

特集 「マテリアルズインフォマティクス」

無機材料へのマテリアルズインフォマティクスの取り組みと課題

Application of Materials Informatics to Inorganic Compounds

森川 幸治

パナソニック株式会社テクノロジーイノベーション本部

Koji Morikawa

Technology Innovation Division, Panasonic Corporation.
morikawa.koji@jp.panasonic.com**Keywords:** materials informatics, inorganic compounds, descriptor.

1. はじめに

新材料開発においては、材料組成の複雑化が進み探索すべき組合せが増加している。このため所望の特性を備えた材料の特定には、膨大な空間からの探索が必要になる。これまでは担当者の経験やシミュレーションによって新たな材料候補を決定し、実際に合成して特性評価を行い、この繰り返しで試行錯誤的に探索を進めることが多かった。この作業には膨大な手間と時間が必要で、材料開発の完了見通しも立てにくいという課題があった。

本稿では、無機材料にマテリアルズインフォマティクスを適用した取り組み事例とその課題などについて紹介する。材料研究では、これまで理論科学・実験科学・計算科学などのアプローチが組み合わされて進められてきた。それに加えてビッグデータ活用の有効性が材料開発にも認知され、データ科学という呼び方で第4のパラダイムとして位置付けられている [Agrawal 16]。

マテリアルズインフォマティクスの実践のためには、解きたい問題の定式化、データの収集、予測器構築、新規材料候補探索などに取り組む必要があり、各ステップに求められる活動を順次紹介していく。

これまで機械学習による材料特性予測、例えば [Ramprasad 17] では、材料の組成式に対して構成元素や結晶構造から記述子と呼ばれる特徴量を定義し、これらの特徴量を入力、目的の特性値を出力として機械学習を適用する。この記述子として構成元素の特性を示す特徴量（原子量や電気陰性度、熱容量など）の平均値や分散値を利用することで、構成元素の数によらず記述子長が一定となり、一般的な機械学習が適用可能となる。

記述子改善に関しては、熱電材料を対象にした取り組みを紹介する。熱電材料の特性に合わせた記述子設計をすることで、材料特性の予測精度が向上できることを示す。

最後にマテリアルズインフォマティクスの課題を説

明する。機械学習が内包する課題として、学習データが存在しない探索領域の予測は不定であるという問題がある。深層学習応用が先行して進んでいる画像認識の分野では、予測したい状況や領域に対して大量の学習データを準備することで性能を維持してきた。一方、材料開発においては、これまでに知られていない材料を探索するため、目標特性を満たす領域のデータは与えられない、という課題があり、本来の機械学習が苦手とする外挿領域での予測性能の向上、という目的にもなる。この課題に対するアプローチについても期待を込めて説明したい。

2. マテリアルズインフォマティクスの流れ

図1に本稿で取り扱う特性予測器による無機新材料の探索を目的としたマテリアルズインフォマティクスの全体の流れを示す。最初に目標となる特性値の設定と予測器に対する入出力定義を行う。目的の材料が満たすべき特性は複数の組合せになる場合も多く、この優先順位や重み付けも含めて決定しておく。次に続く処理としては、(a) データの収集、(b) 予測器構築、(c) 新規材料候補探索、(d) 合成実験と評価による実証の各ステップによって進められる。

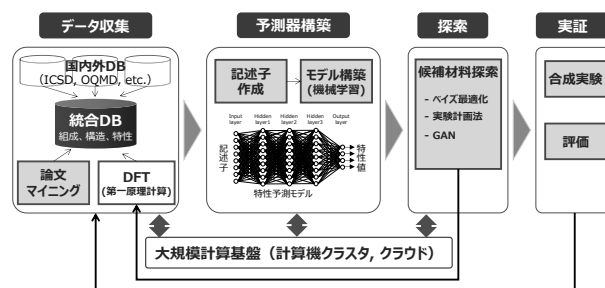


図1 マテリアルズインフォマティクス全体の流れ

2.1 データ収集

材料組成式と、その組成式に対応した特性値データとなるべく幅広く収集するステップである。自分達で実施した実験データ以外にも、論文からの情報抽出、第一原理計算などによる特性計算、公開データベースの参照などがある。

