(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7642497号
(P7642497)

(45)発行日　令和7年3月10日(2025.3.10)

(24)登録日　令和7年2月28日(2025.2.28)

(51)国際特許分類
Ｇ０６Ｎ　20/00　(2019.01)

ＦＩ
Ｇ０６Ｎ　20/00

請求項の数　8　（全19頁）

(21)出願番号　特願2021-143871(P2021-143871)
(22)出願日　令和3年9月3日(2021.9.3)
(65)公開番号　特開2023-37242(P2023-37242A)
(43)公開日　令和5年3月15日(2023.3.15)
審査請求日　令和6年2月14日(2024.2.14)

(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　ローレンス アレクサンダー
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　高田 実佳
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　坂田 匡通
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　馬場 恒彦

最終頁に続く

(54)【発明の名称】　データ管理システム及びデータ管理方法

(57)【特許請求の範囲】
【請求項１】
データ管理システムであって、
１以上の記憶装置と、
１以上の演算装置と、を含み、
前記１以上の記憶装置は、入力データ及び第１学習モデルを格納し、
前記入力データは、複数のレコードを含み、
前記複数のレコードの各レコードは、複数の特徴量の値を含み、
前記１以上の演算装置は、
前記複数の特徴量から、指定されている人関連ＩＤと異なる特徴量を選択し、
前記選択された特徴量のデータによって、第１学習モデルの訓練を実行し、
訓練済みの前記第１学習モデルの推定結果に基づき、前記選択した特徴量の少なくとも一部の、前記第１学習モデルの推定に対する影響度を算出し、
前記影響度及び前記少なくとも一部の特徴量の情報を、ユーザに提示する、データ管理システム。
【請求項２】
請求項１に記載のデータ管理システムであって、
前記１以上の演算装置は、
前記選択された特徴量のデータの間の相関を決定し、
他の特徴量との相関が閾値を超える特徴量から、前記影響度の算出から除外する特徴量

を決定する、データ管理システム。
【請求項３】
請求項１に記載のデータ管理システムであって、
前記１以上の演算装置は、
前記選択された特徴量と、指定されているセンシティブ特徴量との間の相関を決定し、
前記センシティブ特徴量との間の相関が閾値を超える特徴量から、前記影響度の算出から除外する特徴量を決定する、データ管理システム。
【請求項４】
請求項１に記載のデータ管理システムであって、
前記１以上の記憶装置は、前記入力データによる訓練の目的とする第２学習モデルを格納し、
前記第１学習モデルの訓練の処理負荷は、前記第２学習モデルの訓練の処理負荷より小さい、データ管理システム。
【請求項５】
請求項４に記載のデータ管理システムであって、
前記演算装置は、
前記影響度及び前記少なくとも一部の特徴量の情報を前記ユーザに提示した後、前記入力データにおいて前記ユーザに選択された特徴量のデータによって、前記第２学習モデルを訓練し、
訓練された前記第２学習モデルの推定結果に基づき、前記第２学習モデルの公平性及び／又は正確性の指標を決定して、前記ユーザに提示する、データ管理システム。
【請求項６】
請求項１に記載のデータ管理システムであって、
前記１以上の演算装置は、前記第１学習モデルの推定結果から得られる混同行列に基づいて、前記影響度を算出する、データ管理システム。
【請求項７】
請求項１に記載のデータ管理システムであって、
前記１以上の記憶装置は、過去の分析処理履歴を格納し、
前記１以上の演算装置は、前記選択された特徴量の情報を前記分析処理履歴において検索し、
前記選択された特徴量と同一特徴量の情報が前記分析処理履歴において見つかると、前記分析処理履歴から取得した影響度を前記ユーザに提示する、データ管理システム。
【請求項８】
システムがデータを管理する、管理方法あって、
前記システムは入力データ及び第１学習モデルを格納し、
前記入力データは、複数のレコードを含み、
前記複数のレコードの各レコードは、複数の特徴量の値を含み、
前記管理方法は、前記システムが、
前記複数の特徴量から、指定されている人関連ＩＤと異なる特徴量を選択し、
前記選択された特徴量のデータによって、第１学習モデルの訓練を実行し、
訓練済みの前記第１学習モデルの推定結果に基づき、前記選択した特徴量の少なくとも一部の、前記第１学習モデルの推定に対する影響度を算出し、
前記影響度及び前記少なくとも一部の特徴量の情報を、ユーザに提示する、ことを含む、管理方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、機械学習モデルの訓練に使用されるデータの管理に関し、特に、機械学習モデルの推定にバイアスを与え得る特徴量の評価に関する。
【背景技術】

【０００２】
本開示の背景技術として、例えば、特許文献１がある。特許文献１は、効率的に特徴量データを準備する方法を開示する。この方法は、予め格納されているモデルと患者の特徴量データとの類似性を決定する。類似性の決定において、予め格納されているモデルのデータベースが類似性を評価するために分析される。類似性は、予め格納されているモデルの特徴量準備が患者の特徴量データと両立するかを示す。
【０００３】
この方法は、利用すべき特徴量準備を示す類似性について、類似していると判定された予め格納されているモデルに基づいて、患者の特徴量データに対する特徴量準備を行う。特徴量準備は、類似の予め格納されているモデルに関連する再使用可能な特徴量を取得する。再使用可能な特徴量は、類似の予め格納されているモデルの予め計算されている特徴量を含む。特徴量準備の結果及び患者の特徴量データを用いて、機械学習モデルを生成する。
【０００４】
機械学習の分野において、公平性が求められている。機械学習モデルが、例えば、特定のグループに対して偏見（バイアス）を伴う推定を行わないように、その機械学習モデルを生成することが求められている。より公平な推定を行えることは、より高精度な推定を行えることにつながる。
【先行技術文献】
【特許文献】
【０００５】
【文献】米国特許出願公開第２０１９／０３０４６０３号
【発明の概要】
【発明が解決しようとする課題】
【０００６】
従来、機械学習モデルの公平性及び／又は正確性（精度）に対する特徴量の評価は、頻繁なトライアンドエラーを引き起こしている。また、データレイクのような大きなデータストアの高次元データに対しては、常に新しいデータが大量に追加される。そのため、新たに追加されるデータにおける特徴量の評価に、多大な時間が必要とされている。
【０００７】
従って、ユーザ又は他のシステムが機械学習モデルの訓練に使用する特徴量を選択する前に、新たなデータにおける特徴量において、機械学習モデルの推定処理にバイアスを与える可能性が高い特徴量を検出することが望まれる。
【課題を解決するための手段】
【０００８】
本発明の一態様は、データ管理システムであって、１以上の記憶装置と、１以上の演算装置と、を含み、前記１以上の記憶装置は、入力データ及び第１学習モデルを格納し、前記入力データは、複数のレコードを含み、前記複数のレコードの各レコードは、複数の特徴量の値を含み、前記１以上の演算装置は、前記複数の特徴量から、指定されている人関連ＩＤと異なる特徴量を選択し、前記選択された特徴量のデータによって、第１学習モデルの訓練を実行し、訓練済みの前記第１学習モデルの推定結果に基づき、前記選択した特徴量の少なくとも一部の、前記第１学習モデルの推定に対する影響度を算出し、前記影響度及び前記少なくとも一部の特徴量の情報を、ユーザに提示する。
【発明の効果】
【０００９】
本発明の一態様によれば、入力データから、機械学習モデルの公平性及び／又は正確性にバイアスを与える可能性が高い特徴量を検出することができる。
【図面の簡単な説明】
【００１０】
【図１】本明細書の一実施形態に係る特徴量推奨システムのハードウェア構成例を示す。

