

# 柔軟形状の点事象集積領域検出手法の開発

井上 亮\*

\* 東北大学 大学院情報科学研究科 人間社会情報科学専攻

近年、詳細な位置情報を有する地理空間情報の整備と公開が進んでおり、小領域を対象とする空間分析の実行環境が整いつつある。特に社会経済活動に関する小地域分析を行う場合、道路等ネットワーク形状を考慮し空間的近接性を評価することが欠かせない。そのため、ネットワーク状の分布を分析する手法が提案されてきたが、地理空間情報の一種である点データの空間分布から集積領域を検出する分析に関しては、ネットワーク形状を考慮した手法は十分に議論されていない。

そこで本研究は、ネットワーク上の点集積領域検出について新たな手法を提案する。ネットワーク上の各リンクを集積候補領域の最小単位と考え、リンクの接続関係を利用してリンク集合を集積領域として抽出する手法を開発し、その有効性を電話帳の店舗位置データへの適用を通して確認する。

**Key Words :** *cluster detection, network, point event data, connectivity of links*

## 1. はじめに

近年、測位技術の向上・普及や政府の情報公開施策の進展に伴い、詳細な位置精度を持つ多種多様な空間データを入手できる環境の整備が急速に進んでいる。

この空間データの利用環境の変化を生んでいる要因の一つに、測位技術の向上および社会一般への普及によって、専門家以外でも簡単に空間データを作成できる環境が整備されてきたことが挙げられる。GPS（全地球測位システム）を用いた測位機能の携帯電話への搭載が一般化し、経路探索を始めとする位置情報を活用した様々なサービスが提供されるようになった今、日常生活の中で測位を行い位置情報サービスを利用することは珍しいことではなくなった。また、測位技術を活用し、ある地点に関する撮影画像やコメントに対して位置情報（ジオタグ）を付与し、その情報を他者とネット上でリアルタイムで共有する行為も普及しつつある。空間データの作成は、数年前までは専門家など限られた者だけが行う作業であったが、現在では、誰もが簡単に空間データを作成し、即時的に共有できる環境が整っている。このように、詳細な位置情報を持つ空間データが日々大量に生み出されるようになっており、これらを活用してより詳細な地域分析を即時的に実施できる環境が整いつつある。

また、政府が持つ統計・行政情報を公開し、社会で有効に活用することを目指した施策の推進も、空間データの利用環境に大きな変革をもたらしている。まず、2007年に施行された統計法改正に向け、政府が収集する統計情報を国民の共有財産と考え、有効に活用することを目

的に利用環境の整備が図られてきており、国勢調査を始めとする様々な空間的な位置に関する情報が入手しやすくなってきた。更に、「政府の透明性向上・市民参加・官民連携」を旗頭にオープンガバメント施策が世界各国で進められており、政府が所有する様々な行政情報も一般に公開されつつある。世界に先駆けオープンガバメントに取り組んだオバマ政権は、国勢・経済・環境など各種の行政情報を提供するサイト「Data.gov」を開設し、特に空間データに関しては「Geo.Data.gov」上で情報提供が行われている。日本でも、2010年5月に高度情報通信ネットワーク社会推進戦略本部が決定した「新たな情報通信技術戦略」で、行政情報の公開や政策決定への国民参加などオープンガバメントに向けた取り組みを推進していくことが定められており、今後の情報提供が期待される。

また、オープンガバメント施策では、公開した情報の利用を促進するため、コンピュータが解釈しやすいデータ形式で提供し、データの二次利用を促すことに注力されている。政府が公開するデータと民間が作成するデータを融合し、付加価値の高い情報を生むサービスの提案などが行われている。行政が提供する犯罪発生地点や公共交通サービス・飲食店の情報、飲食店におけるtwitterのつぶやきを利用して、飲み屋の情報とそこからの安全な帰宅経路を提供するサービスを提案した“Stumble Safely<sup>2)</sup>”など、詳細な位置情報を持つ空間データに基づくサービスや地域分析の可能性は膨らんでいる。

以上のように、主にデータ取得や公開に関する環境の整備に伴い、詳細な位置情報を有する空間データを活用した小地域単位の空間分析を行う環境が整いつつある。

本研究は、このようなデータを対象とした分析手法の中で、点事象の空間分布からその集積領域を抽出する手法に着目する。

点事象の分布は、事象間の距離を用いて分析することが一般的であるが、通常、ユークリッド距離が用いられることが多い。しかし、詳細な位置情報を活用し小地域における空間分布分析を行う場合、特に社会経済活動に関する分析では、道路などのネットワーク形状を考慮することが不可欠である。そこで、ユークリッド空間上において点事象の集積領域を検出する手法である空間スキャン統計<sup>3)</sup>をネットワーク上に拡張した研究<sup>4)5)</sup>が行われている。しかし、これらの研究では、任意の地点から一定のネットワーク距離内の範囲を集積領域として設定する手法を提案しているが、点事象が分布していない領域も含む集積領域を出力する可能性を有するという弱点を備えている。

