

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7680469号
(P7680469)

(45)発行日 令和7年5月20日(2025.5.20)

(24)登録日 令和7年5月12日(2025.5.12)

(51)国際特許分類		F I	
G 0 6 N	20/00 (2019.01)	G 0 6 N	20/00
G 0 1 N	21/73 (2006.01)	G 0 1 N	21/73
G 0 1 N	21/74 (2006.01)	G 0 1 N	21/74
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00 3 5 0 B
請求項の数 15 (全30頁)			
(21)出願番号	特願2022-562597(P2022-562597)	(73)特許権者	511051982 ヴィト ナムローゼ フェンノートシャップ V I T O N V ベルギー, ベー - 2 4 0 0 モル, ボー レンタンク 2 0 0 Bo er e n t a n g 2 0 0 , B - 2 4 0 0 M o l , B e l g i u m
(86)(22)出願日	令和3年4月16日(2021.4.16)	(74)代理人	100126572 弁理士 村越 智史
(65)公表番号	特表2023-521229(P2023-521229 A)	(72)発明者	グルツ, ルーラント ベルギー モル 2 4 0 0 ブーレタンク 2 0 0 内
(43)公表日	令和5年5月23日(2023.5.23)	審査官	宮司 卓佳
(86)国際出願番号	PCT/EP2021/059942	最終頁に続く	
(87)国際公開番号	WO2021/209614		
(87)国際公開日	令和3年10月21日(2021.10.21)		
審査請求日	令和6年3月18日(2024.3.18)		
(31)優先権主張番号	20169824.8		
(32)優先日	令和2年4月16日(2020.4.16)		
(33)優先権主張国・地域又は機関	欧州特許庁(EP)		

(54)【発明の名称】 マテリアルストリームのコンポーネントを分類するための機械学習モデルをトレーニングする方法およびシステム

(57)【特許請求の範囲】
【請求項1】

複数の未知のコンポーネントを含む不均質なマテリアルストリーム中のコンポーネントの特徴付けを行うように構成された機械学習モデルをトレーニングする方法であって、
前記複数の未知のコンポーネントを含む前記マテリアルストリームの画像化を行うように構成されたセンサシステムによって、前記マテリアルストリームを走査する工程と、
前記マテリアルストリームの前記画像および/または前記マテリアルストリームの前記画像から抽出された前記複数の未知のコンポーネントの一つ以上の特徴を入力として受け取るように構成された機械学習モデルによって、前記マテリアルストリーム中の前記複数の未知のコンポーネントのそれぞれについて、一つ以上の予測ラベルおよび関連するラベル予測確率を予測する工程と、
前記マテリアルストリーム中の前記複数の未知のコンポーネント内の未知のコンポーネントそれぞれに関連付けられるトレーニング報酬を決定する工程と、
前記複数の未知のコンポーネントに関連付けられた前記トレーニング報酬に少なくとも部分的に基づいて、前記マテリアルストリーム中の前記複数の未知のコンポーネントから少なくとも一つの未知のコンポーネントを選択する工程と、を備え、
前記少なくとも一つの未知のコンポーネントのグランドトゥルースを決定するには、物理的分離における分析を必要とし、選択された前記少なくとも一つの未知のコンポーネントは、分離ユニットによって前記マテリアルストリームから物理的に分離され、前記分離ユニットは、前記選択された少なくとも一つの未知のコンポーネントを、アクセス可能な

別の区画に移動させるように構成され、

前記方法は更に、

分離された前記少なくとも一つの未知のコンポーネントのグランドトゥールスラベルを決定するために当該少なくとも一つの未知のコンポーネントを分析する工程と、

前記物理的に分離された少なくとも一つの未知のコンポーネントの決定された前記グランドトゥールスラベルを使用して、前記機械学習モデルのインクリメントバージョンをトレーニングする工程と、を備え、

化学分析に少なくとも部分的に基づいて前記グランドトゥールスラベルを決定するために、前記マテリアルストリームから分離された前記少なくとも一つの未知のコンポーネントの前記化学分析が行われる、方法。

10

【請求項 2】

前記機械学習モデルは、前記マテリアルストリームの前記画像から抽出された前記複数の未知のコンポーネントの一つ以上のユーザ定義の特徴を入力として受け取るように構成され、前記少なくとも一つの未知のコンポーネントを選択する工程における選択について、ユーザが作成した選択基準を採用する、請求項 1 に記載の方法。

【請求項 3】

前記分離ユニットは、異なる複数の分離技術を採用する複数のサブユニットを含み、前記分離ユニットは、少なくとも第 1 のサブユニットと第 2 のサブユニットとを有し、前記マテリアルストリームの前記画像から抽出された前記複数の未知のコンポーネントの一つ以上の特徴に基づいて、前記選択された少なくとも一つの未知のコンポーネントの物理的分離のために、前記第 1 のサブユニットおよび前記第 2 のサブユニットのうち的一方が選択される、請求項 1 に記載の方法。

20

【請求項 4】

前記第 1 のサブユニットは、前記マテリアルストリーム中のより小さいおよび / またはより軽いコンポーネントの物理的分離に使用され、前記第 2 のサブユニットは、前記マテリアルストリーム中のより大きいおよび / またはより重いコンポーネントの物理的分離に使用される、請求項 3 に記載の方法。

【請求項 5】

前記第 1 のサブユニットは、コンポーネントを前記アクセス可能な別の区画へと吹き飛ばすために、前記コンポーネントに向かって流体ジェットを向けて前記コンポーネントを分離するように構成され、前記第 2 のサブユニットは、機械的マニピュレーションデバイスを使用してコンポーネントを分離するように構成される、請求項 3 に記載の方法。

30

【請求項 6】

前記マテリアルストリーム中の前記複数の未知のコンポーネントそれぞれについて、質量を示すデータを算出する、請求項 1 に記載の方法。

【請求項 7】

前記流体ジェットによって発生する力は、前記選択された少なくとも一つの未知のコンポーネントの前記質量に基づいて調整される、請求項 5 に記載の方法。

【請求項 8】

前記分離ユニットによって前記マテリアルストリームから前記未知のコンポーネントを物理的に分離することの難易度を示す値が決定され、前記複数の未知のコンポーネントそれぞれに前記難易度を示す値が関連付けられ、前記マテリアルストリーム中の前記複数の未知のコンポーネントから少なくとも一つの未知のコンポーネントを選択する工程は、前記難易度を示す値に基づいて行われる、請求項 1 に記載の方法。

40

【請求項 9】

前記複数の未知のコンポーネントに関連付けられた前記トレーニング報酬に基づいて、前記マテリアルストリーム中の前記複数の未知のコンポーネントから、複数の上位の未知のコンポーネントが選択され、前記分離ユニットによる物理的分離の実行の前記難易度を示す値に基づいて、前記複数の上位の未知のコンポーネントのサブセットが物理的分離のために選択される、請求項 8 に記載の方法。

50

【請求項 10】

前記アクセス可能な別の区画は、前記分離された未知のコンポーネントの手動除去を可能にし、

前記アクセス可能な別の区画内に位置する前記分離された未知のコンポーネントについて前記機械学習モデルの内部参照の表示が提供され、

前記少なくとも一つの選択された未知のコンポーネントの前記分析は、少なくとも部分的に人間によるアノテーションによって実行される、請求項 1 に記載の方法。

【請求項 11】

前記分離された未知のコンポーネントが分析ユニットで分析され、特徴付けに基づいて前記グラントゥールスラベルを決定するために、前記分析ユニットは、前記アクセス可能な別の区画内の前記分離された未知のコンポーネントの前記特徴付けを自動的に実行するように配置される、請求項 1 に記載の方法。

10

【請求項 12】

前記分離されたコンポーネントの破壊測定に少なくとも部分的に基づいて前記グラントゥールスラベルを決定するために、前記分析ユニットは前記分離されたコンポーネントの前記破壊測定を実行するように構成される、請求項 11 に記載の方法。

【請求項 13】

前記分析ユニットは、エネルギー型または波長分散型蛍光 X 線分光法、試金法、誘導結合プラズマ発光分光法、誘導結合プラズマ原子発光分光法、誘導結合プラズマ質量分析法、レーザ誘起破壊分光法、赤外線分光法、ハイパースペクトル分光法、X 線回折分析法、走査電子顕微鏡法、核磁気共鳴法およびラマン分光法のうちの少なくとも一つを実行するように構成される、請求項 11 に記載の方法。

20

【請求項 14】

前記一つ以上の特徴は、体積、寸法、直径、形状、質感、色および偏心のうちの少なくとも一つに関するものである、請求項 1 に記載の方法。

【請求項 15】

複数の未知のコンポーネントを含む不均質なマテリアルストリーム中のコンポーネントの特徴付けを行うように構成された機械学習モデルをトレーニングするシステムであって、

前記システムは、プロセッサ、コンピュータ可読記憶媒体、センサシステムおよび分離ユニットを備え、前記コンピュータ可読記憶媒体は命令を格納しており、

30

前記命令は前記プロセッサによって実行されると、前記プロセッサに、

前記複数の未知のコンポーネントを含む前記マテリアルストリームの撮像を行うべく、前記マテリアルストリームを走査するように前記センサシステムを動作させる工程と、

前記マテリアルストリームの前記画像および/または前記マテリアルストリームの前記画像から抽出された前記複数の未知のコンポーネントの一つ以上の特徴を入力として受け取るように構成された機械学習モデルによって、前記マテリアルストリーム中の前記複数の未知のコンポーネントのそれぞれについて、一つ以上のラベルおよび関連するラベル確率を予測する工程と、

前記マテリアルストリーム中の前記複数の未知のコンポーネント内の未知のコンポーネントそれぞれに関連付けられるトレーニング報酬を決定する工程と、

40

前記複数の未知のコンポーネントに関連付けられた前記トレーニング報酬に少なくとも部分的に基づいて、前記マテリアルストリーム中の前記複数の未知のコンポーネントから少なくとも一つの未知のコンポーネントを選択する工程であって、前記少なくとも一つの未知のコンポーネントのグラントゥールスを決定するには物理的分離における分析を必要とする、前記選択する工程と、

選択された前記少なくとも一つの未知のコンポーネントを前記マテリアルストリームから物理的に分離するために前記分離ユニットを動作させる工程であって、前記分離ユニットは前記選択された未知のコンポーネントをアクセス可能な別の区画へと移動させる、前記分離ユニットを動作させる工程と、

分析を行うことにより、分離された少なくとも一つの未知のコンポーネントについて

50

決定されたグラントゥールスラベルを受け取る工程であって、分離された少なくとも一つの未知のコンポーネントについて決定されたグラントゥールスラベルをトレーニングデータベースに追加する、工程と、

前記物理的に分離された少なくとも一つの未知のコンポーネントの決定された前記グラントゥールスラベルを使用して、前記機械学習モデルのインクリメントバージョンをトレーニングする工程と、を実行させる命令が含まれ、

化学分析に少なくとも部分的に基づいて前記グラントゥールスラベルを決定するために、前記マテリアルストリームから分離された前記少なくとも一つの未知のコンポーネントの前記化学分析が行われるように構成されるシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数の未知のコンポーネントを含むマテリアルストリーム中のコンポーネントの特徴付けを行うように構成された機械学習モデルをトレーニングするための方法およびシステムに関する。本発明はさらに、コンピュータプログラム製品に関する。

【背景技術】

【0002】

効果的にデータ分類を行うことは、多くの用途において重要な役割を果たすようになっている。例えば、コンピュータビジョンアプリケーションでは、コンポーネントまたは対象物を認識するために、撮像した画像およびビデオストリームに分類器および統計モデル（例えば、機械学習モデル、回帰モデル）を適用してもよい。分類器の信頼性を保証するべく、分類器をラベル付けされた複数の例を使用してトレーニングする必要がある。このようなシステムは、データを手動でラベル付けしていることが多く、人間の労働力に依存している。

【0003】

マテリアルストリーム内の一つまたは複数のコンポーネントに対して特徴付けを行う重要な用途としては、例えば、リサイクルプロセス、農業プロセス、食品生産プロセス等が挙げられる。例えば、品質管理、価値評価、ならびに、プロセスエンジニアリングおよびプロセス制御において、特徴付けを利用できる。例えば、廃棄物処理では、特徴データが不足しているために、従来より多くの廃棄物ストリームが次善の方法でリサイクルされている。改善するには、異種の物質からなるマテリアルストリーム（例えば、大量の固形廃棄物の流れ）の適切な特徴付け技術が必要である。

