

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 14 日現在

機関番号：15301

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500352

研究課題名(和文) 時空間情報に対する同位相分類の多次元・多変量への拡張とその応用

研究課題名(英文) Classification for multivariate and multidimensional spatio-temporal data based on echelon hierarchical structure and its application

研究代表者

栗原 考次 (Kurihara, Koji)

岡山大学・その他の研究科・教授

研究者番号：20170087

交付決定額(研究期間全体)：(直接経費) 3,800,000円、(間接経費) 1,140,000円

研究成果の概要(和文)：本研究では、多次元・多変量時空間情報に対して同位相を持つ領域に分類するとともに、それらを階層構造で表現し、空間情報におけるピークとファウンデーションから生成されるホットスポットの候補を検出するための新手法を確立した。さらに、その階層構造を基に、生態学、森林学、疫学、経済学などの分野で直面する圏(ゾーン)やパッチの検出などの地域構造の分析のためのアルゴリズムの作成及びソフトウェア化を行った。

研究成果の概要(英文)：In this research, we explore the classification for multivariate and multidimensional spatio-temporal data based on echelon structure. We newly develop the technique to detect the candidate of hotspot based on echelon peaks and foundations. We also develop the algorithm to identify the zone and the patch based on practical definition for the relation of peaks and foundations. In addition, we demonstrate some illustrations to detect the zones and the patch for spatio-temporal data.

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：ホットスポット 時空間情報 エシェロン解析

1. 研究開始当初の背景

各種の分野で得られる種々の情報に対して、内存するメカニズムを的確に表現するモデルの構築やその理論的性質を調べることが重要である。特に、多くの分野のデータは、時間・空間的位置情報をともなって得られることが多く、それらのメカニズムを解明していく必要がある。2次元空間データの位相的構造を見つける最近の手法として、エシェロン解析法(1997)がある。栗原らは、その応用として、リモートセンシングデータに基づく植物の生態分布と人口密度の関連(Kurihara, 2000)、空間スキャン統計量とエシェロン解析を利用したホットスポット検出(栗原考次, 2003)、空間データの同位相分割のアルゴリズムとその応用(Kurihara, 2004)、多変量空間データ(Kurihara et al., 2006)、3次元時間データのホットスポット検出(Ishioka et al., 2007,2010)、ゲノムデータにおけるLDブロックの同定(Tomita et al., 2007)、医学データのホットスポットの検出(Kurihara et al., 2008)などの研究を行ってきた。

2. 研究の目的

本研究では、空間情報が時系列的に相互作用を受けるマルコフ確率場において、同位相を持つ領域(エシェロン)に分類するとともにそれらを階層構造で表現し、空間情報の集積性や複雑性に関する新手法を確立することを目的としている。さらに、その階層構造を基に、生態学、森林学、疫学、経済学などの分野で直面するホットスポットやパッチの検出などの地域構造の分析のためのアルゴリズムの作成などを行い、その応用分野の発展に寄与することを意図している。具体的には、以下のような問題を取り扱う。

- (1) 医学・環境データに対する確率場を想定した時空間同位相領域分類
- (2) 多変量空間データの分類とホットスポット検出
- (3) 多次元空間データの分類とホットスポット検出
- (4) 森林の空間的な生態構造分析

3. 研究の方法

空間的・時間的構造をもつデータのメカニズムやホットスポット(集積性)を解明するために、研究の目的で記述した4つの問題に以下のような研究の方法で取り組む。

- (1) 医学・疫学データに対する確率場を想定した時空間同位相領域分類

本研究では適用分野を医学・疫学に絞り、小地域における疾病率や死亡率のマルコフ確率場における構造や同位相領域分類について考える。こうしたデータについては、時間及び空間上での時間的・空間的構造に基づき近傍情報を定義し、時空間上の立体全体の階層構造及びホットスポットの検出を行う必要がある。母標準疾病率や母死亡率を想定し、新たに抽出された標本に基づく手法及び母標準化疾病率になめらかな連続事前分布を仮定するベイジアン接近法に基づく手法を新たに開発する。

- (2) 多変量空間データの分類とホットスポット検出

多変量空間データの分類や候補地(ホットスポット)の選定に対して、次の2つの接近法を考えている。最初の方法は、空間的な位置情報は保持しながら、多変量データの次元の縮約を行い、分類を行う方法である。すなわち、空間データと

