

明 細 書

発明の名称：材料特性予測装置および材料特性予測方法

技術分野

[0001] 本発明は、材料の物性予測のための機械学習システムに関する。

背景技術

[0002] 従前は、化合物の物性を知るためには、合成実験により実際に生成し、直接実験によって測定することが必要であった。しかし現代では、数多くの化合物物性データが蓄積されてくるにつれ、そのデータを機械学習アルゴリズムで処理することによって未知の化合物に対する物性を予測するという試みも行われるようになってきている。

[0003] 未知の化合物の組み合わせは膨大であるから、このように実験を経ずして望ましい性質を持つ物質群を選別することができれば、材料開発の効率を大幅に向上させることが可能になると期待される。これを実現するためには機械学習による予測の信頼性が高いことが必要である。しかし、現状では化合物の構造を機械学習に適した扱いやすい特徴量（記述子）に変換する方法が十分でないため、高い信頼性を持った化合物物性予測システムが未だ実現していない。

[0004] 例えば非特許文献1では、公知の変分オートエンコーダを用いて化合物の構造情報を多変量ベクトルに変換し物性予測に用いる方法が開示されている。

[0005] また特許文献1では、有機化合物の分子構造を複数種類のフィンガープリント法によって表記して物性予測に用いる方法が開示されている。特許文献2では、化合物の情報を潜在変数として公知のオートエンコーダによって符号化する薬物設計用機械学習システムが開示されている。

先行技術文献

特許文献

[0006] 特許文献1：国際公開 WO 2019/048965 A1

特許文献2：特表2019-502988号公報

非特許文献

- [0007] 非特許文献1：R. Gomez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernandez-Lobato, B. Sanchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, “Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules”, ACS Cent. Sci. 2018, 4, 268-276.

発明の概要

発明が解決しようとする課題

- [0008] ある特定の物性を有する物質を機械学習法によって探索したい状況では、その物性を有効に表現しコンピュータに的確に理解させるために必要十分な特徴量（記述子）を化合物の構造式から生成することが求められる。
- [0009] しかしながら、従前の方法、例えばRDKitのようなフリーソフトウェアにより構造式から特徴量を生成する場合には、汎用で一般的な特徴量の集合が生成されるため、所望の分析目的にとっては冗長である。化合物の冗長な特徴量から、所望の物性の予測にとって重要なものだけを選別するためには、十分な分量の教師データ（化合物と教師ラベル（所望の物性値）のペア）を準備して予測モデルをその上で訓練し、その予測モデルが有用と判定した特徴量を抽出することが望ましい。しかし、これは教師データが十分に手に入らない状況においては実行困難である。
- [0010] また、公知のオートエンコーダのように構造式を連続多変量ベクトルに変換できる機械学習手段を用いる際は、そのモデル（ニューラルネットワーク等）の学習にChEMBLのような化合物の大規模オープンデータベースを利用することが普通であり、そのようなデータセットには特定の物性を有する物質以外にも数多くの物質が含まれているため、そこから生成された連続多変量ベクトルは所望の分析目的にとって必ずしも最適なものにならない。
- [0011] 本発明は、上記のような課題に鑑みて成されたものであり、専門家の知見を反映した効果的な化合物特徴量（説明変数）を効率的に生成することがで

きるようにし、それによって未知の化合物の有する物性を精度よく予測することを目的とする。

課題を解決するための手段

[0012] 本発明の好ましい一側面は、案件データベースを複数格納した案件別材料データベースを用いて、材料特性の予測を行うための装置である。案件データベースは、材料の構造に関する構造情報および材料の特性に関する材料特性を関連付けて記録したレコードを複数含む。この装置は、少なくとも一つの案件データベースの指定を受け付ける、化合物空間指定部と、化合物空間指定部の受付けた案件データベースに対応する構造情報を多変量に変換するオートエンコーダを生成するオートエンコーダ学習部と、オートエンコーダ学習部の生成したオートエンコーダが変換した多変量を用いて材料特性を予測する材料特性予測部と、を持つことを特徴とする。