2.2 記述子の作成

材料を表現する組成式は文字列として表現され、組成式だけでは材料特性の予測に貢献する情報はあまり多くない。このため機械学習適用には、材料特性を含む数値列に変換する必要がある。この変換した数値列を記述子(descriptor)と呼び、この記述子が機械学習への入力となる。記述子は、ターゲットとなる材料特性に関する情報が含まれている必要があり、記述子の設計次第で材料の予測性能が大きく変化する。4章で記述子設計の工夫の事例について紹介する。

2.3 予測器の構築

予測器の構築では、収集したデータセットをもとに特性予測器の学習を行う。特性予測器の学習には、各種回帰手法が適用可能である。深層学習、ランダムフォレスト、ブースティングなどの手法が用いられる。データ数の多寡と、記述子空間での問題構造によって適切な手法を選択する必要がある。

2.4 材料の探索

構築した予測器を元に、材料候補群に対する特性値予測を行うことで材料の探索を行う。最初に探索したい新材料候補の組合せからなる材料候補の探索空間を設定する。一般的に予測器の学習には時間を要するが、予測器構築後の予測値の計算には時間がかからないため、探索空間が大きくない場合には、すべての候補材料に対して特性予測値を算出し、最適な特性を与える候補材料を求める。

一方、候補材料の探索空間の全候補の特性値予測が実行時間内では終了できない場合には、バイズ最適化などの探索手法が用いられる [Seko 14]。この探索手法を組み合わせることで、全探索点の特性値予測を行わなくても、効率的にターゲット特性をもつ候補材料を特定できる。

2.5 実験による実証

材料候補が特定された後に、その材料特性を実際に確認するのが本ステップである。本特性評価で実際の材料が予測どおりの特性を示しているかを確認できる。ここで目的となる特性が達成されていれば材料探索は終了となる。もしも目的を達成しない場合には、失敗データとして学習データに追加され、次の予測器構築、材料探索に進み、これらの繰返しで予測の精度を向上させる。

3. データ収集

材料データは自己で保有しているデータを基本に、さまざまな情報源からのデータを集約することで大きなデータを対象にした解析が可能になる。機械学習では、一般に多くのデータを用いて学習したほうが予測精度が良いため、データ収集もマテリアルズインフォマティクスの主要な活動になる。図2にデータ収集方法の一覧を示す。

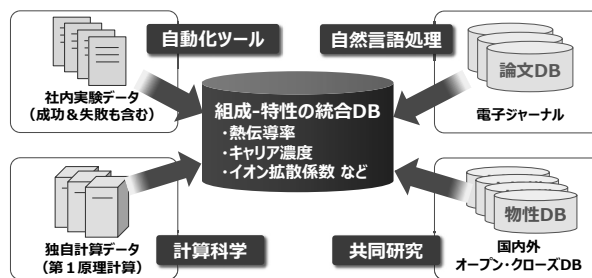


図2 さまざまなデータ収集方法

3.1 公開データベース

一般向けに公開されているデータベースは入手が容易であり活用しやすい。無機結晶構造データベースとしては AtomWork [Xu 11] や ICSD (Inorganic Crystal Structure Database) [Allmann 07] などがあげられる。これらは論文などから収集した結晶構造や各種特性データが収録されている。収録データ数(結晶構造ベース)は AtomWork で8万件以上、ICSD で19万件以上である。また計算データベースとしては、Materials Project [Jain 13] や OQMD [Saal 13] があげられる。第一原理計算は電子状態を量子力学の方程式に従って数値計算により求める手法であり、結晶構造が明らかな材料に対して計算される。第一原理計算では、ターゲットの材料特性ごとに計算が行われ、データベースごとに収録される特性が異なっていることがある。収録データ数(組成式ベース)は、Materials Project では約13万件、OQMD では、約56万件である。データの使用方法に関しては、無償のものと有償のものがあり、ユーザ登録が必要なものなど、使用条件・契約条件の確認には注意が必要である。ここにあげたデータベースはデータ登録を継続的に行っており、収録データ数は日々増加しているため適宜再確認するとよい。

3.2 計算によるデータ生成

公開データベースに収録されていない材料やその特性は、第一原理計算を自前の計算によって収集することも可能である。例えば、社内で開発した未発表の材料組成や、まだ合成を行っていない材料候補については、データベースに登録されていないので自前で計算する必要がある。また、オープンデータベースでは収録されている