【0004】

通常、マテリアルストリームの特徴付けには、人によって、例えば、特別な施設で働く工場職員によって、対象物の手動検査が行われる。この方法は、時間がかかり、主観的であり、費用もかかり、最終的には、マテリアルストリーム内の粒子（particle）に関する情報がほとんど提供されない。別の従来の方法では、サンプルが採取され、例えば、実験室で試験／分析される。このプロセスには時間がかかり（化学分析には数日、数週間～数か月かかる）、コストが増加する場合がある。さらに、マテリアルストリーム内のコンポーネント／物質／対象物の総量のごく一部のみに対して特徴付けが行われる。典型的には、物質の品質を測定するのが難しいため、多くのマテリアルストリームは最適に識別されているとは言えない。より詳細なレベルでデータを提供する、高速で客観的かつ／または自動化された方法が必要である。マテリアルストリームの特徴付けの一例として、廃棄物の特徴付けが挙げられる。

【0005】

機械学習モデルは、人間および／または実験によってラベル付け可能な大量のデータを使用してトレーニングできる統計的分類器である。このようなラベル付けは、労働集約的および／または費用のかかるプロセスになる場合がある。正確な統計システムを構築する際のボトルネックの1つは、高品質のラベル付きデータを取得するために、（手動で）ラベル付けに長い時間を費やさなくてはならないことである。通常、ラベル付けされるサン

10

20

30

40

50

ブル（新しいデータポイントを参照）は、トレーニングデータがテストセットと一致するようにランダムに選択される。

【 0 0 0 6 】

このように、機械学習モデルのトレーニングにおいてグラウンドトゥールズ（ground truth：正解）を決定することは、様々なケースおよび用途において労力と時間がかかることから、必要な労力やコストを削減しつつ、十分にトレーニングされた予測モデルを効果的に取得することが強く求められている。このような予測モデルをより効率的に取得することが望まれている。

【 発 明 の 概 要 】

【 0 0 0 7 】

本発明の目的は、上述の欠点の少なくとも一つを回避する方法およびシステムを提供することである。

【 0 0 0 8 】

これに加えてまたは代えて、本発明の目的は、複数の未知のコンポーネントを含むマテリアルストリーム中のコンポーネントの特徴付けを改善することである。

【 0 0 0 9 】

これに加えてまたは代えて、本発明の目的は、コンポーネントラベル予測モデル等の機械学習モデルのトレーニング効率を改善することである。

【 0 0 1 0 】

これに加えてまたは代えて、本発明の目的は、改善された廃棄物処理を提供することである。

【 0 0 1 1 】

本発明は、複数の未知のコンポーネントを含むマテリアルストリーム中のコンポーネントの特徴付けを行うように構成された機械学習モデルをトレーニングする方法を提供する。方法は、前記複数の未知のコンポーネントを含む前記マテリアルストリームの画像化を行うように構成されたセンサシステムによって、前記マテリアルストリームを走査する工程と、前記マテリアルストリームの前記画像および/または前記マテリアルストリームの前記画像から抽出された前記複数の未知のコンポーネントの一つ以上の特徴を入力として受け取るように構成された機械学習モデルによって、前記マテリアルストリーム中の前記複数の未知のコンポーネントのそれぞれについて、一つ以上の予測ラベルおよび関連するラベル予測確率を予測する工程と、前記マテリアルストリーム中の前記複数の未知のコンポーネントの未知のコンポーネントそれぞれに関連付けられるトレーニング報酬を決定する工程と、前記複数の未知のコンポーネントに関連付けられた前記トレーニング報酬に少なくとも部分的に基づいて、前記マテリアルストリーム中の前記複数の未知のコンポーネントから少なくとも一つの未知のコンポーネントを選択する工程であって、選択された前記少なくとも一つの未知のコンポーネントは、分離ユニットによって前記マテリアルストリームから物理的に分離され、前記分離ユニットは、選択された前記少なくとも一つの未知のコンポーネントを、アクセス可能な別の区画に移動させるように構成される、前記選択する工程と、分離された前記少なくとも一つの未知のコンポーネントのグラントゥールズラベルを決定するために当該少なくとも一つの未知のコンポーネントを分析する工程であって、前記分離された少なくとも一つの未知のコンポーネントの決定された前記グラントゥールズラベルはトレーニングデータベースに追加される、前記分析する工程と、前記物理的に分離された少なくとも一つの未知のコンポーネントの決定された前記グラントゥールズラベルを使用して、前記機械学習モデルのインクリメントバージョンをトレーニングする工程と、を備える。

【 0 0 1 2 】

トレーニング報酬は、マテリアルストリームの複数のコンポーネントに結び付けられた一つ以上のグラントゥールズラベルをデータベースに追加することによる、（機械学習）モデル/分類器の性能の向上の予測であってもよい。トレーニング報酬は、性能指標によって示される機械学習モデルの性能の上昇の予測であってもよい。例えば、アプリケー

10

20

30

40

50

ションに応じて、異なる性能指標を用いることができる。例えば、性能指標は、精度、純度、歩留まり等とすることができる。多くの異なる性能指標やスコアを使用することができる。

【 0 0 1 3 】

選択されたコンポーネントをマテリアルストリームから分離することにより、グラントゥールースラベルを決定するための分析を実行することができる。得られた情報は、トレーニングプロセス中に機械学習モデルにフィードバックすることができる。このようにして、漸進的にトレーニングされた機械学習モデルの精度を高めることができる。グラントゥールース決定のためにラベル付けされた場合、機械学習モデルによる分類（ラベル付けの予測を参照）の性能および／または精度を最大限に改善するであろう、マテリアルストリーム中のコンポーネントを１つまたは複数選択してもよい。システムは、マテリアルストリーム中の他のコンポーネントよりもモデルの学習改善をもたらすであろうコンポーネントを選択し、物理的に分離してもよい。

10

【 0 0 1 4 】

機械学習モデルを学習するための手動の／実験の労力は、能動学習、半教師付き学習、教師なし学習の技術を組み合わせることで効果的に削減することができる。本システムは、更なる分析のために１つ以上の未知のコンポーネントを物理的に分離するための分離吸値ユニットを備える。１つ以上の未知のコンポーネントの選択は、信頼度スコア、予測確率、エントロピー、特徴空間における密度等を用いて実施してもよい。

【 0 0 1 5 】

20

機械学習モデルは、トレーニングのために能動学習を採用してもよい。機械学習モデルは、学習データを能動的に選択することができる学習器と見なすことができる。物理的な能動学習は、選択された学習データを提供するために、分析されるべきマテリアルストリーム中の選択されたコンポーネントを物理的に分離するための手段を含んでもよい。能動学習では、実験、結果の学習、次の実験のコンポーネントの選択というサイクルが繰り返されるため、実験の総量を減らすことが可能である。結果の学習および次の実験のコンポーネントの選択は、コンピュータが行う。このシステムには、選択されたコンポーネントをマテリアルストリームから物理的に分離するための分離装置ユニットを備える。少ない実験回数または実験量で、多くの結果を得ることができる。この物理的な能動学習は、コスト、労力、時間のかかるコンポーネントを分析するための実験を適切に設計するための実験設計に利用することができる。

30

【 0 0 1 6 】

前記機械学習モデルは、前記マテリアルストリームの前記画像から抽出された前記複数の未知のコンポーネントの一つ以上のユーザ定義の特徴を入力として受け取るように構成されてもよい。

【 0 0 1 7 】

前記少なくとも一つの未知のコンポーネントを選択する工程における選択について、ユーザが作成した選択基準を採用してもよい。

【 0 0 1 8 】

前記分離ユニットは、異なる複数の分離技術を採用する複数のサブユニットを含んでもよい。

40

【 0 0 1 9 】

前記分離ユニットは、少なくとも第１のサブユニットと第２のサブユニットとを有し、前記マテリアルストリームの前記画像から抽出された前記複数の未知のコンポーネントの前記一つ以上の特徴に基づいて、前記選択された少なくとも一つの未知のコンポーネントの物理的分離のために、前記第１のサブユニットおよび前記第２のサブユニットのうちの一方が選択されてもよい。

【 0 0 2 0 】

複数の未知のコンポーネントのある特性に応じて、分離ユニットの適切なサブユニットを使用して、マテリアルストリームから未知のコンポーネントを分離することができる。

50

例えば、未知のコンポーネントの質量や大きさ等に応じて、異なる分離技術が必要になる場合がある。例えば、紙は流体吹き付け手段を用いて分離した方がよく、金属の塊は機械的手段を用いて分離した方がよい。未知のコンポーネントの1つ以上の特徴がマテリアルストリームの画像から抽出されるため、このデータを取得可能であり、適切なサブユニットを選択するために有利に使用することができる。

【0021】

前記第1のサブユニットは、前記マテリアルストリーム中のより小さいおよび/またはより軽いコンポーネントの物理的分離に使用され、前記第2のサブユニットは、前記マテリアルストリーム中のより大きいおよび/またはより重いコンポーネントの物理的分離に使用されてもよい。

10

【0022】

いくつかの例では、機械学習モデルは更に、マテリアルストリームから選択された未知のコンポーネントを物理的に分離するのにどの分離技術が最も適切であるかを制御することが可能である。

【0023】

前記第1のサブユニットは、コンポーネントを前記アクセス可能な別の区画へと吹き飛ばすために、前記コンポーネントに向かって流体ジェットを向けて前記コンポーネントを分離するように構成され、前記第2のサブユニットは、機械的マニピュレーションデバイスを使用してコンポーネントを分離するように構成されてもよい。

【0024】

20

前記第2のサブユニットの前記機械的マニピュレーションデバイスは、少なくとも一つのロボットアームを含んでもよい。

【0025】

前記マテリアルストリーム中の前記複数の未知のコンポーネントそれぞれについて、質量を示すデータを算出してもよい。

【0026】

必要に応じて、少なくとも低エネルギーX線画像と高エネルギーX線画像とを得るためのマルチエネルギーイメージングを行うように構成されたX線センサを含むセンサシステムにより、マテリアルストリーム中のコンポーネントの走査が行われてもよい。センサシステムによって得られた画像は、画像中の区別可能な一つ以上の対象物を分離するために分割されてもよく、セグメント化された対象物の面積を示すデータが決定される。分割された対象物のそれぞれについて、面積密度を示すデータおよび原子番号を示すデータが、低エネルギーX線画像および高エネルギーX線画像の解析によって決定されてもよい。面積密度および原子番号が既知の異なる複数のマテリアルを使用してマルチエネルギーX線イメージングを行うことにより校正されるモデルによって、面積密度および原子番号を示すデータが決定される。面積密度を示すデータおよび面積を示すデータに基づいて、セグメント化された対象物のそれぞれについて質量を示すデータが計算されてもよい。

30

【0027】

前記流体ジェットによって発生する力は、前記選択された少なくとも一つの未知のコンポーネントの前記質量に基づいて調整されてもよい。

40

【0028】

前記分離ユニットによって前記マテリアルストリームから前記未知のコンポーネントを物理的に分離することの難易度を示す値が決定され、前記複数の未知のコンポーネントそれぞれに前記難易度を示す値が関連付けられてもよい。前記マテリアルストリームから前記未知のコンポーネントを物理的に分離することの困難さを示す値に基づいて、前記マテリアルストリーム中の複数の未知のコンポーネントから前記少なくとも一つの未知のコンポーネントを選択されてもよい。

【0029】

分離の難易度をランク付けすることで、機械モデルの学習効率を大幅に向上させることができる場合がある。マテリアルストリームからの分離が困難なコンポーネントを分離す

50

ることは、グランドトゥールース分析に失敗する可能性がある。このような場合、その他の未知のコンポーネント候補を選択すること可能であったにもかかわらず選択しなかった結果として、学習性能が低下してしまうことがある。本発明では、分離の難しさも考慮することにより、このような事態を効果的に防止することができる。例えば、未知のコンポーネントを物理的に分離することが困難な場合、どの程度未知のコンポーネントが遮られているか（例えば、周りにあるものによって）が重要な役割を果たす場合がある。また、その他のコンポーネントが付着している可能性もあるため、解析に影響を与える可能性もある。いくつかの例では、分離精度および／または分離純度の予測または推定が、どの少なくとも一つの未知のコンポーネントが物理的分離のためにマテリアルストリーム中の複数の未知のコンポーネントから選択されるかに基づいて決定される。

10

【0030】

前記複数の未知のコンポーネントに関連付けられた前記トレーニング報酬に基づいて、前記マテリアルストリーム中の前記複数の未知のコンポーネントから、複数の上位の未知のコンポーネントが選択され、前記分離ユニットによる物理的分離の実行の前記難易度を示す値に基づいて、前記複数の上位の未知のコンポーネントのサブセットが物理的分離のために選択されてもよい。

【0031】

必要に応じて、トレーニング報酬は、1つ以上の報酬指標に基づいて計算されてもよい。トレーニング報酬は、ユーザによって提供されることも想定される（例えば、専門家の知識に基づく推定）。

