

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 27 日現在

機関番号：10101

研究種目：挑戦的萌芽研究

研究期間：2012～2014

課題番号：24650144

研究課題名(和文) 3次元動体データに対する関数データ解析の方法の開発

研究課題名(英文) Functional data analysis method for 3 dimensional movings

研究代表者

小宮 由里子 (Komiya, Yuriko)

北海道大学・情報基盤センター・助教

研究者番号：40241393

交付決定額(研究期間全体)：(直接経費) 2,000,000円

研究成果の概要(和文)：関数データ解析およびシンボリックデータ解析では、従来の多変量データ解析では扱いが難しいデータを解析対象にすることができる。関数データ解析では、データを関数として捉えることで例えば微分を用いる解析を可能とする。シンボリックデータ解析はデータの内部変動や構造を取り入れることでデータを集約する。大量で変動を含むことを特徴とするビッグデータに対し、関数データ解析およびシンボリックデータ解析の枠組みで手法を開発した。環境空間線量率測定データ、ネットワーク遅延データ、交通データ等実際のデータに適用し、手法の有用性を示した。

研究成果の概要(英文)：With "Functional Data Analysis(FDA)", many data sets have been studied, that with conventional statistical approaches could not answer to. "Symbolic Data Analysis(SDA)" also gives a wider viewpoint for data studies. In SDA, symbolic data have internal variation and structure which must be taken account for when analysing target datasets. We propose methods of FDA and SDA for "Big Data" that has features of big size and variation within it. Our methods are usefully applied to environmental radioactivity data, network delay data and traffic data.

研究分野：計算機統計学

キーワード：統計学

1. 研究開始当初の背景

研究開始当初は、動体追跡技術に基づく3次元空間放射線治療データについて、関数データ解析による、実際の解析手法を開発することを目的としていた。

2. 研究の目的

関数データ解析では、従来の多変量データ解析の枠組みを超えた解析を行うことが期待できる。さらに、従来の多変量データ解析では表現できない、多様なデータを柔軟に扱うことができるシンボリックデータ解析の枠組みへ、研究を進めることにした。近年注目を集めるビッグデータにおける多様性を扱うことが可能になると考えた。

3. 研究の方法

関数データ解析とシンボリックデータ解析の分野で手法を新たに提案した。

(1) 関数データ解析

Ramsay を中心に提唱された関数データ解析法は、1990年代以降、理論および応用の両面から活発な研究がなされている。従来の多変量データ解析では、データを多次元数ベクトルとして表現しているが、関数データ解析では、データを関数とみなして解析を行うものであり、応用の範囲は多くの分野に渡る。データを関数として扱う利点に、微分が利用できることがあげられる。

近年、情報機器の発達によりセンシングデータのような時間に依存するデータは、観測の頻度を上げ、それらデータを大量に保存できるようになっている。従来の多変量解析では、1地点についてその複数時点における離散的な観測値の集まりを観測値ベクトルとして1個の多変量データとして捉える。一方、関数データ解析では、1地点における大量の観測値を時間に依存する1つの関数として捉えることにより、観測値の変化、異常検出などについて新しい知見が得られることが期待できる。

(2) シンボリックデータ解析

Didayらによって提唱され1990年代以降盛んに研究されているシンボリックデータ解析では、データの捉え方を、従来とは異なり、データは「内部変動」や「構造」を持つ、より複雑なものとしている。従来の個々のデータを「インディヴィデュアル」と呼ぶ。ある特徴に基づくインディヴィデュアルの集まり「オブジェクト」をデータの単位とする。たとえば、従来の多変量データ解析では、データは多次元数ベクトルとして表現される。ベクトルの各成分はスカラー値である。シンボリック

クデータ解析の立場では、たとえばベクトルの各成分が変動を含むと考え、オブジェクトとして、各成分が区間値、分布値などをもつと考える。シンボリックデータ解析では、データの個々の値を、集約して捉える。利点の一例として、膨大なデータの蓄積に対し、ある基準を導入してそれらを集約したデータ表現を取ることによってデータサイズを減らすことができることが挙げられる。