特性は基本的なものが多く、独自のターゲット特性を設定した場合には自分で特性値を計算する必要がある。

第一原理計算の実行に関しては、設定したセル内の元素数が多い場合や、計算したい特性が複雑なモデルに基づく場合には、計算時間が数時間、パラメータによっては数日間必要な場合もある。これらの計算を多くの材料に対して実施するには、大量の計算資源が必要になる。計算機資源の確保に関しては、(a) PCを購入して自社内に計算クラスターを構築する以外にも、(b) クラウド上の計算資源の利用や、(b) 産業技術総合研究所や理研などの研究機関が構築しているスーパーコンピュータの利用申請によっても、活用可能である。著者らは計算目的に応じて、上記手段を組み合わせて計算を実行してデータ蓄積を推進している。

3.3 論文からのデータ抽出

論文誌からのデータ収集も、データの入手方法として有効である。ほぼすべての論文誌は電子的にアクセスが可能になってきており、一部の電子ジャーナルの出版社はプログラムによる広範囲な検索を可能にするAPIを公開している。例えば、Springer社やElsevier社はAPIを契約の締結によって利用可能になる。これによって、特定の組成や応用分野の論文を広範囲に検索でき、また目的に合った論文に関してはダウンロードも可能である。

ダウンロードした論文は、テキストマイニング技術と連動させることで、組成式の特定制や材料特性の抽出も実施可能である。ただし、材料特性はグラフや写真として掲載されている場合もあるため、これらの自動認識は精度との認識精度との兼ね合いから完全な自動化はまだ難しい段階にある。

また材料の合成プロセス条件も材料特性予測に関して有効な情報になり得る。合成の温度など検出しやすい条件に絞ったマイニング事例はできつつあるが[Kim 17]、合成のプロセス全体を抽出するには、文章として記述されている合成方法の記載を読み解く必要がある。合成方法の記載は著者によって異なること、条件のうち著者が当然とみなし得る部分の省略も多いことから、文章からの実用的な情報抽出はまだ課題が残っている。

3.4 実験からのデータ収集

論文には本当に探索したい未知の材料候補の特性は記載されていないこと、また、第一原理計算では、材料を構成するマクロな状態（微量な添加物や、相や界面の影響など）をシミュレーションすることは困難である。このため必要な材料特性を把握するには、合成を実際に行ったうえでの特性評価によるデータ収集が必要になる場合もある。

これまで材料研究者は、ターゲットとなる材料特性の値が目標をクリアしたかどうかを主に注目していたが、

マテリアルズインフォマティクスの観点からは、特性の良くないデータ、いわゆる失敗データや、材料合成の各ステップの詳細（例えば温度変化や、混練条件、圧力条件など）も、記録しておくことが望ましい。論文には失敗データは掲載されていないことが多いので、自部署で実施した実験データは機械学習にとって貴重な情報源となり得る。

著者らは、実験データを円滑に収集するために、合成装置や評価装置をネットワークで接続し、各装置の合成条件や評価条件を統一的に扱えるID管理にも配慮した、データを自動的に回収する枠組みを構築している。これにより実験者の負担が少なく、かつ、人為的なミスが少ないデータ蓄積システムを運用している。

3.5 データベースへの統合

これまでに述べた各データソースから収集したデータはそれぞれデータフォーマットが異なるため、フォーマット変換により統一的に扱える必要がある。データ標準化を目指していくつかのフォーマットが提案されているが（例えば、NIST Material Data Curation System [Dima 16] など）、データ管理および研究者の参照用を想定した設計になっており、マテリアルズインフォマティクスの運用に適した標準的なフォーマットはまだない。

4. 記述子変更による予測精度の改善

本章では、著者らの取組みの例として記述子変更による予測精度の改善[Hagawa 18]について紹介する。記述子は、材料特徴を表現した変数で、材料の組成式や結晶構造と、元素の特性などから作成される。記述子は数十から数百の数値列として表現され、この数値列が機械学習の入力となる。機械学習の入力変数として直接使われるため、この記述子の設計によって予測性能は大きく異なる。このため特性予測に効果的な記述子の設計は、マテリアルズインフォマティクスの主要テーマの一つになる。