して与えられる多変量データを変量毎の合計や主成分分析により次元を縮約した1変量データに基づき空間データの分類の方法である。2番目の方法は、多変量空間における類似度を基に新たな空間構造を考え、分類を行う方法である。すなわち、多変量データの主成分分析を行い、寄与率の高い次元の主成分空間に各格子を布置し、そのポロノイ分割に基づき近傍情報を定義し、総合指標化されたデータに対して分類を行う方法である。

(3) 多次元空間データの分類とホットスポット検出

2次元空間規則格子データでは4、8近傍系、非規則格子データでは隣接や距離に基づき近傍を定義し、同位相分類及びホットスポット検出を行う。3次元空間では、規則格子データのみを取り扱い、6、26近傍系における同位相分類を行うとともに、具体的な3次元データに対してホットスポット検出を行う。4次元空間については、3次元空間に加え時間軸を4次元目の空間として考え、隣接する時間の同位置を近傍として定義し、4次元時空間データの同位相分類及びホットスポットを求める。

(4) 森林の空間的な生態構造分析

森林における生態構造分析では、パッチが重要な意味をもつ。本研究では、森林のパッチを検出するために、樹木の大きさにより森林を層別し、空間的な階層構造を求め、同位相をもつ近隣樹木群に基づきパッチを検出し、森林の生態構造分析を行う方法について研究する。

4. 研究成果

(1) 医学・疫学データに対する確率場を想定した時空間同位相領域分類

日本の4大死亡原因による都道府県データに対して、年齢構成を調節した指標SMRを用いたEchelon解析に基づく階層構造により、疾病毎の時系列的に地域が変化するホットスポットを検出した。また、寄与率の高い低次元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た。

疾患に関連するSNPの検出には、遺伝子頻度が有意に高い頻度で出現する連鎖不平衡に基づく方法が有効である。そこで、SNPのエシェロン階層構造に基づき関連ブロック遺伝子を検出し、ブロックの代表SNPを選択する方法について研究した。さらに、これらの関連解析を行うために、モデルを選択する方法について研究を行った。

日本原子力研究開発機構(JAEA)が公開している福島第一原子力発電所から80km圏内の空間線量率データに対し、空間集積性の観点から、空間線量率が有意に高い及び低い領域の検出を行った。領域をスキャンする方式として、circular scan法及び同位相分類に基づくechelon scan法を利用するとともに、echelon scan法の方がより尤度が高い領域が検出できることを示した。

(2) 多変量空間データの分類とホットスポット検出

本研究では、(1)空間データとして与えられる多変量データを変量毎の合計や主成分分析により次元を縮約した1変量データに基づく分類、(2)多変量データの主成分分析を行い、寄与率の高い次元の主成分空間に各格子を布置し、そのポロノイ分割に基づき近傍情報を定義し、総合指標化されたデータに対する分類、について研究を行った。

(3) 多次元空間データの分類とホットスポット検出

空間データの同位相分類においては、空間データの値を近傍データと比較することが必要である。規則格子データでは1次元では2近傍系、2次元では4、8近傍系、非規則格子データでは隣接や距離に基づき近傍を定義し、同位相分類及びホットスポット検出を行った。3次元空間では、規則格子データのみを取り扱い、6、26近傍系における同位相分類を行った。4次元空間については、3次元空間に加え時間軸を4次元目の空間と考えた8近傍系での分類を環境汚染物質データを基に同位相分類及びホットスポット検出を行った。さらに、4次元目の空間を同等に考える8、80近傍系を定義し、人工データでの同位相分類及びホットスポット検出を行った。

(4) 森林の空間的な生態構造分析

森林の樹木の動態調査や定量的評価を行うため、森林において生態学的に同じ共通属性を持つパッチに基づく構造分析について議論した。すなわち、森林を構成する全樹種の生態や森林の更新を反映したパッチを同定するために、エシェロン同位相分割領域の階層構造に基づき生態構造分析を行う手法について研究した。また、具体的に試験地データに適用し、提案した手法の優位性を示した。

これらの研究成果は、韓国中央大学における国際会議での招待講演、日伊国際分類学会での招待講演、国際分類学会(IFCS2013)での招待講演、ISI2011、IASC2011、ヨーロッパ地区国際計算機統計学会(COMPSTAT2012)、アジア地区国際計算機統計学会(ARS2013)等などの国際会議、統計関連学会連合などの

