

令和元年6月12日現在

機関番号：17102

研究種目：基盤研究(B) (一般)

研究期間：2015～2018

課題番号：15H02670

研究課題名(和文) 自然現象や社会現象から得られる時空間データの統計モデリングと現象の理解の研究

研究課題名(英文) Statistical modeling of spatio-temporal data on natural and social phenomena, and those understanding

研究代表者

西井 龍映(Nishii, Ryuei)

九州大学・マス・フォア・インダストリ研究所・教授

研究者番号：40127684

交付決定額(研究期間全体)：(直接経費) 12,100,000円

研究成果の概要(和文)：以下の研究を行った。

1) IGARSS 2015 で、コロンビア大学とリモートセンシング画像の社会応用に関する特別セッションを企画した。2) 太陽活動に関する物理量を用いて 磁気嵐指数を予測する時系列回帰モデルを推定した。3) 試験植物の遺伝子発現の時系列データに、自己回帰型回帰モデルの疎推定により遺伝子間ネットワークを推定した。4) 超高解像度土地被覆画像に対し、出力層の階層的構造と入力層の従属性を考慮したマルチラベル判別について機械学習に基づく判別器を提案した。

研究成果の学術的意義や社会的意義

研究成果の意義は次である。1) リモートセンシング画像と社会的なデータとのリンクにより新しい知見が得られた。2) 従来の物理モデルを凌駕する磁気嵐指数の予測モデルが得られた。3) 疎推定で特定されたハブ遺伝子を改変すると、植物の成長を加速することが実験的に確認できた。今後植物燃料への適応が期待される。4) 超高解像度画像に対する高精度なマルチラベル判別は、圃場をドローンで観測した画像への応用が期待される。

研究成果の概要(英文)：The following studies were conducted.

1) In IGARSS 2015, a special session about the social application of the remote sensing imagery was organized with Columbia University. 2) Disturbance storm index was accurately predicted by regression model based on physical measurements about the solar activity. 3) Gene regulatory network was established by sparse modeling of regression model on time-series of global gene expressions. 4) For super high-resolution land covering imagery, multi-label classification was proposed by machine learning by the use of hierarchical structure of the output layer and spatial dependency of the input layer.

研究分野：統計学

キーワード：スパースモデリング 太陽活動 磁気嵐指数 遺伝子間制御ネットワーク マルチラベル判別 構造的VM 自己回帰型回帰モデル

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

センサの高性能化・小型軽量化・低廉化，データ転送・保存技術の進化，ネットワークの広域化・高速化により，従来は得られなかった高次元・高頻度の巨大な時空間データが時々刻々と蓄積されてきている．このビッグデータから新しい知を抽出する科学的手法を確立することが現在の急務となっている．ここでは (1) 地球環境を観測した時空間データ，(2) 植物の遺伝子発現時系列，(3) 企業の製品開発現場や市場から得られる時系列・画像・感性評価データ、について解析し、さらに(4) 一般の時空間データの解析手法の提案により時間的空間的ダイナミクスを統計モデルに取り込み、モデルを通して現象の特徴を把握し、現象を解明することは統計学における重要な研究課題である．

2. 研究の目的

前項の背景で述べた課題について、次を研究目的とする。

- (1) 地球環境の時空間データに対する統計モデルの構築・推定と評価方法を開発し、その時間的ダイナミクスを統計モデルにより検証する．また太陽活動が地球に及ぼす影響、たとえば磁気嵐や地震の予測モデルの構築や影響の程度を評価する．
- (2) 植物遺伝子発現量に関する短い超高次元の時系列データに基づいて、遺伝子間ネットワークの構造推定、ターゲット遺伝子を他で予測するモデルを研究する．
- (3) 企業の製品開発現場から得られる時空間データ (多次元高頻度データ、画像・感性データを含む) に対する統計モデルを開発し、その統計学的性質を研究する．
- (4) 時空間データのモデリングに関して、手法の提案や各分担者固有の分野・手法での統計研究を行う．

