(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7594951号
(P7594951)

(45)発行日　令和6年12月5日(2024.12.5)

(24)登録日　令和6年11月27日(2024.11.27)

(51)国際特許分類
G06N 3/08 (2023.01)

FI
G06N 3/08

請求項の数　9　（全23頁）

(21)出願番号　特願2021-45937(P2021-45937)
(22)出願日　令和3年3月19日(2021.3.19)
(65)公開番号　特開2022-144778(P2022-144778A)
(43)公開日　令和4年10月3日(2022.10.3)
審査請求日　令和5年8月28日(2023.8.28)

(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
(74)代理人　110001678
藤央弁理士法人
(72)発明者　刑部 好弘
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　淺原 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
審査官　北川 純次

最終頁に続く

(54)【発明の名称】　候補アイディアを生成するシステム

(57)【特許請求の範囲】
【請求項１】
候補アイディアを生成するシステムであって、
１以上のプロセッサと、
前記１以上のプロセッサが実行するプログラムを格納する１以上の記憶装置と、を含み、
前記１以上のプロセッサは、
第１入力アイディアベクトルを、学習済みモデルを使用して、多変量空間に属する第１縮約ベクトルに変換し、
前記第１縮約ベクトルの前記多変量空間における第１所定成分の値に基づいて、前記多変量空間から１以上の第２縮約ベクトルを抽出し、
前記１以上の第２縮約ベクトルそれぞれに基づき、学習済みモデルを使用して候補アイディアを表す１以上の第１出力アイディアベクトルを出力し、
前記学習済みモデルは、縮約ベクトルの入力アイディアベクトルと出力アイディアベクトルとの間の類似度を表す項、及び、前記第１所定成分と第１所定指標値との類似度を表す項、を含む目的関数を使用して学習済みである、システム。
【請求項２】
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、前記第１縮約ベクトルの前記第１所定成分及び第２所定成分に基づいて前記１以上の第２縮約ベクトルを抽出し、
前記目的関数は、前記第２所定成分と第２所定指標値との類似度を表す項をさらに含む、

システム。
【請求項３】
請求項１に記載のシステムであって、
前記第１所定成分と前記第１所定指標値との類似度は、前記第１所定成分と前記第１所定指標値との相関を示し、
前記入力アイディアベクトルと前記出力アイディアベクトルとの類似度は、ユークリッド距離で表される、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以上のプロセッサは、前記１以上の第２縮約ベクトルの抽出において、前記第１縮約ベクトルの前記第１所定成分に、１以上の所定バイアスを与える、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、複数のアイディアから、前記第１所定指標値と目標値との差異が所定範囲内にあるアイディアを選択し、前記選択したアイディアから前記第１入力アイディアベクトルを生成する、システム。
【請求項６】
請求項１に記載のシステムであって、
前記１以上の第２縮約ベクトルは複数の第２縮約ベクトルであり、
前記複数の第２縮約ベクトルの抽出において、前記１以上のプロセッサは、前記第１縮約ベクトルの前記第１所定成分に対して、前記複数の第２縮約ベクトルそれぞれのために異なるバイアスを与える、システム。
【請求項７】
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、提案済みのアイディアを管理し、
前記１以上の第２縮約ベクトルは複数の第２縮約ベクトルであり、
前記１以上のプロセッサは、前記提案済みのアイディアの前記多変量空間における縮約ベクトルと前記複数の第２縮約ベクトルとの間の類似度に基づいて、ユーザに提案する候補アイディアの第２縮約ベクトルを選択する、システム。
【請求項８】
請求項１に記載のシステムであって、
前記アイディアは、材料又は語句の少なくとも一方を含み、
前記１以上のプロセッサは、前記１以上の第１出力アイディアベクトルを、アイディアを表す表現に変換して表示する、システム。
【請求項９】
システムが、候補アイディアを生成する方法であって、
前記システムが、第１入力アイディアベクトルを、学習済みモデルを使用して、多変量空間に属する第１縮約ベクトルに変換し、
前記システムが、前記第１縮約ベクトルの前記多変量空間における第１所定成分の値に基づいて、前記多変量空間から１以上の第２縮約ベクトルを抽出し、
前記システムが、前記１以上の第２縮約ベクトルに基づき、学習済みモデルを使用して候補アイディアを表す１以上の第１出力アイディアベクトルを出力し、
前記学習済みモデルは、縮約ベクトルの入力アイディアベクトルと出力アイディアベクトルとの間の類似度を表す項、及び、前記第１所定成分と第１所定指標値との類似度を表す項、を含む目的関数を使用して学習済みである、方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、候補アイディアを生成するシステムに関する。
【背景技術】

【０００２】
現実世界での観測データをサイバー空間で数値化し定量的に分析することで、システムの効率化や知的生産性向上などを目指す仕組みをサイバーフィジカルシステム（ＣＰＳ）と呼ぶ。近年では、ＩｏＴ（Ｉｎｔｅｒｎｅｔ　ｏｆ　Ｔｈｉｎｇｓ）などのセンシング技術や、大量のデータを収集し分析することのできるコンピューティング技術などが成熟してきたことにより、サイバー空間とフィジカル空間をより緊密に連携させることが可能になってきている。
【０００３】
特にＣＰＳを用いたシステムの最適化は、現状の改善策を立案するフェーズと、その改善策を講じた場合に実際にどうなったか観測するフェーズを交互に行うことで実現される。観測結果、すなわちアイディアの質を評価してフィードバックすることで、次なる立案フェーズでより効果の高い改善策を提案することを目指す。
【０００４】
近年では、人工知能技術を用いた内挿的探索手法をこの立案フェーズで用いて、事前に改善確度の高いアイディアを候補施策として提示する技術が提案されている。たとえば、ベイズ最適化は既知情報から算出される期待値をもとにして、次に試すべきアイディアを割り出す手法である。非特許文献１は、材料開発において次に実験すべき候補化合物を探索するためにベイズ最適化を用いる手法を開示する。
【先行技術文献】
【特許文献】
【０００５】
【文献】特開２０２０－９５４５２号公報
【非特許文献】
【０００６】
【文献】R. Griffiths et al., “Constrained Bayesian optimization for automatic chemical design using variational autoencoders,” Chemical Science, 11(2), 577-586 (2020)
【発明の概要】
【発明が解決しようとする課題】
【０００７】
現在の人工知能技術は既知情報を基にした内挿的探索手法が主である。これは初期値に強く依存する探索手法である。つまり、過去の試行条件が偏っている場合、次に試すべき候補アイディアも過去の試行に類似しすぎてしまい、アイディアの多様性を欠くという問題がある。
【０００８】
たとえば、試行錯誤の末に発見されたアイディアの現状改善効果ならびにその経済効果が極めて高い場合、アイディアの試行に多少のコストがかかっても十分利益を回収できる。こうしたケースでは、内挿的探索手法ではなく、むしろこれまで試してこなかったアイディアを試行する方が効率は良いと思われる。しかし、単なるランダム探索等は過去の試行結果を考慮しないため、これもまた探索効率が悪いという課題がある。
【０００９】
したがって、次に質を評価すべき候補アイディアを探索するときに、候補アイディアの新規性と、既知データに基づいて予測される候補アイディアの質の両者を同時に考慮して探索する技術が望まれる。
【課題を解決するための手段】
【００１０】
本発明の一態様は、次に評価する候補アイディアを生成するシステムであって、１以上のプロセッサと、前記１以上のプロセッサが実行するプログラムを格納する１以上の記憶装置と、を含み、前記１以上のプロセッサは、アイディアを表す第１入力アイディアベクトルを、第１縮約ベクトルに変換し、前記第１縮約ベクトルが属する多変量空間において

、前記第１縮約ベクトルの前記多変量空間における第１所定成分の値に基づいて、１以上の第２縮約ベクトルを生成し、前記１以上の第２縮約ベクトルそれぞれから、ユーザに提示する候補アイディアを表す１以上の第１出力アイディアベクトルを生成する、前記多変量空間は、縮約ベクトルを生成する入力アイディアベクトルと前記縮約ベクトルから生成される出力アイディアベクトルとの類似度と、前記縮約ベクトルの第１所定成分と前記第１所定指標値の類似度とが、維持されるように構成されている。

【発明の効果】

【００１１】

本発明の一態様によれば、候補アイディアの新規性と質とに基づき、次の試行で評価するべき候補アイディアを生成できる。

【図面の簡単な説明】

【００１２】

【図１】本明細書の一実施形態に係る候補アイディア提案システムの論理構成の一例を示す。

【図２】候補アイディア生成装置、データ保存装置及びユーザインタフェースのハードウェア構成例を示す。

【図３】パラメータデータベースに含まれる、パラメータデータの例を示す。

【図４】アイディア表現データベースの構成例を示す。

【図５】アイディアベクトルデータベースの構成例を示す。

【図６】アイディアベクトル化モジュールの処理例のフローチャートである。

【図７】アイディア表現である化学構造式から、アイディアベクトルである構造式ベクトルを生成する例を示す。

【図８】縮約ベクトルデータベースの構成例を示す。

【図９】アイディア縮約モジュールの処理例を示すフローチャートである。

【図１０】変分オートエンコーダの構成例を示す。

【図１１】縮約空間探索モジュールの処理例を示すフローチャートである。

【図１２】候補アイディアデータベースの構成例を示す。

【図１３】候補アイディア生成モジュールの処理例のフローチャートである。

【図１４Ａ】実施例２に係る発想支援アプリケーションのユーザインタフェース（画面）の例を示す。

【図１４Ｂ】実施例２に係る発想支援アプリケーションのユーザインタフェース（画面）の例を示す。

【図１５】評価結果テーブルの構成例を示す

【図１６】単語ベクトルの例を示す。

【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】

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。

〔概略〕

【００１６】

以下において、次の試行で質を評価するべき候補アイディアを生成する技術が開示される。アイディアは、様々な表現形式で表され得る。材料開発における化学構造式や実験条件は、次に実験すべきアイディアの一形態である。化学構造式の表現形式は、例えば、一定の文法規則に従って記述された文字列や、行列であり得る。文法規則の例は、ＳＭＩＬＥＳ（Ｓｉｍｐｌｉｆｉｅｄ Ｍｏｌｅｃｕｌａｒ Ｉｎｐｕｔ Ｌｉｎｅ Ｅｎｔｒｙ Ｓｙｓｔｅｍ）である。以下に説明する実施例は、化学構造式を記述する文法規則の一例として、ＳＭＩＬＥＳを使用する。

