

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7063079号
(P7063079)

(45)発行日 令和4年5月9日(2022.5.9)

(24)登録日 令和4年4月25日(2022.4.25)

(51)国際特許分類		F I			
G 0 6 Q	10/06	(2012.01)	G 0 6 Q	10/06	
G 0 6 N	3/08	(2006.01)	G 0 6 N	3/08	
G 0 6 Q	10/04	(2012.01)	G 0 6 Q	10/04	
G 0 6 Q	10/10	(2012.01)	G 0 6 Q	10/10	3 2 0

請求項の数 6 (全18頁)

(21)出願番号	特願2018-81900(P2018-81900)	(73)特許権者	000005223 富士通株式会社
(22)出願日	平成30年4月20日(2018.4.20)		神奈川県川崎市中原区上小田中4丁目1 番1号
(65)公開番号	特開2019-191778(P2019-191778 A)	(74)代理人	110002147 特許業務法人酒井国際特許事務所
(43)公開日	令和1年10月31日(2019.10.31)	(72)発明者	新田 泉 神奈川県川崎市中原区上小田中4丁目1 番1号 富士通株式会社内
審査請求日	令和3年1月13日(2021.1.13)	(72)発明者	松尾 達 神奈川県川崎市中原区上小田中4丁目1 番1号 富士通株式会社内
		審査官	橘 均憲

最終頁に続く

(54)【発明の名称】 機械学習プログラム、機械学習方法および機械学習装置

(57)【特許請求の範囲】

【請求項1】

コンピュータに、

複数の従業員に関する、各レコードがカレンダーにおける期間に対応するとともに複数の項目を有する複数のレコードより構成される出勤簿データの入力を受け付け、

前記出勤簿データから、前記複数の従業員それぞれが個別に設定する個別休日に対応するレコード、および、前記複数の従業員それぞれに共通に設定される共通休日に対応するレコードを除外した除外データと、前記出勤簿データに該当する従業員が療養したか否かを示すラベルとを対応付けたデータを生成し、

前記除外データからカレンダー情報、および、前記複数の項目それぞれを別次元としてテンソル化したテンソルデータと前記ラベルとを含む訓練データを生成し、

入力されたテンソルデータをテンソル分解して得られたコアテンソルをニューラルネットワークに入力して従業員が療養するか否かを予測する予測結果を出力する学習モデルに対し前記訓練データを入力することによって、前記ニューラルネットワークの訓練および前記テンソル分解のパラメータの更新を行う、

処理を実行させる機械学習プログラム。

【請求項2】

前記データを生成する処理は、

前記出勤簿データに含まれる前記個別休日および前記共通休日のうち、該当する従業員が出勤していない休日を除外した前記除外データを生成し、前記除外データと前記ラベルと

を対応付けた前記データを生成する処理を含み、

前記訓練データを生成する処理は、

前記除外データから前記テンソルデータと前記ラベルとを含む訓練データを生成する処理を含み、

前記訓練および前記パラメータの更新を行う処理は、

前記学習モデルに対し前記訓練データを入力することによって、前記ニューラルネットワークの訓練および前記テンソル分解のパラメータの更新を行う処理を含む、

請求項 1 に記載の機械学習プログラム。

【請求項 3】

前記訓練データを生成する処理は、

前記テンソルデータの生成時に、除外した休日部分に空の要素を付加することで、前記複数の従業員それぞれの出勤簿データから生成される各テンソルデータのサイズを揃えた前記訓練データを生成する処理を含む、

請求項 1 に記載の機械学習プログラム。

【請求項 4】

前記訓練および前記パラメータの更新を行う処理は、

前記入力されたテンソルデータをテンソル分解する際に、前記空の要素に該当するテンソル部分の重みを 0 にして、前記テンソルデータからコアテンソルを生成する際の対象から除外して前記コアテンソルを生成し、

前記コアテンソルを前記ニューラルネットワークに入力して前記予測結果を出力する前記学習モデルに対し前記訓練データを入力することによって、前記ニューラルネットワークの訓練および前記テンソル分解のパラメータの更新を行う、処理を含む、

請求項 3 に記載の機械学習プログラム。

【請求項 5】

コンピュータが、

複数の従業員に関する、各レコードがカレンダーにおける期間に対応するとともに複数の項目を有する複数のレコードより構成される出勤簿データの入力を受け付け、

前記出勤簿データから、前記複数の従業員それぞれが個別に設定する個別休日に対応するレコード、および、前記複数の従業員それぞれに共通に設定される共通休日に対応するレコードを除外した除外データと、前記出勤簿データに該当する従業員が療養したか否かを示すラベルとを対応付けたデータを生成し、

前記除外データからカレンダー情報、および、前記複数の項目それぞれを別次元としてテンソル化したテンソルデータと前記ラベルとを含む訓練データを生成し、

入力されたテンソルデータをテンソル分解して得られたコアテンソルをニューラルネットワークに入力して従業員が療養するか否かを予測する予測結果を出力する学習モデルに対し前記訓練データを入力することによって、前記ニューラルネットワークの訓練および前記テンソル分解のパラメータの更新を行う、

処理をする機械学習方法。

【請求項 6】

コンピュータが、

複数の従業員に関する、各レコードがカレンダーにおける期間に対応するとともに複数の項目を有する複数のレコードより構成される出勤簿データの入力を受け付ける受付部と、前記出勤簿データから、前記複数の従業員それぞれが個別に設定する個別休日に対応するレコード、および、前記複数の従業員それぞれに共通に設定される共通休日に対応するレコードを除外した除外データと、前記出勤簿データに該当する従業員が療養したか否かを示すラベルとを対応付けたデータを生成する第 1 生成部と、

前記除外データからカレンダー情報、および、前記複数の項目それぞれを別次元としてテンソル化したテンソルデータと前記ラベルとを含む訓練データを生成する第 2 生成部と、入力されたテンソルデータをテンソル分解して得られたコアテンソルをニューラルネットワークに入力して従業員が療養するか否かを予測する予測結果を出力する学習モデルに対

10

20

30

40

50

し前記訓練データを入力することによって、前記ニューラルネットワークの訓練および前記テンソル分解のパラメータの更新を行う学習部とを有する機械学習装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、機械学習プログラム、機械学習方法および機械学習装置に関する。

【背景技術】

【0002】

従業員の出勤簿データから、数か月先のメンタル不調を予測し、カウンセリング等の対応を早期に実施することで、休職（療養）を回避することが行われている。一般的には、専任スタッフが、頻繁な出張、長時間残業、連続する急な欠勤、無断欠勤、これらの組合せなどの特徴的な特徴パターンの勤務状態に該当する従業員を目視で探している。このような特徴パターンは、各専任スタッフにより基準が異なることもあり、明確に定義することが難しい。近年では、決定木、ランダムフォレスト、SVM（Support Vector Machine）などを用いた機械学習によって、メンタル不調を特徴づける特徴パターンを学習し、専任スタッフが行っている判断を機械的に予測することが行われている。

10

【先行技術文献】

【特許文献】

【0003】

【文献】国際公開第2017/073373号

特開2010-198189号公報

特開2016-151979号公報

特開2005-332345号公報

20

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、上記技術では、間違っただ特徴パターンの学習が進んでしまい、学習精度が劣化する場合がある。例えば、土日祝日、会社が支給する夏季休暇やリフレッシュ休暇などの休日に出勤せずに休んだ日が、多くの人に共通な特徴パターンと認識されてしまって間違っただ学習が進み、体調不良への影響が大きい真の特徴パターンが見逃されて真の学習が進まない可能性がある。

