

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24年 5月9日現在

機関番号：15301

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21700305

研究課題名（和文） 空間データの構造解析とその集積性の研究

研究課題名（英文） Study on the structure analysis and spatial clustering for spatial data

研究代表者

石岡 文生（ISHIOKA FUMIO）

岡山大学・大学院法務研究科・助教

研究者番号：20510770

研究成果の概要（和文）：

ある郡における病気の発生率等の様な領域毎に得られる各種の空間データに対して、有意に高い領域群（集積性、クラスター）を検出する新たな手法を確立した。具体的には、空間データが本来もつ階層構造の上位に位置する領域から優先的にホットスポット領域を探索していく。これにより、大量の領域からなる空間データや、多次元空間データ、時空間データなど、先行研究では困難であった各種の空間データに対するホットスポット検出を可能にした。

研究成果の概要（英文）：

Several approaches are used to detect hotspot for various kinds of spatial data. We have proposed a technique using an echelon analysis as the hotspot detection method. The echelon's hotspot detection method enables the detection of spatial clusters having various shapes and significant likelihood ratios, because its searches are based on the essential spatial structure of data. Accordingly, the echelon's hotspot detection method can be applied to extensive spatial data with thousands of regions and detect multi-dimensional or space-time hotspot.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,000,000	300,000	1,300,000
2010年度	700,000	210,000	910,000
2011年度	500,000	150,000	650,000
総計	2,200,000	660,000	2,860,000

研究分野：総合領域

科研費の分科・細目：情報学、統計科学

キーワード：空間・環境統計、エシエロン解析、ホットスポット

1. 研究開始当初の背景

ある郡における病気の発生率や、ある区画の有害物質の汚染濃度のように、領域毎に得られる各種のデータ（空間データ）に対する集積性（ホットスポット）を検出することは、環境状況の把握や、将来の環境や健康へ

の影響を早期に発見するためにも大変重要といえる。これまで各種のホットスポット検出手法が提案されてきたが、大量の領域からなる空間データや、多次元空間データ、時空間データに対しては適用が困難であった。

2. 研究の目的

本研究の目的は、Myers 博士が 1997 年に提案した Echelon 解析を利用することにより、これまでと異なる新たなホットスポット検出手法を確立することにある。具体的には、(1) エシェロンを用いたホットスポット検出の信頼性についての検証、(2) エシェロンを利用した新たな空間分類手法の提案、(3) エシェロンによるホットスポット検出手法のアルゴリズムを発展させる事によって、前述した大量の領域からなる空間データや、多次元空間データ、時空間データに対するホットスポット検出を可能にする。さらに、それらを具体的なデータに適用する。

3. 研究の方法

エシェロン解析とは、空間データに対して、空間的な位置を表面上のデータの高低に基づき分割し、空間データの位相的な構造を系統的かつ客観的に見つける解析法である。得られたデータに対しエシェロン解析を行い、それによって作られた位相的な階層構造に基づいてスキャンを行うことで、任意の形状をしたホットスポットの検出が可能となる。

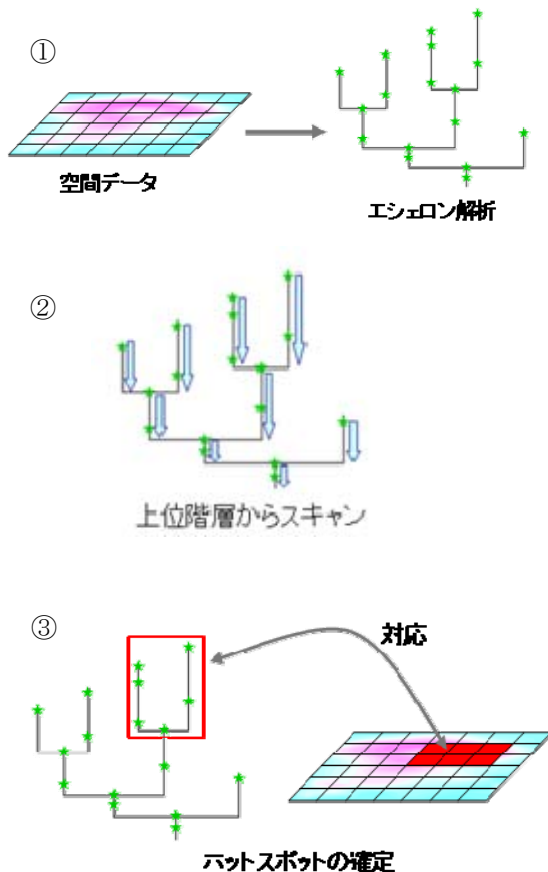


図 1. エシェロン-ホットスポット検出法

具体的な手順は、①エシェロン解析によって、位相的な階層構造を明らかにし、②上位

の階層を構成する領域から順にスキャンしていき、③最も空間スキャン統計量が高くなった領域までをホットスポットとする。また、得られたホットスポットの有意性はモンテカルロ検定による p 値で判定する (図 1)