20

【0032】

機械学習モデルまたは学習機械は、明示的にプログラムされていないタスクを実行するために、1つ以上の機械学習アルゴリズムに依存する計算上の実体として理解され得る。特に、機械学習モデルは、その挙動を環境に適応させることができる。マテリアルストリーム中のコンポーネントの特徴付けおよび検出のコンテキストにおいて、マテリアルストリームはしばしば変化する条件と要件に直面するため、この適応能力は非常に重要である。システムは、リアルタイムで動作すべく、新しい受信データを取り込むように構成されてもよい。新しいデータポイントを追加することにより、どの時点においても機械学習モデルの知識を増やすことができる。バッチモードでは、大規模なデータセットを収集し、データセット全体を一度に処理することができる。インクリメントモードでは、機械学習モデルを随時新しいデータで増強することができる（軽量で適応性が高い）。

30

【0033】

前記機械学習モデルのインクリメントバージョンは、前記分離された少なくとも一つの未知のコンポーネントの前記グランドトゥールースラベルの前記分析から定期的に得られるデータを用いて定期的にトレーニングされてもよい。

【0034】

本発明によれば、能動学習を採用することにより、ラベル付けするトレーニング例の数を大幅に減らすことができる。したがって、ラベル付けされていない例を調査し人間（例えば、少なくとも部分的に手動）および／または機械がラベル付けするために、与えられたコスト関数に関連して最も有益な例が選択的にサンプルされる。有利には、能動学習アルゴリズムは、最大の性能向上をもたらす（物理的）ラベル付けのための例を効果的に選択する方法を提供し得る。

40

【0035】

いくつかの例では、次に分析されるべきコンポーネントは、特徴空間におけるその他のコンポーネントからの（クラスタの）距離に基づいて選択される。しかしながら、その他の技術も使用できる。例えば、統計的手法を採用することができ、この手法では、将来の機械学習モデルの少なくとも一つの統計的特性（例えば、学習器分散）が最適化されるように、コンポーネントが分析のために選択され分離される。いくつかの例では、コンポーネントの選択は、分類器のアンサンブルの不一致量のレベルに基づいて実行される。マテリアルストリームにおいて特定されたコンポーネントのそれぞれに結び付けられたトレー

50

ニング報酬を決定するために、その他の手法も想定されることは理解されよう。

【 0 0 3 6 】

前記複数の未知のコンポーネントを一つ以上のクラスタに分け、当該クラスタのそれぞれが類似の特徴および／または特性を有するコンポーネントを含み、特定された未知のコンポーネントには、前記一つ以上のクラスタからの距離に少なくとも部分的に基づいて前記トレーニング報酬が割り当てられてもよい。

【 0 0 3 7 】

本発明は、例えばニューラルネットワークの実装を有する機械学習モデルにおける確実性予測によってトレーニング報酬の制御を可能にする。しかしながら、機械学習モデルによる予測ラベルの予測の不確実性／信頼性に基づいてトレーニング報酬を決定してもよいことが理解されよう。また、その他の実装も想定される。場合によっては、機械学習モデルによって以前に処理されたことのないマテリアルストリームのコンポーネントが観察されることがある。このような場合、モデルは、それがある分類（予測ラベル）に属するとかなり高い確実性で判定可能であるが、実際には、そのコンポーネントは、モデルがまだ観察していない分類に属する。これは、コンポーネントの不確実性ではなく、特徴空間におけるクラスタを分析することで克服できるかもしれない。コンポーネントが特徴空間において現在観測されている全てのクラスタから遠い場合、コンポーネントの位置またはクラスタまでの距離に基づいて選択を行うことができる。これは、異常や外れ値の検出とみなすことができる。

【 0 0 3 8 】

有利には、機械学習モデルのトレーニングにおいて多様性が効果的に考慮され得る。例えば、物理的な能動的機械学習システムは、観測された全ての未知のコンポーネント（すなわち、ラベル付けされていない観測値）にわたる多様な新しい未知のコンポーネント（すなわち、ラベル付けされていない観測値）を選択してもよい。このようにして、（物理的な）能動的機械学習システムは、より質の高いトレーニングセットを組み立てることができる。

【 0 0 3 9 】

クラスタ化アルゴリズムを使用して、異なる複数のクラスタを区別し、クラスタが互いに本質的に異なるかどうかを確認してもよい。クラスタの深部に位置していると特定されたコンポーネントは低い不確実性を有し、一方、コアから遠く離れた粒子またはいくつかのクラスタの間にある粒子は高い不確実性を有し得る。1つ以上の次元でクラスタおよびその境界をより正確に決定するように、トレーニング報酬をクラスタに結び付けることができる。

【 0 0 4 0 】

前記トレーニング報酬は、少なくとも部分的に信頼度スコアに基づいてもよい。

【 0 0 4 1 】

マテリアルストリーム中の特定されたコンポーネントに結び付けられたラベルなしデータポイントそれぞれの不確実性と多様性とに基づいて、トレーニングスコアを決定することができる。いくつかの例では、上位 n 個のスコアを有するデータポイントがバッチで選択され、ここで n はバッチサイズに対応する。バッチサイズは、ラベル付けされていないマテリアルストリームから選択されるデータポイント（成分を参照）の数として理解できる。手動および／または実験的アノテーションのために、選択されたコンポーネントをマテリアルストリーム内のその他のラベル付けされていないコンポーネントから分離してもよい。

【 0 0 4 2 】

物理的能動学習を使用して、機械学習モデルをより効率的にトレーニングすることが可能である。モデルは、手動／人間によるラベル付けデータではなく、実験的にラベル付けされたデータ（例えば、手動判定、自動化判定）を用いて、マテリアルストリーム中の複数の未知のコンポーネントから選択されたセットでトレーニングさせることができる。

【 0 0 4 3 】

10

20

30

40

50

能動学習プロセスにおいて、機械学習モデルは最初に、グラントゥールスデータの第1のセットを使用してトレーニングされてもよい。この第1のセットは、例えば、手動で生成されるまたはモデルによって自動的に生成される小さなセットであってよい。センサシステムを用いて、複数の未知のコンポーネントから1つ以上の未知のコンポーネントを、トレーニングデータを提供するための候補として認識してもよい。例えば、機械学習モデルによって現在誤って認識されている候補を予測するために、トレーニング報酬（例えば、信頼度尺度）が採用されてもよい。例えば、選択された1つ以上の未知のコンポーネントは、使用された（トレーニングされた）機械学習モデルによって認識誤りを有する可能性が高いケースに対応してもよい。その後、1つ以上の未知のコンポーネントは、分離ユニット（例えば、分離タスクを実行するための1つ以上のセンサを有するロボット構成）により物理的に分離され、グラントゥールスを決定するための更なる分析を可能にする。例えば、人間は、分離された選択された一つ以上の未知のコンポーネントを手動で確認してもよい。これに加えてまたは代えて、分離された選択された一つ以上の未知のコンポーネントを実験的に決定するために機械および/またはその他のセンサ装置を使用することができる。

10

【0044】

必要に応じて、複数のコンポーネントを同時に選択し、一度に一つのインスタンスで分離してもよい。これは、特徴空間とターゲット（ラベル/従属変数）空間との間に弱い相関がある場合に相当する。必要に応じて、分離のために複数のコンポーネントを選択することは、機械学習モデルによって予測されたラベルに基づいてもよい。

20

【0045】

必要に応じて、複数の特定された未知のコンポーネントは、選択的にサンプリングされた順序を導出するために、選択のための候補としてトレーニング報酬に基づいて順序付けられてもよい。選択的にサンプリングされた順序に従う上位の識別された未知のコンポーネントは、そのグラントゥールスラベルを決定するために分離および分析される。機械学習モデルのインクリメントバージョンは、グラントゥールスラベルに基づいてトレーニングされてもよい。

【0046】

必要に応じて、機械学習モデルは、プールベースの能動学習を採用するように構成されてもよい。機械学習モデルが、マテリアルストリーム中の特定されたコンポーネントに結び付けられたラベル付けされていないデータポイントのプールにさらされる。機械学習モデルは、グラントゥールスを決定するための少なくとも部分的に手動および/または少なくとも部分的に自動（例えば、測定装置を用いて）のアノテーションのために、マテリアルストリーム中の複数のコンポーネントのうちの一以上のコンポーネントを反復して選択するように構成されてもよい。

30

【0047】

前記アクセス可能な別の区画は、前記分離された未知のコンポーネントの手動除去を可能にし、前記アクセス可能な別の区画内に位置する前記分離された未知のコンポーネントについて前記機械学習モデルの内部参照の表示が提供され、前記少なくとも一つの選択された未知のコンポーネントの前記分析は、少なくとも部分的に人間によるアノテーションによって実行されてもよい。

40

【0048】

必要に応じて、機械学習モデルは順次クエリベースであってもよく、一度に一つのコンポーネントが選択され、更なる分析のために分離される。

【0049】

必要に応じて、機械学習モデルはバッチモードベースであってもよく、機械学習モデルを更新する前に、（例えば同時）分析のためにコンポーネントのバッチが選択され分離される。

【0050】

前記分離された未知のコンポーネントが分析装置で分析されてもよい。

50

【 0 0 5 1 】

特徴付けに基づいて前記グランドトゥールスラベルを決定するために、前記分析ユニットは、前記アクセス可能な別の区画内の前記分離された未知のコンポーネントの前記特徴付けを自動的に実行するように配置されてもよい。いくつかの例では、分離された未知のコンポーネントは、分析ユニットによって自動的に分析されてもよい。

【 0 0 5 2 】

前記分離されたコンポーネントの化学分析に少なくとも部分的に基づいて前記グランドトゥールスラベルを決定するために、前記分析ユニットは前記分離されたコンポーネントの前記化学分析を実行するように構成されてもよい。

【 0 0 5 3 】

前記分離されたコンポーネントの破壊測定に少なくとも部分的に基づいて前記グランドトゥールスラベルを決定するために、前記分析ユニットは前記分離されたコンポーネントの前記破壊測定を実行するように構成されてもよい。

【 0 0 5 4 】

前記分析ユニットは、エネルギー型または波長分散型蛍光 X 線分光 (X R F) 法、試金法、誘導結合プラズマ発光分光 (I C P - O E S) 法、誘導結合プラズマ原子発光分光 (I C P - A E S) 法、誘導結合プラズマ質量分析 (I C P - M S) 法、レーザ誘起破壊分光 (L I B S) 法、(近) 赤外線 (N I R) 分光法、ハイパースペクトル分光法、X 線回折分析 (X R D) 法、走査電子顕微鏡 (S E M) 法、核磁気共鳴 (N M R) 法およびラマン分光法のうちの少なくとも一つを実行するように構成されてもよい。また、グランドトゥールスを決定するために、測定技術を組み合わせてもよい。

【 0 0 5 5 】

前記分析ユニットは、センサシステムに関してオフラインモードで測定を実行するように構成されてもよい。また、前記分析ユニットは、分離された対象物のグランドトゥールスラベルを決定するためにバッチ処理モードで動作するように構成されてもよい。いくつかの例では、前記分析ユニットは、ある時間遅延 (例えば、数分) を伴って、ほぼリアルタイムで測定を実行するように構成される。いくつかの例において、前記分析ユニットはまた、比較的速いフィードバックを提供するように構成されてもよく、例えば、リアルタイムまたは準リアルタイムで動作する (例えば、オンライン測定) ことが理解されよう。

【 0 0 5 6 】

いくつかの例では、前記分析ユニットは、遅延測定 (例えば、非リアルタイム) を実行するように構成されてもよい。

【 0 0 5 7 】

必要に応じて、前記分析ユニットは、選択された対象物のグランドトゥールスを決定するための不連続的、周期的および / または間欠的な測定を実行するように構成される。前記分析ユニットによって実行される測定技術は、延長された、継続的なまたは比較的長期の測定プロセスを必要とする場合がある。

【 0 0 5 8 】

前記分析ユニットによって採用される一つ以上の測定技術は、時間がかかるおよび / または少なくとも部分的に破壊的であり得る準備段階を必要とする場合がある。いくつかの例では、一つ以上の測定技術は、リアルタイムで実行されない。採用される測定技術は、比較的高価であってもよいおよび / または人手を必要とする。

【 0 0 5 9 】

必要に応じて、前記分析ユニットは、非画像化測定を実行する。いくつかの例では、前記分析ユニットは、例えば、画像を生成しない等、(光学) 画像化技術を実行しない。例えば、分析ユニットは、化学分析に基づく測定を実行するように構成されてもよい。

【 0 0 6 0 】

必要に応じて、センサシステムは、少なくとも低エネルギー X 線画像と高エネルギー X 線画像とを得るためのマルチエネルギーイメージングを行うように構成された X 線センサを含んでもよい。前記センサシステムによって得られた画像は、当該画像中の 1 つまたは