30

【0005】

一つの側面では、学習精度の劣化を抑制することができる機械学習プログラム、機械学習方法および機械学習装置を提供することを目的とする。

【課題を解決するための手段】

【0006】

第1の案では、機械学習プログラムは、コンピュータに、複数の従業員に関する、各レコードがカレンダーにおける期間に対応するとともに複数の項目を有する複数のレコードより構成される出勤簿データの入力を受け付ける処理を実行させる。機械学習プログラムは、コンピュータに、前記出勤簿データから、前記複数の従業員それぞれが個別に設定する個別休日に対応するレコード、および、前記複数の従業員それぞれに共通に設定される共通休日に対応するレコードを除外した除外データを生成する処理を実行させる。機械学習プログラムは、コンピュータに、前記除外データから、カレンダー情報、および、前記複数の項目それぞれを別次元としてテンソル化した、テンソルデータを生成する処理を実行させる。機械学習プログラムは、コンピュータに、前記テンソルデータを入力テンソルデータとしてテンソル分解し、深層学習する際に、ニューラルネットワークに入力し、前記ニューラルネットワークの深層学習を行うとともに、前記テンソル分解の方法を学習する処理を実行させる。

40

【発明の効果】

50

【 0 0 0 7 】

一実施形態によれば、学習精度の劣化を抑制することができる。

【 図面の簡単な説明 】

【 0 0 0 8 】

【 図 1 】 図 1 は、実施例 1 にかかる機械学習の全体例を説明する図である。

【 図 2 】 図 2 は、グラフ構造とテンソルとの関係の一例を示す図である。

【 図 3 】 図 3 は、部分グラフ構造の抽出の一例を示す図である。

【 図 4 】 図 4 は、ディープテンソルの学習例を説明する図である。

【 図 5 】 図 5 は、実施例 1 にかかる学習装置の機能構成を示す機能ブロック図である。

【 図 6 】 図 6 は、出勤簿データ DB に記憶される出勤簿データの例を示す図である。

10

【 図 7 】 図 7 は、ラベル設定例を説明する図である。

【 図 8 】 図 8 は、テンソル化の具体例を説明する図である。

【 図 9 】 図 9 は、該当休日をテンソルの要素から除外した場合のテンソル表現を説明する図である。

【 図 1 0 】 図 1 0 は、除外した該当休日部分のテンソル表現を説明する図である。

【 図 1 1 】 図 1 1 は、実施例 1 におけるコアテンソルの生成例を説明する図である。

【 図 1 2 】 図 1 2 は、一般的な機械学習による問題点を説明する図である。

【 図 1 3 】 図 1 3 は、学習処理の流れを示すフローチャートである。

【 図 1 4 】 図 1 4 は、予測手法の別例を説明する図である。

【 図 1 5 】 図 1 5 は、ハードウェア構成例を説明する図である。

20

【 発明を実施するための形態 】

【 0 0 0 9 】

以下に、本願の開示する機械学習プログラム、機械学習方法および機械学習装置の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの発明が限定されるものではない。また、各実施例は、矛盾のない範囲内で適宜組み合わせることができる。

【 実施例 1 】

【 0 0 1 0 】

[全体例]

図 1 は、実施例 1 にかかる機械学習の全体例を説明する図である。図 1 に示すように、実施例 1 にかかる学習装置 1 0 0 は、機械学習装置の一例であり、社員の日々の出勤、退勤時間、休暇取得、出張などの状況を含む出勤簿データを機械学習して学習モデルを生成し、学習後の学習モデルを用いて、予測対象のある社員の出勤簿データから、当該社員が療養するか療養しないかを予測するコンピュータ装置の一例である。なお、ここでは、学習装置 1 0 0 が学習と予測とを実行する例で説明するが、別々の装置に実行することもできる。

30

【 0 0 1 1 】

具体的には、学習装置 1 0 0 は、療養（休職）したことがある体調不良者の出勤簿データ（ラベル＝療養あり）と、療養（休職）したことがない通常者の出勤簿データ（ラベル＝療養なし）とを教師有データとして、グラフ構造のデータを深層学習（ディープラーニング・Deep Learning ; DL）するディープテンソルによって学習モデルを生成する。その後、学習結果を適用した学習モデルを用いて、新規なグラフ構造のデータに対して正確な事象（ラベル）の推定を実現する。

40

【 0 0 1 2 】

例えば、学習装置 1 0 0 は、複数の従業員に関する、各レコードがカレンダーにおける期間に対応するとともに複数の項目を有する複数のレコードより構成される出勤簿データを取得する。そして、学習装置 1 0 0 は、出勤簿データから、複数の従業員それぞれが個別に設定した個別休日に対応するレコード、および、土日祝日などのように複数の従業員それぞれに共通に設定される共通休日を含む指定休日のうち出勤していない休日に対応するレコードを除外し、カレンダー情報、および、複数の項目を別次元としてテンソル化した、テンソルデータを生成する。その後、学習装置 1 0 0 は、テンソルデータを入力として

50

テンソル分解してニューラルネットワークに入力する学習モデルに対し、ニューラルネットワークの深層学習およびテンソル分解の方法の学習を行う。このようにして、学習装置100は、出勤簿データから複数の学習データを生成し、各学習データのテンソルデータから、「療養する」か「療養しない」かを分類する学習モデルを生成する。

【0013】

その後、学習装置100は、判別対象の従業員の出勤簿データについても同様に、指定休日のうち出勤していない休日を除外した後に、テンソル化してテンソルデータを生成し、学習済みの学習モデルに入力する。そして、学習装置100は、当該従業員が「療養する」か「療養しない」かの予測結果を示す出力値を出力する。

【0014】

すなわち、学習装置100は、休日を休んでいる日の特徴を学習データおよび予測データから除外する。ここで、休日とは、例えば土日祝日などのように各従業員に共通に設定される共通休日と、夏季休暇やリフレッシュ休暇などのように個人により所得日が異なる個別休日と、会社が支給した休暇などを含む。また、休日に出勤した場合は学習データおよび予測データから除外せずに残す。このようにすることで、学習装置100は、出勤した日の特徴に絞り込むことでノイズを除外して、学習精度の向上を図る。

【0015】

ここで、ディープテンソルについて説明する。ディープテンソルとは、テンソル(グラフ情報)を入力とするディープラーニングであり、ニューラルネットワークの学習とともに、判別に寄与する部分グラフ構造を自動的に抽出する。この抽出処理は、ニューラルネットワークの学習とともに、入力テンソルデータのテンソル分解のパラメータを学習することによって実現される。

【0016】

次に、図2および図3を用いてグラフ構造について説明する。図2は、グラフ構造とテンソルとの関係の一例を示す図である。図2に示すグラフ20は、4つのノードがノード間の関係性(例えば「相関係数が所定値以上」)を示すエッジで結ばれている。なお、エッジで結ばれていないノード間は、当該関係性がないことを示す。グラフ20を2階のテンソル、つまり行列で表す場合、例えば、ノードの左側の番号に基づく行列表現は「行列A」で表され、ノードの右側の番号(囲み線で囲んだ数字)に基づく行列表現は「行列B」で表される。これらの行列の各成分は、ノード間が結ばれている(接続している)場合「1」で表され、ノード間が結ばれていない(接続していない)場合「0」で表される。以下の説明では、この様な行列を隣接行列ともいう。ここで、「行列B」は、「行列A」の2,3行目および2,3列目を同時に置換することで生成できる。ディープテンソルでは、この様な置換処理を用いることで順序の違いを無視して処理を行う。すなわち、「行列A」および「行列B」は、ディープテンソルでは順序性が無視され、同じグラフとして扱われる。なお、3階以上のテンソルについても同様の処理となる。