【００１７】

研究開発戦略立案におけるレポート（文書）は、次に研究開発すべき指針を示すアイディアの一形態である。文書に含まれる文章、あるいは文章に対して形態素解析を行ったあとに得られるベクトル表現なども、アイディアの一形態である。

【００１８】

図１は、本明細書の一実施形態に係る候補アイディア提案システムの論理構成の一例を示す。候補アイディア提案システムは、候補アイディア生成装置Ｍ０１、データ保存装置Ｍ０２、及びユーザインタフェースＭ０５を含む。これらは、ネットワークを介して互いに通信可能である。候補アイディア生成装置Ｍ０１は、ユーザに提案する候補アイディアを生成する。候補アイディア生成装置Ｍ０１は、アイディアベクトル化モジュールＰ０１、アイディア縮約モジュールＰ０２、縮約空間探索モジュールＰ０３、及び候補アイディア生成モジュールＰ０４を含む。

【００１９】

データ保存装置Ｍ０２は、候補アイディア生成装置Ｍ０１が参照又は生成するデータを格納する。図１の構成例において、データ保存装置Ｍ０２は、各種データベースを格納する。具体的には、アイディア表現データベースＤＢ０１、アイディアベクトルデータベースＤＢ０２、縮約ベクトルデータベースＤＢ０３、候補アイディアデータベースＤＢ０４、モデルデータベースＤＢ０５、及びパラメータデータベースＤＢ０６が格納されている。データ保存装置Ｍ０２は、さらに、データベースＤＢ０１からＤＢ０６を管理する。データ管理モジュールＰ０６を含む。

【００２０】

ユーザインタフェースＭ０５は、ユーザによりデータ入力及びユーザに対する情報の提示を行う。ユーザインタフェースＭ０５は、表示部Ｐ０７及び入力受付部Ｐ０８を含む。表示部Ｐ０７は、例えば、候補アイディア生成装置Ｍ０１が生成した情報をユーザに対して提示する。入力受付部Ｐ０８は、例えば、候補アイディア生成装置Ｍ０１が処理を行うために必要な情報をユーザから受け付ける。

【００２１】

候補アイディア生成装置Ｍ０１のアイディアベクトル化モジュールＰ０１は、所定の形式で表されたアイディア表現を予め決められている所定の方法に従って実数ベクトル、すなわちアイディアベクトルに変換する。

【００２２】

アイディア縮約モジュールＰ０２は、アイディアベクトル化モジュールＰ０１からのアイディアベクトルを、別の実数ベクトル、すなわち縮約ベクトルに変換する。アイディア縮約モジュールＰ０２は、ユーザによって事前に設定される１以上の指標値の類似度が、

縮約ベクトルによって構成される多変量空間における距離と対応するように変換する。例えば、アイディア縮約モジュールＰ０２は、入力されたアイディアベクトルの類似度と、アイディアの質の指標値の類似度とが相関するように、縮約ベクトルによって構成される多変量空間を変換する。

【００２３】

アイディア縮約モジュールＰ０２は、例えば、１つの変分オートエンコーダ（ＶＡＥ）を含むことができる。ＶＡＥは、オートエンコーダの一種であり、エンコーダとデコーダの二つのニューラルネットワークから構成される深層生成モデルである。エンコーダは、入力（ベクトル）を実数ベクトルに変換する。実数ベクトルが属する空間は潜在空間と呼ばれ、所定の分布、例えば正規分布に従うことが仮定される。デコーダは、その実数ベクトルを逆変換し、入力と等しい次元のベクトルを出力する。

【００２４】

一般的なエンコーダ及びデコーダは、入力と出力とが等しくなるように訓練される（学習する）。中間出力の実数ベクトルから入力を再構成できることは、入力の十分な特徴が実数ベクトルに反映されていることを意味する。潜在空間の次元は、入力の次元よりも小さくなるように設定される。そのため、エンコーダは、入力の特徴量を抽出すると共に、入力の次元を圧縮することができる。

【００２５】

中間出力のベクトルは、潜在変数又は潜在ベクトルと呼ばれ、入力されたアイディアベクトルから抽出した特徴を表す抽象表現である。したがって、ＶＡＥの潜在ベクトルは縮約ベクトルである。潜在ベクトルは、所定の分布、例えばガウス分布に従うことが仮定される。したがって、デコーダは、例えば、ノイズが加えられたベクトルを受け取った場合に、入力されたベクトルを高い精度で復元することができる。このように、ＶＡＥは、生成モデルとして高いロバスト性を有する。

【００２６】

ＶＡＥが潜在空間を構成する学習（訓練）過程では、損失関数の出力値（損失）が小さくなるようにニューラルネットワークのパラメータが逐次的に最適化される。一般的には、入力と出力との誤差、すなわち再構成誤差を小さくするように訓練される。本明細書の一実施形態において、再構成誤差に別の項を付与した損失関数を定義する。

【００２７】

第一には、潜在ベクトルにおける１以上の所定の成分と、所定の指標値、たとえば、アイディアの質を表す指標値との類似度を示す項を、損失関数に含める。類似度を表す項は、例えば、相関を表す項や、ある特定のデータからのユークリッド距離を表す項等があり得る。例えば、相関を表す値が大きくなると損失が小さくなるような項を損失関数に付け加える。この損失関数を用いて訓練されたＶＡＥの潜在空間は、入力されたアイディアベクトルとそれに対応する出力ベクトルの類似度と、潜在ベクトルにおける所定の成分と所定の指標値との類似度の両方が維持された、多変量空間である。

【００２８】

この他にも、候補アイディアの探索目的に応じた項を損失関数に付け加えることで、所望の類似度を反映した潜在空間をＶＡＥに構成させることができる。たとえば、過去に本システムが提案した候補アイディアの実際の質の指標値（真値）が判明している場合、提案時に予測されたアイディアの質（予測値）と、指標値の真値との誤差を損失関数に加えることができる。これにより、予測値と真値との類似度を反映した潜在空間を構成することができるため、予測が大きく外れやすい領域をあらかじめ避けることが可能となる。

【００２９】

さらに、実験によって求めた真値のかわりに、別のデータセットによって訓練された機械学習システムに判定させた予測値や、物理学ないし化学に基づく理論計算によって算出された理論値を用いてもよい。

【００３０】

縮約空間探索モジュールＰ０３は、縮約ベクトルによって構成される多変量空間におい

て、目標値、たとえば探索時点で最も質の高い既知アイディアと対応する縮約ベクトルの近傍を勾配法に基づき探索することで、さらに望ましい縮約ベクトルを発見し得る。上述のように、潜在ベクトルにおける所定の成分とアイディアの質の指標値とは相関しているため、所定の成分を所定の方法で変更することで、その軸に沿った勾配法に基づく探索を実現することができる。例えば、所定の成分に対して、所定の値が加減算される。

【００３１】

候補アイディア生成モジュールＰ０４は、縮約空間探索モジュールＰ０３が発見した縮約ベクトルを受け取り、事前にユーザが設定した出力形式へと変換する。例えば、縮約ベクトルをまずアイディアベクトルへ、次にアイディア表現へと順次逆変換する。これにより、入力されたアイディア表現と同じ表現形式に復元することで、縮約ベクトルをユーザが理解できる表現形式に直す。

【００３２】

本明細書の一実施形態は、主に以下のステップを実行する。まず、システムは、ユーザ設定とデータとを受け取り、アイディアベクトル化モジュールＰ０１を実行する。続いて、システムは、アイディア縮約モジュールＰ０２を構成するＶＡＥを学習する。ＶＡＥの学習が完了すると、システムは、縮約空間探索モジュールＰ０３及び続く候補アイディア生成モジュールＰ０４によって、次に質を評価すべき候補アイディアを生成する。生成された候補合いでは、ユーザインタフェースＭ０５を介してユーザに提示される。

【００３３】

ＶＡＥの学習は本システムが稼働する際にまず行われなければならない。２回目以降の探索では、学習のステップは省略してもよい。また、アイディアの新しい試行結果が入力されたときには、そのデータを用いて追加で学習を実施してもよい。

【００３４】

本システムを運用すると、過去に提案したアイディアの質の指標値の真値を得ることができる。このとき、ユーザはユーザインタフェースＭ０５を介してその真値を本システムに入力することができる。データ管理モジュールＰ０６は、保存されているアイディアの検索機能と、候補アイディアの質の実際の指標値の入力を受け付けて保存する機能を有する。図２は、候補アイディア生成装置Ｍ０１、データ保存装置Ｍ０２、及びユーザインタフェースＭ０５のハードウェア構成例を示す。候補アイディア生成装置Ｍ０１は、演算性能を有するプロセッサＵ１１１と、プロセッサＵ１１１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ＤＲＡＭＵ１１２と、を含む。