ここでは熱電材料をターゲット事例として取り上げた。熱電材料とは、熱および電気エネルギーを相互に変換する特性をもつ無機材料であり、熱発電やペルチェ冷却などに使用される。

熱電材料の変換性能は ZT という特性で表現され、より高い ZT の材料の発見が求められている。 ZT は複数の基礎特性の組合せで、 $ZT = (PF/K)T$ と表現される。ここで、 PF はパワーファクタ、 K は熱伝導率、という材料特性、 T は温度である。 PF はさらに $PF = S^2\sigma$ の関係があり、 S はゼーベック係数、 σ は電気伝導率、という特性で表現される。このようにターゲットの材料特性は複数の別の材料特性で表現可能なことも多く、どの特性を機械学習の予測ターゲットにするかは、データの入手性などによって選択可能である。

熱電材料は一般に母物質に添加物を加えることで合成され、この添加物次第で特性が大きく変化することが特徴である。例えば材料 CaYb0.05Mn0.95O3 は、母物質 CaMnO3 における Mn の一部を添加物 Yb0.05 で置換することにより合成され、Yb0.05 という微量の元素によって特性が左右される。表 1 に添加物と熱電特性（パワーファクタ）の関係を例示する。添加物の種類や量を変えることで特性値が大きく変動していることがわかる。

表 1 熱電特性に対する添加物の影響の例

母物質	添加物	組成式	Power Factor [$\mu\text{W}/\text{cmK}^2$]
CaMnO3	なし	CaMnO3	0.42
	Ru0.02	CaRu0.02Mn0.98O3	1.84
	Yb0.05	CaYb0.05Mn0.95O3	3.21
	Yb0.1	CaYb0.1Mn0.9O3	2.11

通常の記述子は、組成式に含まれる各元素の加重平均により算出されることが多く、添加物の影響が十分に考慮できず熱電特性の予測が難しい原因となっていた。これに対して構成元素情報を均等に扱う記述子を並置する手法が提案されている [Furmanchuk 18]。組成式内の添加物は、その比率が 0.01 や 0.05 など割合が小さいことが課題であるとし、加重平均ではなく構成元素の種類のみを記述子としている。しかし構成元素比率の影響は考慮できなかった。

そこで著者らは構成元素の比率も表現できる記述子を設計して学習に適用した。熱電材料の合成過程では母物質に添加物を加えることに着目し、母物質と添加物を区別した記述子表現を採用した。熱電材料の公開データベースを用いた比較実験の結果、この記述子表現によって、特に添加物の影響が大きい特性である電気伝導率の予測誤差の低減が確認された。

4.1 使用した DB と記述子の構成方法

熱電材料の特性値をまとめた公開データベース UCSB-MRL thermoelectric database (UCSB) [Gaultois 13] を使用した。本データベースには、261 種類の組成式に対する材料特性（電気伝導率または電気抵抗率、ゼーベック係数、パワーファクタ、熱伝導率、 ZT ）と評価温度条件が記載されている。ここから電気伝導率とゼーベック係数をターゲット特性として設定した。

記述子の生成方法は、以下の 3 ステップで構成した。

Step 1：組成式から母物質と記述子を分離

UCSB データベースに収録の組成式は、母物質情報と添加物が分離されておらず、基準を設けて組成式から母物質と添加物を分離した。

母物質は組成式の係数がすべて整数となるもの、添加物は組成式と対応する母物質との差分であると定義した。例えば組成式 CaMn0.9Ru0.1O3 は、母物質 CaMnO3 のうち、Mn の一部を Ru に置き換えたと考え

え、添加物は Ru0.1 と Mn-0.1 と設定した。このように置き換えられた物質は、減少分の係数にマイナスを付与し、添加物の一つとした。

母物質を特定後、その母物質が Inorganic Crystal Structure Database (ICSD) [Allmann 07] に存在する場合のみを今回の評価対象のデータとして使用した。

Step 2：母物質と添加物に対する各記述子の算出

母物質と添加物のそれぞれから記述子を導出した。元素由来の記述子導出には XenonPy [Yoshida 18a] を用いた。XenonPy は組成式に含まれる各元素がもつさまざまな既知物理量から 58 種類の記述子を生成できる。

母物質由来の記述子として、母物質組成式の係数によって加重した各元素の特徵量の最大・最小・合計・平均・標準偏差値の 5 種類を使用した。それぞれ記述子は 58 個ずつ出力されるので母物質由来の記述子は $58 \times 5 = 290$ 個の数値列になる。

