

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7339924号
(P7339924)

(45)発行日 令和5年9月6日(2023.9.6)

(24)登録日 令和5年8月29日(2023.8.29)

(51)国際特許分類 F I
G 0 6 F 30/10 (2020.01) G 0 6 F 30/10
G 0 6 F 30/27 (2020.01) G 0 6 F 30/27

請求項の数 10 (全16頁)

(21)出願番号	特願2020-79793(P2020-79793)	(73)特許権者	000005108 株式会社日立製作所 東京都千代田区丸の内一丁目6番6号
(22)出願日	令和2年4月28日(2020.4.28)	(74)代理人	110001678 藤央弁理士法人
(65)公開番号	特開2021-174403(P2021-174403 A)	(72)発明者	金澤 拓也 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
(43)公開日	令和3年11月1日(2021.11.1)	(72)発明者	森田 秀和 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
審査請求日	令和4年11月14日(2022.11.14)	(72)発明者	浅原 彰規 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内
		(72)発明者	林 貴之 最終頁に続く

(54)【発明の名称】 材料の特性値を推定するシステム

(57)【特許請求の範囲】

【請求項1】

材料の特性値を推定するシステムであって、
1以上のプロセッサと、
1以上の記憶装置と、を含み、
前記1以上の記憶装置は、第1学習モデルと第2学習モデルとを格納し、
前記1以上のプロセッサは、
複数の材料それぞれの、所定数の要素からなる低次元記述子を生成し、
前記第1学習モデルによって、前記低次元記述子から前記複数の材料それぞれの特性値を推定し、
前記特性値に基づいて、前記複数の材料から一部の材料を選択し、
前記一部の材料それぞれの、要素数が前記所定数より多い高次元記述子を生成し、
前記第2学習モデルによって、前記高次元記述子から前記一部の材料それぞれの特性値を推定する、システム。

10

【請求項2】

請求項1に記載のシステムであって、
前記低次元記述子の全ての要素の種類は、前記高次元記述子の要素の種類に含まれる、システム。

【請求項3】

請求項1に記載のシステムであって、

20

前記 1 以上のプロセッサは、

前記高次元記述子の要素の種類の特性値推定に対する重要度を予め設定された方法により計算し、

前記重要度に基づいて、前記高次元記述子の要素の種類から前記低次元記述子の要素の種類を選択する、システム。

【請求項 4】

請求項 1 に記載のシステムであって、

前記 1 以上の記憶装置は、複数の材料の特性測定値を含む測定データを格納し、

前記 1 以上のプロセッサは、

前記測定データから選択した第 1 学習データを使用して、前記第 1 学習モデルの学習を実行し、

前記測定データから選択した前記第 1 学習データより多い第 2 学習データを使用して、前記第 2 学習モデルの学習を実行する、システム。

【請求項 5】

請求項 1 に記載のシステムであって、

前記 1 以上のプロセッサは、前記一部の材料及び前記第 2 学習モデルが推定した特性値をモニタに出力する、システム。

【請求項 6】

システムが材料の特性値を推定する方法であって、

前記システムは、

1 以上のプロセッサと、

1 以上の記憶装置と、を含み、

前記 1 以上の記憶装置は、第 1 学習モデルと第 2 学習モデルとを格納し、

前記方法は、

前記 1 以上のプロセッサが、複数の材料それぞれの、所定数の要素からなる低次元記述子を生成し、

前記 1 以上のプロセッサが、前記第 1 学習モデルによって、前記低次元記述子から前記複数の材料それぞれの特性値を推定し、

前記 1 以上のプロセッサが、前記特性値に基づいて、前記複数の材料から一部の材料を選択し、

前記 1 以上のプロセッサが、前記一部の材料それぞれの、要素数が前記所定数より多い高次元記述子を生成し、

前記 1 以上のプロセッサが、前記第 2 学習モデルによって、前記高次元記述子から前記一部の材料それぞれの特性値を推定する、ことを含む方法。

【請求項 7】

請求項 6 に記載の方法であって、

前記低次元記述子の全ての要素の種類は、前記高次元記述子の要素の種類に含まれる、方法。

【請求項 8】

請求項 6 に記載の方法であって、

前記 1 以上のプロセッサが、前記高次元記述子の要素の種類の特性値推定に対する重要度を予め設定された方法により計算し、

前記 1 以上のプロセッサが、前記重要度に基づいて、前記高次元記述子の要素の種類から前記低次元記述子の要素の種類を選択する、ことを含む方法。

【請求項 9】

請求項 6 に記載の方法であって、

前記 1 以上の記憶装置は、複数の材料の特性測定値を含む測定データを格納し、

前記方法は、

前記 1 以上のプロセッサが、前記測定データから選択した第 1 学習データを使用して、前記第 1 学習モデルの学習を実行し、

10

20

30

40

50

前記 1 以上のプロセッサが、前記測定データから選択した前記第 1 学習データより多い第 2 学習データを使用して、前記第 2 学習モデルの学習を実行する、ことを含む方法。

【請求項 10】

請求項 6 に記載の方法であって、

前記 1 以上のプロセッサが、前記一部の材料及び前記第 2 学習モデルが推定した特性値をモニタに出力する、方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、材料の特性値を推定するシステムに関する。

