

令和 4 年 6 月 13 日現在

機関番号：82108
研究種目：基盤研究(C)（一般）
研究期間：2019～2021
課題番号：19K11988
研究課題名（和文）材料データの探索的可視化分析と探索シナリオの再利用を支援するシステムの研究開発

研究課題名（英文）Integrated Platform Supporting Exploratory Visual Analysis and Resuse of
Exploration Scenarios for Material Informatics

研究代表者
藤間 淳（FUJIMA, Jun）

国立研究開発法人物質・材料研究機構・統合型材料開発・情報基盤部門・主幹エンジニア

研究者番号：30447081
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、ウェブ上に探索的可視化分析環境を構築し、新規材料探索・設計のための統合プラットフォームを実現した。既存材料データや科学計算によるシミュレーション結果に対して、様々な統計解析や機械学習手法を即座に適用し、それらを用いたインタラクティブなデータ可視化・分析を可能とする統合プラットフォームの研究開発を行った。また、本プラットフォーム上で行われた可視化・分析過程を再利用することによって、特定の種類の材料データ解析に共通に用いられる探索手法の抽出や、新たな分析シナリオの構築を支援することを可能にするシステムを構築した。

研究成果の学術的意義や社会的意義

従来のデータ分析や可視化を用いた新たな知識の発見は、職人技に近いものであり、革新的な発見の多くは、研究者が別のものを探索しているときや間違っただけで偶然に発見されたものであることも多い。従来のデータ分析手法に加えて、このような職人技、あるいは偶然の発見過程を、研究者が自在にアクセスし、自身の目的のために再利用することができれば、新技術の開発や科学の発展におおいに貢献することが期待される。本研究では、材料分野における新規材料の発見、設計のための柔軟な解析環境を提供し、その環境内での利用者の行動を解析することで、それを再利用する基盤を提供しようとするものである。

研究成果の概要（英文）：In this study, an exploratory visual analytics environment was constructed on the web to realize an integrated platform for novel material exploration and design. We developed an integrated platform that enables interactive data visualization and analytics by applying various statistical analysis and machine learning methods to given material data and simulation results from scientific calculations. In addition, by reusing the visualization and analysis processes performed on this platform, we have developed a system that enables the extraction of exploration methods commonly used for specific types of material data analysis and supports the construction of new analysis scenarios. A part of the developed platform is available at <https://cads.eng.hokudai.ac.jp/>.

研究分野：マテリアルズ・インフォマティクス、データベース、情報可視化

キーワード：探索的可視化分析 マテリアルズ・インフォマティクス 可視化分析プラットフォーム 操作履歴管理

1. 研究開始当初の背景

近年の急速な計算機技術の発達と計算科学の発展により、物質・材料分野においてもハイスループット計算・ハイスループット実験手法の発展により短期間で大量のデータの生成が可能になってきている。そのようにして得られた大量のデータから、新たな材料設計のための意味のある情報を得るための方法論として様々な提案がなされており、従来のデータの解析に加えて情報科学の積極的な活用をしようという取り組みが材料情報学(マテリアルズ・インフォマティクス)として進められている。

情報学的手法をうまく生かして新規物質・材料を効率的に発見・設計するためには、文献や実験、計算シミュレーションなど様々なソースから得られるデータを統合的に情報処理可能な形でデータベースを構築・管理し、それに対して記述子の計算、クラスタリングなどの解析手法、機械学習による物性値の予測、様々な可視化手法を探索的に組み合わせ統合的に扱うことができるプラットフォームの構築が不可欠である。科学計算や機械学習などの様々な機能がフリーソフトとして公開され、研究者は Python などのプログラミング言語を駆使し、充実したライブラリの中から必要なものを組み合わせてデータの解析・可視化が可能になった。しかし、そのためにはデータ科学的解析手法の知識に加えて高度なプログラミングやライブラリの知識が必要であり、それらを適切に組み合わせる目的に応じた分析をすることは難しい。近年では専門のデータサイエンティストを雇って分析を依頼する方法も可能となっているが、ドメイン知識の伝達の難しさや、データ取得から解析、実験条件を変えて実験のようなプロセスの効率化のために、別の人を介さず、自身で簡単な分析を行いたい実験化学の研究者も多く、そのような研究者でも簡便に利用可能な統合プラットフォームへのニーズも高まっている。

また、従来のデータ分析や可視化を用いた新たな知識の発見は、職人技に近いものであり、革新的な発見の多くは、研究者が別のものを探索しているときや間違っただけに偶然に発見されたものであることも多い。従来のデータ分析手法に加えて、このような職人技、あるいは偶然の発見過程を、研究者が自在にアクセスし、自身の目的のために再利用することができれば、新技術の開発や科学の発展におおいに貢献することが期待される。物質・材料研究分野でも新規材料の発見や設計に役立つ。しかしながら、そのための方法論は未だ確立されていない。

