形状複雑度設定に応じた ネットワーク上における点事象集積領域検出法

塚原 元英¹・井上 亮²

1学生会員 東北大学 大学院情報科学研究科(〒980-8577 宮城県仙台市青葉区片平2-1-1) E-mail:tsukahara@plan.civil.tohoku.ac.jp

²正会員 東北大学准教授 大学院情報科学研究科(〒980-8577 宮城県仙台市青葉区片平2-1-1) E-mail:rinoue @plan.civil.tohoku.ac.jp

近年,空間スキャン統計と呼ばれる点事象集積領域の検出法は、ネットワーク上の分析へと拡張されている。検出される領域の形状に着目すると、単純な形状の集積領域を検出する手法と、柔軟な形状の集積領域を検出する手法に大別される。前者は検出された集積領域内に点事象分布が希薄な地域を含む可能性、後者は複雑な形状の集積領域を検出する可能性を有しており、その利用に制約がある。利用しやすい分析法を提供するためには、領域形状を制御して解釈のしやすい集積領域を検出する分析法が必要である。そこで本研究は、ネットワーク上の領域形状に対し評価指標を提案した上で、その指標を用いて形状を制限し、集積領域を検出する手法を提案する。提案手法の実行可能性は、東京および仙台の施設立地データに対する適用を通じて確認する。

Key Words: point-events, shape-complicasy, network, cluster, spatial scan statistic

1. はじめに

近年、詳細な位置の情報を持つ地理情報の整備、公開が進んでいる。行政機関が持つ統計情報の有効活用を目的とした「オープンガバメント」施策では、一般に公開されているデータの中に、詳細な位置の情報を備えたものも多く存在する。例えば、警察が公開している犯罪発生マップ、事故発生マップでは、事象発生箇所の詳細な情報を見ることができる。このような行政からの情報公開の他にも、GPS付き携帯端末の普及により、個人のレベルで詳細な位置の情報を持った地理情報を、作成・共有することも可能になってきている。

以上のように、誰もが詳細な位置の情報を備えた地理情報を入手し、利用できる環境づくりが図られている. こうしたデータは、小領域を対象とした地域分析に活用することができる. そこで本研究では、詳細な位置情報を活用した分析手法として、点事象の分布集積検出法に焦点を当てて研究を行う.

点事象集積検出手法に関する研究はこれまで、地域単位に集計されたデータを対象にして行われてきた。それらの手法は、空間スキャン統計^{1),2)}という統計的な手法に基づいており、検出される集積領域の形状に着目すると、①単純な形状での集積領域の検出、②柔軟な形状で

の集積領域の検出、に分類することができる.

単純な形状での検出は、円形り、円形の領域とその周囲の領域を組み合わせるもの3、楕円形りといった、空間上の事象の広がりを考慮した形状で集積領域を検出するものである。この手法による検出結果は、集積の中心地、分布の範囲を解釈のしやすいという利点を備えている。しかし、点集積は全てが単純な形状で分布しているとは限らない。例えば、ある公害の集積分析をするとしよう。その原因物質が川を流れていた場合、その公害分布も川の流域に沿った、細長い形状であると考えることができる。このように、川や山、道路といった、点の分布に影響を与える可能性のある地理的条件は多い。単純な形状での検出では、こうした地理的な条件を結果に反映することができないため、本来は集積箇所とはいえない地域も、集積領域として検出してしまうことがある。

柔軟な形状での検出は、複数の隣接地域を連結して自由度の高い形状を持つ候補領域として、集積検出を行うものである。候補領域の作成は、焼きなまし法⁹、遺伝的アルゴリズム⁹、あるいは専用のアルゴリズム⁷⁾を用いるものなど、様々な手法が提案されてきた。単純な形状では考慮できなかった地理的な条件を反映して、集積領域を検出することができる。しかし、形状の自由度が高いため、集積領域を複雑すぎて解釈が難しい形状で検

出する,必要以上に大きな領域として検出するという問題が生じる場合がある.