3. 研究の方法

研究手法：(1) - (3) は説明変数が与えられたとき目的変数を予測するという回帰モデルの枠組みで議論できる．線形モデルが中心ではあるが、非線形性が強いときは非線形回帰モデル、またカーネル関数に基づくカーネル回帰、あるいは機械学習のモデルであるサポートベクター回帰 (SVR) を用いた．また変数選択のために、AIC 等の情報量規準や L1 罰則等を用いた疎推定を用いた．(4) については、今までの統計解析や機械学習を応用してきた経験を生かして新手法を提案する．

共同研究の進め方：九州大学内の代表者・分担者とは定期的な打合せで、学外の分担者とはメール、スカイプ会議、出張等で相互に密接に情報を交換した．得られた研究成果の中間報告は九州大学の統計科学セミナー等で口頭発表し、今後の研究指針について討論した．さらに国内外のシンポジウム、学会で口頭発表し、学術雑誌で発表した．

4. 研究成果

(1) コロンビア大は 2016 年に GPWv4 (高解像度の人口密度メッシュデータ) を発表した．それを見越して同データとリモートセンシングデータの統合により多方面での知の発見を探索するため、ミラノ (イタリア) で開催された国際会議 IEEE IGARSS 2015 において A. Sherbinin (コロンビア大) と研究代表者が共同で Invited Session を組織し、本科研費で外国人発表者をミラノへ派遣した．

メッシュデータは、行政区域のように不規則な領域でのデータと異なり、人工衛星データとの統合や空間依存性を評価することが容易であることなど様々なメリットがある．ここではメッシュデータについての回帰分析について、階段関数による平均構造の近似、それに基づく非線形回帰関数の提案、対数変換を含む冪変換の利用、モデル選択等の一連の流れを整理し、実データで検証を総合報告した．また今後の展望を示した．

(2) 森林減少のモデリングでは、パラメトリックな非線形関数やスプライン関数を用いた統計アプローチを用いてきた．本論文では機械学習アプローチによる手法だけを用いて、統計手法との比較を試みた．まず森林被覆率 F を 3 群に分類する判別手法として SVM を用いた．なお空間依存性を判別に組み込むため、人口密度や起伏量や近傍での平均を変数として追加した．これにより統計モデルによる誤判別率を 0.5% 程度改善した．また $(0, 1)$ 区間の F のロジット変換した目的変数について、人口密度や起伏量および近傍での平均を用いたガウスカーネルに基づく SVR (Support Vector 回帰) を用いると、スプライン関数で得られた平均構造に近いものが得られた．

(3) Dst 指数は、太陽から高速で放出される荷電粒子によって引き起こされる地表面での磁気嵐の強さの尺度である．この研究は太陽風の観測値に基づいた統計的回帰モデルにより、Dst 指数を予測することを目的とした．Dst 指数のモデルとして、誤差が t 分布に従い、自己回帰項を含む回帰モデル (ARX) を用い、太陽風の陽子密度、速度、惑星間磁場 B_z を外部変数 (説明変数) とした．また t 分布の自由度やスケールも一般化線形モデルでモデル化した．実データへの適応した結果、提案手法は物理モデルから得られるモデルの予測能力より優れ、Dst の

時間依存の特徴を記述するのに非常に有効であることがわかった。

(4) 科学研究費は日本の研究者の競争的研究資金として定着したシステムである。ここでは科研応募者の申請が採択されることと応募者の属性(分野, 学位, 職階, 異動回数等)との関係を複数の統計モデル(ポアソン分布, 負の2項分布およびそれらの一般化線形モデル, ロジスティック回帰)を用いて解析した。その結果, ゼロインフレーション分布が有効(採択されない確率が大きい)であること, 分野間の違いが大きいこと, 職階や所属大学群(論文の被引用数で国立大学を4群に分類したもの)の影響が大きいこと, 移動回数が多すぎると採択率が下がる等興味深い知見が得られた。

(5) サポートベクターマシン(SVM)は, 高精度な分類性能を持つ機械学習の手法である。本論文ではラッパー法に基づいたSVMの2つの変数選択基準を提案する。提案基準は, 各々の変数の判別に用いる目的関数への貢献度を測る。なお各変数の重要度は, 測定量に基づいて定量化される。提案手法は再帰的計算なしで変数の重要性を評価するため, 高い計算効率を持つ。提案手法をいくつかの人工データおよび実データに適用した結果, 既存の方法より優れていることが分かった。