【図２】特徴量推奨システムのソフトウェア構成例を示す。
【図３Ａ】患者個人情報テーブルの構成例を示す。
【図３Ｂ】通報イベントテーブルの構成例を示す。
【図３Ｃ】病院患者情報テーブルの構成例を示す。
【図４】人関連ＩＤ情報テーブルの構成例を示す。
【図５】準識別子情報テーブルの構成例を示す。
【図６】センシティブ特徴量情報テーブルの構成例を示す。
【図７】バイアス特徴量情報テーブルの構成例を示す。
【図８】バイアス特徴量候補情報テーブルの構成例を示す。
【図９】影響度情報テーブルの構成例を示す。
【図１０】特徴量推奨システムが実行する処理の概要を示すフローチャートである。
【図１１】ユーザが分析対象のデータについて情報を入力するための、ＧＵＩ（Ｇｒａｐｈｉｃａｌ　Ｕｓｅｒ　Ｉｎｔｅｒｆａｃｅ）画像の例を示す。
【図１２】図１０のステップＳ１１の詳細のフローチャートを示す。
【図１３】特徴量の影響度の推定処理の例を示すフローチャートである。
【図１４】特徴量の影響度の推定処理の例を示すフローチャートである。
【図１５】フルスケールモデルの訓練結果を示す画像の例を示す。

【発明を実施するための形態】

【００１１】

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形態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。

【００１２】

本明細書の一実施形態に係るシステムは、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以上の演算装置又はプロセッサを含む。

【００１３】

プログラムが演算装置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、機能は演算装置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、演算装置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。

【００１４】

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。本明細書の実施形態は、他のシステム又は処理から独立して又は組み合わせて利用することができ、実施形態の機能は、所望の実装に従って任意の手段により実現できる。

【００１５】

本明細書の一実施形態は、機械学習システムのデータ管理に関する。本明細書の一実施形態に係るシステムは、未分析の新たなデータにおいて、機械学習モデルの訓練に利用され得る特徴量を抽出する。システムは、抽出された特徴量において、機械学習モデルの推定アルゴリズムにバイアスを与える可能性があるバイアス特徴量の候補を抽出する。これ

により、機械学習モデルによる推定アルゴリズムの公平性、性能及び精度等の目的のため機械学習モデルの最適化を促進でできる。

【００１６】

本明細書の一実施形態に係るシステムは、入力データと分析処理履歴とを比較することで、入力データにおける特徴量を効率的に評価することができる。また、システムは、入力データの特徴量の機械学習モデルの推定に対する影響度を算出して、その評価結果をユーザに提示することができる。これにより、訓練データに使用するユーザによる特徴量選択を効果的に支援できる。

【００１７】

機械学習モデルの学習に利用される特徴量の選択は、グラフィカルユーザインタフェース（ＧＵＩ）若しくは他の入力手段を介して人間のユーザによって実行される、又は所望のアルゴリズムに従って本システム又は他のシステムよって実行され得る。

【００１８】

本明細書一実施形態は、機械学習モデルを継続的に改良することに関連し、公平性や精度等の尺度についての機械学習モデルの最適化に関連する。本明細書の一実施形態は、様々な目標尺度に対する特徴量の影響について、エンドユーザに改善された洞察を提供する。これにより、特徴量の選択を改善し、機械学習モデルの改良が促進され得る。本明細書の一実施形態は、例えば、救急医療サービス（ＥＭＳ）決定システム内の機械学習モデルの訓練に適用することができる。

【００１９】

機械学習モデルの入力は、例えば、救急車の出動記録（時刻、チーム番号、出動先場所、患者情報（氏名・年齢・症状1・症状２等）、配送先）、天候（場所、気温、等）、配送先病院情報である。出力は、例えば、救急車の需要（患者発生数）、予測結果（救急システム管理官向け）、救急配送の最適なルート推薦（救急車両向け）、病気診断予測（コールセンタ、病院向け）である。

【００２０】

本明細書の一実施形態は、公平性及び／又は正確性等の機械学習モデルの評価値を改善し、例えば、医療決定システムの日常的なシステム管理を容易にすることができる。医療決定は、例えば、待機している救急車の最適配置や、看護師による患者への様々な治療順序をトリアージ方式で決定する等の例示的なユースケースを含むことができるが、これらに限定されない。

【００２１】

本明細書の一実施形態に係るシステムは、機械学習モデルの訓練フェーズの前に、機械学習モデルの訓練におけるバイアスを低減するために、未分析の入力データ（新データ）を前処理するように構成されている。システムは、記憶装置に格納されている情報を参照して、入力データを分析する。

【００２２】

入力データは、例えば新たにデータベースに追加された患者に関するデータであり、検出前の潜在的なバイアス特徴量候補を含む。分析されるデータは、記憶装置に格納されている、入力データ自体でもよく、入力データの特徴量の情報を示すメタデータであってもよい。

【００２３】

入力データの分析のために参照される情報は、例えば、過去の分析処理履歴を含むことができる。分析処理履歴は、分析されたデータ（それぞれ複数特徴量からなる多数のサンプル）、ユーザに指定されたバイアス特徴量（センシティブ特徴量と呼ぶことがある）や、過去の入力データの特徴量についての分析結果、及びこれらの管理のために参照され得るメタデータ（管理情報）を含むことができる。

【００２４】

分析処理履歴は、過去の分析処理により、バイアス特徴量候補と判定された特徴量及び非バイアス特徴量候補と判定された特徴量を、示してよい。分析処理履歴は、入力データ

における特徴量の、機械学習モデルによる推定に対する影響度の計算結果を含んでよい。

【００２５】

図１は本明細書の一実施形態に係る特徴量推奨システムのハードウェア構成例を示す。図１に示す特徴量推奨システム１００は、データを管理するデータ管理システムであり。特徴量推奨システム１００は、ユーザが機械学習モデルの訓練を行うために、訓練データに含めてよい又は訓練データから除くべき特徴量を決定し、それらをユーザに提示する。具体的には、パーソナルコンピュータの他、タブレット、スマートフォンなどの携帯型情報端末、サーバやクラウド上の仮想情報処理装置などを使用できる。

【００２６】

特徴量推奨システム１００は、ＲＡＭなど揮発性記憶素子で構成されるメモリ（主記憶装置）１０１、ＳＳＤ（Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅ）やハードディスクドライブなど適宜な不揮発性記憶素子で構成される補助記憶装置１０２を含む。特徴量推奨システム１００は、さらに、補助記憶装置１０２に保持されるプログラム１０３をメモリ１０１に読み出すなどして実行し装置事態の統括制御を行うとともに、各種判定、演算及び制御処理を行うＣＰＵなどの演算装置１０４を含む。

【００２７】

さらに、特徴量推奨システム１００は、ユーザからの入力動作を受け付けるキーボードやマウス、タッチパネルなどの入力装置１０５、利用者に対して処理結果を表示するディスプレイなどの出力装置１０６、及び、適宜なネットワークに接続しデータをやり取りするための通信装置１０７を備えている。なお、特徴量推奨システム１００がスタンドアロンマシンとして稼働する場合、通信装置１０７は省略してもよい。