このように、単純な形状・柔軟な形状の検出にはどちらにも利点・欠点があり、どちらの手法が適しているかは、分析の対象・目的によって異なる。そこで、集積領域の形状複雑度を分析者が調節し、任意の形状複雑度で検出ができる手法³が提案されている。分析者が解釈しやすい結果を検出するよう調整することができるため、自由度の高い地域単位の集積分析が可能である。

そして近年、詳細な地理情報を活用した集積検出法と して、点間の近接関係を道路のネットワーク距離で評価 し、集積領域をネットワークとして検出する手法が提案 されている. 地域単位での検出では, 近接関係をユーク リッド距離で評価していた. 広域での分析に関しては, これでも特に不都合は生じない. だが、詳細な地理情報 を活用した, 小領域中での分析ではどうだろうか. 都市 空間における人間活動は、道路などの交通ネットワーク と依存関係にある。また、ネットワーク空間上の移動を 考えた場合, ユークリッド距離とネットワーク距離では 大きな違いがある. そのため、詳細な地理情報を活用し た小領域中での分析では、近接関係をネットワーク距離 で考慮する必要がある. ネットワーク上の集積検出法に 関しても、ネットワーク中の任意の位置から一定のネッ トワーク距離の範囲内を集積領域とする単純な形状での 検出法8や、リンクの接続関係に基づいてリンクの組み 合わせで構成される柔軟な形状の領域を拡大して集積領 域を探す検出法9が提案されている.しかし,両者の間 で任意に形状複雑度を調節して集積領域を検出できる手 法はまだ提案されておらず、ネットワーク上での集積領 域検出において分析の幅を拡げるには、この手法の開発 が不可欠である.

そこで本研究では、ネットワーク上の点事象集積分析において、集積領域の形状を、関数で評価できる形状指標を導入し、集積領域を任意の形状複雑度で検出できる手法を提案する。また、この形状指標を適用する手法として、地域単位で遺伝的アルゴリズムを用いて、柔軟な形状の集積検出法をネットワーク上の検出に拡張した手法を提案する。これによって、詳細な位置の情報を持つ地理情報を活用したネットワーク上における点事象集積分析の自由度を高めることが、本研究の目的である。

2. 形状複雑度に基づく集積検出の既往研究

本研究の提案手法を含め、多くの集積検出法は、空間スキャン統計^{1),2}に基づき集積領域を検出する。そこでまず(1)では、空間スキャン統計の解説を行う。この空間スキャン統計に基づいた手法に、地域単位における集

積検出において、形状複雑度を制御し、集積領域を検出する手法³が提案されている。この手法は、領域の形状複雑度を関数を用いて評価し、形状複雑度が高い領域が検出されるのを制限する手法である。形状複雑度を評価する関数は、柔軟な形状の集積領域検出法と組み合わせて利用する。分析者が任意の強さの形状複雑度制限を設定することで、一定以上のコンパクトさを持つ集積領域を検出できる。

次に、(2)では、地域単位で検出される集積領域の形状を評価する関数として提案されたCompactness Value³と、それを用いた集積領域の形状複雑度の制御手法を、また(3)では、Compactness Valueを適用する柔軟な形状の集積領域検出手法として、遺伝的アルゴリズムを用いた柔軟な形状の集積検出法³の解説を行う.

(1) 空間スキャン統計^{1), 2)}

空間スキャン統計は、集積領域を統計的に特定する手 法の一つであり、多くの集積検出法がこの手法に基づい て提案されている.

ここで、集積の評価基準となる点事象密度算出の分母となる集積地域全域の大きさをN、全事象数をCと標記する。複数の隣接地域を結合させて作成する集積の候補領域Zと表し、Zの大きさをN2、事象数をC2とする。Z2を構成する地域を変、さまざまな大きさ、形状の候補領域作成して集積地域を探索する。Zが集積領域かどうかは

帰無仮説 H_0 : Z内とZ外の事象発生確率は変わらない. 対立仮説 H_1 : Z内はZ外よりも事象発生確率が高い.