4. 研究成果

関数データ解析の枠組みで新たな手法を提案し、福島県における空間線量率測定データに適用し、有用な結果を得た。また、近年の情報技術革新により、膨大な量で頻繁に更新されつつ蓄積されるビッグデータに、最近、注目が集まっている。このようなデータの多様性を扱うため、シンボリックデータ解析によるさまざまな解析手法を提案した。これら手法を具体的なデータに適用し有用な結果を得た。これらについて国内外の学術雑誌に掲載され、また国内外の学会で発表を行った。今後、関数データ解析およびシンボリックデータ解析の枠組みで種々の手法の開発が期待できる。

- (1) 関数クラスタリングにおいて、定義域を逐次的に移動させ解析する、移動関数 k -means 法を提案した。従来の関数クラスタリングでは定義域は対象の関数全体に固定して扱われている。移動関数 k -means 法は、部分的な区間ごとの関数データに着目して分析することで、従来クラスタリング手法では検出されにくいクラスターを検出することができた。福島県の空間線量率データを測定した大規模センシングデータに対し適用し、積雪のある冬期間にクラスターの構成が変化する様子を捉えることを示した。
- (2) 福島県の複数地点に設置されている、空間線量率センサーによるセンシングデータを同時に扱うため、値域を多次元とする関数データを導入した。すなわち、ある時刻に対して、複数箇所のセンシングデータを多次元値として対応させ、多次元値関数データとする。ここで多次元値関数データにおける異常領域の検出を問題にする。一般にセンシングデータにおいて、測定環境の変化や測定の誤差を見つけるため、データの「外れ値」に注意することは重要である。ここで、多次元値関数データにおいて、値のどの次元においても「外れ値」となる「同時異常領域」を見つける手法を提案した。提案手法を、空間線量率センシングデータ

に適用し、同時異常領域となる積雪期間を検出することを示した。

- (3) シンボリックデータ解析の枠組みで、分布値データに注目した。分布値非類似度を導入し、新たに階層的クラスタリング法を提案した。この手法を複数サイト間における通信時間を測定したデータの解析に適用し、特徴的クラスターを同定し、その解釈から手法の有用性を示した。
- (4) シンボリックデータ解析の枠組みで、異なる尺度が混在している尺度混在型データを扱う手法を提案した。異なる尺度により得られた観測値の間の関連性を調べるため、尺度ごとに解析対象間に非類似度を定義し、これを用いて、多次元尺度構成法によって解析対象をユークリッド空間上に布置した。解析対象の布置座標に対し主成分分析法を適用し、その関連性を分析する。福島県各地域で複数の方法によって得られた空間線量率データに対し、それぞれの測定方法の特性を考慮し、分布値と関数値によるデータ表現として再構成し、手法を適用した。得られた結果の解釈から手法の有用性が示された。
- (5) シンボリックデータ解析の枠組みで、分布値データに対する非階層的クラスタリングを提案した。分布値データとして一次元正規混合分布を用いる。分析対象をとするオブジェクトを道路を単位とし、1道路における車両の走行速度ごとの発生確率を分布値データとする。約17万本の道路における車両走行データに適用し、道路の特性に基づくクラスタリングを行い、分類が実現できることを示した。

5. 主な発表論文等

〔雑誌論文〕(計5件)

松井 佑介、小宮由里子、南 弘征、水田正弘、移動関数 k -means 法の提案とその応用、データ分析の理論と応用、査読有、Vol.4, No.1, 2015, pp.43-55

松井 佑介、小宮由里子、南 弘征、水田正弘、多次元値関数データにおける異常領域の検出法とセンシングデータへの応用、計算機統計学、査読有、Vol.27, No.2, 2014, pp.65-77

Matsui, Y., Minami, H. and Mizuta, M., Symbolic Cluster Analysis for Distribution Valued Dissimilarity, Communications for Statistical Applications and Methods, 査読有、Vol. 21, No.3, 2014, pp.225-234
DOI:10.5351