エシェロン解析によって、データが本来もつ空間位相構造を明らかにし、その得られた階層構造のピークを形成する領域から優先的に走査 (スキャン) することによって、空間データの構造に基づいたホットスポット検出が可能となる。また、ホットスポットとなりそうな領域を総当たりにスキャンしていくのではなく、空間的に近い領域でいてかつ値の高い領域から優先的にスキャンすることも可能となる。これより、スキャンされる領域の総数を格段に縮減することができ、大量データへの適用が可能となる。本研究では、(1) ランダムに値を付与した空間データを用いたシミュレーションを利用してエシェロン-ホットスポット検出法の信頼性について検討する。また、研究目的 (2) のエシェロンを利用した新たな空間分類法については、以下の方法を提案する。いま、地形図の断面図のような一次元空間データを考える。図 2 は、A から W と名前が付けられた 23 個の領域と、その領域が持つ値を示している。図 3 は、図 2 の空間データの断面図を表している。

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
ID	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
$h(i)$	1	2	3	4	3	4	5	4	3	2	3	4	5	6	5	6	7	6	5	4	3	2	1

図 2. 一次元空間データ

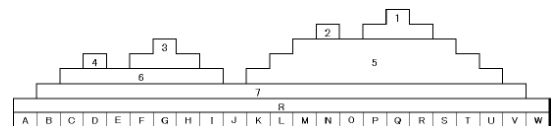


図 3. 同じ位相領域への分割

このような断面図が与えられた場合、位相的に同じ領域 (エシェロン) へ分けることができる。図で与えられている番号がエシェロン番号であり、1 から 4 までのピークと 5 から 8 までのファウンデーションから構成される。図 2 のデータに対し、4 個のピークを分類の中心と考え、4 つのクラスター (圏) に分類することができる。これらの分類結果を図 4 に示す。本研究では、これら手法を具体的な地域データに対して適応する。

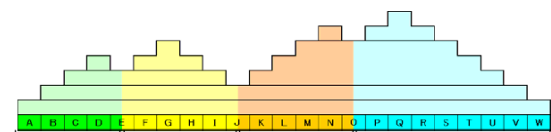


図 4. ピークを中心とした圏による分類結果

さらに、研究目的の(3)に関して、多次元におよぶ空間データであっても、異なる次元の領域間に隣接情報を与えてやることにより、2次元のデンドログラムで表現することが可能になる。また、多次元になることでデータ量は膨大となるが、階層構造に基づいてスキャンすることにより、ホットスポット検出のための計算時間が大幅に短縮できる。ここでは、エシェロン解析を時空間データに適用させるため、各領域のもつ時系列的な隣接情報を様々な形で定義し、空間と時間の両方を同時に考慮した階層構造を実現する。これより、ホットスポットの時系列的な推移の表現を可能にする。

4. 研究成果

(1) エシェロン解析は、測定誤差など、値のわずかな差異によってその結果に変化が生じてしまう。従って、コンピュータによってランダムな値を付置した複数の空間データを作成し、我々の提唱しているエシェロンホットスポット検出法を適用した。何通りものデータを作成してシミュレーションを行う事により、我々の手法の特性を明らかにし、他の手法との差別化を明確にした。その一例として、ここでは平成21年度の都道府県人口10万人当たりの刑法犯罪認知件数のデータに対し、そのホットスポットの信頼性について検討した結果を示す。

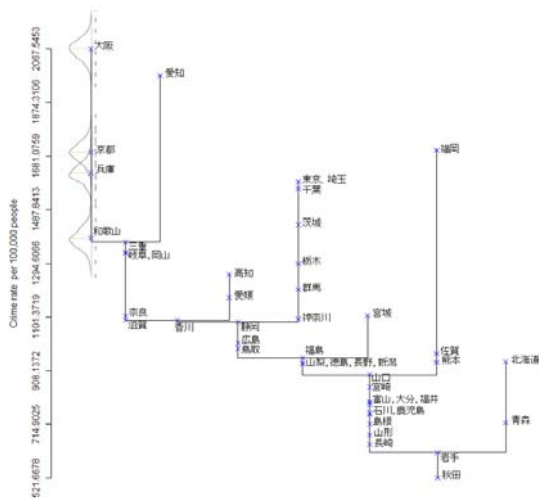


図5. 犯罪率のエシェロンデンドログラム

図5は、母比率 p の二項分布に従うランダムに値を発生させることにより、値に幅をもたせたエシェロンデンドログラムを表している。その階層構造に基づいてホットスポットを検出することにより、犯罪発生件数のホットスポットの信頼性のマップを作成した(図6)。乱数によるわずかな変動でホットスポットとならなくなる領域の存在、逆に影響されないロバスタなホットスポットの存在

