

研究報告書

「実験・計算データのマイニングと精密結晶構造解析との融合による逆問題可解な材料設計技術の開発」

研究期間：2015年12月～2019年3月

研究者：ダム・ヒョウ・チ

1. 研究のねらい

本研究では、情報科学理論分野で発展目覚ましいデータ駆動型アプローチであるデータマイニング・機械学習と、材料科学分野で広く用いられている理論アプローチである第一原理計算、実験アプローチである精密結晶構造解析とを融合させ、材料科学の複雑データを整理・記述する方法を構築することである。さらに、構築した記述方法を活用して物質の構造物性の各要素からなる関係構造や物理法則を帰納的に解明する手法の開発とそれを用いた新材料の探索・設計手法を開発する。

本研究では、第一原理計算の結果は、物理方程式を通して得られるデータと捉えた上で、実験結果と対比、統合させることが可能である。これにより、理論、計算、実験という手法によらず、全ての結果は広義の「データ集合」として一元的に扱えるようになる。このデータ集合に対し主観や経験に頼ることなくデータ科学各技術を駆使して、設計プロセスにおける物質群の構造物性の各要素からなる関係構造(各変数間の連動性・セマンティック構造)を、帰納的にデータから学習することにより、構造物性に関する物理法則を可視化し、材料設計方針の決定を助ける。また、逆問題である材料設計をデータ科学技術の活用によって実行可能にする手法を提案する。本研究のねらいはデータ駆動型の固体材料設計方法の基盤技術を築くことと、開発した手法を合金磁石材料の研究へ適用することであり、新規磁石材料の設計やその磁気特性を予測する。

本研究は理論手法の確立と新規磁石材料の設計をねらいとして、以下の3点の目標を定める。

① 磁石材料の計算データと実験データからなる集合構築、及び結晶学の知識を活用した材料の記述子の設計と、物質のデータからデータマイニング・機械学習を駆使した記述子の選択・抽出

② 線形及び非線形推定法を活用したスパースモデリングによる、構造物性の各要素からなる関係構造(各変数間の連動性・セマンティック構造)を解析する手法の確立

③ 開発した手法の活用による構造推定の解法の確立

本研究で確立する解析手法は、磁石材料のみならず幅広い固体材料の研究に適用できる。また、実験データと第一原理計算からのモデル化が難しい物質群の現象への解明に大いに貢献できる。さらに、理論計算と実験を方法とデータの両側面での統合的な解析方法にプレイクスルーを促し、ポスト次世代スパコン時代を見据えた解析手法開発の発展に貢献する。

2. 研究成果

(1) 概要

本研究は磁性材料を舞台に、材料研究とデータ科学の融合を円滑・効率的に行うために、下記の基礎技術を開発した。

- 1) 多種元素が混在する複雑材料の構造物性を表現できる原子軌道と配位子場の情報を取り入れた OFM 記述子
 - 2) 開発した記述子 OFM によってデータから説明性の高い構造物性の法則学習技術
 - 3) 開発した記述子 OFM によってデータから推論可能な物質空間の可視化技術
- OFM 記述子は構造物性の研究開発に予測性と説明性を両備された材料記述子であり、材料物性データを開拓する有名なライブラリである Matminer にも実装され、高い評価を得ている。人間の理解を助ける手法の開発を目的に、独自に開発したデータから物質の構造物性に関する記述型データマイニング技術はデータに潜む構造を簡潔で人間の理解が容易な形で引き出し、可視化する汎用性の高いであることを確認した[3,4,5]。さらに、開発してきたデータ科学技術を用いて、
- 4) 元素置換規則によって Nd-Fe-X 化合物の安定構造の探索を行い、第一原理計算を用いて理論評価を行なった結果、開発した手法を用いると、効率的に安定構造を探索できることが確認した。

