

科学研究費助成事業 研究成果報告書

令和 4 年 6 月 17 日現在

機関番号：82111
研究種目：基盤研究(C)（一般）
研究期間：2019～2021
課題番号：19K06308
研究課題名（和文）NIR3.0にむけたAIを活用した分光ビッグデータの解析手法の確立

研究課題名（英文）Data cleansing method for NIR spectral data

研究代表者

源川 拓磨（GENKAWA, Takuma）

国立研究開発法人農業・食品産業技術総合研究機構・食品研究部門・主任研究員

研究者番号：10571698

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究では、スペクトルデータ向けのデータクレンジング法（以下、DC法）を開発し、開発したDC法の有用性をモデルデータおよび実サンプルデータを使って明らかにした。モデルデータ（糖水溶液）の回帰分析では、DC法の適用により、糖由来のバンドの強度変化を直接反映した回帰モデルを構築できた。その結果、回帰モデルの複雑さ（説明変数の数）を低減しつつ高い回帰精度を実現した。実サンプルデータでの回帰分析では、11種類の青果物（トマト、バナナ、モモ、キウイ等）の果汁についてBrix糖度の回帰分析を行ったところ、DC法を適用することで決定係数が0.70から0.99に、RMSEが3.3%から0.5%に改善した。

研究成果の学術的意義や社会的意義

本研究ではスペクトルデータに対する機械学習におけるデータクレンジング（データ前処理）に焦点を当て、これまで皆無であったスペクトルデータ向けのデータクレンジング法を新たに開発した。本データクレンジング法を適用することによって、これまで十分に検討されてこなかったスペクトルデータに対するデータクレンジングを効率的に実施することができ、機械学習によるスペクトル解析の解析結果を改善することが可能となった。これにより、農産物の品質を非破壊で計測する分光センサーの開発期間を大幅に短縮することができ、分光センシングに基づく高度な食料生産体制の構築に貢献することが期待される。

研究成果の概要（英文）：In this study, a data cleansing method for spectral data (DC method) was developed and clarified the usefulness of the DC method using model data and real sample data. In the regression analysis of model data (aqueous sugar solution), the application of the DC method enabled us to construct a regression model that directly reflects changes in the intensity of sugar-derived bands, resulting in high regression accuracy while reducing the complexity (number of explanatory variables) of the regression model. In the regression analysis on actual sample data, Brix sugar content was regressed on fruit juices of 11 different fruits and vegetables (tomatoes, bananas, peaches, kiwi, etc.), and the application of the DC method improved a coefficient of determination from 0.70 to 0.99 and an RMSE from 3.3% to 0.5%.

研究分野：ポストハーベスト工学、分析化学

キーワード：非破壊計測 スペクトル解析 機械学習 データクレンジング AI

1 . 研究開始当初の背景

農産物の非破壊評価技術の中で、近赤外分光法を用いた内部成分の評価は最も成功したものの1つである。例えば果物糖度の測定や小麦粉のタンパク質含量の測定は広く実用化されており、加工や流通の現場での日常分析に採用されている。この近赤外分光法による非破壊評価技術は、これまでに大きな2回の革新 (NIR1.0 , NIR2.0) を経て発展してきたが、いま、3回目の変革期 (NIR3.0) を迎えようとしている。

NIR1.0 : 多変量解析との融合による非破壊評価技術としての確立

近赤外光の存在は19世紀には知られていたが、長らく研究対象とはならなかった。可視光や中赤外光と比べて近赤外光のスペクトルは複雑な形状であり、ピーク波長の帰属に基づく伝統的なスペクトル解析手法の適用が困難であったためである。しかし、1950年代にアメリカのKarl Norrisが近赤外スペクトルの解析に多変量解析を適用したことによって、小麦の水分とタンパク質を非破壊で定量できることが発見された。スペクトル解析手法の常識にとらわれなかったKarl Norrisの柔軟な発想によって1つ目の革新が起こり、近赤外分光法による非破壊評価技術は農産物・食品をはじめ、高分子化学や製薬といった様々な分野へと波及した。

NIR2.0 : ポータブル分光計の誕生による生産現場での活用

2000年代に入る頃までに、近赤外分光法による非破壊評価技術は世界中に普及し、公定法として採用されるまでになっていた。この頃、分光計はベンチトップ型で、試料をホルダーに詰めて測定が行われていたが、2000年代初頭に持ち運びが可能なポータブル分光計が開発されたことを契機に、試料を分光計の状態(場所やホルダーの形状)に合わせるのではなく、分光計を試料の状態に合わせて近赤外スペクトルを取得できるようになった。これにより、近赤外分光法による非破壊評価技術は生産現場での「あるがまま分析」へと発展した。

