

氏 名 岩山 めぐみ

学位(専攻分野) 博士(統計科学)

学位記番号 総研大甲第 2408 号

学位授与の日付 2023 年 3 月 24 日

学位授与の要件 複合科学研究科 統計科学専攻
学位規則第6条第1項該当

学位論文題目 材料科学における多次元出力変数の教師あり学習に関する
研究

論文審査委員 主 査 日野 英逸
統計科学専攻 教授
藤澤 洋徳
統計科学専攻 教授
Stephen Wu
統計科学専攻 准教授
津田 宏治
東京大学 大学院新領域創成科学研究科 教授

(様式3)

博士論文の要旨

氏名 岩山 めぐみ

論文題目 材料科学における多次元出力変数の教師あり学習に関する研究

はじめに

材料科学分野においては、データ科学を用いた構造・物性・機能の予測による開発フロー・製造プロセスの効率化を進めることが益々重要になってきており、マテリアルズインフォマティクス (MI) への注目が高まっている。MI の機械学習モデルでは、化学組成や配合、製造時のプロセス条件を入力変数として扱い、記述子と呼ばれるベクトルに変換した上で、製品として要求される物性や機能、構造の特徴を出力変数として予測する。その際に、出力変数として用いられる機能、物性、構造の特徴量は、スカラーを対象とすることが多い。しかしながら、材料研究で活用されるデータは、スカラーには限らず、例えば、光吸収の特性を明らかにするために測定される紫外線可視近赤外光 (UV-Vis-NIR) 吸収スペクトルのような関数や、走査電子顕微鏡を用いて観察される材料の表面や界面構造の電子顕微鏡像のような行列・テンソルといった多次元配列で表現されるデータも数多く存在している。しかしながら、これらの多次元配列を出力変数とした教師あり学習の研究は材料科学分野ではほとんど行われてきておらず、体系的に整理されていない。本研究は、材料科学分野に潜んでいる高次元出力変数の問題発掘と、高次元出力変数にも対応し得る方法論の提案を目的とする。材料科学における多くの多次元出力変数の潜在的な回帰問題設定に適用可能な統一的なフレームワークとして、Conditional GAN (cGAN) と関数データ回帰に着想を得て独自にモデリングした Kernel Regression with Functional Outputs (KRFO) の性能を、UV-Vis-NIR 吸収スペクトルと走査電子顕微鏡像の微細構造の予測から評価する。

研究方法

UV-Vis-NIR 吸収スペクトルの関数データを対象に、有機物の化学構造からスペクトル関数を予測する実験と、金属薄膜の電子顕微鏡像の画像データを対象に、組成およびプロセス条件から電子顕微鏡像の微細構造を予測する実験を行う。使用するデータについては、UV-Vis-NIR 吸収スペクトルの予測では、高速液体クロマトグラフィーを用いて測定された 949 種類の市販化合物の UV-Vis スペクトル (データセット I)、分光光度計で測定された 2,222 種類の市販医薬品分子の UV-Vis スペクトル (データセット II)、米国地質先行研究で用いられたデータセット I とデータセット II およびアメリカ地質調査所のスペクトルライブラリから得られた 68 種の有機物の NIR スペクトル (USGS) と、データ量が異なる 3 種のデータセットを用いた。電子顕微鏡像の予測では、マグネトロンスパッタリングにより生成された薄膜材料の組成とプロセス条件、走査電子顕微鏡像が組み合わさった 123 種のデータセットを用いた。

結果

データセット I と II に対して実験した結果、cGAN については、入力変数に用いるノイズ変動の影響により、予測したスペクトル関数は揺らぎを有することが分かった。一方、KRFO ではスペクトルの細かいピーク様の山や関数の形状についても多くの分子でほぼ完全に予測できることが分かった。場合によっては、小さなピークや緩やかな丘状のピークのような目視でも明確に認識しづらい特徴についても予測されることが示され、化学的な部分構造の有無が分子の UV-Vis スペクトルの主要因であると考えられる。極端にデータ数が少ない USGS に対する実験についても、KRFO は十分に高い予測精度を達成することができ、データ量が限定された場合においても KRFO は多次元出力変数を高精度に予測し得ることが示唆された。走査電子顕微鏡像の実験では、わずか 100 程の学習データであったにも関わらず、KRFO では、実際の走査電子顕微鏡像から観察される微細構造の粒子サイズや形状といった形態学的な特徴の違いを捉えられることが分かった。cGAN は画像生成手法として有名であるが、極端にデータ量が少ないため、微細構造の特徴を予測することはできなかった。UV-Vis-NIR 吸収スペクトル予測および微細構造の予測においても、データ量が非常に少ないデータ量にも関わらず KRFO では学習が成功した背後にあるメカニズムとして、マルチタスク学習に共通する学習メカニズムが存在すると考えられる。そこで、データセット I に対して、KRFO による関数全体を予測するケースと、ピーク位置と強度をスカラー値として従来実施されてきたような単純な多層パーセプトロンを用いて予測するケースと、波長に対してのスペクトルの強度をそれぞれ独立に多層パーセプトロンを用いて予測するケースとを比較し、それぞれのピーク位置、強度の正解率を比較した。結果、KRFO による関数全体の予測がいずれの方法よりも、予測性能が圧倒的に上回ることが分かった。よって、独立に予測するのではなく、関数全体を同時に予測することが、予測性能を向上させる鍵であり、マルチタスク学習との関連性が示唆された。

