

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7384705号  
(P7384705)

(45)発行日 令和5年11月21日(2023.11.21)

(24)登録日 令和5年11月13日(2023.11.13)

(51)国際特許分類

F I

G 1 6 H 10/00 (2018.01)

G 1 6 H 10/00

G 0 6 N 20/00 (2019.01)

G 0 6 N 20/00 1 3 0

請求項の数 9 (全24頁)

(21)出願番号	特願2020-33769(P2020-33769)	(73)特許権者	000005108
(22)出願日	令和2年2月28日(2020.2.28)		株式会社日立製作所
(65)公開番号	特開2021-135930(P2021-135930 A)	(74)代理人	東京都千代田区丸の内一丁目 6 番 6 号 110001678
(43)公開日	令和3年9月13日(2021.9.13)		藤央弁理士法人
審査請求日	令和4年4月8日(2022.4.8)	(72)発明者	垂水 信二
			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	竹内 渉
			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	チャルキディス ジョージ
			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	由井 俊太郎
			最終頁に続く

(54)【発明の名称】 分析装置、分析方法、および分析プログラム

(57)【特許請求の範囲】

【請求項 1】

プログラムを実行するプロセッサと、前記プログラムを記憶する記憶デバイスと、を有する分析装置であって、  
前記プロセッサは、  
入力層、1 以上の中間層、および出力層を有し、前記入力層に割り当てられる対象者への実施行為に関する訓練データ群と、前記出力層に割り当てられる前記実施行為前の状態に関する実施行為前正解データ群および前記実施行為後に関する実施行為後正解データ群と、に基づいて学習された学習モデルに、前記訓練データ群の各々の訓練データを入力することにより、前記中間層から第 1 特徴量データ群を算出する第 1 算出処理と、  
前記学習モデルの予測対象データを入力することにより、前記中間層から第 2 特徴量データを算出する第 2 算出処理と、  
前記第 1 算出処理によって算出された第 1 特徴量データ群の中から、前記第 2 算出処理によって算出された第 2 特徴量データに類似する特定の第 1 特徴量データを検索する検索処理と、  
前記検索処理によって検索された特定の第 1 特徴量データの算出元となる特定の訓練データを前記実施行為に関する訓練データ群から抽出する抽出処理と、  
前記第 1 特徴量データ群を複数のクラスタに分類するクラスタリング処理と、  
前記クラスタリング処理によって分類された複数のクラスタの各々について、前記クラスタ内の第 1 特徴量データの算出元となる特定の訓練データおよび当該特定の訓練データに

10

20

対応する特定の実施行為前正解データを含むクラスタリングに関する訓練データと、当該特定の訓練データに対応する特定の実施行為後正解データであるクラスタリングに関する正解データとに基づいて、予測モデルを生成する生成処理と、  
前記複数のクラスタの中から前記第 2 特徴量データの所属先クラスタを特定する特定処理と、

前記特定処理によって特定された所属先クラスタの予測モデルを前記生成処理によって生成された複数の予測モデルの中から取得する取得処理と、

前記取得処理によって取得された予測モデルに前記予測対象データを入力することにより、予測結果データを出力する出力処理と、

を実行することを特徴とする分析装置。

10

【請求項 2】

請求項 1 に記載の分析装置であって、

前記検索処理では、前記プロセッサは、前記第 1 特徴量データ群の各々の第 1 特徴量データと前記第 2 特徴量データとの類似度を算出し、前記類似度に基づいて、前記第 1 特徴量データ群の中から前記特定の第 1 特徴量データを検索する、

ことを特徴とする分析装置。

【請求項 3】

請求項 2 に記載の分析装置であって、

前記検索処理では、前記プロセッサは、前記類似度が所定のしきい値以上となる第 1 特徴量データを前記特定の第 1 特徴量データとして検索する、

ことを特徴とする分析装置。

20

【請求項 4】

請求項 1 に記載の分析装置であって、

前記抽出処理では、前記プロセッサは、前記特定の訓練データに対応する特定の実施行為前正解データおよび前記特定の実施行為後正解データを前記実施行為前正解データ群および前記実施行為後正解データ群から抽出する、

ことを特徴とする分析装置。

【請求項 5】

請求項 1 に記載の分析装置であって、

前記プロセッサは、

前記特定の訓練データに関する統計値を算出する統計処理を実行する、

ことを特徴とする分析装置。

30

【請求項 6】

請求項 1 に記載の分析装置であって、

前記プロセッサは、

前記特定の実施行為前正解データに関する統計値を算出する統計処理を実行する、

ことを特徴とする分析装置。

【請求項 7】

請求項 1 に記載の分析装置であって、

前記プロセッサは、

前記特定の実施行為後正解データに関する統計値を算出する統計処理を実行する、

ことを特徴とする分析装置。

40

【請求項 8】

プログラムを実行するプロセッサと、前記プログラムを記憶する記憶デバイスと、を有する分析装置が実行する分析方法であって、

前記分析方法は、

前記プロセッサが、

入力層、1 以上の中間層、および出力層を有し、前記入力層に割り当てられる対象者への実施行為に関する訓練データ群と、前記出力層に割り当てられる前記実施行為前の状態に関する実施行為前正解データ群および前記実施行為後に関する実施行為後正解データ群と

50

に基づいて学習された学習モデルに、前記訓練データ群の各々の訓練データを入力することにより、前記中間層から第1特徴量データ群を算出する第1算出処理と、  
前記学習モデルの予測対象データを入力することにより、前記中間層から第2特徴量データを算出する第2算出処理と、  
前記第1算出処理によって算出された第1特徴量データ群の中から、前記第2算出処理によって算出された第2特徴量データに類似する特定の第1特徴量データを検索する検索処理と、  
前記検索処理によって検索された特定の第1特徴量データの算出元となる特定の訓練データを前記実施行為に関する訓練データ群から抽出する抽出処理と、  
前記第1特徴量データ群を複数のクラスタに分類するクラスタリング処理と、  
前記クラスタリング処理によって分類された複数のクラスタの各々について、前記クラスタ内の第1特徴量データの算出元となる特定の訓練データおよび当該特定の訓練データに対応する特定の実施行為前正解データを含むクラスタリングに関する訓練データと、当該特定の訓練データに対応する特定の実施行為後正解データであるクラスタリングに関する正解データとに基づいて、予測モデルを生成する生成処理と、  
前記複数のクラスタの中から前記第2特徴量データの所属先クラスタを特定する特定処理と、  
前記特定処理によって特定された所属先クラスタの予測モデルを前記生成処理によって生成された複数の予測モデルの中から取得する取得処理と、  
前記取得処理によって取得された予測モデルに前記予測対象データを入力することにより、予測結果データを出力する出力処理と、  
を実行することを特徴とする分析方法。

【請求項9】

プロセッサに、  
入力層、1以上の中間層、および出力層を有し、前記入力層に割り当てられる対象者への実施行為に関する訓練データ群と、前記出力層に割り当てられる前記実施行為前の状態に関する実施行為前正解データ群および前記実施行為後に関する実施行為後正解データ群と、  
に基づいて学習された学習モデルに、前記訓練データ群の各々の訓練データを入力することにより、前記中間層から第1特徴量データ群を算出する第1算出処理と、  
前記学習モデルの予測対象データを入力することにより、前記中間層から第2特徴量データを算出する第2算出処理と、  
前記第1算出処理によって算出された第1特徴量データ群の中から、前記第2算出処理によって算出された第2特徴量データに類似する特定の第1特徴量データを検索する検索処理と、  
前記検索処理によって検索された特定の第1特徴量データの算出元となる特定の訓練データを前記実施行為に関する訓練データ群から抽出する抽出処理と、  
前記第1特徴量データ群を複数のクラスタに分類するクラスタリング処理と、  
前記クラスタリング処理によって分類された複数のクラスタの各々について、前記クラスタ内の第1特徴量データの算出元となる特定の訓練データおよび当該特定の訓練データに対応する特定の実施行為前正解データを含むクラスタリングに関する訓練データと、当該特定の訓練データに対応する特定の実施行為後正解データであるクラスタリングに関する正解データとに基づいて、予測モデルを生成する生成処理と、  
前記複数のクラスタの中から前記第2特徴量データの所属先クラスタを特定する特定処理と、  
前記特定処理によって特定された所属先クラスタの予測モデルを前記生成処理によって生成された複数の予測モデルの中から取得する取得処理と、  
前記取得処理によって取得された予測モデルに前記予測対象データを入力することにより、予測結果データを出力する出力処理と、  
を実行させるための分析プログラム。

【発明の詳細な説明】

10

20

30

40

50

## 【技術分野】

## 【0001】

本発明は、データを分析する分析装置、分析方法、および分析プログラムに関する。

## 【背景技術】

## 【0002】

実世界ヘルスケアデータを活用した効果的かつ効率的な診断支援やヘルスケア関連サービスの実現が求められている。従来型のヘルスケアは、臨床知見に基づく統一ガイドラインに則って実施されるが、この知見は全症例の10%以下の臨床研究にのみ基づくとの報告もあり、本来あるべき個別最適なヘルスケアの実現には至っていなかった。

## 【0003】

そこで、最適なヘルスケアサービスを提供するため、実世界ヘルスケアデータ分析に基づき、医師や保健指導者などのサービス提供者を支援する技術の実現が期待されている。特に、個人に実施した医療サービス（投薬、生活習慣指導、介護サービスなど）の効果や質を、実データを活用して評価および予測する技術に注目が集まっている。たとえば、以下の技術が開示されている。

