様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　　１　日現在

機関番号：　１２６１２
研究種目：　若手研究(B)
研究期間：　2015～2017
課題番号：　１５Ｋ１５９４７

研究課題名（和文）スパース学習に基づく情報統合型多変量統計手法の研究

研究課題名（英文）Multivariate statistical modeling for information integration via sparse learning

研究代表者

川野　秀一（KAWANO, SHUICHI）

電気通信大学・大学院情報理工学研究科・准教授

研究者番号：５０６１１４４８

交付決定額（研究期間全体）：（直接経費）　　3,000,000円

研究成果の概要（和文）：データ間の構造学習と得られた構造に基づく予測を同時に可能とする，情報統合型統計的モデリング手法の理論・方法論の開発研究に取り組んだ．特に，主成分分析モデルと回帰モデルを統合した1段階法による主成分回帰モデルを軸として研究を進め，スパースモデリングやオンライン学習に基づくデータ解析手法を提案するとともに，スパースモデルに含まれる調整パラメータの選択方法を情報量およびベイズ理論の観点から開発することができた．開発したモデリング手法は，コンソミック系統マウスデータを含む様々な実データに応用した．

研究成果の概要（英文）：We developed statistical methods for information integration that can learn a structure between data and provide a predictive model based on the given structure simultaneously. In particular, we presented a one-stage procedure for principal component regression and developed various data analysis techniques based on sparse modeling and online learning. We also proposed a procedure for selecting a value of tuning parameters in sparse modeling from the viewpoint of information theory and Bayesian approach. We applied the proposed methods into various datasets including mouse consomic strain data.

研究分野：　統計科学

キーワード：　機械学習　スパース学習　主成分回帰　一般化線形モデル　オンライン学習　順序カテゴリカルデータ　情報量規準　ベイズモデリング

１．研究開始当初の背景

データ取得技術の高度な発展に伴い，諸科学のあらゆる分野でビッグデータが取得されはじめていた．得られたビッグデータから有効な情報を抽出するためには，統計モデルの利用が必要不可欠である．統計モデルを用いる利点は，将来の「予測」またはデータ間の「構造学習」を行うことができることである．予測を行う統計モデルの例としては，回帰モデルや判別モデルなどを挙げることができ，構造学習を行う統計モデルの例としては，主成分分析モデル，因子分析モデル，グラフィカルモデルなどを挙げることができる．

元来，予測と構造学習を行う各モデル達はそれぞれ独立に研究されてきたが，これらを同時に可能にする統計モデルを必要とする気運が諸科学の様々な分野で高まっていた．例えば，生命科学分野においては，遺伝子と疾患の状態との関係をモデル化する際，遺伝子個々と疾患の状態とを直接モデル化するよりも，遺伝子の集まり (例えば，遺伝子ネットワークなどの遺伝子間の構造) をまずモデル化し，それを介して疾患の状態とをモデル化した方が，解析結果を容易に解釈できる．しかし，研究開始当初において，国内外を含め，予測と構造学習を同時に可能にする統計モデルの研究は，多くが既存の統計モデルをアドホックに組み合わせたものであり，確かな理論基盤の上でモデルを構成し，有効な情報抽出のための理論および実際面でも有用な方法論については，十分に研究は進んでいるとは言い難かった．

２．研究の目的

データの多様性などを十分に考慮に入れた統計的モデリング手法，特に，データ間の構造学習と得られた構造に基づく予測を同時に可能とする情報統合型統計的モデリング手法の理論・方法論について開発・研究することを目的とした．具体的には，以下の 3 点

(1) 離散値，順序カテゴリー，関数といった様々な種類のデータ形式に対する統計モデルを構築することが重要になること．
(2) lasso 等のスパース推定法は，変数選択とパラメータ推定を同時に行う統計的推定法であり，特に高次元データ解析に対して有効に働くことが知られている．本研究で考えている情報統合型統計モデルにおいても，得られるデータが高次元になることは往々として考えられるため，スパース推定法に基づくパラメータ推定は重要になること．
(3) 構造学習を行う際，従来では無向グラフによりモデル化していたが，構造内の因果をより詳細に記述するためには，有向グラフを用いたモデリングが必要となること．

に着目し，情報統合が可能な統計科学的理論の構築，および高次元かつ多種多様なデータを解析するための方法論の開発を目的として研究を推進した．

３．研究の方法

主成分分析モデル，回帰モデル，オンライン学習等の統計的機械学習手法を融合させ，推定すべきパラメータの自動選択を誘導するスパース推定法に基づく統計的モデリング手法について研究を行った．具体的には，指数型分布族に従うデータ形式まで扱うために一般化線形モデルまでのモデルを想定し，スパース推定法としては lasso に基づく凸罰則による方法を採用した．提案したモデリング手法は，人工データによる検証後，実問題への適用を通してその有効性を検証した．

