

## 明 細 書

**発明の名称 : 保健指導対象者選定条件作成支援装置**

### 技術分野

[0001] 本発明は、保険者の保健指導事業に関し、特に保健指導対象者を選定する条件を決定する保険者の作業を支援する装置に関する。

### 背景技術

[0002] 近年、保険者が被保険者に対して負担する医療費の増加が保険者の財政を圧迫しており、医療費削減に向けた施策作りが急務である。その施策の一つとして、保険者は被保険者の健康増進に向けた保健指導を実施している。具体的には、被保険者の健康診断結果に対して一定の基準を設定し、基準を満たす人（該当者）に対して保健指導を行う。例えば、生活習慣の改善が必要な人や、将来の傷病リスクが高い人などが該当するように基準が設定される。一方、基準を満たさない人（非該当者）に対しては、生活習慣を改善する必要なしと判断して、保健指導は行わない。

[0003] 平成20年4月から保険者（国保・被用者保険）には、40歳以上の被保険者・被扶養者を対象として、内臓脂肪型肥満に着目した健診及び保健指導の事業実施が義務づけられている。この事業で用いられている保健指導基準は厚生労働省が定めている。

[0004] しかし、保険者ごとに非保険者の特徴が異なるため、一律の条件で保健指導対象者を絞り込んだ場合には、効果に大きな差が生じると考えられる。そのため、国が定めた保健指導対象者とは別に、保険者独自の基準を設けてその基準に当てはまる非保険者を対象にした保健指導を行うことが望ましいと考えられている。

[0005] 特に、問題になっているのは、新規に保健指導基準に該当する人への対応である。当該年度の保健指導の基準を満たさない人（非該当者）の中には、次年度に保健指導の基準を満たして該当者になる人もいる。そのため、保険者が当該年度の該当者に対して保健指導を行い、該当者の健康増進を実現し

ても、毎年度、新規に該当者が増えてしまうと、被保険者全体では被保険者の健康増進を達成できない。

[0006] 被保険者全体の健康増進のために、保険者が被保険者全員に健康増進対策を行うのが望ましいが、膨大なコストがかかるため、保険者は、保健指導基準とは別に、保険者独自の基準を設定して、当該年度の被保険者の中から、次年度に該当する被保険者を選定し、選定した限られた群に対して、次年度に基準に該当しないように当該年度に保健指導を実施したいと考えている。

[0007] 保健指導対象者を選定する技術の一例が、特許文献 1 に記載されている。特許文献 1 に記載の技術は、保険者が対象者の選定条件を設定することで、その条件を満たす保健指導対象群に対して指導をした場合の医療費削減効果の予測値を出力する機能を提供している。保険者は、この支援システムを使うことで、ある特定の選定条件を設定した場合の医療費の削減効果を知ることができる。

## 先行技術文献

### 特許文献

[0008] 特許文献1：特開 2 0 0 7 - 2 5 7 5 6 5 号公報

### 発明の概要

### 発明が解決しようとする課題

[0009] しかし、特許文献 1 に記載の技術のように、保健指導によって医療費削減効果がある群を選定する方法では、非該当者の中から、次年度に該当しないにもかかわらず、保健指導による医療費削減効果が高いということで選定してしまう可能性があり、必ずしも、次年度に該当する人を選ぶとは限らない。

[0010] 本発明の目的は、上述したような課題、すなわち、次年度に保健指導基準に該当する見込みがあるかないかの観点から保健指導対象者を選定する条件を作成することが困難である、という課題を解決する保健指導対象者選定条件作成支援装置を提供することにある。

## 課題を解決するための手段

- [0011] 本発明の第 1 の観点にかかる保健指導対象者選定条件作成支援装置は、
- 第 1 の期間における個人の健康診断データである第 1 の健康診断データと、前記個人が前記第 1 の期間の次の期間である第 2 の期間の健康診断データで予め定められた保健指導基準に該当したか否かを表すラベル値とを記憶するメモリと、前記メモリに接続されたプロセッサとを備え、
- 前記プロセッサは、
- 前記健康診断データの複数の健康診断項目を複数の説明変数とし、前記説明変数と前記説明変数毎の係数とから構成される多項式で表現され、前記個人が前記第 2 の期間の健康診断データで前記保健指導基準に該当するか否かを判別するための判別モデルを、前記第 1 の健康診断データと前記ラベル値とを用いて学習し、
- 前記学習後の前記判別モデルにおける前記複数の説明変数としての前記複数の健康診断項目と前記係数の値との組み合わせを保健指導対象者選定条件として生成する
- ようにプログラムされている、といった構成を採る。
- また本発明の第 2 の観点にかかる保健指導対象者選定条件作成支援方法は、
- 第 1 の期間における個人の健康診断データである第 1 の健康診断データと、前記個人が前記第 1 の期間の次の期間である第 2 の期間の健康診断データで予め定められた保健指導基準に該当したか否かを表すラベル値とを記憶するメモリと、前記メモリに接続されたプロセッサとを備えた装置が実行する保健指導対象者選定条件作成支援方法であって、
- 前記プロセッサが、
- 前記健康診断データの複数の健康診断項目を複数の説明変数とし、前記説明変数と前記説明変数毎の係数とから構成される多項式で表現され、前記個人が前記第 2 の期間の健康診断データで前記保健指導基準に該当するか否かを判別するための判別モデルを、前記第 1 の健康診断データと前記ラベル値

とを用いて学習し、

前記学習後の前記判別モデルにおける前記複数の説明変数としての前記複数の健康診断項目と前記係数の値との組み合わせを保健指導対象者選定条件として生成する、  
といった構成を採る。