10

20

30

40

50

複数の異なる未知のコンポーネントを分離するために分割される。分割された前記対象物の面積を示すデータが決定される。分割された未知のコンポーネントそれぞれについて、面積密度を示すデータ並びに原子番号および／または化学成分を示すデータが、低エネルギーX線画像および高エネルギーX線画像の解析によって決定されてもよい。面積密度および原子番号が既知の異なる複数のマテリアルを使用してマルチエネルギーX線イメージングを行うことにより校正されるモデルによって、面積密度および原子番号を示すデータが決定される。分割された対象物の面積密度を示すデータおよび面積を示すデータに基づいて、分割された未知のコンポーネントそれぞれについて質量を示すデータが計算されてもよい。

【0061】

前記X線センサは、デュアルエネルギーX線センサであってもよい。

【0062】

前記センサシステムは、分割されたオブジェクトの体積を示すデータを決定するための深度画像化ユニットをさらに含む。

【0063】

前記深度画像化ユニットは、3次元レーザ三角測量ユニットまたは3次元カメラのうちの少なくとも一方を含む。

【0064】

前記センサシステムは、分割された対象物のカラー画像を撮影するように構成されたカラー画像処理ユニットをさらに含んでもよい。

【0065】

1つ以上の分割された対象物のそれぞれについて特徴を決定する前に、前記センサシステムの異なるサブシステムからのデータはアライメントされてもよい。

【0066】

1つ以上の分割された対象物のそれぞれについて、体積、寸法、直径、形状、テクスチャ、色または偏心の少なくとも1つに関する特徴量をさらに決定してもよい。

【0067】

マテリアルストリームはコンベアで搬送され、前記マテリアルストリーム中の対象物を特徴付けるために、当該マテリアルストリームがセンサシステムによって走査される。

【0068】

デジタルツインモデルを構築するために、1つ以上の分割された対象物の特徴を保存してもよい。

【0069】

第1のデジタル識別マーカを決定するためにマテリアルストリームが搬送前に特徴付けられた後、離れた場所へ搬送され、第2のデジタル識別マーカを決定するために前記マテリアルストリームが特徴付けられ、輸送中のコンテンツの変化を決定するために前記第1および第2のデジタル識別マーカを互いに対して比較する。

【0070】

前記マテリアルストリームは、不均質であってもよい。

【0071】

マテリアルストリームは、固形廃棄物、生産物、農産物または電池からなる群から選択されてもよい。

【0072】

一般的に、マテリアルストリームにおける対象物／コンポーネントの特徴づけに対する従来のアプローチは、人が手作業で対象物を検査することである。これは、廃棄物の流れの中で行われることが多い。さらにこの作業は、小さすぎる、つまり代表的でないサンプルの表面的な目視検査によって行われることがよくある。この作業は面倒で時間がかかるだけでなく、その主観的な性質から、結果として得られる結論が必ずしも信頼できるとは限らない。品質管理はマテリアルストリーム（例えば、廃棄物の流れ）のコンポーネントの特徴づけの重要な目的であり、中古品の品質が変動すると市場の関心が低下してしまう

10

20

30

40

50

ので、このような作業は循環経済への移行を阻害するものである。本発明は、より詳細なレベルのデータを利用した、高速で客観的かつ正確な自動化手法を提供するものである。自動検査は人工知能（ＡＩ）によって行われ、マテリアルストリームのコンポーネント特徴付けを、データ駆動型かつ自動化されたアプローチへと移行させる。

【００７３】

前記機械学習モデルは、オンライン学習モデルまたは連続学習モデルであってもよく、選択された分離された未知のコンポーネント（サンプルを参照）の新規分析の各々で更新されるように構成される。分析は、例えば、ユーザによる分類（例えば、少なくとも部分的に手動）または分析ユニット（例えば、実験的決定）を用いた完全自動化で実行されてもよい。グラントゥールスラベルを決定するための自動解析と手動解析との組み合わせも想定される。

10

【００７４】

必要に応じて、深層学習機械学習モデルが採用される。深層学習とは表現学習方法を採用した機械学習技術のクラスであり、機械に生データを与えてデータ分類に必要な表現を決定することを可能にしている。深層学習は、深層学習機械モデルの内部パラメータ（例えば、ノード重み）を変更するのに使用される逆伝播アルゴリズムを用いて、データセットにおける構造を確認することができる。深層学習機械は、様々な多層アーキテクチャおよびアルゴリズムを利用してもよい。

【００７５】

ニューラルネットワーク環境での深層学習は、ニューロンと呼ばれる多数の相互接続されたノードを含み得る。外部から活性化された入力ニューロンは、ニューラルネットワークのパラメータによって支配されるその他のニューロンへの接続に基づいて、その他のニューロンを活性化する。ニューラルネットワークは、それ自身のパラメータに基づいてある種の振る舞いを行うことができる。深層学習モデルをトレーニングすることで、ニューラルネットワークが望ましい振る舞いを行うように、ネットワーク内のニューロン間の接続を表すモデルパラメータを改良する（例えば、マテリアルストリーム中のコンポーネントを分類する等、意図するタスクでより良い振る舞いを行うようにする）。

20

【００７６】

深層学習は、多くのデータセットが、低レベルの特徴（例えば、エッジ）から高レベルの特徴（例えば、パターン、オブジェクト等）までの特徴の階層を含んでいるという理解に基づいて動作する。例えば、画像を調べている時に、対象物を探すのではなく、モデルは探している対象物を形成するモチーフを形成するエッジを探すようになる。学習した観測可能な特徴とは、機械学習モデルが学習したオブジェクトおよび定量的な規則性等である。機械学習モデルは、良好に分類された大量のデータを与えられると、新しいデータをうまく分類するのに必要な特徴を区別して抽出することができるようになる。

30

【００７７】

前記機械学習モデルは、畳み込みニューラルネットワーク（CNN）を利用してもよい。いくつかの例では、深層学習は、データ内の学習された観測可能な特徴を見つけ出し識別するために、畳み込みニューラルネットワークのセグメント化を利用できる。CNNアーキテクチャの各フィルタまたは層は、入力データを変換して、データの（特徴）選択性および頑健性を向上させることができる。このようにデータを抽象化することで、機械は分類しようとしているデータ中の特徴に集中し、無関係な背景情報を無視することができる。畳み込みニューラルネットワーク（CNN）を用いた深層学習機械モデルは、画像解析に利用することができる。

40

【００７８】

本発明の別の側面によれば、複数の未知のコンポーネントを含むマテリアルストリーム中のコンポーネントの特徴付けを行うように構成された機械学習モデルをトレーニングするシステムを提供する。前記システムは、プロセッサ、コンピュータ可読記憶媒体、センサシステムおよび分離ユニットを備え、前記コンピュータ可読記憶媒体は命令を格納しており、前記命令は前記プロセッサによって実行されると、前記プロセッサに、前記複数の

50

未知のコンポーネントを含む前記マテリアルストリームの撮像を行うべく、前記マテリアルストリームを走査するように前記センサシステムを動作させる工程と、前記マテリアルストリームの前記画像および／または前記マテリアルストリームの前記画像から抽出された前記複数の未知のコンポーネントの１つ以上の特徴を入力として受け取るように構成された機械学習モデルによって、前記マテリアルストリーム中の前記複数の未知のコンポーネントのそれぞれについて、一つ以上のラベルおよび関連するラベル確率を予測する工程と、前記マテリアルストリーム中の前記複数の未知のコンポーネント内の未知のコンポーネントそれぞれに関連付けられるトレーニング報酬を決定する工程と、前記複数の未知のコンポーネントに関連付けられた前記トレーニング報酬に少なくとも部分的に基づいて、前記マテリアルストリーム中の前記複数の未知のコンポーネントから少なくとも一つの未知のコンポーネントを選択する工程と、選択された前記少なくとも一つの未知のコンポーネントを前記マテリアルストリームから物理的に分離するために前記分離ユニットを動作させる工程であって、前記分離ユニットは前記選択された未知のコンポーネントにアクセス可能な別の区画へと移動させる、前記分離ユニットを動作させる工程と、前記分離された少なくとも一つの未知のコンポーネントについて、分析を行うことにより前記決定されたグラントゥールスラベルを受け取る工程であって、前記分離された少なくとも一つの未知のコンポーネントについて前記決定されたグラントゥールスラベルをトレーニングデータベースに追加する、前記受け取る工程と、物理的に分離された少なくとも一つの未知のコンポーネントについて決定されたグラントゥールスラベルを使用して、機械学習モデルのインクリメントバージョンをトレーニングする工程と、を実行させる命令が含まれる。

10

20

【 0 0 7 9 】

トレーニングセットに含めるべきマテリアルストリームの特定のコンポーネントを選択する方法は、特定された各コンポーネントをトレーニングセットに含めることによって得られる「報酬」の推定値（性能上昇の推定値）に基づいてもよい。そして、機械学習モデルをさらにトレーニングするために、マテリアルストリームの選択された特定のコンポーネントを分離して分析してもよい。報酬は、マテリアルストリームのラベル付けされていないコンポーネントに関連する不確実性に基づいてもよい。しかしながら、トレーニング報酬は、特徴空間におけるコンポーネントの特定されたクラスタリングに基づいてもよい。

【 0 0 8 0 】

30

能動学習は、アルゴリズムが新しいデータポイントについて所望の出力（例えば、マテリアル特性、マテリアルの種類、マテリアルの特徴、化学分析、色、形状特性、質量、密度等の少なくとも一つ）を得るために情報源に対話的に問い合わせることができる機械学習の特定の分野である。本発明によって提供される物理的能動学習では、分離装置ユニットが、一つ以上の新しいデータポイントを提供する更なる分析のために、マテリアルストリーム中の一つ以上の選択された未知のコンポーネントを物理的に分離するために使用される。物理的能動学習モデルは、入力データポイントの「最適性」を示すトレーニング報酬（例えば、加重スコア）に従って、行うべき測定を決定することが可能である。いくつかの例では、このトレーニング報酬は、入力データ情報のみを用いて決定および／または計算されてもよい。

40

【 0 0 8 1 】

ユーザ（例えば、専門家またはオペレータ）は、分離および解析のために、コンポーネントの選択に追加の基準を課してもよい。いくつかの例では、トレーニング報酬は常に計算されるわけではないが、ユーザによって仮定されてもよい。所定の（例えば、経験／知識ベースの）仮定に基づいて機械学習モデルをトレーニングするためのグラントゥールスを決定するために、コンポーネントが選択され分離および分析されてもよい。マテリアルストリーム中のコンポーネントの物理的分離は、例えば、コンポーネントの特性（例えば、形状、密度、．．．）に少なくとも部分的に基づいて実行することができる。

【 0 0 8 2 】

いくつかの例では、センサシステムは、２Ｄ－カメラ、３Ｄ－カメラ、Ｘ線撮像システ

50

ム等の1つ以上の撮像装置を含む。また、コンピュータ断層撮影（ＣＴ）システム、磁気共鳴画像（ＭＲＩ）システム等のその他の撮像装置を使用することも可能である。また、複数の撮像装置を組み合わせ使用することも可能である。例えば、３ＤカメラシステムとＸ線システムとの組み合わせも可能である。

【００８３】

一態様によれば、本発明は、グラントゥールスラベルを決定するべくさらなる分析のために、マテリアルストリーム中の1つ以上のコンポーネントを選択的に分離する手段を有するシステムおよび方法を提供する。1つ以上のコンポーネントはその特定の特徴に基づいて選択され、グラントゥールスラベルは、マテリアルストリーム中のコンポーネントそれぞれについて予測ラベルを得るために使用する機械学習モデルの能動学習トレーニングに使用される。いくつかの例では、グラントゥールス決定のための分離および更なる分析のために1つ以上のコンポーネントを選択することは、例えば、ユーザによって提供される選択基準（例えば、高密度のコンポーネント、色等のいくつかの視覚特性を有するコンポーネント、特定の形状を有するコンポーネント等の選択）に基づいてもよい。例えば、専門家は、選択されたコンポーネントについてグラントゥールスを決定することが、モデルに大きなトレーニング報酬を与えるか否かを推定することができる。有利には、機械学習モデルのトレーニングに使用される選択されたコンポーネントの物理的分離を提供するセンサを備えた分離装置を得てもよい。

10

【００８４】

機械学習モデルは、ランダムに選ばれた複数の例から学習する（受動学習）のではなく、ラベル付けされる例に動作するようにすることができ、これは能動学習と見なすことができる。能動学習を用いると、トレーニングデータのサブセットを用いて、より良い性能を得ることができる。本発明は、マテリアルストリームからコンポーネントを物理的に分離するように配置された分離装置ユニットが提供される物理的能動学習を採用する。