(6) テスト植物 プラキポディウムディスタチオンの全遺伝子発現量を4時間おきに観測した時系列をえた。これに基づき昼夜の周期が明瞭な遺伝子を同定した。さらに遺伝子の共発現解析を行い, ARXモデル(外部入力つき自己回帰モデル)をグループSCAD(Smoothly Clipped Absolute Deviation)でスパース推定して, 向きの付いた安定した遺伝子間ネットワークが得られた。これにより転写因子間の制御関係が得られ, ハブ遺伝子の重要度も明らかとなった。

(7) 条件付き不均一の時系列モデル(ARCH)の空間データへの拡張(S-ARCH)を提案した。S-ARCHモデルは近傍の観測値が与えられた条件つき分散を規定する。S-ARCHモデルのパラメータは最小二乗法と擬似最大尤推定法のツーステップ手順によって推定する。この推定法により一貫性, および漸近正規性を持つ母数推定となることが示される。提案手法はシミュレーションおよび東京地域での土地価格のデータ分析によって, 経験的に得られている特徴が確認できた。

(8) ドローンで地表面を観測した超高解像度画像は, 解像度が数cmとなる。そのため各画素の判別ではなく, タイルと呼ばれる領域(例:5m四方)に, どのようなカテゴリ(車, 駐車場, 花壇等)が含まれているかを複数答えることが求められ, これをマルチラベル判別という。構造化SVMはデータの構造を反映するように構成された判別関数を定義し, L2損失に誤判別に対するペナルティを加えた目的関数最小化で判別を行う。提案手法は, さらに隣接タイル間の判別結果が大きく異ならないというルールを埋め込んだ構造化SVMを提案し, 複数の実画像で構造化SVMより優れた性能を示した。

(9) ハイスループットな遺伝子発現データに基づき, 植物の遺伝子ネットワーク(GRN)を推定する統計的手法や機械学習手法が進歩してきている。このGRNは細胞内の現象である。大きなGNRは遺伝子の機能を導き出す基本的道具であり, 機能解析においてはハブ遺伝子の検出が求められている。ここでは60以上のGNR推定に関連した論文を, 統計手法のみ, 機械学習のみ, 両者を併用した論文に大別してレビューした。またGNRは植物の成長過程に応じて変化するため, 時系列を意識してGNR推定を行う論文にも注目した。

(10) 植物の画像やビデオから成長の成長過程に関する有益な情報を抽出するため, コンピュータビジョンはキーテクノロジーである。ここでは画像の前処理, 画像分割(葉と茎の同定, 個別の葉の同定等), 成長過程に関する特徴抽出, 成長過程の判別といったプロセスが求められる。画像解析における最近の発展は深層学習等の機械学習の発展によるところが大きい。また人工衛星, 航空機, ドローンやトラクター搭載のセンサで, 複数の波長帯で観測するリモートセンシングが穀物の成長やデジタル農業に用いられている。本論文ではコンピュータビジョンを用いて植物の表現型を評価する131編の論文についてレビューした。

5. 主な発表論文等

[雑誌論文](計35件)

S. Koda, F. Melgani, A. Zeggada and R. Nishii (2018).

Spatial and structured SVM for multilabel image classification.

IEEE Transactions on Geoscience and Remote Sensing, 査読有, 56 (10), 5948 - 5960.

DOI:10.1109/TGRS.2018.2828862.

K. Mochida, S. Koda, K. Inoue and R. Nishii (2018).

Statistical and Machine Learning approaches to predict gene regulatory networks from transcriptome datasets. Frontiers in Plant Science, 査読有, 29, 1 - 7.

DOI:10.3389/fpls.2018.01770

K. Mochida, S. Koda, K. Inoue, T. Hirayama, S. Tanaka, R. Nishii, F. Melgani (2018).

Computer vision-based phenotyping for improvement of plant productivity.