【0017】

図3は、部分グラフ構造の抽出の一例を示す図である。図3に示すグラフ21は、6つのノードがエッジで結ばれたものである。グラフ21は、行列(テンソル)で表すと行列22に示すように表現できる。行列22に対して、特定の行および列を入れ替える演算、特定の行および列を抽出する演算、ならびに、隣接行列における非ゼロ要素をゼロに置換する演算を組み合わせることで、部分グラフ構造を抽出できる。例えば、行列22の「ノード1,4,5」に対応する行列を抽出すると、行列23となる。次に、行列23の「ノード4,5」間の値をゼロに置換すると、行列24となる。行列24に対応する部分グラフ構造は、グラフ25となる。

【0018】

このような部分グラフ構造の抽出処理は、テンソル分解と呼ばれる数学的演算によって実現される。テンソル分解とは、入力されたn階テンソルをn階以下のテンソルの積で近似する演算である。例えば、入力されたn階テンソルを1つのn階テンソル(コアテンソルと呼ばれる。)、および、より低階のn個のテンソル(n>2の場合、通常は2階のテン

10

20

30

40

50

ソル、つまり行列が用いられる。)の積で近似する。この分解は一意ではなく、入力データが表すグラフ構造中の任意の部分グラフ構造をコアテンソルに含める事ができる。

【0019】

続いて、ディープテンソルの学習について説明する。図4は、ディープテンソルの学習例を説明する図である。図4に示すように、学習装置100は、「療養あり」などの教師ラベル(ラベルA)が付された出勤簿データからテンソルデータを生成する。そして、学習装置100は、生成したテンソルデータを入力テンソルとしてテンソル分解を行って、初回にランダムに生成されたターゲットコアテンソルに類似するようにコアテンソルを生成する。そして、学習装置100は、コアテンソルをニューラルネットワーク(NN: Neural Network)に入力して分類結果(ラベルA: 70%、ラベルB: 30%)を得る。その後、学習装置100は、分類結果(ラベルA: 70%、ラベルB: 30%)と教師ラベル(ラベルA: 100%、ラベルB: 0%)との分類誤差を算出する。

10

【0020】

ここで、学習装置100は、誤差逆伝搬法を拡張した拡張誤差伝搬法を用いて予測モデルの学習を実行する。すなわち、学習装置100は、NNが有する入力層、中間層、出力層に対して、分類誤差を下層に伝搬させる形で、分類誤差を小さくするようにNNの各種パラメータを修正する。さらに、学習装置100は、分類誤差をターゲットコアテンソルまで伝搬させ、予測に寄与するグラフの部分構造、すなわち休職者の特徴を示す特徴パターンもしくは通常者の特徴を示す特徴パターンに近づくように、ターゲットコアテンソルを修正する。このようにすることで、最適化されたターゲットコアテンソルには予測に寄与する部分パターンが抽出されるようになる。

20

【0021】

なお、予測時には、テンソル分解により、入力テンソルをコアテンソル(入力テンソルの部分パターン)に変換し、コアテンソルをニューラルネットに入力することで、予測結果を得ることができる。テンソル分解では、コアテンソルがターゲットコアテンソルに類似するように変換される。つまり、予測に寄与する部分パターンを持つコアテンソルが抽出される。

【0022】

[機能構成]

図5は、実施例1にかかる学習装置100の機能構成を示す機能ブロック図である。図5に示すように、学習装置100は、通信部101、記憶部102、制御部110を有する。

30

【0023】

通信部101は、他の装置との通信を制御する処理部であり、例えば通信インターフェースである。例えば、通信部101は、管理者の端末から、処理開始指示、出勤簿データなどを受信する。また、通信部110は、管理者の端末に対して、学習結果や予測対象のデータの予測結果などを出力する。

【0024】

記憶部102は、プログラムやデータを記憶する記憶装置の一例であり、例えばメモリやハードディスクなどである。この記憶部102は、出勤簿データDB103、テンソルDB104、学習結果DB105、予測対象DB106を記憶する。

40

【0025】

出勤簿データDB103は、ユーザ等により入力された、社員等の出勤に関する出勤簿データを記憶するデータベースである。ここで記憶される出勤簿データは、複数の従業員に関する、各レコードがカレンダーにおける期間に対応するとともに複数の項目を有する複数のレコードより構成される出勤簿データである。また、出勤簿データは、各企業で使用される出勤簿をデータ化したものであり、公知の様々な出勤管理システムなどから取得することができる。図6は、出勤簿データDB103に記憶される出勤簿データの例を示す図である。図6に示すように、出勤簿データは、各月(月度)の日単位のレコードより構成され、各レコードは出勤情報として、「出欠区分、出張有無、出勤時刻、退勤時刻」などの各項目の値を対応付けて記憶する。図6の例では、「9月1日は、出張ではなく、9

50

時に出勤し 21 時に退勤した」ことを示す。

【0026】

なお、出欠区分の項目では、例えば入社、療養（休職）、積立休暇、有給休暇、リフレッシュ休暇などの値が設定される。出張有無の項目には、出張の有無の値が設定され、出張ありまたは出張なしのいずれかに対応した値が格納される。なお、これらの値は、番号等により区別することができる。例えば、入社 = 0、療養 = 1、積立休暇 = 2、有給休暇 = 3 などのように区別することができる。なお、出勤簿データのレコードの単位は、日単位とは限らず、週単位や月単位であってもよい。また、休暇が時間単位で取ることができる場合に対応して、時間休暇 = 4、という値を設けてもよい

【0027】

ここで記憶される出勤簿データは、学習データであり、教師ラベルが付加されている。図 7 は、ラベル設定例を説明する図である。図 7 の (a) は、ラベルとして「療養あり」が付加されている、休職した体調不良者の出勤簿データである。図 7 の (b) は、ラベルとして「療養なし」が付加されている、休職しなかった通常者の出勤簿データである。例えば、6 か月の出勤簿データを 1 つの訓練データとしテンソル化し、その後の 3 か月以内に療養した療養期間がある場合に、ラベルとして「療養あり」を設定し、その後の 3 か月以内に療養期間がない場合に、ラベルとして「療養なし」を設定することができる。

【0028】

テンソル DB 104 は、各従業員の出勤簿データから生成された各テンソル（テンソルデータ）を記憶するデータベースである。このテンソル DB 104 は、各テンソルとラベルとを対応付けた訓練データを記憶する。例えば、テンソル DB 104 は、「テンソルデータ、ラベル」として「テンソルデータ 1、ラベル（療養あり）」や「テンソルデータ 2、ラベル（療養なし）」などを記憶する。

【0029】

なお、上述の学習データにおけるレコードの項目やテンソルデータのラベルの設定は一例であり、「療養あり」と「療養なし」の値およびラベルに限らず、「休職者」と「通常者」、「休職あり」と「休職なし」などのように、体調不良者の有無を区別できる様々な値およびラベルを用いることもできる。