[0013] 本発明の好ましい他の一側面は、材料の構造に関する構造情報を記録したレコードを複数含む第1のデータベースを準備する第1のステップ、第1のステップで準備した第1のデータベースから、構造情報を抽出する第2のステップ、第2のステップで抽出した構造情報を用いて、構造情報を多変量に変換するオートエンコーダを訓練する第3のステップ、材料の構造に関する構造情報および材料の特性に関する材料特性を関連付けて記録したレコードを複数含む第2のデータベースを準備する第4のステップ、第4のステップで準備した第2のデータベースから、構造情報を抽出する第5のステップ、オートエンコーダを用いて、第5のステップで抽出された構造情報を多変量に変換する第6のステップ、第6のステップで変換された多変量に基づいて説明変数を得、第2のデータベースから抽出した材料特性に基づいて目的変数を得る第7のステップ、説明変数と目的変数を用いて、説明変数から目的変数を推定する予測モデルを生成する第8のステップ、を実行する材料特性予測方法である。

発明の効果

[0014] 本発明によれば、専門家の知見を反映した効果的な化合物特徴量を効率的

に生成することができ、それによって未知の化合物の有する物性を精度よく予測することが可能になる。

図面の簡単な説明

- [0015] [図1]実施例における材料特性予測装置の機能構成を示すブロック図である。
- [図2]実施例における材料特性予測装置の処理のフローチャートである。
- [図3]実施例における実験データ受付部の表示画面の一例のイメージ図である。
- 。
- [図4]実施例における案件別材料データベースのデータ構造の一例の表図である。
- [図5]実施例における化合物空間指定部の表示画面の一例のイメージ図である。
- 。
- [図6]実施例における案件別材料データベースからオートエンコーダ学習部への出力のデータ構造の一例の表図である。
- [図7]実施例におけるオートエンコーダの構成を説明する概念図である。
- [図8]実施例における材料特性予測受付部の表示画面の一例のイメージ図である。
- [図9]実施例における材料特性予測受付部への入力データのデータ構造の一例の表図である。
- [図10]実施例における材料特性予測部の処理のフローチャートである。
- [図11]実施例における表示部の表示画面の一例のイメージ図である。
- [図12]実施例の利用イメージを説明する概念図である

発明を実施するための形態

- [0016] 以下、本発明の実施例について、図面を参照しながら詳細に説明する。ただし、本発明は以下に示す実施の形態の記載内容に限定して解釈されるものではない。本発明の思想ないし趣旨から逸脱しない範囲で、その具体的構成を変更し得ることは当業者であれば容易に理解される。
- [0017] 以下に説明する発明の構成において、同一部分又は同様な機能を有する部分には同一の符号を異なる図面間で共通して用い、重複する説明は省略する

ことがある。

[0018] 同一あるいは同様な機能を有する要素が複数ある場合には、同一の符号に異なる添字を付して説明する場合がある。ただし、複数の要素を区別する必要がない場合には、添字を省略して説明する場合がある。

[0019] 本明細書等における「第1」、「第2」、「第3」などの表記は、構成要素を識別するために付するものであり、必ずしも、数、順序、もしくはその内容を限定するものではない。また、構成要素の識別のための番号は文脈毎に用いられ、一つの文脈で用いた番号が、他の文脈で必ずしも同一の構成を示すとは限らない。また、ある番号で識別された構成要素が、他の番号で識別された構成要素の機能を兼ねることを妨げるものではない。

[0020] 図面等において示す各構成の位置、大きさ、形状、範囲などは、発明の理解を容易にするため、実際の位置、大きさ、形状、範囲などを表していない場合がある。このため、本発明は、必ずしも、図面等に関連された位置、大きさ、形状、範囲などに限定されない。

[0021] 本明細書で引用した刊行物、特許および特許出願は、そのまま本明細書の説明の一部を構成する。

[0022] 本明細書において単数形で表される構成要素は、特段文脈で明らかに示されない限り、複数形を含むものとする。

実施例 1

[0023] 以下で説明される実施例の一態様は、材料特性の予測を行うための装置であって、案件別に材料の構造式と実験条件と材料特性とを関連付けて記録する案件別材料データベースと、利用者から単数もしくは複数の案件の指定を受け付ける化合物空間指定部と、化合物空間指定部の受付けた案件に対応する構造式の集合に対し、構造式を多変量に変換した後、多変量から構造式を復元できるような性質を備えたモデルであるオートエンコーダを生成する機能を持ったオートエンコーダ学習部と、オートエンコーダ学習部の生成したオートエンコーダを用いて説明変数を生成し材料特性を予測する材料特性予測部を有する。