10

【背景技術】

【0002】

一般に、新材料探索タスクに対してバーチャルスクリーニングの手法が利用される。この手法は、既知の材料（化合物）のデータに機械学習モデルを適用し、材料の特性値を推定する特性値推定モデルを構成する。より具体的には、材料の化学構造式から、材料の特徴を多変量で表現した記述子を生成する。さらに、当該記述子と特性値との間の関係を学習することで、特性値推定モデルを構成する。特性値推定モデルは、入力された記述子に対して特性値を推定する。記述子は、複数の要素（特徴量）で構成されており、各要素が対応する特徴、例えば、分子量や元素混合比を表す。

【0003】

20

次に、バーチャルスクリーニングの手法は、特性値が未知の多くの化合物の化学構造式から記述子を生成し、それら記述子に対して、上記特性値推定モデルを適用する。こうして計算された特性値をもとにスクリーニングを行い、閾値を超える特性値を持つと期待される化学構造式を、実験又はシミュレーションの候補となる有望化合物として提示する。

【0004】

ユーザは候補の中から選択した材料の実験やシミュレーションを行い、それら材料を評価する。このように、バーチャルスクリーニングによって、材料の必要な実験やシミュレーションの回数を低減し、効率的に所望の特性値を有する材料を得ることができる。

【0005】

材料の記述子を生成する手法は、例えば、例えば非特許文献 1 又は非特許文献 2 に開示されている。これらは、数千～数万×の記述子要素（特徴量）から、無機化学分野において、推定に有用な少数の記述子要素の組み合わせからなる記述子を発見するための手法を開示する。

30

【先行技術文献】

【非特許文献】

【0006】

【文献】L. M. Ghiringhelli et al., "Big Data of Materials Science: Critical Role of the Descriptor", Phys. Rev. Lett. 114, 105503 (2015)

R. Ouyang et al., "SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates", Phys. Rev. Materials 2, 083802 (2018)

40

【発明の概要】

【発明が解決しようとする課題】

【0007】

所望の特性値を有する化合物を見つけるため、バーチャルスクリーニングは、膨大な数の候補化合物の記述子を生成し、特性値推定モデルによって生成した記述子の特性値を推定する。また、化学構造式の特徴を高精度に表す記述子の次元数（要素数）は非常に大きく、一般に、1000～4000程度である。

【0008】

そのため、膨大な数の候補化合物の全ての記述子を計算することは、非常に多くの計算

50

リソース及び時間を必要とする。また、候補化合物の多くは、所望の特性値を得ることができない不適な化合物であり、それらの記述子を計算することは無駄な計算リソース及び時間を消費することになる。

【0009】

従って、より効率的に所望の特性値を有すると期待できる有望な材料を選択することができる技術が望まれる。

【課題を解決するための手段】

【0010】

本発明の一態様は、材料の特性値を推定するシステムであって、1以上のプロセッサと、1以上の記憶装置と、を含む。前記1以上の記憶装置は、第1学習モデルと第2学習モデルとを格納する。前記1以上のプロセッサは、複数の材料それぞれの、所定数の要素からなる低次元記述子を生成する。前記1以上のプロセッサは、前記第1学習モデルによって、前記低次元記述子から前記複数の材料それぞれの特性値を推定する。前記1以上のプロセッサは、前記特性値に基づいて、前記複数の材料から一部の材料を選択する。前記1以上のプロセッサは、前記一部の材料それぞれの、要素数が前記所定数より多い高次元記述子を生成する。前記1以上のプロセッサは、前記第2学習モデルによって、前記高次元記述子から前記一部の材料それぞれの特性値を推定する。

10

【発明の効果】

【0011】

本発明の一態様によれば、より効率的に所望の特性値を有すると期待できる有望な材料を選択することができる。

20

【図面の簡単な説明】

【0012】

【図1】本明細書の実施例に係る材料特性推定装置の論理構成例を模式的に示す。

【図2】材料特性推定装置のハードウェア構成例を示す。

【図3】材料特性推定装置の全体的な処理の例のフローチャートを示す。

【図4】モニタに表示される、材料実験データを入力するためのグラフィカルユーザインタフェースの例を模式的に示す。

【図5】実験済み材料データベースの構成例を示す。

【図6】モニタに表示される、材料特性値推定対象の材料リストを入力するためのグラフィカルユーザインタフェースの例を模式的に示す。

30

【図7】材料構造式データベースの構成例を示す。

【図8】低次元材料特性推定モデルの学習処理の詳細のフローチャートを示す。

【図9】記述子計算部が材料特性推定モデル学習部に渡す記述子リストの構成例を示す。

【図10】材料特性推定モデル学習部が実験済み材料データベースから取得する材料特性測定値（材料特性測定値リスト）の例を示す。

【図11】材料特性推定部から材料選択部に渡される、材料特性推定結果（材料特性推定値リスト）の構成例を示す。

【図12】高次元材料特性推定モデルの学習処理の詳細のフローチャートを示す。

【図13】材料特性推定結果表示部がモニタにおいて表示する材料特性推定結果の画像例を示す。

40

【発明を実施するための形態】

【0013】

以下においては、便宜上その必要があるときは、複数のセクションまたは実施例に分割して説明するが、特に明示した場合を除き、それらは互いに無関係なものではなく、一方は他方の一部または全部の変形例、詳細、補足説明等の関係にある。また、以下において、要素の数等（個数、数値、量、範囲等を含む）に言及する場合、特に明示した場合及び原理的に明らかに特定の数に限定される場合等を除き、その特定の数に限定されるものではなく、特定の数以上でも以下でもよい。