の、仮説検定を行うことで確認できる。ただし、各地域に対して検定を行うことは、多重検定問題が発生するため、好ましくない。そこで空間スキャン統計では、帰無仮説と対立仮説の尤度比を用いて集積領域を特定する。点の分布をポアソン分布と仮定すると、尤度比*LR(Z)*は式(1)で表される。

$$LR(Z) = \frac{L_1(Z)}{L_0} = \begin{cases} \left(\frac{C_Z}{N_Z}\right)^{C_Z} \left(\frac{C - C_Z}{N - N_Z}\right)^{C - C_Z} & \text{if } C_Z < N_Z \\ 1 & \text{otherwise} \end{cases}$$
 (1)

各候補領域に対し尤度比を算出し、最も尤度比が高かった地域がMLC(Most Likely Cluster)となる。MLCに対してモンテカルロ検定を行い、集積領域の判定を行う。

(2) 地域単位での分析における形状複雑度設定

a) 形状複雑度評価関数

Duczmalら³は、地域単位での分析において、集積領域の形状を調整する手法を提案した。候補領域の形状複雑度をCompactness Value K(Z)と呼ばれる関数で評価し、その指標値を用いて任意の形状複雑度を持つ集積領域を検出する。対象とする領域をZ、その面積をA(Z)、Zの凸包の外周長をH(Z)とする。K(Z)は、候補領域の面積と、そ

の凸包と同じ外周長を持つ円の面積との比をとった無次元数であり、その値は対象領域の形状のみに依存する. K(Z)は式(Z)で表される.

$$K(Z) = A(Z) / \pi \left(\frac{H(Z)}{2\pi}\right)^2 \tag{2}$$

対象領域が複雑な形状をしているほどK(Z)は0に近い値をとり、対象領域が円形に近いコンパクトな形状であるほどK(Z)は1に近い値をとる.

b) 形状複雑度評価関数を用いた集積検出

Compactness Value K(Z)は,柔軟な形状の集積領域検出法と合わせて用いられる.空間スキャン統計を用いて集積領域を求める際,通常,各候補領域は尤度比LR(Z)で評価される.ここで,K(Z)を用いて形状複雑度を調節する場合は,尤度比LR(Z)にかえて $LR(Z)^{K(Z)^e}$ を評価基準としている.aは分析者が設定するパラメータであり,この値を0から1の間で変化させることで,検出される集積領域の形状複雑度を調節することができる.aが1に近いほどK(Z)の制限が強まり,コンパクトな形状の集積領域を,aが0に近いほど制限が弱まり,柔軟な形状の集積領域を検出できる.

(3) 遺伝的アルゴリズムを用いた地域単位における柔軟な形状の集積検出法⁶

地域単位での集積検出において、柔軟な形状の集積領域を検出する手法の1つである.遺伝的アルゴリズム(GA)を用いて、集積領域となる地域の組み合わせを探す.GAの各個体は候補領域を表し、その領域を構成する地域の組み合わせを染色体情報として持っている.空間スキャン統計によって集積を特定するため、個体は候補領域の尤度比で評価される.

GAを用いて良い解を算出するためには、問題に適した交叉方法を導入することが重要である。次に、地域単位分析におけるアルゴリズム中の交叉方法を紹介する.

ランダムに選択した2個体X,Yに共通する地域があれば交叉を開始する.共通地域を核とし,Xの要素中で最も核から遠い1地域を除いたものと,Yの要素中で最も核に近い1地域を結合し,子とする.次に,現在残っているXの要素中で最も核から遠い1地域をはずし,Yの要素中で最も核に近い1地域を結合し,子とする.この操作をXの要素が残り1地域になるか,Yの要素がも核に最も遠い1地域を残して全て結合されるまで繰り返し,この間に生成されたX,Yの各要素を含む組み合わせはそれぞれが1個体となり,子世代を形成していく.

アルゴリズムが終了するまで以上の交叉を各世代ごとに繰り返し、最終世代中で最も優秀な値を持つ個体がMLCとなる.