【００２８】

なお、補助記憶装置１０２内には、本実施形態の特徴量推奨システム１００として必要な機能を実装するためのプログラム群１０３、各種処理に必要なデータを格納した情報データベース１５０が格納されている。補助記憶装置１０２は、さらに、特徴量推奨システム１００に登録されている各種電子ファイルの保存領域１７０を含む。ファイル保存領域１７０は、特徴量推奨システム１００に固有のものであってもよいし、また企業や組織内で運用されているＮＡＳ（Ｎｅｔｗｏｒｋ　Ａｔｔａｃｈｅｄ　Ｓｔｏｒａｇｅ）などのネットワークストレージや、ウェブストレージなどに確保されてもよい。

【００２９】

図２は、特徴量推奨システム１００のソフトウェア構成例を示す。図２は、メモリ１０１にロードされているプログラム群１０３及び補助記憶装置１０２に格納されているデータベース１５０に格納されている情報の例を示す。演算装置１０４に実行されるプログラム群１０３は、ユーザインタフェース（ＵＩ）部１３１、分析処理履歴検索部１３２、バイアス特徴量候補認識部１３３、データ管理部１３４、影響度推定部１３５、及び推奨特徴量認識部１３６を含む。演算装置１０４は、これらプログラムの命令コードを実行することによって、これらプログラムにより実現される機能部として動作する。上述のように、プログラム群１０３は、例えば、補助記憶装置１０２からメモリ１０１にロードされ、演算装置１０４によって実行される。

【００３０】

データベース１５０は、特徴量推奨システム１００が参照又は生成する種々のデータ及び情報を格納している。データベース１５０内のデータは、データ管理部１３４によって管理される。図２に示す例において、データベース１５０は、患者個人情報テーブル１５１、通報イベントテーブル１５２、病院患者情報テーブル１５３、モデル管理情報テーブル１５７を格納している。データベース１５０は、さらに、人関連ＩＤ情報テーブル１５８、準識別子情報テーブル１５９、センシティブ特徴量情報テーブル１６０、バイアス特徴量情報テーブル１６１、バイアス特徴量候補情報テーブル１６２、影響度情報テーブル１６３、フルスケールモデル訓練結果情報テーブル１６４を格納している。

【００３１】

患者個人情報テーブル１５１、通報イベントテーブル１５２、及び病院患者情報テーブ

ル１５３は、医療決定システムの例において収集され得るデータの例であり、特徴量推奨システム１００の分析対象データの例である。医療決定システムにおける学習モデル他ための訓練データは他のデータ、例えば、消防署や気候についてのデータを含むことができる。

【００３２】

図３Ａ、３Ｂ及び３Ｃは、それぞれ、患者個人情報テーブル１５１、通報イベントテーブル１５２、及び病院患者情報テーブル１５３の構成例を示す。分析対象の入力データの例であり、また、分析後においては、これらテーブルの一部又は全レコードが分析に用いられたことが、分析処理履歴に記録される。患者個人情報テーブル１５１は、患者ＩＤで一意に同定される患者の個人情報を格納している。例えば、患者の指名、年齢、性別、住所、収入、症状などのカラムが含まれる。これらカラムは特徴量である。

【００３３】

通報イベントテーブル１５２は、消防署への１１９の通報の情報を示す。各通報はコールＩＤで一意に識別される。通報イベントテーブル１５２は、通報者及び対応者の氏名の他、患者の患者ＩＤや症状等の情報を含む。病院患者情報テーブル１５３は、特定の病人における患者の情報を格納している。患者は、患者ＩＤによって一意に同定される。三つのテーブル１５１、１５２、１５３におけるおいて、共通の患者ＩＤが使用される。患者ＩＤの他、例えば、患者の指名、年齢、性別、銃所、人種、症状などのカラムが、病院患者情報テーブル１５３に含まれる。

【００３４】

図４は、人関連ＩＤ情報テーブル１５８の構成例を示す。人関連ＩＤ情報テーブル１５８は、人関連ＩＤの情報を管理する。人関連ＩＤは、人を一意に識別する識別子である。後述するように、ユーザは、人関連ＩＤを指定することができる。図４の例において、人関連ＩＤ情報テーブル１５８は、番号カラム５８１、テーブルカラム５８２及びカラム名カラム５８３を有する。

【００３５】

番号カラム５８１は、人関連ＩＤ情報テーブル１５８のレコード番号を示す。テーブルカラム５８２は、人関連ＩＤが格納されているテーブル（入力データ）を示す。カラム名カラム５８３は、人関連ＩＤのカラム名を示す。カラム名は、カラムを一意に同定する。

【００３６】

図５は、準識別子情報テーブル１５９の構成例を示す。準識別子情報テーブル１５９は、準識別子の情報を管理する。準識別子は、それ自体で人を同定しないが、複数の準識別子の組み合わせは、当該人の識別子となり得る。

【００３７】

本明細書の一実施形態は、分析対象データにおいて人関連ＩＤ以外の特徴量（カラム）を、準識別子と判定する。図５の例において、準識別子情報テーブル１５９は、番号カラム５９１、テーブルカラム５９２及びカラム名カラム５９３を有する。番号カラム５９１は、準識別子情報テーブル１５９のレコード番号を示す。テーブルカラム５９２は、準識別子が格納されているテーブル（入力データ）を示す。カラム名カラム５９３は、準識別子のカラム名を示す。カラム名は、カラムを一意に同定する。

【００３８】

図６は、センシティブ特徴量情報テーブル１６０の構成例を示す。センシティブ特徴量情報テーブル１６０は、センシティブ特徴量の情報を管理する。センシティブ特徴量は、ユーザによって指定されたバイアス特徴量である。図６の例において、センシティブ特徴量情報テーブル１６０は、番号カラム６０１、テーブルカラム６０２及びカラム名カラム６０３を有する。番号カラム６０１は、センシティブ特徴量情報テーブル１６０のレコード番号を示す。テーブルカラム６０２は、センシティブ特徴量が格納されているテーブル（入力データ）を示す。カラム名カラム６０３は、センシティブ特徴量のカラム名を示す。カラム名は、カラムを一意に同定する。

【００３９】

図７は、バイアス特徴量情報テーブル１６１の構成例を示す。バイアス特徴量情報テーブル１６１は、特徴量推奨システム１００によってバイアス特徴量と判定された特徴量の情報を管理する。センシティブ特徴量は、学習モデルの推定にバイアスを与え、公平性及び／又は正確性を低下させるバイアス特徴量において、ユーザに指定された特徴量である。図７の例において、バイアス特徴量情報テーブル１６１は、番号カラム６１１、テーブルカラム６１２及びカラム名カラム６１３を有する。番号カラム６１１は、バイアス特徴量情報テーブル１６１のレコード番号を示す。テーブルカラム６１２は、バイアス特徴量が格納されているテーブル（入力データ）を示す。カラム名カラム６１３は、バイアス特徴量のカラム名を示す。カラム名は、カラムを一意に同定する。

【００４０】

図８は、バイアス特徴量候補情報テーブル１６２の構成例を示す。バイアス特徴量候補情報テーブル１６２は、特徴量推奨システム１００によってバイアス特徴量と判定された特徴量の情報を管理する。図８の例において、バイアス特徴量候補情報テーブル１６２は、番号カラム６２１、テーブルカラム６２２及びカラム名カラム６２３を有する。番号カラム６２１は、バイアス特徴量候補情報テーブル１６２のレコード番号を示す。テーブルカラム６２２は、バイアス特徴量候補が格納されているテーブル（入力データ）を示す。カラム名カラム６２３は、バイアス特徴量候補のカラム名を示す。カラム名は、カラムを一意に同定する。