【0014】

50

本システムは、物理的な計算機システム（一つ以上の物理的な計算機）でもよいし、クラウド基盤のような計算リソース群（複数の計算リソース）上に構築されたシステムでもよい。計算機システムあるいは計算リソース群は、1以上のインタフェース装置（例えば通信装置及び入出力装置を含む）、1以上の記憶装置（例えば、メモリ（主記憶）及び補助記憶装置を含む）、及び、1以上のプロセッサを含む。

【0015】

プログラムがプロセッサによって実行されることで機能が実現される場合、定められた処理が、適宜に記憶装置及び/またはインタフェース装置等を用いながら行われるため、機能はプロセッサの少なくとも一部とされてもよい。機能を主語として説明された処理は、プロセッサあるいはそのプロセッサを有するシステムが行う処理としてもよい。プログラムは、プログラムソースからインストールされてもよい。プログラムソースは、例えば、プログラム配布計算機または計算機が読み取り可能な記憶媒体（例えば計算機読み取り可能な非一過性記憶媒体）であってもよい。各機能の説明は一例であり、複数の機能が一つの機能にまとめられたり、一つの機能が複数の機能に分割されたりしてもよい。

10

【0016】

以下において、バーチャルスクリーニングにおいて、所望の材料特性を有することが期待できる材料を効率的に選択することができる技術が開示される。材料特性推定装置は、候補材料の母集団に対して、2段階の絞り込みを行う。

【0017】

具体的には、材料特性推定装置は、全候補材料それぞれの低次元の記述子を計算する。さらに、材料特性推定装置は、低次元記述子それぞれから、簡易的な機械学習モデルを使用して、材料特性値を推定する。材料特性推定装置は、材料特性推定値に基づき一部の材料を選択する。

20

【0018】

次に、材料特性推定装置は、選択した材料それぞれの高次元の記述子を計算する。材料特性推定装置は、高次元記述子それぞれから、より高精度の機械学習モデルを使用して、材料特性値を推定する。材料特性推定装置は、これら材料特性推定値に基づき、最終的に候補としてユーザに提示する材料を選択する。このように、低次元記述子からの材料特性推定結果に基づき、高次元記述子を生成する材料を選択することで、所望の材料特性を有することが期待できる材料を効率的及び高速に選択することができる。

30

【0019】

以下においては、本明細書の実施例をより具体的に説明する。以下においては、化学構造式で表される有機化合物の特性値を推定する例が説明される。本明細書において説明する材料特性推定装置は、有機化合物及び無機化合物のいずれにも適用可能である。記述子は化学式、つまり、構造式及び組成式のいずれから生成することができる。

【実施例1】

【0020】

図1は、本明細書の実施例に係る材料特性推定装置の論理構成例を模式的に示す。材料特性推定装置100は、材料構造式データベース105、実験済み材料データベース106、及び選択済み構造式データベース112を格納している。

40

【0021】

材料特性推定装置100は、実験データ受付部103、材料リスト受付部104、記述子計算部107、材料特性推定モデル学習部108、材料特性推定部109、材料選択部110、及び材料特性推定結果表示部111を含む。これらは、プログラムであり、材料特性推定装置100の1以上のプロセッサは、これらプログラムを実行することで、それぞれに対応する機能部として動作することができる。なお、材料特性推定装置100の任意の機能は任意のプログラムに実装できる。

【0022】

実験データ受付部103は、入出力装置を介してユーザ200に入力された、様々な材料の特性値を示す実験データを受け取り、実験済み材料データベース106に格納する。

50

材料リスト受付部 104 は、入出力装置を介してユーザ 200 に入力された多くの材料の化学構造式のデータを受け取り、材料構造式データベース 105 に格納する。材料構造式データベース 105 は、実験済み材料データベース 106 に格納されていない材料（化学構造式）のデータを格納している。

【0023】

記述子計算部 107 は、化学構造式から、所定の方法によって、記述子を生成する。記述子は、化学構造式が示す材料の特徴を表す。記述子は、複数の要素（特徴量）で構成されたベクトルで表される。各要素が対応する特徴、例えば、分子量や元素混合比を表す。後述するように、記述子計算部 107 は、一つの化学構造式から、要素数が少ない低次元記述子と、要素数が多い高次元記述子を生成することができる。なお、記述子計算部 107 は、低次元記述子と高次元記述子それぞれのためのモジュールに分割されてもよい。

10

【0024】

低次元記述子及び高次元記述子それぞれの要素数及び要素の種類は一定であり、低次元記述子の要素数は、高次元記述子の要素数より少ない。低次元記述子の要素の全ての種類が、高次元記述子の要素の種類に含まれていてもよく、低次元記述子の要素の一部又は全部が、高次元記述子の要素の種類と異なってもよい。

【0025】

低次元記述子の要素の全ての種類が、高次元記述子の要素の種類に含まれている場合、高次元記述子から特性値を推定する学習モデルの観点から、低次元記述子のより適切な要素の種類を決定することができる。記述子計算部 107 は、高次元記述子の要素の重要度を決定し、それら重要度に基づき、低次元記述子を構成する要素を決定してもよい。これにより、低次元記述子からの材料特性値推定により、より適切な候補材料を選択することができる。

20

【0026】

例えば、記述子計算部 107 は、高次元記述子を用いて、ランダムフォレストや勾配ブースティング等の決定木系アンサンブル学習器で学習を行い、高次元記述子の各要素の重要度を算出する。記述子計算部 107 は、重要度の高い要素から順に所定数の記述子を選択する。