2. 研究の目的

本研究では第一に、データベース、高度な科学計算・シミュレーション、統計・解析手法、機械学習を含めた様々な機能を、プログラミングの知識を持たない物質・材料科学研究者を含めて幅広いユーザがウェブ上で直接編集操作のみで連携定義し探索的に実行・利用可能な、材料探索・設計のための統合プラットフォームを構築することを目標とする。第二に、構築された統合プラットフォーム上でのユーザの一連の操作履歴を「探索シナリオ」としてモデル化し、再利用可能にする。ユーザの可視化分析環境上での操作履歴から、意味のある固まりをパターンとして取り出し、それを用いて探索過程を支援するための枠組みの構築を目指す。

ウェブ上でのビジュアルな部品として各要素機能を統一的に操作可能なかたちで提供する。部品として提供された解析・可視化機能をユーザがカスタマイズ可能な協調的マルチビュー環境に埋め込み統合することにより、ユーザが簡便に各部品の実行や部品間の連携を定義することを可能にする。協調的マルチビューは、複数のビューを組み合わせる解析対象の可視化を行う手法であり、複数のビューで多面的に解析対象を可視化することが探索的なデータ分析には有効であることが指摘されている。これにより、個々のデータベース、分析ツール、可視化ツール、機械学習ツールなどを探索的可視化のための部品として複数のユーザ間で共有可能になる。また、可視化・分析結果を複合マルチメディア文書として保存・共有することで、それらのツールの利用コンテキストも合わせてユーザに提供することが可能になり、ツールの再利用が促進されると期待される。また、部品に対する基本操作(生成、リサイズ、移動、削除など)を、探索過程をモデル化するための単位として用いることが可能になり、クリックやマウスカーソルの場所などよりも上位のモデル化が可能になる。

また、本研究で目指す探索シナリオのモデル化は、ユーザの一連の操作の列から「一番大きな値をさがす」や「例外を見つける」といった、探索過程の中でユーザの人的活動に着目したユーザの「意図」を推論を可能にすることを目指す。それにより、プリミティブな操作履歴の解析よりも高度な応用が期待できる。

3. 研究の方法

本研究では、材料探索・設計のための探索的可視化分析と探索シナリオの再利用を支援するプラットフォームの構築を行った。具体的には、科学研究費の交付期間内で、下記の2つの課題を達成した。

- (1) 協調的マルチビューに基づく可視化分析プラットフォームの構築と運用
- (2) 探索シナリオのモデル化と再利用の枠組みの構築

上記(1)では、簡単な操作でカスタマイズ可能な強調的マルチビューシステムの基本設計と開発を行った。ここでは、マルチビューの協調動作のためのバックエンドデータベースの統一的な演算フレームワークを確立し、それを基盤としてマルチビューの直接操作とバックエンドデータベースのデータ演算とのマッピング、カスタマイズされたマルチビューの保存と共有のため枠組みの設計と開発に取り組んだ。また、ユーザが自身の持つデータを簡単にシステム上で利用できるように、簡易のファイル管理システムと統合し、アップロードされたファイルに対するアクセス権の管理機能も実装した。様々な環境で利用可能にするため一般的なノートPCや大画面ディスプレイ環境での動作検証・評価を行った。

上記(2)では、(1)で構築されたプラットフォーム上でユーザが実際に行ったデータ探索過程における部品の操作履歴を捕捉し、それをを用いてモデル化を行った。モデル化には、操作の履歴と探索結果のマルチビュー画面状態の関連付けを行い、操作履歴からユーザの「意図」を推論することを可能にすることを試みた。ここでは、機能推論や機械学習など様々な手法を用いて推論を行う試作システムを構築し、その結果得られた仮説を実際の可視化・分析結果との比較、材料研究の専門家へのインタビューなどを行い評価した。さらに、確立された枠組みを用いた応用システムを実装し、プラットフォームに統合した。

4. 研究成果

2019年度は、統合プラットフォームの基盤技術の開発をおもに行い、簡便な操作でカスタマイズ可能な協調マルチビューシステムの基本設計と開発、マルチビューの協調動作のためのバックエンドデータベースの統一的な演算フレームワークの確立を行った。さらに、協調マルチビューの直接操作とバックエンドデータベースのデータ演算とのマッピング、カスタマイズされたマルチビューの保存と共有のための基本機能の設計・開発に取り組んだ。スキャタープロットなどの可視化部品、ヒストグラムやクラスタリングなどを計算する分析部品、回帰や分類などを実行する機械学習部品をインタラクティブな基本部品として実装し、提供した。これにより、ユーザがこれらの部品をオンデマンドで組み合わせ、自分なりの可視化分析を行うことができるようになった。また、探索シナリオのモデル化に関しては、ユーザの操作履歴の補足を行うための手法を開発し、プラットフォームに組み込んだ。さらに、開発されたプラットフォームを触媒データの可視化分析にも応用し、CADS (Catalyst Acquisition by Data Science) として一般公開し、運用した(図1)。 <https://cads.eng.hokudai.ac.jp> のURLでアクセス可能である。