3. 形状複雑度を考慮したネットワーク上の集積 領域検出法の提案

(1) ネットワーク上における形状指標関数の提案

地域単位の分析では、候補領域Zの面積と、コンパクトな形状の面積の比をとった関数で形状を評価した。ネットワーク上における分析では、地域単位での「面積」に相当する、リンクの「長さ」をの比を用いることで、候補領域Zの形状指標関数K_b(Z)を定義する。

まず、K₁(Z)を定義するためには、ネットワーク上におけるコンパクトな形状を考える必要がある。本研究では、ネットワーク上におけるコンパクトな形状を「中心リンクから各端点までの距離が全て等しいリンクの集合」と定義する。これは、ユークリッド空間上における一般的な円の「中心からの距離が全て等しい」という性質を参考にしている。図-1はコンパクトな形状の例である。赤線が中心リンクを表し、青線が中心リンクから延びている枝を表す。中心リンクから端点ノードp1~p6までの距離が等しい。

さて、中心リンクが事前に与えられれば、コンパクトな形状のネットワークは容易に作ることはできる。しかし、ネットワーク上の形状指標を考える場合には、与えられた候補領域Zの形状を評価することであり、そのためにはZと同じ中心リンク、同じ総延長を持つコンパクトな形状のネットワークを作らなければならない。

本研究では、ネットワークの中心リンクを、「ネットワーク中で最長となる最短経路中で、その距離の1/2に位置するリンク」と定義し、与えられたネットワーク形状から中心リンクを探す。

図-2は、中心リンク検索の手順を表す、左図で、黒線と赤線で構成されてネットワークを、候補領域Zとする.

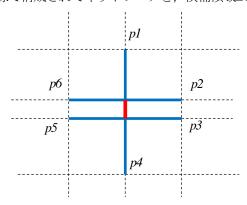


図-1 円形ネットワークの例

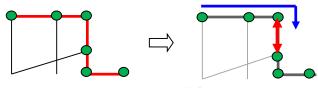


図-2 中心リンク検索手順

この中で最長の最短経路が赤線で示されており、5本の リンクで構成されている. 右図の青矢印は、経路端から その経路長の半分を表しており、赤矢印のリンクがこの 候補領域Zの中心リンクと選ばれる.

Zの中心リンクが選別できたら、Zのネットワーク長 を越えるまで、中心リンクからのネットワーク距離の短 い順にリンクを結合し、ネットワークを作成する.

以上の手順で見つかったネットワークを、候補領域Z に対応したコンパクトなネットワークとする. ここで, L(Z)をZ中で最長となる最短経路長、CL(Z)をZと同じ中心 リンク、同じ総延長を持つコンパクトなネットワーク中 で最長となる最短経路長とすると、ネットワーク上の形 状評価関数は式(3)で表される.

$$K_n(Z) = \frac{CL(Z)}{L(Z)} \tag{3}$$

図-3に $K_n(Z)$ の値の例を載せる. $K_n(Z)$ はZが複雑な形状 であるほど0に近く、円形に近いコンパクトな形状であ るほど1に近い値をとる.

(2) 遺伝的アルゴリズム (GA) を用いたネットワーク上 における柔軟な形状の集積領域検出法

形状指標関数Kn(Z)は、柔軟な形状の集積領域検出法 と組み合わせて使用する. 本研究では、地域単位分析で GAを用いて柔軟な形状の集積領域を検出する手法を参 考に、ネットワーク上の柔軟な形状の集積領域検出法を 提案し、これに形状指標関数を適用する.

なお, 各個体は候補領域となるネットワークを表し, ネットワークを構成するリンクの集合を染色体情報とし て持ち、尤度比で評価される.以下に、GAの各段階に ついての解説を行う.

a) 初期世代の生成

点事象が存在しているリンクを1本選び、個体の生成 を行う. 隣接リンク中から、点事象が存在しているリン クを優先的に結合し、ネットワークを拡大する. この操 作をリンク数が任意の値に達するか、個体の尤度比が任 意の値を下回るまで続け、個体を生成する。 同様の操作 を分析対象地域中における、点事象が存在する全リンク に対して行い, 初期世代を作成する.