## 【0004】

特許文献1には、「投薬効果情報送信装置17は、電子カルテサーバー16から患者への投薬終了を表す投薬終了通知が送信された場合に、投薬により患者に表れた投薬効果に関する投薬効果情報を、画像サーバー14及び電子カルテサーバー16から収集する。投薬効果情報送信装置17で収集された投薬効果情報は、投薬効果情報データベース54に格納される。クライアント端末18により、薬剤名及び患者の属性情報に基づいて投薬効果情報サーバー55を検索すると、投薬効果情報サーバー55は、検索した薬剤の平均的な投薬効果を表す平均投薬効果情報をクライアント端末18に送信する。クライアント端末18は、平均投薬効果情報を時系列にモニターに表示する。」との記載がある。

## 【0005】

特許文献2には、「プロセッサにより実行可能な命令のセットを記憶する非一時的なコンピュータ可読記憶媒体。命令のセットは、現在の患者に関連するデータの現在の患者のセットを受信し、現在の患者のデータセットを複数の以前の患者データセット（それぞれが以前の患者に対応）と比較し、選択された以前の患者のデータセットと現在の患者のデータセットとの間の類似性のレベルに基づいて、以前の患者のデータセットの1つを選択し、選択した以前の患者のデータセットをユーザーに提供するように動作可能である。」との記載がある。

## 【先行技術文献】

## 【特許文献】

## 【0006】

【文献】特開2014-71592号公報

【文献】国際公開第2012/080906号

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【0007】

医療サービスのような介入手段は、その種類によって実施の頻度が異なるため、過去の実施例が少ない場合、介入手段を訓練データとして当該医療サービスの効果を分析することは容易ではない。特に、複数の医療サービスを組み合わせた場合の効果を分析する場合、この課題は顕著であり、当該の組み合わせと完全に合致する実施例が0件となることもある。このような問題は、医療サービスに限らず、他のサービスでも起こりえる問題である。

## 【0008】

本発明は、過去の訓練データとの異同や実施回数にかかわらず、分析精度の向上を図ることを目的とする。

## 【課題を解決するための手段】

## 【 0 0 0 9 】

本願において開示される発明の一側面となる分析装置は、プログラムを実行するプロセッサと、前記プログラムを記憶する記憶デバイスと、を有する分析装置であって、入力層、1 以上の中間層、および出力層を有し、前記入力層に割り当てられる対象者への実施行為に関する訓練データ群と、前記出力層に割り当てられる前記実施行為前の状態に関する実施行為前正解データ群および前記実施行為後に関する実施行為後正解データ群と、に基づいて学習された学習モデルに、前記訓練データ群の各々の訓練データを入力することにより、前記中間層から第1 特徴量データ群を算出する第1 算出処理と、前記学習モデルの予測対象データを入力することにより、前記中間層から第2 特徴量データを算出する第2 算出処理と、前記第1 算出処理によって算出された第1 特徴量データ群の中から、前記第2 算出処理によって算出された第2 特徴量データに類似する特定の第1 特徴量データを検索する検索処理と、前記検索処理によって検索された特定の第1 特徴量データの算出元となる特定の訓練データを前記実施行為に関する訓練データ群から抽出する抽出処理と、前記第1 特徴量データ群を複数のクラスタに分類するクラスタリング処理と、前記クラスタリング処理によって分類された複数のクラスタの各々について、前記クラスタ内の第1 特徴量データの算出元となる特定の訓練データおよび当該特定の訓練データに対応する特定の実施行為前正解データを含むクラスタリングに関する訓練データと、当該特定の訓練データに対応する特定の実施行為後正解データであるクラスタリングに関する正解データとに基づいて、予測モデルを生成する生成処理と、前記複数のクラスタの中から前記第2 特徴量データの所属先クラスタを特定する特定処理と、前記特定処理によって特定された所属先クラスタの予測モデルを前記生成処理によって生成された複数の予測モデルの中から取得する取得処理と、前記取得処理によって取得された予測モデルに前記予測対象データを入力することにより、予測結果データを出力する出力処理と、を実行することを特徴とする。

10

20

## 【 発明の効果 】

## 【 0 0 1 0 】

本発明の代表的な実施の形態によれば、過去の介入手段との異同や実施回数にかかわらず、分析精度の向上を図ることができる。前述した以外の課題、構成及び効果は、以下の実施例の説明により明らかにされる。

## 【 図面の簡単な説明 】

30

## 【 0 0 1 1 】

【 図 1 】 図 1 は、実施例 1 にかかる分析装置によるヘルスケアデータ分析例を示す説明図である。

【 図 2 】 図 2 は、分析装置のハードウェア構成例を示すブロック図である。

【 図 3 】 図 3 は、未整形ヘルスケア情報の一例を示す説明図である。

【 図 4 】 図 4 は、実施例 1 にかかる分析装置による学習モデル生成処理手順例を示すフローチャートである。

【 図 5 】 図 5 は、整形ヘルスケア情報の一例を示す説明図である。

【 図 6 】 図 6 は、ニューラルネットワークの一例を示す説明図である。

【 図 7 】 図 7 は、ニューラルネットワークの他の例を示す説明図である。

40

【 図 8 】 図 8 は、特徴量情報生成処理手順例を示すフローチャートである。

【 図 9 】 図 9 は、実施例 1 にかかる予測対象未整形ヘルスケア情報の一例を示す説明図である。

【 図 1 0 】 図 1 0 は、実施例 1 にかかる予測対象整形ヘルスケア情報の一例を示す説明図である。

【 図 1 1 】 図 1 1 は、実施例 1 にかかる分析装置による分析処理手順例を示すフローチャートである。

【 図 1 2 】 図 1 2 は、統計処理（ステップ S 1 1 1 0 ）の結果例を示す説明図である。

【 図 1 3 】 図 1 3 は、実施例 2 にかかるクラスタの一例を示す説明図である。

【 図 1 4 】 図 1 4 は、図 1 3 に示したクラスタを用いた予測モデルの生成例を示す説明図

50

である。

【図 1 5】図 1 5 は、実施例 2 にかかる分析装置による予測モデル生成処理手順例を示すフローチャートである。

【図 1 6】図 1 6 は、実施例 2 にかかる分析装置による予測処理手順例を示すフローチャートである。

【図 1 7】図 1 7 は、実施例 2 にかかる予測対象未整形ヘルスケア情報の一例を示す説明図である。

【図 1 8】図 1 8 は、実施例 2 にかかる予測対象整形ヘルスケア情報の一例を示す説明図である。

【発明を実施するための形態】

【実施例 1】

【0012】

＜ヘルスケア分析例＞

図 1 は、実施例 1 にかかる分析装置によるヘルスケアデータ分析例を示す説明図である。(1) 分析装置は、訓練データとなる介入手段情報 101 と、正解データ 102 である患者背景情報 102A および介入効果情報 102B の組み合わせと、を取得する。介入手段情報 101 とは、介入手段を示す情報であり、介入手段データの集合である。介入手段情報 101 の各レコードが、患者ごとの介入手段データである。図 1 では、介入手段情報 101 は、患者 a ~ c の 3 人の介入手段データ 101a、101b、101c により構成される。

【0013】

介入手段とは、対象者（たとえば、患者や被験者）に介入する医療サービス（投薬、生活指導、介護サービスなど）である。「介入」とは、健康増進、疾病予防、病気の治療などを目的とした健康指導、介助、自立支援、投薬、外科的な処置などを対象者に実施する行為を意味する。介入手段データは、たとえば、投与した医薬品の有無や生活指導といった医療サービスの実施有無（「1」が有り、「0」が無し）を含む。すなわち、介入手段データは、患者に提供された 1 以上の医療サービスの組み合わせを規定する。

【0014】

患者背景情報 102A は、患者の背景を示す情報であり、患者ごとの患者背景データの集合である。患者背景情報 102A の各レコードが、患者ごとの患者背景データである。患者背景情報 102A は、患者 a ~ c の 3 人の患者背景データ 102Aa、102Ab、102Ac により構成される。「Pre HbA1c」は、介入前の HbA1c の値を示す。

【0015】

介入効果情報 102B は、介入効果を示す情報であり、介入効果データの集合である。介入効果とは、介入によりもたらされる結果であり、たとえば、治療アウトカムである。介入効果情報 102B の各レコードが、患者ごとの介入効果データである。図 1 では、介入効果情報 102B は、患者 a ~ c の 3 人の介入効果データ 102Ba、102Bb、102Bc により構成される。「Post HbA1c」は、介入後の HbA1c（ヘモグロビンエーワンシー）の値を示す。

【0016】

分析装置は、介入手段情報 101（訓練データ）および正解データ 102 を訓練データセットとしてニューラルネットワーク 103 の入力層 131 および出力層 133 に与えて学習し、学習モデルを生成する。学習モデルは、学習パラメータ（重みパラメータおよびバイアス。ハイパーパラメータを含めてもよい（以下同様）。）が設定されたニューラルネットワーク 103（以降、学習モデル 103 と称す。）である。

【0017】

(2) 分析装置は、特徴量情報 104 を生成する。特徴量情報 104 は、学習モデル 103 の内部表現である。特徴量情報 104 は、患者ごとの特徴量データの集合である。特徴量データは学習モデル 103 の 1 以上の中間層の中の特定の間層 132 を構成する各

10

20

30

40

50

ニューロンの計算結果である。特徴量データは、ニューロン数分の次元の特徴量空間内の点になる。2人の患者間の患者背景データおよび介入効果データがともに類似すると、2人の患者間の特徴量データも類似する。

【0018】

また、図1では、中間層132のニューロンは3個であるため、特徴量1～3の値を含む特徴量データが患者ごとに計算される。このため、特徴量情報104は、患者a～cの3人の特徴量データ104a、104b、104cにより構成される。

【0019】

(3)分析装置は、予測対象介入手段情報111内の患者zの予測対象介入手段データ111zを、学習モデル103に入力し、学習モデル103の特定の間層132から患者zの特徴量データ114zを算出する。