【００３５】

候補アイディア生成装置Ｍ０１は、さらに、本システムにおける他の装置を含む他の装置とデータ通信をおこなう通信装置Ｕ１１３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Ｕ１１４と、を含む。また、候補アイディア生成装置Ｍ０１は、ユーザからの操作を受け付ける入力装置Ｕ１１５と、各プロセスでの出力結果をユーザに提示するモニタＵ１１６（出力装置の例）と、を含む。

【００３６】

例えば、候補アイディア生成装置Ｍ０１は、アイディアベクトル化モジュールＰ０１、アイディア縮約モジュールＰ０２、縮約空間探索モジュールＰ０３、及び候補アイディア生成モジュールＰ０４等のプログラムを格納する。プロセッサＵ１１１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Ｕ１１４からＤＲＡＭＵ１１２にロードされる。

【００３７】

データ保存装置Ｍ０２を構成するハードウェア要素は、候補アイディア生成装置Ｍ０１と同様でよい。例えば、データ保存装置Ｍ０２は、図１に示す各種データベースＤＢ０１からＤＢ０６及びデータ管理モジュールＰ０６等のプログラムを格納し、データ管理モジュールＰ０６を実行する。ユーザインタフェースＭ０５を構成するハードウェア要素は、候補アイディア生成装置Ｍ０１と同様でよい。例えば、ユーザインタフェースＭ０５は、

表示部Ｐ０７及び入力受付部Ｐ０８を実行する。

【００３８】

なお、複数の装置に分かれている機能を一つの装置に統合してもよく、上記複数の装置機能をさらに多くの装置に分散してもよい。このように、候補アイディア提案システムは、１以上の記憶装置及び１以上のプロセッサを含む。

【実施例１】

【００３９】

実施例１では、次に実験すべき候補化合物の構造式を探索し提示する。ユーザは、いずれかの装置の入力装置を介して、パラメータデータベースＤＢ０６を設定することができる。パラメータデータベースＤＢ０６は、候補アイディア生成装置Ｍ０１の処理に必要なデータを格納する。

【００４０】

パラメータデータベースＤＢ０６は、例えば、アイディアベクトル化モジュールＰ０１が、アイディアベクトルを生成するために参照する情報を含むことができる。一例として、ＳＭＩＬＥＳ記法によって表現された化学構造式を行列へ変換するための語彙情報を含む。また、パラメータデータベースＤＢ０は、アイディア縮約モジュールＰ０２の構成情報を含むことができる。具体的には、パラメータデータベースＤＢ０６は、アイディア縮約モジュールＰ０２の構造及び学習パラメータなどを記載する。

【００４１】

図３は、パラメータデータベースＤＢ０６に含まれる、パラメータデータの例を示す。例えば、セクション６１の「ＴａｂｌｅＮａｍｅｓ」は、ＳＭＩＬＥＳ記法によって表現された化学構造式を行列へ変換するための語彙情報を格納するテーブルを示す。セクション６２の「ＴａｒｇｅｔＰｒｏｐｅｒｔｉｅｓ」は、化合物の質を示す特性値（目的変数）についてのパラメータを示す。具体的には、ｍｗｔ及びｌｏｐＰについての特性値が示されている。これら特性値は、アイディア縮約モジュールＰ０２の学習や、縮約空間探索モジュールＰ０３による縮約ベクトルの探索において参照される。

【００４２】

セクション６３の「ＶＡＥＩｎｉｔｉａｌＰａｒａｍｓ」は、アイディア縮約モジュールＰ０２のネットワーク構成のパラメータを示す。本例のアイディア縮約モジュールＰ０２はＶＡＥを含むため、ＶＡＥの構成パラメータが示されている。例えば、「ＣｏｒｒＤｉｍＩｎｄｅｘ」は、どの次元に相関をもたせるかを規定し、「ＣｏｒｒＤｉｒ」は、正負どちらにバイアスかけるのが望ましい方向か、を規定する。

【００４３】

セクション６４の「ＳｅａｒｃｈＰａｒａｍｓ」は、縮約空間探索モジュールＰ０３による縮約ベクトルの探索において参照されるパラメータを示す。本例において、提示する候補縮約ベクトルをフィルタリングするための条件が示されている。

【００４４】

アイディア表現データベースＤＢ０１を説明する。ユーザは、ユーザインタフェースＭ０５を介して、アイディア表現データベースＤＢ０１にデータを格納することができる。アイディア表現データベースＤＢ０１は、過去の実験条件ならびに実験結果の情報を含む。アイディア表現データベースＤＢ０１は、これまで試行したアイディアを格納する。どのようなアイディアを試し、どういうＫＰＩが得られたのかが、対応付いて蓄積されている。例えば、材料開発においては、実験データが該当する。

【００４５】

図４は、アイディア表現データベースＤＢ０１の構成例を示す。図４に示す例において、ＩＤ欄Ｔ１Ｃ１はレコードの識別子を表す。ＳｕｇｇｅｓｔＩＤ欄Ｔ１Ｃ２１は、過去に本システムが提案した候補アイディアの識別子を示す。この欄Ｔ１Ｃ２１は、候補アイディアデータベースＤＢ０４のＳｕｇｇｅｓｔＩＤ欄Ｔ４Ｃ１と紐づいている。ＳｕｇｇｅｓｔＩＤ欄Ｔ１Ｃ２１において、過去に提示していない実験については空欄または空欄を示す記号「-」が入力される。このように、アイディア表現データベースＤＢ０１は、過

去に提案済みの材料（実験結果）と未提案の材料（実験結果）の双方を格納している。
【００４６】
ＴｉｍｅＳｔａｍｐ欄Ｔ１Ｃ２は実験した日付を表す。ｍｗｔ欄Ｔ１Ｃ３は分子量を示す。本例において、分子量は、改善したい目的変数の１つであって、値が大きい方が望ましい量である。ｌｏｇＰ欄Ｔ１Ｃ４は、水分配係数を示す。本例において、ｌｏｇＰは改善したい目的変数の１つであって、値が小さい方が望ましい量である。ＳＭＩＬＥＳ欄Ｔ１Ｃ５はＳＭＩＬＥＳ記法により表された化学構造式を示す。ｃａｔａｌｙｓｔ欄Ｔ１Ｃ６は触媒を示し、実験条件の１つである。ｔｅｍｐ欄Ｔ１Ｃ７は合成温度を示し、実験条件の１つである。
【００４７】
アイディア表現データベースＤＢ０１は、化合物の特性のいくつかの推定値の欄を含む。後述するように、アイディア縮約モジュールＰ０２のＶＡＥは、化合物の特性値を入力として受け取り、それら特性値の予測値を出力する。
【００４８】
具体的には、ｐｒｅｄ＿ｍｗｔ欄Ｔ１Ｃ８は本システムが過去に提案した候補アイディアのｍｗｔの予測値を示す。「ｐｒｅｄ＿ｌｏｇＰ」は本システムが過去に提案した候補アイディアのｌｏｇＰの予測値である。「ｐｒｅｄ＿ｃａｔａｌｙｓｔ」欄Ｔ１Ｃ１０は、本システムが過去に候補アイディアとして提案した際のｃａｔａｌｙｓｔの予測値を示す。「ｐｒｅｄ＿ｔｅｍｐ」欄Ｔ１Ｃ１１は、本システムが過去に提案した候補アイディアのｔｅｍｐの予測値を示す。ｐｒｅｄ＿を含む欄Ｔ１Ｃ８からＴ１Ｃ１１において、過去に提示していない実験については、空欄または空欄を示す記号「-」が入力される。
【００４９】
上述のように、図４に示す構成例は、ＩＤ及びＴｉｍｅＳｔａｍｐ(実験日)という識別情報、ＳＭＩＬＥＳという有機化合物の構造式情報、ｃａｔａｌｙｓｔ、ｔｅｍｐというプロセス条件情報、そして、ｍｗｔ、ｌｏｇＰというＫＰＩ（目的変数）を記載している。１又は複数のＫＰＩ（目的変数）を設定することができる。複数のＫＰＩを設定することで、それらＫＰＩに即した候補アイディアを提案できる。
【００５０】
次に、アイディアベクトル化モジュールＰ０１の処理を説明する。アイディアベクトル化モジュールＰ０１は、アイディア表現データベースＤＢ０１から、アイディアベクトルデータベースＤＢ０２を生成する。
【００５１】
図５は、アイディアベクトルデータベースＤＢ０２の構成例を示す。アイディアベクトルデータベースＤＢ０２は、アイディア表現データベースＤＢ０１のデータに加えて、ｉｄｅａ＿ｖｅｃｔｏｒ欄Ｔ２Ｃ５を含む。ｉｄｅａ＿ｖｅｃｔｏｒ欄Ｔ２Ｃ５は、アイディア表現のベクトル、本例において、ＳＭＩＬＥＳ記法により表現された化学構造式のベクトルを示す。
【００５２】
図６はアイディアベクトル化モジュールＰ０１の処理例のフローチャートである。アイディアベクトル化モジュールＰ０１は、パラメータデータベースＤＢ０６からパラメータデータを読み込む。続いて、アイディア表現データベースＤＢ０１からアイディア表現データを読み込む（Ｓ１０１）。
【００５３】
続いて、アイディアベクトル化モジュールＰ０１は、パラメータデータに記載された語彙情報を用いて、ＳＭＩＬＥＳ記法によって表現された化学構造式をベクトル（行列）へと変換する。なお、行列とベクトルは配列の表現方法の例であり、同一の配列をベクトル及び行列で表すことができる。この変換方法は、たとえばｏｎｅ－ｈｏｔエンコードとして広く知られている方法を用いることができる。この方法によれば、化学構造式とベクトルは１対１対応しており逆変換が可能である。
【００５４】