世代中からランダムに2個体X'、Y'を選び出し、共通 リンクが存在すれば交叉を開始する. 共通リンクを核に し、核とX'の要素中で最も核に近いリンク1本とを連結 させたものを用意する. そこにY'の要素を核に近い順に 1リンクずつ、結合していく. それぞれの組み合わせが それぞれ1個体となる. Y'の要素がすべて結合されたら 次に、核とX'の要素中で最も核に近いリンク2本を結合 させたものに対し、再びY'の要素を結合していく、この 操作を、X'の全要素、Y'の全要素が結合した個体が生ま れるまで繰り返すことで、1組の交叉が完了する.

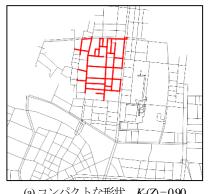
c) 淘汰

本研究で提案した交叉方法は、非常に多くの子を生成 する. そのため子世代の中には、わずかな遺伝子の違い しかない、非常に似通った個体が多く存在する可能性が ある、似た個体のみで次世代が構成されてしまうと、世 代中の個体の多様性が失われ、局所解に陥る可能性が高 い. そこで本研究では、1組の交叉から次世代へ進める 個体数に制限を設ける、1組の交叉終了時に、最も優秀 な個体のみを取り出し、子のリストへ入れる. これを各 交叉で行い、それぞれの交叉で最も優秀な個体のみで構 成される子のリストができる、次世代を構成する個体は、 親世代の上位90%の個体と、子のリスト中、上位から残 りの10%分の数の個体で構成するこことする.

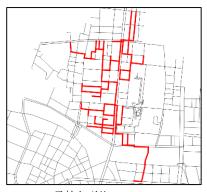
以上の操作を終了条件まで繰り返し、最終世代で最も 優秀な個体をMLCとする.

(3) 形状制限の適用

形状制限は、GAの交叉の段階で行う. 生成された子 の個体毎に形状指標値K,(Z)を計算し、事前に設定した 閾値を上回る個体のみ生き残るようにする. その結果, 最終的に検出される集積領域も閾値以上の形状指標値を もつ個体となり、閾値の設定によって任意の形状複雑度 を持つ集積領域を検出できる.



(a) コンパクトな形状 K_n(Z)=0.90



(b)柔軟な形状 K_n(Z)=0.27

図-3 ネットワーク形状による形状指標値の違い

4. 適用例

本研究の適用可能性を,仙台市青葉区と東京都千代田 区の道路ネットワーク上の店舗集積検出で検証した.

道路ネットワークデータは、2012年12月にOpen Street Mapで作成されたデータを用いた.店舗データは、座標付き電話帳「テレポイント®Pack!」に基づいており、仙台市青葉区については2010年9月に電話帳発行された全件版、東京都千代田区については2010年3月に電話帳発行された全件版を使用した.仙台市青葉区に関しては居酒屋の店舗データ、東京都千代田区に関しては精密機器店の店舗データを用いる.

図-4は、仙台市青葉区の道路ネットワークと居酒屋店舗位置を表す。道路ネットワークは総延長1,358km、11,368本のリンクで構成される。居酒屋店舗数は501店舗である。仙台市青葉区に関しては、以後図-4中の赤枠内だけを表示する。

また、図-5は. 東京都千代田区の道路ネットワークと、緑点が精密機器店の位置を表す. 道路ネットワークの総延長は277km, リンク数は4,501本である. 精密機器店の店舗数は772店舗である. 東京都千代田区のデータに関しては、以後図-5中の赤枠内だけを表示する.

まず、仙台市青葉区における居酒屋の集積領域を、形状複雑度を調節しながら検出した.ケース1が形状制限なし、ケース2~ケース4にかけて、形状制限を強くして検出している.図-5中の緑点が居酒屋位置、黒線が道路ネットワーク、赤線が検出された集積領域を表し、表-1は検出結果の指標値を示す.