を立証した。

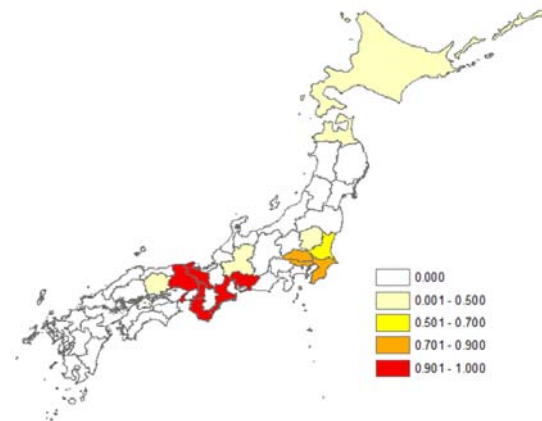


図6. 犯罪率ホットスポットの信頼性マップ

これらの成果を2009年度統計関連学会連合大会、科学研究費補助金による研究集会「時空間統計解析その理論と応用」、JKSC2010、日本計算機統計学会第24回大会、2010年度統計関連学会連合大会などで報告した。

(2) エシェロン解析に基づいた新たな空間データの分類法として、ピークを中心とする圏への分類を新たに提唱した。この手法をノースカロライナ州の乳幼児突然死症候群(SIDS)データに適用した結果を図7に示す。19のピークに基づいて、19の圏に分類することができる。この新手法は、各種の地域データにおける圏の客観的な検出が容易に行える。今後の課題として、種々の地域データ、時空間データなどへの適用が期待される。



図7. SIDSデータの圏への分類

これらの成果をThe 10th China-Japan Symposium on Statisticsなどで報告した。

(3) 解析対象のデータを、時間と空間の広がりの中で観測される時空間データに拡張し、そのデータにエシェロン解析を応用することで、従来の手法では困難であったホットスポットの時系列的な移動、分割、融合、吸収などの推移を明らかにした。特に本研究期間においては、内閣府・国立精神・神経センター精神保健研究所から提供された日本における自殺データへの適用に力を入れた。簡単にデータについて説明すると、354に分けられた日本全国の二次医療圏(2008年3月時点)において、1973年から2007年までの20年間

の年齢別の男女の自殺者のデータになっている。このような大多数な時系列空間データの場合、既存のホットスポット検出手法での適用は困難であり、適用できた場合でも時系列的なホットスポットの推移までを表現することは不可能であった。しかし、本研究によって確立してきたエシェロン解析を利用する事により、このような大容量かつ時系列的な空間データに対するホットスポットの検出が可能である事を証明した。我々はこの自殺データに対し、多角的な角度からホットスポット検出をおこなった。その結果の一つを簡単に紹介すると、男性の自殺は主に東北北部と南九州にホットスポットが検出され、時系列的な変化では初期ではホットスポットが点々と少なかったものが、年を経ていくにつれて大きく広がっており、特に 1998 年以降が急激に増加していることが示された(図 8)。今後も、さらに適用可能なデータを収集し、我々の手法を適用して行く事が望まれる。