【0030】

学習結果 DB 105 は、学習結果を記憶するデータベースである。例えば、学習結果 DB 105 は、制御部 110 による学習データの判別結果（分類結果）、機械学習やディープラーニングによって学習された、NN の各種パラメータやディープテンソルの各種パラメータなどを記憶する。

【0031】

予測対象 DB 106 は、学習された予測モデルを用いて、療養（休職）の有無を予測する対象の出勤簿データを記憶するデータベースである。例えば、予測対象 DB 106 は、予測対象の出勤簿データ、または、予測対象の出勤簿データから生成されたテンソルデータなどを記憶する。

【0032】

制御部 110 は、学習装置 100 全体の処理を司る処理部であり、例えばプロセッサなどである。この制御部 110 は、除外部 111、テンソル生成部 112、学習部 113、予測部 114 を有する。なお、除外部 111、テンソル生成部 112、学習部 113、予測部 114 は、プロセッサなどが有する電子回路やプロセッサなどが実行するプロセスの一例である。

【0033】

除外部 111 は、出勤簿データから、複数の従業員それぞれが個別に設定する個別休日に対応するレコード、および、複数の従業員それぞれに共通に設定される共通休日に対応するレコードを除外した除外データを生成する処理部である。すなわち、除外部 111 は、学習対象のデータである各出勤簿データから、土日祝日などの共通休日、会社が支給して従業員が個別に設定する夏季休暇やリフレッシュ休暇などの個別休日、会社が休日と指定

10

20

30

40

50

する休日などを含む指定休日のうち、出勤しなかった指定休日（該当休日）を除外する。具体的には、除外部 1 1 1 は、出勤簿データ DB 1 0 3 に記憶される各出勤簿データを読み出し、各出勤簿データについて、出勤しなかった指定休日（該当休日）を除外して、テンソル生成部 1 1 2 に出力する。

【 0 0 3 4 】

例えば、除外部 1 1 1 は、出勤簿データの出欠区分を参照し、土日祝日などの指定休日を特定する。続いて、除外部 1 1 1 は、特定した各指定休日について、出欠区分が「出勤」に該当しない指定休日を特定する。そして、除外部 1 1 1 は、特定した指定休日を該当休日として出勤簿データから除外（削除）し、除外後の出勤簿データをテンソル生成部 1 1 2 に出力する。

10

【 0 0 3 5 】

テンソル生成部 1 1 2 は、除外部 1 1 1 から入力された、出勤していない指定休日（該当休日）が除外された各出勤簿データをテンソル化したテンソルデータを生成する処理部である。テンソル生成部 1 1 2 は、各出勤簿データに含まれるカレンダー情報、および、各項目「月度、日付、出欠区分、出張有無、出勤時刻、退勤時刻」のそれぞれを次元としてテンソル化する。そして、テンソル生成部 1 1 2 は、ユーザ等から指定されたラベル（療養ありもしくは療養なし）、または、出勤簿データの出欠区分から特定したラベル（療養ありもしくは療養なし）を、テンソル化されたテンソル（テンソルデータ）と対応付けてテンソル DB 1 0 4 に格納する。ここで、生成されたテンソルデータを入力としてディープテンソルによる学習を実行される。なお、ディープテンソルでは、学習時に、予測に影響を与える学習データの部分パターンを識別するターゲットコアテンソルを抽出し、予測時には、抽出したターゲットコアテンソルに基づいて予測を実行する。

20

【 0 0 3 6 】

具体的には、テンソル生成部 1 1 2 は、頻繁な出張、長時間残業、連続する急な欠勤、無断欠勤、頻繁な休日出勤、これらの組合せなど療養する傾向を特徴づけると想定される各項目を各次元として、学習データからテンソルを生成する。例えば、テンソル生成部 1 1 2 は、月度、日付、出欠区分、出張有無の 4 要素を用いた 4 次元の 4 階テンソルを生成する。なお、4 か月分のデータである場合は、月度の要素数は「4」、各月の曜日数の最大値が 3 1 であることから日付の要素数は「3 1」、出欠の種類が出社・休暇・休日であることから出欠区分の要素数は「3」、出張はありとなしであることから出張有無の要素数は「2」となる。したがって、学習データから生成されるテンソルは、「 $4 \times 3 1 \times 3 \times 2$ 」のテンソルとなり、学習データの各月度、日付における出欠区分、出張有無に対応する要素の値が 1、そうでない要素の値が 0 となる。なお、テンソルの次元とする項目は、任意に選択することができ、過去の事例等から決定することもできる。

30

【 0 0 3 7 】

図 8 は、テンソル化の具体例を説明する図である。図 8 に示すように、テンソル生成部 1 1 2 が生成するテンソルは、横方向に月度、縦方向に日付、奥行き方向に出欠区分、左から出張ありのデータで途中から出張なしのデータとなる。日付は、上から 1 日目を順に示し、出欠区分は、手前から出社、休暇、休日を示す。例えば、図 8 の (a) は、月度 1 の 1 日目に本社出張した要素を示し、図 8 の (b) は、月度 1 の 2 日目に休暇を取得して出張しなかった要素を示す。

40

【 0 0 3 8 】

なお、本実施例では、上述したテンソルを簡略化して図 8 の (c) のように記載することとする。すなわち、月度、日付、出欠区分、出張有無の各要素を重ねたキューブ状で表現することとし、各月度かつ日付の出張有無を区別して表現し、各月度かつ日付の出欠区分を区別して表現することとする。

【 0 0 3 9 】

学習部 1 1 3 は、テンソルデータを入力テンソルデータとしてテンソル分解してニューラルネットワークに入力する学習モデルに対し、ニューラルネットワークの深層学習およびテンソル分解の方法の学習を行う処理部である。すなわち、学習部 1 1 3 は、各学習デー

50

タから生成された各テンソルデータおよびラベルを入力として、ディープテンソルによる学習モデルの学習を実行する。具体的には、学習部 1 1 3 は、図 4 で説明した手法と同様、入力対象のテンソルデータ（入力テンソル）からコアテンソルを抽出して NN に入力し、NN からの分類結果と入力テンソルに付与されているラベルとの誤差（分類誤差）を算出する。そして、学習部 1 1 3 は、分類誤差を用いて、NN のパラメータの学習およびターゲットコアテンソルの最適化を実行する。その後、学習部 1 1 3 は、学習が終了すると、各種パラメータを学習結果として学習結果 DB 1 0 5 に格納する。

【 0 0 4 0 】

予測部 1 1 4 は、学習結果を用いて、判別対象のデータのラベルを予測する処理部である。具体的には、予測部 1 1 4 は、学習結果 DB 1 0 5 から各種パラメータを読み出し、各種パラメータを設定したニューラルネットワーク等を含むディープテンソルを構築する。そして、予測部 1 1 4 は、予測対象 DB 1 0 6 から予測対象の出勤簿データを読み出し、除外部 1 1 1 と同様の手法で該当休日を除外した後にテンソル化し、ディープテンソルに入力する。その後、予測部 1 1 4 は、療養ありまたは療養なしを示す予測結果を出力する。そして、予測部 1 1 4 は、予測結果をディスプレイに表示したり、管理者端末に送信したりする。