また添加物は、各添加物の元素由来特徴量および係数値の全組成に対する割合値を記述子として利用する。例えば、組成式 CaMn0.9Ru0.1O3 における添加物 Ru0.1 により導出される記述子は、各元素由来特徴量の値および係数の総和に対する割合値 0.02 となる。

複数の添加物に対応するため、最大 5 個分の添加物の記載領域を確保した。係数情報と合わせて添加物由来の記述子は全体では $59 \times 5 = 295$ 個の数値列になった。

Step 3：各物質の記述子を結合

提案手法においては、母物質由来の記述子 290 個と添加物由来の記述子 295 個を足し合わせて 585 個の記述子となった。なお、記述子は提案手法の係数情報以外すべて平均 0、分散 1 に規格化した。

また、UCSB のデータベースで、組成式情報とともに収録されていた温度情報も記述子として追加した。この記述子によって、添加物の種類の影響と添加物の割合の変動の影響の両方を考慮した熱電材料の特性予測が可能になる。

4.2 比較結果

比較実験に使用したデータは UCSB から抽出した。収録データのうち、母物質と添加物が分離できた 456 データを学習データとして使用した。データには組成式 110 種類、母物質 46 種類が含まれていた。データのうち組成式や実験条件がすべて同じデータが複数個含まれる場合は特性の最良値を採用した。

予測対象の特性値は、添加物によって値が変動することが知られている電気伝導率 σ とゼーベック係数 S の二つを選択した。

予測手法には全結合ニューラルネットワークを使用した。ネットワークは中間層の素子数が記述子数 $\times 2$ の 4

層、活性化関数は ReLU、中間層に 50%の割合でドロップアウトを適用した。学習率は 0.01 から 0.001 まで線形に変化させ、30 epoch の学習を行った。10 fold のクロスバリデーションを実施し、fold ごとに用いるデータは、従来手法・提案手法で同一とした。また、母物質の設定に関して同じ母物質が学習データとテストデータの両方に含まれないよう母物質単位で学習・テストデータを分割した。これにより、予測器は学習時に与えられていない母物質に対する予測を行うことになる。

評価は、電気伝導率とゼーベック係数の予測誤差 (RMSE)、および、決定係数 R^2 を比較した。従来法として [Furmanchuk 18] を対象とし、組成式そのもの由来の記述子 290 個と、係数を考慮しない記述子 290 個を合わせた 580 個の記述子となった。また、予測手法については、上記と同様の構成のニューラルネットワークを使用した。

表 2 に従来法と提案手法における RMSE 値および R^2 の比較結果を示す。表 2 より電気伝導率は提案手法が、ゼーベック係数は従来手法が高い予測精度を示し、特性値ごとに有効な記述子が分かれる結果となった。

この原因を検討するため、特性ごとの同一母物質内の元データの標準偏差を比較した。表 3 に正解特性値を平均 0、分散 1 に正規化した際の温度帯ごとの標準偏差の平均値を示す。各セル内では、同一母物質・温度条件であるため、添加物の違いに起因した値の標準偏差を示している。表 3 に示す値が高いことは、添加物の違いによる特性値の変動が強いことを示している。電気伝導率はゼーベック係数に比べ標準偏差が大きく、添加物による特性値への影響が強いといえる。これより、提案手法は添加物による変化が大きい特性値の予測に対して有効であるといえる。一方で、ゼーベック係数に関しては添加物の影響が少ないため性能の改善は見られなかった。このように、記述子の設計はターゲットの特性とともに適切に設定される必要があることがわかる。

表 2 予測性能

	Electrical Conductivity [log10(S/mm)]		Seebeck Coefficient [μ V/K]	
	RMSE	R^2	RMSE	R^2
Compared	0.91	0.07	176	0.19
Proposed	0.84	0.12	184	0.10

表 3 元データの標準偏差の比較

	300K	400K	700K	1000K
Electrical Conductivity	0.52	0.53	0.52	0.54
Seebeck Coefficient	0.44	0.40	0.32	0.37

5. 未知材料の探索を目指して

本章では、マテリアルズインフォマティクス適用における二つの大きな課題について述べる。一つは、データの数量が少ない場合にどのように材料探索を進めるか、もう一つは探索したい材料系のデータが存在しない場合にどうするかである。