続いて、アイディアベクトル化モジュールＰ０１は、アイディア表現データベースＤＢ０１を複製し、ｉｄｅａ＿ｖｅｃｔｏｒ欄Ｔ２Ｃ５を追加して、対応する変換後のベクトルを格納し、アイディアベクトルデータベースＤＢ０２として書き出す。

【００５５】

図７は、アイディア表現である化学構造式から、アイディアベクトルである構造式ベクトルを生成する例を示す。アイディアベクトル化モジュールＰ０１は、ｏｎｅ－ｈｏｔエンコードによって、ＳＭＩＬＥＳ記法によって表現された化学構造式をベクトルへと変換する。

【００５６】

次に、アイディア縮約モジュールＰ０２の処理を説明する。アイディア縮約モジュールＰ０２は、アイディアベクトルデータベースＤＢ０２から、縮約ベクトルデータベースＤＢ０３を生成する。

【００５７】

図８は、縮約ベクトルデータベースＤＢ０３の構成例を示す。縮約ベクトルデータベースＤＢ０３は、は、アイディアベクトルデータベースＤＢ０２のデータに加えて、ａｂｓｔ＿ｖｅｃｔｏｒ欄Ｔ３Ｃ５を含む。ａｂｓｔ＿ｖｅｃｔｏｒ欄Ｔ３Ｃ５は、アイディアベクトルの縮約ベクトルを格納する。本例において、アイディアベクトルは、ＳＭＩＬＥＳ記法により表現された化学構造式を表すベクトルである。

【００５８】

図９は、アイディア縮約モジュールＰ０２の処理例を示すフローチャートである。アイディア縮約モジュールＰ０２は、まずパラメータデータベースＤＢ０６から、パラメータデータを読み込む（Ｓ１５１）。さらに、アイディア縮約モジュールＰ０２は、モデルデータベ-スＤＢ０５からモデルデータを読み込む（Ｓ１５２）。さらに、アイディア縮約モジュールＰ０２は、アイディアベクトルデータベースＤＢ０２から、アイディアベクトルデータを読み込む（Ｓ１５３）。

【００５９】

続いて、アイディア縮約モジュールＰ０２は、ＶＡＥモデルをインスタンス化する（Ｓ１５４）。学習済みモデルが保存されている場合には、その学習済みモデルをインスタンス化する。もし学習済みのＶＡＥが保存されていない場合、アイディア縮約モジュールＰ０２は、パラメータデータに記載されている初期値を用いて（ニューラルネットワーク）ＶＡＥを初期化する。

【００６０】

図１０は、ＶＡＥの構成例を示す。ＶＡＥのエンコーダは、例えば、外側エンコーダ２１１と、その後段の内側エンコーダ２１２で構成できる。外側エンコーダ２１１は複数の１次元畳み込み層で構成され、内側エンコーダ２１２は、複数の全結合層で構成され得る。エンコーダは、Ｍ×Ｎ次元の行列で表すことができる化学構造式のベクトル２０２と、ＳＭＩＬＥＳ以外の説明変数を連結したベクトル２０３と、を入力データ２０１として受け取る。ＶＡＥのエンコーダは、入力データ２０１を、Ｌ×１次元のベクトル２１４（以下、単にＬ次元ベクトルと記載する）に変換する。このベクトル２１４は、潜在空間２１３における縮約ベクトルである。

【００６１】

エンコーダは、外側エンコーダ２１１（畳み込み層）の最終レイヤの出力ベクトルと、アイディアベクトルデータベースＤＢ０２に含まれるＳＭＩＬＥＳ以外の説明変数を連結したベクトルと、を連結してベクトルＶ１を生成する。エンコーダは、ベクトルＶ１を畳み込み層の最終レイヤに続く内側エンコーダ２１２（全結合層）に入力して、Ｌ次元ベクトル２１４を生成する。

【００６２】

本例において、ＫＰＩであるｍｗｔ及びｌｏｇＰ並びに実験条件のｃａｔａｌｙｓｔ及びｔｅｍｐが、ベクトル２０３（ベクトルＶ１）に含まれ、エンコーダの内側エンコーダ２１２に入力される。こうすることで、特性値や実験条件も含めてエンコードすることが

できる。他の構成例において、化合物の特性推定値等が不要であれば、ＳＭＩＬＥＳ以外の説明変数は入力データに含まれていなくてもよい。

【００６３】

デコーダは、例えば、内側デコーダ２１５と、その後段の外側デコーダ２１６で構成できる。内側デコーダ２１５は複数の全結合層で構成され得、外側デコーダ２１６はＲＮＮ（Ｒｅｃｕｒｒｅｎｔ　Ｎｅｕｒａｌ　Ｎｅｔｗｏｒｋ）で構成できる。デコーダは、Ｌ次元ベクトル２１４を入力として受け取り、Ｍ×Ｎ次元の行列に対応する化学構造式を表すベクトル２２２と、化学構造式以外の説明変数からなるベクトル２２３に、逆変換する。出力される化学構造式２２２は、入力される化学構造式とわずかに異なるものであり得る。また、化学構造式２２２と合わせて、入力される化学構造式以外の説明変数の予測値２２３が出力データ２２１に含まれる。

【００６４】

エンコーダがＳＭＩＬＥＳを変換した行列以外の説明変数もエンコードしているため、全結合層の最終レイヤの出力ベクトルＶ２を分割する。これにより、復元された説明変数、例えば、ｍｗｔ、ｌｏｇＰ、ｃａｔａｌｙｓｔ、及びｔｅｍｐ（の予測値）も得ることができる。Ｖ１の結合位置とＶ２の分割位置は、一致している。

【００６５】

分割された一方は、外側デコーダ２１６に入力されて、化学構造を示すベクトル（行列）２２２が出力される。他方は、化学構造以外の入力された変数の予測値２２３として出力される。なお、図１０はアイディア縮約モジュールＰ０２が縮約ベクトルを形成するためのモデルの一例を示すものであって、設計に応じて任意のモデルを使用することができる。

【００６６】

図９に戻って、ステップＳ１５５以降のステップは、深層学習で一般に用いられているバッチ処理に基づく再帰的な処理（学習処理）が続く。つまり、学習データのうちバッチサイズによって指定されたデータ数ずつ処理される。すべてのバッチが１回ずつ処理されるまでを１エポックとよび、実際の学習は、複数エポック反復実行される。以降は１バッチで行われる処理を説明する。

【００６７】

まず、アイディア縮約モジュールＰ０２は、学習対象となるアイディアベクトルをバッチ数分だけＶＡＥに入力する（Ｓ１５５）。エンコーダの入力とデコーダの出力との誤差から、再構成誤差Ｕ０を算出する（Ｓ１５６）。再構成誤差Ｕ０の計算方法は任意であり、例えば、バッチに含まれるサンプルそれぞれのＶＡＥの入力と出力の誤差の平均値でもよい。本例のように、ＳＭＩＬＥＳだけでなく、特性値（ｍｗｔ、ｌｏｇＰ）や実験条件（ｃａｔａｌｙｓｔ、ｔｅｍｐ）も含めてＶＡＥを学習させている場合、その部分も含めた再構成誤差が算出される。

【００６８】

次に、アイディア縮約モジュールＰ０２は、アイディア縮約ベクトルにおける所定の成分と目的変数（ＫＰＩ）との相関係数を計算する（Ｓ１５７）。例えば、アイディア縮約モジュールＰ０２は、エンコーダの出力ベクトル（縮約ベクトル）の所定の第１成分とｍｗｔとの相関係数Ｒ０を計算する。次に、エンコーダの出力ベクトル（縮約ベクトル）の所定の第２成分と、ｌｏｇＰとの相関係数Ｒ１を計算する。

【００６９】

どの目的変数とどの成分との相関をとるかは、パラメータデータの「ＣｏｒｒＤｉｍＩｎｄｅｘ」で指定する。相関の計算方法は任意である。例えば、相関を算出するカーネル関数を定義してもよく、共分散と平均に基づき算出される相関係数を使ってもよく、異なる方法で算出された相関係数と重みの積和を使用してもよい。

【００７０】

続いて、アイディア縮約モジュールＰ０２は、特性値の真値と予測値との間の誤差（差分）を算出する。真値はＶＡＥへの入力であり、予測値は出力である。例えば、アイディ

ア縮約モジュールＰ０２は、アイディアベクトルデータベースＤＢ０２のｍｗｔ欄Ｔ１Ｃ３の値とｐｒｅｄ＿ｍｗｔ欄Ｔ１Ｃ８の値の差分、及びｌｏｇＰ欄Ｔ１Ｃ４の値とｐｒｅｄ＿ｌｏｇＰ欄Ｔ１Ｃ９の値の差分を算出し、それらの和Ｗ０を算出する（Ｓ１５８）。

【００７１】

和Ｗ０は、例えば、複数のサンプルの特性値の差分の和である。異なる特性の値の差分の和は、単純な和でもよいし、パラメータデータによって指定した重みを乗じたあとの和でもよい。なお、特性値（ｍｗｔ、ｌｏｇＰ）の誤差Ｗ０の算出は省略してもよい。その構成において、ＶＡＥの入力データ及び出力データは、特性値（ＫＰＩ）を含まなくてもよい。実験条件は、誤差がＬｏｓｓ関数に含まれていないので省略可能である。