ケース1では、街の中心部全体の範囲から大きな集積を検出している。ケース2、ケース3では、それぞれの形状制限に応じてコンパクトな形状の集積領域が検出できている。ケース4では強めの形状制限を行ったが、円形に近い形状ということで、本来は集積があるとは言えないであろうリンクも含めて検出してしまっている。おそらく固定された形状での検出ではこのような結果が検出されるので、ケース2、ケース3のような、コンパクトな形状であり、その中でも一定の柔軟性を持った結果を検出できるのはこの手法ならではのものだと言えるだろう。

続いて、東京都千代田区における精密機器店の集積を、形状制限を行いながら検出する. 図-7、表-2は検出結果である. ケース1が形状制限なし、ケース2~ケース4にかけて、形状制限を強くして検出している. この適用例でも、各形状制限時に、より集積が強い箇所をコンパクトに検出することができている. ケース4では、仙台市青葉区での適用例と同様に、本来は集積があるとは言えないリンクも含めて検出されている. それに対してケース2、ケース3は、コンパクトな形状ながらも一定の柔軟性を持った集積領域を検出できていることがわかる.

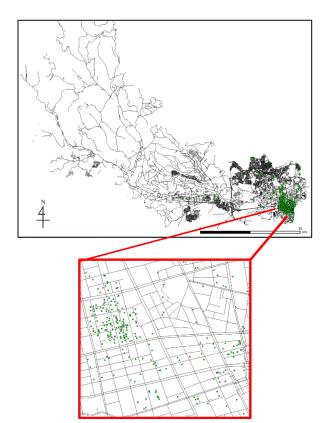


図-4 仙台市青葉区道路ネットワークと居酒屋店舗位置

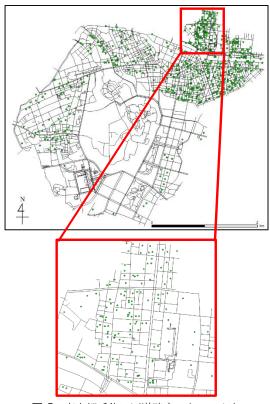
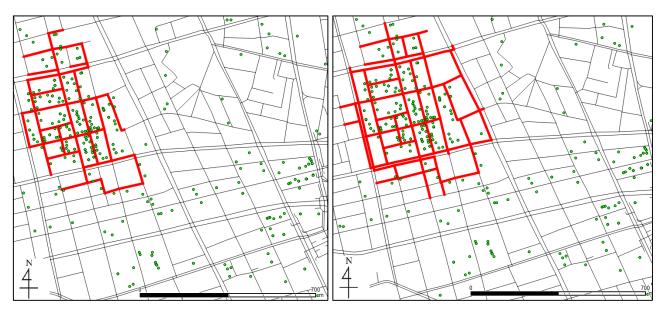


図-5 東京都千代田区道路ネットワークと 精密機器店店舗位置



(a) ケース1: 形状制限なし

(b) ケース2: 形状指標値0.6以上



(c) ケース 3: 形状指標値 0.7以上

(d) ケース4: 形状指標値0.9以上

図-6 仙台市青葉区,居酒屋の集積検出結果

表-1 仙台市青葉区,居酒屋の集積検出結果

	形状制限	対数尤度比	店舗数(店)	総延長(m)	検出結果の Kn
ケース 1	なし	1,444.6	350	9,051	0.35
ケース 2	$K_n \ge 0.6$	1,115.7	270	6,025	0.60
ケース 3	$K_n \ge 0.7$	1,042.4	245	4,653	0.71
ケース 4	$K_n \ge 0.9$	970.5	243	5,979	0.90

5. 結論

本研究では、ネットワーク上の集積領域形状指標を提案し、任意の形状複雑度設定に応じた点事象集積領域の 検出手法を提案した. 仙台市青葉区、東京都千代田区の 店舗位置データを用いて集積検出を行い提案手法の実行 可能性を確認した.

