補となる時空間領域(以後、候補領域と呼ぶ)の形状を設定し、候補領域を時空間内に走査し、集積領域を探索する。候補領域の大きさと候補領域内の点事象数から、最も集積が見られる候補領域を選別する。なお、点事象が市町村界などの空間単位で集計されている場合には、重心が候補領域内に含まれる空間集計単位内の点事象数と、当該の空間集計単位の大きさを用いて、集積の有無を判断する。

これまで候補領域の形状として、円形(Kulldorff, 1997)や楕円形(Kulldorff *et al.*, 2006)が提案されてきた。また、Tango and Takahashi(2005)は、候補領域形状の柔軟性を高めるため、ある一定サイズの円形領域内に含まれた空間単位の組み合わせを候補領域として設定している。また、空間スキャン統計と類似の方法において、Mori and Smith(2009)は、交通ネットワーク上の所要時間から規

定される空間上で“凸領域”を考え、候補領域として設定する方法を提案している。

これらの候補領域設定方法は、基本的には、ある地点を中心とした影響の波及構造を仮定し、空間集積を検出することを目指している。しかし、特に大縮尺の分析を行う場合には、空間の持つ特性が均一である、あるいは、滑らかに変化していることを前提条件として設定できるとは限らない。そのため、影響が等方的あるいは連続的に波及して点事象が発生していることを仮定した分析は必ずしも妥当ではない。

一方、Duczmal and Assunção (2004)は、隣接する空間集計単位の組み合わせを候補領域と考え、焼き鈍し法を用いて空間集積を検出する方法を提案している。任意の隣接空間集計単位の組み合わせを候補とするこの方法では、波及構造を柔軟に想定することが可能で、柔軟な形状の集積領域を検出することができる。Duczmal and Assunção (2004)は、Belo Horizonte の殺人事件発生地点データに対して提案手法を適用し、高速道路に沿った劣悪な環境の地域に、細長い集積を検出できることを示している。この適用例では空間集計単位数が 240 と比較的少ないことから、短時間で集積の検出および検定を実行できている。しかし、空間集計単位数が大きくなると、隣接空間集計単位の組み合わせの数は爆発的に増大するため、広範囲にわたる集積領域が考えられうる場合には、焼き鈍し法に基づく手法で集積検出を行うことは難しいと予想される。また、既存研究では、空間的な集積のみを検討しているが、焼き鈍し法に基づく方法を時空間集積検出へと拡張することを考えた場合、有意性の高い集積領域を得るには多くの計算時間を要することが予想される。

そこで、本研究では、空間集計単位数が多いデータから時空間集積を検出する方法について、新たな方法の提案を行う。また、不動産取引地点データに対する適用を通して、提案手法の有用性を確認する。

2. 空間スキャン統計の集積検出法

Kulldorff and Nagarwalla (1995)は、集積検出の候補となる領域を設定した上で、候補領域内外の点事象発生率に差があるかを尤度比で評価する空間スキャン統計を提案した。ここで、尤度比を用いた集積検出法について概略を示す。

点事象が分布する分析対象領域を G 、集積の候補となる領域 G の部分集合を Z とする。また、領域 G 内の事象数を N 、事象発生の母数を S_G 、領域 Z 内の事象数を n 、母数を S_Z と表す。

点事象分布が Poisson 分布に従うとの仮定の下で、対立仮説「領域 Z 内の点事象発生率は外よりも大きい」に対して、帰無仮説「領域 Z の内外で点事象発生率は等しい」を設定する。この仮説の尤度比は、式(1)で求められる。ただし、 $I()$ は指示関数で、括弧内が真の時 1、偽の時 0 を返す。

$$\lambda(Z) = \frac{(n/S_Z)^n ((N-n)/(S_G-S_Z))^{N-n}}{(N/S_G)^N} I\left(\frac{n}{S_Z} > \frac{N-n}{S_G-S_Z}\right) \quad (1)$$