【００４１】

図９は、影響度情報テーブル１６３の構成例を示す。影響度情報テーブル１６３は、特徴量推奨システム１００によって計算された、特徴量の学習モデルの推定に対する影響度の情報を管理する。図９の例において、影響度情報テーブル１６３は、番号カラム６１１、テーブルカラム６１２、カラム名カラム６１３、正確性影響度カラム６３４、公平性影響度カラム６３５、及びモデルＩＤカラム６３６を有する。

【００４２】

番号カラム６３１は、影響度情報テーブル１６３のレコード番号を示す。テーブルカラム６３２は、影響度が計算された特徴量が格納されているテーブル（入力データ）を示す。カラム名カラム６３３は、影響度が計算された特徴量のカラム名を示す。カラム名は、カラムを一意に同定する。正確性影響度カラム６３４は、特徴量推奨システム１００により計算された、学習モデルの正確性への特徴量の影響度を示す。公平性影響度カラム６３５、特徴量推奨システム１００により計算された、学習モデルの正確性への特徴量の影響度を示す。モデルＩＤカラム６３６は、影響度の計算のために使用された学習モデルのＩＤを示す。

【００４３】

特徴量の影響度の計算に使用される機械学習モデル（第１学習モデル）は、ユーザが実際に訓練し、実装することを目的とするフルスケールの機械学習モデル（第２学習モデル）と比較して、プログラムサイズが小さく、訓練に必要とされる処理負荷（処理量及び処理時間）が小さい身軽な機械学習モデル（例えばランダムフォレスト）であってよい。これによって、効率的に影響度を計算することができる。

【００４４】

人関連ＩＤ情報テーブル１５８、準識別子情報テーブル１５９、センシティブ特徴量情報テーブル１６０、バイアス特徴量情報テーブル１６１、バイアス特徴量候補情報テーブル１６２、影響度情報テーブル１６３は、分析処理履歴に含まれる。分析処理履歴の情報は、新たな入力データの分析において参照される。

【００４５】

図１０は、特徴量推奨システム１００が実行する処理の概要を示すフローチャートである。特徴量推奨システム１００は、ユーザからの分析対象についての入力を受け付け（Ｓ１１）、入力データを読み出す（Ｓ１２）。図１１は、ユーザが分析対象のデータについて情報を入力するための、ＧＵＩ（Ｇｒａｐｈｉｃａｌ　Ｕｓｅｒ　Ｉｎｔｅｒｆａｃｅ）画面３００の例を示す。ＧＵＩ３００は、ＵＩ部１３１によって出力装置１０６に出力さ

れ、入力されたデータは、ＵＩ部１３１によって受信され、分析処理履歴の一部として、記憶装置に格納される。

【００４６】

ユーザは、ＧＵＩ画面３００において分析対象の入力データを指定する他、人関連ＩＤやセンシティブ特徴量を指定することができる。具体的には、セクション３０１は、分析対象データの指定、及び分析対象データにおけるバイアス候補特徴量の学習モデルの推定に対する影響度の計算の指示を受け付けるためのオブジェクトを含む。図３Ａ～３Ｃの例を参照して説明したように、分析対象データは複数のレコードで構成され、また、複数のカラム（特徴量）で構成されている。各レコードは、複数の特徴量で構成されている。

【００４７】

セクション３０２は、人関連ＩＤの指定を受け付けるためのオブジェクトを含む。例えば、セクション３０２は、指定された分析対象データの特徴量名（カラム名）を表示し、その中から人関連ＩＤの指定を受け付ける。人関連ＩＤ以外の特徴量（カラム）は、準識別子と認識されて、潜在的なバイアス特徴量候補として、特徴量推奨システム１００による分析の対象となる。

【００４８】

セクション３０３は、センシティブ特徴量の指定を受け付けるためのオブジェクトを含む。例えば、セクション３０３は、指定された分析対象データの特徴量名（カラム名）を表示し、その中からセンシティブ特徴量の指定を受け付ける。センシティブ特徴量は、学習モデルの推定にバイアスを与える望ましくないバイアス特徴量としてユーザに指定された特徴量である。

【００４９】

セクション３０４、３０５及び３０６は、特徴量推奨システム１００による分析結果を示す。セクション３０４は、特徴量推奨システム１００によってバイアス特徴量と判定された特徴量を示す。セクション３０５は、特徴量推奨システム１００によって非バイアス特徴量と判定された特徴量を示す。セクション３０６は、バイアス特徴量候補と判定された特徴量の影響度の計算結果を示す。上述のように、セクション３０１において影響度計算を指示するオブジェクトが選択されると、影響度が計算され、セクション３０６に表示される。セクション３０４、３０５及び３０６が示す情報の生成方法の詳細は後述する。

【００５０】

セクション３０６は、さらに、ユーザが目的とする学習モデルの訓練に使用する特徴量の指定を受け付ける。ユーザは、オブジェクト３０７を選択することで、特徴量推奨システム１００に、セクション３０６において選択した特徴量のデータによって、学習モデルを訓練することを指示する。

【００５１】

図１２は、図１０のステップＳ１１の詳細のフローチャートを示す。ＵＩ部１３１は、図１１に示すＧＵＩ画面３００を、出力装置１０６である表示装置において表示する（Ｓ１０１）。ＵＩ部１３１は、指定された対象データの特徴量を表示して、入力装置１０５を介して人関連ＩＤの指定を受け付ける（Ｓ１０２）。ＵＩ部１３１は、受信した人関連ＩＤの情報を、人関連ＩＤ情報テーブル１５８に格納する（Ｓ１０３）。

【００５２】

バイアス特徴量候補認識部１３３は、人関連ＩＤ以外のカラムを準識別子と認識する（Ｓ１０４）。バイアス特徴量候補認識部１３３は、人関連ＩＤと異なる準識別子を潜在的バイアス候補カラムとして、準識別子情報テーブル１５９に格納する（Ｓ１０５）。なお、人関連ＩＤの指定を受けないこともあり得る。その場合、全ての特徴量が準識別子と判定されてよい。

【００５３】

図１０に戻って、分析処理履歴検索部１３２は、補助記憶装置１０２に格納されている分析処理履歴を検索する（Ｓ１３）。本明細書の一実施形態において、分析処理履歴検索部１３２は、入力データの準識別子のセットと一致する準識別子のセットの分析処理済み

データを検索する。同一の準識別子セットの分析処理済みデータが存在する場合（Ｓ１４：ＹＥＳ）、分析処理履歴検索部１３２、その分析結果を分析処理履歴から読み出す（Ｓ１５）。これにより、分析処理の効率が改善される。

【００５４】

本明細書の他の一実施形態において、分析処理履歴検索部１３２は、分析対象の準識別子のセットを含む分析処理済み準識別子セットが存在する場合、その分析結果を読み出してもよい。または、分析処理履歴検索部１３２は、入力データの準識別子毎に、使用できる分析処理結果、例えば、影響度の計算結果を分析処理履歴において検索してもよい。

【００５５】

同一の準識別子セットの分析処理済みデータが存在しないと判定された場合（Ｓ１４：ＮＯ）、影響度推定部１３５は、入力データにおける特徴量の影響度の推定を実行する（Ｓ１６）。以下において、影響度の計算方法の例を説明する。バイアス特徴量候補認識部１３３は、ＵＩ部１３１を介して、センシティブ特徴量のユーザからの指定を受け付ける。センシティブ特徴量の情報は、センシティブ特徴量情報テーブル１６０に格納される。なお、センシティブ特徴量の指定を受けないこともあり得る。