NIR3.0 : IoTのセンサー端末としての役割と分光ビッグデータの誕生

そして現在、近赤外分光法による非破壊評価技術は3回目の変革期を迎えようとしている。小型化と低価格化が進むスマートフォンと同様、分光計は手のひらに収まるサイズまで小さくなり、価格も数万~数10万円に低下した。例えば、5万円以下の近赤外分光計で、測定したスペクトルデータをスマートフォンに送り、スマートフォンの中のアプリを使ってその場で解析が行えるものが既に市販されている。この技術革新によって、いつでも、誰でも、どこでもスペクトルデータを取得でき、さらにインターネット経由で大規模にスペクトルデータを収集することも可能となっている。つまり、現在は分光ビッグデータと呼ぶべき膨大な数のスペクトルデータがいつ誕生してもおかしくない状況にあるのである。NIR3.0においては、近赤外分光計はIoTにおけるセンサー端末として活用され、分光ビッグデータの解析結果に基づいた高度な食料生産体制の構築に貢献することが期待されている。

2 . 研究の目的

本研究は、AIを活用した分光ビッグデータの解析手法を確立することを目的とし、機械学習によるスペクトル解析の要となるスペクトルデータ向けのデータクレンジング法を開発する。

3 . 研究の方法

(1) データクレンジング法の開発

本研究ではまず、スペクトルデータ向けのデータクレンジング法を開発した。データクレンジングでは、オリジナルデータ(図1A)から解析波長域を選択し(図1B、この例では600波長点から50波長点を選択)、選択した波長域についてスペクトル前処理条件を最適化する(図1C、この例では2条件の組み合わせ)。最適化された前処理条件を施したスペクトルデータ(図1D)を使って回帰分析等の機械学習を行う。

(2) モデルデータおよび実サンプルデータでの検証

開発したデータクレンジング法の有効性を明らかにするために、解析結果を検証可能なモデルサンプルの近赤外スペクトルを用いた解析と、実サンプルでの性能評価のための農産物や食品素材の近赤外スペクトルを用いた解析を行った。データクレンジング法の適用の有無による機械学習による解析結果を比較し、定量分析の場合は決定係数とRMSE、判別分析の場合は判別率の中率を指標として、開発したデータクレンジング法の有用性を評価した。

4 . 研究成果

(1) データクレンジング法の開発

スペクトルデータ向けのデータクレンジング法を開発した。オリジナルデータ(原スペクトル)から解析波長域を選択するために、Moving-window PLS 回帰法とSNV処理を組み合わせた解析ア

ルゴリズム (Searching Region SNV)) を採用し, さらに選択された波長域について, グリッドサーチ法により 2 次微分処理の計算パラメーターを最適化する手法を採用した。これらを組み合わせた手法を, オリジナルデータからベースライン変動とバンド重複の外乱因子を排除するデータクレンジング法とした。