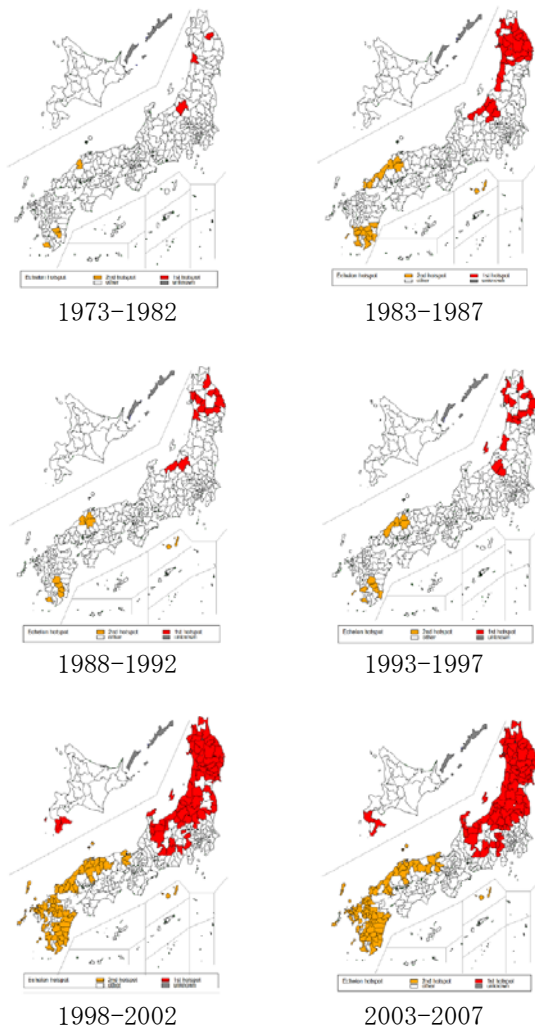


図 8. 日本人男性自殺の時空間ホットスポット

これらの成果を、日本計算機統計学会第 23 回大会、2009 年度新領域融合プロジェクト研究による研究会「大規模データ・リンケージ、データマイニングと統計手法」、日本計算機統計学会第 24 回シンポジウム、研究会「大規模医学データの統計的解析とその応用」、計算機統計学和文誌 23 巻(1)、COMPSTAT2010、ISI2011、2011 年度統計関連学会連合大会、第 11 回西東京統計研究会などで報告した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 25 件)

- ① 石岡文生、栗原考次、Echelon 解析に基づくスキューン法によるホットスポット検出について、統計数理、査読有、60 巻(1)、2012、in press
- ② 富田誠、石岡文生、藤田利治、日本の自殺データにおける時空間解析、計算機統計学、査読有、23 巻(1)、2010、25-43
- ③ Fumio Ishioka、Makoto Tomita、Toshiharu Fujita、Detection of Spatial Cluster for Suicide Data using Echelon Analysis、COMPSTAT2010 Proceedings in Computational Statistics (Edited by Lechevallier, Y. and Saporta, G.)、査読有、2010、1159-1166
- ④ 石岡文生、矢木良、栗原考次、空間データの階層構造の信頼性について、日本計算機統計学会 第 24 回大会論文集、査読無、2010、37-40
- ⑤ 石岡文生、栗原考次、空間データの階層構造の安定性評価について、2010 年度統計関連学会連合大会講演報告集、査読無、2010、166
- ⑥ 石岡文生、栗原考次、大規模空間データに対する集積製の研究、2009 年度新領域融合プロジェクト研究による研究会「大規模データ・リンケージ、データマイニングと統計手法」、査読無、2009、29-36
- ⑦ 石岡文生、栗原考次、ホットスポット検出法の性質評価について、2009 年度統計関連学会連合大会講演報告集、査読無、2009、211

[学会発表] (計 15 件)

- ① Fumio Ishioka、A new method of spatial clustering based on echelon analysis、The 10th China-Japan Symposium on Statistics、2010 年 10 月 16 日、Southwestern University of Finance and Economics, Chengdu, CHINA

- ② Fumio Ishioka、Detection of Spatial Cluster for Suicide Data using Echelon Analysis、COMPSTAT2010 (Computational Statistics 2010)、2010年8月24日、Conservatoire National des Arts et des Metiers, Paris, FRANCE
- ③ 石岡文生、空間データの階層構造の信頼性について、日本計算機統計学会第24回大会、2010年5月14日、東京都立川市統計数理研究所
- ④ Fumio Ishioka、The properties of hotspot detection based on echelon、JKSC2010 (Joint Meeting of Japan-Korea Special Conference of Statistics and The 2nd Japan-Korea Statistics Conference of Young Researchers)、2010年2月2日、50th Anniversary Hall of Okayama University, Okayama, JAPAN
- ⑤ 石岡文生、ホットスポット検出法の性質評価について、2009年度統計関連学会連合大会、2009年9月8日、京都府京田辺市同志社大学京田辺キャンパス

[図書] (計 件)

[産業財産権]

○出願状況 (計 件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 出願年月日：
 国内外の別：

○取得状況 (計◇件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 取得年月日：
 国内外の別：

[その他]

ホームページ等

6. 研究組織

- (1) 研究代表者
 石岡 文生 (ISHIOKA FUMIO)
 岡山大学・大学院法務研究科・助教
 研究者番号：20510770
- (2) 研究分担者
 ()
 研究者番号：
 ()
 研究者番号：
 ()
 研究者番号：
 ()
 研究者番号：
- (3) 連携研究者 ()
 研究者番号：