

研究種目：若手研究（B）

研究期間：2007～2010

課題番号：19700258

研究課題名（和文） 勾配写像を利用した非ガウス型時空間モデルの設計法と評価法

研究課題名（英文） Construction and evaluation of non-Gaussian spatio-temporal models through the use of gradient maps

研究代表者

清 智也（SEI TOMONARI）

東京大学・大学院情報理工学系研究科・助教

研究者番号：20401242

研究成果の概要（和文）：

時間と空間の関数として与えられるデータに対し、その背後に潜む複雑な相関関係を適切に抽出するための統計モデルを提案し、その有効性について検討した。方法として、勾配写像と呼ばれる特殊な変数変換を利用した。特に、機械学習の分野で盛んに研究されているスパースな構造推定法を適用するとともに、その対象範囲を方向データにも拡張した。提案したモデルの理論的な利点を、推定の損失という側面から明らかにした。

研究成果の概要（英文）：

For data in the form of function of time and space, we proposed a statistical model to extract complex structures behind the data. We used the gradient map as a special transformation. In particular, we apply a sparse estimation method, which is an analogy of recently developed methodology in machine learning theory. Furthermore, we generalize our models to those of directional data. We also clarify a theoretical merit of our models from the viewpoint of loss of estimation.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	800,000	0	800,000
2008年度	800,000	240,000	1,040,000
2009年度	800,000	240,000	1,040,000
2010年度	800,000	240,000	1,040,000
年度			
総計	3,200,000	720,000	3,920,000

研究分野：統計科学

科研費の分科・細目：情報学・統計科学

キーワード：勾配写像，凸最適化問題，輸送問題，方向統計学。

1. 研究開始当初の背景

時空間統計解析の中心的役割を果たす統計的モデルはガウスモデルであり、多くの好ましい性質を持っている。しかし、ガウス性の

仮定は強く、現実的には使えないような場面もしばしば存在する。特に、3つ以上の地点（時点）における交互作用を詳しく記述したいとき、これを多変量ガウスモデルで表現することには限界がある。ガウス性の仮定を除

くために有効な考え方として、変数変換がある。ここでいう変数変換とは例えば、「正の値しかとらないデータに対して対数関数を施し、変換後のデータがガウス分布に従うと仮定する」といった類のものであり、実際によく用いられている。しかし、Box-Cox 変換を含む既存の変換方法は、各地点における変量を別々に変換するというものがほとんどであり、依然として多次元交互作用をうまく表現しているとは言いがたい。そこで、多数の地点におけるデータを同時に変数変換することが思いつく。ところが、このような多変数の変換を場当たりに構成しようとしても、多変数であるが故に直観が働かず、具体的な構成法は考えられていないのが現状であった。

これに対し本研究では、変数変換として多変数凸関数の勾配写像を用いる、という着想を得た。ここで、勾配写像を用いる理由のひとつは、「任意の連続確率分布に対し、ある勾配写像が一意的に存在して、ガウス分布に変換することができる」という数学的事実である。この意味で、勾配写像は対数関数のような単調増加関数の自然な一般化と見ることもできる。ただし、勾配写像に基づく統計的モデリングの大まかな枠組みは構築できても、時空間データの解析という観点からは、多くの課題が残されていた。

2. 研究の目的

項目1で述べたように、勾配写像に基づく統計的モデリングのアイデアは大まかに構築されていたが、その特性や実用性については未知の点が多かった。これらの点を解明することが本研究の当初の目的であった。具体的には、階層的ベイズモデルとの比較、モデル選択規準の整備、数値計算の容易性・安定性、関連する理論の発展、の4つが課題であり、これらの課題を一つずつ解決することが目的であった。

3. 研究の方法

本研究で用いる勾配写像は、最適輸送理論と呼ばれる確率論・幾何学・偏微分方程式論の分野と密接に関連しているため、これらの調査から研究を始めた。それにより、対象となるデータの特性（例えばデータの範囲）に応じて、使うべき勾配写像を定めていった。研究成果の欄で述べる、フーリエ級数を用いたモデルでは、フーリエ級数の稠密性を利用して、有限フーリエ級数に基づく勾配写像モデルを導いた。さらに球面上のデータへも適用できることに気づき、方向統計学の調査を進めた。同時に多様体上の最適輸送理論における最新の結果も必要になり、関係する論文の

検討を行なった。

4. 研究成果

年度ごとの研究成果を述べる。以下、勾配写像に基づいた統計モデルのことを単に「勾配モデル」と呼ぶ。

(1) 2007 年度

勾配モデルをグラフィカルモデリングに役立てる方法を研究した。グラフィカルモデリングとは多変量の統計データの交互作用を、視覚的かつ統計的に表現するための一連の方法論である。従来、連続変量データに対して適用されるグラフィカルモデルはガウス分布に基づくものが主だったが、本研究で提案したモデルにより、非ガウス型のグラフィカルモデルを近似的に実現できた。具体的には、高次元の交互作用を検出するために勾配写像へ高次項を追加した。また、実際のデータに対してグラフィカルモデルを選択するための、パラメータ推定の手順およびモデル選択の手順を与えた。ここで、パラメータ推定において許容領域の判定条件について考察する必要性が生じたが、モデルを限定することにより、一定の十分条件を得ることができた。