10

【 0 0 4 1 】

なお、予測部 1 1 4 は、該当休日を残したままである予測対象の出勤簿データをそのまま学習モデルに入力した場合でも、該当休日を削除した予測対象の出勤簿データを学習モデルに入力した場合でも、予測することができる。また、予測対象の出勤簿データを 6 か月ごとのデータに区切って入力することもでき、予測対象の出勤簿データそのままを入力とすることもできる。

20

【 0 0 4 2 】

[具体例]

次に、図 9 から図 1 1 を用いて、出勤していない指定休日を示す該当休日の除外からテンソル化までの具体例を説明する。図 9 は、該当休日をテンソルの要素から除外した場合のテンソル表現を説明する図である。図 1 0 は、除外した該当休日部分のテンソル表現を説明する図である。図 1 1 は、実施例 1 におけるコアテンソルの生成例を説明する図である。

【 0 0 4 3 】

図 9 の (a) の例では、除外部 1 1 1 は、従業員 1 の出勤簿データから指定休日である 1 日 (日曜日)、7 日 (土曜日)、8 日 (日曜日)、1 1 日 (水曜日) から 1 3 日 (金曜日) などの休日 (例えばリフレッシュ休暇や慶事休暇)、1 4 日 (土曜日) を特定する。続いて、除外部 1 1 1 は、特定した指定休日のうち、出勤した 1 4 日 (土曜日) を除く、1、7、8、1 1 から 1 3 日を該当休日として削除した出勤簿データを生成する。その後、テンソル生成部 1 1 2 は、1、7、8、1 1 から 1 3 日が削除された出勤簿データからテンソルデータを生成する。

30

【 0 0 4 4 】

また、図 9 の (b) の例では、除外部 1 1 1 は、従業員 1 の出勤簿データから指定休日である 1 日 (日曜日)、2 日 (月曜日) から 4 日 (水曜日) などの休日 (例えばリフレッシュ休暇や慶事休暇)、7 日 (土曜日)、8 日 (日曜日)、1 4 日 (土曜日) を特定する。続いて、除外部 1 1 1 は、特定した指定休日のうち出勤した日がないので、特定されたすべての指定休日を該当休日として削除した出勤簿データを生成する。その後、テンソル生成部 1 1 2 は、指定休日 が 削除された出勤簿データからテンソルデータを生成する。

40

【 0 0 4 5 】

ここで、月度毎の日数の違い、個人選択により期間が異なる休日、休日出勤の有無等により月度ごとの日付の数が異なる場合がある。このような場合はテンソルデータのサイズ (大きさ) が異なることになる。そこで、テンソル上で値のない日付部分に相当する箇所は空の要素とすることで、テンソル表現では、月度によらず同じ日付サイズで表現する。

【 0 0 4 6 】

具体的には、図 1 0 に示すように、テンソル生成部 1 1 2 は、該当休日 が 除外された月度

50

2、月度3、月度4を含む出勤簿データをテンソル化する際、月度ごとの日付のサイズを揃えて月度1と同じサイズにするために、月度2と月度3と月度4のそれぞれにおいて、削除された該当休日に対応する値のない日付部分については空の要素を設定する。そして、テンソル生成部112は、値のない日付部分のテンソルの要素の重みを0とし、値のない日付部分をコアテンソルとして抽出しないように学習させる。ここで重みを0にすると、例えばコアテンソルを生成する際の要素行列の値を0にすることである。

【0047】

ディープテンソルでは、このように重みを0にすることで、コアテンソルを抽出する際の対象から除外することができる。図11に示すように、該当休日を除くことにより、該当休日部分がコアテンソルとして認識されてしまうケース(図11の(a))を回避し、重要な部分(図11の(b))が認識されやすくなる。つまり、学習部113は、ディープテンソルの性質を利用して、従業員1および従業員2それぞれの出勤簿データをテンソル化するとき、重みが0で値の対部分(図11の(c))をコアテンソルの対象から除外することができるので、予測に影響を与える部分パターンを識別するコアテンソルを正確に抽出することができる。この結果、予測精度の向上が図れる学習を実行することができる。

10

【0048】

なお、一般的な機械学習では、入力固定長の特徴ベクトルが前提であり、該当休日部分を単純に削除しても学習処理が正常に処理されない。このため、削除した日数分を補完することも考えられるが、補完したとしても、学習精度の向上が期待できない。図12は、一般的な機械学習による問題点を説明する図である。図12では、療養期間がある従業員Aと療養期間がない従業員Bの各出勤簿データを例にして説明する。

20

【0049】

図12に示すように、学習対象である従業員Aの出勤簿データのうち、1日から13日が削除対象の該当休日であり、予測対象である従業員Bの出勤簿データのうち、2日から4日が削除対象の該当休日であるとする。

【0050】

図12の(1)の例は、該当休日の要素を削除せずに、該当休日の要素をNULL値として0を入れるなどで補完する例である。一般的な機械学習では、従業員Aを学習済みデータ、従業員Bを予測したいデータとすると、特徴ベクトル上で同じ位置にある属性値が一致するほど、類似度が高いと判断される。したがって、この例では、囲んだ部分の要素が多いほど、従業員Bを従業員Aとの類似性が高いと判断し、従業員Bを体調不良と予測する可能性が高くなる。

30

【0051】

図12の(2)の例は、該当休日の要素を削除した後、入力データの固定長にあわせるために、各要素を左詰めにして後方に空の要素を補完した例である。この場合、従業員Aを学習済みデータ、従業員Bを予測したいデータとすると、該当休日分を補完した部分が同じ属性値を持つ要素として認識される。したがって、(1)の場合に比べると、従業員Bと従業員Aとの類似性がより高くなり、従業員Bを体調不良と判断する可能性が高くなる。

40

【0052】

このように、一般的な機械学習では、該当休日を削除した場合、入力データが固定長であることの制約のために、空の要素を補完することになる。しかし、補完した空の要素が学習データ間で類似することになり、却って学習精度の低下を招く。

【0053】

[処理の流れ]

図13は、学習処理の流れを示すフローチャートである。図13に示すように、処理開始の指示を受け付けると(S101:Yes)、除外部111は、出勤簿データDB103から出勤簿データを読み込み(S102)、出勤していない指定休日(該当休日)があるか否かを判定する(S103)。

【0054】

50

そして、出勤していない指定休日がある場合（S103：Yes）、除外部111が、出勤簿データから出勤していない指定休日を削除し（S104）、テンソル生成部112が、削除した部分に空の要素を挿入して、出勤簿データをテンソル化する（S105）。

【0055】

一方、出勤していない指定休日がない場合（S103：No）、テンソル生成部112は、出勤簿データをそのままテンソル化する（S106）。

【0056】

その後、未処理の出勤簿データがある場合（S107：Yes）、次の出勤簿データについて、S102以降を繰り返す。一方、未処理の出勤簿データがない場合（S107：No）、学習部113は、テンソルDB104に記憶される各テンソルおよびラベルを用いて、学習処理を実行する（S108）。

10

【0057】

[効果]

上述したように、学習装置100は、指定休日を休んでいる日の特徴パターンを学習データから除外して学習処理を実行することができるので、各学習データ内の共通項目である指定休日を含む特徴パターンではなく、指定休日以外の特徴パターンを抽出して学習処理を実行することができる。この結果、学習装置100は、間違っただけの特徴パターンの学習が進むことを抑制でき、学習精度の劣化を抑制することができる。