これまでネットワーク上の点事象集積領域検出では, 強い形状制限を持ち円形に近いコンパクトな形状の集積 領域の検出,あるいは形状に制約を課さず,柔軟な形状

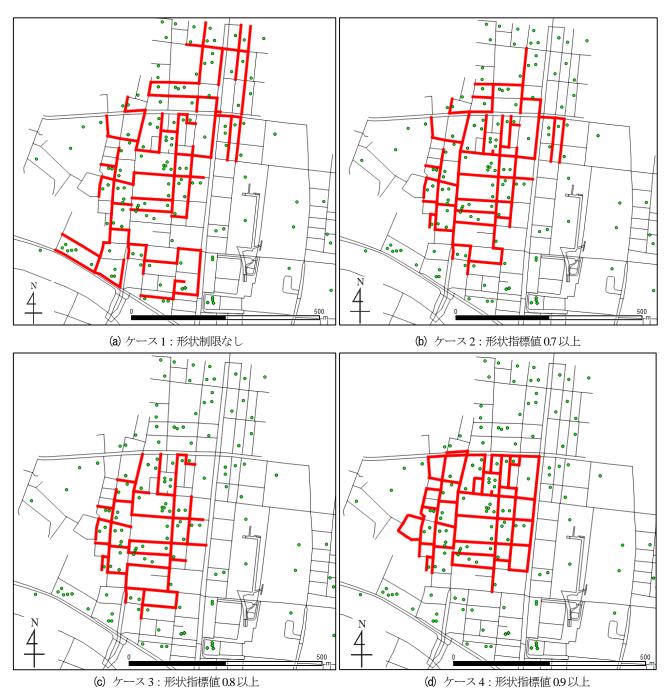


図-7 東京都千代田区、精密機器店の集積検出結果

表-2 東京都千代田区、精密機器店の集積検出結果

	形状制限	対数尤度比	店舗数(店)	総延長(m)	検出結果の Kn
ケース 1	なし	238.9	155	5,271	0.35
ケース 2	<i>K</i> _n ≥0.7	154.8	109	4,104	0.70
ケース 3	<i>K</i> _n ≥0.8	123.3	81	2,677	0.80
ケース 4	<i>K</i> _n ≥0.9	100.6	77	3,222	0.91

の集積領域の検出のいずれかしか行うことができなかった. コンパクトな形状での検出は時に本来集積領域とは言えない領域を含んでしまうことがあり、また、柔軟な形状での検出は時に解釈不可能な結果を検出することが

あった. 本研究で提案した手法は、その両者の中間の形状で集積領域を検出することができ、ネットワーク上の分析の自由度を高めることができた.

謝辞:本研究は、東京大学空間情報科学研究センターの空間データ利用に伴う共同研究(No. 456)による成果であり、座標付き電話帳「テレポイント®Pack!」のデータを利用した。

参考文献

- Kulldorff M and Nagarwalla N. 1995. Spatial disease clusters: Detection and inference. Statistics in Medicine 15, 707-715.
- Kulldorff M. 1997. A spatial scan statistic. Communications in Statistics-Theory and Methods 26, 1481-1496.
- Tango T and Takahashi K. 2005. A flexible shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*, 3-17.
- Duczmal L, Kulldorff M and Huang M. 2006. Evaluation of spatial scan statistic for irregularly shaped clusters. *Journal of Computational and Graphical Statistics* 15:2, 428-442.

- Duczmal Land Assunção R. 2004. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics* and Data Analysis 45, 269-286.
- Duczmal L, Cançad A, Takahashi R and Bessegato L. 2007. A genetic algorithm for irregularly shaped spatial scan statistics. Computational Statistics and Data Analysis 52,43-52.
- Yao Z, Tang J and Zhan B. 2011. Detecting 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.
- Shiode S 2011. Street-level spatial scan statistic and STAC for Analysing street crime concentrations. *Transactions in GIS* 15(3), 365-383.
- 9) 井上 亮, 古郡美佳: リンクの接続関係を利用したネットワーク上の点事象集積領域の検出法の開発, 土木計画学研究・講演集, Vol.47, CD-ROM.

(2014.4.25 受付)