3章で述べたように、各種の公開データベースでは大量のデータが公開されているが、必ずしも必要な特性のデータが豊富にあるわけではない、という課題がある。例えば熱電の ZT やゼーベック係数などは、熱電材料固有の特性であり、関連データを大量に収集するのは容易ではない。

これに対して転移学習の適用が考えられる。転移学習は、あるタスクの学習モデルを別のタスクに流用することを目的とする方法論の総称であり [吉田 18b], ターゲット特性の材料データが少ない場合に、データを多く準備できる類似の特性に対して作成した予測器を転用することが考えられる。

二つ目は、ほしい領域のデータが必ずしも存在するわけではない、という課題である。図 3 にほしい領域と既存領域の関係を示す。既存データは、既存特性限界以下にしか存在せず、本来ターゲットにしている既存特性以上の特性値を示す領域の予測性能は十分ではない、といういわゆる外挿問題に対応する。

これらに対して著者らは二つのアプローチの可能性があると考えている。図 3 には繰り返し計算による材料探索

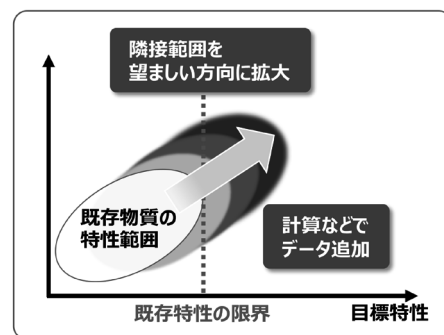


図 3 繰り返し計算による材料探索の概念図

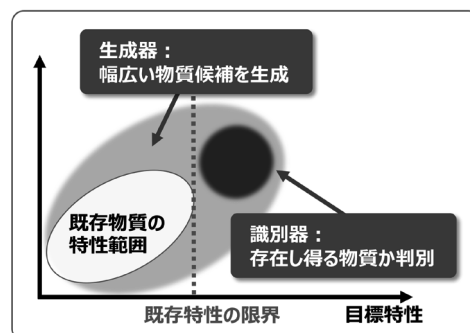


図 4 生成モデルによる材料探索の概念図

の概念図を示す。有機材料分野に対する外挿問題への対応を目指した取組みとしてはベイズに基づく分子設計の取組みがある [Ikebata 17]。分子材料候補を ASCII の文字列によって表現する SMILES 表記法を用いて、そのうえで HOMO-LUMO gap と呼ばれる特性の最適値を探索するものである。また図 4 には、生成モデル (GAN) による対応方法もあると考えられる。生成器と識別器の競合学習によって、まだデータの無い部分にも候補物質を生成し得る可能性を考えている。例えば有機分野では、生成モデルによる探索事例が報告されている [De Cao 18]。

6. おわりに

本稿では無機材料に対するマテリアルズインフォマティクスの適用事例とその課題、取り得る対応策などについて紹介した。新材料発見には、データ面の工夫、アルゴリズム面の工夫を適切に組み合わせて材料探索のフレームを構築する必要があり、予測モデル構築や実験・評価においては、理論や実験の材料研究者との連携も必須になる。マテリアルズインフォマティクスの適用によって、新材料が発見された事例も徐々に発表されつつあり、今後ますます取組みが増え、新材料の研究開発の効率化に貢献を果たしていくと考えている。

今後は、マテリアルズインフォマティクス適用の大きな課題としてあげた、データ不足への対応とアルゴリズムによる支援による総合的な解決を推進していきたい。これにより多様な材料探索への展開や合成プロセスも含めた材料開発、材料設計にまで応用範囲を拡大し、開発効率の向上が実現されると考えている。