スパース推定により得られたモデルの性能は，正則化パラメータの値に依存しているため，この値の決定がモデリングの過程において重要となる．スパース推定に含まれる正則化パラメータの値を客観的に選択するために，情報量・ベイズ理論の両観点に基づくモデル評価基準に関する研究を行った．開発した評価基準は，モンテカルロ・シミュレーションや実データ解析などの数値的検証に基づきその有用性を評価した．

４．研究成果

(1) 目的変数に関して得られるデータが実数のときにおける 1 段階法に基づく主成分回帰モデルを開発した．従来，主成分回帰モデルでは，まず主成分分析を実行し，得られた主成分得点を説明変数として回帰分析を行う 2 段階法が採られている．提案モデルでは，2 乗損失関数と主成分損失関数の加重平均に基づきモデルを構成し，1 段階法による主成分回帰モデルを実現した．さらに，損失関数にスパース正則化を課したパラメータ推定法を採用することにより，変数選択ならびにパラメータの一意性を保証した推定方式を確立した．人工データによる数値実験や実データへの適用を通して，2 段階法による主成分回帰モデルや部分最小二乗回帰モデルと比較することにより，提案手法の有効性を検証した．
(2) 目的変数に関して得られるデータが指数型分布族に従う場合における 1 段階法に基づく主成分回帰モデルを開発した．一般化線形モデルに着目し，そのモデルから自然に導入される尤度ベースの損

失関数と主成分損失関数およびスパース正則化項を用いることにより，1 段階法による主成分回帰モデルを提案した．提案手法と様々な種類のスパース主成分分析手法や一般化線形モデルに対する部分最小二乗回帰モデルを数値的に比較することにより，提案手法の有効性を多角的に検証した．また，コンソミック系統マウスデータの解析に提案手法を適用し，現象の解明に寄与することができた．研究成果(1)と(2)で開発したモデルを計算するソフトウェアを統計解析ソフトウェア R のパッケージ spcr として作成し，一般に公開した．

(3) 順序付きカテゴリカルデータを扱うことが可能な連続比ロジットモデルに着目し，平行性仮定と説明変数の自動選択を目的とした，スパース推定に基づく順序ロジットモデリング手法を開発した．具体的には，連続比ロジットモデルから導入される対数尤度関数と一般化 fused lasso に基づく目的関数を提案し，交互方向乗数法による計算アルゴリズムを開発した．企業の格付けデータへの適用を通して，提案手法の有効性を検証した．

(4) 逐次的にデータが得られ，その都度データを解析しモデルを構築する機械学習手法は，オンライン学習と呼ばれている．オンライン学習アルゴリズムの一つである adaptive regularization of weight vectors (AROW) に着目し，特徴選択を有したスパース AROW を開発した．スパース推定法には lasso を採用し，座標降下法を組み込んだ計算アルゴリズムを開発した．

また，分布更新に基づく決定木のオンライン学習モデルを提案した．決定木のオンライン学習はこれまでミニバッチ学習の枠組みでしか研究されてこなかった．本研究では，それぞれの群に属するデータの情報を確率分布で捉え，カルバック-ライブラーダイバージェンスに基づいた損失関数を導出し，逐次的に得られるデータのみでモデルの更新が可能なオンライン決定木学習手法を提案した．

(5) スパース推定により得られる推定量の漸近分布を考えることにより，スパース推定された一般化線形モデルの評価基準を導出した．特に，lasso により得られたモデルの評価基準を導出し，その評価基準の有効性を，モンテカルロ・シミュレーションや実データへ適用することにより評価した．提案した評価基準を計算するソフトウェアを，統計解析ソフトウェア R のパッケージ sAIC として作成し，一般に公開した．

(6) lasso をベイズ理論の枠組みで捉えたベイジアン lasso に対し，そのベイズモデルに含まれるハイパーパラメータの値を客観的に選択する方法を開発した．具体的には，ベイズ型予測分布に基づいた予測情報量規準を提案するとともに，効率的にパラメータを選択するための最適化アルゴリズムを提案した．

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　８　件）

① 川野秀一 (2018) スパース正則化に基づく回帰モデリングとその計算アルゴリズム，計算機統計学，印刷中，査読有．

② Kawano, S., Fujisawa, H., Takada, T. and Shiroishi, T. (2018) Sparse principal component regression for generalized linear models, Computational Statistics & Data Analysis, 124, 180-196, DOI (10.1016/j.csda.2018.03.008), 査読有．