- [0024] 本実施例によれば、十分な量の化合物教師データが無い状況においても専門家の知見を取り入れた効果的な特徴量を生成し、それを以て精度の高い物性予測を行うことを可能にする分析システムを提供することができる。
- [0025] 図1は、本実施例における材料特性予測装置の機能構成の一例を示す機能構成ブロック図である。図1において、材料特性予測装置101は、利用者102から材料実験データを受け取る実験データ受付部106、材料実験データを案件別に蓄積するための案件別材料データベース107、利用者102から化合物空間の指定を受け取る化合物空間指定部103、オートエンコーダの作成と学習を実行するオートエンコーダ学習部104、オートエンコーダ108、利用者102から予測対象の材料リストを受け取る材料特性予測受付部105、材料特性の予測を行う材料特性予測部109、予測結果を利用者102に表示する表示部110を具えている。
- [0026] なお、材料特性予測装置101は、ハードウェアとしては、一般的な情報処理装置である、プロセッサとメモリと記憶装置と通信部（インターフェース）を有する装置によって実現される。すなわち、実験データ受付部106、化合物空間指定部103、材料特性予測受付部105は通信部によってデータ入力を受け付け、案件別材料データベース107は記憶装置にデータを保存する。また、オートエンコーダ学習部104、オートエンコーダ108、材料特性予測部109は、プロセッサにより、メモリに格納されたプログラムを実行するソフトウェア処理で実行される。また、表示部110は、一般的に用いられる操作部と表示部を兼用しており、例えばディスプレイ、キーボード、マウス等からなり、また、タッチパネルを有したディスプレイでもよい。
- [0027] 図2は、本実施例における材料特性予測装置101の処理のフローチャートである。図2において、ステップS201では実験データ受付部106が利用者102から材料実験データを受け取る。
- [0028] 図3に実験データ受付部106の入力受付画面の例を示す。典型的な例では、材料実験データはすでに電子ファイルとして記憶媒体などに格納されて

いる。図3に示すように、利用者は、材料実験データのファイル名をマウスやキーボード等を用いて指定してアップロードし、OKボタンを押すことによって確定する。またキャンセルボタンにより、入力を訂正できる。

[0029] 図2に戻って、ステップS202では案件別材料データベース107が実験データ受付部106から材料実験データを取得し、案件ごとに保存する。

[0030] 図4は、案件別材料データベース107の保存するデータの形式を示す。図4に示すように、このデータは案件番号401、通し番号402、化合物の構造式403、実験条件404および405、材料特性値406および407を含む。化合物の構造式はSMILES (Simplified molecular-input line-entry system) 形式を用いることによって簡便に表記することができるが、必ずしもこれに限られるものではなく、例えば分子の構造をグラフ形式のデータで扱ってもよい。また実験条件はデータに含まれていなくとも構わず、材料特性値は複数でも単一でもよいものとする。

[0031] 図4に示すように、案件別材料データベース107は案件データ（案件データベース）ごとに分割して保存されている。本実施例では、一つの案件データでは、実験条件404および405、材料特性値406および407が同じ定義あるいは種類のデータで統一された複数のレコードを含むものとする。この例では、一つのレコードは、一つの材料構造に対応している。

[0032] 各案件データは、対象とする材料、材料特性の定義、材料の作成主体、材料の作成目的、材料の作成時期、材料の作成設備などの少なくとも一つが異なるデータであって、例えば異なるテーマに関連して行った実験結果のデータである。よって、材料の構造や製造・実験条件や材料特性は、その定義や種類が案件毎に異なるものが含まれる場合がある。データの作成主体、作成目的、作成時期、作成設備、テーマ等の付随的な情報は、例えばテキスト情報として案件データに関連付けて記憶しておき、利用者が参照したり検索したりできるようにしてもよい。この場合、化合物空間指定部103が、案件データをキーワード等で検索を行うためのGUI (Graphical User Interface) を備える。利用者は検索機能を用いて、使用する案件データを抽出するこ