20

【００８５】

機械学習モデルは、コンピュータの処理能力を使用してアルゴリズムを実行し、データの挙動または特性の予測因子を学習してもよいことが理解されよう。機械学習技術は、特定の特性／特徴を示すことが知られているコンポーネントのセット等、既知の分類またはラベルを有するトレーニングサンプルのセット（トレーニングセット）に対してアルゴリズムを実行し、未知のコンポーネントが特定の分類またはグループに属するかどうか等、未知のものの動作または特徴を予測する特徴を学習してもよい。

30

【００８６】

ラベル付けは、様々な方法で様々なエンティティによって実行され得ることが理解されよう。例えば、ラベル付けは、機械学習モデルによって実行され得る（すなわち、予測ラベルを提供する）。一方で、ラベル付けは、グラントゥールスラベルを決定するために実行することもできる（例えば、分析者、人間の注釈者、実験セットアップ等によって実施される）。

【００８７】

トレーニング報酬は、トレーニング報酬または能動的報酬と見なすことができることが理解されるであろう。グラントゥールスでラベル付けされ機械学習モデルのトレーニングに使用された場合に、その性能がどの程度向上するかについての予測および／または示唆として理解することができる。トレーニング報酬は、マテリアルストリームの選択されたコンポーネントに関連する決定されたグラントゥールスラベルを使用してトレーニングすることによる機械学習モデルの改善を示してもよい。トレーニング報酬は、機械学習プロセスにおける学習報酬として理解することができる。

40

【００８８】

様々な能動学習技術を実装できることが理解されよう。能動学習技法は、トレーニングセットを選択する際に、知識または「ノウハウ」の最大利得を提供するアクションを選択するように構成されてもよい。能動学習技術は、「知識」および知識の利得が定量化される方法に関して様々なあってもよい。また、どのアクションが最大の知識獲得につながり

50

やすいかを決定する方法も様々であってもよい。様々な実装が可能である。

【 0 0 8 9 】

方法の観点で説明された態様、特徴およびオプションは何れも、本発明のシステムおよび説明されたりサイクル装置に等しく適用されることが理解されるであろう。上記の態様、特徴、およびオプションの何れか一つ以上の組み合わせが可能であることも明らかとなるであろう。

【図面の簡単な説明】

【 0 0 9 0 】

本発明を、図面に示される例示的な各実施形態に基づいて更に説明する。例示的な各実施形態は、非限定的な例示として与えられる。各図面は本発明の実施形態を模式的に示したものに過ぎず、非限定的な例として与えられることに留意すべきである。

【 0 0 9 1 】

【図 1】システムの一実施形態の模式図を示す。

【図 2】システムの一実施形態の模式図を示す。

【図 3】方法の一実施形態の模式図を示す。

【図 4】特徴空間の一例を示す。

【図 5】複数の異なるコンポーネント種類の特徴の分布を示す。

【図 6】例示的な学習プロセス指標を示す。

【図 7】システムの模式図を示す。

【図 8】方法の模式図を示す。

【発明を実施するための形態】

【 0 0 9 2 】

教師あり機械学習では、オブジェクト各々にラベルが付された（大きな）マテリアルストリームでモデルのトレーニングを行う。ラベルは、マテリアルストリーム中で識別されたコンポーネント / 対象物それぞれの材質の種類（金属、木材、ガラス、セラミック等）を示してもよく、マテリアルストリーム内のコンポーネント / 対象物をどのように正しく分類すべきか機械学習モデルが学習するために使用できる。複数の異種の物質からなるストリームから何千ものコンポーネントまたは粒子を丹念に選択することは時間とコストのかかる作業であるため、ラベル付きデータの決定および / または準備が、トレーニングプロセスのボトルネックとなることが多い。マテリアルストリーム中には、ラベルの付いていないデータが豊富にあり容易に取得できるが、ラベルの付いたデータは少なく取得が困難な場合がある。さらに、新しいマテリアルストリームを考えるたびに、ラベル付けプロセス全体を最初から最後まで繰り返さなければならない場合がある。本発明は、トレーニングデータセット全体を使用する教師ありモデルと同等の精度を実質的に確保しながら、ラベル付コストを大幅に削減できる、マテリアルストリーム中のコンポーネントのデータ主導型の特徴付けを採用するものである。能動学習を採用することにより、機械学習モデル自身が、マテリアルストリーム中のラベル付けが必要なコンポーネント（物体、粒子等）の最適な小さな部分集合（サブセット）を選択することができるようになる。このラベル付けされた小さな部分集合のみを使って機械学習モデルをトレーニングさせると、マテリアルストリーム中のコンポーネント全てを使ってトレーニングした場合に匹敵する性能を持つモデルが得られる。

【 0 0 9 3 】

図 1 は、未知のコンポーネント 3 i を含むマテリアルストリーム 3 中のコンポーネントの特徴付けを行うように構成された機械学習モデルをトレーニングするためのシステム 1 の実施形態の概略図である。システム 1 は、プロセッサ、コンピュータ可読記憶媒体、センサシステム 5 および分離装置ユニット 1 0 0 を含む。コンピュータ可読記憶媒体には、プロセッサによって実行されるとプロセッサに以下に記す工程を行わせる命令が格納されている。複数の未知のコンポーネント 3 i を有するマテリアルストリーム 3 の撮像を行うべく、マテリアルストリーム 3 を走査するようにセンサシステム 5 を動作させる工程と；マテリアルストリーム 3 の画像および / またはマテリアルストリーム 3 の画像から抽出さ

10

20

30

40

50

れた未知のコンポーネントの一つ以上の特徴を入力として受け取るように構成された機械学習モデルによって、マテリアルストリーム 3 中の未知のコンポーネント 3 i それぞれについて、一つ以上のラベルおよび関連するラベル確率を予測する工程と；マテリアルストリーム 3 中の複数の未知のコンポーネント 3 i のコンポーネント 3 i 各々に関連するトレーニング報酬を決定する工程と；未知のコンポーネント 3 i に関連付けられたトレーニング報酬に少なくとも部分的に基づいて、マテリアルストリーム 3 中の複数の未知のコンポーネント 3 i から少なくとも一つの未知のコンポーネントを選択する工程と；選択された少なくとも一つの未知のコンポーネントをマテリアルストリーム 3 から物理的に分離するために分離装置ユニット 100 を動作させる工程であって、分離装置ユニット 100 は、選択された未知のコンポーネントをアクセス可能な別の区画 101 に移動させる工程と；分析を行うことにより、分離された少なくとも一つの未知のコンポーネントについて決定されたグラントゥールスラベルを受け取る工程であって、分離された少なくとも一つの未知のコンポーネントについて決定されたグラントゥールスラベルをトレーニングデータベースに追加する、工程と；物理的に分離された少なくとも一つの未知のコンポーネントについて決定されたグラントゥールスラベルを使用して、機械学習モデルのインクリメントバージョンをトレーニングする工程と、を実行させる命令が含まれる。

10

【0094】

この実施形態の一例では、分離装置ユニットは、選択されたコンポーネントを自動的に区画 101 内へと分離するロボットアームを備える。グラントゥールス決定に関してさらに分析するために、選択されたコンポーネントをマテリアルストリーム 3 から区画 101 へと選択的に移動させるには、他の手段も採用され得ることが理解されよう。これは、例えば、物理的に分離を行うロボット手段を含む、様々な方法で実行することができる。様々な他の技術も採用することができる。例えば、マテリアルストリームから選択されたコンポーネントの放出は、エアジェット（例えば、エアノズルを使用）により達成してもよい。また、（例えば、マテリアルストリームから分離／隔離する部品の大きさに応じて）複数の技術を組み合わせて使用してもよい。例えば、大きな部品はロボットアームを用いて物理的に分離することができ、小さな部品は流体ノズルを用いた流体ジェットによって分離することができる。

20

【0095】

マテリアルストリーム 3 には大量のコンポーネントが含まれるため、人間がコンポーネントそれぞれを手作業でラベル付けする（大規模データセット）ことは現実的でないと思われる。データ分類器のトレーニングに伴うラベル付作業を最適化するために、手動でラベル付けするために有望かつ模範的なコンポーネントのみを選択する能動学習法が採用される。マテリアルストリーム中の選択されたコンポーネントは、分離装置ユニット 100 によって自動的に物理的に分離される。本実施例では、ロボットアームが配置される。しかし、上述したように、一つ以上のその他の手段も採用することができる。

30

【0096】

機械学習モデルは、ラベル付けのために、コンポーネントを物理的に分離するべく選択関数を適用する能動学習器であってもよい。選択に基づいて、グラントゥールスを決定するための手動および／または実験的ラベル付けのために、マテリアルストリーム 3 から構成要素をアクセス可能な別の区画 101 内へと分離してもよい。新たにラベル付けされたデータを用いて機械学習モデル（分類器を参照）を再トレーニングすることができ、このプロセスは、例えば、予め定義された停止基準が満たされるまで継続してもよい。機械学習モデルをトレーニングするためにラベル付けされるコンポーネントは、トレーニング報酬に基づいて選択され分離されるので、新しいデータポイントに基づいて分類器を再トレーニングする時間のかかるプロセスを回避することができる。そのため、機械学習モデルをより効率的に学習することができる。

40

【0097】

図 2 は、図 1 に示す実施形態と同様なシステム 1 の実施形態を示す模式図である。本実施形態において、分離装置ユニット 100 は、マテリアルストリーム 3 の経路に配置され

50

た操作可能な蓋 103 を備える。例えば、マテリアルストリームは、操作可能な蓋 103 が配置されたコンベアベルトによって搬送されてもよい。システムは、マテリアルストリーム 3 中の一つ以上のコンポーネントを分離するために、蓋 103 を選択的に開くように構成されてもよい。マテリアルストリーム 3 から一つ以上のコンポーネントを分離するべく、蓋 103 がいつ開かれるかを検出するために、必要に応じて光学ユニット 105（例えばカメラ）を使用してもよい。例えば、光学ユニット 105 を使用しない等、その他の変形例も可能であることが理解されよう。光学ユニット 105 は必要に応じて設けられてもよく、例えば、いくつかの例示的な実施形態では、センサシステム 5 からのデータは、マテリアルストリーム 3 から一つ以上のコンポーネントを分離するべく蓋 103 がいつ開かれるかを検出するために使用されてもよい。いくつかの例では、必要に応じて設けられる光学ユニット 105 はまた、蓋 103 が開かれる時の反応時間長くするために上流に配置されてもよい。

10

【0098】

マテリアルストリーム中の複数の特定されたコンポーネントと結び付けられた最も適切なデータポイントを選択して、当該コンポーネントの分離、手動および/または実験的なラベル付けを行い、グラントゥールスを決定することができる。得られたグラントゥールスは、その後、機械学習モデルのさらなる訓練に使用することができる。選択はトレーニング報酬に基づいて行われるため、最小限の人間によるラベル付け作業で、機械学習モデルの最大限の汎化能力を獲得することができる。

【0099】

20

図 3 は、方法 20 の一実施形態の模式図である。本方法は、外部エンティティによって選択されたサンプルを受動的に受け取るのではなく、トレーニングデータを受け取ることが望まれるマテリアルストリーム 3 のサンプルのセットが選択されるような能動的な機械学習を採用する。例えば、機械学習モデルが学習する際に、外部の人間の専門家または外部システムのみに依存してサンプルを特定および提供するのではなく、機械学習モデルが学習に最も役立つ（トレーニングに関連している）と判定したサンプルを選択することを可能にしている。

【0100】

プールベースの能動学習サイクルが図 3 に示されている。ラベル付けされたトレーニングセット 21 を使用して、機械学習モデル 23 をトレーニングしてもよい。機械学習モデルには、ラベルの付いていないプール 25 が提示されてもよい。機械学習モデルは、マテリアルストリーム中のコンポーネントに関連付けられるラベルおよびトレーニング報酬を予測してもよい。次に、分析 27（人間によるアノテーションおよび/または実験）のために、クエリを選択してもよい。選択は、トレーニング報酬に基づいてもよい。選択されたコンポーネントを、ラベル付けのために物理的に分離してもよい。分析/ラベル付けの結果は、機械学習モデルのさらなるトレーニングセット（ラベル付きトレーニングセット 21 を参照）として使用することができる。