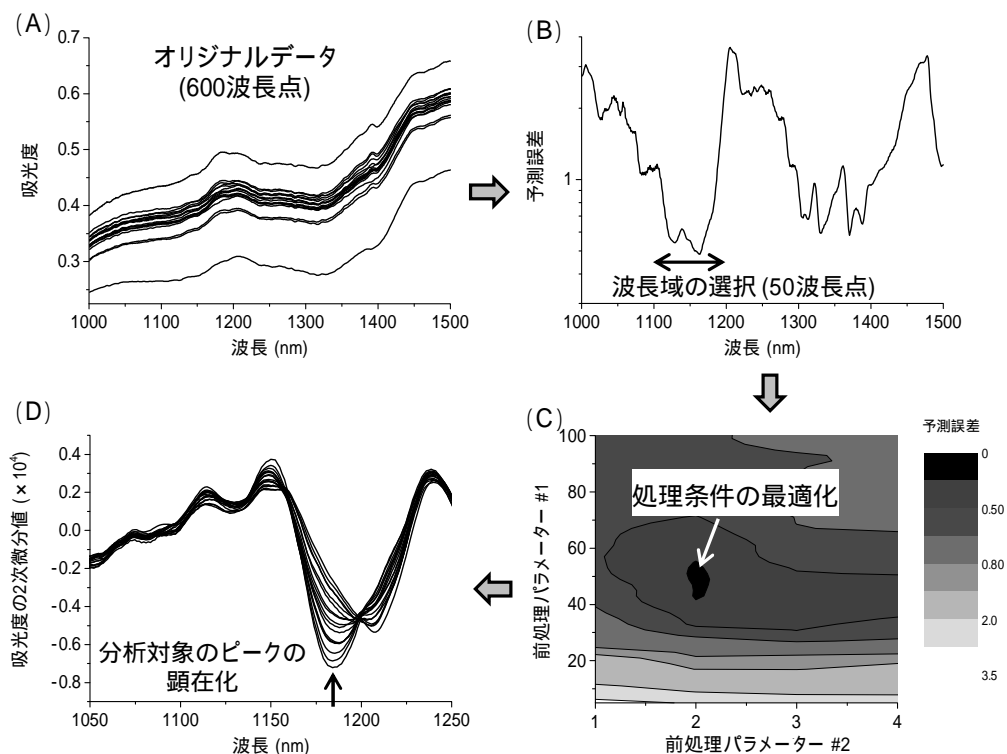


図 1 データクレンジング法の概要

(2) モデルデータでの検証

開発したデータクレンジング法の有効性を明らかにするために, 解析結果を検証可能なモデルサンプル (糖濃度を段階的に調整した水溶液, 図 2A) の近赤外スペクトルを用いて機械学習の 1 つである回帰分析を行った。データクレンジング法の適用により (図 2B, 図 2C), 溶媒である水由来のバンドに基づく間接的な回帰モデルではなく, 基質である糖由来のバンド (2274 nm) の強度変化を直接反映した回帰モデルを構築することが可能であった (図 2D)。その結果, 回帰モデルの複雑さ (説明変数の数) は低減され (LV=1), 機械学習によって構築された回帰モデルについて科学的な根拠を検討することが可能であった。回帰精度についても, データクレンジングを適用した方が良い結果を示した (図 2E: データクレンジング法を適用した場合, 図 2F: データクレンジング法を適用せずに説明変数を 1 とした場合) ことから, 開発したデータクレンジング法がスペクトルデータの機械学習に有効であることが確かめられた。

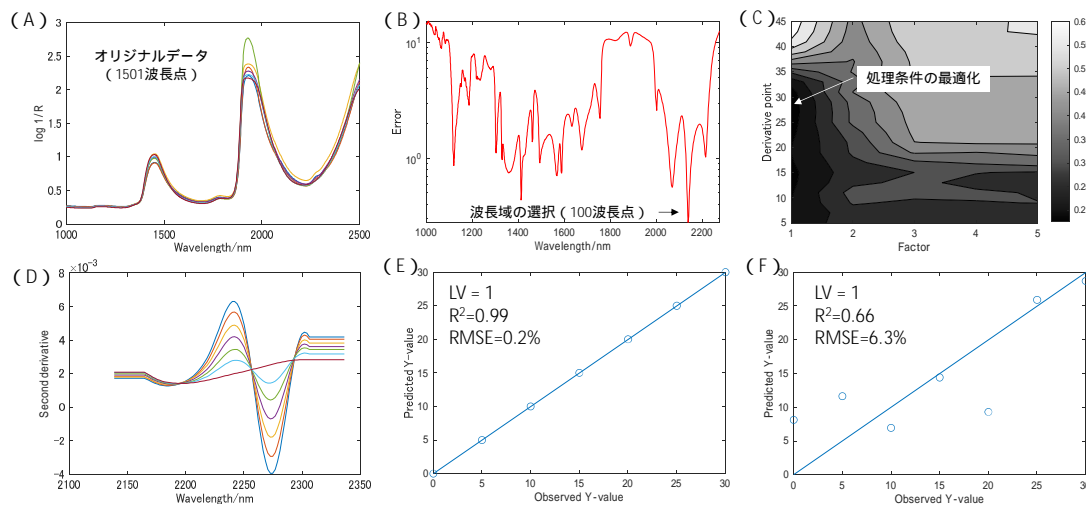


図 2 モデルデータにデータクレンジング法を適用した結果

(3) 実サンプルデータでの検証

実サンプルとしてアサリ貝および青果物の果汁を用いて、前者は鮮度判別を、後者では Brix 糖度の回帰分析を行った。アサリ貝の鮮度判別においては、市販のアサリ貝を購入して4日間冷蔵し、貯蔵中のアサリ貝の近赤外スペクトルを殻付きのまま測定した(図3A, 貯蔵前:青, 貯蔵後:赤)。購入日のものを貯蔵前、冷蔵したもの(1~4日)を貯蔵後として、PLS 判別法を用いて貯蔵の有無の判別モデルを構築した(図3B~3D, 図3Dの貯蔵前:青, 貯蔵後:赤)。スペクトル前処理を行わない場合、判別の中率は87%であったが(図3F)、データクレンジング法の適用によって100%に向上した(図3E)。