国内学会で講演するとともに、国内外の査読論文誌において公表した。

5. 主な発表論文等

〔雑誌論文〕(計10件)

- Kim, S., Hayashi, K. and Kurihara, K., Parameter Estimation for Semivariogram of Spatial Data with Outliers, *Proceedings of the 2013 IASC Satellite for the ISI WSC and 8th IASC-ARS Conference*, (Edited by Sang-Hoon Cho), 査読有, 2013, 415-421.
- Na, M., Tamura, T., Hayashi, K. and Kurihara, K., A Discriminant Classifier Based on Tuning Parameters Controlled by Cross-Validation Method, *Proceedings of the 2013 IASC Satellite for the ISI WSC and 8th IASC-ARS Conference*, (Edited by Sang-Hoon Cho), 査読有, 2013, 141-144.
- 石岡文生, 栗原考次, Echelon 解析に基づくスキャ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, 2012, 60(1), 93-108.
- 小田牧子, 石岡文生, 正木隆, 栗原考次, エシェロン階層構造を利用した森林の分類, *データ分析の理論と応用*, 査読有, 2012, 2(1), 17-31.
- Ishioka, F. and Kurihara, K., Detection of spatial clustering using echelon scanning method, *Proceedings of COMPSTAT2012*, (Edited by A. Colubi, K. Fokianos and E. J. Kontoghiorghes) Physica-Verlag, 査読有, 2012, 341-352.
- Hayashi, K., Ishioka, F., Suito, H. and Kurihara, K., Diagnostics and improvements on statistical sensitivity analysis in a subspace method, *Proceedings of COMPSTAT2012*, (Edited by A. Colubi, K. Fokianos and E. J. Kontoghiorghes)

Physica-Verlag, 査読有, 2012, 303-316.

羅明振, 荒木勝, 栗原考次, 大学生の社会貢献・地域連携に対する意識, 岡山大学環境理工学部研究報告, 査読無, 2012, 17(1), 7-21.

Nakatamari, H., Ueda, T., Ishioka F., Raman, B., Kurihara, K., Rubin, G., Ito, H. and Sze, D., Discriminant Analysis of Native Thoracic Aortic Curvature: Risk Prediction for Endoleak Formation After Thoracic Endovascular Aortic Repair, *Journal of Vascular and Interventional Radiology*, 査読有, 2011, 22(7), 974-979.
doi:10.1016/j.jvir.2011.02.031.

Tomita, M., Kurihara, K. and Moon, S., An Application to Select Tag Loci by Using Hierarchical Structures of DNA Markers, *Journal of the Korean Data Analysis Society*, 査読有, 2011, 13(6), 2749-2762.

石橋雄一, 原敦子, 岡安勲, 栗原考次, 画像特徴量とテキスト情報による画像検索可能な病理情報データベースの構築, *計算機統計学*, 査読有, 2011, 24(1), 3-21.

〔学会発表〕(計26件)

石岡文生, 栗原考次, 福島第一原子力発電所事故に伴う放射線物質の集積性の検討, 日本計算機統計学会第27回シンポジウム, 2013.11.15, 熊本市市民会館.

林邦好, 石岡文生, 井上幸平, 中玉利葉月, 植田琢也, 水藤寛, 栗原考次, 樹木モデルを用いた肝臓癌の診断論理に関する統計解析, 日本計算機統計学会第27回シンポジウム, 2013.11.15, 熊本市市民会館.

金秀柱, 栗原考次, 崔承培, 空間データのバリオグラム推定における信頼性解析, 日本計算機統計学会第27回シンポジウム, 2013.11.15, 熊本市市民会館.

石岡文生, 林邦好, 植田琢也, 水藤寛,

栗原考次, 胸部大動脈瘤ステントグラフト治療における統計的アプローチ, 2013年度統計関連学会連合大会, 2013.9.9, 大阪大学.

Kurihara, K. and Ishioka, F., Spatial Clustering based on Hierarchical Structure of Multidimensional Lattice Data, Conference of the International Federation of Classification Societies, 2013.7.15. (Invited paper), Tilburg, Netherland.

林邦好, 石岡文生, 植田琢也, 水藤寛, 栗原考次, 胸部大動脈瘤ステントグラフト治療に関わる予後予測 - sigmoid関数により血管形状曲率を重み付けした場合のエンドリークの統計的予測 -, 日本計算機統計学会第27回大会, 2013.5.17, 弘前大学.