10

【0020】

(4)分析装置は、(2)で算出した特徴量情報104の中から(3)で算出した特徴量データに類似する特徴量データを検索する。特徴量データは、特定の間層を構成するニューロン数分の要素(図1では、特徴量1～3の3要素)を持つ1次元ベクトルであるため、特徴量空間内の特徴量データ114zからのベクトル間距離が所定距離以内である特徴量情報104内の特徴量データは、特徴量データ114zに類似する特徴量データとなる。図1では、特徴量データ104aが特徴量データ114に類似する特徴量データに該当したとする(以下、類似特徴量データ104aと称す)。分析装置は、類似特徴量データ104aの算出元となる同一患者aの介入手段データ101aを取得する。

20

【0021】

このように、特徴量データ114z、104aが類似するということは、患者zの予測対象介入手段データ111zと、患者aの介入手段データ101aとが、患者背景と介入効果のふたつの観点で共に類似することを意味する。介入効果を分析する際には、過去に類似した介入を行った事例を分析することが好ましいため、分析装置は患者zに対する予測対象介入手段と類似した介入が行われた患者aの介入手段データ101a、患者背景情報102Aa、介入効果情報102Ba、を取得する。

【0022】

患者zに対する予測対象介入手段データ111zの効果や妥当性を分析するため、分析装置が抽出した類似した介入を行った事例を分析する。本例では、介入効果情報102Baに基づき介入効果に関する統計情報や推定値を提供したり、患者背景情報102Aaに基づき患者背景に関する統計情報や推定値を提供する。

30

【0023】

なお、上述した例では、複数の異なる種類の医療サービス(医薬品処方および生活指導)の適否を示すデータ列として、介入手段情報101の介入手段データ101a～101cおよび予測対象介入手段情報111の予測対象介入手段データ111zを例示した。複数の異なる種類の医療サービスは、医薬品処方や生活指導に限らず、たとえば、治療や手術もあるため、医薬品処方、生活指導、治療および手術のうち2種類以上の組み合わせでもよい。また、これら以外の医療サービスの2種類以上の組み合わせでもよい。

【0024】

また、介入手段情報101の介入手段データ101a～101cおよび予測対象介入手段情報111の予測対象介入手段データ111zは、1種類の医療サービスにおける複数のサービス属性の適否を示すデータ列としてもよい。

40

【0025】

また、上述した介入手段情報101および予測対象介入手段情報111では、サービス属性：医薬品処方として、「医薬品A」、「医薬品B」、...の処方の有無としたが、「医薬品A 医薬品B」のように、過去に「医薬品A」を処方していたが「医薬品B」に変更したというサービス属性が含まれていてもよい。これにより、介入手段情報101および予測対象介入手段情報111を詳細に規定することができる。

【0026】

50

#### < 分析装置のハードウェア構成例 >

図 2 は、分析装置のハードウェア構成例を示すブロック図である。分析装置 200 は、プロセッサ 201 と、記憶デバイス 202 と、入力デバイス 203 と、出力デバイス 204 と、通信インターフェース（通信 I/F）205 と、を有する。プロセッサ 201、記憶デバイス 202、入力デバイス 203、出力デバイス 204、および通信 I/F 205 は、バス 206 により接続される。プロセッサ 201 は、分析装置 200 を制御する。記憶デバイス 202 は、プロセッサ 201 の作業エリアとなる。また、記憶デバイス 202 は、各種プログラムやデータを記憶する非一時的なまたは一時的な記録媒体である。記憶デバイス 202 としては、たとえば、ROM（Read Only Memory）、RAM（Random Access Memory）、HDD（Hard Disk Drive）、フラッシュメモリがある。入力デバイス 203 は、データを入力する。入力デバイス 203 としては、たとえば、キーボード、マウス、タッチパネル、テンキー、スキャナがある。出力デバイス 204 は、データを出力する。出力デバイス 204 としては、たとえば、ディスプレイ、プリンタ、スピーカがある。通信 I/F 205 は、ネットワークと接続し、データを送受信する。

#### 【0027】

##### < 未整形ヘルスケア情報 >

図 3 は、未整形ヘルスケア情報の一例を示す説明図である。未整形ヘルスケア情報 300 は、記憶デバイス 202 に格納されている。また、分析装置 200 は、通信 I/F 205 を介して通信可能な他のコンピュータに格納されている未整形ヘルスケア情報 300 を取得してもよい。

#### 【0028】

未整形ヘルスケア情報 300 は、基本情報 301 と、検査情報 302 と、医薬品情報 303 と、処置情報 304 と、関連サービス情報 305 と、を含む。基本情報 301 とは、個人 ID、生年月日、性別など、患者の基本的な情報である。個人 ID は、患者を一意に特定する識別情報である。

#### 【0029】

検査情報 302 とは、個人 ID、個人 ID で特定される患者に対する検査の実施年月日、検査内容を示す実施項目、検査結果など、検査に関する情報である。医薬品情報 303 とは、個人 ID、個人 ID で特定される患者に医薬品を使用した実施年月日、当該使用した医薬品を示す実施項目など、医薬品に関する情報である。処置情報 304 とは、個人 ID、個人 ID で特定される患者に対する処置の実施年月日、処置の内容を示す実施項目など、処置に関する情報である。関連サービス情報 305 とは、個人 ID、個人 ID で特定される患者に対する関連サービスの実施年月日、関連サービスの内容を示す実施項目など、関連サービスに関する情報である。

#### 【0030】

基本情報 301、検査情報 302、医薬品情報 303、処置情報 304、および関連サービス情報 305 において、同一個人 ID でかつ同一実施年月日のレコードを、未整形ヘルスケアデータと称す。個人 ID および実施年月日の両方が同一である必要があるため、同一個人 ID でも実施年月日が異なれば、それらは異なる未整形ヘルスケアデータとなる。

#### 【0031】

##### < 学習モデル生成処理 >

図 4 は、実施例 1 にかかる分析装置 200 による学習モデル生成処理手順例を示すフローチャートである。分析装置 200 は、記憶デバイス 202 または通信可能な他のコンピュータから未整形ヘルスケア情報 300 を取得する（ステップ S401）。つぎに、分析装置 200 は、未整形ヘルスケア情報 300 をデータ整形し、整形ヘルスケア情報を生成する（ステップ S402）。

#### 【0032】

図 5 は、整形ヘルスケア情報の一例を示す説明図である。整形ヘルスケア情報 500 は、レコード ID 501 と、個人 ID 502 と、介入日 503 と、患者背景情報 102A と

10

20

30

40

50



、介入手段情報 1 0 1 と、介入効果情報 1 0 2 B と、を有する。整形ヘルスケア情報 5 0 0 の各レコードが整形ヘルスケアデータである。

【 0 0 3 3 】

レコード I D 5 0 1 は、整形ヘルスケアデータを一意に特定する識別情報である。個人 I D 5 0 2 は、未整形ヘルスケア情報 3 0 0 で規定された個人 I D である。介入日 5 0 3 は、個人 I D 5 0 2 で特定される患者に介入した年月日である。介入日 5 0 3 は、未整形ヘルスケア情報 3 0 0 を構成する基本情報 3 0 1、検査情報 3 0 2、医薬品情報 3 0 3、処置情報 3 0 4、および関連サービス情報 3 0 5 のいずれかの実施年月日である。

【 0 0 3 4 】

患者背景情報 1 0 2 A は、たとえば、性別、年齢、体重、空腹時血糖値のような個人情報を含む。患者背景情報 1 0 2 A は、たとえば、基本情報 3 0 1 から整形される。患者背景情報 1 0 2 A の各レコードを、患者背景データと称す。

10

【 0 0 3 5 】

介入手段情報 1 0 1 は、たとえば、医薬品 A、医薬品 B、処置 X、介護サービス a のような各種介入手段を含む。介入手段情報 1 0 1 は、たとえば、医薬品情報 3 0 3、処置情報 3 0 4、および関連サービス情報 3 0 5 から整形される。介入手段情報 1 0 1 の各レコードを、介入手段データと称す。

【 0 0 3 6 】

介入効果情報 1 0 2 B は、たとえば、アウトカム取得日、空腹時血糖値、医療費のような各種介入効果を含む。介入効果情報 1 0 2 B は、たとえば、検査情報 3 0 2 から整形される。介入効果情報 1 0 2 B の各レコードを、介入効果データと称す。

20

【 0 0 3 7 】

図 4 に戻り、分析装置 2 0 0 は、未選択の整形ヘルスケアデータがあるか否かを判断する（ステップ S 4 0 3）。未選択レコード（整形ヘルスケアデータ）がある場合（ステップ S 4 0 3：Yes）、分析装置 2 0 0 は、未選択レコードを選択し（ステップ S 4 0 4）、選択レコードから介入手段、背景データ、および介入効果データを抽出する（ステップ S 4 0 5～S 4 0 6）。

【 0 0 3 8 】

分析装置 2 0 0 は、抽出した介入手段データを訓練データとし、抽出した背景データおよび介入効果データを正解データとして、ニューラルネットワーク 1 0 3 に与え、ニューラルネットワーク 1 0 3 の学習パラメータを更新し（ステップ S 4 0 8）、ステップ S 4 0 3 に戻る。未選択レコードがない場合（ステップ S 4 0 3：No）、分析装置 2 0 0 は、ニューラルネットワーク 1 0 3 は、最新の学習パラメータが設定された学習モデル 1 0 3 となる。これにより、学習モデル生成処理が終了する。

30

【 0 0 3 9 】

<ニューラルネットワーク>

図 6 は、ニューラルネットワーク 1 0 3 の一例を示す説明図である。ニューラルネットワーク 1 0 3 は、入力層 1 3 1、1 つの中間層 1 3 2、および出力層 1 3 3（1 3 3 A、1 3 3 B）により構成される。入力層 1 3 1 には、介入手段データが入力される。介入手段データを、n 次元のベクトル x とする。