とができる。

- [0033] 図2に戻って、ステップS203では化合物空間指定部103が化合物空間の指定を利用者102の入力から取得する。
- [0034] 図5は、化合物空間指定部103の入力受付画面の例を示す。図5に示すように、利用者は、案件別材料データベース107から、利用する化合物案件データの選択を行うことができる。
- [0035] 図2に戻って、ステップS204では、オートエンコーダ学習部104が化合物空間指定部103から化合物空間の指定を取得し、それに該当する材料実験データを案件別材料データベース107から読み出し、その材料実験データを用いてオートエンコーダ108の学習を行う。
- [0036] 図6は、オートエンコーダ学習部104が案件別材料データベース107から受け取るデータの構造である。このデータは案件番号601、化合物を指定する通し番号602、化合物の構造式603を含む。
- [0037] 図7は、オートエンコーダ108の構成および学習方法の説明である。オートエンコーダはニューラルネットワークを用いた公知の次元削減器である。オートエンコーダ内部において、エンコーダと呼ばれるニューラルネットワークが高次元の入力情報を低次元化し、それを受け取ったデコーダと呼ばれる別のニューラルネットワークがそれを復元して最初の高次元入力に近い出力をする。この入出力の誤差を最小化するようにこれらのニューラルネットワークは訓練される。
- [0038] 構造式がSMILESのような文字情報として入力され出力される場合には、リカレント（再帰型）ニューラルネットワークがエンコーダおよびデコーダとして適している。本実施例では、オートエンコーダはオートエンコーダ学習部104が案件別材料データベース107から読み出した化合物の構造式情報を低次元数値ベクトルに変換するべく訓練される。なお本実施例ではオートエンコーダを用いたが、代わりに公知の変分オートエンコーダを用いても構わない。また、図7に示したオートエンコーダの構成は一例であって、ニューラルネットワークの内部構造（層の数やユニットの数）は図7で

示したものに限定されない。

- [0039] 図2に戻って、ステップS205では材料特性予測受付部105が利用者102から材料特性予測対象の材料リストを取得する。
- [0040] 図8は、材料特性予測受付部105の入力受付画面の例である。利用者はマウスおよびキーボード操作によって材料リストをアップロードすることが可能であり、OKボタンを押すことによって入力を確定し、キャンセルボタンを押すことで入力を訂正できる。
- [0041] 図9は、材料特性予測受付部105が受け取る材料リストのデータ構造である。図9に示すように、このデータは番号901と化合物の構造式情報902を含む。
- [0042] 図2に戻って、ステップS206では材料特性予測部109が材料特性予測を行い、予測結果を表示部110に出力する。
- [0043] 図10はステップS206で材料特性予測部109が行う処理のフローチャートである。図10において、まずステップS1001で、材料特性予測部109が材料特性予測受付部105から、前記図9に示した構造の予測対象材料リストを受け取る。
- [0044] ステップS1002では、材料特性予測部109が案件別材料データベース107から材料実験データを取得する。この材料実験データの構造は前記図4に示されている。
- [0045] ステップS1003では、材料特性予測部109が学習済みのオートエンコーダ108に材料実験データの構造式情報を入力することで化合物の特徴量（記述子）を生成する。そして、材料特性予測部109がこの特徴量と材料実験データの材料特性値（目的変数）とを機械学習モデルにインプットして予測モデルの訓練を行う。
- [0046] この機械学習には、例えば線形回帰や決定木、サポートベクトルマシン、ニューラルネットワーク、ランダムフォレスト、ガウス過程回帰、勾配ブースティング、ロジスティック回帰、k最近傍アルゴリズム、等の任意の機械学習アルゴリズムを用いることができる。