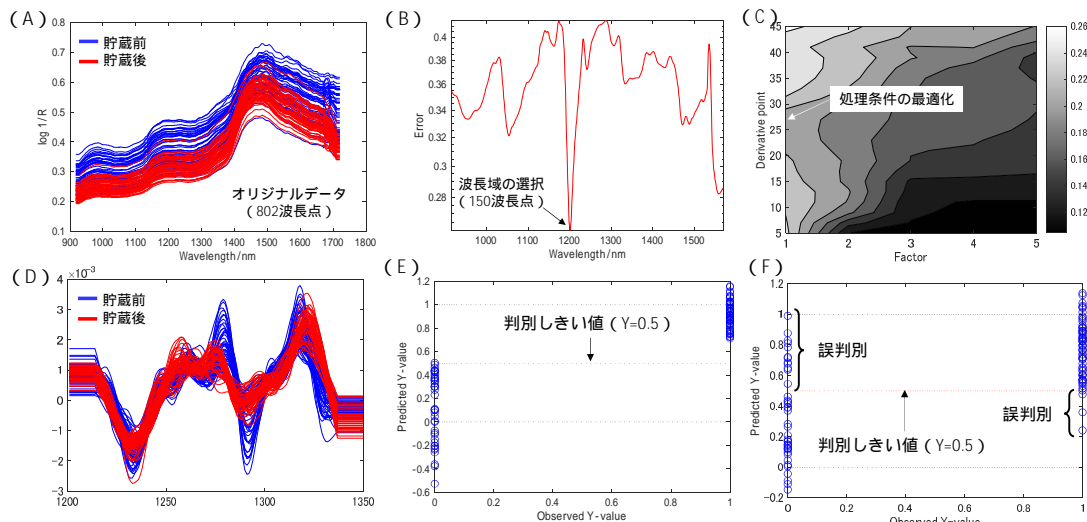


図3 実サンプルデータ(アサリ貝)にデータクレンジング法を適用した結果

青果物果汁の糖度分析においては、11種類の青果物(トマト, バナナ, モモ, キウイ等)の果汁について Brix 糖度(5~25%)と透過反射スペクトルを測定し(図4A), PLS 回帰法を用いて Brix 糖度の検量モデルを構築した。スペクトル前処理を行わない場合, 決定係数は0.70, RMSEは3.3%であったが(図4F), データクレンジング法の適用によって決定係数が0.99, RMSEが0.5%に向上した(図4B~図4E)。