(2) 2008 年度

立方体上の多変量データに対する構造的なモデルを提案した。ここでは、立方体上では関数がフーリエ級数を用いて表されるという一般的な事実を利用した。フーリエ級数を用いたことにより、これまでは本質的に周波数が0か1の構造しか考えていなかったものが、高次の周波数まで扱えることになった。例えば、周波数ベクトルが2と1を含むものを考えると、回帰分析における不等分散性に類似した関係性を記述できることが明らかになった。さらに、不等分散性を持つ非対称な関係を利用して、従来のガウスモデルでは到底検出できないようなデータの依存関係を検出できることが分かった。以上の方法を用いる上でネックとなる点は、パラメータ空間の許容領域が陽に表されない点である。この問題に対しては、許容領域を内側から近似する集合列を陽に得たほか、許容領域より狭い集合ながらも計算が容易である集合を得た。後者については、副産物として、解がスパース性を持つことを数値的に確認できた。このスパース性を利用して、近年注目を浴びているラッソ型の推定量を構成した。その結果、実データに対して、ガウス型モデルよりも良い予測性能を持つモデルを選択できるケースがあることを指摘した。

(3) 2009 年度

球面上に値をとるデータに対するモデルを提案した。球面上のデータは、方向統計学と呼ばれる分野で扱われており、扱いやすいモデルの構築が急がれている。特に、既存のモデルであるフィッシャー・ビンガムのモデルは、柔軟であるものの、密度関数の正規化定数が陽に計算できないという問題がある。本研究で提案したモデルは、最適輸送理論で発展している多様体上の輸送写像を利用したものであり、尤度が陽に書け、最尤法が凸最適化問題になることが示された。これにより、フィッシャー・ビンガムモデルと同等な柔軟性を持ち、かつ扱いやすいモデルが構築された。数学的には、最適輸送理論の最新の成果を用いており、関連分野にもインパクトを与えている。以上の結果、実データに対して、これまで扱いやすかったフィッシャーモデルよりも良い予測性能を持つモデルを選択できるケースがあることを指摘した。

(4) 2010 年度

勾配モデルの中でも様々な交互作用を入れた構造的勾配モデル (Structural Gradient Model, 以下 SGM と略記する) に対して, Efron の統計的曲率と呼ばれる量を評価した。Efron の統計的曲率とは、考えている統計モデルが指数型分布族からどの程度離れた分布族であるかを定量的に測ったものであり、推定や検定の理論においても重要な役割を持っている。モデルの特定の点 (一様分布に対応) において統計的曲率を評価した結果、同じ点で同じスコア関数ベクトルを持つ混合分布族に比べ、SGM は小さい曲率を持っていることが示された。両モデルとも、尤度が陽に書け、最尤法が凸最適化問題になるという利点を持っているが、この結果から、SGM の方が有利な面もあることが示された。次に、SGM は多変量データに対して適用可能なモデルであるが、これをさらに一般化して、時系列モデルも考えられるようにした。

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 4 件)

1. Sei, T. (2011). Efron's curvature of the structural gradient model, *Journal of the Japan Statistical Society*, accepted. 査読有.
2. Sei, T. (2011). A Jacobian inequality for

gradient maps on the sphere and its application to directional statistics, *Communications in Statistics - Theory and Methods*, accepted. 査読有.

3. Sei, T. (2010). A structural model on a hypercube represented by optimal transport, *Statistica Sinica*, accepted. 査読有.

4. Sei, T. (2009). Gradient modeling for multivariate quantitative data, *Annals of the Institute of Statistical Mathematics*, accepted. 査読有.

[学会発表] (計 9 件)

1. Sei, T., Structural gradient model for time series, *The International Symposium on Statistical Analysis of Spatio-Temporal Data*, November 6, 2010, Kamakura, Japan.

2. 清 智也, 時系列解析のための構造的勾配モデル, 統計関連学会連合大会, 2010 年 9 月 6 日, 早稲田大学.

3. Sei, T., Optimal transportation and statistics, *BIRS 5-day Workshop - Optimal transportation and applications*, April 21, 2010, Banff, Canada.

4. 清 智也, 勾配写像で表される球面分布族とその性質, 研究集会「高次元データの統計学 - 理論・方法論・関連分野への応用 -」, 2009 年 12 月 15 日, 筑波大学.

5. Sei, T., Spherical distributions defined by gradient maps, *Directional, asymptotic, differential-geometric statistics and related area*, November 6, 2009, Osaka, Japan.

6. 清 智也, 球面上の勾配写像で表される分布族, 統計関連学会連合大会, 2009 年 9 月 8 日, 同志社大学.

7. 清 智也, 凸関数のフーリエ級数表示と勾配モデル, 統計関連学会連合大会, 2008 年 9 月 10 日, 慶應義塾大学.

8. Sei, T. Gradient modeling for multivariate analysis, *The Pyrenees International Workshop on Statistics, Probability and Operations Research (SPO 2007)*, September 12, 2007, Jaca, Spain.

9. 清 智也, Brenier 写像を利用した近似的
グラフィカルモデリング, 統計関連学会連
合大会, 2007年9月8日, 神戸大学.

[その他]
ホームページ等

6. 研究組織

(1) 研究代表者

清 智也 (東京大学・大学院情報理工学系
研究科・助教)

研究者番号 : 20401242

(2) 研究分担者

()

研究者番号 :

(3) 連携研究者

()

研究者番号 :