【0027】

他の例において、記述子計算部 107 は、Permutation Importance や LASSO による線形回帰を使用して、高次元記述子の要素の重要度を決定してもよい。低次元記述子の要素を選択するための機械学習モデルは、高次元記述子から材料特性値を推定する機械学習モデルと、同一でもよく又は異なってもよい。これらの間のアルゴリズムも、同一でもよく又は異なってもよい。

30

【0028】

材料特性推定モデル学習部 108 は、化学構造式（材料）の記述子から、所定の特性値を推定することができる材料特性推定モデル（学習モデル）の学習を行う。後述するように、本明細書の実施例に係る材料特性推定装置 100 は、低次元記述子から 1 以上の所定種類の特性値を推定する低次元材料特性推定モデル（第 1 学習モデル）と、高次元記述子から同様の種類の特性値を推定する高次元材料特性推定モデル（第 2 学習モデル）とを用意する。

40

【0029】

設計により、複数の低次元材料特性推定モデルが用意されてもよく、それらの次元数は異なっても共通でもよく、モデル間において要素種類の組み合わせは同一でも異なってもよい。いずれの低次元材料特性推定モデルの低次元記述子の次元も、高次元記述子の次元数より小さい。

【0030】

材料特性推定モデルは、1 又は複数の種類の特性値を推定するように構成できる。以下に説明する例においては、材料特性推定モデルは単一の特性値を推定（出力）するものとする。低次元材料特性推定モデル及び高次元材料特性推定モデルそれぞれが利用する回帰

50

アルゴリズムは任意であり、これらのアルゴリズムは同一でも異なってもよい。例えば、ランダムフォレスト、サポートベクタマシン、ガウス過程回帰、ニューラルネットワークを含む様々な回帰アルゴリズムから、任意のアルゴリズムを選択できる。

【0031】

材料特性推定部109は、学習済みの低次元材料特性推定モデルを使用して、低次元記述子から、推定材料特性値を得る。さらに、材料特性推定部109は、学習済みの高次元材料特性推定モデルを使用して、高次元記述子から、推定材料特性値を得る。以下に説明する例において、低次元記述子は、材料構造式データベース105に格納された全ての化学構造式(材料)から生成される。高次元記述子は、推定材料特性値が理想に近い値を示す低次元記述子の材料に対して生成される。

10

【0032】

材料選択部110は、低次元記述子に対して推定された材料特性値に基づき、高次元記述子を生成する材料(化学構造式)を選択して、その情報(化学構造式)を選択済み構造式データベース112に格納する。材料選択の基準は、材料特性値の性質やユーザの要求に依存し、例えば、材料特性値が目標値に近い特性推定値を有する所定数の材料や、所定範囲内に含まれる材料が選択されてもよい。

【0033】

材料特性値が高いほど好ましい場合、材料選択部110は、材料特性値が最も高い値を示す所定数の材料を選択してもよいし、材料特性値が所定の閾値を超える材料を選択してもよい。材料特性値が低いほど好ましい場合、材料選択部110は、材料特性値が最も低い値を示す所定数の材料を選択してもよいし、材料特性値が所定の閾値未満の材料を選択してもよい。

20

【0034】

複数の低次元材料特性推定モデルが用意されている場合、これら複数の低次元材料特性推定モデルの推定値の統計値(例えば加重平均値(平均値含む))に基づいて、高次元記述子を生成する材料が選択されてもよい。

【0035】

材料特性推定結果表示部111は、高次元材料特性値推定モデルによる、選択された材料の高次元記述子からの材料特性値推定結果を取得する。材料特性推定結果表示部111は、対応する化学構造と共に、材料特性値推定結果を表示することで、有望な材料をユーザ200に提示する。材料特性推定結果表示部111は、選択された全ての材料の推定結果を表示してもよく、所定の基準で選択された好ましい推定値を示す一部の材料のみの推定結果を表示してもよい。

30

【0036】

図2は、材料特性推定装置100のハードウェア構成例を示す。材料特性推定装置100は、演算性能を有するプロセッサ151と、プロセッサ151が実行するプログラム及びデータを格納する揮発性一時記憶領域を与えるDRAM152と、を含む。材料特性推定装置100は、さらに、他の装置とデータ通信をおこなう通信装置153と、HDD(Hard Disk Drive)やフラッシュメモリなどを利用した永続的な情報記憶領域を与える補助記憶装置154と、を含む。

40

【0037】

例えば、補助記憶装置154は、実験データ受付部103、材料リスト受付部104、記述子計算部107、材料特性推定モデル学習部108、材料特性推定部109、材料選択部110、及び材料特性推定結果表示部111等のプログラムを格納する。補助記憶装置154は、さらに、材料構造式データベース105、実験済み材料データベース106、及び選択済み構造式データベース112等の各種データを格納する。プロセッサ151が実行するプログラム及び処理対象のデータは、補助記憶装置154からDRAM152にロードされる。

【0038】

また、材料特性推定装置100は、ユーザからの操作を受け付ける入力装置155と、

50

各プロセスでの出力結果をユーザに提示するモニタ156（出力装置の例）と、を含む。なお、材料特性推定装置100の機能は、複数の装置に分かれて実装されてもよい。このように、材料特性推定装置100は、1以上の記憶装置及び1以上のプロセッサを含む。