- [0047] ステップS 1 0 0 4では、材料特性予測部 1 0 9がまず予測対象材料リスト内の構造式をオートエンコーダ 1 0 8に入力して特徴量（記述子）を生成し、次にこれを前ステップS 1 0 0 3で訓練した予測モデルに入力して、予測対象材料に対する材料特性値の予測を行う。ステップS 1 0 0 5では、材料特性予測部 1 0 9が予測モデルの出力した材料特性予測値を表示部 1 1 0に出力する。
- [0048] 図 2に戻って、ステップS 2 0 7では表示部 1 1 0が材料特性予測結果を利用者 1 0 2に表示する。
- [0049] 図 1 1は、表示部 1 1 0の結果表示画面の例を示す。ここには予測対象の化合物とその材料特性予測値とのリストが表示され、利用者は保存ボタンによって結果を保存することができ、終了ボタンによって終了することができる。
- [0050] 図 1 2は、本実施例の利用イメージを説明する概念図である。本実施例の一つの特徴は、オートエンコーダ 1 0 8の学習にあたって、任意の材料データベースではなく利用者が選択した化合物空間に該当する材料データベースを用いる点である。
- [0051] 仮に、利用者の目的が材料物性 Aの高い材料を探索することであるとする。このとき理想的には、材料物性 Aが高い材料集合を記述するのに最も有用な一群の特徴量があると考えられ、それらを用いて機械学習を行うことが予測精度を上げるためには最も望ましい。
- [0052] さて、過去に別の案件で収集した材料データベース Xがあり、その時の利用の目的は材料物性 Bが高い材料の探索であったとする。ここで、もし、専門家知識により、材料物性 Aと Bには相関があることが知られていたならば、材料データベース Xは材料物性 Aが高い材料集合と重なりがある可能性が高いことになる。よって、材料データベース Xを用いてオートエンコーダを学習し、材料特徴量を抽出することによって、材料物性 Aの高い材料を探索しやすくなると考えられる。
- [0053] そこで本実施例での具体的な利用例を説明すると、図 1 2 (a)に示すように

、案件別材料データベース107から、案件データの一つである材料データベースX1201を抽出し、その構造式1202で、特徴量1203を得るためのオートエンコーダ108を学習する。

[0054] 案件別材料データベース107では、案件データ毎に、テキスト形式の説明文やタグを付しておき、利用者が案件データの目的やテーマ、材料の主成分あるいは格納されている材料特性等を表示および検索できるようにしてもよい。また、当該案件データの過去の利用履歴として利用目的、利用者、その他の情報を格納して表示および検索できるようにしてもよい。オートエンコーダ108は、例えばRNN（リカレントニューラルネットワーク）やDNN（ディープニューラルネットワーク）を用いる。

[0055] 図12(a)では、材料物性Aについての検討を進めようとしている利用者が、案件別材料データベース107から、過去に材料物性Bについて検討したデータである材料データベースX1201を抽出した例を示している。専門的な知識を持つ利用者は、材料物性AとBには関係があることを知っており、上述の検索機能を用いて材料データベースX1201を抽出する。材料データベースX1201は、例えば構造式と材料物性Bのデータを含んでいる。オートエンコーダ108の学習には、そのうちの構造式1202を用いる。

[0056] このようにして、構造式から適切な特徴量を得るオートエンコーダ108を準備したら、材料物性Aを推定するための予測モデルを学習する。図12(b)に示すように、教師データ1204として、構造式と測定済みの材料物性Aを含むデータを準備する。教師データ1204は、案件別材料データベース107から選択してもよいし、案件別材料データベース107以外から新しいデータを取得してもよい。

[0057] 教師データ1204から構造式1205を取得してオートエンコーダ108に入力し、特徴量1206を得る。そして、特徴量1206と材料物性Aのデータ1207の組を教師データとして用いて、予測モデル1208を学習する。材料物性Aのデータは、適当な特徴量に変換してもよい。予測モデ

ル 1 2 0 8 は例えば R N N や D N N を用い、学習は公知の教師あり学習を用いてよい。

[0058] なお、教師データ 1 2 0 4 が構造式と材料物性 A のデータ以外のデータ（例えば材料の製造条件）を含んでいる場合、それを教師データに加えてもよい。このようにして、材料物性 A を推定する予測モデル 1 2 0 8 が学習できた。予測モデル 1 2 0 8 は、材料特性予測部 1 0 9 に実装される。