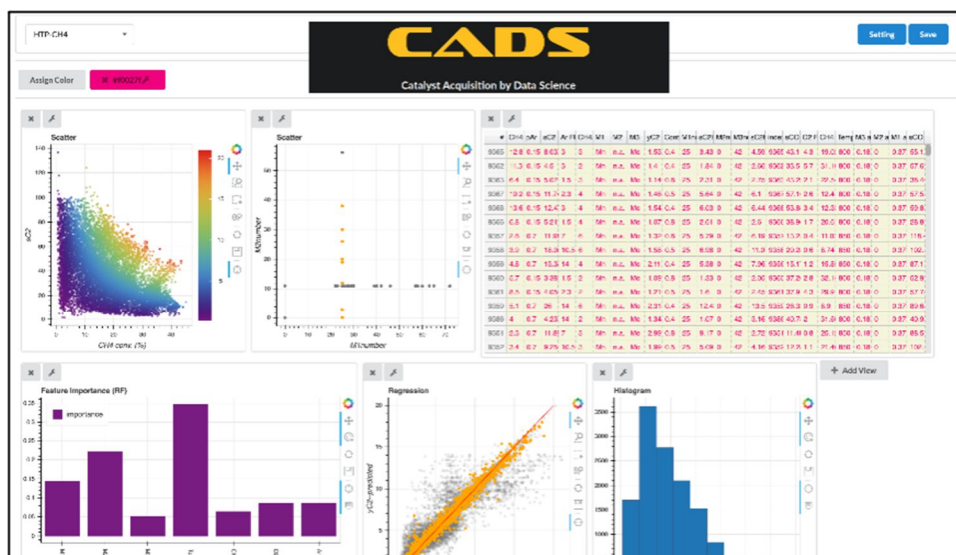


図1 公開されたプラットフォームのスクリーンショット。
協調マルチビューによる解析画面。

2020年度は、前年度に構築された統合プラットフォームをもとに、可視化・分析部品の拡充とその他の機能改善を行った。また、アップロードされたデータファイルのアクセス管理機能を実装した。これにより、ウェブ上でのデータ共有、共有されたデータを用いた可視化・分析結果の共有が可能になった。さらに、このプラットフォームに関する論文として成果発表した [1]。探索シナリオのモデル化に関しては、プラットフォーム上でのユーザの操作履歴からモデル化の方法を検討し、グラフ構造を用いたモデリング手法を構築した。そのモデルを用いてグラフ理論やデータマイニングの手法を用いてユーザの意図を推論する手法の実装と応用に取り組んだ。

2021年度には、可視化プラットフォームの安定運用に継続して取り組み、公開データとして10の触媒データをホストし順調に登録利用者を増やすとともに、国内外の100を超える研究者が非公開のデータをアップロードし研究に利用している。機能面では3次元プロットなど可視化部品のさらなる拡充を行った。ユーザの操作履歴のモデル化と推論システムにおいては、実際の利用者のプラットフォーム上での操作履歴からグラフ構造としてそれをモデル化し、様々なユーザによる操作履歴のモデル化を行った。その操作履歴モデルをグラフ解析し、似ている操作列やそのなかで頻繁に用いられている可視化手法などを解析可能にした。また、この手法を用いて現在のユーザの操作履歴から次にやるべきアクションを決定し、ユーザに推薦するシステムの試作も行った。

[1] Jun Fujima, Yuzuru Tanaka, Itsuki Miyazato, Lauren Takahashi, and Keisuke Takahashi. 2020. Catalyst Acquisition by Data Science (CADS): A Web-Based Catalyst Informatics Platform for Discovering Catalysts. Reaction Chemistry & Engineering 5(5):903-11. doi: 10.1039/D0RE00098A.

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 0件/うちオープンアクセス 0件）

| | |
|---|-------------------------|
| 1. 著者名 Fujima Jun、Tanaka Yuzuru、Miyazato Itsuki、Takahashi Lauren、Takahashi Keisuke | 4. 巻 5 |
| 2. 論文標題 Catalyst Acquisition by Data Science (CADs): a web-based catalyst informatics platform for discovering catalysts | 5. 発行年 2020年 |
| 3. 雑誌名 Reaction Chemistry & Engineering | 6. 最初と最後の頁 903 ~ 911 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1039/D0RE00098A | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

| | |
|--|---------------------------|
| 1. 著者名 Takahashi Keisuke、Fujima Jun、Miyazato Itsuki、Nakanowatari Sunao、Fujiwara Aya、Nguyen Thanh Nhat、Taniike Toshiaki、Takahashi Lauren | 4. 巻 12 |
| 2. 論文標題 Catalysis Gene Expression Profiling: Sequencing and Designing Catalysts | 5. 発行年 2021年 |
| 3. 雑誌名 The Journal of Physical Chemistry Letters | 6. 最初と最後の頁 7335 ~ 7341 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.jpcllett.1c02111 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

〔学会発表〕 計0件

〔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 |
|---------|---------|
|---------|---------|