【0039】

図3は、材料特性推定装置100の全体的な処理の例のフローチャートを示す。ステップS101において、実験データ受付部103が、入力装置155を介して、ユーザ102から材料実験データを受け取り、実験済み材料データベース106に格納する。さらに、ステップS102において、材料リスト受付部104が、入力装置155を介して、ユーザ102から材料リストを受け取り、材料構造式データベース105に格納する。

【0040】

図4は、モニタ156に表示される、材料実験データを入力するためのグラフィカルユーザインタフェース（GUI）201の例を模式的に示す。ユーザは、入力装置155からGUI201に必要な情報を入力する。ユーザは、GUI201における「参照ボタン」を使用して実験データを格納しているファイルを指定し、「OK」ボタンを選択することで、当該ファイルを実験データ受付部103に指示することができる。実験データ受付部103は、指定されたファイルのデータを実験済み材料データベース106に格納する

【0041】

図5は、実験済み材料データベース106の構成例を示す。実験済み材料データベース106は、材料と材料の特性値の実験結果とを関連付ける。具体的には、実験済み材料データベース106は、番号カラム251、構造式（SMILES）カラム252、及び材料特性測定値カラム253を含む。

【0042】

番号カラム251は、実験済み材料データベース106においてレコードそれぞれを識別する。構造式（SMILES）カラム252は、材料の化学構造式を示す。図4の例において、化学構造式は、SMILES（Simplified Molecular Input Line Entry System）記法に従って表現されている。記述子を生成することができる、化学構造式の任意の表現形式を使用することができる。材料特性測定値カラム253は、化学構造式それぞれの所定の特性値の実験結果を示す。実験済み材料データベース106（測定データベース）が格納する測定値の一部又は全部はシミュレーション結果の測定値であってもよい。

【0043】

図6は、モニタ156に表示される、材料特性値推定対象の材料リストを入力するためのGUI202の例を模式的に示す。ユーザは、入力装置155からGUI202に必要な情報を入力する。ユーザは、GUI202において、「参照ボタン」を使用して材料リストを格納しているファイルを指定し、「OK」ボタンを選択することで、当該ファイルを材料リスト受付部104に指示することができる。材料リスト受付部104は、指定されたファイルのデータを材料構造式データベース105に格納する。

【0044】

図7は、材料構造式データベース105の構成例を示す。材料構造式データベース105は、材料特性値推定対象の化学構造式を格納する。ここで説明する例において、材料構造式データベース105に格納されている全て材料の低次元記述子が生成され、その材料特性値が推定される。他の例において、材料構造式データベース105に格納されている一部の材料（化学構造式）が、低次元記述子を生成するために選択されてもよい。例えば、ランダムに所定数の材料が選択されてもよい。

【0045】

図7に示す例において、材料構造式データベース105は、番号カラム261、及び構造式（SMILES）カラム262を含む。番号カラム261は、材料構造式データベース105においてレコードそれぞれを識別する。構造式（SMILES）カラム262は、材料の化学構造式のSMILES表現を示す。

【0046】

10

20

30

40

50

図3に戻って、ステップS103において、材料特性推定モデル学習部108が低次元材料特性推定モデルの学習(訓練)を行い、学習済み材料特性推定部109に渡す。ここでは、一つの低次元材料特性推定モデルが構成されるとする。

【0047】

図8は、低次元材料特性推定モデルの学習処理(S103)の詳細のフローチャートを示す。まず、ステップS201において、材料特性推定モデル学習部108の指示に応じて、記述子計算部107は、実験済み材料データベース106から一部又は全ての材料(化学構造式)を取得し、それぞれの低次元記述子を計算する。

【0048】

一例において、低次元材料特性推定モデルの学習のために実験済み材料データベース106から取得される材料の数(学習データ量)は、後述する高次元材料特性推定モデルのために取得される材料の数(学習データ量)より少ない。低次元材料特性推定モデルの次元数は、高次元材料特性推定モデルの次元数より小さいため、高次元材料特性推定モデルより少ない学習データで効率的かつ適切に学習を行うことができる。

【0049】

実験済み材料データベース106から取得されるデータは、実験済み材料データベース106の番号カラム251の値及び構造式カラム252の値を示す。低次元記述子を構成する記述子要素の種類及び数は、上述のように、装置内に予め設定されている、又は、高次元記述子の要素から重要度に基づき選択され、設定されている。

【0050】

次に、ステップS202において、材料特性推定モデル学習部108は、記述子計算部107から計算された低次元記述子を受け取り、実験済み材料データベース106から、計算された低次元記述子に対応する化学構造式の材料特性値(材料特性測定値リスト)を取得する。

【0051】

図9は、記述子計算部107が材料特性推定モデル学習部108に渡す記述子リストの構成例を示す。図9は、記述子計算部107が低次元記述子の記述子リスト300の例を示すが、高次元記述子リストのテーブル構成も、記述子要素数が少ないことを除き同様である。

【0052】

記述子リスト300は、番号カラム301と記述子要素それぞれのカラムを含む。番号カラム301の値は、実験済み材料データベース106の番号カラム251の値に対応する。実験データ図9の例においては、記述子は4000の記述要素で構成され、四つの記述子要素のカラムが、例として、符号302~305で指示されている。

【0053】

図10は、材料特性推定モデル学習部108が実験済み材料データベース106から取得する材料特性測定値(材料特性測定値リスト)330の例を示す。材料特性測定値リストは、番号カラム331及び材料特性測定値カラム332を含む。番号カラム331の値は、実験済み材料データベース106の番号カラム251の値に対応する。材料特性測定値カラム332の値は、材料特性測定値カラム253の値に対応する。