【００７２】

続いて、アイディア縮約モジュールＰ０２は、再構成誤差Ｕ０と相関係数Ｒ０と相関係数Ｒ１と目的変数の真値と予測値の差分Ｗ０の和から、損失Ｌｏｓｓを計算する。Ｌｏｓｓが小さくなるようにニューラルネットワークは学習するため、適切に符号を設定する。たとえば、Ｒ０およびＲ１の符号はパラメータデータの「ＣｏｒｒＤｉｒ」であらかじめ設定される。

【００７３】

アイディア縮約モジュールＰ０２は、このＬｏｓｓに基づき、ＶＡＥは、パラメータを最適化するように更新する（Ｓ１５９）。本例においては、再構成誤差Ｕ０及び特性値誤差Ｗ０が小さくなり、相関Ｒ１、Ｒ２が大きくなるように、ＶＡＥのパラメータが更新される。なお、Ｌｏｓｓ関数における変数の符号は設計により決定され、例えば、相関が小さくなるときにＬｏｓｓ関数が小さくなるように符号が決定されてもよい。

【００７４】

次に、アイディア縮約モジュールＰ０２は、学習が終了したか判定する（Ｓ１６０）。本例では、所定のエポック数分だけ前記の学習処理を終わらせた場合に終了フラグが立つ。そうでない場合は（Ｓ１６０：ＮＯ）、ステップＳ１５５に戻って処理を続ける。

【００７５】

学習プロセスが終了したあと（Ｓ１６０：ＹＥＳ）、アイディア縮約モジュールＰ０２は、学習済みモデルのパラメータをモデルデータベースＤＢ０５に書き出し保存する（Ｓ１６１）。

【００７６】

続いて、アイディア縮約モジュールＰ０２は、学習済みのＶＡＥもう一度すべてのアイディアベクトルを入力し、エンコーダの出力を取得することで縮約ベクトルへと変換する（Ｓ１６２）。

【００７７】

続いて、アイディア縮約モジュールＰ０２は、アイディアベクトルデータベースＤＢ０２の構成にａｂｓｔ＿ｖｅｃｔｏｒ欄Ｔ３Ｃ５を追加して、縮約ベクトルを格納し、このテーブルを縮約ベクトルデータベースＤＢ０３として書き出す（Ｓ１６３）。

【００７８】

次に、縮約空間探索モジュールＰ０３の処理を説明する。図１１は、縮約空間探索モジュールＰ０３の処理例を示すフローチャートである。縮約空間探索モジュールＰ０３は、まずパラメータデータベースＤＢ０６から、パラメータデータを読み込む（Ｓ２０１）。続いて、縮約空間探索モジュールＰ０３は、縮約ベクトルデータベースＤＢ０３を読み込む（Ｓ２０２）。

【００７９】

続いて、縮約空間探索モジュールＰ０３は、パラメータデータに記載された探索条件に基づき、縮約ベクトルデータベースＤＢ０３から、探索中心となる縮約ベクトルを抽出する（Ｓ２０３）。図３に示す例では、パラメータデータのセクション６２の「ＴａｒｇｅｔＶａｌｕｅ」に、各目的変数の目標値が記載されている。

【００８０】

したがって、縮約ベクトルデータベースＤＢ０３において、ｍｗｔが１．６に近く、さ

らに、ｌｏｇＰが－１．５に近いサンプル（レコード）が、抽出される。どの程度の近さまで抽出するかを示す所定範囲は、パラメータデータの「ＦｉｌｔｅｒｉｎｇＲａｎｇｅ」で指定される。抽出する数は任意であって、パラメータデータが指定してもよい。なお、ｍｗｔ又はｌｏｇＰの少なくとも一方が目標値に近いレコードを抽出してもよく、目標値と異なる基準、例えば、ユーザに指定されたレコードが抽出されてもよい。

【００８１】

続いて、縮約空間探索モジュールＰ０３は、抽出した縮約ベクトルの第１成分及び第２成分に微小な実数を所定バイアスとして加えることにより、新しい縮約ベクトルを生成する（Ｓ２０４）。新しい縮約ベクトルの目的変数の値は、もとになる縮約ベクトルの目的変数の値とわずかに異なる。各縮約ベクトル与える、１以上の成分のバイアスからなるバイアスセットの数は、一つでも２以上でもよい。各縮約ベクトルに異なるバイアスセットを加えて、各縮約ベクトルから複数の新しい縮約ベクトルを生成することができる。

【００８２】

各成分に加えるバイアスは、予め設定された値でよく、一つ又は異なるバイアスを各抽出した縮約ベクトルに加えることができる。加えるバイアスセットは、異なる成分で共通でも異なっていてもよい。抽出された縮約ベクトル及びバイアスにより新たに生成された縮約ベクトルは、ユーザに提示する化合物（アイディア）の候補である。

【００８３】

続いて、縮約空間探索モジュールＰ０３は、縮約ベクトルデータベースＤＢ０３のＳｕｇｇｅｓｔＩＤ欄Ｔ１Ｃ２１に有効な値が格納された全て又は一部のレコードを抽出する。これらは、過去に提案済みのレコードである。

【００８４】

縮約空間探索モジュールＰ０３は、上記抽出又は生成した縮約ベクトルの各縮約ベクトルと、抽出した提案済みレコードの縮約ベクトルそれぞれとの間の類似度を計算する。類似度は、例えば、ユークリッド距離で表される。一例において、類似度の計算から成分０及び成分１の値が除かれる。これらの値が、類似度の計算に含まれていてもよい。さらに、縮約空間探索モジュールＰ０３は、算出した類似度を用いて、上記目標値に基づいて抽出した縮約ベクトル及び新たに生成した縮約ベクトルにフィルタリングをかける（Ｓ２０５）。

【００８５】

本例では候補アイディアの新規性を重視し、全ての提案済みのアイディアから所定の閾値以上離れている縮約ベクトルが残るようにフィルタリングする。閾値はパラメータデータの「ｔｈｒｅｓｈｏｌｄ」で指定する。これにより、目的変数（ＫＰＩ）が目標値近傍にあり、提案済みの化合物（アイディア）から離れた化合物（アイディア）を選択できる。

【００８６】

続いて、縮約空間探索モジュールＰ０３は、縮約ベクトルデータベースＤＢ０３のＴｉｍｅＳｔａｍｐ欄Ｔ１Ｃ２を参照し、現在日時から一定期間内のレコードを抽出する。

【００８７】

縮約空間探索モジュールＰ０３は、抽出したレコードそれぞれに含まれた縮約ベクトルと、上記フィルタリングにより選択された各候補縮約ベクトルと、の類似度を計算する。類似度は、上述のように、例えば、ユークリッド距離で表すことができる。類似度の計算は、提示済みアイディアと最近のアイディアとの間で同一でも異なっていてもよい。縮約空間探索モジュールＰ０３は、類似度を用いてフィルタリングをかけて、候補縮約ベクトルから一部を選択する（Ｓ２０６）。

【００８８】

本例では候補アイディアの時系列的な新規性を重視し、現在日時から一定期間内の全てのレコードから閾値以上離れている候補縮約ベクトルが残るようにフィルタリングする。閾値はパラメータデータの「ｔｈｒｅｓｈｏｌｄ＿Ｔｉｍｅ」で指定でき、本例では３０日間を設定する。なお、提示済みアイディアとの類似度又は最近のアイディアとの類似度の一方のみについてフィルタリングが実行されてもよく、これらフィルタリングが独立に

実行されてもよい。

【００８９】

続いて、縮約空間探索モジュールＰ０３は、フィルタリング後に残った候補縮約ベクトルを、候補アイディア生成モジュールＰ０４に渡す（Ｓ２０７）。

【００９０】

以下において、候補アイディア生成モジュールＰ０４の処理を説明する。候補アイディア生成モジュールＰ０４は、候補アイディアを生成して、候補アイディアデータベースＤＢ０４に格納する。

【００９１】

図１２は、候補アイディアデータベースＤＢ０４の構成例を示す。図１２に示す例において、ＳｕｇｇｅｓｔＩＤ欄Ｔ４Ｃ１は、提案されたアイディアのＩＤを示す。ＳｕｇｇｅｓｔＤａｔｅ欄Ｔ４Ｃ２は、提案した日付を表す。ｐｒｅｄ＿ｍｗｔ欄Ｔ４Ｃ３は、候補アイディアのｍｗｔの予測値を示す。ｐｒｅｄ＿ｌｏｇＰ欄Ｔ４Ｃ４欄は、本システムが過去に提案した候補アイディアのｌｏｇＰの予測値を示す。

【００９２】

ＳＭＩＬＥＳ欄Ｔ４Ｃ５は、本システムが過去に候補アイディアとして提案した候補化合物（候補アイディア）のＳＭＩＬＥＳ表記を示す。ｐｒｅｄ＿ｃａｔａｌｙｓｔ欄Ｔ４Ｃ６は、本システムが過去に候補アイディアとして提案した際のｃａｔａｌｙｓｔの予測値を示す。ｐｒｅｄ＿ｔｅｍｐ欄Ｔ４Ｃ７は、本システムが過去に提案した候補アイディアのｔｅｍｐの予測値を示す。欄Ｔ４１ＣからＴ４Ｃ２は、それぞれ、アイディア表現データベースＤＢ０１の同名の欄と関連づいており、同一提案の値は同一である。

【００９３】

図１３は、候補アイディア生成モジュールＰ０４の処理例のフローチャートである。候補アイディア生成モジュールＰ０４は、まずパラメータデータを読み込み（Ｓ２５１）、続いて縮約空間探索モジュールＰ０３から、候補アイディアの抽象表現である縮約ベクトルを受け取る（Ｓ２５２）。