【００５６】

まず、センシティブ特徴量の指定がない場合、又は、指定されたセンシティブ特徴量を区別することなく準識別子の影響度を計算する処理の例を説明する。図１３は、特徴量の影響度の推定処理の例を示すフローチャートである。図１３の処理例において、特徴量推奨システム１００は、センシティブ特徴量の情報を参照することなく準識別子の影響度を計算する。

【００５７】

バイアス特徴量候補認識部１３３は、分析対象の入力データの準識別子のカラム（データ）を読み出す（Ｓ１５１）。バイアス特徴量候補認識部１３３は、カラム間の相関マトリックスを生成する（Ｓ１５２）。カラム（準識別子）間の相関は、例えば、ピアソンの相関係数によって表すことができる。ピアソンの相関係数は、２つのカラムのデータの共分散をそれぞれの標準偏差の積で割ることで算出できる。なお、相関を表す係数の種類は限定されない。

【００５８】

バイアス特徴量候補認識部１３３は、各カラムｉについて、ステップＳ１５３及び１５４を実行する。ステップＳ１５３において、バイアス特徴量候補認識部１３３は、カラムｉとの相関値が閾値を超えるカラムｊをバイアス特徴量候補として選択する。カラムの相関性が高いということは、ある種の「冗長性」を示唆している。これは、情報が繰り返されており、より小さな要素にまとめることができることを意味している。これにより、バイアスを与え得るカラムを検出できる。

【００５９】

ステップＳ１５４において、バイアス特徴量候補認識部１３３は、カラムｊから、最も高い相関のカラムをバイアス特徴量として削除する。これにより、この後の処理効率を改善しつつ、特徴量の減少による訓練の精度の低下を抑制できる。バイアス特徴量候補認識部１３３は、この特徴量の情報を、バイアス特徴量情報テーブル１６１に格納する。

【００６０】

他の例において、相関値が閾値を超える全てのカラムを削除してもよい。これらの例のように、バイアス特徴量は、他の特徴量との間の相関が閾値を超える特徴量から選択される。また、ステップＳ１５３及び１５４は省略されてもよい。削除されたカラムは、この後のステップＳ１５３及びＳ１５４の対象から除外されてよい。

【００６１】

次に、影響度推定部１３５は、残っているカラムのデータによって、影響度を計算するために、機械学習モデル（第１学習モデル）を訓練する（Ｓ１５５）。残っているカラムは、ステップＳ１５３及び１５４において削除されたカラム以外のカラムである。本例において、残っているカラムは、バイアス特徴量候補と判定され、それらの情報はバイアス

特徴量候補情報テーブル１６２に格納される。残っているカラムのデータの全て又は一部によって、学習モデルを訓練できる。

【００６２】

訓練されるモデルは、モデル管理情報テーブル１５７で管理されている。モデル管理情報テーブル１５７は、例えば、モデルの種類、訓練の負荷（影響度計算用か否か）、格納位置等の情報を含むことができる。

【００６３】

上述のように、訓練される学習モデルは、訓練に必要とされる処理量及び時間が小さい身軽な機械学習モデルから選択される。例えば、決定木が使用できる。なお、処理効率に係わらず、訓練する学習モデルが選択されてよい。影響度を計算するための１以上の学習モデルが予め用意されている。例えば、訓練データのデータ量やカラムの数に応じて、それらに関連付けられている一つの学習モデルが選択されてもよい。

【００６４】

影響度推定部１３５は、学習モデルを訓練した後、残っている各カラムｋについて、ステップＳ１５６及びＳ１５７を実行する。ステップＳ１５６において、影響度推定部１３５は、訓練済みの学習モデルを使用して、カラムｋに対するモデルの公平性及び／又は正確性を計算する。この検証に使用されるデータは、上記残っているカラムのデータにおいて、学習モデルの訓練に使用されなかったデータ又は訓練に使用されたデータであってよい。さらに、ステップＳ１５７において、影響度推定部１３５は、計算結果を、カラムｋについての影響度として、影響度情報テーブル１６３に格納する。

【００６５】

影響度の計算方法の例を説明する。なお、影響度の計算方法は、以下に説明する例に限定されない。グループの公平性は、２つのグループの平等な扱いを示す。不公平であることは、望ましくないバイアスの結果、特定の恵まれたグループが他の恵まれていないグループよりも有利になることである。

【００６６】

学習モデルの推定を評価するために、混同行列を使用することができる。陽性又は陰性のラベルを有するテストデータを機械学習モデルに入力する。その推定結果と、ラベルとの比較結果により、以下の数値を得ることができる。陽性を陽性と推定した数（真陽性：Ｔｐ）、陽性を陰性と推定した数（偽陽性：Ｆｐ）、陰性を陰性と推定した数（真陰性：Ｔｎ）、そして、陰性を陽性と推定した数（偽陰性：Ｆｎ）である。これらの数には、２×２の混同行列にまとめることができる。

【００６７】

第１グループと第２グループの間の公平性を探るために、公平性の指標として、例えば、ＥＯＤ（Ｅｑｕａｌ Ｏｐｐｏｒｔｕｎｉｔｙ Ｄｉｆｆｅｒｅｎｃｅ）やＡＯＤ（Ａｖｅｒａｇｅ Ｏｄｄｓ Ｄｉｆｆｅｒｅｎｃｅ）のスコアを使用することができる。分類タスクでは、これらの指標は、混同行列から算出することができる。

【００６８】

ＡＯＤとＥＯＤを算出するために、影響度推定部１３５は、訓練済み学習モデルへ、ラベル付きテストデータを入力し、学習モデルの混同行列を生成する。さらに、影響度推定部１３５は、混同行列から、感度（Ｔｒｕｅ Ｐｏｓｉｔｉｖｅ Ｒａｔｅ：ＴＰＲ）と、偽陽性（Ｆａｌｓｅ Ｐｏｓｉｔｉｖｅ Ｒａｔｅ：ＦＰＲ）のスコアを算出する。

ＴＰＲ＝Ｔｐ／（Ｔｐ＋Ｆｎ）

ＦＰＲ＝Ｆｐ／（Ｆｐ＋Ｔｎ）

【００６９】

ＥＯＤは、第２グループの感度（ＴＰＲｇ２）と、第１グループの感度（ＴＰＲｇ１）から、次のように計算することができる。

ＥＯＤ＝ＴＰＲｇ２－ＴＰＲｇ１

【００７０】

一方、ＡＯＤは、第２グループの偽陽性（ＦＰＲｇ２）及び感度（ＴＰＲｇ２）と、第

１グループの偽陽性（ＦＰＲｇ１）及び感度（ＴＰＲｇ１）から、以下のように計算することができる。

ＡＯＤ＝（ＦＰＲｇ２－ＦＰＲｇ１）＋（ＴＰＲｇ２－ＴＰＲｇ１）／２

【００７１】

ＥＯＤとＡＯＤはともに－１から１の範囲の値を返し、０の値は２つのグループの間で同等であることを意味する。０未満の場合は、第１のグループが第２のグループから利益を得ていることになる。０以上の場合、第２のグループが第１のグループの恩恵を受けていることになる。

【００７２】

精度のスコアには、ＴＮＲ（Ｔｒｕｅ　Ｎｅｇａｔｉｖｅ　Ｒａｔｅ）や感度（ＴＰＲ）を使用することができる。ＴＮＲは誤分類の指標であり、ＴＰＲは正しい分類の指標である。ＴＮＲは，以下のように、真陰性を、偽陽性と真陰性の合計で割った値である。