(2) 詳細

研究テーマ1: 多種元素が混在する複雑材料の構造物性を表現できる原子軌道と配位子場の情報を取り入れた記述子の設計

この研究テーマの実行に当たって、まず磁石化合物データベースの整理・構築を行った。構造記述子として原子軌道場行列を化合物の結晶構造情報から計算を実行し、物性記述子として全磁気モーメント、局所磁気モーメント、生成エネルギーなどの追加物性値計算を行って、データベースを構築した。強力磁石化合物の探索手法を確立するために、次元削減手法を用いて、構造記述子や物性値記述子から化合物の物性値を予測する方法を開発した。カーネル非線形回帰法と全探索による変数選択法の組み合わせを用いて、構造記述子としての原子軌道場行列から化合物の全磁気モーメント、局所磁気モーメント、生成エネルギーを高い精度で予測できることがわかった。

① データベースの構築

658 種類の希土類遷移金属化合物(RT)の構造が Materials Project データベースから獲得した。これらの化合物を構成する遷移金属 T は[Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Tc, Ru, Rh, Pd, Ag, Cd, Rh, Pd, Ag, Cd, Hf, Ta, W, Re, Os, Ir, Pt, Au]から選ばれ、希土類金属 R は[La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu]から選ばれる。物性値に関しては金属サイトの局所磁気モーメントに注目し、密度汎関数法[Jain 28,29]を用いて計算し、データベースに加えた。さらにデータベースに 1510 種類の RTX 化合物の構造や 1311 種類の TTX 化合物や 707 種類の TT 化合物の構造とそれらの密度汎関数法を用いた計算から得られた生成エネルギー値を加えた。

② 構造記述子の設計

中性原子の電子配置情報から、各原子軌道の電子占有状態を辞書とし ($\{s^1, s^2, p^1, p^2, \dots, p^6, d^1, d^2, \dots, d^{10}, f^1, f^2, \dots, f^{14}\}$), one-hot ベクトルを用いて記述し、局所構造は原子軌道行列を用いて記述する. [1]

例) Na 原子の電子配置は $[\text{Ne}]3s^1$ と塩素原子の電子配置は $[\text{Ne}]3s^2 3p^5$ から図1で示したように塩素で形成される八面体サイトに入る Na 原子の局所構造を原子軌道行列を使って記述する事ができる. 行列の各行と列は電子が占有する原子軌道状態に対応し、行列の成分は局所構造における原子軌道間の Voronoi 解析から得られる配位数に対応する.