40

【 0 0 4 0 】

中間層 1 3 2 は、学習パラメータとして、重みパラメータ W 1 とバイアス b 1 とを有し、下記式（1）の演算を実行する。式（1）の演算結果が、図 1 に示したような特徴量データとなる。

【 0 0 4 1 】

$$x_2 A = W_1 \cdot x + b_1 \cdots (1)$$

【 0 0 4 2 】

W 1 は、下記式（2）の m × n 行列で表現される。ただし、式（2）の  $v_{1i}$ （1 ≤ i ≤ n）は、m 次元列ベクトルである。また、バイアス b 1 も、m 次元列ベクトルである。

【 0 0 4 3 】

50

$$W1 = (v11, v12, v13, \dots, v1n) \dots (2)$$

【0044】

第1出力層133Aには、式(1)の実行結果( $x2A = W1 \cdot x + b1$ )が入力される。第1出力層133Aは、学習パラメータとして、重みパラメータW2とバイアスb2とを有し、下記式(3)の演算を実行する。

【0045】

$$y1 = W2 \cdot x2A + b2 \dots (3)$$

【0046】

W2は、下記式(4)の $1 \times n$ 行列で表現される。ただし、式(4)の $v2i(1 \leq i \leq 1)$ は、1次元列ベクトルである。また、バイアスb2も、1次元列ベクトルである。

10

【0047】

$$W2 = (v21, v22, v23, \dots, v21) \dots (4)$$

【0048】

第2出力層133Bには、式(1)の実行結果( $x2A = W1 \cdot x + b1$ )が入力される。第2出力層133Bは、学習パラメータとして、重みパラメータW3とバイアスb3とを有し、下記式(5)の演算を実行する。

【0049】

$$y2 = W3 \cdot x2A + b3 \dots (5)$$

【0050】

W3は、下記式(6)の $k \times n$ 行列で表現される。ただし、式(6)の $v3i(1 \leq i \leq k)$ は、k次元列ベクトルである。また、バイアスb3も、k次元列ベクトルである。

20

【0051】

$$W3 = (v31, v32, v33, \dots, v3k) \dots (6)$$

【0052】

図7は、ニューラルネットワーク103の他の例を示す説明図である。図6のニューラルネットワーク103の中間層は1層であったのに対し、図7のニューラルネットワークの中間層はm層有する。図1の特定の間層132と同様、m層の中間層のうち特定の間層132が特徴量情報104を生成する。

【0053】

<特徴量情報生成処理>

30

図8は、特徴量情報生成処理手順例を示すフローチャートである。分析装置200は、整形ヘルスケア情報500に未選択の介入手段データがあるか否かを判断する(ステップS801)。未選択の介入手段データがある場合(ステップS801: Yes)、分析装置200は、整形ヘルスケア情報500から未選択の介入手段データを取得する(ステップS802)。

【0054】

分析装置200は、取得した介入手段データを、学習モデル103に入力する(ステップS803)。分析装置200は、学習モデル103の特定の間層132により特徴量データを算出して記憶デバイス202に格納し、ステップS801に戻る(ステップS804)。この算出された特徴量データを、便宜的に第1特徴量データと称す。

40

【0055】

ステップS802で取得された介入手段データが図1に示した介入手段データ101aであれば、第1特徴量データとして特徴量データ104aが算出される。ステップS801において、未選択の介入手段データがない場合(ステップS801: No)、特徴量情報生成処理は終了する。このようにして、図1の(2)に示したような特徴量情報104が生成される。

【0056】

<予測対象未整形ヘルスケア情報>

図9は、実施例1にかかる予測対象未整形ヘルスケア情報の一例を示す説明図である。予測対象未整形ヘルスケア情報900は、記憶デバイス202に格納されている。また、

50

分析装置 200 は、通信 IF 205 を介して通信可能な他のコンピュータに格納されている予測対象未整形ヘルスケア情報 900 を取得してもよい。

【0057】

予測対象未整形ヘルスケア情報 900 は、医薬品情報 903 と、処置情報 904 と、関連サービス情報 905 と、を含む。医薬品情報 903、処置情報 904、および関連サービス情報 905 は、図 3 に示した医薬品情報 303、処置情報 304、および関連サービス情報 305 と同一項目を有する。また、検査は実施されていないため、予測対象未整形ヘルスケア情報 900 には検査情報は含まれない。未整形ヘルスケア情報 300 の少なくとも一部を予測対象未整形ヘルスケア情報 900 として用いる場合、基本情報 301 および検査情報 302 を除外すればよい。

10

【0058】

医薬品情報 903、処置情報 904、および関連サービス情報 905 において、同一個人 ID でかつ同一実施年月日のレコードを、予測対象未整形ヘルスケアデータと称す。個人 ID および実施年月日の両方が同一である必要があるため、同一個人 ID でも実施年月日が異なれば、それらは異なる予測対象未整形ヘルスケアデータとなる。

【0059】

図 10 は、実施例 1 にかかる予測対象整形ヘルスケア情報の一例を示す説明図である。予測対象整形ヘルスケア情報 1000 は、レコード ID 501 と、個人 ID 502 と、介入日 503 と、予測対象介入手段情報 111 と、を有する。予測対象整形ヘルスケア情報 1000 の各レコードが予測対象整形ヘルスケアデータである。

20

【0060】

< 分析処理 >

図 11 は、実施例 1 にかかる分析装置 200 による分析処理手順例を示すフローチャートである。分析装置 200 は、記憶デバイス 202 または通信可能な他のコンピュータから予測対象未整形ヘルスケア情報 900 を取得する（ステップ S1101）。つぎに、分析装置 200 は、予測対象未整形ヘルスケア情報 900 をデータ整形し、予測対象整形ヘルスケア情報 1000 を生成する（ステップ S1102）。

【0061】

分析装置 200 は、予測対象整形ヘルスケア情報 1000 から予測対象整形ヘルスケアデータを選択する（ステップ S1103）。分析装置 200 は、選択した予測対象整形ヘルスケアデータから予測対象介入手段データを抽出する（ステップ S1104）。

30

【0062】

なお、分析装置 200 は、ステップ S1101 ~ S1104 のように介入手段データを取得するのではなく、ユーザ操作により、入力デバイス 203 または通信 IF 205 を介して通信可能な他のコンピュータから、たとえば、「医薬品 A and 医薬品 X」のように、介入手段データの入力を受け付けてもよい。

【0063】

分析装置 200 は、抽出した予測対象介入手段データを学習モデル 103 に入力する（ステップ S1105）。分析装置 200 は、学習モデル 103 の特定の中間層 132 により特徴量データを算出する（ステップ S1106）。この特徴量データを、ステップ S804 の第 1 特徴量データと区別するため、便宜的に、第 2 特徴量データと称す。ステップ S1104 で取得された予測対象介入手段データが図 1 に示した予測対象介入手段データ 111z であれば、第 2 特徴量データとして特徴量データ 114z が算出される。

40

【0064】

分析装置 200 は、第 2 特徴量データに類似する特定の第 1 特徴量データを検索する（ステップ S1107）。具体的には、たとえば、分析装置 200 は、第 1 特徴量データの各々と第 2 特徴量データとの類似度を算出する。類似度が類似度しきい値以上であれば、その第 1 特徴量データは、第 2 特徴量データに類似する特定の第 1 特徴量データとなる。

【0065】

類似度は、たとえば、特徴量空間内における第 1 特徴量データと第 2 特徴量データとの

50

距離である。算出距離の逆数が類似度しきい値以上であれば、その第 1 特徴量データは、第 2 特徴量データに類似する特定の第 1 特徴量データとなる。

【 0 0 6 6 】

なお、類似度しきい値は、あらかじめ分析装置 2 0 0 に設定された値でもよく、ユーザ操作により、入力デバイス 2 0 3 または通信 I F 2 0 5 を介して通信可能な他のコンピュータから受け付けた値でもよい。

【 0 0 6 7 】

分析装置 2 0 0 は、特定の第 1 特徴量データに対応する特定のヘルスケアデータを抽出する（ステップ S 1 1 0 8 ）。具体的には、たとえば、分析装置 2 0 0 は、特定の第 1 特徴量データの算出元となる介入手段データを含む整形ヘルスケアデータを、整形ヘルスケア情報 5 0 0 から抽出する。図 1 の（ 5 ）の例では、特定の第 1 特徴量データが特徴量データ 1 0 4 a だとすると、特定のヘルスケアデータとして、介入手段情報 1 0 1 から介入手段データ 1 0 1 a が抽出される。

10

【 0 0 6 8 】

このあと、分析装置 2 0 0 は、統計処理を実行し（ステップ S 1 1 1 0 ）、統計処理（ステップ S 1 1 1 0 ）の結果を出力する（ステップ S 1 1 1 1 ）。統計処理（ステップ S 1 1 1 0 ）の詳細については後述する。

【 0 0 6 9 】

図 1 2 は、統計処理（ステップ S 1 1 1 0 ）の結果例を示す説明図である。結果画面 1 2 0 0 は、統計処理（ステップ S 1 1 1 0 ）の結果を表示する画面である。結果画面は、入力領域 1 2 0 1 と、出力領域 1 2 0 2 と、を有する。

20

【 0 0 7 0 】

入力領域 1 2 0 1 は、編集ボタン 1 2 1 1 と、分析ボタン 1 2 1 2 と、介入手段入力欄 1 2 1 3 と、しきい値入力欄 1 2 1 4 と、を有する。編集ボタン 1 2 1 1 は、押下により介入手段入力欄 1 2 1 3 およびしきい値入力欄 1 2 1 4 への文字列の入力を可能にするボタンである。分析ボタン 1 2 1 2 は、押下により図 1 1 に示した分析処理を実行するためのボタンである。