[0059] その後、図 1 2 (c) に示すように、利用者は、材料物性 A を推定したい材料の構造式 1 2 0 9 を準備する。構造式 1 2 0 9 は、材料特性予測受付部 1 0 5 から入力される。構造式をオートエンコーダ 1 0 8 に入力して特徴量 1 2 1 0 を得る。特徴量 1 2 1 0 を予測モデル 1 2 0 8 に入力して、推定された材料物性 A 1 2 1 1 が得られる。

[0060] このように、制約のない一般的なオープンデータ等を用いるのではなく、現在の材料分析の目的に関連のある過去の案件で収集された材料データベースを活用することによって特徴量を生成することにより、精度の良い推定が可能になる。

[0061] 本実施例では、オートエンコーダ 1 0 8 の学習および材料特性予測部 1 0 9 での予測モデルの学習の両方に同一の案件別材料データベース 1 0 7 を用いたが、これは別々のデータベースを用いることを妨げるものではない。すなわち、材料データベース A と材料データベース B を用意し、オートエンコーダ 1 0 8 の学習には材料データベース A を用い、予測モデルの学習には材料データベース B を用いるという形態であっても構わない。この場合、材料データベース B には化合物ごとの材料特性値のデータが必要であるが、材料データベース A にはそれが必要ない。また、さらに別の形態として、オートエンコーダ 1 0 8 の学習には材料データベース A および B の両方を用い、予測モデルの学習には材料データベース B のみを用いるという形態であっても構わない。

[0062] 実施例中で生成したオートエンコーダ 1 0 8 や予測モデルは、内容を説明するテキストデータを付して、記憶装置に記憶しておいてもよい。このよう

に過去のモデルをライブラリ化しておくことにより、必要に応じた再利用が可能になる。

[0063] 以上のように、分析目的と関連のある材料データを集めた案件別材料データベースから化合物空間を利用者が選択し、オートエンコーダの学習を行うことによって、より分析目的に合致した効果的な材料特徴量が生成され、それによって材料特性の高精度な予測が容易になる。

[0064] 以上の実施例について説明したが、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、必ずしも説明した全ての構成を具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。

符号の説明

[0065] 101：材料特性予測装置、103：化合物空間指定部、104：オートエンコーダ学習部、105：材料特性予測受付部、106：実験データ受付部、107：案件別材料データベース、108：オートエンコーダ、109：材料特性予測部

請求の範囲

- [請求項1] 案件データベースを複数格納した案件別材料データベースを用いて、材料特性の予測を行うための装置であって、
- 前記案件データベースは、材料の構造に関する構造情報および材料の特性に関する材料特性を関連付けて記録したレコードを複数含み、
- 少なくとも一つの案件データベースの指定を受け付ける、化合物空間指定部と、
- 前記化合物空間指定部の受付けた案件データベースに対応する構造情報を多変量に変換するオートエンコーダを生成するオートエンコーダ学習部と、
- 前記オートエンコーダ学習部の生成したオートエンコーダが変換した多変量を用いて材料特性を予測する材料特性予測部と、
- を持つことを特徴とする材料特性予測装置。
- [請求項2] 前記オートエンコーダは、
- 前記構造情報を多変量に変換した後、前記多変量から前記構造情報を復元できるような性質を備えたモデルである、
- 請求項1記載の材料特性予測装置。
- [請求項3] 前記材料特性予測部は、
- 材料の構造に関する構造情報および材料の特性に関する材料特性を関連付けて記録したレコードを複数含む教師データを入力とし、
- 前記教師データに対応する構造情報を前記オートエンコーダに入力して多変量に変換して説明変数とし、
- 前記教師データに対応する材料特性を目的変数とし、前記説明変数と前記目的変数を用いて予測モデルを訓練する、
- 請求項1記載の材料特性予測装置。
- [請求項4] 特性を予測すべき材料の構造に関する構造情報を受け付ける材料特性予測受付部を備え、
- 前記材料特性予測部は、