【００９４】

続いて、候補アイディア生成モジュールＰ０４は、モデルデータベースＤＢ０５から学習済みモデルを読み込む（Ｓ２５３）。ここで読み込むモデルは、アイディア縮約モジュールＰ０２で用いたＶＡＥと同じものである。

【００９５】

続いて、候補アイディア生成モジュールＰ０４は、縮約空間探索モジュールＰ０３からからの縮約ベクトルを、学習済みＶＡＥのデコーダに入力して、アイディアベクトルである行列を得る（Ｓ２５４）。本例においては、アイディア縮約モジュールＰ０２は、ＳＭＩＬＥＳ以外の説明変数である、特性ｍｗｔ及びｌｏｇＰ並びに実験条件ｃａｔａｌｙｓｔ及びｔｅｍｐをエンコードしている。そのため、候補アイディア生成モジュールＰ０４は、上記中間出力ベクトルＶ２を分割することで、ｍｗｔ及びｌｏｇＰ並びにｃａｔａｌｙｓｔ及びｔｅｍｐの出力を得る。ＳＭＩＬＥＳのみがエンコードされている場合、この処理は実行されない。

【００９６】

続いて、候補アイディア生成モジュールＰ０４は、逆変換されたアイディアベクトルを所定の方法に従いアイディア表現に逆変換する（Ｓ２５５）。本例は、復号した行列をｏｎｅ－ｈｏｔデコードすることにより、アイディア表現であるＳＭＩＬＥＳ記法の化学構造式を得る。

【００９７】

続いて、候補アイディア生成モジュールＰ０４は、復号されたＳＭＩＬＥＳと特性値及び実験条件をまとめ、候補アイディアデータベースＤＢ０４に新しいレコードとして格納する。このとき、一意に定まる識別子をＳｕｇｇｅｓｔＩＤ欄４Ｔ１Ｃに、実行日時をＳｕｇｇｅｓｔＤａｔｅ欄Ｔ４Ｃ２にそれぞれ補う。

【００９８】

候補アイディア生成モジュールＰ０４は、候補アイディアデータベースＤＢ０４に追加した更新分のデータを、ユーザインタフェースＭ０５の表示部Ｐ０７を介して表示し、ユーザに候補アイディアを提示する。

【実施例２】

【００９９】

以下において、実施例２として文書解析技術をもとにした候補アイディア生成システムの実施例を開示する。特許文献、学術論文、社内報告書など各種文書や、各文書内の構成要素である文章も、アイディアの表現形態の１つである。本システムによれば、こうした既知のアイディアをもとに、新しい候補アイディアを生成及び提示して、ユーザの発想を支援することができる。

【０１００】

たとえば、人のアイディア整理を支援するシステムを考える。人がアイディアを整理するプロセスは、抽象的なイメージを言語化するプロセスと捉えることができる。そして、誰が読んでも理解できる言語化に成功したアイディアが、整理されたアイディアだと位置づけることができる。つまり、ユーザが頭の中に抱いている暗黙的なアイディアを説明するための語句選択を、適切にサポートする次のようなシステムを考える。

【０１０１】

図１４Ａ及び１４Ｂは、実施例２に係る発想支援アプリケーションのユーザインタフェース（画面）の例を示す。図１４Ａが示す画面Ｈ１０Ａにおけるユーザ操作の後、図１４Ｂに示す画面Ｈ１０Ｂでのユーザ操作が実行される。

【０１０２】

画面Ｈ１０Ａにおいて、オブジェクトＨ１３は、「あなたの手持ち」のキーワードを示す。オブジェクトＨ１３において、４種類のキーワードが、単語カードとして並んでいる。これらが、ユーザのアイディアを説明するための現時点での最有力な単語群である。

【０１０３】

システムは、「新単語」オブジェクトＨ１２内に、新しい単語（単語カード）を提示する。ユーザはまず、現在の手持ちカードがアイディアを言語化するうえで十分かどうかを評価し、たとえば５段階評価で「現手持ちカードの説明度」Ｈ１１が示すボタンを選択することでシステムに入力する。

【０１０４】

その後、ユーザは、手持ち単語と新単語を比較し、自分のアイディアに最も無関係な単語カードを選択し、「次へ」ボタンＨ１４を押下することで、選択した単語カードを破棄する動作を実行する。システムは必要に応じて単語カードの入れ替えを行ったのち、図１４Ｂに示す次の画面Ｈ１０Ｂを表示する。

【０１０５】

システムは、新しい単語カードを、図１４Ｂに示す画面Ｈ１０Ｂの「新単語」オブジェクトＨ１２に表示する。オブジェクトＨ１３は、「あなたの手持ち」のキーワードを示す。図１４Ａを参照して説明した操作と同様に、ユーザは、現在の手持ちカードがアイディアを言語化するうえで十分かどうかを評価し、５段階評価で「現手持ちカードの説明度」Ｈ１１が示すボタンを選択することでシステムに入力する。ユーザは、手持ち単語と新単語を比較し、自分のアイディアに最も無関係な単語カードを選択し、「次へ」ボタンＨ１４を押下することで、選択した単語カードを破棄する動作を実行する。システムは必要に応じて単語カードの入れ替えを行ったのち、不図示の次の画面を表示する。

【０１０６】

手持ち単語群と、その単語群に対するユーザ評価は、ステップごとに記録され、評価結果テーブルに格納される。図１５は、評価結果テーブルＴ６０の構成例を示す。ステップ欄Ｔ６Ｃ１は、ユーザ評価のステップの識別子を示す。

【０１０７】

単語欄Ｔ６Ｃ２は、評価ステップそれぞれの手持ち単語群を示す。この単語群を表すベクトルは、単語ベクトルの例である。この単語ベクトルは、評価ステップにおける手持ち

単語に含まれている単語の有無を０と１によって表している。
【０１０８】
評価指標欄Ｔ６Ｃ３は、評価ステップの評価結果を示す。評価指標欄Ｔ６Ｃ３は、差分評価欄Ｔ６Ｃ５及び絶対評価欄Ｔ６Ｃ６で構成されている。評価画面において入力された各評価ステップのユーザ評価は、絶対評価欄Ｔ６Ｃ６に格納される。前ステップの単語ベクトルの絶対評価と、その次の単語ベクトルの絶対評価の差が、差分評価欄Ｔ６Ｃ５に格納される。
【０１０９】
ユーザは、図１４Ａ及び１４Ｂを参照して説明した単語群の評価を繰り返すことで、自分のアイディアを説明するために適した単語カードを手持ちとして残すことができ、アイディアの言語化を進めることができる。
【０１１０】
上術の動作を実現するために、実施例２におけるアイディアベクトル化モジュールＰ０１は、単一の文章ないしは複数の文章を１単位とするアイディア表現を受け取って、実数ベクトルであるアイディアベクトルに変換する。こうした自然言語の表現を実数ベクトルに変換する方法は、たとえばＷｏｒｄ２ＶｅｃやＤｏｃ２Ｖｅｃが広く知られる。
【０１１１】
あるいは、文書中に出現する単語の出現頻度や共起確率をもとに関係性の強い語彙を抽出し、特徴量としての実数ベクトルに変換する方法が知られている。これらのような、単語と直接紐づくアイディアベクトルを、単語ベクトルと呼ぶ。図１６は、単語ベクトルＴ７０の例を示す。変換対象となる文書が１レコード（１単語ベクトル）であり、１単語ベクトルは、文書中に含まれている単語の有無を０と１によって表している。
【０１１２】
システムは、運用初期段階では、例えば、単語ベクトル間の類似度（距離）のみを頼りに、最も近い単語群を検索し、次に示す新単語を決定する。たとえば図１６の３レコードは１、成分のみ値の異なる単語ベクトル群である。現手持ち単語群が、１番目のレコードであり、システムは、類似する単語ベクトルの検索を行い、２番目と３番目のレコードを抽出したとする。システムは、ランダムにその中から３番目のレコードを選択し、差分である「分解能」を、次の新単語オブジェクトＨ１２に提示する。
【０１１３】
システムの運用が一定回数を超えると、単語ベクトルとユーザ評価の組、すなわち図１５の多数のレコードが蓄積される。図１５の評価結果テーブルＴ６０は、アイディアベクトルデータベースＤＢ０２の例である。実施例２のアイディア縮約モジュールＰ０２は、実施例１と同様に、ＶＡＥを使用してもよい。アイディア縮約モジュールＰ０２は、単語ベクトルを入力とし、エンコーダによって得られる縮約ベクトルの所定の成分とユーザ評価値との間の相関を強める損失関数を用いて、ネットワークを学習する。
【０１１４】
すると、実施例１と同様に、縮約空間探索モジュールＰ０３は、アイディアの整理に寄与する有益な単語ベクトルを検索することができる。候補アイディア生成モジュールによって新しい単語ベクトルを得ることができるため、ユーザの癖や関心対象に応じた思考整理を支援できる。
【０１１５】
ユーザ評価に変えて、単語ベクトルを生成する際のもとになった文書に対する評価値や、他ユーザからの評価など、アイディアのクオリティに関する指標値であれば、どのようなものであっても構わない。
【０１１６】
さらに、複数ユーザのアイディアベクトルデータベースを統合することもできる。他の人の思考プロセスも情報として取込みアイディア縮約モジュールに学習させることで、他の人の発想にインスパイアされた思考の遷移、つまり、発想転換の支援も可能になる。
【０１１７】