【0058】

また、学習装置100は、指定休日などの休日情報を除いた学習モデル（予測モデル）と、休日情報を含んだ学習モデルを併用することにより、予測結果の優先度をつけることができる。図14は、予測手法の別例を説明する図である。図14に示すように、学習装置100は、休日情報を含む学習データを用いた学習処理によって学習モデルAを生成するとともに、休日情報を除外した学習データを用いた学習処理によって学習モデルBを生成する。

20

【0059】

そして、学習装置100は、予測対象の出勤簿データを学習モデルAと学習モデルBのそれぞれに入力して、各学習モデルからの予測結果に応じて、従業員の療養の危険度を判定する。具体的には、学習装置100は、「両方の学習モデルが体調不良と予測する場合」を危険度高と判定し、「いずれか一方の学習モデルのみが体調不良と予測する場合」を危険度中と判定し、「どちらの学習モデルも体調不良でないと予測する場合」を危険度小と判定する。

30

【0060】

図14の例では、学習モデルAにより体調不良があると予測され、学習モデルBにより体調不良がないと予測されているので、入力された出勤簿データに対応する従業員に対して、危険度中と判定される。なお、休日情報を残したままの出勤簿データを各学習モデルへ入力することもでき、学習モデルAへは休日情報を残したままの出勤簿データを入力し、学習モデルBへは休日情報を削除した出勤簿データを入力することもできる。

【実施例2】

【0061】

さて、これまで本発明の実施例について説明したが、本発明は上述した実施例以外にも、種々の異なる形態にて実施されてよいものである。

40

【0062】

[学習]

上述した学習処理は、任意の回数だけ実行することもできる。例えば、すべての訓練データを用いて実行することもでき、所定回数だけ実行することもできる。また、分類誤差の算出方法は、最小二乗法など公知の算出手法を採用することができ、NNで用いられる一般的な算出手法を採用することもできる。なお、学習データを用いて、事象（例えば療養ありと療養なし）を分類できるように、テンソルデータをニューラルネットワークに入力してニューラルネットワークの重み等を学習することが学習モデルの一例に対応する。

50

【 0 0 6 3 】

また、6か月の出勤簿データを予測に用いるデータ例として説明したが、これに限定されるものではなく、4か月など任意に変更することができる。また、6か月の出勤簿データに対して、その後3か月以内に療養（休職）したか否かによってラベルを付与する例を説明したが、これに限定されるものではなく、2か月以内など任意に変更することができる。また、テンソルデータも四次元に限らず、四次元未満のテンソルデータを生成することもでき、五次元以上のテンソルデータを生成することもできる。また、出勤簿データに限らず、従業員等の出退勤や休暇の取得状況などがわかる勤怠データであれば、データの形式を問わず採用することができる。

【 0 0 6 4 】

また、上記例では、除外部111は、指定休日のうち出勤していない休日である該当休日を出勤簿データから除外する例を説明したが、これに限らず、指定休日をすべて除外することもできる。

【 0 0 6 5 】

[ニューラルネットワーク]

本実施例では、RNN (Recurrent Neural Networks) やCNN (Convolutional Neural Network) など様々なニューラルネットワークを用いることができる。また、学習の手法も、誤差逆伝播以外にも公知の様々な手法を採用することができる。また、ニューラルネットワークは、例えば入力層、中間層（隠れ層）、出力層から構成される多段構成であり、各層は複数のノードがエッジで結ばれる構造を有する。各層は、「活性化関数」と呼ばれる関数を持ち、エッジは「重み」を持ち、各ノードの値は、前の層のノードの値、接続エッジの重みの値（重み係数）、層が持つ活性化関数から計算される。なお、計算方法については、公知の様々な手法を採用できる。

【 0 0 6 6 】

また、ニューラルネットワークにおける学習とは、出力層が正しい値となるように、パラメータ、すなわち、重みとバイアスを修正していくことである。誤差逆伝播法においては、ニューラルネットワークに対して、出力層の値がどれだけ正しい状態（望まれている状態）から離れているかを示す「損失関数（loss function）」を定め、最急降下法等を用いて、損失関数が最小化するように、重みやバイアスの更新が行われる。

【 0 0 6 7 】

[システム]

上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。また、実施例で説明した具体例、分布、数値などは、あくまで一例であり、任意に変更することができる。

【 0 0 6 8 】

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散や統合の具体的形態は図示のものに限られない。つまり、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。さらに、各装置にて行なわれる各処理機能は、その全部または任意の一部が、CPUおよび当該CPUにて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。

【 0 0 6 9 】

[ハードウェア]

図15は、ハードウェア構成例を説明する図である。図15に示すように、学習装置100は、通信装置100a、HDD (Hard Disk Drive) 100b、メモリ100c、プロセッサ100dを有する。また、図15に示した各部分は、バス等で相互に接続される。

【 0 0 7 0 】

通信装置100aは、ネットワークインタフェースカードなどであり、他のサーバとの通

10

20

30

40

50

信を行う。HDD100bは、図5に示した機能を動作させるプログラムやDBを記憶する。

【0071】

プロセッサ100dは、図5に示した各処理部と同様の処理を実行するプログラムをHDD100b等から読み出してメモリ100cに展開することで、図5等で説明した各機能を実行するプロセスを動作させる。すなわち、このプロセスは、学習装置100が有する各処理部と同様の機能を実行する。具体的には、プロセッサ100dは、除外部111、テンソル生成部112、学習部113、予測部114等と同様の機能を有するプログラムをHDD100b等から読み出す。そして、プロセッサ100dは、除外部111、テンソル生成部112、学習部113、予測部114等と同様の処理を実行するプロセスを実行する。

10

【0072】

このように学習装置100は、プログラムを読み出して実行することで学習方法を実行する情報処理装置として動作する。また、学習装置100は、媒体読取装置によって記録媒体から上記プログラムを読み出し、読み出された上記プログラムを実行することで上記した実施例と同様の機能を実現することもできる。なお、この他の実施例でいうプログラムは、学習装置100によって実行されることに限定されるものではない。例えば、他のコンピュータまたはサーバがプログラムを実行する場合や、これらが協働してプログラムを実行するような場合にも、本発明を同様に適用することができる。

【0073】

このプログラムは、インターネットなどのネットワークを介して配布することができる。また、このプログラムは、ハードディスク、フレキシブルディスク(FD)、CD-ROM、MO(Magneto-Optical disk)、DVD(Digital Versatile Disc)などのコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行することができる。

20

【符号の説明】

【0074】

- 100 学習装置
- 101 通信部
- 102 記憶部
- 103 出勤簿データDB
- 104 テンソルDB
- 105 学習結果DB
- 106 予測対象DB
- 110 制御部
- 111 除外部
- 112 テンソル生成部
- 113 学習部
- 114 予測部