【 0 0 7 1 】

介入手段入力欄 1 2 1 3 は、ユーザ操作により、入力デバイス 2 0 3 または通信 I F 2 0 5 を介して通信可能な他のコンピュータから、たとえば、「医薬品 A and 医薬品 X」のように、予測対象介入手段データ 1 1 1 z の入力を受け付ける入力欄である。しきい値入力欄 1 2 1 4 は、ユーザ操作により、入力デバイス 2 0 3 または通信 I F 2 0 5 を介して通信可能な他のコンピュータから、たとえば、「 0 . 8 0 」のように、類似度を示す数値の入力を受け付ける入力欄である。

30

【 0 0 7 2 】

出力領域 1 2 0 2 は、類似介入手段情報 1 2 2 1 と、類似患者背景情報 1 2 2 2 と、類似介入効果情報 1 2 2 3 と、を有する。類似介入手段情報 1 2 2 1 は、類似介入手段 1 2 3 1 ごとに、類似度 1 2 3 2 や症例数 1 2 3 3 といった情報を有する。類似介入手段 1 2 3 1 は、介入手段入力欄 1 2 1 3 に入力された介入手段データが学習モデル 1 0 3 に入力された結果、ステップ S 1 1 0 6 で特定の中間層 1 3 2 から算出された第 2 特徴量データに類似する特定の第 1 特徴量データ（ステップ S 1 1 0 7 ）の算出元となる介入手段データであり、ステップ S 1 1 0 8 で抽出される。

40

【 0 0 7 3 】

類似度 1 2 3 2 は、介入手段入力欄 1 2 1 3 に入力された予測対象介入手段データが学習モデル 1 0 3 に入力された結果、ステップ S 1 1 0 6 で特定の中間層 1 3 2 から算出された第 2 特徴量データと、類似介入手段 1 2 3 1 との類似度（たとえば、算出距離の逆数）であり、ステップ S 1 1 0 7 で算出される。類似介入手段情報 1 2 2 1 として表示される類似介入手段は、たとえば、その類似度 1 2 3 2 が類似度しきい値以上となる介入手段データである。

【 0 0 7 4 】

50

症例数 1 2 3 3 は、類似介入手段 1 2 3 1 を介入手段データとして持つ整形ヘルスケアデータの計数値であり、統計処理（ステップ S 1 1 0 9）で算出される。

【 0 0 7 5 】

類似患者背景情報 1 2 2 2 は、類似介入手段 1 2 3 1 ごとに、年齢 1 2 4 1 や体重 1 2 4 2、空腹時血糖 1 2 4 3 といった情報を有する。

【 0 0 7 6 】

年齢 1 2 4 1 は、類似介入手段から特定される 1 以上の年齢の統計値（たとえば、平均値 ± 標準偏差）である。具体的には、たとえば、分析装置 2 0 0 は、類似介入手段 1 2 3 1 となる介入手段データごとに、当該介入手段データを含む整形ヘルスケアデータ（以下、類似整形ヘルスケアデータ）を整形ヘルスケア情報 5 0 0 から特定する。分析装置 2 0 0 は、類似整形ヘルスケアデータごとに、類似整形ヘルスケアデータ内の患者背景データに含まれる年齢を抽出する。分析装置 2 0 0 は、類似整形ヘルスケアデータごとに抽出した年齢の平均値と標準偏差を算出し、類似患者背景情報 1 2 2 2 における類似介入手段 1 2 3 1 に対応する年齢 1 2 4 1 として表示する。年齢 1 2 4 1 は、統計処理（ステップ S 1 1 0 9）で算出される。

10

【 0 0 7 7 】

体重 1 2 4 2 も年齢 1 2 4 1 と同様、類似介入手段 1 2 3 1 から特定される 1 以上の体重の統計値（たとえば、平均値 ± 標準偏差）である。具体的には、たとえば、分析装置 2 0 0 は、類似整形ヘルスケアデータごとに、類似整形ヘルスケアデータ内の患者背景データに含まれる体重を抽出する。分析装置 2 0 0 は、類似整形ヘルスケアデータごとに抽出した体重の平均値と標準偏差を算出し、類似患者背景情報 1 2 2 2 における類似介入手段 1 2 3 1 に対応する体重 1 2 4 2 として表示する。体重 1 2 4 2 は、統計処理（ステップ S 1 1 0 9）で算出される。

20

【 0 0 7 8 】

空腹時血糖 1 2 4 3 も年齢 1 2 4 1 と同様、類似介入手段 1 2 3 1 から特定される 1 以上の介入前の空腹時血糖の統計値（たとえば、平均値 ± 標準偏差）である。具体的には、たとえば、分析装置 2 0 0 は、類似整形ヘルスケアデータごとに、類似整形ヘルスケアデータ内の背景データに含まれる空腹時血糖を抽出する。分析装置 2 0 0 は、類似整形ヘルスケアデータごとに抽出した空腹時血糖の平均値と標準偏差を算出し、類似患者背景情報 1 2 2 2 における類似介入手段 1 2 3 1 に対応する空腹時血糖 1 2 4 3 として表示する。空腹時血糖 1 2 4 3 は、統計処理（ステップ S 1 1 0 9）で算出される。

30

【 0 0 7 9 】

類似介入効果情報 1 2 2 3 は、類似介入手段 1 2 3 1 ごとに、空腹時血糖 1 2 5 1 や医療費 1 2 5 2 といった情報を有する。

【 0 0 8 0 】

空腹時血糖 1 2 5 1 は、類似介入手段 1 2 3 1 から特定される 1 以上の介入後の空腹時血糖の統計値（たとえば、平均値 ± 標準偏差）である。具体的には、たとえば、分析装置 2 0 0 は、類似整形ヘルスケアデータごとに、類似整形ヘルスケアデータ内の介入効果データに含まれる空腹時血糖を抽出する。分析装置 2 0 0 は、類似整形ヘルスケアデータごとに抽出した空腹時血糖の平均値と標準偏差を算出し、類似介入効果情報 1 2 2 3 における類似介入手段 1 2 3 1 に対応する空腹時血糖 1 2 5 1 として表示する。空腹時血糖 1 2 5 1 は、統計処理（ステップ S 1 1 0 9）で算出される。

40

【 0 0 8 1 】

医療費 1 2 5 2 も空腹時血糖 1 2 5 1 と同様、類似介入手段から特定される 1 以上の医療費の統計値（たとえば、平均値 ± 標準偏差）である。具体的には、たとえば、分析装置 2 0 0 は、類似整形ヘルスケアデータごとに、類似整形ヘルスケアデータ内の介入効果データに含まれる医療費を抽出する。分析装置 2 0 0 は、類似整形ヘルスケアデータごとに抽出した医療費の平均値と標準偏差を算出し、類似介入効果情報 1 2 2 3 における類似介入手段に対応する医療費 1 2 5 2 として表示する。医療費 1 2 5 2 は、統計処理（ステップ S 1 1 0 9）で算出される。

50

## 【 0 0 8 2 】

以上説明したように、実施例 1 によれば、ある患者  $z$  に提供された 1 以上の医療サービスの組み合わせである予測対象介入手段データ  $1\ 1\ 1\ z$  と一致しない介入手段データ群から、類似介入手段情報  $1\ 2\ 2\ 1$  のような類似した介入を行った事例と、類似介入効果情報  $1\ 2\ 2\ 3$  のような当該の事例の介入効果に関する統計情報と、類似背景情報  $1\ 2\ 2\ 2$  のような患者背景に関する統計情報（たとえば、平均値や標準偏差）と、を提供することができる。

## 【 実施例 2 】

## 【 0 0 8 3 】

つぎに、実施例 2 について説明する。実施例 1 では、類似した介入を行った事例の情報を提供するため、分析装置 200 は、類似した介入を行った事例と、当該の事例の介入効果に関する統計情報と、患者背景に関する統計情報（たとえば、平均値や標準偏差）と、を提供した。実施例 2 では、似たような介入手段データには、似たような介入効果があるため、分析装置 200 は、そのような予測をすることができる予測モデルを生成する。

## 【 0 0 8 4 】

そして、分析装置 200 は、この予測モデルに患者  $z$  の予測対象介入手段データ  $1\ 1\ 1\ z$  と患者  $z$  の予測対象患者背景データ  $1\ 1\ 1\ A\ z$  を入力することにより、患者  $z$  の介入効果データを予測することができる。なお、特に指摘しない限り、実施例 2 では実施例 1 で説明した図 1 ~ 図 10 の内容が適用される。また、実施例 1 と同一構成には同一符号を付し、その説明を省略する。

## 【 0 0 8 5 】

## &lt; クラスタ &gt;

図 13 は、実施例 2 にかかるクラスタの一例を示す説明図である。特徴量空間  $1\ 3\ 0\ 0$  の特徴量データ  $1\ 0\ 4\ a \sim 1\ 0\ 4\ m$  は、介入手段データ  $1\ 0\ 1\ a \sim 1\ 0\ 1\ m$  が学習モデル  $1\ 0\ 3$  に入力された結果、特定の中間層  $1\ 3\ 2$  から算出される。クラスタ  $C\ 1 \sim C\ 5$  は、類似する特徴量データを含む。