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。

【０１１８】

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。

【０１１９】

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。

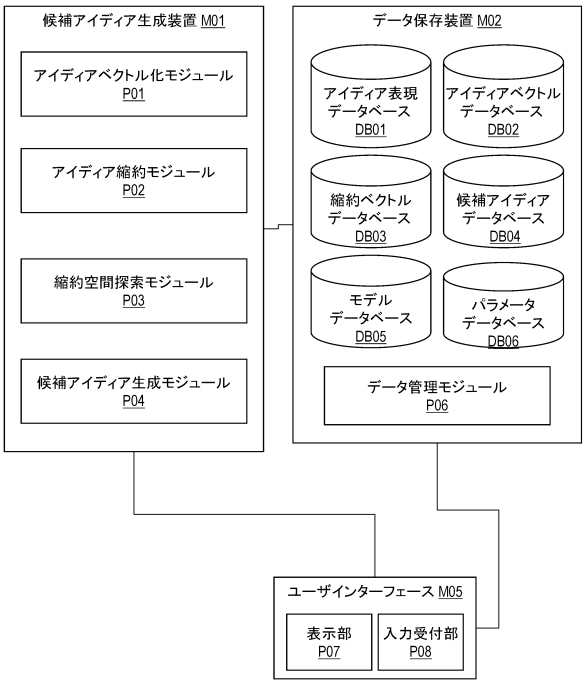
【符号の説明】

【０１２０】

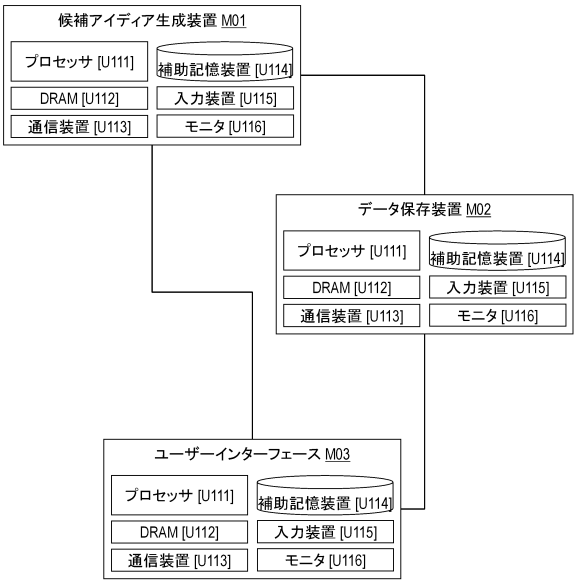
２０１　入力データ
２１１　外側エンコーダ
２１２　内側エンコーダ
２１３　潜在空間
２１４　潜在ベクトル
２１５　内側デコーダ
２１６　外側デコーダ
２２１　出力データ
ＤＢ０１　アイディア表現データベース
ＤＢ０２　アイディアベクトルデータベース
ＤＢ０３　縮約ベクトルデータベース
ＤＢ０４　候補アイディアデータベース
ＤＢ０５　モデルデータベース
ＤＢ０６　パラメータデータベース
Ｍ０１　候補アイディア生成装置
Ｍ０２　データ保存装置
Ｍ０５　ユーザインタフェース
Ｐ０１　アイディアベクトル化モジュール
Ｐ０２　アイディア縮約モジュール
Ｐ０３　縮約空間探索モジュール
Ｐ０４　候補アイディア生成モジュール
Ｐ０６　データ管理モジュール
Ｐ０７　表示部
Ｐ０８　入力受付部
Ｕ１１１　プロセッサ
Ｕ１１３　通信装置
Ｕ１１４　補助記憶装置
Ｕ１１５　入力装置
Ｕ１１６　モニタ

【図面】

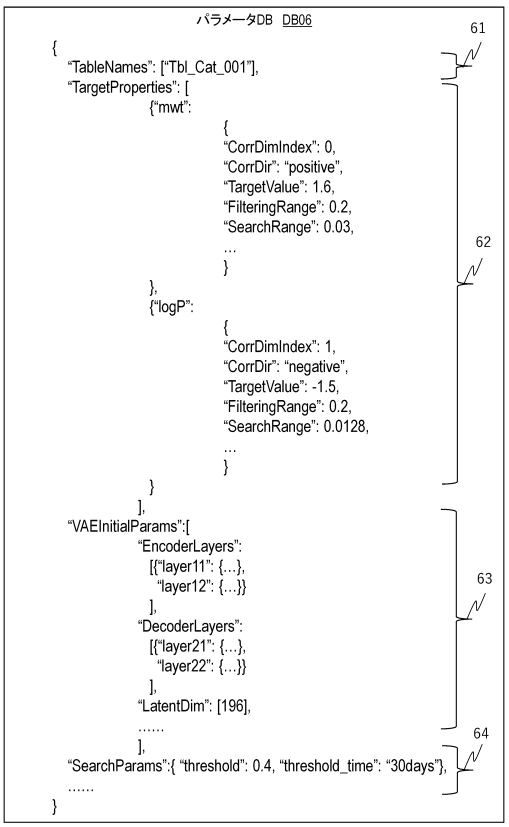
【図１】



【図２】



【図３】

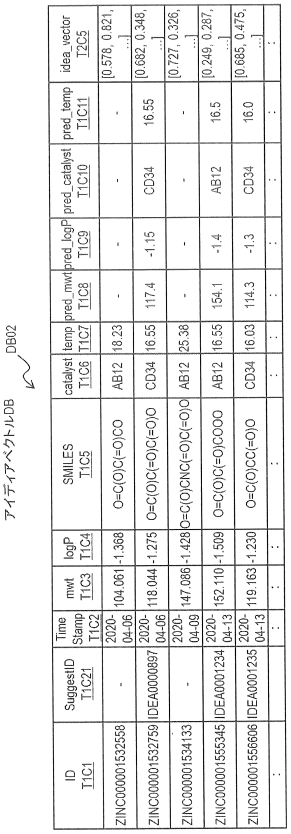


【図４】

アイディア表現DB DB01

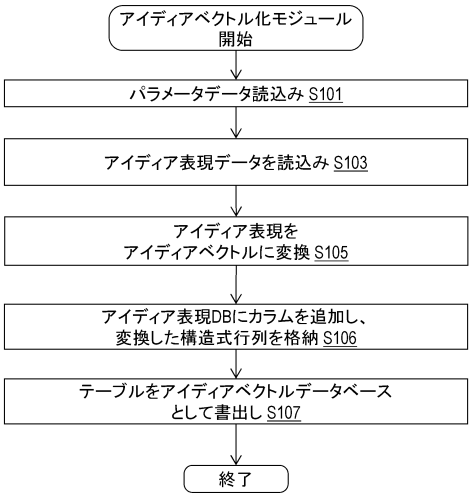
ID T1C1	SuggestID T1C21	TimeStamp T1C2	mwt T1C3	logP T1C4	SMILES T1C5	catalyst T1C6	temp T1C7	pred_mwt T1C8	pred_logP T1C9	pred_catalyst T1C10	pred_temp T1C11
ZINC000001532558	-	2020-04-06	104.061	-1.368	O=C(O)C(=O)CO	AB12	18.23	-	-	-	-
ZINC000001532759	IDEA0000897	2020-04-06	118.044	-1.275	O=C(O)C(=O)C(=O)O	CD34	16.55	117.4	-1.15	CD34	16.55
ZINC000001534133	-	2020-04-09	147.086	-1.428	O=C(O)CNC(=O)C(=O)O	AB12	25.38	-	-	-	-
ZINC000001555345	IDEA0001234	2020-04-13	152.110	-1.509	O=C(O)C(=O)COCO	AB12	16.55	154.1	-1.4	AB12	16.5
ZINC000001556606	IDEA0001235	2020-04-13	119.163	-1.230	O=C(O)CC(=O)O	CD34	16.03	114.3	-1.3	CD34	16.0
:	:	:	:	:	:	:	:	:	:	:	:

【図５】

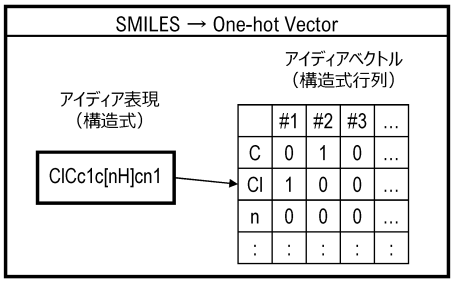


ID T1C1	SuggestID T1C21	Time Stamp T1C2	mwt T1C3	logP T1C4	SMILES T1C5	catalyst T1C6	temp T1C7	pred_mwt T1C8	pred_logP T1C9	pred_catalyst T1C10	pred_temp T1C11	idea_vector T2C5
ZINC000001532558	-	2020-04-06	104.061	-1.368	O=C(O)C(=O)CO	AB12	18.23	-	-	-	-	[0.578, 0.821, ...]
ZINC000001532759	IDEA0000897	2020-04-06	118.044	-1.275	O=C(O)C(=O)C(=O)O	CD34	16.55	117.4	-1.15	CD34	16.55	[0.682, 0.348, ...]
ZINC000001534133	-	2020-04-09	147.086	-1.428	O=C(O)CNC(=O)C(=O)O	AB12	25.38	-	-	-	-	[0.727, 0.326, ...]
ZINC000001555345	IDEA0001234	2020-04-13	152.110	-1.509	O=C(O)C(=O)COOO	AB12	16.55	154.1	-1.4	AB12	16.5	[0.249, 0.287, ...]
ZINC000001556606	IDEA0001235	2020-04-13	119.163	-1.230	O=C(O)CC(=O)O	CD34	16.03	114.3	-1.3	CD34	16.0	[0.685, 0.475, ...]
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