空間スキャン統計は、候補領域 Z を分析対象領域 G 内で走査し、式(1)の尤度比が最大となる候補領域 \hat{Z} を検出する。検出された集積の有意性は、ランダムな点事象分布から得られる最大尤度比分布をモンテカルロ法で算出し判定する。

なお、時空間に拡張する場合も同様に、空間領域に加えて期間を設定した候補領域を分析対象領域内に走査させて、同様に最大尤度比を出力する領域を検索する。

3. 柔軟形状の集積領域検出手法の提案

時空間スキャン統計で集積検出を行うには、時空間上を走査する集積候補領域形状を定める必要がある。本研究では、柔軟な空間形状の集積検出のため、行政界などの空間単位の隣接関係を利用し、集積候補領域の空間形状を設定する。

隣接空間単位の組み合わせは、空間単位が多くなると指数関数的に増加する。空間単位数が N の時、隣接空間単位の組み合わせ数が最小の場合、すなわち全空間単位が一列に並び、両隣の 2 つの

空間単位とのみ隣接する場合でも，組み合わせ数は $2^N - 1$ になる．メッシュや町丁目界など，通常，分析の対象として用いられる空間単位では，各単位の隣接数が多く，組み合わせ数はより大きくなるため，全組み合わせを走査することは不可能であろう．Duczmal and Assunção (2004)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．

以上のように，対象領域内の隣接空間単位の全組み合わせを走査することは現実的ではない．そこで，本研究は分析対象領域を階層的に分割した上で，分割された領域内で空間単位を結合して集積検出を行い，得られた集積を階層的に結合して全対象領域における集積を検出するアルゴリズムを提案する．以下にアルゴリズムの流れを示す．

- (a) 対象領域全体のうち，点事象が存在する空間単位のみを選定する(図-1)．すなわち，点事象がない空間単位は，集積検出対象としない．
- (b) 対象領域全体を小領域に分割する(図-2)．
- (c) 小領域を集積検出単位とし，隣接空間単位の組み合わせをすべて考え，尤度比を算出する．図-3の小領域Aでは， $\{1\}$ ， $\{1, 2\}$ ， $\{1, 2, 3\}$ ， $\{1, 2, 4\}$ ， $\{1, 2, 3, 4\}$ ， $\{2\}$ ， $\{2, 3\}$ ， $\{2, 4\}$ ， $\{3\}$ ， $\{3, 4\}$ ， $\{4\}$ の11組み合わせを考える．なお，尤度比算出時は，内部の点事象発生時点をを用い，尤度比が最大となる期間を探索する．
- (d) 各集計検出単位で，最も尤度比の高い空間単位の組み合わせを選定し，集積候補領域とする(例えば，図-3の集計検出単位Aの $\{3, 4\}$)．また，他の集計検出単位と隣接している空間単位の組み合わせも，他の小領域内で見つかった集積候補領域と結合することにより，有意な集積となる可能性がある．そこで，各小領域から最も尤度比の高い領域を含む組み合わせを除いた後に，小領域の境界と接する空間単位の組み合わせを集積候補領域に追加す

る(例えば，図-3の集計検出単位Aの $\{1\}$)．

- (e) 隣接する小領域を結合し，集積検出単位とする．その集積検出単位内の集積候補領域を選び，隣接する候補領域の組み合わせを考え，操作(c)(d)を繰り返す．操作(c)と同様に，最大尤度比の集積候補領域を選定し，上位の集積候補領域とする(例えば，図-4で集積検出単位A+Bを考えした場合，赤の領域を指す)．また，操作(d)と同様に，他の集積検出単位と隣接する集積候補領域も上位の集積候補領域とする．以上の作業を対象領域全体で繰り返し，集積候補領域を選定する．

この提案手法では，集計検出単位内の候補領域数を限定することによって，隣接領域の組み合わせ数を小さくすることができるため，集積検出に要する時間を大幅に短縮することが可能である．