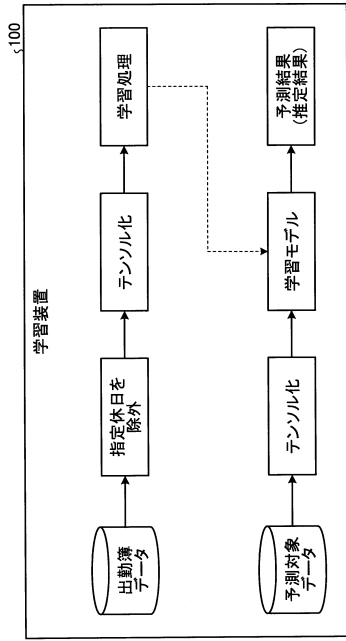
30

40

50

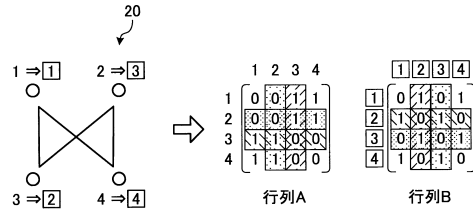
【図面】
【図 1】

実施例11にかかる機械学習の全体例を説明する図



【図 2】

グラフ構造とテンソルとの関係の一例を示す図

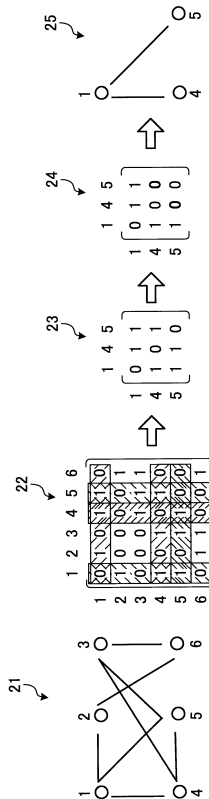


10

20

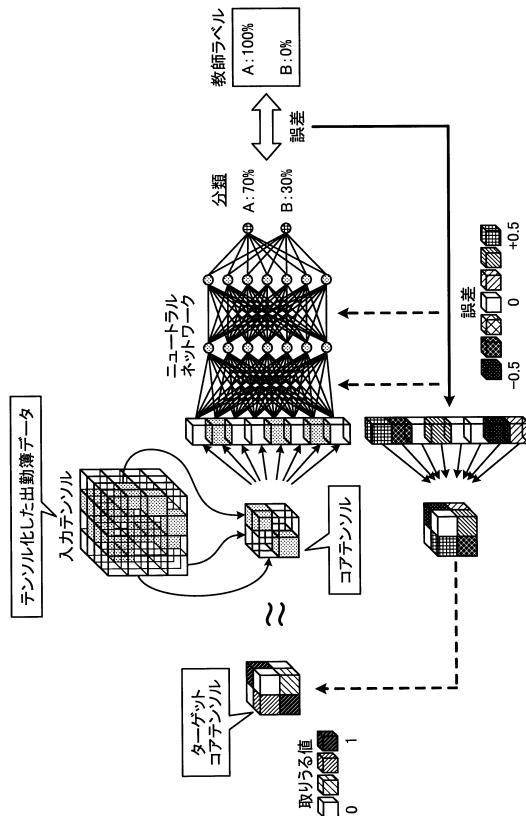
【図 3】

部分グラフ構造の抽出の一例を示す図



【図 4】

テンソルネットワークの学習例を説明する図



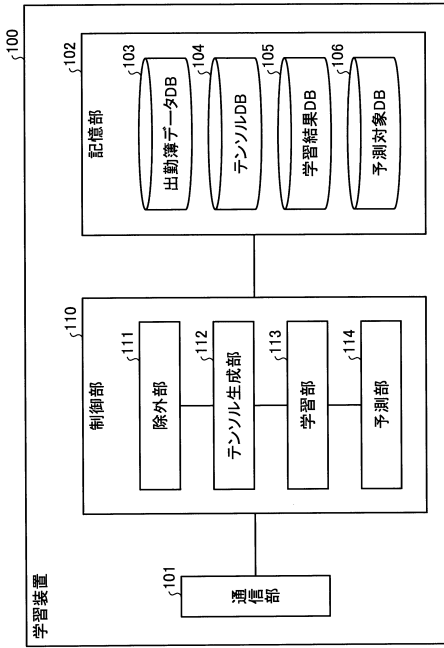
30

40

50

【図 5】

実施例11にかかる学習装置の機能構成を示す機能ブロック図



【図 6】

出勤簿データベースに記憶される出勤簿データの例を示す図

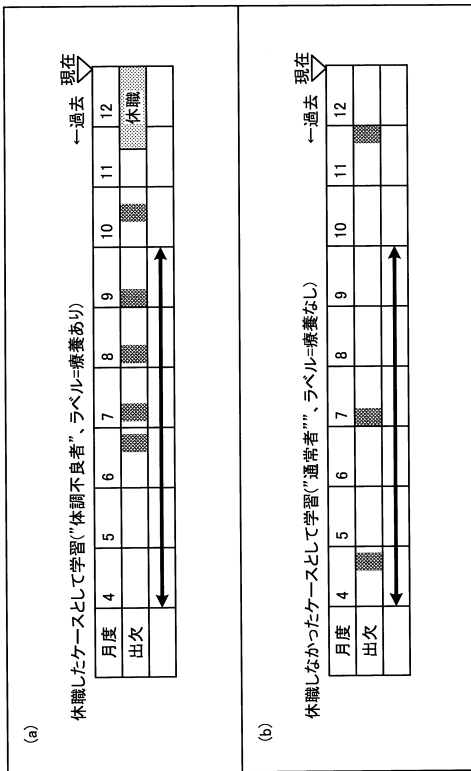
9月				
項目	1日	2日	3日	...
出欠区分	出勤	欠勤	出勤	...
出張有無	なし	なし	あり	...
出勤時刻	9:00	-	8:50	...
退勤時刻	21:00	-	19:30	...