【0054】

図8に戻って、ステップS203において、材料特性推定モデル学習部108は、取得した低次元記述子と材料特性値とから、低次元材料特性推定モデルの学習を行う。材料特性推定モデル学習部108は、低次元材料特性推定モデルの初期構成の情報を予め保持しており、その情報に従って、低次元材料特性推定モデルを構成する。上述のように、低次元材料特性推定モデルとして、任意の種類 of 機械学習モデルが利用できる。

【0055】

さらに、材料特性推定モデル学習部108は、低次元材料特性推定モデルに低次元記述子を順次入力し、出力された材料特性値推定値を取得する。材料特性推定モデル学習部108は、材料特性値推定値と上記取得した材料特性測定値との誤差に基づき、低次元材料

10

20

30

40

50

特性推定モデルのパラメータを更新することで、低次元材料特性推定モデルを最適化する。最後に、ステップS 204において、材料特性推定モデル学習部108は、学習済み低次元材料特性推定モデルを材料特性推定部109に渡す。

【0056】

図3に戻って、ステップS 104において、材料特性推定部109からの指示に応じて、記述子計算部107は、材料構造式データベース105から一部又は全ての化学構造式(レコード)を取得して、それぞれの低次元記述子を計算する。記述子計算部107が材料構造式データベース105から取得するデータにおいて、番号と化学構造式の対応は、図7に示す材料構造式データベース105と同様である。

【0057】

ステップS 105において、材料特性推定部109は、記述子計算部107から計算された低次元記述子を受け取り、材料特性推定を行う。具体的には、材料特性推定部109は、取得した低次元記述子それぞれを、学習済みの低次元材料特性推定モデルに入力して、対応する特性推定値を取得する。

【0058】

次に、ステップS 106において、材料選択部110は、材料特性推定部109から材料特性推定結果を受け取り、さらに、受け取った推定結果が示す番号の化学構造式を材料構造式データベース105から取得する。材料選択部110は、材料特性推定結果に基づいて材料(化学構造式)の選択を行い、選択した化学構造式を選択済み構造式データベース112に格納する。選択済み構造式データベース112のデータ構造は、材料構造式データベース105と同様であり、化学構造式に対応付けられる番号も同様でよい。

【0059】

図11は、材料特性推定部109から材料選択部110に渡される、材料特性推定結果(材料特性推定値リスト)340の構成例を示す。材料特性推定値リスト340は、番号カラム341及び材料特性推定値カラム342を含む。番号カラム341の値は、材料構造式データベース105の番号カラム261の値に対応する。材料特性推定値カラム342は、番号カラム341が示す番号の化学構造式それぞれの材料特性推定値を示す。

【0060】

材料選択部110は、材料特性推定値リスト340を参照し、予め設定された条件に合致する特性推定値を有する材料を選択し、選択した材料の化学構造式を選択済み構造式データベース112に格納する。

【0061】

図3に戻って、ステップS 107において、材料特性推定モデル学習部108は、高次元材料特性推定モデルの学習(訓練)を行い、学習済みの高次元材料特性推定モデルを材料特性推定部109に渡す。

【0062】

図12は、高次元材料特性推定モデルの学習処理(S 107)の詳細のフローチャートを示す。まず、ステップS 301において、材料特性推定モデル学習部108の指示に応じて、記述子計算部107は、実験済み材料データベース106から一部又は全ての材料(化学構造式)を取得し、それぞれの高次元記述子を計算する。実験済み材料データベース106から取得されるデータは、実験済み材料データベース106の番号カラム251の値及び構造式カラム252の値を示す。高次元記述子を構成する記述子要素の種類及び数は予め設定されている。

【0063】

次に、ステップS 302において、材料特性推定モデル学習部108は、記述子計算部107から計算された高次元記述子(記述子リスト)を受け取る。記述子計算部107が材料特性推定モデル学習部108に渡す記述子リストの構造は、図9に示す記述子リスト300と同様であり、これらの間のレコードは同一でも異なってもよい。

【0064】

さらに、材料特性推定モデル学習部108は、実験済み材料データベース106から、

10

20

30

40

50

計算された高次元記述子に対応する化学構造式の材料特性値（材料特性測定値リスト）を取得する。材料特性測定値リストの構造は、図10に示す材料特性測定値リスト330と同様である。材料特性測定値リストの番号（対応する材料）は、高次元記述子リストの番号（対応する材料）と一致する。

【0065】

ステップS303において、材料特性推定モデル学習部108は、取得した高次元記述子と材料特性値とから、高次元材料特性推定モデルの学習を行う。材料特性推定モデル学習部108は、高次元材料特性推定モデルの初期構成の情報を予め保持しており、その情報に従って、高次元材料特性推定モデルを構成する。上述のように、高次元材料特性推定モデルとして、任意の種類機械学習モデルが利用できる。

10

【0066】

さらに、材料特性推定モデル学習部108は、高次元材料特性推定モデルに高次元記述子を順次入力し、出力された材料特性値推定値を取得する。材料特性推定モデル学習部108は、材料特性値推定値と上記取得した材料特性測定値との誤差に基づき、高次元材料特性推定モデルのパラメータを更新することで、高次元材料特性推定モデルを最適化する。最後に、ステップS304において、材料特性推定モデル学習部108は、学習済み高次元材料特性推定モデルを材料特性推定部109に渡す。