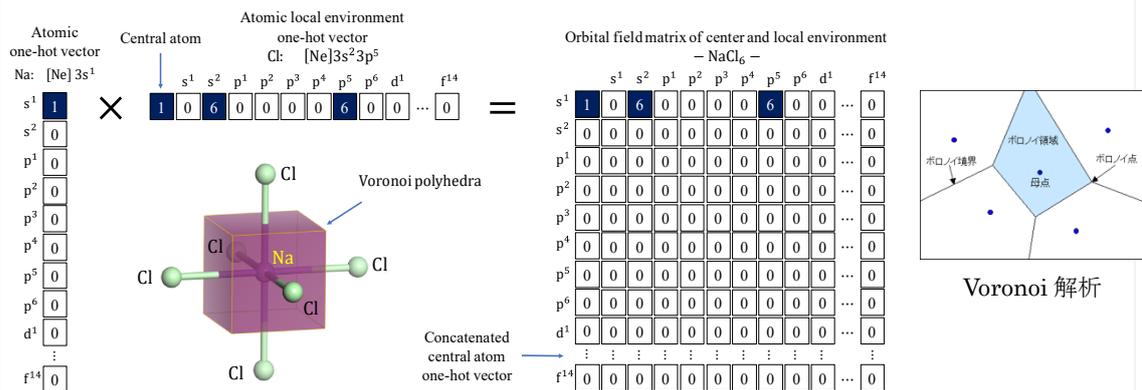


図1. 原子軌道行列による材料の局所構造の記述

③ データベースからの学習

非線形回帰法と変数選択法の組み合わせを用いて、上記の構造記述子としての原子軌道場行列から化合物の局所磁気モーメント、生成エネルギーの予測機を学習した. 高い精度で局所構造から局所磁気モーメントを予測できる事と材料の構造から生成エネルギーを予測できることがわかった. 研究分野で開発される最新構造記述子との比較を行なった. 組成及び結合状態が複雑でない材料系に対しては新規開発した原子軌道場行列記述子の適用で同程度の物性の予測精度を得られる事がわかった. また、組成及び結合状態が複雑材料系に対しては新規開発した原子軌道場行列記述子が原子軌道情報を含む事によって、従来の構造記述子と比べて、圧倒的な高い予測精度をもたらす事がわかった [1,2].

研究テーマ2: 開発した記述子 OFM によってデータから説明性の高い構造物性の法則学習技術や推論可能な物質空間の可視化技術の開発

開発した OFM 構造記述子を基にした幾何学的な距離が類似度評価として活用することが妥当であることが確認できたため、局所距離を活用した多様体学習法 (ISOMAP) を活用して、物質群の構造記述子による多次元空間から次元縮約を行い可視化および整理を行った (図2). 結晶構造のデータは、数値の構造記述子ベクトルの形をしているためユークリッド空間上の点集合と同一視できる. しかし、安定な結晶構造のデータはユークリッド空間の一部 (データ空間) にのみ存在し、多次元空間の多様体を成すと仮定する. 通常、データ間距離はユークリッド距離を用いるが、ISOMAP 法では k 近傍グラフの最短経路長を用いる. k 近傍部分は低次元ユークリッド距離で近似され、局所的な距離をつなぎあわせて大域的

な距離空間を作成することによって、局所座標系を張り合わせて構成されるリーマン多様体の推定が可能となる。ISOMAP法の導入によって、安定な物質群のデータの構造記

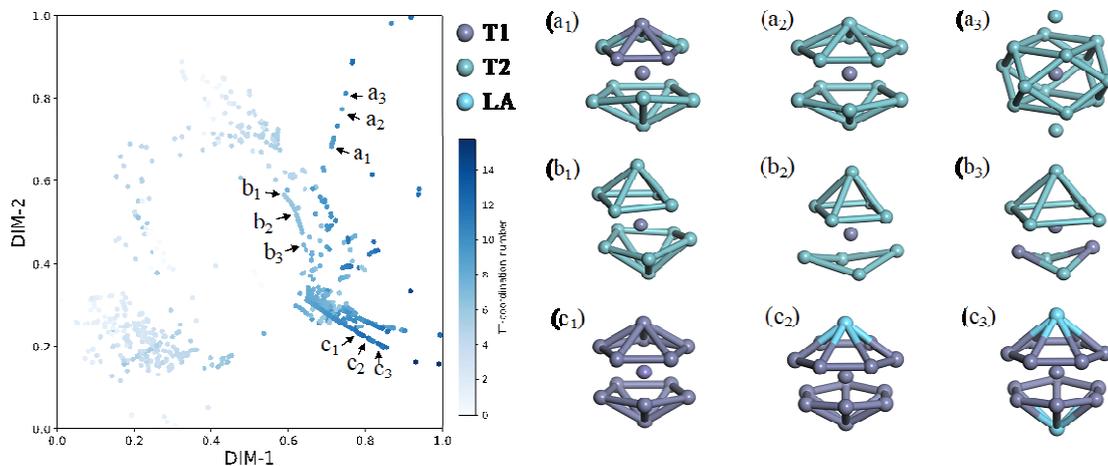


図2. 多様体学習を用いた物質群の局所構造空間の可視化および整理

述子の多次元空間における多様体性質を抽出し、多次元空間から次元縮約を行い可視化に成功した[2].