ＴＮＲ＝Ｔｎ（Ｆｐ＋Ｔｎ）

【００７３】

上記例は、削除された特徴量以外の全ての特徴量の影響度を計算している。本明細書の一実施形態において、バイアス特徴量候補認識部１３３は、相関値が閾値を超える特徴量をバイアス特徴量候補とし、他の特徴量を非バイアス特徴量と判定してもよい。影響度推定部１３５は、バイアス特徴量候補の影響度を計算し、非バイアス特徴量の影響度の計算を省略してもよい。

【００７４】

影響度の計算方法は、上記例に限定されない。公平性の指標及び正確性の指標の一方のみが計算され、ユーザに提示されてもよい。上記二つの公平性の指標のうちの一方のみが算出及び提示されてもよく、上記二つの正確性の指標の一方のみが算出及び提示されてもよい。

【００７５】

バイアス特徴量候補認識部１３３は、センシティブ特徴量との相関が閾値を超え、最も高い相関を示す準識別子を削除してもよい。バイアス特徴量候補認識部１３３は、センシティブ特徴量との相関が閾値を超える準識別子を、バイアス特徴量候補と判定し、それ以外の準識別子を非バイアス特徴量と判定してもよい。影響度は、バイアス特徴量候補について計算される。

【００７６】

次に、特徴量の影響度の推定処理の他の例を説明する。以下において、ユーザに指定されたセンシティブ特徴量との相関に基づいて、バイアス特徴量又はバイアス特徴量候補を判定する。図１４は、特徴量の影響度の推定処理の例を示すフローチャートである。図１４の処理例において、特徴量推奨システム１００は、センシティブ特徴量の情報を参照して準識別子の影響度を計算する。これにより、学習モデルの推定の公平性及び／又は正確性を低下させ得るバイアスと特徴量をより適切に検出できる。

【００７７】

バイアス特徴量候補認識部１３３は、分析対象の入力データの準識別子のカラム（データ）を読み出す（Ｓ１７１）。バイアス特徴量候補認識部１３３は、各センシティブカラムについて、ステップＳ１７２及びＳ１７３を実行する。センシティブ特徴量は、入力データに対してユーザによって指定されている。他の例において、分析処理履歴に格納されているセンシティブ特徴量の情報を参照してもよい。

【００７８】

ステップＳ１７２において、バイアス特徴量候補認識部１３３は、センシティブ特徴量と、他の準識別子との間の相関を計算する。相関の計算方法は上述の通りである。ステップＳ１７３において、バイアス特徴量候補認識部１３３は、相関が閾値を超えるカラム（準識別子）を、バイアス特徴量として削除する。これにより、学習モデルにバイアスを与える可能性が高い特徴量を削除できる。

【００７９】

バイアス特徴量は、バイアス特徴量情報テーブル１６１に格納される。他の例において、バイアス特徴量候補認識部１３３は、閾値を超えてかつ最も高い相関を示すカラムのみを、削除してもよい。バイアス特徴量は、センシティブ特徴量との間の相関が閾値を超える特徴量から選択される。バイアス特徴量は、後述する影響度計算から除外され、処理を効率化できる。

【００８０】

ステップ１７４において、バイアス特徴量候補認識部１３３は、全センシティブ特徴量それぞれとの相関が第２閾値未満のカラムを、非バイアス特徴量と判定する。第２閾値は、ステップＳ１７３における閾値より小さい。バイアス特徴量候補は、センシティブ特徴量、バイアス特徴量又は非バイアス特徴量と判定された準識別子、以外の準識別子である。バイアス特徴量候補認識部１３３は、バイアス特徴量候補の情報を、バイアス特徴量候補情報テーブル１６２に格納する。

【００８１】

次に、影響度推定部１３５は、残っているカラムによって、影響度を計算するための機械学習モデルを訓練する（Ｓ１７５）。残っているカラムは、ステップＳ１７２及び１７３において削除されたカラム以外のカラムである。訓練されるモデルは、上記例と同様である。

【００８２】

影響度推定部１３５は、学習モデルを訓練した後、バイアス特徴量候補のカラムについて、ステップＳ１７６及びＳ１７７を実行する。ステップＳ１７６において、影響度推定部１３５は、訓練済みの学習モデルを使用して、バイアス特徴量候補ｋに対するモデルの公平性及び／又は正確性を計算する。さらに、ステップＳ１７７において、影響度推定部１３５は、計算結果を、バイアス特徴量候補ｋについての影響度として、影響度情報テーブル１６３に格納する。なお、非バイアス特徴量を判定することなく、センシティブ特徴量及びバイアス特徴量以外の全ての準識別子の影響度を計算してもよい。

【００８３】

図１０に戻って、ＵＩ部１３１は、分析処理結果を、出力装置１０６である表示装置においてユーザに提示する（Ｓ１７）。具体的には、ＵＩ部１３１は、入力データの分析結果を、図１１に示すＧＵＩ画面３００において表示する。図１１に示すように、セクション３０４は、特徴量推奨システム１００によってバイアス特徴量と判定された特徴量を示す。セクション３０５は、特徴量推奨システム１００によって非バイアス特徴量と判定された特徴量を示す。セクション３０６は、バイアス特徴量候補と判定された特徴量の影響度の計算結果を示す。

【００８４】

セクション３０４、３０５は省略され得る。このように、特徴量の影響度を提示することで、ユーザは、目的の機械学習モデルの訓練から除外する（訓練に使用する）特徴量を適切に選択することができる。図１１に示す例において、バイアス特徴量は訓練から排除され、及び非バイアス特徴量は訓練に使用される。他の例において、これら特徴量において、訓練に使用する特徴量がユーザによって選択可能であってもよい。

【００８５】

上述のように、図１１に示すＧＵＩ画面を介して、ユーザは、データの入力、センシティブ特徴量の定義、バイアス特徴量の検出／除去、訓練時間が短い効率的なモデル（ランダムフォレストなど）の訓練による影響度の計算を、行うことができる。ユーザが入力データの分析結果に満足すると、次に、ユーザは、入力データによる訓練の目的とする、実際のフルスケールのモデル（ＸＢｏｏｓｔ、Ａｄａｂｏｏｓｔ、ＬｉｇｈｔＧＢＭなど）を訓練することができる。ＵＩ部１３１は、オブジェクト３０７の選択によって、フルスケールモデルの訓練の指示を受け付ける。

【００８６】

影響度推定部１３５は、入力データにおいて選択された準識別子のカラムによって、フルスケールモデルの訓練を行うと共に、公正さや精度のスコアを算出する。影響度推定部

１３５は、訓練の結果をフルスケールモデル訓練結果情報テーブル１６４に格納する。ＵＩ部１３１は、フルスケールモデルの訓練結果を、出力装置１０６である表示装置において表示する。

【００８７】

図１５は、フルスケールモデルの訓練結果を示す画像の例を示す。図１５は、訓練データからバイアス特徴量が除去され、その訓練データによってモデルが訓練され、各モデルの公平性／正確性のスコアが算出された後の出力を示す。図１５の例において、ＥＯＤ、ＡＯＤが公平性スコアとして算出され、ＴＰＲ、ＴＮＲが精度のスコアとして提示されている。図１５の例は、さらに、適合率（ｐｒｅｃｉｓｉｏｎ）＝ＴＰ／（ＴＰ＋ＦＰ）と、Ｆ１値を示す。Ｆ１値は、感度と適合率から計算され、これらのバランスを示す。