30

【0101】

能動学習やクエリ学習は、人間の注釈者および/または自動分析器等のオラクルによってラベル付けされる予定の未ラベル事例をクエリとすることで、トレーニング過程におけるラベル付けのボトルネックを克服することが可能である。このように、能動学習器はできるだけ少ないラベル付きインスタンスを使って高い精度を達成することを目指し、それによってラベル付きデータを得るためのコストを最小化する。多くのクエリ戦略が存在する。例えば、いわゆるプールベースの能動学習が採用され、トレーニングデータは、一方では（小さな）ラベル付きデータセットに、他方ではラベル無しインスタンスの大きなプールに分けられる。能動学習器は貪欲に動作してもよい。ラベルなしプールにおける全てのインスタンスを同時に評価することによって、アノデータに問い合わせるサンプルを選択してもよい。ある基準を最大化するコンポーネント（サンプル）は、アノテーションのためにオラクルに送られ、ラベル付きトレーニングセットに追加される。モデルから送られる更新された結果により、人間のアノデータのために、能動学習器がクエリの新しい選

40

50

択をすることが可能になる。

【 0 1 0 2 】

能動学習器は、アノテーションのために分離および分析する新たなコンポーネントを選択するために、一つまたは複数の基準を採用することができる。これには、様々なアプローチが存在する。いくつかの有利な実施形態において採用されるクエリ戦略は、不確実性サンプリングに基づくものである。能動学習器は、ラベル付けする方法が最も不確実であって未だラベル付けされていないプールのインスタンスを問い合わせる。以下の式において、 \mathbf{x} を、マテリアルストリームのコンポーネントのラベル付けされていないプール中に存在する、ある一つのコンポーネントを記述する特徴ベクトルとする。モデルの下では、そのマテリアル分類、すなわち、粒子のラベルを、全てのクラス y の中で最も事後確率

10

【数 1】

$$\hat{y}|\mathbf{x} = \arg \max_{y|\mathbf{x}} P_{\theta}(y|\mathbf{x}). \quad (1)$$

【 0 1 0 3 】

クエリ戦略の一例としては、ラベル付けされていないプールにおける全てのコンポーネントについて上記の式 (1) を計算し、式 (2) に従ってコンポーネント一つを選択することにより、最も予測信頼性の低いコンポーネントを選択することである。

20

【数 2】

$$\mathbf{x}^* = \arg \max_{\mathbf{x}} (1 - P_{\theta}(\hat{y}|\mathbf{x})). \quad (2)$$

【 0 1 0 4 】

この基準は、機械学習モデルが \mathbf{x} に誤ったラベルを付ける可能性が最も高いサンプルを選択することと等価である。すなわち、最もラベル付される可能性の高いサンプルが、クエリに利用できるラベル付けされていない複数のコンポーネントの中で最も可能性の低いサンプルであるということである。欠点は、機械学習モデルが最も可能性の高いラベルに関する情報のみを考慮するため、事後分布の残りの部分に関する情報を捨ててしまうことである。

30

【 0 1 0 5 】

上記の欠点を解決する代替的なサンプリング戦略として、シャノンエントロピーを不確実性指標とするものがある。

【数 3】

$$\mathbf{x}^* = \arg \max_{\mathbf{x}} \mathcal{H}(\mathbf{y}|\mathbf{x}) = \arg \max_{\mathbf{x}} - \sum_i P_{\theta}(y_i|\mathbf{x}) \log P_{\theta}(y_i|\mathbf{x}). \quad (3)$$

40

ここで、 $\mathbf{y} = (y_1, \dots, y_6)^T$ は、図 1 の例に示すように、6 つのクラス全てのラベルを含むベクトルである。当然、その他のクラス (種類、種別) も使用することができる。エントロピーは変数の平均的な情報量の尺度であるため、機械学習における不確実性または不純物の尺度として一般的に使用される。

【 0 1 0 6 】

図 4 は、特徴空間の一例を示している。本発明は、マテリアルストリーム中のコンポーネント / 対象物に結び付けられたラベルを予測するのに使用された機械学習モデルの更なるトレーニングのために、次いで用いられるグラントゥールスラベルを決定するために、マテリアルストリーム中のこれらのコンポーネント / 対象物のみが選択され分離される物理的能動学習を採用する。このようにして、特定の選択され分離されたコンポーネント

50

／対象物の自動分析を用いて、より効果的に機械学習モデル（例えば、分類モデル）をトレーニングすることができる。

【0107】

その後、コンポーネント一つ一つのランドトゥールスラベルを個別に決定するのは、あまりにも労力がかかりすぎる場合がある。有利なことに、現在では、より少ないデータでモデルを非常によくトレーニングすることができる。システムは、ランドトゥールスラベルを決定するために、マテリアルストリーム中のコンポーネントを自動的に選択して分離し、さらに分析を行うことができる。これは、例えば、一つ以上の廃棄物のストリームを伴う廃棄物処理に非常に有効である。例えば、システムは、廃棄物の特徴付けを行うように構成されてもよく、システムは、採用された機械学習モデルの効率的な更なるト

10

【0108】

ランドトゥールスの決定は、例えば、部分的に手動ラベル付け（例えば、少なくとも部分的に人間によって分析が行われる）を含む、様々な方法で確立されてもよい。また、例えば、化学実験を行う等、自動的に決定してもよい。複数の手法の組み合わせを採用することも可能であり、例えば、ランドトゥールスラベルを導出するために異なる複数の特性を決定する場合等の異なる複数の技術を必要とする場合に、手法を組み合わせてもよい。ランドトゥールスを決定するために、異なる複数の特性パラメータを決定してもよい（例えば、質量、化学反応、重量、幾何学的特性、等）。

20

【0109】

マテリアルストリームは、異種の物質またはコンポーネントの流れであってもよい。どの粒子が機械学習モデルのトレーニングに最も貢献するかを決定するために、様々なアルゴリズムと技術を使用することができる。この目的のために、様々な能動学習法を適用することができる。

【0110】

ランドトゥールスラベル付けを行うために、次の点を選択するべく、（例えば、分析によって）異なる戦略を採用することができる。図4に示す例では、システムは、分離装置ユニット100をクラスタまでの距離に基づいて動作させて、ランドトゥールスラベル付けのために選択される一つ以上のコンポーネントを分離する（サンプリングを参照）ように構成されている。例えば、以下のような様々な手法を用いることができる。

30

- 全てのクラスタ中心から最も遠くに位置するサンプルを選択する。これにより、潜在的な新しい（サブ）クラスを検出することができる。

- クラスタ間に位置するサンプルを選択する（例えば、2つのクラスタに対して等距離に位置するサンプル）。これにより、クラス間の決定を洗練させることができる。

大多数のサンプル／クラスタから最も離れた場所に位置するサンプルを選択する（すなわち、孤立したサンプル）。これにより、新しい（サブ）クラスを表す可能性のある外れ値／異常値を識別することができます。

【0111】

40

上記の技術の組み合わせも使用可能である。その他の選択戦略を採用してもよいことが理解されよう。

【0112】

図5は、異なる複数のコンポーネントクラスに対する特徴の分布を示す。マテリアルストリーム3中のコンポーネント3*i*は、例えば、紙、木材、ガラス、石、鉄金属（*f e r r o*）、非鉄金属（*n o n - f e r r o*）といった、異なる複数のクラス（種類）に分類することができる。例示的なクラスが図5に示されている。機械学習モデルは、これらの異なる複数のクラスを区別するように学習した分類モデルであってもよい。図5に示すグラフは、4つの特徴量の一変量分布と二変量分布を表している。期待されるように、ある特徴量はその他の特徴量よりも、ある材料を識別するのに適している。例えば、原子番号

50

は紙と非鉄金属を良好に識別できるが、石とガラスを区別することはできない。しかしながら、平均密度の場合は逆である。本例では、全部で31個の特徴量を組み合わせて分類器をトレーニングすることで、学習能力を最大化している。

【0113】

斜めに配置されているグラフは、データセットから得られた4つの特徴量の分布についてのカーネル密度推定値を表している。それ以外のグラフは、特徴量それぞれの散布図であり、平均原子番号 $\langle Z \rangle$ 、平均密度 $\langle \rho \rangle$ の対数、高さの標準偏差 $\langle \sigma \rangle$ の対数、および、コンポーネントの周囲長の対数である。

【0114】

グラントゥールスラベル分析のための、マテリアルストリーム中のコンポーネントの選択および分離は、マテリアルストリーム中のラベルの付いていない特定されたコンポーネントについての現在の機械学習モデル（分類器参照）の信頼度に基づいて行うことができる。

10

【0115】

図6は、学習プロセスの指標を様々なグラフで示したものである。図6の上段には、アノテーションのために分離および分析されるべき新たなコンポーネントを選択するために、異なる基準を採用する複数の例示的モデル（ランダムサンプリング、ならびに、最小信頼度およびエントロピー基準に基づく不確実性サンプリング、それぞれ式（2）および（3）を参照）の学習曲線が示されている。これらの曲線は、各モデルのテスト性能が、マテリアルストリーム3中のクエリされた（分類のために選択、分離、分析された）コンポーネントの数の関数として、または、ラベル付きトレーニングセットのサイズの関数として、どのように変化するかを示している。本例では、サンプルサイズを1ずつインクリメントさせ、分類器として、径方向ベースのカーネルを持つサポートベクターマシン（SVM）を使用した。

20

【0116】

式（2）の信頼性基準に基づく不確実性サンプリングと、式（3）のエントロピー基準に基づくランダムサンプリングとを比較する。後者の場合、マテリアルストリームのコンポーネントは、ある不確かさ基準に基づいて問い合わせを行うのではなく、完全にランダムに問い合わせる。

【0117】

30

一般に、ラベル付けされたデータが多いほど情報量が多いので、サンプルサイズが大きくなるほどモデルの性能は上がると考えられている。しかし、これは全てのモデルで同じスピードで起こるわけではない。グラフは、エントロピーおよび信頼性に基づくサンプリング技術の結果は同等であるが、ランダムサンプリングはマテリアルストリームのコンポーネントの分類において明らかに劣ることを示している。サンプルサイズの大きい極限では、全てのモデル性能は、学習データセット全体を使用するモデルの「最適」な値に収束する。この性能は、能動学習モデルが競うべきものであり、図6ではベースライン精度0.988として示されている。図6から明らかなように、能動学習器の性能がベースライン精度に近づくにつれ、小さな絶対的ゲインを得るにはより多くのサンプルが必要となる。例えば、不確実性サンプリングでは、ベースライン精度の99%に達するには77個のラベル付きインスタンスが必要であり、ベースライン精度の99.9%に達するには195個のラベル付きサンプルが必要である。つまり、わずか0.9%ポイントの精度向上のために、ラベル付きサンプルの数を2倍以上にする必要がある。このことは、ラベル付けにかかるコストと分類誤りを犯す可能性のあるコストとのトレードオフという、実用上の重要な問題を提起している。

40

【0118】

グラフの曲線は、250種類のランダムな初期条件の平均結果を示し、斜線領域の境界は10%および90%の分位点で定義されている。さらに、学習プロセスの3つの異なる段階における平均原子番号 $\langle Z \rangle$ および密度 $\langle \rho \rangle$ によって画定される特徴空間の断面図が示されている。1列目は、その時点までどのサンプルが照会されたかを示す。2列目と3列目は

50

、この特徴空間の２次元断面における最小信頼度とエントロピー測定 of の振る舞いを示している。残りのラベル付けされていないサンプルも示されており、最も不確実性の高いものが十字で示されている。これは、（例えば、人間のアノテータや実験によって）分離して分析されるべき次のコンポーネントである。

【 0 1 1 9 】

図 6 のその他のグラフでは、更に検討された学習曲線上に 3 つの位置が示されている。これらは、平均原子番号 Z と密度 ρ の対数とで表される 3 次元の特徴空間の断面を示している。1 列目は、学習プロセスの 3 つの段階ごとに、どのサンプルがその時点までにラベル付けされたかが示されている。能動学習の開始時点では、空間内でサンプルがほぼ均等に描かれており、この時点でランダムモデルとの差がそれほど大きくならない理由である。

10

【 0 1 2 0 】

しかし、より多くのデータが利用可能になると、能動学習器は異なるマテリアルクラス間の境界領域を認識し始め、これらのクラス境界のすぐ近くにあるサンプルを主に問い合わせる。この現象は図 6 の 2 列目と 3 列目から観察することができ、それぞれ 2 次元モデルに対する信頼度およびエントロピーサンプリング基準の挙動を表している。ラベル付けされたトレーニングデータが増えるにつれて、異なるマテリアルクラスを分ける境界は、より高い不確実性を持つ領域として顕著になる。クラス境界は信頼度基準よりもエントロピー基準の方が滑らかに見えるが、共に形状はほぼ同じである。このことは、なぜ類似のサンプルが照会され、両者の性能がほぼ同じであるかを説明している。

20

【 0 1 2 1 】

一般的に、不確実性指標の最適な選択は、手元のデータセットに強く依存する。しかし、分類が単純に多数決によって実行される場合、信頼基準がおそらく適切であると言える。コンポーネントは、最も高い事後確率を持つクラスに割り当てられる。しかし、より複雑なルールが使用される場合（例えば、不均衡なデータセットの場合）、間違いなくエントロピーがより明白な選択となる。