【0067】

図3に戻って、ステップS108において、材料特性推定部109の指示に回答して、記述子計算部107は、選択済み構造式データベース112から化学構造式を取得して、それらの高次元記述子を計算する。低次元記述子による特性推定値に基づき選択された材料のみの高次元記述子を計算することで、高次元記述子の計算及びその後の特性推定値の計算を高速化できる。

20

【0068】

次に、ステップS109において、材料特性推定部109は、記述子計算部107から、選択済み構造式データベース112に格納されている化学構造式の高次元記述子を受け取り、材料特性推定を行う。具体的には、材料特性推定部109は、取得した高次元記述子を順次、学習済み高次元材料特性推定モデルに入力する。高次元材料特性推定モデルは、入力された高次元記述子それぞれの材料特性推定値を出力する。

【0069】

次に、ステップS110において、材料特性推定結果表示部111は、材料特性推定部109から、選択された化学構造式の材料特性推定結果を受け取る。材料特性推定結果表示部111は、さらに、選択済み構造式データベース112から化学構造式を取得する。材料特性推定結果表示部111は、取得した材料特性推定結果と化学構造式とを、ユーザに提示する。

30

【0070】

図13は、材料特性推定結果表示部111がモニタ156において表示する材料特性推定結果の画像例を示す。図13の例において、画像は、選択された材料の化学構造式と、それらに対応する材料特性値の推定値とを示す。ユーザは、表示された化学構造式及び材料特性値を参考に、実際に実験又はシミュレーションを実行する化学構造式を決定することができる。保存ボタンにより推定結果が保存される。

40

【0071】

なお、本発明は上記した実施例に限定されるものではなく、様々な変形例が含まれる。例えば、上記した実施例は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明したすべての構成を備えるものに限定されるものではない。また、ある実施例の構成の一部を他の実施例の構成に置き換えることが可能であり、また、ある実施例の構成に他の実施例の構成を加えることも可能である。また、各実施例の構成の一部について、他の構成の追加・削除・置換をすることが可能である。

【0072】

また、上記の各構成・機能・処理部等は、それらの一部又は全部を、例えば集積回路で

50

設計する等によりハードウェアで実現してもよい。また、上記の各構成、機能等は、プロセッサがそれぞれの機能を実現するプログラムを解釈し、実行することによりソフトウェアで実現してもよい。各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリや、ハードディスク、SSD (Solid State Drive) 等の記録装置、または、ICカード、SDカード等の記録媒体に置くことができる。

【0073】

また、制御線や情報線は説明上必要と考えられるものを示しており、製品上必ずしもすべての制御線や情報線を示しているとは限らない。実際には殆どすべての構成が相互に接続されていると考えてもよい。

【符号の説明】

【0074】

100 材料特性推定装置、103 実験データ受付部、104 材料リスト受付部、105 材料構造式データベース、106 実験済み材料データベース、107 記述子計算部、108 材料特性推定モデル学習部、109 材料特性推定部、110 材料選択部、110 材料特性推定結果表示部、112 選択済み構造式データベース