小田牧子, 石岡文生, 栗原考次, 階層的空間構造を利用した森林パッチの境界検討, 日本計算機統計学会第26回シンポジウム, 2012.11.2, 東京大学.

林邦好, 石岡文生, 植田琢也, 水藤寛, 栗原考次, 胸部大動脈瘤ステントグラフト治療に関わる予後予測, 日本計算機統計学会第26回シンポジウム, 2012.11.2, 東京大学.

栗原考次, 時空間データの同位相分類と多次元・多変量への拡張, 研究会「時系列・時空間統計解析の新たなる展開」, 2012.9.25, 東京大学.

石岡文生, 栗原考次, 空間データの分類とその応用, 2012年度統計関連学会連合大会, 2012.9.11, 北海道大学.

Tomita, M. and Kurihara, K., Association Study and Comprehended Hierarchical Structures on DNA markers, Joint meeting of Japanese and Italian Classification Societies, 2012.9.4, Anacapri(Italy), (Invited paper).

Ishioka, F. and Kurihara, K., Spatial Classification and Clustering Approaches based on Echelon Analysis, Joint meeting of Japanese and Italian Classification Societies, 2012.9.4, Anacapri(Italy), (Invited paper).

Kurihara, K., Spatial Clustering Based on Hierarchical Structure of Lattice Data and Its Applications, The 2nd Institute of Mathematical Statistics Asia Pacific Rim Meeting, 2012.7.2, Tsukuba(Japan).

林邦好、石岡文生、植田琢也、水藤寛、栗原考次、胸部大動脈瘤ステントグラフト治療に関わる予後予測、日本計算機統計学会第26回大会、2012.5.12、高松市。

Al-Sultan, Ahmed Thanyan, Ishioka, F. and Kurihara, K., Aircrafts Scheduling Algorithms at Airport Terminal, 日本計算機統計学会第26回大会、2012.5.12、高松市。

Na, M. and Kurihara, K., The Site Selection of Landfill Using Fuzzy Set and AHP Theory, Joint Meeting of the Korea-Japan Conference of Computational Statistics and the 25th Symposium of Japanese Society of Computational Statistics, 2011.11.11, Busan(Korea).

Al-Sultan, Ahmed Thanyan, Ishioka, F. and Kurihara, K., A Comparative Study of Aircrafts Scheduling Algorithms Base on a Simulation Approach, Joint Meeting of the Korea-Japan Conference of Computational Statistics and the 25th Symposium of Japanese Society of Computational Statistics, 2011.11.11, Busan(Korea).

羅明振、栗原考次、空間データを利用した最終処分場の立地選定、2011年度統計関連学会連合大会、2011.9.7、九州大学。

石橋雄一、原敦子、岡安勲、栗原考次、病事情報データベースにおける画像とテ

キストによる分類の方法、2011年度統計関連学会連合大会、2011.9.6、九州大学。

石岡文生、植田琢也、水藤寛、栗原考次、大動脈形状の分類と予後予測、2011年度統計関連学会連合大会、2011.9.6、九州大学。

- 21 小田牧子、正木隆、栗原考次、樹木群集のパッチ検出法、2011年度統計関連学会連合大会、2011.9.6、九州大学。
- 22 林邦好、水藤寛、栗原考次、エントロピー基準による次元縮小に基づくパターン認識手法に対する統計的感度分析法、2011年度統計関連学会連合大会、2011.9.5、九州大学。
- 23 Al-Sultan, Ahmed Thanyan , Ishioka, F. and Kurihara, K., Aircrafts Scheduling algorithms: A simulation based approach, 2011年度統計関連学会連合大会、2011.9.5、九州大学。
- 24 Kurihara, K., Spatial Clustering for Spatial-temporal Lattice Data and its Applications in Modern Industry, 2011 international conference: Applications of Statistical Methods in the Modern Industry, 2011.6.4, Chung-Ang University(Seoul), (Invited paper).
- 25 羅明振、荒木勝、栗原考次、大学生の社会貢献・地域連携に対する意識、日本計算機統計学会第25回大会、2011.5.8、函館市。
- 26 富田誠、栗原考次、文勝浩、DNAデータにおけるLD集積性と代表SNPの選択、日本計算機統計学会第25回大会、2011.5.7、函館市。

6 . 研究組織

(1)研究代表者

栗原 考次 (KURIHARA KOJI)
 岡山大学・大学院環境生命科学研究所・教授
 研究者番号：20170087