そこで本研究では、既存手法の問題点解決を目指し、分析対象ネットワークの各リンクを集積検出の最小単位、接続関係にあるリンクの組み合わせを集積候補領域と設定し、分析する手法を提案する。提案手法の有効性を仙台市中心部の店舗立地データを用いて検討を行う。

## 2. 既往の点事象集積の検出法

本章では、ユークリッド空間上における点事象の集積領域検出法である空間スキャン統計<sup>3)</sup>について説明を行った後、ネットワーク上での集積領域検出に拡張した一連の研究<sup>4)5)</sup>について紹介する。

空間スキャン統計では、まず、点事象の空間分布のモデルを仮定するが、通常は、点事象分布がポアソン過程に従うと仮定する。その仮定の下で、任意の集積の候補領域について、対立仮説 $H_1$ 「候補領域内の点事象の発生率は領域外より大きい」と、帰無仮説 $H_0$ 「候補領域の内外の点事象の発生率は等しい」の尤度比を算出し、各候補領域の集積性を判断する。分析対象領域内を走査して、尤度比が最大となる候補領域を検索し、抽出された領域を最も有意性の高い領域MLC (Most Likely Cluster)とする。

MLCの有意性は、Monte Carlo simulationで作成した点事象の無作為分布から得られるMLCの尤度比の分布と、分析対象の点事象分布から抽出されたMLCの尤度比を

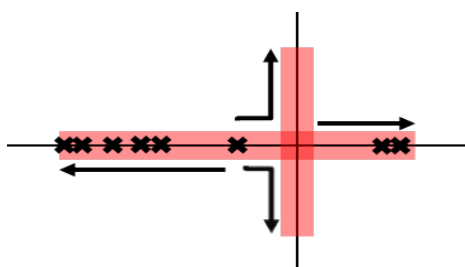


図-1 ネットワーク距離で拡大する集積候補領域

比較し、検定を行う。

この空間スキャン統計を、ネットワーク上の点事象集積検出に拡張したのがShiode<sup>3)</sup>である。空間スキャン統計と同様に、まず点事象がネットワーク上でポアソン分布に従って分布していることを仮定した上で、帰無仮説・対立仮説の尤度比を算出し、候補領域の集積を評価する。ネットワーク上集積の候補領域は、ネットワーク上の任意の基点から一定のネットワーク距離以内の領域として設定(図-1)することを提案しており、対象ネットワーク上で候補領域を走査してMLCを抽出し、分析対象のネットワーク上で作成した点事象の無作為分布の下で得られる最大尤度比分布との比較を通して、MLCの有意性を検定する。

Shiode<sup>3)</sup>は、作為的にネットワーク上の集積を作成した点事象分布のシミュレーションデータを利用し、提案手法でネットワーク上の点集積が検出できること、また、ユークリッド空間上の集積検出方法では誤った集積を検出する可能性があることを例示している。また、Shiode and Shiode<sup>4)</sup>は、上記の手法を時間方向にも拡張し、ネットワーク空間と時間上での集積検出を提案している。

ただし、この手法は、集積候補領域をネットワーク距離を用いて設定するため、強い集積がみられる領域から距離が近いが点事象が分布していない領域も、集積領域として抽出される場合があるという限界を有している。例えば、ひたつくりの多発地点を抽出する場合を考える。ひたつくりは、通常、人通りの少ない裏通りで多発し、人通りの多い大通りでは発生しない。しかし、集積候補領域をネットワーク上の任意の点から一定のネットワーク距離以内の範囲として設定する場合、ひたつくり多発地域に近いが実際には犯罪が発生していない大通りも集積候補領域として一緒に選ばれてしまう可能性がある。

また、ネットワーク上の任意地点を中心として、任意のネットワーク距離以内の範囲を集積の候補領域と設定することを提案しているが、基点やネットワーク距離の設定は膨大となるため、可能な候補領域すべてを検索することは現実的ではない。既往研究においてデータに対して適用する際には、分析対象範囲内に格子状に基点を配置し、各基点から一定の刻みで設定したネットワーク距離以内の範囲として候補領域を作成して、尤度比算出・MLCの検索を行っている。なお、限定された地点を基点に候補領域を設定する場合でも、ネットワーク上の領域分割が必要であるため、計算負荷は大きい。