GigaScience, 査読有, 8 (1), 1 - 12. DOI:10.1093/gigascience/giy153

T. Sato and Y. Matsuda (2017). Spatial autoregressive conditional

heteroskedasticity models. Journal of the Japan Statistical Society, 査読有,

47(2), 221 - 236. DOI:10.14490/jjss.47.221

- S. Koda, Y. Onda, H. Matsui, K. Takahagi, Y. Uehara-Yamaguchi, M. Shimizu, K. Inoue, T. Yoshida, T. Sakurai, H. Honda, S. Eguchi, R. Nishii and K. Mochida (2017). Diurnal transcriptome and gene network represented through sparse modeling in *Brachypodium distachyon*. *Frontiers in Plant Science*, 査読有, 8, Article 2055. DOI:10.3389/fpls.2017.02055
- S. Koda, R. Nishii and Y. Fukasawa (2016). Novel wrapper approach for variable selection in support vector machine. *Bulletin of Informatics and Cybernetics*, 48, 105 - 117.
- M. Hosotsubo and R. Nishii (2016). Relation between awarding of grants-in-aid for scientific research and characteristics of applicants in Japanese universities. *Scientometrics*, 査読有, 109 (2), 1097--1116. DOI:10.1007/s11192-016-2074-3.
- R. Nishii and S. Tanaka (2015). Unified modeling based on SVM and SVR for prediction of forest area ratio by human population density and relief energy. *Proc. IEEE International Geoscience and Remote Sensing Symposium 2015*, 査読有, 2552 - 2555. DOI:10.1109/IGARSS.2015.7326332
- S. Tanaka and R. Nishii (2015). Incorporation of gridded data into the analysis of remotely-sensed images: Basic quantitative strategy to analyze deforestation by population growth. *Proc. IEEE International Geoscience and Remote Sensing Symposium 2015*, 査読有, 2548 - 2551. DOI:10.1109/IGARSS.2015.7326331
- P. Qin and R. Nishii (2015). Statistical prediction of Dst index by solar wind data and t-distributions. *IEEE Transactions on Plasma Sciences*, 査読有, 43 (11), 3908 - 3915. DOI:10.1109/TPS.2015.2485661.

〔学会発表〕(計 21 件)

- 西井 龍映, ビッグデータ時代の機械学習と統計手法, 電気情報通信学会 九州支部専門講習会, 長崎市, 2018.
- 上田 勇祐, 西井 龍映, 江田 智尊, Kullback-Leibler 情報量に基づく Alternating Direction Method of Multipliers による精度行列のスパース推定, 統計関連学会連合大会, 東京都文京区, 2018.
- S. Koda, Y. Onda, H. Matsui, K. Takahagi, Y. Uehara-Yamaguchi, M. Shimizu, K. Inoue, T. Yoshida, T. Sakurai, H. Honda, S. Eguchi, R. Nishii and K. Mochida, Diurnal transcriptome and gene network represented through sparse modeling in *Brachypodium distachyon*, 第 59 回日本植物生理学会年会, 日本植物生理学会, 札幌市, 2018.
- S. Koda, Y. Onda, H. Matsui, R. Nishii and K. Mochida, Statistical gene data analysis and applications to plant growth, Data driven crop design technology, 福岡市, 2018.
- S. Koda, R. Nishii, K. Mochida, and Y. Onda, Statistical gene analysis and contributions to low-carbon society, I2CNER Annual Symposium, 九州大学カーボンニュートラル・エネルギー国際研究所, 福岡市, 2018.
- R. Nishii, 太陽風が弱い地震の引き金であることの統計的検証, 研究集会 データサイエンスの発展と実社会への応用, 科研費研究集会, 福岡市, 2017.
- 江田 智尊, Abdallah Zeggada, Farid Melgani, 西井 龍映, Structured support vector machine を用いた土地被覆画像のマルチラベル判別, 日本分類学会第 36 回大会, 日本分類学会, 2017.
- 喜久山 良式, 西井 龍映, サポートベクター回帰における変数選択, 統計関連学会連合大会, 2017.
- R. Nishii, R. Kikuyama and P. Qin, Feature selection of support vector regression

based on information theoretic criteria, The SIAM Workshop on Parameter Space Dimension Reduction, Society for Industrial and Applied Mathematics, ピッツバーグ, アメリカ, 2017.