また、非線形回帰法と変数選択法の組み合わせにアンサンブル学習法の導入により、図6で示したように、合金化合物の T_c の予測精度の向上に成功した。合金化合物の T_c を事例として、データベースに入っている物質群の線形回帰モデル解析に基づいたグルーピング手法を開発した。さらに、得られた各グループの分類結果を分析するために決定木の学習を行い、それぞれのグループの決定する要素を明確にした。この一連の解析を行うことによって物質群の構造物性を実験家が馴染み・イメージしやすい形で自動的かつ系統的に可視化・整理する方法を提案した[3,4].

研究テーマ3: 開発した手法の活用による構造推定の解法の確立

さらに、開発してきたデータ科学技術を用いて、元素置換規則によって Nd-Fe-X 化合物の安定構造の探索を行い、第一原理計算を用いて理論評価を行なった結果、開発した手法を用いると、効率的に安定構造を探索できることが確認した。

参考文献

- 1) T. L. Pham, H. Kino, T. Miyake, and *H. C. Dam, 等計 7 名, *STAM*, **18** (2017).
- 2) T. L. Pham, H. Kino, T. Miyake, and *H. C. Dam, 等計 6 名, *J. Chem. Phys.*, **148**
- 3) H. C. Dam, T. Miyake, *H. Kino, 等計 7 名, *J. Phys. Soc. Jpn.*, **87** (2018).
- 4) D. N. Nguyen and *H. C. Dam, 等計 7 名, *IUCrJ*, **5** (2018)

3. 今後の展開

本研究は、高い精度で物性値を予測できる材料の記述子、材料構造物性機構を解明するために研究者の直観が働きやすい物質空間を可視化する手法、物質の構造物性の物理法則を帰納的に解明する手法一連の要素技術の開発・準備を行ってきた。本研究課題を遂行して、データ科学を取り入れて、材料開発研究を加速化するためには確立された実験・理論計算の知識や技術をデータ科学に合理的に取り入れることによって、各技術が協奏的に融合できる首尾一貫したフレームワークを確立する必要があることが明確になった。また、機械学習は基本的に、結果や導出の過程を人間が判断できない形式になっている。そのために、本研究では記述子作成・データ比較・学習の各段階で、人間の理解を受け入れる形式を取り入れた。これによって得られるメリットは、2つある。まず仮説生成・仮説検証・仮説評価といった、人間の介入が必要な過程において、人による評価・判断が容易なことである。データ科学は相関関係しか導けず、現象を解明する為には適切に人間が介入する必要があるが、合理的に技術を適用することによってそれを支援し、解明の過程を加速することができる。次に、新規物質・新規現象の、従来データが無い領域では、機械学習は内挿問題から外挿問題に移行する。そして外挿問題に対応するデータ科学的手法は現在のところ存在しない。データ科学による新規物質の探索で、人間の評価が可能か否かは本質的な問題である。

それを踏まえて、本研究の今後の展開としては構造物性機構への説明性を備えたデータ駆動型 AI の革新的な融合フレームワークを確立する。特に、応用面での成果だけに注目せず、研究者の閃き・理解・判断の要素も最大限に開拓することによって、人間と人工知能が共創するフレームワークの確立を目指す。最終的には、実験による蓄積された経験・勘、理論計算による経験・論理、データ科学の帰納法的なアプローチによる知識の形式化を、合理的に統合する。

4. 自己評価

研究目的の達成状況:

本研究は理論手法の確立と新規磁石材料の設計をねらいとして、(1)有効な記述子の設計、(2)データからの構造物性の法則性を抽出できる技術、(3)新規材料の構造を推定できる解法の確立、3点の目標を定めた。(1)の研究目的に対しては構造物性の研究開発に予測性と説明性を両備された OFM 記述子を導入でき、材料物性データを開拓する有名なライブラリである Matminer にも実装され、高い評価を得ている。(2)の研究目的に対しては、独自に開発したデータから物質の構造物性に関する記述型データマイニング技術はデータに潜む構造を簡潔で人間の理解が容易な形で引き出し、可視化する汎用性の高いであることを確認した。さらに、(3)の研究目的に対しては、開発してきたデータ科学技術を用いて、元素置換規則によって Nd-Fe-X 化合物の安定構造の探索を行い、第一原理計算を用いて理論評価を行なった結果、開発した手法を用いると、効率的に安定構造を探索できることが確認した。全体として、全ての研究目的を達成できたと言える。

研究の進め方(研究実施体制及び研究費執行状況):

研究実施体制として、研究補助員や学生研究補助を使って研究を加速化できた。また、予定通りの研究執行を行い研究に必要な機材を計画通りに導入した。上述2点を計画した研究計画通りにできたため、研究目的を達成することができた。

研究成果の科学技術及び社会・経済への波及効果(今後の見込みを含む):
本研究の研究成果から材料科学の研究へ説明性を備えたデータ駆動型 AI の革新的な融合フレームワークの確立が期待できる。特に、応用面での成果だけに注目せず、研究者の閃き・理解・判断の要素も最大限に開拓することによって、人間と人工知能が共創するフレームワークの確立することで社会・経済両面に大きな波及効果があると期待できる。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

1. T. L. Pham, H. Kino, T. Miyake, and * <u>H. C. Dam</u> , 等計 7 名, “Machine learning reveals orbital interaction in materials”, <i>STAM</i> , 2017 年, 18 , 1, 756-765
2. T. L. Pham, H. Kino, T. Miyake, and * <u>H. C. Dam</u> , 等計 6 名, “Learning structure-property relationship in crystalline materials: A study of lanthanide transition metal alloys”, <i>J. Chem. Phys.</i> , 2018 年, 148 , 20, 204106
3. D. N. Nguyen and * <u>H. C. Dam</u> , 等計 7 名, “Committee machine that votes for similarity between materials”, <i>IUCrJ</i> , 2018 年, 5 , 6, 830-840
4. <u>H. C. Dam</u> , T. Miyake, *H. Kino, 等計 7 名, “Important descriptors and descriptor groups of Curie temperatures of rare-earth transition-metal binary alloys”, <i>J. Phys. Soc. Jpn.</i> , 2018 年, 87 , 113801, Awarded Editor’s Choice
5. タム・ヒョウ・チ, 水上 卓, “第一原理計算とデータ科学的アプローチ”, 化学と教育, 67 巻 2 号 in press

(2) 特許出願

研究期間累積件数: 0 件(公開前の出願件名については件数のみ記載)
該当なし

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- 1) DAM Hieu-Chi, “Mining magnetic materials data”, 招待講演, 日本磁気学会・第 40 回日本磁気学会学術講演会, 金沢大学(角間キャンパス), 2016 年 9 月 5 日~8 日
- 2) DAM Hieu-Chi, “データマイニングによる材料設計”, 招待講演, 日本セラミックス学会・第 29 回秋季シンポジウム, 広島大学(東広島キャンパス), 2016 年 9 月 7 日~9 日
- 3) DAM Hieu-Chi, “Accelerating materials science research with data mining”, 招待講演, Car-Parrinello Molecular Dynamics CPMD2017 Workshop, Tsukuba, 2017/10/18
- 4) DAM Hieu-Chi, “Determining physical and chemical properties for material systems with data mining”, 招待講演, 255th American Chemistry Society National Meeting, Division of Catalysis Science and Technology, Machine Learning for Catalysis Research, New Orleans, LA, USA, March 18-22, 2018.
- 5) DAM Hieu-Chi, “データ科学手法の活用による物質空間の理解に向けて”, 基調講演, 第 163 回 日本金属学会 2018 年秋期講演, 仙台, 2018 年 9 月 19 日~21 日