10

20

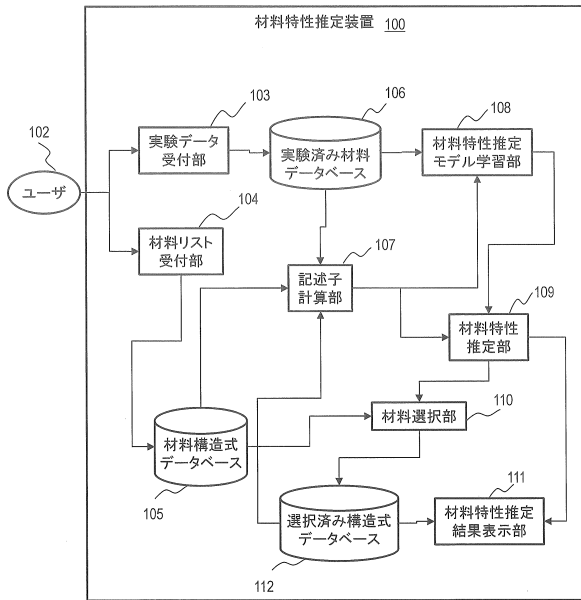
30

40

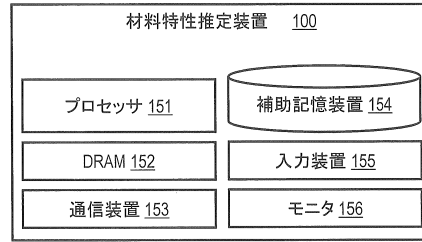
50

【 図 面 】

【 図 1 】



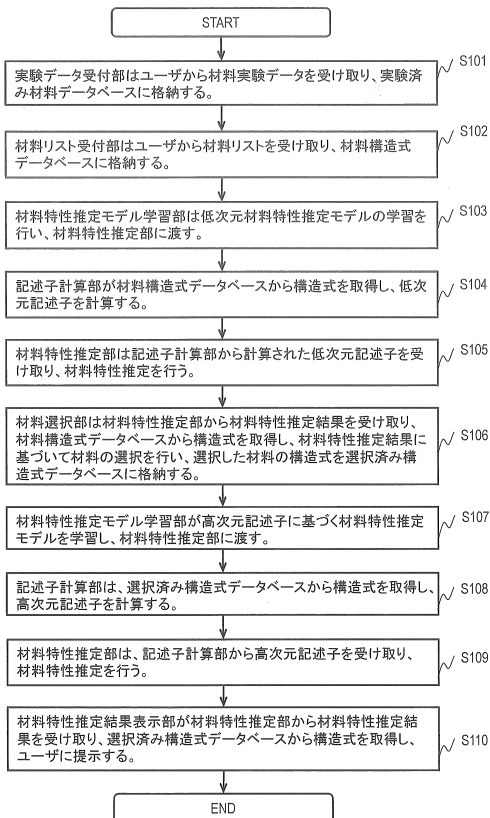
【 図 2 】



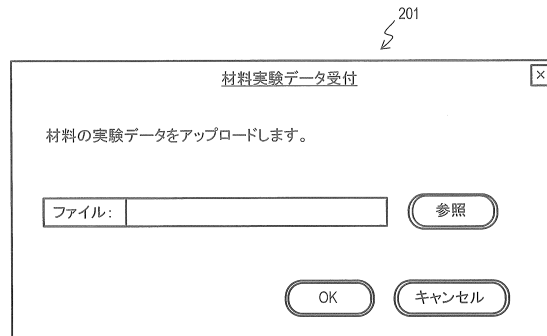
10

20

【 図 3 】



【 図 4 】



30

40

50

【図 5】

番号	構造式 (SMILES)	材料特性測定値
1	<chem>CO[Si](C)(OC)N1CCCCC1</chem>	15.372
2	<chem>CCN(C)[Si](C)(OC)OC</chem>	25.097
3	<chem>O=C1C=CC(=O)C=C1</chem>	10.406
⋮	⋮	⋮

実験済み材料データベース

【図 6】

材料リスト受付

材料特性値推定対象の材料リストをアップロードします。

ファイル: 参照

OK キャンセル

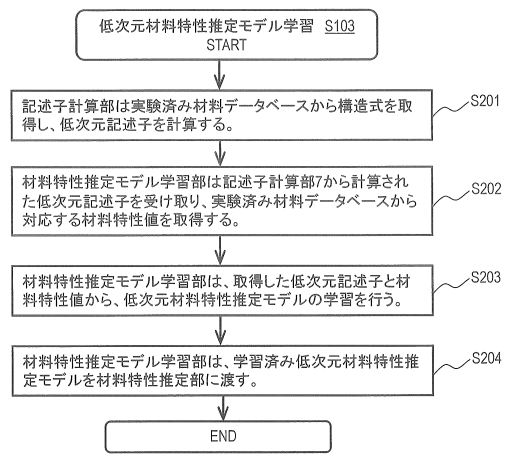
10

【図 7】

番号	構造式 (SMILES)
1	<chem>C1C=CC2C1C3CC2C=C3</chem>
2	<chem>CN1C=NC2=C1C(=O)N(C(=O)N2C)C</chem>
3	<chem>CC(C)(C)c1cc(C)cc(c1O)C(C)(C)C</chem>
⋮	⋮

材料構造式データベース

【図 8】



20

【図 9】

番号	記述子要素1	記述子要素2	記述子要素3	...	記述子要素200
1	15.372	-6.534	6.478	...	153.248
2	25.097	-5.409	12.036	...	78.501
3	10.406	-8.241	10.114	...	92.664
⋮	⋮	⋮	⋮	⋮	⋮

記述子リスト

【図 10】

番号	材料特性測定値
1	15.372
2	25.097
3	10.406
⋮	⋮

材料特性測定値リスト

30

40

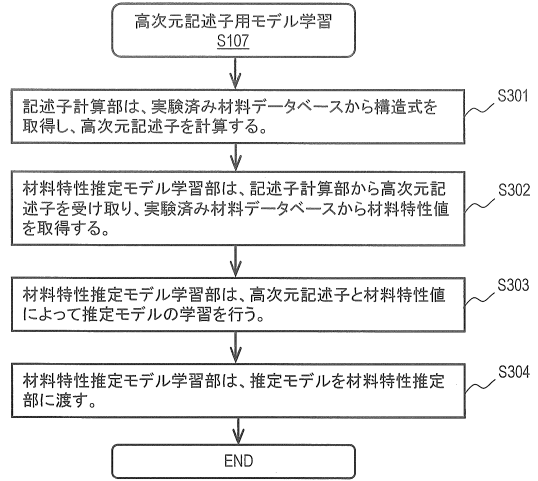
50

【図 1 1】

番号	材料特性推定値
1	15.013
2	18.954
3	22.677
⋮	⋮

材料特性推定値リスト

【図 1 2】



10

【図 1 3】

材料特性推定結果

番号	構造式 (SMILES)	材料特性推定値
1	<chem>C1C=CC2C1C3CC2C=C3</chem>	25.412
2	<chem>COC1=CC=C(C=C1)O</chem>	24.037
3	<chem>O=C1C=CC(=O)C=C1</chem>	23.891
⋮	⋮	⋮

20

30

40

50

フロントページの続き

東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

審査官 合田 幸裕

- (56)参考文献 国際公開第2020/031671(WO,A1)
国際公開第2020/075573(WO,A1)
国際公開第2019/060268(WO,A1)
国際公開第2018/098588(WO,A1)
森川幸治, 無機材料へのマテリアルズインフォマティクスの取組みと課題, 人工知能, 日本, 一般社団法人人工知能学会, 2019年05月01日, 第34巻, 第3号, pages 364-369, DOI: https://doi.org/10.11517/jjsai.34.3_364
- (58)調査した分野 (Int.Cl., DB名)
G06F 30/10
G06F 30/27
IEEE Xplore
JSTPlus (JDreamIII)