【図６】



【図７】



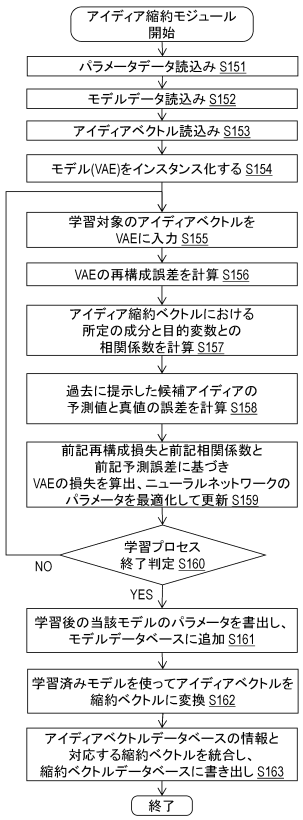
【図８】

縮約ベクトルDB DB03

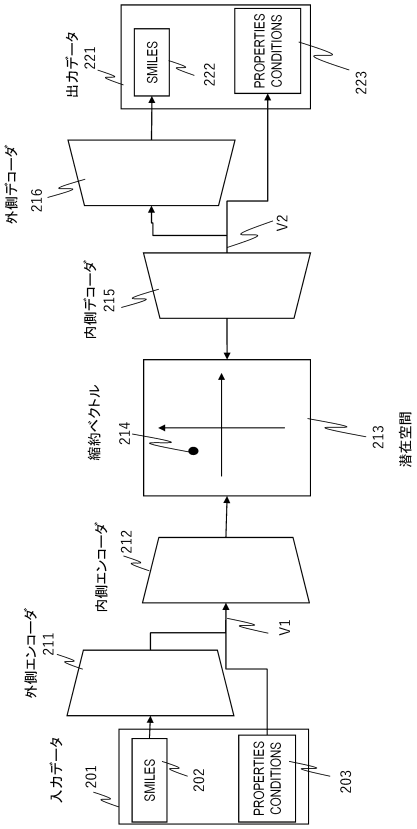
ID T1C1	SuggestID T1C21	TimeStamp T1C2	mwt T1C3	logP T1C4	SMILES T1C5	catalyst T1C6
ZINC000001532558	-	2020-04-06	104.061	-1.368	O=C(O)C(=O)CO	AB12
ZINC000001532759	IDEA0000897	2020-04-06	118.044	-1.275	O=C(O)C(=O)C(=O)O	CD34
ZINC000001534133	-	2020-04-09	147.086	-1.428	O=C(O)CNC(=O)C(=O)O	AB12
ZINC000001555345	IDEA0001234	2020-04-13	152.110	-1.509	O=C(O)C(=O)COOO	AB12
ZINC000001556606	IDEA0001235	2020-04-13	119.163	-1.230	O=C(O)CC(=O)O	CD34
⋮		⋮	⋮	⋮	⋮	⋮

temp T1C7	pred_mwt T1C8	pred_logP T1C9	pred_catalyst T1C10	pred_temp T1C11	idea_vector T2C5	abst_vector T3C5
18.23	-	-	-	-	[0.578, 0.821, ...]	[0.487,0.628, ...]
16.55	117.4	-1.15	CD34	16.55	[0.682, 0.348, ...]	[0.321, 0.542, ...]
25.38	-	-	-	-	[0.727, 0.326, ...]	[0.627, 0.597, ...]
16.55	154.1	-1.4	AB12	16.5	[0.249, 0.287, ...]	[0.498, 0.873, ...]
16.03	114.3	-1.3	CD34	16.0	[0.685, 0.475, ...]	[0.775, 0.735, ...]
⋮	⋮	⋮	⋮	⋮	⋮	⋮

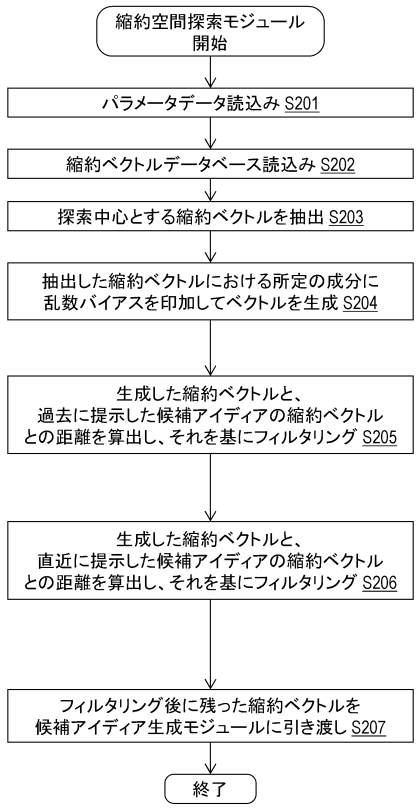
【図９】



【図１０】



【図１１】

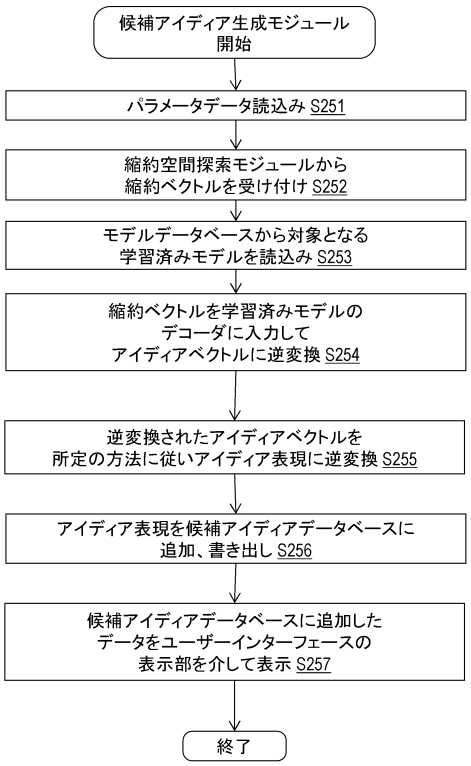


【図１２】

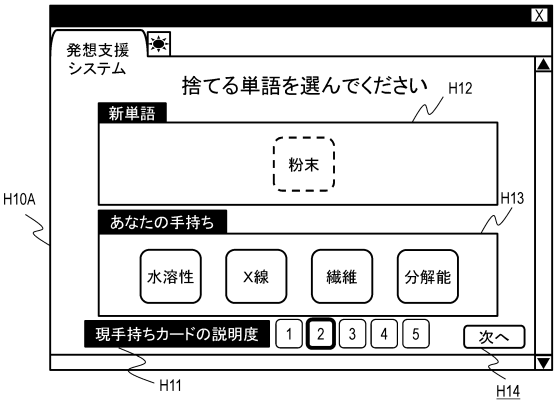
候補アイディアDB DB04

SuggestID T4C1	SuggestDate T4C2	pred_mwt T4C3	pred_logP T4C4	SMILES T4C5	pred_catalyst T4C6	pred_temp T4C7
IDEA0001234	2020-04-11	154.1	-1.4	O=C(O)C(=O)COOO	AB12	16.5
IDEA0001235	2020-04-11	114.3	-1.3	O=C(O)CC(=O)O	CD34	16.0
IDEA0001236	2020-04-12	143.9	-1.6	O=C(O)CNC(=O)O	AB12	22.5
IDEA0001237	2020-04-17	90.7	-1.2	CNC(=O)C(N)=O	CD34	26.0
IDEA0001238	2020-04-17	131.4	-1.1	NC(=O)N1CC1	CD34	28.5
：	：	：	：	：	：	：

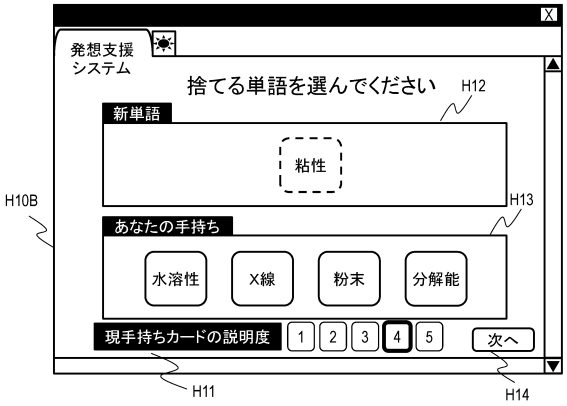
【図１３】



【図１４Ａ】



【図１４Ｂ】



【図１５】

評価結果テーブル T60

ステップ T6C1	単語 T6C2								評価指標 T6C2	
	粉末	X線	水溶性	凝固	繊維	粘性	分解能	…	差分評価 T6C5	絶対評価 T6C6
…	…	…	…	…	…	…	…	…		
1	0	1	1	0	1	0	1	…		
2	1	1	1	0	0	0	1	…		
…	…	…	…	…	…	…	…	…		

【図１６】

T70

単語ベクトル

ID	単語							
	粉末	X線	水溶性	凝固	繊維	粘性	分解能	…
1	1	1	1	0	0	1	0	…
2	1	1	0	1	0	1	0	…
3	1	1	0	0	0	1	1	…
…	…	…	…	…	…	…	…	…

フロントページの続き

(56)参考文献 特開２０２０－１８７６４２（ＪＰ，Ａ）
特開２００５－０６３３１２（ＪＰ，Ａ）
米国特許出願公開第２０１８／０３３６４９０（ＵＳ，Ａ１）
GOMEZ-BOMBARELLI, Rafael et al.，Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules，ACS central science，ACS Publications，2018年01月12日，pages 268-276

(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０６Ｎ　３／００ － ９９／００
Ｇ０６Ｑ　１０／００
Ｇ０６Ｆ　３０／００