【 0 1 2 2 】

図 7 は、システム 1 の一実施形態の模式図を示す。この例では、必要に応じて設けられる（カラー）カメラ 7 および必要に応じて設けられる 3 D レーザ三角測量ユニット 9 のうちの少なくとも一つが、区画に分けられたオブジェクトのそれぞれに結びつけられる追加の特性を決定できるように配置されている。したがって、いくつかの例では、材料の種類、質量等に関する特徴 / 特性に次いで、サイズ、形状、色、テクスチャ、視覚的洞察等の少なくとも一つに基づいて特定されたおよび / またはセグメント化されたオブジェクトを区別することも可能である。また、このような情報により、仮想的な実験が可能となり得る。本例では、センサユニット 5 は、デュアルエネルギー X 線撮像を行うための 2 つの X 線サブユニット 1 1 a、1 1 b を有する X 線センサ 1 1 を含む。さらに、カメラ 7 および 3 D レーザ三角測量ユニット 9 は、センサユニット 5 と一体化されている。このようにして、センサユニット 5 は、例えば、コンピュータユニット 1 3 によって位置合わせおよび / または合成可能な複数の画像を提供する。異なる複数のカメラ / 検出器から得られた画像データの位置合わせおよび / または合成により、セグメント化されたオブジェクトの特徴 / 特性をより良く決定することを可能にする。1 つ以上のマテリアルがセグメント化されており、その関連する特徴 / 特性を決定するために、セグメント化された個々のオブジェクト 3 i の分析が行われる。本例では、密度、材料、形状、サイズおよび質量といった特徴量 1 5 が、セグメント化されたオブジェクトそれぞれについて決定される。その他の特徴量セットも可能であることは理解されよう。データから、セグメント化されたオブジェクトそれぞれの相対的な重量パーセントを導出することも可能である。

30

40

【 0 1 2 3 】

本発明によるシステムは、人間による（労働集約的な）入力 that 少なくて済む一方で、一つまたは複数のマテリアルの特徴づけをより速く、より自律的に行うことができる。このシステムは、廃棄物の特徴づけの用途において重要な利点を提供できる。

50

【 0 1 2 4 】

廃棄物の画像を認識し、異なる複数のカテゴリへと分類するモデルを開発するためには、機械学習モデルに多数の画像を見せ、それぞれの画像に何が写っているかを示すラベルを付けることでトレーニングさせることができる。このように、全てのデータに予めラベルを付けておく従来の手法を教師あり学習という。ラベル付けされたデータは、機械学習アルゴリズムの燃料となる。廃棄物の特徴付け技術では、ラベル付けされたデータは典型的には、物理的な「純粋な」単一物質ストリームを走査することで生成できる。このストリームは、同質でない廃棄物ストリームから何千もの個々の粒子を丹念に選択して手動で準備されることが多い。

【 0 1 2 5 】

リサイクル産業において、廃棄物の特徴付けにはいくつかの重要な用途があります。それは、価値評価に利用可能である。完全なマテリアルストリームを迅速かつ確実に特徴付けすることにより、商品株式市場の変動にさらされるリスクを減らすことができる。さらに、品質管理にも利用できる。循環型経済では、リサイクル製品の品質が保証されることが望ましい。特徴付け技術は、市場の信頼を確立するのに役立つ。さらに、プロセスエンジニアリングにも利用できる。廃棄物リサイクルプロセスの技術的、経済的実現可能性や、仮想実験による新しいプロセス設計を評価することができる。さらに、オンラインによるプロセスの最適化にも利用できる。選別工程をその場で測定し、制御し、最適化することができる。

【 0 1 2 6 】

いくつかの例では、マテリアルを定性的に（マテリアルの種類、化学、純度等）、定量的に（質量バランス、物理特性等）を評価する、直接的なインライン特徴付け技術を提供することができる。このようなインライン特徴付けシステムは、不均質で複雑なマテリアルストリームを完全に評価するように構成することができ、サブサンプリングの必要性を排除することができる。さらに、物質収支はオンザフライで作成することができる。実際、マテリアルオブジェクトのそれぞれについて、デジタルツインを作成し、仮想的に評価することができる。

【 0 1 2 7 】

本発明は、トレーニングデータを収集する際のラベル付けの労力を大幅に低減することができる物理的能動学習を用いたデータ駆動型のマテリアル特徴付けを提供することができる。従来の機械学習アルゴリズムは、トレーニングのために、完全にラベル付けされた大量のデータセットを必要とするが、良好な予測を行うためには、このデータのほんの一部しか必要ないことが分かっている。能動学習では、アルゴリズムが選択した小さな部分集合でモデルをトレーニングすることができ、完全なデータセットでモデルをトレーニングさせた場合と同等の精度を得ることが可能である。いくつかの例では、能動学習により、全てラベル付けされたデータセットで学習した場合の 99% の精度を維持しながら、ラベル付けコストを 70% 削減することができる。

【 0 1 2 8 】

本発明によるシステムおよび方法は、様々なマテリアルストリームに使用できることは理解されよう。いくつかの例では、マテリアルストリームは、建築廃棄物および解体廃棄物を含む。しかしながら、その他の廃棄物の流れにも使用することができる。

【 0 1 2 9 】

図 8 は、方法 30 の模式図である。第 1 のステップ 31 において、一つまたは複数のマテリアルの対象物またはコンポーネントが特定され、セグメント化される。これは、オブジェクト検出アルゴリズムおよび / またはセグメント化アルゴリズムによって実行することができる。画像は、センサユニット 5 を使用して取得される。また、セグメント化される画像は、例えば、センサユニット 5 の異なるセンサまたはサブユニットから提供される異なる複数の画像のアライメントおよび / または合成を実行した後に、取得されてもよい。本例では、セグメント化された対象物 3 i の周囲にボックス 20 が配置される。第 2 のステップ 33 では、セグメント化された対象物 3 i のそれぞれについて、特性 / 特徴量 1

10

20

30

40

50

5 が決定される。この例では、質量、体積、原子番号が決定される。第 3 のステップでは、機械学習モデルによってラベルを予測してもよい。ステップ 37 に示すように、これは、出力として（予測された）ラベル 17 を得るようにトレーニングされたニューラルネットワーク 25 に、入力としてデータを提供することによって行うことができる。この例では、トレーニングされたニューラルネットワークは、深層学習モデルである。しかしながら、例えば、サポートベクターマシン（SVM）、決定木ベースの学習システム、ランダムフォレスト、回帰モデル、オートエンコーダクラスタリング、最近傍（例えば kNN）機械学習アルゴリズム等の、その他の機械学習モデルも使用することが可能である。いくつかの例では、人工ニューラルネットワークの代わりに、代替の回帰モデルが使用される。

【0130】

本発明は、使用される機械学習モデル（例えば、ディープニューラルネットワーク）のより効率的なトレーニングを提供する。能動学習によって、（マテリアルストリーム中の）ラベル付けされていないデータのサブセットを選択的にサンプリングすることにより、（手動で）ラベル付けされるべきトレーニングサンプルの数を減らすことが可能である。これは、人間によるラベル付けおよび/または実験的ラベル付けのために、ラベル付けされていないサンプルを検査し、所定のコスト関数に関して最も情報量の多いものを選択することによって減らすことができる。能動学習機械学習モデルは、性能の最も大きな改善をもたらすことができるサンプルを選択することができ、それによって、人間によるラベル付けおよび/または実験的ラベル付けの労力を低減できる。マテリアルストリーム中の複数のコンポーネントから選択的にコンポーネントをサンプリングすることは、複数のコンポーネントのうちラベル付けするコンポーネント候補のプールが存在することを想定している。マテリアルストリーム中には、新しい比較的ユニークなコンポーネントの定常的なストリームが存在し得ることから、ストリームは機械学習モデルの性能を継続的かつ効果的に向上させるためのソースを提供している。有利なことに、選択されたコンポーネントは、分離ユニットを使用したシステムによって自動的に分離されてもよい。能動学習機械モデルは、人間によるラベル付けおよび/または実験的ラベル付けのために、マテリアルストリームから収集された全コンポーネントのうちの小さなサブセットを導出してよい。

【0131】

例えば、人間のアノテーションによって得られた分類済データセットを用いて、最初の深層学習ニューラルネットワークをトレーニングすることができる。このデータセットによって、ニューラルネットワークの最初のパラメータが構築され、これが教師あり学習の段階となる。教師あり学習の段階で、所望の動作が達成されたかどうかのニューラルネットワークのテストを行ってもよい。所望のニューラルネットワークの動作が達成されると（例えば、機械学習モデルが指定された閾値に従って動作するようにトレーニングされると）、実際の使用のために機械学習モデルを配備することができる（例えば、「実際の」データでマシンをテストする）。ニューラルネットワークの動作を改善し続けるために、動作中、（例えば、専門家ユーザ、専門家システム、参照データベース等により）ニューラルネットワークが行う分類を確認または否定することができる。その後、ニューラルネットワークの振る舞いを決定する分類についてのパラメータが、進行中のインタラクションに基づいて更新されるため、本例のニューラルネットワークは転送学習の状態にあると言える。いくつかの例では、機械学習モデルのニューラルネットワークは、別のプロセスに直接フィードバックを提供することができ、例えば、廃棄物リサイクルプロセスの制御パラメータを変更することができる。いくつかの例では、ニューラルネットワークは、（例えば、クラウドなどを介して）バッファリングされ検証されたデータを別のプロセスに提供する前に出力する。

【0132】

データ取得は、様々な方法で行うことができる。センサシステムは、様々なセンサを含んでもよい。一例として、マテリアルストリーム（例えば、廃棄物ストリーム）中の粒子のマテリアル特性に関するデータは、マルチセンサ特徴づけ装置により収集される。まず

10

20

30

40

50

、デュアルエネルギー X 線透過法 (DE - XRT) により、マテリアルを「透視」し、平均原子番号や密度等の、特定のマテリアル特性を判定してもよい。これの利点は、部品の表面だけでなく、体積全体を検査できることである (例えば、廃棄物は汚れていることが多く、表面特性は必ずしもマテリアルの大部分を代表するものではない)。次に、これに加えてまたは代えて、3D レーザ三角測量ユニットを利用して、対象物の形状を高解像度 (例えば、サブミリ精度) で測定してもよい。これにより、DE - XRT から収集した情報を補完する、3次元形状および体積等の追加情報を得ることができる。次いで、これに加えてまたは代えて、RGB 検出器を使用してもよく、これによりマテリアルストリームに含まれるコンポーネントを色および形に関して区別することができる。いくつかの例では、上述した複数のセンサを一緒に使用する。必要に応じて、画像を個々の成分にセグメント化するために画像処理を使用してもよい。これらのセグメント化された画像から、対象物の形状を記述する様々な特徴を計算してもよい。例えば、コンポーネントの面積、偏心、周囲長等である。いくつかの例では、これは、全てのセンサから得られた全ての画像に対して行ってもよい。

10

【0133】

様々なニューラルネットワークモデルおよび/またはニューラルネットワークアーキテクチャを使用することができる。ニューラルネットワークは、センサデータおよび/または前処理されたデータ (例えば、セグメント化された対象物について決定された特徴の特性) を処理する、例えば、分類する能力を有する。ニューラルネットワークは、コンピュータ化されたシステムに実装されてもよい。ニューラルネットワークは、複雑なデータ入力を処理するための様々な機械学習アルゴリズムのフレームワークとして機能してもよい。このようなニューラルネットワークシステムは、一般に、タスクに特化した規則でプログラムされるのではなく、複数の例を考慮することによってタスクを実行するように「学習」することができる。ニューラルネットワークは、ニューロンと呼ばれる接続されたユニットまたはノードの集合体をベースとしている。各接続は、ニューラルネットワーク内の一つのニューロンから別のニューロンへと信号を伝達可能である。信号を受け取ったニューロンはその信号を処理し、そのニューロンに接続された更なるニューロンへと信号を送ることができる (アクティベーションを参照)。ニューロンそれぞれの出力は通常、入力の和の非線形関数によって計算される。複数の接続はそれぞれの重みを持つことができ、重みは学習が進むにつれて調整される。また、バイアスのようなその他のパラメータが存在する場合もある。一般に、ニューロンはレイヤへと集約される。異なるレイヤはその入力に対して異なる種類の変換を行い、ディープニューラルネットワークを形成することができる。