10

20

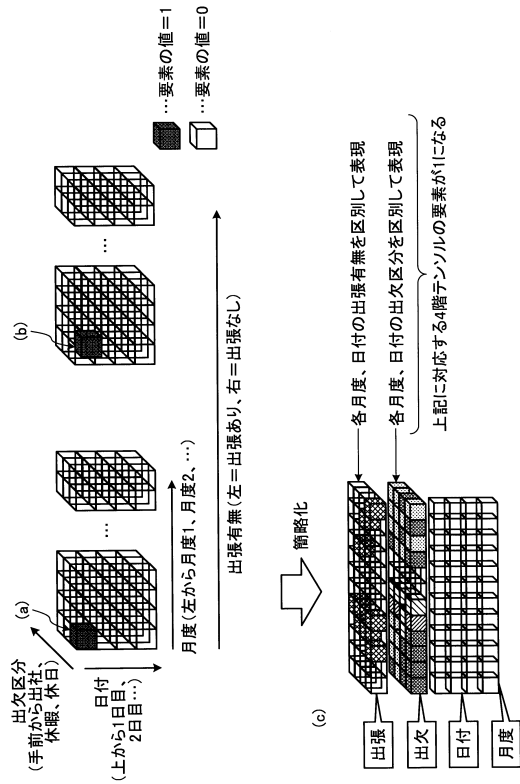
【図 7】

ラベル設定例を説明する図



【図 8】

テンソル化の具体例を説明する図



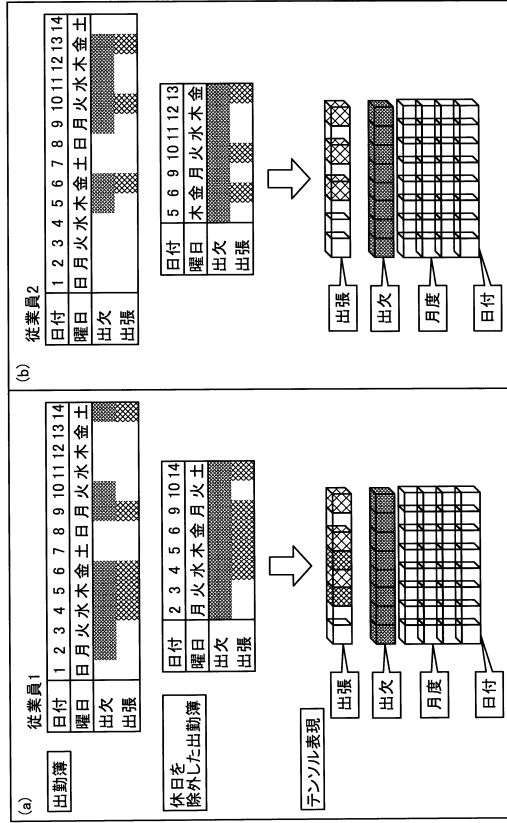
30

40

50

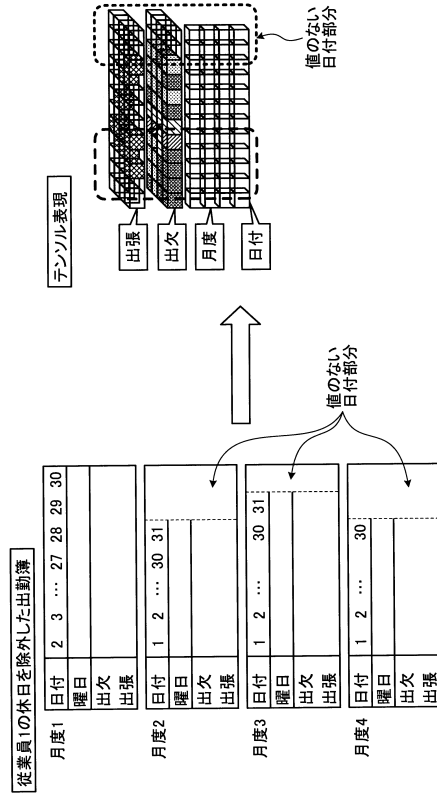
【図 9】

該当休日をテンソルの要素から除外した場合のテンソル表現を説明する図



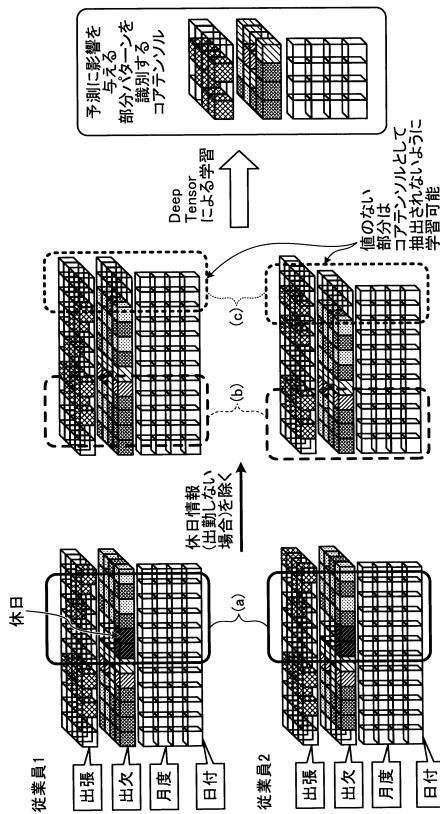
【図 10】

除外した該当休日部分のテンソル表現を説明する図



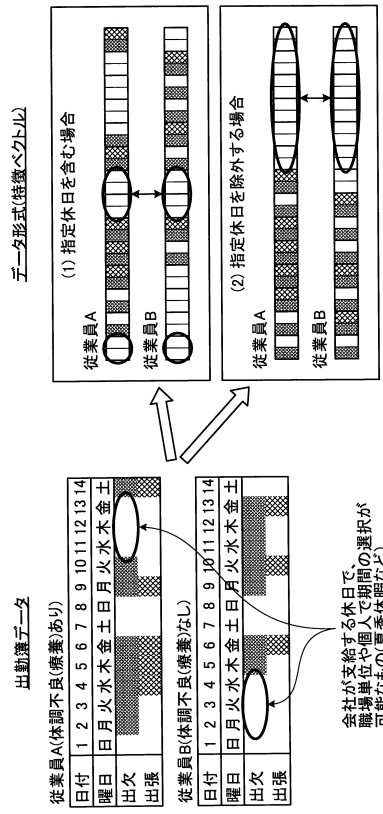
【図 11】

実施例1におけるコアテンソルの生成例を説明する図

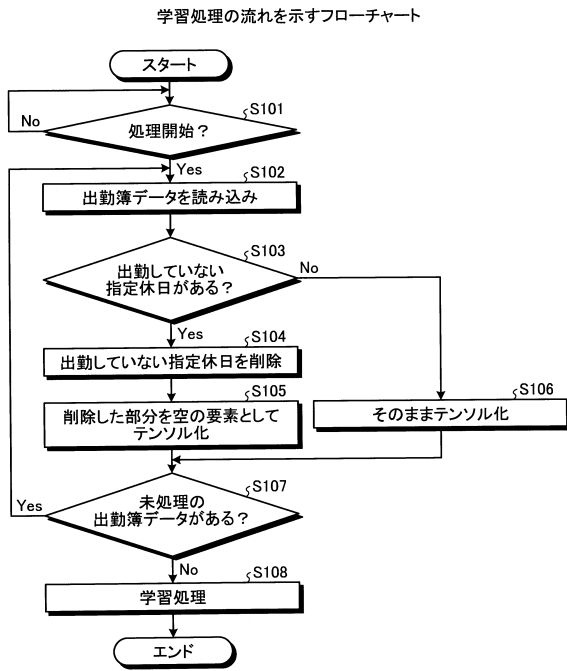


【図 12】

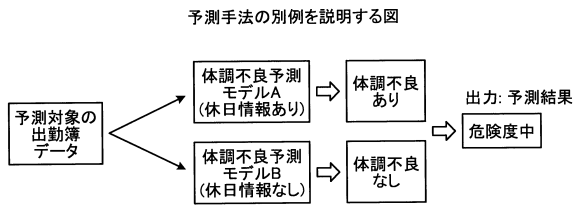
一般的な機械学習による問題点を説明する図



【 図 1 3 】



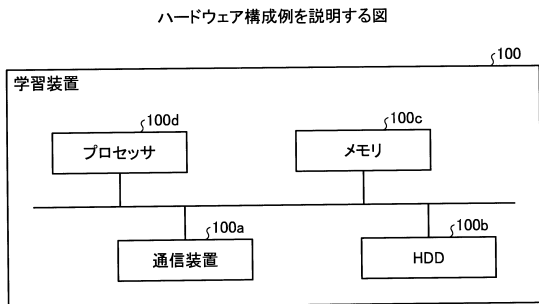
【 図 1 4 】



10

20

【 図 1 5 】



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2016/0350721 (US, A1)
国際公開第2016/035336 (WO, A1)
丸橋 弘治, 研究開発最前線, FUJITSU Vol. 68 No. 5, 富士通株式会社
第68巻
- (58)調査した分野 (Int.Cl., DB名)
G06Q 10/00 - 99/00
G06N 3/08