## 【 0 0 8 6 】

## &lt; 予測モデル &gt;

図 14 は、図 13 に示したクラスタを用いた予測モデルの生成例を示す説明図である。予測モデル  $M\ 1 \sim M\ 5$  の生成は、学習モデル  $1\ 0\ 3$  の生成と同じように実行される。たとえば、クラスタ  $C\ 5$  の場合、分析装置 200 は、クラスタ  $C\ 5$  に所属する特徴量データ  $1\ 0\ 4\ h \sim 1\ 0\ 4\ k$  の算出元となる介入手段データ  $1\ 0\ 1\ h \sim 1\ 0\ 1\ k$ 、患者背景データ  $1\ 0\ 2\ A\ h \sim 1\ 0\ 2\ A\ k$  をニューラルネットワークの入力層に与え、対応する介入効果データ  $1\ 0\ 2\ B\ h \sim 1\ 0\ 2\ B\ k$  を出力層に与えて学習し、学習パラメータ（重みパラメータおよびバイアス）を取得する。

## 【 0 0 8 7 】

ニューラルネットワークにこの学習パラメータが設定すると予測モデル  $M\ 5$  となる。予測モデル  $M\ 1 \sim M\ 4$  についても、クラスタ  $C\ 1 \sim C\ 4$  を用いて、同様に生成される。なお、予測モデル  $M\ 1 \sim M\ 5$  は、その他の予測分析モデル、たとえば、線形回帰モデルでもよい。

## 【 0 0 8 8 】

## &lt; 予測モデル生成 &gt;

図 15 は、実施例 2 にかかる分析装置 200 による予測モデル生成処理手順例を示すフローチャートである。分析装置 200 は、各介入手段データを整形ヘルスケア情報  $5\ 0\ 0$  から抽出し、学習モデル  $1\ 0\ 3$  に入力することで特定の中間層  $1\ 3\ 2$  から算出された特徴量データを取得する（ステップ  $S\ 1\ 5\ 0\ 1$ ）。これにより、図 13 に示したような特徴量データ  $1\ 0\ 4\ a \sim 1\ 0\ 4\ m$  が得られる。

## 【 0 0 8 9 】

分析装置 200 は、ステップ  $S\ 1\ 5\ 0\ 1$  で取得された特徴量データ群についてクラスタリングを実行する（ステップ  $S\ 1\ 5\ 0\ 2$ ）。たとえば、階層的クラスタリングを実行する

10

20

30

40

50

場合、分析装置 200 は、(1) 個々の特徴量データを 1 つのクラスタとして設定し、(2) クラスタ間の類似度を計算して、最も類似するクラスタを併合し、(3) クラスタ数が所定数に収束するまで、(2) を実行する。(2) のクラスタ間の類似度は、たとえば、最近隣法、最遠隣法、または重心法により算出されたクラスタ間距離の逆数である。これにより、図 13 に示したようなクラスタ C1 ~ C5 が生成される。また、分析装置 200 は、k 平均法などの非階層的クラスタリングを実行して、クラスタ C1 ~ C5 を生成してもよい。

#### 【0090】

分析装置 200 は、クラスタごとにクラスタに所属する特徴量データの算出元となる介入手段データを取得する(ステップ S1503)。たとえば、クラスタ C5 の場合、分析装置 200 は、クラスタ C5 に所属する特徴量データ 104h ~ 104k の算出元となる介入手段データ 101h ~ 101k を取得する。

10

#### 【0091】

そして、分析装置 200 は、クラスタ群のうち未選択クラスタがあるか否かを判断する(ステップ S1504)。未選択クラスタがある場合(ステップ S1504: Yes)、分析装置 200 は、未選択クラスタを選択する(ステップ S1505)。

#### 【0092】

分析装置 200 は、選択クラスタの介入手段データごとに、対応する患者背景データおよび介入効果データを整形ヘルスケアデータから抽出する(ステップ S1506)。たとえば、クラスタ C5 の場合、分析装置 200 は、介入手段データ 101h ~ 101k に対応する患者背景データ 102Ah ~ 102Ak, 介入効果データ 102Bh ~ 102Bk を患者 h ~ k のヘルスケアデータから抽出する。

20

#### 【0093】

分析装置 200 は、選択クラスタについて予測モデルを生成し(ステップ S1507)、ステップ S1504 に戻る。たとえば、クラスタ C5 の場合、分析装置 200 は、クラスタ C5 に所属する特徴量データ 104h ~ 104k の算出元となる介入手段データ 101h ~ 101k をニューラルネットワークの入力層に与え、介入手段データ 101h ~ 101k に対応する患者背景データ 102Ah ~ 102Ak, 介入効果データ 102Bh ~ 102Bk をニューラルネットワークの出力層に与え、予測モデル M5 を生成する。

#### 【0094】

ステップ S1504 において、未選択クラスタがない場合(ステップ S1504: No)、予測モデル生成処理が終了する。生成された予測モデル群(たとえば、予測モデル M1 ~ M5)は、記憶デバイス 202 または通信 IF 205 を介して通信可能な他のコンピュータに格納される。

30

#### 【0095】

##### < 予測処理 >

図 16 は、実施例 2 にかかる分析装置 200 による予測処理手順例を示すフローチャートである。図 17 は、実施例 2 にかかる予測対象未整形ヘルスケア情報の一例を示す説明図である。予測対象未整形ヘルスケア情報 1700 は、記憶デバイス 202 に格納されている。また、分析装置 200 は、通信 IF 205 を介して通信可能な他のコンピュータに格納されている予測対象未整形ヘルスケア情報 1700 を取得してもよい。

40

#### 【0096】

予測対象未整形ヘルスケア情報 1700 は、基本情報 1701 と、検査情報 1702 と、医薬品情報 1703 と、処置情報 1704 と、関連サービス情報 1705 と、を含む。基本情報 1701 ~ 関連サービス情報 1705 は、図 3 に示した基本情報 301 ~ 関連サービス情報 305 と同様な情報である。

#### 【0097】

図 18 は、実施例 2 にかかる予測対象整形ヘルスケア情報の一例を示す説明図である。予測対象整形ヘルスケア情報 1800 は、図 5 に示した整形ヘルスケア情報 500 と同様、レコード ID 501 と、個人 ID 502 と、介入日 503 と、患者背景情報 102A と

50

、介入手段情報 1 0 1 と、を有する。ただし、予測対象であるため、介入効果情報 1 0 2 B は含まれない。患者背景情報 1 0 2 A には、患者 z の予測対象患者背景データ 1 1 1 A<sub>z</sub> が含まれる。介入手段情報 1 0 1 には、患者 z の予測対象介入手段データ 1 1 1 z が含まれる。

【0 0 9 8】

図 1 6 に戻り、分析装置 2 0 0 は、記憶デバイス 2 0 2 または通信可能な他のコンピュータから予測対象未整形ヘルスケア情報 1 7 0 0 を取得する（ステップ S 1 6 0 1）。つぎに、分析装置 2 0 0 は、予測対象未整形ヘルスケア情報 1 7 0 0 をデータ整形し、予測対象整形ヘルスケア情報 1 8 0 0 を生成する（ステップ S 1 6 0 2）。分析装置 2 0 0 は、予測対象整形ヘルスケア情報 1 8 0 0 から予測対象整形ヘルスケアデータを選択する（ステップ S 1 6 0 3）。分析装置 2 0 0 は、選択した予測対象整形ヘルスケアデータから予測対象介入手段データおよび予測対象患者背景データを抽出する（ステップ S 1 6 0 4）。ステップ S 1 1 0 5 とステップ S 1 1 0 6 は、図 1 1 のステップ S 1 1 0 5 とステップ S 1 1 0 6 と同じ処理であるため、説明を省略する。

10

【0 0 9 9】

なお、分析装置 2 0 0 は、ステップ S 1 6 0 1 ~ S 1 6 0 4 のように介入手段データおよび患者背景データを取得するのではなく、ユーザ操作により、入力デバイス 2 0 3 または通信 I F 2 0 5 を介して通信可能な他のコンピュータから、たとえば、「医薬品 A and 医薬品 X」や「SET 年齢 = 7 9」のように、介入手段データおよび患者背景データのを受け付け、その予測対象介入手段データおよび予測対象患者背景データを学習モデル 1 0 3 に入力して、特定の間層 1 3 2 から第 2 特徴量データを算出してもよい。

20

【0 1 0 0】

ステップ S 1 1 0 6 のあと、分析装置 2 0 0 は、第 2 特徴量データの所属クラスタを特定する（ステップ S 1 6 0 7）。第 2 特徴量データの所属クラスタとは、特徴量空間において第 2 特徴量データを包含するクラスタである。所属クラスタが存在しない場合、分析装置 2 0 0 は、その旨を出力し、予測対象整形ヘルスケアデータの再選択（または予測対象介入手段データの再入力）を促してもよく、また、第 2 特徴量データが包含するまでクラスタ総数が減少するようクラスタリングを再実行してもよい。

【0 1 0 1】

分析装置 2 0 0 は、予測モデル群の格納先からステップ S 1 6 0 7 で特定したクラスタの予測モデルを取得する（ステップ S 1 6 0 8）。分析装置 2 0 0 は、取得した予測モデルに、ステップ S 1 6 0 4 で抽出した予測対象介入手段データと予測対象患者背景データをすることにより、予測モデルから予測結果データを出力する（ステップ S 1 6 1 0）。予測結果データは、介入効果データの予測値を含む。これにより、予測処理が終了する。

30

【0 1 0 2】

このように、実施例 2 によれば、クラスタ C 1 ~ C 5 ごとに生成された予測モデル M 1 ~ M 5 から、予測対象介入手段データが該当するクラスタの予測モデルが特定される。したがって、特定した予測モデルを用いることにより、予測対象介入手段データと一致する介入手段データがなくても、予測対象介入手段データと患者背景データから構築された予測モデルに基づき、介入効果データの予測結果を得ることができる。

40

【実施例 3】

【0 1 0 3】

実施例 3 について説明する。特に指摘しない限り、実施例 3 では、実施例 1 で説明した図 1 ~ 図 1 2 の内容、および、実施例 2 で説明した図 1 3 および図 1 5 の内容が適用される。また、実施例 1 および実施例 2 と同一構成には同一符号を付し、その説明を省略する。