米森 力、南 弘征、水田正弘、分布値データに対する非階層的シンボリッククラスタリングと交通ビッグデータへの適用について、日本統計学会誌、査読有、Vol.43, 2013, pp.59-67

藤田将成、伊藤浩二、小林 稔、南 弘征、水田正弘、GPS 移動履歴とスポット情報閲覧履歴の統合解析に基づいたスポット推薦手法の提案、データ分析の理論と応用、査読有、Vol.2, 2012, pp.53-68

〔学会発表〕(計12件)

Mizuta, M. and Matsui, Y., Analysis of sensing data with moving functional methods, The 21st International Conference on Computational Statistics, 2014年8月19日~2014年8月22日、Geneva (Switzerland)

Matsui, Y. and Mizuta, M., SDA for mixed-type data and its application to analysis environmental radio activity data, The 21st International Conference on Computational Statistics, 2014年8月19日~2014年8月22日、Geneva (Switzerland)

松井佑介、南 弘征、水田正弘、放射性物質の分布状況等調査データベースを用いた空間線量率分布に関する考察、2013年度統計関連学会連合大会、2013年9月8日~2013年9月11日、大阪大学(大阪府・豊中市)

Matsui, Y. Minami, H. and Mizuta, M., Symbolic Cluster Analysis for Distribution Valued Data, Joint Meeting of the IASC Satellite Conference and the 8th Conference of Asian Regional Section of the IASC, 2013年8月22日~2013年8月23日、Seoul (Korea)

Matsui, Y., Minami, H. and Mizuta, M., Hierarchical Symbolic Cluster Analysis with Quantile Function Representation, Conference of the International Federation of Classification Societies IFCS-2013, 2013年7月14日~2013年7月17日、Tilburg (the Netherlands)

Matsui, Y., Komiya, Y., Minami, H. and Mizuta, M., Cluster analysis of distribution valued data and its application, Hokkaido Univ.-Korea Univ. The second Joint Workshop in Statistics, 2013年6月25日~2013年6月26日、Seoul (Korea)

Igarashi, K., Komiya, Y., Minami, H. and Mizuta, M., Visualization of Fukushima radioactive substances distribution data, Hokkaido Univ.-Korea Univ. The second Joint Workshop in Statistics, 2013年6月25日~2013年6月26日、Seoul (Korea)

Seishin Takao, Masahiro Mizuta,
Hiroyuki Date, Yuriko Komiya,
Kenneth L. Sutherland, Rikiya
Onimaru and Hiroyuki Shirato, A
Mathematical Method for
Estimating Optimal Fractionation
Regimen in Radiotherapy, Third
International Conference on
Real-time Tumor-tracking Radiation
Therapy with 4D Molecular Imaging
Technique, 2013年2月7日~2013年
2月8日、札幌コンベンションセンタ
ー(北海道・札幌市)

水田正弘、高尾聖心、伊達広行、白土
博樹、腫瘍の放射線治療における数理
モデルについて、2012年度統計関連学
会連合大会、2012年9月9日~2012
年9月12日、北海道大学(北海道・
札幌市)

松井佑介、南弘征、水田正弘、SDA
によるネットワークトラフィックの解
析、2012年度統計関連学会連合大会、
2012年9月9日~2012年9月12日、
北海道大学(北海道・札幌市)

鈴木和之、小宮由里子、水田正弘、サ
ポートベクターマシンにおける直交変
換に対する解の不変性について、2012
年度統計関連学会連合大会、2012年9
月9日~2012年9月12日、北海道大
学(北海道・札幌市)

妹尾いづみ、松井佑介、小宮由里子、
水田正弘、ライフログを用いた個人健
康管理における異常検知手法の適用、
2012年度統計関連学会連合大会、2012
年9月9日~2012年9月12日、北海
道大学(北海道・札幌市)

6. 研究組織

(1) 研究代表者

小宮 由里子 (KOMIYA, Yuriko)

北海道大学・情報基盤センター・助教

研究者番号：40241393

(2) 研究分担者

水田 正弘 (MIZUTA, Masahiro)

北海道大学・情報基盤センター・教授

研究者番号：70174026