以上のように、ネットワーク上における点事象の集積領域を抽出する手法は提案されているものの、分析には限界を有している。

### 3. リンク単位の候補領域設定による集積領域抽出手法の提案

本研究では、分析対象ネットワークの各リンクを集積検出の最小単位とし、接続関係にあるリンクの組み合わせを集積候補領域として設定する方法を提案する。

各リンクを集積検出の最小単位とすることで、候補領域設定の基点設定に関する問題や、ネットワーク距離に基づく領域分割という計算負荷の大きい操作から解放される利点を備えている。更に、リンク毎に候補領域に含めるかを考慮することができるため、基点からのネットワーク距離を用いて候補領域を設定する手法とは異なり、強い集積がみられる領域から距離が近いが点事象が分布していない領域を含めない候補領域設定も可能になる。

本研究で提案する、ネットワーク上の集積領域検出法は以下の手順で行う。

提案手法では、空間スキャン統計と同様に、ネットワーク上の点事象分布はポアソン分布に従い、各事象は互いに独立に分布していると仮定し、尤度比を用いたMLCの検出を行う。

候補領域設定では、リンクの接続関係を用いて、分析対象範囲内のリンクの組み合わせを作成する。例えば、**図-2**の点事象分布を分析する場合、

- {1}, {2}, {3}, {4}, {1,2}, {1,3}, {1,4}, {2,3}, {2,4}, {3,4}
- {1,2,3}, {1,2,4}, {1,3,4}, {2,3,4}

の14通りの候補領域を設定し、尤度比算出を行う。ただし、点事象が分布していないリンクを集積候補領域に含めると、尤度比は必ず低下するため、分析対象から除外し候補数を減らしたい。一方、2つの候補領域を接続して新たな候補領域を作成する場合、点が分布していないリンクを含めても、より尤度比の大きい結果が得られる可能性がある。候補数の削減と尤度比の高い候補領域設定のトレードオフを考慮し、両端に点が分布するリンクが接続していないリンクを候補から外すこととする。**図-2**のネットワークの場合、リンク2・4を候補から外し

- {1}, {3}, {1,3}

の3候補を作成して尤度比を算出し、最大尤度比を算出する{1,3}をMLCとして抽出する。

この候補数の削減を実施しても、対象領域のネットワークが大きくなると、候補領域数は指数関数的に増加するため、対象領域全域から一度に候補領域を構成することは現実的ではない。そこで、ユークリッド空間上での点事象集積をを目指した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<sup>9</sup>を参考に、対象領域を小領域に分割し、各小領域で尤度比の高い集積候補領域を抽出、その後、抽出された候補領域の組み合わせから更に尤度比の高い領域を探索するという操作を繰り返し、対象領域全体での集積領域を検出する。