【0 1 0 4】

実施例 2 では、分析装置 2 0 0 は、クラスタ C 1 ~ C 5 ごとに予測モデル M 1 ~ M 5 を生成した。これに対し、実施例 3 では、分析装置 2 0 0 は、図 1 3 に示したようにクラスタ C 1 ~ C 5 を生成するが（ステップ S 1 5 0 2）、予測モデル M 1 ~ M 5 を生成（ステ

50



ップS 1 5 0 7) せずに、実施例 1 のように、特定のヘルスケアデータを抽出して(ステップS 1 1 0 8)、統計処理を実行する(ステップS 1 1 0 9)。

【0 1 0 5】

具体的には、たとえば、分析装置 2 0 0 は、第 2 特徴量データの所属先クラスタをクラスタ C 1 ~ C 5 から特定する。たとえば、分析装置 2 0 0 は、第 2 特徴量データと最も類似する(距離が近い)クラスタを、たとえば、上述した最近隣法、最遠隣法、または重心法により特定する。そして、分析装置 2 0 0 は、ステップ S 1 1 0 7 のように、特定した所属先クラスタ内の第 1 特徴量データ分群の中から第 2 特徴量データに類似する特定の第 1 特徴量データを検索する。これ以降は、分析装置 2 0 0 は、実施例 1 と同様、ステップ S 1 1 0 8 ~ S 1 1 1 0 を実行する。

10

【0 1 0 6】

このように、実施例 3 によれば、第 2 特徴量データとクラスタとの類似度により第 2 特徴量データの所属先クラスタを特定するため、個々の第 1 特徴量データとの類似度を算出する必要がない。したがって、計算処理の効率化を図ることができる。

【0 1 0 7】

なお、上述した実施例 1 ~ 実施例 3 では、医療サービスを用いた分析について説明したが、分析に用いる訓練データセットは、医療機関由来に限定されず、広く健康に関わるものでも良い。たとえば、診療報酬明細書などの医療費支払いに関するデータでもよい。さらに、医療サービスに限らず他のサービスでもよい。たとえば、スポーツ競技の支援サービスに適用してもよい。この場合、介入手段情報 1 0 1 は、選手の練習方法や栄養剤の処方となり、患者背景情報 1 0 2 A は、選手の背景情報となり、介入効果情報 1 0 2 B は、筋力や走力といった測定結果となる。

20

【0 1 0 8】

また、工作機械の修理サービスに適用してもよい。この場合、介入手段情報 1 0 1 は、修理項目となり、患者背景情報 1 0 2 A は、工作機械の背景情報(製造年月日、使用年数など)となり、介入効果情報 1 0 2 B は、工作機械の部位(たとえばアーム)の可動範囲となる。また、株式投資に適用してもよい。この場合、介入手段情報 1 0 1 は、銘柄となり、患者背景情報 1 0 2 A は、投資家の背景情報(経験年数や投資額など)となり、介入効果情報 1 0 2 B は、保有株式数や売買価格となる。

【0 1 0 9】

30

なお、本発明は前述した実施例に限定されるものではなく、添付した特許請求の範囲の趣旨内における様々な変形例及び同等の構成が含まれる。たとえば、前述した実施例は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに本発明は限定されない。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てもよい。また、ある実施例の構成に他の実施例の構成を加えてもよい。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、または置換をしてもよい。

【0 1 1 0】

また、前述した各構成、機能、処理部、処理手段等は、それらの一部又は全部を、たと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く、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。

40

【0 1 1 1】

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、ハードディスク、SSD(Solid State Drive)等の記憶装置、又は、IC(Integrated Circuit)カード、SDカード、DVD(Digital Versatile Disc)の記録媒体に格納することができる。

【0 1 1 2】

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、実装上必要な全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、ほとんど全ての構成が相互に接