20

30

【0134】

深層学習ニューラルネットワークは、複数のレベルの表現を持つ表現学習法とみなすことができる。この表現学習法は、生の入力から始まるあるレベルの表現を、より高い、やや抽象的なレベルの表現にそれぞれ変換する、単純だが非線形なモジュールを構成することによって得られる。ニューラルネットワークは、従来の手法では困難なパターンを識別することができる。したがって、ある印刷条件で構造体を印刷するという問題に特化したカスタムコードを書く代わりに、例えば、分類アルゴリズムを使用して、異なるおよび/または変化する構造体の印刷条件を処理できるようにネットワークをトレーニングすることができる。印刷プロセスを効率的に制御するための分類ロジックを決定できるように、ニューラルネットワークにトレーニングデータを供給してもよい。

40

【0135】

方法の一の工程が別の工程の後に実施されると説明された場合、そうでないと明記されない限り、それは別の工程の後に一の工程が直接続くまたは一の工程を実行する前に一つまたは複数の中間工程を実行してもよいことを意味する。同様に、ニューラルネットワークのニューロン等のコンポーネント間の接続が記述されている場合、この接続は、特に指定されない限りまたは文脈によって除外されない限り、直接的にまたはその他のニューロンや論理演算等の中間コンポーネントを介して確立されてもよいことが理解されるである

50

う。

【0136】

「ラベル」という用語は、カテゴリ変数（例えば、ニューラルネットワークを使用）および連続変数（例えば、回帰モデルを使用）の両方として理解され得ることは明らかである。例えば、連続変数は、不確実性を有する場合がある（例えば、化学分析変数）。

【0137】

本方法は、コンピュータで実装される工程を含み得ることが理解されるであろう。上述したすべての工程は、コンピュータで実装される工程とすることができる。実施形態は、プロセスを実行するコンピュータ装置を含んでもよい。また、本発明は、本発明を実施するために適合化されたコンピュータプログラム、特にキャリア上またはキャリア内のコンピュータプログラムにも及ぶ。プログラムは、ソースコードまたはオブジェクトコードの形態であっても、あるいは本発明によるプロセスの実装に使用するのに適した他の形態であってもよい。キャリアは、プログラムを運搬することができる任意の実体または装置であってもよい。例えば、キャリアは例えば半導体ROMなどのROMまたはハードディスクのような記憶媒体であってもよい。さらに、キャリアは電気信号や光信号などの伝送可能なキャリアであってもよく、電気ケーブルや光ケーブル、無線などの手段、例えばインターネットやクラウドを介して伝達されてもよい。

【0138】

いくつかの実施形態は、例えば、命令または命令セットを格納することができる機械または有形コンピュータ可読媒体または要素を用いて実装されてもよく、これらの命令または命令セットは、機械によって実行された場合に、当該機械に実施形態に従った方法および/または動作を実行させることができる。

様々な実施形態は、ハードウェア要素、ソフトウェア要素、またはその両方の組み合わせを用いて実装することができる。ハードウェア要素の例としては、プロセッサ、マイクロプロセッサ、回路、ASIC (Application Specific Integrated Circuit)、PLD (Programmable Logic Device)、DSP (Digital Signal Processor)、FPGA (Field Programmable Gate Array)、論理ゲート、レジスタ、半導体デバイス、マイクロチップ、チップセットなどがある。ソフトウェアの例としては、ソフトウェアコンポーネント、プログラム、アプリケーション、コンピュータプログラム、アプリケーションプログラム、システムプログラム、マシンプログラム、オペレーティングシステムソフトウェア、モバイルアプリ、ミドルウェア、ファームウェア、ソフトウェアモジュール、ルーチン、サブルーチン、関数、コンピュータで実装された方法、プロシージャ、ソフトウェアインターフェース、アプリケーションプログラムインターフェース (API)、メソッド、命令セット、コンピューティングコード、コンピュータコード等を挙げることができる。

【0139】

本明細書では、本発明の各実施形態の具体的な例を参照して本発明が説明されている。ただし、本発明において、本発明の本質から逸脱することなく、様々な変更、変形、代替、および改変を行うことができることは明白であろう。明確化および簡潔な説明のために、本明細書では同じまたは別々の実施形態の一部として特徴が説明されているが、これらの別々の実施形態において説明される特徴の全てまたは一部の組み合わせを有する代替的な実施形態も、特許請求の範囲によって概説される本発明の枠組み内にあると想定および理解される。従って、各仕様、図面、および例は、制限的な意味ではなく、例示的な意味で解釈されるべきである。本発明は、添付の特許請求の範囲内にある全ての代替、変更、および変形を包含することを意図している。また、説明される要素の多くは、個別の、もしくは分散された構成要素として、または他の構成要素と併用して、任意の好適な組み合わせおよび位置で実装され得る機能的エンティティである。

【0140】

特許請求の範囲において、括弧内のいかなる参照符号も、請求項を限定するものとは解

10

20

30

40

50

釈されないものとする。「含む (comprising)」という単語は、請求項に列挙されたもの以外の特征または工程の存在を排除するものではない。また、単語「a」および「an」は、「一つだけ」に限定されると解釈されるべきではなく、「少なくとも一つ」を意味するために用いられ、複数を排除しない。単に特定の複数の手段が相互に異なる請求項に記載されているという事実は、これらの手段の組み合わせが有利に使用できないことを示すものではない。

【図面】

【図 1】

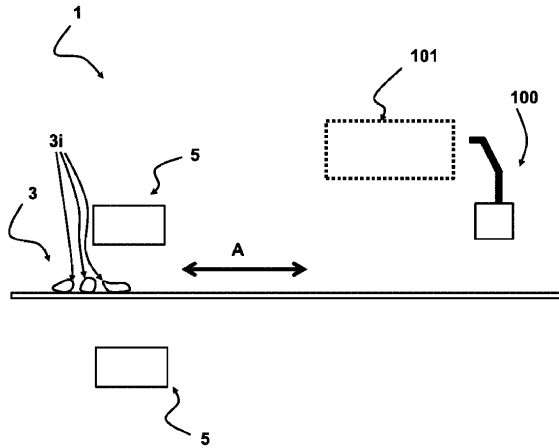


FIG 1

【図 2】

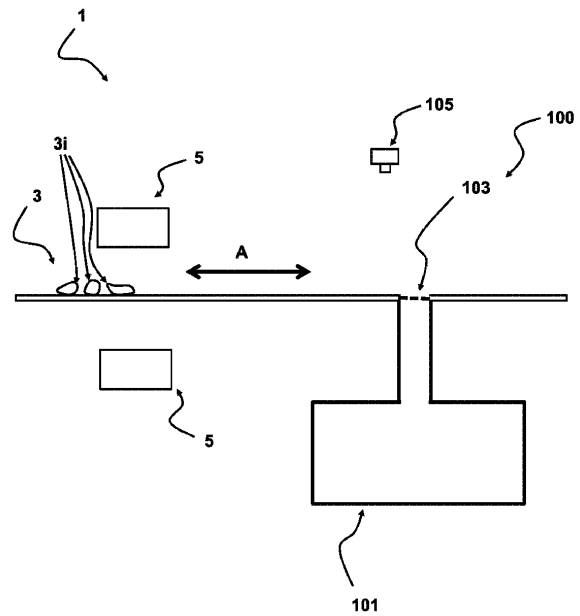


FIG 2

10

20

30

40

50

【図 3】

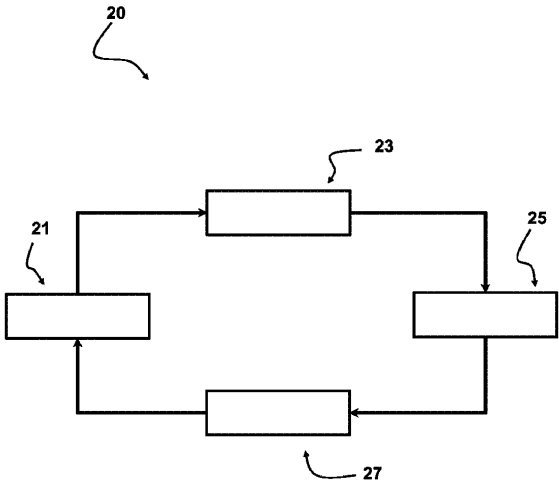


FIG 3

【図 4】

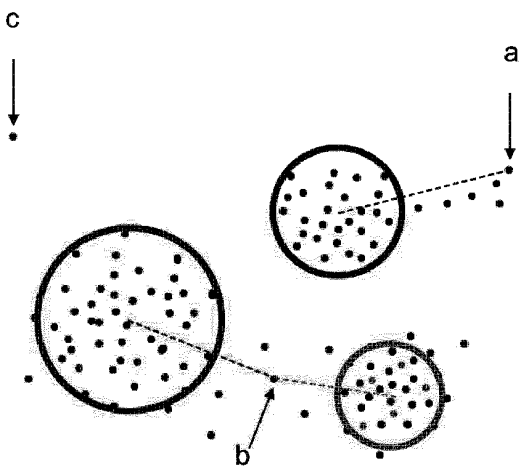


FIG 4

【図 5】

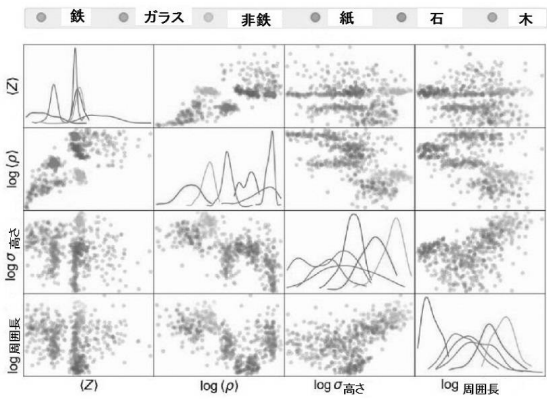


FIG 5

【図 6】

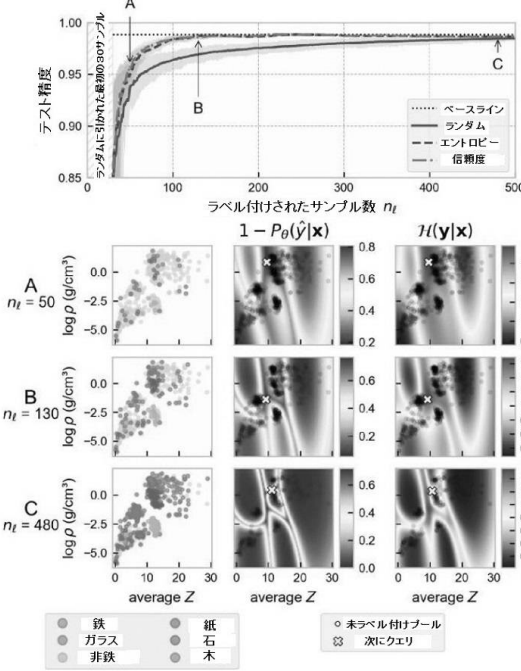
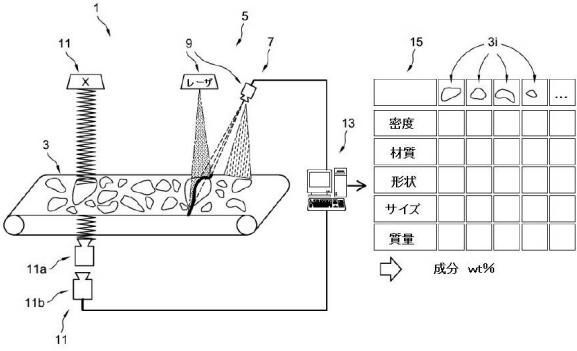
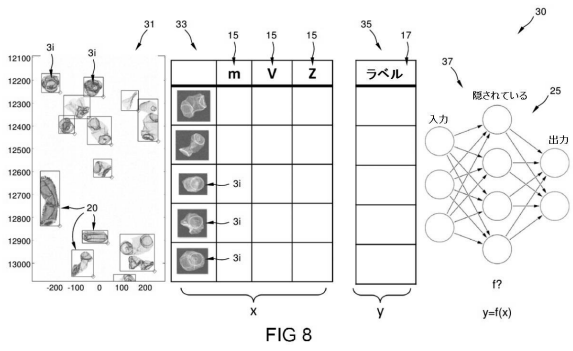


FIG 6

【図 7】



【図 8】



10

20

30

40

50

フロントページの続き

- (56)参考文献 米国特許出願公開第 2 0 1 9 / 0 2 1 7 3 4 2 (U S , A 1)
特許第 6 6 7 9 1 8 8 (J P , B 1)
特開 2 0 1 7 - 1 0 9 1 9 7 (J P , A)
特開 2 0 1 9 - 1 9 5 7 5 1 (J P , A)
中国特許出願公開第 1 1 0 5 1 3 1 6 1 (C N , A)

- (58)調査した分野 (Int.Cl. , D B 名)
G 0 6 N 2 0 / 0 0
G 0 1 N 2 1 / 7 3
G 0 1 N 2 1 / 7 4
G 0 6 T 7 / 0 0