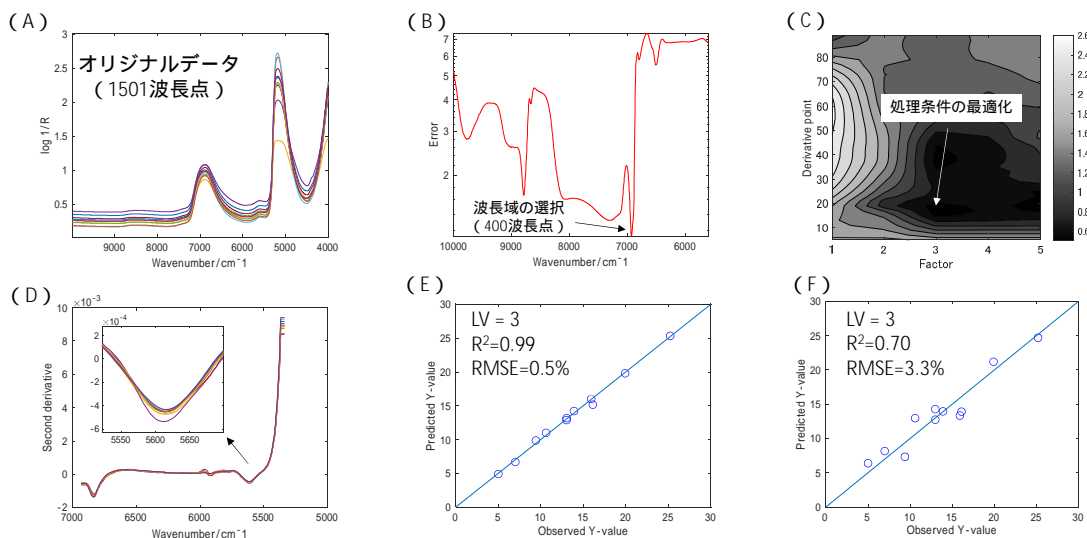


図4 実サンプルデータ(青果物果汁)にデータクレンジング法を適用した結果

(4) 異常サンプルの検出

開発したデータクレンジング法を用いて異常サンプルの検出を試みた。モデルサンプル(糖濃度を段階的に調整した水溶液)の近赤外スペクトルに測定ミスを起こすことで意図的に歪ませたスペクトルを混入させ、これにデータクレンジング法を適用した。その結果、意図的に歪ませたスペクトルデータを異常サンプルとして検出することができた。これはデータクレンジング法によって正常なスペクトルデータでは高い線形関係が得られる一方で、意図的に歪ませたスペクトルデータはこの線形関係から大きく外れるために異常サンプルとして検出されると考えられた。

(5) 分光イメージングのスペクトルデータへの適用：

開発したデータクレンジング法の汎用性を評価するために分光イメージングのスペクトルデータへの適用を検討した。分光イメージングの解析において試料輪郭の抽出を効率的に行うために、データクレンジング法のアルゴリズムを一部変更して、ベースライン変動を補正しつつ第1主成分の寄与率を最大化するデータクレンジング法を開発した。検証用のモデルサンプルとして小麦粉を詰めたガラスシャーレを用いた結果、データクレンジング法を適用することによって試料部分と背景部分とを明瞭に識別可能となることが確かめられた(図 5A と図 5C がデータクレンジング法を適用。図 5B と図 5D はデータクレンジング法の適用無し。)

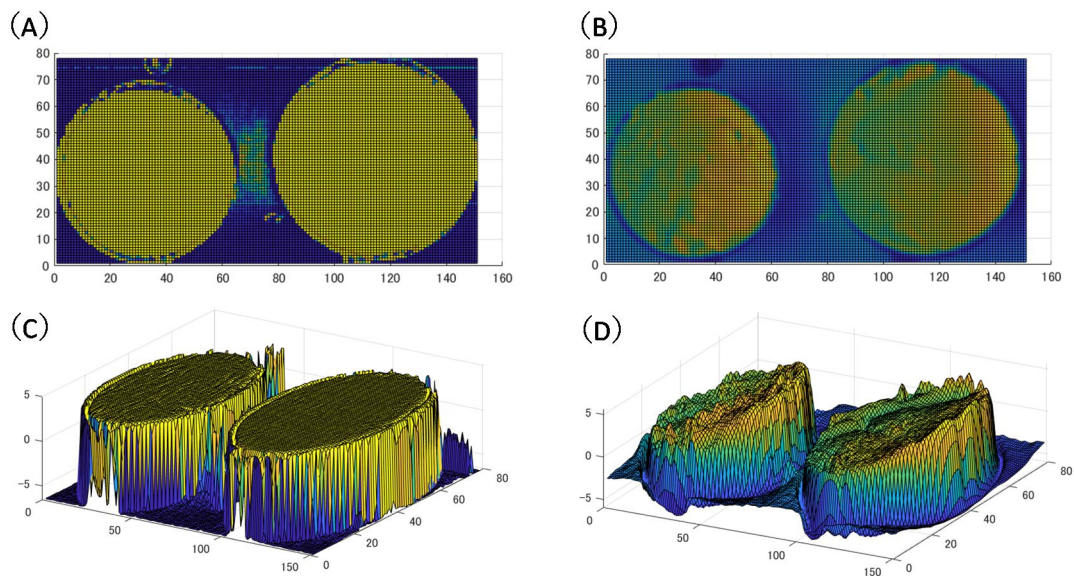


図5 分光イメージングのスペクトルデータにデータクレンジング法を適用した結果

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 宮田彩希、野口良造、Tofael Ahamed、源川拓磨
2. 発表標題 近赤外分光法を用いた貝類の非破壊鮮度評価
3. 学会等名 第36回近赤外フォーラム
4. 発表年 2020年

1. 発表者名 藤井未来、野口良造、Tofael Ahamed、源川拓磨
2. 発表標題 教師あり学習におけるPLS-DAを用いたデータ選別手法のための最適条件の探索
3. 学会等名 第36回近赤外フォーラム
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------