考察

多次元出力変数の回帰問題に対して cGAN と KRFO の二つの方法の性能を検証した。化学構造からの UV-Vis-NIR 吸収スペクトル関数の予測と、組成・プロセス条件からの電子顕微鏡像の微細構造の予測の二つのケースから、KRFO の予測可能性について示し、さらにデータ量が少ない場合にも KRFO は、十分に高い予測精度を保つことを示した。学習が成功したその背景には、独立に関数の値や画像の強度を予測するタスクに比べ、複数のタスクからのデータを同時に使用することで、データ拡張によるタスク固有のノイズを抑えることができたことがあると考えられ、KRFO にはマルチタスク学習に共通する学習メカニズムが存在している。材料科学分野での、多次元出力変数の回帰に対する研究は、まだ始まりを迎えたばかりであり、本研究で提案した方法以外にも有望な方法はあると考えられる。この観点からの研究の出発点として、スペクトル関数と電子顕微鏡像の 2 つのケースに注目したが、材料科学分野の潜在的な問題設定は多数存在するため、今後の発展が期待される。

博士論文審査結果

Name in Full 氏名 岩山 めぐみ

Title 論文題目 材料科学における多次元出力変数の教師あり学習に関する研究

2023 年 1 月 27 日午前 10 時から約 2 時間にわたり岩山めぐみ氏の博士論文審査委員会を開催した。出願者による 1 時間の公开发表による概要説明と質疑応答，さらに約 1 時間の審査委員のみによる審査を行った結果，審査委員会は本論文が学位の授与に値すると判断した。

[論文の概要]

出願論文は日本語で執筆されており，6 章 78 頁からなる。マテリアルズインフォマティクス（以下 MI という）では，出力変数が関数の形で与えられる回帰分析の問題設定が多数存在するが，このような視点に基づく研究はほとんど進んでいない。本研究は，材料科学における関数出力回帰の未開拓問題の発掘と統計的方法論の構築を目的とする。

第 1 章では，MI の近年の動向を概説したのち，材料研究における関数出力回帰の重要性を論じ，研究の問題意識と学術的意義・貢献を説明している。

第 2 章では，分子の光吸収スペクトルの予測，材料微細構造（電子顕微鏡画像）の予測，物性の温度・周波数依存性の予測など，材料研究における関数出力回帰の適用範囲の広さと潜在的可能性を論じ，先行研究のレビューを行っている。

第 3 章では，カーネル回帰と深層学習を組み合わせた関数出力回帰モデルと推定手法を提案している。条件付き敵対的生成ネットワークなどの深層学習に基づく手法や関数データ解析の分野で研究されてきた既存手法を取り上げ，提案手法の特色やデータが少ない状況における利点などを論じている。

第 4 章では，モデルのネットワーク構造の設計などを詳細に説明している。

第 5 章では，有機分子の光吸収スペクトルの予測，複合薄膜材微細構造の画像予測への適用例を通じて，提案手法と既存手法の予測性能を比較している。特にデータが少ない状況において，提案手法は著しく高い汎化性能を示すことが報告されている。この観測結果に対し，関数全体を予測することでマルチタスク学習と同様のメカニズムが働く可能性があることが示唆されている。この仮説を検証するために，通常の教師あり学習で関数の値を独立に予測するケースとスカラーの関数特徴量を予測するケースとの比較を行い，関数全体の予測が圧倒的に高い汎化性能を示すことを実験的に確認している。

第 6 章は，まとめと今後の課題の章である。

[論文の評価]

MI において、少数データに基づく関数出力変数の予測を対象とした体系的な研究はほとんど存在しない。そのような中、データ科学の独自の視点から新しい問題設定を発掘し、カーネル回帰と深層学習に基づく独自の方法論を構築した上でその有用性を実証したことは高く評価できる。また、関数全体を同時に学習・予測することでマルチタスク学習と同様のメカニズムが働き、従来のスカラー変数の予測よりも圧倒的に高い汎化性能を引き出せることが実験的に示された。このような現象を発見したことも本研究の学術的貢献といえる。以上より、本論文は統計科学分野の博士論文として十分に高い水準に達していると判断する。

なお、第 3 章から第 5 章までの研究をまとめた論文が、アメリカ化学会の査読付きジャーナル *Journal of Chemical Information and Modeling* 誌に掲載されている（第一著者）。