西井 龍映, Statistical analysis of global gene expression data and applications to plant growth, I2CNER Annual Symposium 2017, 九州大学カーボンニュートラル・エネルギー国際研究所, 福岡市, 2017.

西井 龍映, 太陽風が弱い地震の引き金であることの統計的検証, 科研費研究集会, 京都市, 2017.

T. Nakamoto, R. Nishii and S. Eguchi, The prediction of ellipsoids for color matching problem, 科研費研究集会, 福岡市, 2016.

江田 智尊, 西井 龍映, 深沢 祐太, Support Vector Machine における変数選択基準, 科研費研究集会, 福岡市, 2016.

R. Nishii and S. Tanaka, Overall review of statistical spatial relations of neighbor cells in deforestation models by human population and topographic attributes, 統計数理研究所共同研究集会, 東京都立川市, 2016.

西井 龍映, ペイジアンネットワークによる構造学習と材料品質への応用ー統計的手法の産業界への貢献ー, 2016 年度日本機械学会年次大会, 日本機械学会, 福岡市, 2016.

西井 龍映, Detection of hub genes by sparse modeling of gene expression time series, I2CNER-IMI Joint Symposium, 九州大学カーボンニュートラル・エネルギー国際研究所, マス・フォア・インダストリ研究所, 2016.

江田 智尊, 松井 秀俊, 西井 龍映, 持田 恵一, 恩田 義彦, 櫻井 哲也, 吉田 拓広, 植物バイオマス増産のための遺伝子データ解析, 情報・統計科学シンポジウム, 統計科学研究会, 福岡市, 2015.

S. Tanaka and R. Nishii, Review on the methodology to explore functional forms of deforestation, 共同研究集会, 統計数理研究所, 東京都立川市, 2015.

江田 智尊, 西井 龍映, Support Vector Machine における変数選択基準, 統計関連学会連合大会, 岡山市, 2015.

R. Nishii and S. Tanaka, Unified modeling based on SVM and SVR for prediction of forest area ratio by human population density and relief energy, IEEE International Geoscience and Remote Sensing Symposium 2015, ミラノ, イタリア, July.29, 2015.

- ⑳ S. Tanaka and R. Nishii, Incorporation of gridded data into the analysis of remotely-sensed images: Basic quantitative strategy to analyze deforestation by population growth, IEEE International Geoscience and Remote Sensing Symposium 2015, ミラノ, イタリア, July.29, 2015.

〔産業財産権〕

出願状況(計1件)

名称: 塗色評価用の等色範囲設定方法及び等色範囲設定装置

発明者: 西井 龍映, 中本 尊元, 山根 貴和, 大友 直人, 石丸 伸吾, 越智 和彦

権利者: 国立大学法人九州大学, マツダ株式会社

種類: 特許

番号: 2018-182317

出願年: 2018 年

国内外の別: 国外

6. 研究組織

(1) 研究分担者

研究分担者氏名：持田 恵一

ローマ字氏名： Mochida Keiichi

所属研究機関名：理化学研究所

部局名：環境資源科学研究センター

職名：チームリーダー

研究者番号(8桁): 90387960

研究分担者氏名：松田 安昌

ローマ字氏名：Matsuda Yasumasa

所属研究機関名：東北大学

部局名：経済学部

職名：教授

研究者番号(8桁): 10301590

研究分担者氏名：田中 章司郎

ローマ字氏名：Tanaka Shojiro

所属研究機関名：広島経済大学

部局名：経済学部

職名：教授

研究者番号(8桁): 00197427

研究分担者氏名：前園 宜彦

ローマ字氏名：Maesono Yoshihiko

所属研究機関名：九州大学

部局名：数理学研究院

職名：教授

研究者番号(8桁): 30173701

研究分担者氏名：増田 弘毅

ローマ字氏名：Masuda Hi roki

所属研究機関名：九州大学

部局名：数理学研究院

職名：教授

研究者番号(8桁): 10380669

(2) 研究協力者

研究協力者氏名：江田 智尊

ローマ字氏名：Koda Satoru

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。