③ 加藤駿典，川野秀一 (2017) Fused Lasso に基づくスパース順序ロジットモデリング，計算機統計学，30，3-16，DOI (10.20551/jscswabun.30.1_3)，査読有．

④ 野崎俊貴，木村拓海，川野秀一 (2016) スパース推定に基づく適応正則化オンライン学習の特徴選択問題，計算機統計学，29，117-131，DOI (10.20551/jscswabun.29.2_117)，査読有．

⑤ Ninomiya, Y. and Kawano, S. (2016) AIC for the Lasso in generalized linear models, Electronic Journal of Statistics, 10, 2537-2560, DOI (10.1214/16-EJS1179), 査読有．

⑥ 嶋村海人，川野秀一，小西貞則 (2015) モデル平均化法による Bayesian lasso 回帰モデリング，応用統計学，44，101-117，DOI (10.5023/jappstat.44.101)，査読有．

⑦ Kawano, S., Hoshina, I., Shimamura, K. and Konishi, S. (2015) Predictive model selection criteria for Bayesian lasso regression, Journal of the Japanese Society of Computational Statistics, 28, 67-82, DOI (10.5183/jjscs.1501001_220), 査読有．

⑧ Kawano, S., Fujisawa, H., Takada, T. and Shiroishi, T. (2015) Sparse principal component regression with adaptive loading, Computational Statistics & Data Analysis, 89, 192-203, DOI (10.1016/j.csda.2015.03.016), 査読

有.

〔学会発表〕（計　１４　件）

① 川野秀一，スパース推定法による統計モデリング（招待講演），第 12 回日本統計学会春季集会，2018 年 3 月，早稲田大学（東京都・新宿区）.
② Kawano, S., Fujisawa, H., Takada, T. and Shiroishi, T., Principal component regression for generalized linear models via L1-type regularization, The 10th International Conference of the ERCIM WG on Computational and Methodological Statistics, 2017 年 12 月，ロンドン（イギリス）.
③ Kawano, S., Fujisawa, H., Takada, T. and Shiroishi, T., A one-stage principal component regression method with sparse regularization（招待講演），2016 CSA & NCCU Joint Statistical Meetings，2016 年 12 月，台北（台湾）.
④ 木村拓海，川野秀一，分布の逐次推定に基づくオンライン決定木学習，第 19 回情報論的学習理論ワークショップ，2016 年 11 月，京都大学（京都府・京都市）.
⑤ 川野秀一，藤澤洋徳，高田豊行，城石俊彦，スパース主成分多項ロジスティック回帰モデリングとその応用，2016 年度統計関連学会連合大会，2016 年 9 月，金沢大学（石川県・金沢市）.
⑥ 木村拓海，川野秀一，識別・判別問題におけるオンライン決定木学習，2016 年度統計関連学会連合大会，2016 年 9 月，金沢大学（石川県・金沢市）.
⑦ Kawano, S., Fujisawa, H., Takada, T. and Shiroishi, T., A one-stage approach for principal component regression via L1-type regularization, The 4th Institute of Mathematical Statistics Asia Pacific Rim Meeting，2016 年 6 月，香港（中国）.
⑧ Kawano, S., Fujisawa, H., Takada, T. and Shiroishi, T., One-stage estimation of principal component regression with sparse regularization, The 8th International Conference of the ERCIM WG on Computational and Methodological Statistics, 2015 年 12 月，ロンドン（イギリス）.
⑨ 川野秀一，スパース推定と統計解析（招待講演），2015 年度統計関連学会連合大会，2015 年 9 月，岡山大学（岡山県・岡山市）.
⑩ 野崎俊貴，川野秀一，適応正則化オンライン学習における特徴選択問題，2015 年度統計関連学会連合大会，2015 年 9 月，岡山大学（岡山県・岡山市）.
⑪ 加藤駿典，川野秀一，Fused Lasso に基づくスパース順序ロジットモデリング，2015 年度統計関連学会連合大会，2015 年 9 月，岡山大学（岡山県・岡山市）.
⑫ 川野秀一，藤澤洋徳，高田豊行，城石俊彦，1 段階法による主成分回帰モデリング，日本計算機統計学会第 29 回大会，2015 年 5 月，山梨県立図書館（山梨県・甲府市）.

〔図書〕（計　０　件）

〔産業財産権〕

○出願状況（計　０　件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　０　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等：
http://kjk.office.uec.ac.jp/Profiles/68/0006701/profile.html

６．研究組織
(1)研究代表者
川野　秀一（KAWANO, Shuichi）
電気通信大学・大学院情報理工学研究科・准教授
研究者番号：５０６１１４４８

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
（　　）

研究者番号：

(4)研究協力者
（　　）