

氏名	石岡 文生
授与した学位	博士
専攻分野の名称	理学
学位授与番号	博甲第3490号
学位授与の日付	平成19年 9月30日
学位授与の要件	自然科学研究科資源管理科学専攻 (学位規則第4条第1項該当)
学位論文の題目	Studies on Hotspot Detection for Spatial Data (空間データのホットスポット検出に関する研究)
論文審査委員	教授 栗原 考次 教授 垂水 共之 教授 梶原 毅

学位論文内容の要旨

近年、環境リスク解析や環境保全のため、リモートセンシングデータや、ある州内の各郡のガン発生率データなどの、各種の空間データに対して、有意に他から逸脱している領域群（ホットスポット）を検出する事は重要な研究分野である。ホットスポット検出法については、これまでも様々な手法が提案されているが、現在主として使われているものに空間スキャン統計量がある。空間スキャン統計量は、データが得られた地点を中心に円状に領域をスキャンし、有意に高い比率を示す領域を見つける。しかし、円状に領域をスキャンするため円状のホットスポットの検出には優れているが、線状や他の形状をしたホットスポットの検出は不可能である。この問題を補うため、我々はエシェロン解析を利用してホットスポットを検出する方法を確立した。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。このエシェロン解析を用いることにより、従来の手法では困難であった多次元空間データ、空間型大量データ、時空間データ、多変量空間データに対するホットスポットが、容易に検出可能となった。

従来の2次元の空間データへの適用例として、アメリカ・ノースカロライナ州100郡におけるSIDS（幼児突然死症候群）のデータを使い、SIDSが有意に高頻度で起こる地域群をエシェロン解析と空間スキャン統計量を用いて検出した。さらに、同データをローカルモラン統計量、Tangoの集積性、GAM、円状スキャン統計量、flexibleスキャン統計量などの他のホットスポット検出法を用いて解析し、それらの結果とエシェロン解析によって得られる結果とを比較した。

また、複数の2次元空間データを重ね合わせることによって立体的な空間を想定し、そこに新たな隣接情報を定義することで、多次元空間データにおけるエシェロン解析を可能にした。適用例として、汚染物質の地下における移流・拡散シミュレーションによる予測データを用い、そのホットスポットを検出した。このアイデアは、時空間データにも適用可能である。適用例として2期間におけるSIDSデータのホットスポットを同時に検出し、時間の経過によるホットスポットの推移を表現した。

多変量空間データに対しては、次元縮約法の一つである主成分分析のアイデアを利用した。多変量データにおいて主成分分析を行い、各主成分で因子を特徴付けられると思われる変量を検出し、その主成分得点を用いる事でエシェロン解析を可能にした。適用例として、2001年から2003年までの韓国統計庁の上位5番目までの死亡原因データを用いた。また、主成分分析を用いるホットスポット検出法の拡張として、主成分の解釈を容易にするためのバリマックス回転後の因子を利用するホットスポット検出法や、主成分空間にボロノイ図を適用することで地理的な制約を受けないホットスポット検出法について述べた。

論文審査結果の要旨

ホットスポットとは、病気の発生率などの疫学データや環境汚染物質において、有意に集積性が見られる地域のことである。ホットスポットの検出は、将来の環境や健康への影響を早期に発見するためにも重要である。本論文では、空間データの中で特に lattice データに対するホットスポット検出に関する問題について取り扱っている。ホットスポット検出に主として用いられる空間スキャン統計量 (Kulldorff, 1997) は、有意に高い比率を示すホットスポット領域を見つける。この統計量は複雑な従属性を持つため、理論特性はモンテカルロ法で評価してきた。Lattice データの場合、ホットスポット領域のパターンが限られ、統計量の理論特性の評価を直接計算することができる。そこで、スキャンする領域の総数を計算するために、セルパターンの総数を効率的に計算するアルゴリズムの開発を行い、統計量の確率分布を計算している。また、ホットスポット検出の対象データを、(1) 2次元空間から時系列多次元空間、(2) 一変量から多変量、へと拡張を行っている。(1)では、時系列多次元空間における新たな隣接情報を定義し、汚染地下水の漏洩モデルにおける時系列 3次元空間データに適用し、時系列多次元空間でのホットスポット検出及び p 値の計算をしている。(2)では、主成分分析により次元を縮小するとともにバリマックス回転により因子の解釈しやすい軸を求める方式を提唱し、2001年から2003年の上位5位までの死亡原因別データに適用し、時系列的に地域が変化するホットスポット検出を行っている。また、主成分空間においてボロノイ分割に基づき隣接情報を定義し、死亡総合指標を用いて、疾病に関する主成分空間におけるホットスポットの検出を行い、地域と疾病との関係を統計的に説明している。さらに、lattice データに対して、エシェロン解析で得られた階層構造に基づきホットスポット検出を行う、一連のホットスポット検出に関するソフトウェアを世界で初めて構築している。

これらの研究成果は、4編の査読付き論文、9編の国際会議論文、20の口頭発表などにおいて公表しており、日本計算機統計学会第19回大会で最優秀学生発表賞、5th IASC Asian Conferenceにおいて Wakimoto Memorial Fund を受賞している。これらの研究は、空間データにおけるホットスポット検出の理論と応用への貢献は大きい。以上により、本論文は博士(理学)に値すると判断した。