◇ 参考文献 ◇

- [Agrawal 16] Agrawal, A. and Choudhary, A.: Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in Materials Science, *APL Materials*, Vol. 4, 053208 (2016)
- [Allmann 07] Allmann, R. and Hinek, R.: The introduction of structure types into the inorganic crystal structure database ICSD, *Acta Crystallographica Section A*, Vol. 63, Part 5, pp. 412-417 (2007)
- [知京 17] 知京豊裕: 「マテリアルズインフォマティクスの現状と課題」—海外の動向と日本の挑戦—, 情報知識学会誌, Vol. 27, No. 4, pp. 297-304 (2017)
- [De Cao 18] De Cao, N. and Kipf, T.: MolGAN: An implicit generative model for small molecular graphs, *ICML Deep Generative Models Workshop* (2018)
- [Dima 16] Dima, A., Bhaskarla, S., Becker, C., Brady, M., Campbell, C., Dessauw, P., Hanisch, R., Kattner, U., Kroenlein, K., Newrock, M., Peskin, A., Plante, R., Li, S. Y., Rigodiat P., Amaral, G. S., Zachary, T., Schmitt, X., Warren, J. and Youssef, S.: Informatics infrastructure for the materials genome initiative, *JOM*, Vol. 68, No. 8, pp. 2053-2064 (2016)
- [Furmanchuk 18] A. Furmanchuk, et al.: Prediction of seebeck coefficient for compounds without restriction to fixed stoichiometry: A machine learning approach, *J. Comput. Chem.*, Vol. 39, No. 4, pp. 191-202 (2018)
- [Gaultois 13] Gaultois, M., Sparks, T., Borg, C., Seshadri, R., Bonificio, W. and Clarke, D.: Data-driven review of thermoelectric materials: Performance and resource considerations, *Chem. Mater.*, Vol. 25, No. 15, pp. 2911-2920 (2013)
- [Hagawa 18] Hagawa, R., Tamaki, H. and Morikawa, K.: Descriptor for separating base-material and additive in machine learning of thermoelectric material property prediction, *NeurIPS 2018 Workshop “Machine Learning for Molecules and Materials”* (2018)
- [Ikebata 17] Ikebata, H., Hongo, K., Isomura, T., Maezono, R. and Yoshida, R.: Bayesian molecular design with a chemical language model, *J. Comput. Aided Mol. Des.*, Vol. 31, No. 4, pp. 379-391 (2017)
- [Jain 13] Jain, A., Ong, S.P., Hautier, G., Chen, W., Richards, W.D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., Persson, K.A.: The materials project: A materials genome approach to accelerating materials innovation, *APL Materials*, Vol. 1, No. 1, 011002 (2013)
- [Kim 17] Kim, E., Huang, K., Saunders, A., McCallum, A., Ceder, G. and Olivetti, E.: Materials synthesis insights from scientific literature via text extraction and machine learning, *Chemistry of Materials*, Vol. 29, No. 21, pp. 9436-9444 (2017)
- [Ramprasad 17] Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakthodi, A. and Kim, C.: Machine learning in materials informatics: Recent applications and prospects, *npj Comput. Mater.*, Vol. 3, Article number 54 (2017)
- [Saal 13] Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. and Wolverton, C.: Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD), *JOM*, Vol. 65, No.11, pp. 1501-1509 (2013)
- [Seko 14] Seko, A., Maekawa, T., Tsuda, K. and Tanaka, I.: Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids, *Phys. Rev. B*, Vol. 89, p. 054303 (2014)
- [Seko 17] Seko, A., Hayashi, H., Nakayama, K., Takahashi, A. and Tanaka, I.: Representation of compounds for machine-learning prediction of physical properties, *Phys. Rev. B*, Vol. 95, No. 14, p. 144110 (2017)
- [Xu 11] Xu, Y., Yamazaki, M. and Villars, P.: Inorganic materials database for exploring the nature of material, *Jpn. J. Appl. Phys.*, Vol. 50, 11RH02 (2011)
- [Yoshida 18a] XenonPy (online) Available at: <http://xenonpy.readthedocs.io/en/latest/> (Accessed 8 March 2019)
- [吉田 18b] 吉田 亮, 山田寛尚, Chang, L., Guo, Z. and Wu, S.: マテリアルズ・インフォマティクス概説, 日本化学会情報化学部会誌, Vol. 36, No.1 (2018)

2019年3月6日 受理

著者紹介



森川 幸治 (正会員)

1996年名古屋大学大学院工学研究科博士課程修了。日本学術振興会特別研究員を経て1997年松下電器産業株式会社入社。現在、パナソニック株式会社テクノロジイノベーション本部副主幹研究員。材料インフォマティクス、脳機能計測、生体品号解析、生体信号センサ開発、認知モデルなどの研究開発に従事。博士 (工学)。