【００８８】

ユーザが結果に満足した場合、モデルを本番環境にデプロイしてもよい。図１５の例において、ユーザは、複数の機械学習モデルをトレーニングして、公正さや精度のスコアを比較することができる。

【００８９】

なお、本発明は上記した実施形態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、上記した実施形態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、ある実施形態の構成の一部を他の実施形態の構成に置き換えることが可能であり、また、ある実施形態の構成に他の実施形態の構成を加えることも可能である。また、各実施形態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。

【００９０】

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。

【００９１】

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】

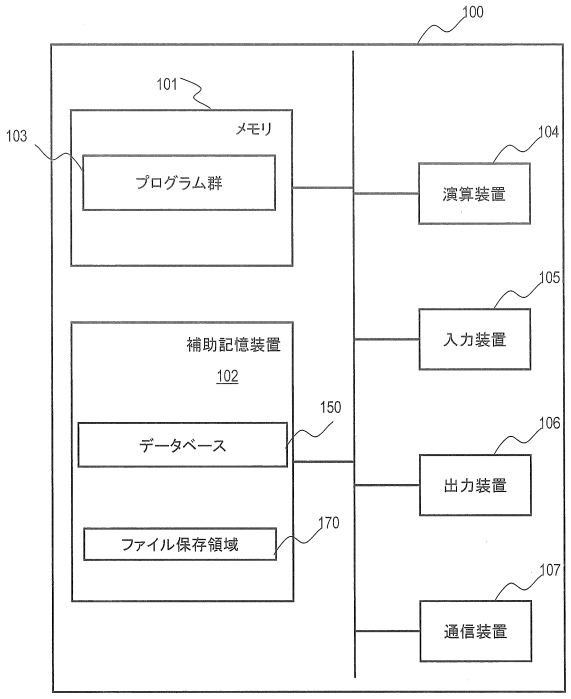
【００９２】

１００　特徴量推奨システム
１０１　メモリ
１０２　補助記憶装置
１０４　演算装置
１０６　出力装置
１０７　通信装置
１５１　患者個人情報テーブル
１５２　通報イベントテーブル
１５３　病院患者情報テーブル
１５７　モデル管理情報テーブル
１５８　人関連ＩＤ情報テーブル
１５９　準識別子情報テーブル
１６０　センシティブ特徴量情報テーブル
１６１　バイアス特徴量情報テーブル
１６２　バイアス特徴量候補情報テーブル
１６３　影響度情報テーブル

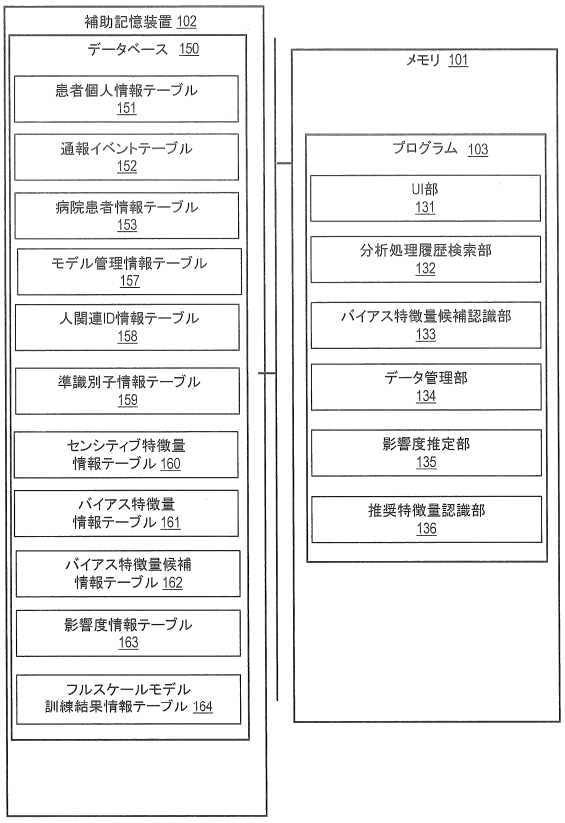
１６４　フルスケールモデル訓練結果情報テーブル

【図面】

【図１】



【図２】



【図３Ａ】

PATIENT ID	NAME	AGE	SEX	ADDRESS	INCOME	SYMPTOM	...
100	HANAKO HITACHI	80	1	1-2-3 KOKUBUJI-SHI, TOKYO	xxxxxxx	yyyy	...
101	TARO HITACHI	40	0	23-10 KOKUBUNJI-SHI, TOKYO	xxxxxxx	yyyy	...
...	...	...	...	...	...	...	...

患者個人情報テーブル　151

【図３Ｂ】

CALL_ID	CALLER	RECEIVER	RECEIVED TIME	PATIENT ID	SYMPTOM	SYPTOM2	...
1000	KEI KITANO	HAJIME ONO	2020-03-20 11:22:33	100	yyyy	yyyy	...
1001	SAE IIDA	HAJIME ONO	2020-03-21 11:22:33	101	yyyy	yyyy	...
...	...	...	...	...	...	...	...

通報イベントテーブル　152

【図３Ｃ】

PATIENT ID	NAME	AGE	SEX	ADDRESS	RACE	SYMPTOM	...
1000	ICHIRO IMAI	80	1	1-2-3 KOKUBUJI-SHI, TOKYO	R1	yyyy	...
1001	MIWA UENO	40	0	23-10 KOKUBUNJI-SHI, TOKYO	R2	yyyy	...
...	...	...	...	...	...	...	...