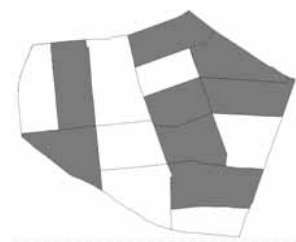


図-1 対象領域の選定

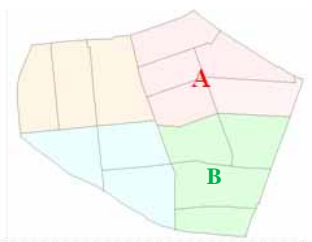


図-2 小領域分割

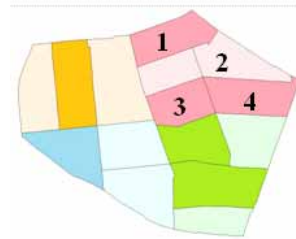


図-3 クラスタ候補の選定



図-4 隣接地域の結合

4. 提案手法の適用

時空間スキャン統計を，不動産取引価格情報に対して適用し，取引位置の集積・取引価格水準の地域的・時期的傾向の抽出を行う．

不動産取引価格情報とは，不動産取引の位置・日および取引物件の属性を記録する情報である．本研究では，1999年1月～2009年12月の東京都区部の更地取引(21796件)の取引位置・日を用いて分析を行う．また，空間単位と

して数値地図 25000(空間データ基盤)の大字町丁目界(3145 地域)を用いる。なお、空間単位が一点でも共有していれば、「隣接」関係にあるとする。また、母数 S には面積(km^2)を用いた。

最も尤度比の大きな集積の検出には Xeon X5670 2.93GHz で 30 分を要したものの、図 5 が得られた。集積の期間は 1999 年 2 月から 2009 年 7 月までと分析対象期間をほぼカバーする結果となった。

図 5 では、小規模宅地が多い地域が全面的に集積地域として抽出されている。取引件数を解釈する母数として空間単位の面積を用いたため、小規模宅地の多い地域では単位面積当たりの取引件数が多くなり、集積として判断されたためと考えられる。

提案手法を用いた分析では、母数の設定について注意を要することが示唆された。



図-5 更地取引の集積地域

5. おわりに

本研究では、階層的な操作を通じて柔軟な空間形状の時空間集積検出を行う手法を提案し、大規模データへ適用可能性を確認した。

しかし、集積結果は不動産取引集積の解釈には不十分で、提案手法を活用した地域分析を行うには課題があることが明らかになった。また、今回使用した尤度比を用いた集積検出では、複数の集

積領域があるという仮説を検討できないため、発生確率の異なる領域を一領域として集約してしまう可能性がある。集積領域の数も含めた検定を行う手法の構築が不可欠である。これらの検討を今後の課題としたい。

謝辞

本論文は、日本学術振興会科学研究費助成事業の基盤研究(A)21241039 および 24241053 の援助を受けて行われた研究の一部である。また、不動産取引価格情報は、東京都不動産鑑定士協会からご提供いただいた。ここに記し、謝意を表す。

参考文献

- Duczmal, L. and Assunção, R., 2004. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics and Data Analysis*, 45, 269-286.
- Kulldorff, M., 1997. A spatial scan statistic. *Communication Statistics Theory and Method*, 26(6): 1481-1496.
- Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R. and Mostashari, F. 2005. A space-time permutation scan statistic for disease outbreak detection. *PLoS Medicine*, 2(3): 216-224.
- Kulldorff, M., Huang, L., Pickle, L. and Duczmal, L. 2006. An elliptic spatial scan statistic. *Statistics in Medicine*, 25: 3929-3943.
- Kulldorff, M. and Nagarwalla, N., 1995. Spatial disease clusters - detection and influence, *Statistics in Medicine*, 14, 799-810.
- Mori, T. and Smith, T.E. 2009. A probabilistic modeling approach to the detection of industrial agglomerations. Discussion paper No.682, Institute of Economic Research, Kyoto University.
- Tango, T. and Takahashi, K., 2005. A flexibly shaped scan statistic for detecting clusters. *International Journal of Health Geographics*, 4: 11.