前記特性を予測すべき材料の構造に関する構造情報を前記オートエンコーダに入力して多変量に変換して説明変数とし、

前記説明変数を前記予測モデルに入力して目的変数である特性を予測する、

請求項 3 記載の材料特性予測装置。

[請求項 5]

前記化合物空間指定部は、

前記案件データベースをキーワードで検索する機能を備える、

請求項 1 記載の材料特性予測装置。

[請求項 6]

材料の構造に関する構造情報を記録したレコードを複数含む第 1 のデータベースを準備する第 1 のステップ、

前記第 1 のステップで準備した第 1 のデータベースから、構造情報を抽出する第 2 のステップ、

前記第 2 のステップで抽出した構造情報を用いて、構造情報を多変量に変換するオートエンコーダを訓練する第 3 のステップ、

材料の構造に関する構造情報および材料の特性に関する材料特性を関連付けて記録したレコードを複数含む第 2 のデータベースを準備する第 4 のステップ、

前記第 4 のステップで準備した第 2 のデータベースから、構造情報を抽出する第 5 のステップ、

前記オートエンコーダを用いて、前記第 5 のステップで抽出された構造情報を多変量に変換する第 6 のステップ、

前記第 6 のステップで変換された多変量に基づいて説明変数を得、前記第 2 のデータベースから抽出した材料特性に基づいて目的変数を得る第 7 のステップ、

前記説明変数と前記目的変数を用いて、説明変数から目的変数を推定する予測モデルを生成する第 8 のステップ、

を実行する材料特性予測方法。

[請求項 7]

前記第 1 のステップでは、

案件データベースを複数格納した案件別材料データベースを用い、
該案件別材料データベースから少なくとも一つの案件データベースを
選択して前記第 1 のデータベースとする、

請求項 6 記載の材料特性予測方法。

[請求項 8]

前記案件別材料データベースでは、テキスト情報を前記案件データ
ベースに関連付けて記憶しておき、

前記第 1 のステップでは、

利用者が前記テキスト情報を検索して少なくとも一つの案件データ
ベースを選択する、

請求項 7 記載の材料特性予測方法。

[請求項 9]

前記第 1 のステップでは、

案件データベースを複数格納した案件別材料データベースを用い、
前記案件データベースは、材料の構造に関する構造情報および材料の
特性に関する材料特性を関連付けて記録したレコードを複数含み、

前記第 1 のステップでは、

前記案件別材料データベースから少なくとも一つの案件データベ
ースを選択して前記第 1 のデータベースとし、

前記第 4 のステップでは、

前記案件別材料データベースから少なくとも一つの案件データベ
ースを選択して前記第 2 のデータベースとする、

請求項 6 記載の材料特性予測方法。

[請求項 10]

前記第 1 のデータベースのレコードに含まれる材料特性と、

前記第 2 のデータベースのレコードに含まれる材料特性とは、

異なる定義の材料特性である、

請求項 9 記載の材料特性予測方法。

[請求項 11]

前記オートエンコーダに、

前記構造情報を多変量に変換した後、前記多変量から前記構造情報
を復元できるような性質を備えたモデルを用いる、

請求項 6 記載の材料特性予測方法。

- [請求項12] 特性を予測すべき材料の構造に関する構造情報を準備する第 9 のステップ、
- 前記オートエンコーダを用いて、前記第 9 のステップで準備した構造情報を多変量に変換する第 10 のステップ、
- 前記第 10 のステップで変換された多変量に基づいて説明変数を得る第 11 のステップ、
- 前記第 11 のステップで得られた説明変数を前記予測モデルに適用して、目的変数である材料特性を推定する第 12 のステップ、
- を実行する請求項 6 記載の材料特性予測方法。
- [請求項13] 前記オートエンコーダおよび前記予測モデルの少なくとも一つを記憶装置に記憶しておき、再利用を行う、
- 請求項 6 記載の材料特性予測方法。
- [請求項14] 前記第 1 のデータベースおよび前記第 2 のデータベースは、ともに材料の構造に関する構造情報および材料の特性に関する材料特性を関連付けて記録したレコードを複数含むが、材料特性に関しては、異なる定義もしくは種類のデータを記録している、
- 請求項 6 記載の材料特性予測方法。