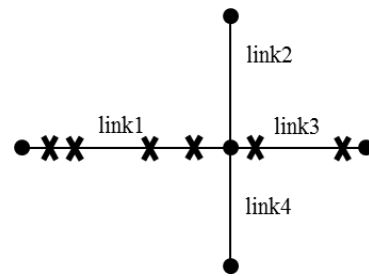


図-2 点事象分布とネットワーク形状例

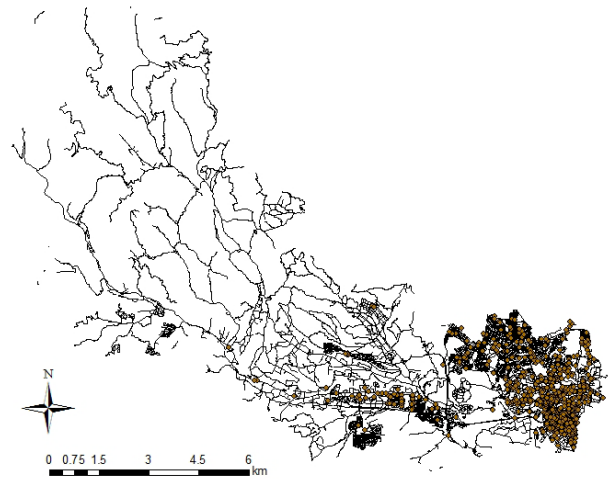


図-3 分布対象ネットワークと居酒屋分布

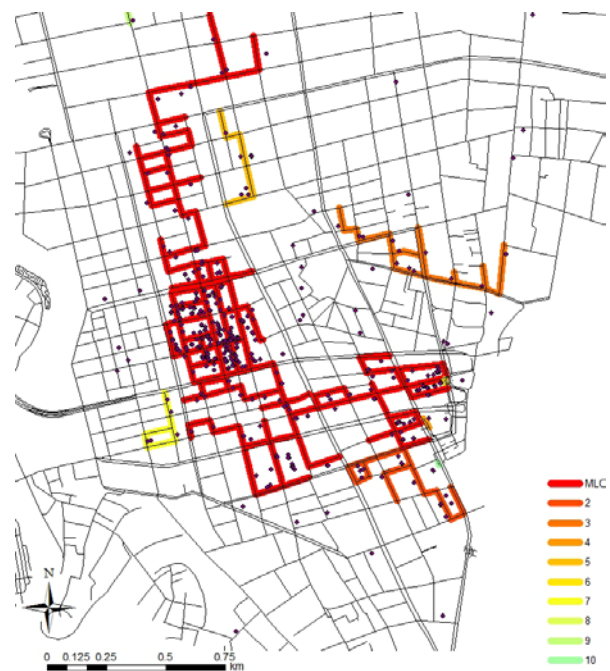


図-4 仙台市中心部における居酒屋集積地域検出結果

なお、空間スキャン統計に基づく本提案手法は、同時に1箇所の集積領域の抽出・検定しか実行できない。これは、尤度比算出に用いる対立仮説・帰無仮説が、集積領域が対象領域内で1箇所しか存在しないことを前提としているためである。空間スキャン統計を用いて、複数

表-1 居酒屋集積地域検出結果

集積領域	対数尤度比	店舗数		リンク長	
		(件)	(%)	(km)	(%)
1	1390	363	72.5	12.99	0.96
2	52	14	2.8	1.27	0.09
3	41	12	2.4	1.62	0.12
4	38.6	6	1.2	0.04	0.00
5	33.7	8	1.6	0.56	0.04
6	29	4	0.8	0.01	0.00
7	19.8	5	1	0.51	0.04
8	17.6	4	0.8	0.27	0.02
9	15.5	4	0.8	0.48	0.04
10	13.6	2	0.4	0.01	0.00

の集積領域を検出する場合には、以前の分析により抽出された集積領域を所与とした条件の下で集積抽出を行い対応することも可能であるが、複数の集積領域が存在することを前提とした分析と異なった結果をもたらすことには注意が必要である。

#### 4. 提案手法の適用

仙台市青葉区内の店舗立地データに対して提案手法を適用し、その有効性を評価する。

Open Street Mapの道路ネットワークデータと、座標付き電話帳「テレポイント@Pack!」の居酒屋立地地点データを使用し、集積領域の検出を行った。総リンク数は11368本、総リンク長は1358km、総店舗数は501件である(図-3)。

提案手法を適用した結果、計10箇所の集積領域が検出された。その中でMLCとして検出された地域を拡大した結果を図-4に示す。黒点が居酒屋の立地点を表し、色が付けられたリンクが集積領域を表す。検出された集積領域の概要を表-1に示す。

提案手法によって、対象領域内で最も集積が顕著な領域が抽出できることが確認された。抽出されたリンクは、定禅寺通りと広瀬通り、それに交差する国分町通りや稲荷小路など仙台市中心部の繁華街である国分町の主な街路であり、居酒屋の集積地として尤もらしい。

また、この適用例では計算は数秒程度で完了しており、一定の適用可能性があることが確認された。

#### 5. おわりに

本研究では、ネットワーク上の点事象集積検出に関して、各リンクを集積検出の最小単位とし、接続関係にあるリンクの組み合わせを集積候補領域として設定して分析する手法を提案した。店舗立地データを用いた適用では、集積性の高い領域のみを抽出することに成功し、提案手法の有効性が確認された。

なお、複数の集積領域の同時抽出法<sup>7)</sup>を利用した拡張や、時空間集積検出への拡張は、今後の課題である。

謝辞：本研究は、本研究はJSPS科研費24241053の助成を受けて行われた。また、東京大学空間情報科学研究センターの空間データ利用を伴う共同研究(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

- 1) BIGAPPS NYC 2013. 2013. <http://nycbigapps.com/>, (accessed on 2013/5/4).
- 2) Stumble Safety, 2008. [www.appsfordemocracy.org/stumble-safely/](http://www.appsfordemocracy.org/stumble-safely/).(accessed on 2013/5/4).
- 3) Kulldorff, M. 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, Vol.26, No.6, pp.148-196.
- 4) Shiode, S. 2011. Street-level spatial scan statistic and STAC for analyzing street crime concentrations. *Transactions in GIS*, Vol.15, No.3, pp.365-383.
- 5) Shiode, S. and Shiode, N. 2012. Network-based space-time search-window technique for hotspot detection of street-level crime incidents. *International Journal of Geographical Information Science*, published online (in print).
- 6) 糟谷志帆, 井上 亮. 2012. 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. 地理情報システム学会講演論文集, Vol. 21, C-6-3, CD-ROM.
- 7) Mori, T. and Smith, T.E. 2009. A probabilistic modeling approach to the detection of industrial agglomerations. *KIER Discussion Paper*.