病院患者情報テーブル 153

【図４】

158

581 NO	582 TABLE	583 COLUMN NAME
1	PATIENT_PRIVATE	NAME
2	HOSPITAL_PATIENT	ID
...	...	...

人関連ID情報テーブル

【図５】

159

591 NO	592 TABLE	593 COLUMN NAME
1	HOSPITAL_PATIENT	AGE
2	HOSPITAL_PATIENT	RACE
...	...	...

準識別子情報テーブル

【図６】

160

601 NO	602 TABLE	603 COLUMN NAME
1	PATIENT_PRIVATE	SEX
2	HOSPITAL_PATIENT	RACE
...	...	...

センシティブ特徴量情報テーブル

【図７】

161

611 NO	612 TABLE	613 COLUMN NAME
1	PATIENT_PRIVATE	INCOME
...	...	...

バイアス特徴量情報テーブル

【図８】

162

621 NO	622 TABLE	623 COLUMN NAME
1	PATIENT_PRIVATE	AGE
2	PATIENT_PRIVATE	ADDRESS
3	PATIENT_PRIVATE	SYMPTOM
...	...	...

バイアス特徴量候補情報テーブル

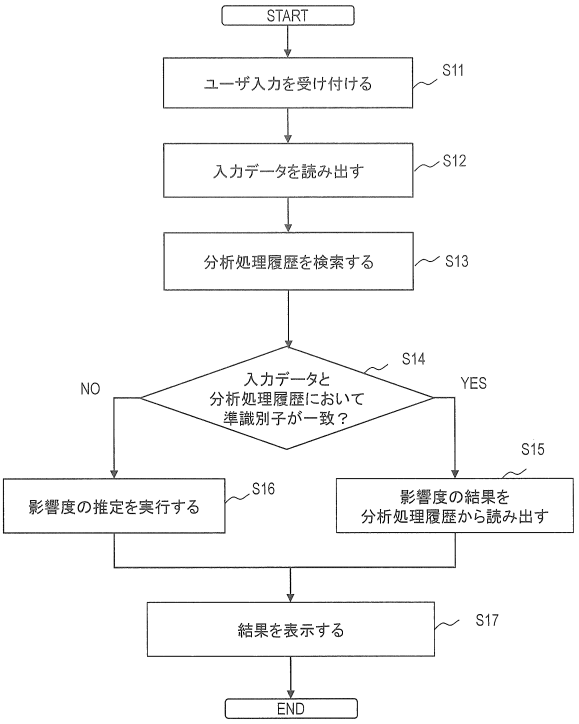
【図９】

163

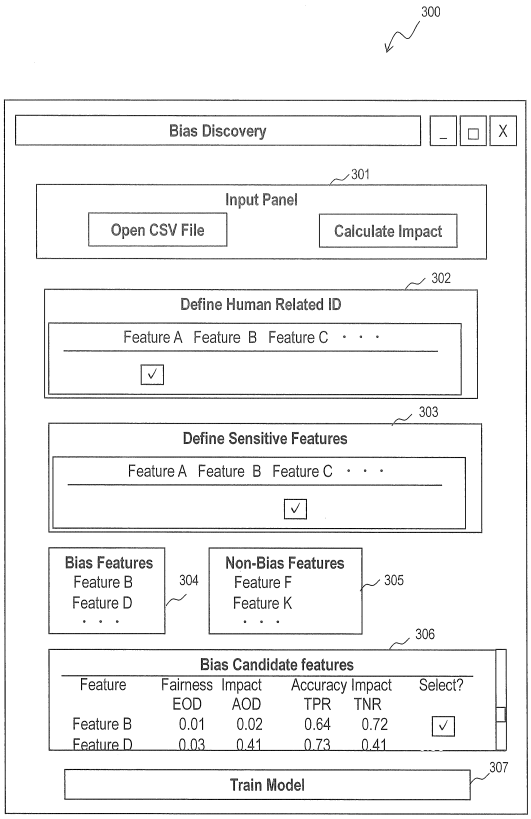
631 NO	632 TABLE	633 COLUMN NAME	634 IMPACT FOR ACCURACY	635 IMPACT FOR FAIRNESS	636 MODEL ID
1	PATIENT_PRIVATE	SYMPTOM	4.3	0.7	12
2	PATIENT_PRIVATE	AGE	4.0	0.1	45
...	...	....	...	...	...

影響度情報テーブル

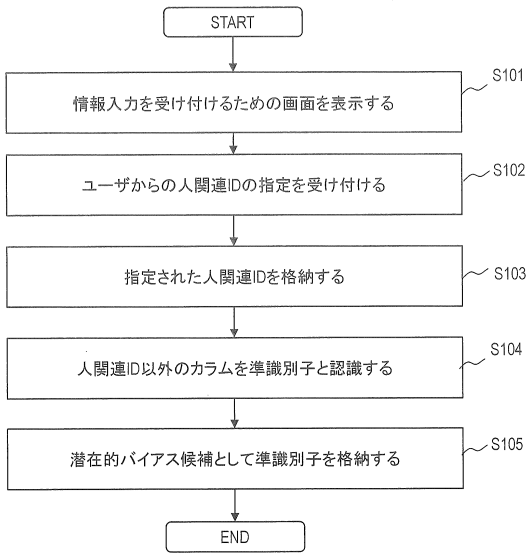
【図１０】



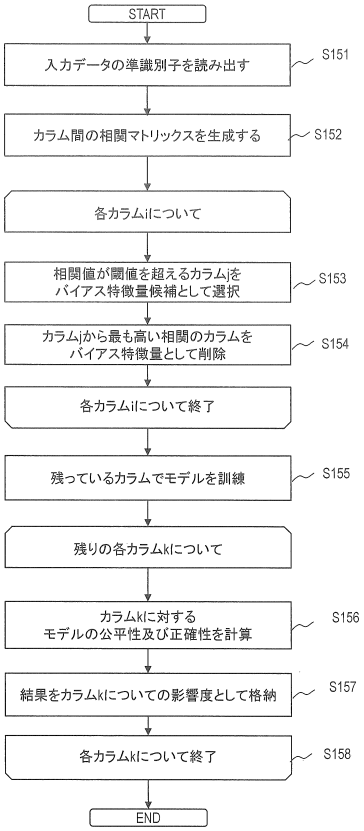
【図１１】



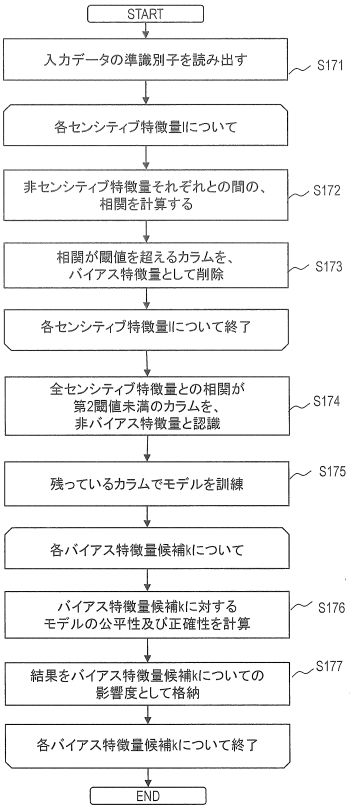
【図１２】



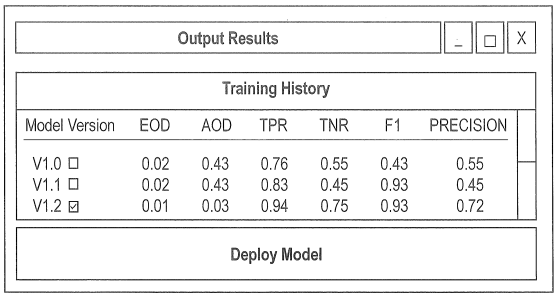
【図１３】


START
入力データの準識別子を読み出す
S151
カラム間の相関マトリックスを生成する
S152
各カラムiについて
相関値が閾値を超えるカラムjを
バイアス特徴量候補として選択
S153
カラムjから最も高い相関のカラムを
バイアス特徴量として削除
S154
各カラムiについて終了
残っているカラムでモデルを訓練
S155
残りの各カラムkについて
カラムkに対する
モデルの公平性及び正確性を計算
S156
結果をカラムkについての影響度として格納
S157
各カラムkについて終了
S158
END

【図１４】


START
入力データの準識別子を読み出す
S171
各センシティブ特徴量Iについて
非センシティブ特徴量それぞれとの間の、
相関を計算する
S172
相関が閾値を超えるカラムを、
バイアス特徴量として削除
S173
各センシティブ特徴量Iについて終了
全センシティブ特徴量との相関が
第2閾値未満のカラムを、
非バイアス特徴量と認識
S174
残っているカラムでモデルを訓練
S175
各バイアス特徴量候補kについて
バイアス特徴量候補kに対する
モデルの公平性及び正確性を計算
S176
結果をバイアス特徴量候補kについての
影響度として格納
S177
各バイアス特徴量候補kについて終了
END

【図１５】


Output Results
_
□
X
Training History
Model Version EOD AOD TPR TNR F1 PRECISION
V1.0 ☐ 0.02 0.43 0.76 0.55 0.43 0.55
V1.1 ☐ 0.02 0.43 0.83 0.45 0.93 0.45
V1.2 ☑ 0.01 0.03 0.94 0.75 0.93 0.72
Deploy Model

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 佐藤 直樹

(56)参考文献 特開２０２１－１１１０６０（ＪＰ，Ａ）
特開２０２０－０９５３９８（ＪＰ，Ａ）

(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０６Ｎ ２０／００