50

続されていると考えてよい。

【符号の説明】

【0113】

C 1 ~ C 5 クラスタ

M 1 ~ M 5 予測モデル

1 0 1 介入手段情報

1 0 2 A 患者背景情報

1 0 2 A 介入効果情報

1 0 2 B 介入効果情報

1 0 3 学習モデル

1 0 4 特徴量情報

1 1 1 予測対象介入手段情報

1 1 1 z 予測対象介入手段データ

1 3 1 入力層

1 3 2 中間層

1 3 3 出力層

2 0 0 分析装置

2 0 1 プロセッサ

2 0 2 記憶デバイス

5 0 0 整形ヘルスケア情報

1 0 0 0 予測対象整形ヘルスケア情報

1 2 3 1 類似介入手段

1 2 3 2 類似度

10

20

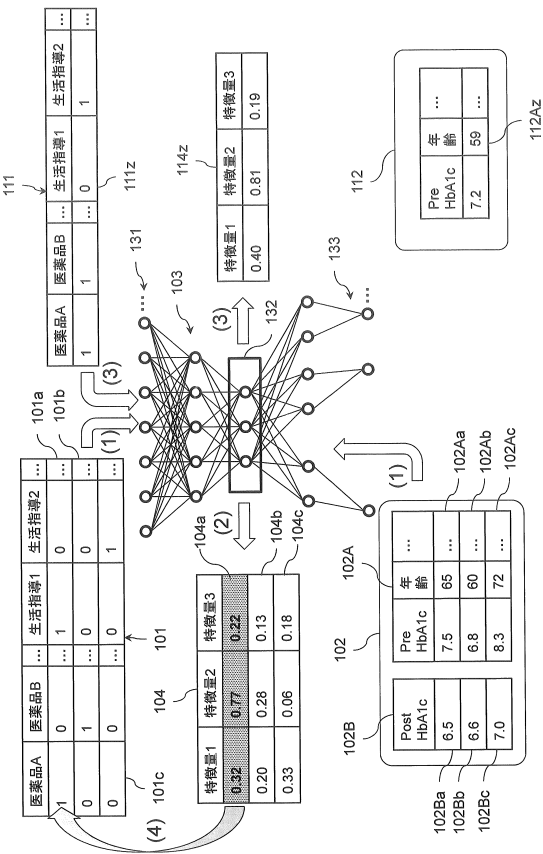
30

40

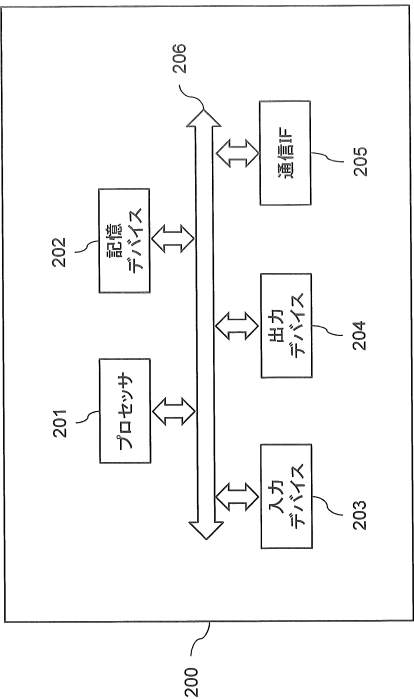
50

【図面】

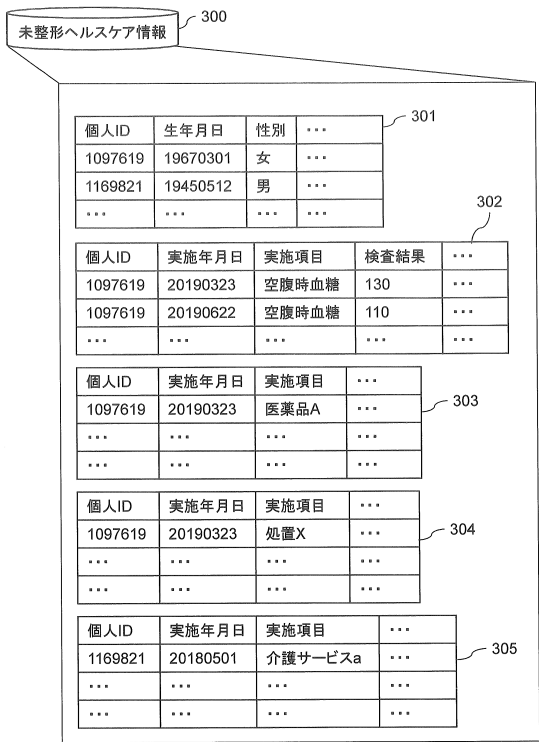
【図 1】



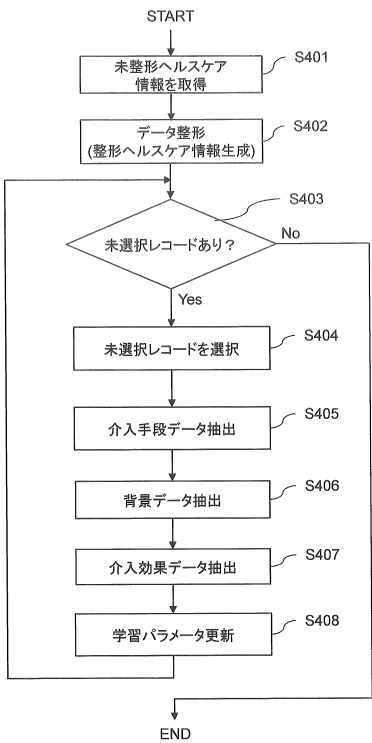
【図 2】



【図 3】



【図 4】



10

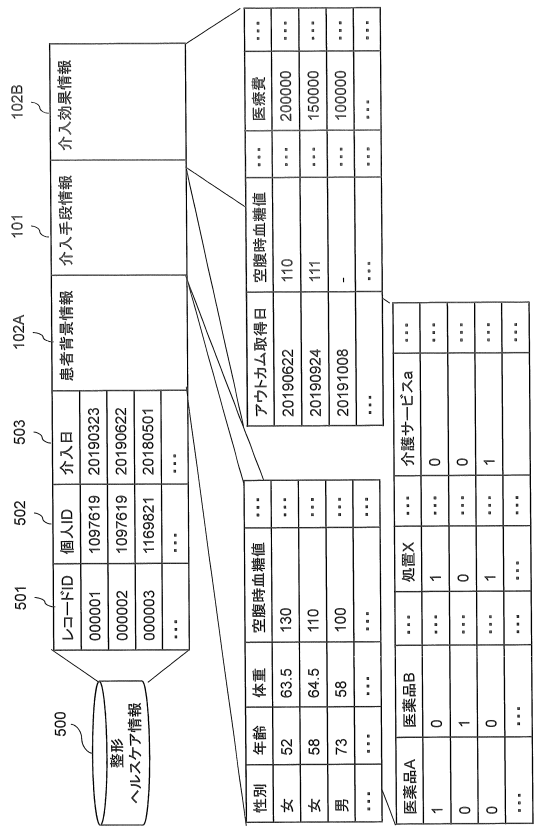
20

30

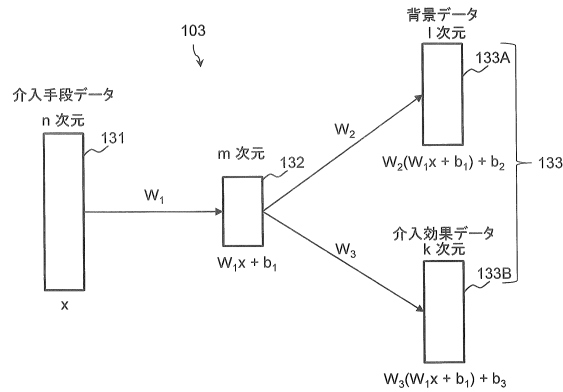
40

50

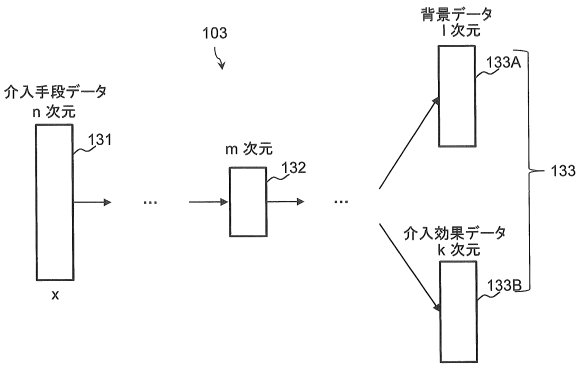
【図 5】



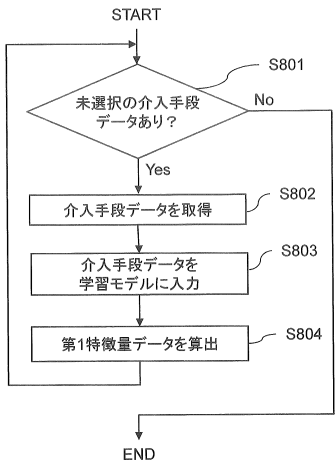
【図 6】



【図 7】



【図 8】



10

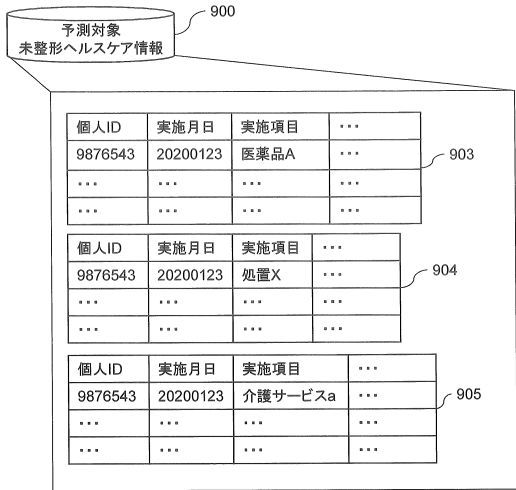
20

30

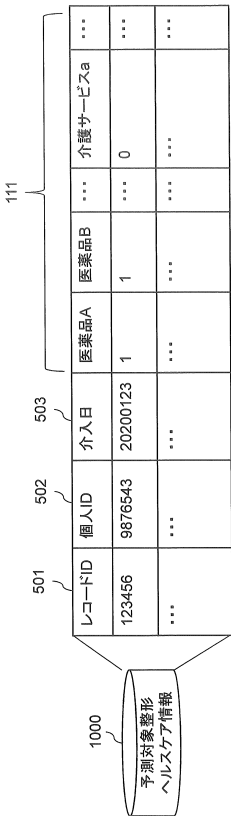
40

50

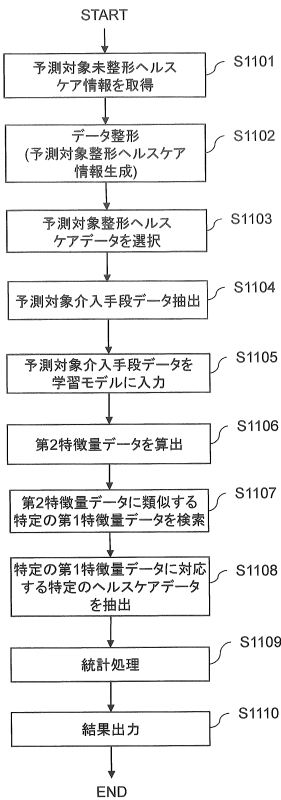
【図 9】



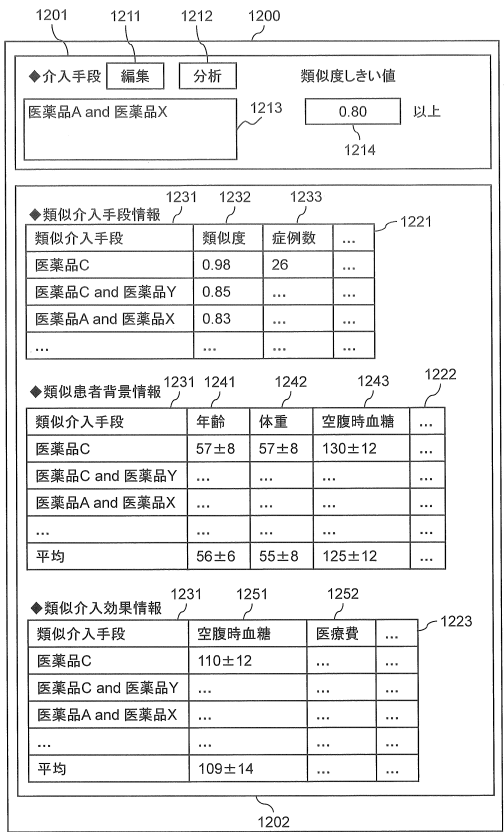
【図 10】



【図 11】



【図 12】



10

20

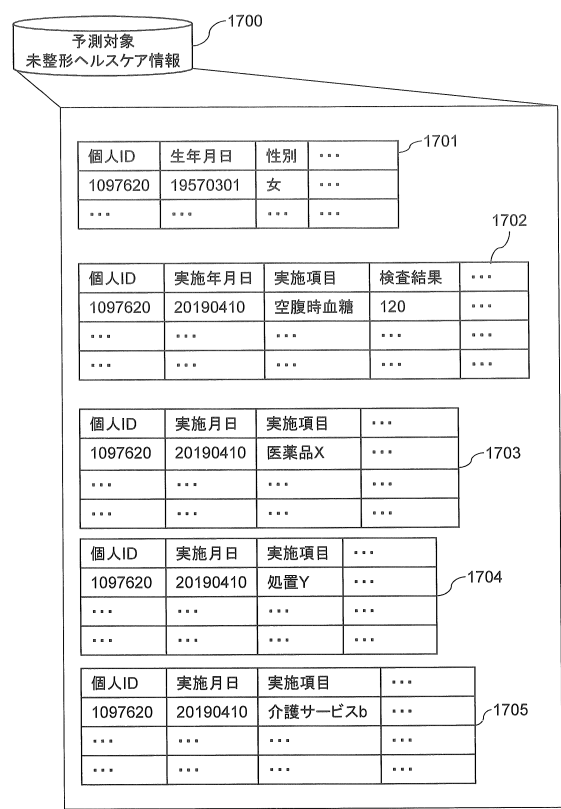
30

40

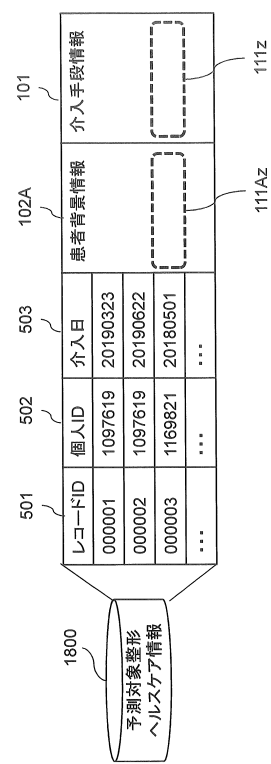
50



【図 1 7】



【図 1 8】



10

20

30

40

50

---

フロントページの続き

東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内

審査官 鹿谷 真紀

(56)参考文献 国際公開第2020/026643(WO,A1)

特開2017-054214(JP,A)

特開2019-046474(JP,A)

(58)調査した分野 (Int.Cl., DB名)

G16H 10/00-80/00

G06N 20/00-20/20