## 発明の効果

[0012] 本発明は上述したような構成を有するため、保険者は、次期間に保健指導基準に該当する見込みがあるかないかの観点から保健指導対象者を選定する条件を作成することが可能になる。

## 図面の簡単な説明

[0013] [図1]本発明の第1の実施形態のブロック図である。

[図2]本発明の第1の実施形態の動作を示すフローチャートである。

[図3]本発明の第1の実施形態における式3で与えられる目的関数を最大化する $W^*$ の値と対応する健康診断項目の例を示す図である。

[図4]本発明の第1の実施形態における動作説明用のグラフラプラシアンおよび正規化グラフラプラシアンを示す図である。

[図5]本発明の第1の実施形態における式4で与えられる目的関数を最大化する $W^*$ の値と対応する健康診断項目の例を示す図である。

[図6]本発明の第1の実施形態において生成された選定条件の例を示す図である。

[図7]本発明の第2の実施形態のブロック図である。

[図8]本発明の第2の実施形態の動作を示すフローチャートである。

## 発明を実施するための形態

[0014] 次に本発明の実施の形態について図面を参照して詳細に説明する。

### [第1の実施形態]

図1を参照すると、本発明の第1の実施形態にかかる保健指導対象者選定条件作成支援装置1は、保健指導基準に該当する見込みがあるかないかの観点から保健指導対象者を選定する条件を作成する機能を有している。

- [0015] 保健指導対象者選定条件作成支援装置 1 は、ハードウェアとして、通信インターフェース部（通信 I / F 部） 1 1、操作入力部 1 2、画面表示部 1 3、記憶部 1 4、およびプロセッサ 1 5 を有する。
- [0016] 通信 I / F 部 1 1 は、専用のデータ通信回路からなり、通信回線（図示せず）を介して接続された図示しない各種装置との間でデータ通信を行う機能を有している。操作入力部 1 2 は、キーボードやマウスなどの操作入力装置からなり、オペレータの操作を検出してプロセッサ 1 5 に出力する機能を有している。画面表示部 1 3 は、LCD や PDP などの画面表示装置からなり、プロセッサ 1 5 からの指示に応じて、操作メニューや選定条件などの各種情報を画面表示する機能を有している。
- [0017] 記憶部 1 4 は、ハードディスクや半導体メモリなどの記憶装置からなり、プロセッサ 1 5 での各種処理に必要な処理情報やプログラム 1 4 P を記憶する機能を有している。プログラム 1 4 P は、プロセッサ 1 5 に読み込まれて実行されることにより各種処理部を実現するプログラムであり、通信 I / F 部 1 1 などのデータ入出力機能を介して外部装置（図示せず）やコンピュータ読取可能な記憶媒体（図示せず）から予め読み込まれて記憶部 1 4 に保存される。記憶部 1 4 に記憶される主な処理情報として、健康診断データ 1 4 A と、フラグ 1 4 B と、保険者の希望条件 1 4 C と、判別モデル 1 4 D と、選定条件 1 4 E とがある。
- [0018] 健康診断データ 1 4 A は、過去の或る年度（以下、基準年度と記す）における個人の健康診断データである。健康診断データ 1 4 A は、個人別に分けられている。1 個人の 1 年度分の健康診断データは、その個人を一意に識別する個人 ID、受診年度、受診年齢、性別等の個人情報と、身長、体重、腹囲、BMI 値、最低血圧、最高血圧、血糖値、中性脂肪などの各種検査値と、「飲酒量が 1 日あたり 500 ml 以上」に該当するか否か等の各種問診結果とを有している。ここでは、1 年を 1 期間としているが、期間は、1 年未満であれば任意である。
- [0019] フラグ 1 4 B は、基準年度の次の年度における健康診断において所定の保

健指導基準に該当したか否かを示す個人別のラベル値である。所定の保健指導基準とは、例えば内蔵脂肪型肥満に着目した保健指導基準のことである。但し、本発明はそのような例に限定されない。

[0020] 保険者の希望条件 1 4 C は、保険者が積極的に保健指導に参加させたい個人に関する条件である。例えば、年齢が 4 0 歳代の個人を積極的に保健指導に参加させたいという条件や、或る特定の健康診断の検査項目の結果が或る条件を満たす個人を積極的に保健指導に参加させたいという条件などである。

[0021] 判別モデル 1 4 D は、個人の健康診断データと個人が基準年度の次年度に所定の保健指導基準に該当するか否かとの関係を示すモデルである。判別モデル 1 4 D は、線形回帰モデル、ロジスティック回帰モデルなどが考えられる。判別モデル 1 4 D は、一般に、複数の説明変数とその係数（パラメータ）とからなる多項式である。個々の説明変数として、健康診断データにおける個々の健康診断項目が使用される。健康診断データにおける全ての健康診断項目を説明変数としても良いし、健康診断データにおける一部の健康診断項目を説明変数としても良い。例えば、健康診断データが、受診年齢、性別、身長、体重、腹囲、BMI 値、最低血圧、最高血圧、血糖値、中性脂肪、「飲酒量が 1 日あたり 5 0 0 m l 以上」に該当するか否かの問診結果との合計 1 1 個の健康診断項目を有する場合、この 1 1 個すべてを説明変数としても良いし、例えば、受診年齢と性別と身長を除く 8 個の健康診断項目だけを説明変数としても良い。

[0022] 選定条件 1 4 E は、学習後の判別モデル 1 4 C から作成される保健指導対象者選定条件である。選定条件 1 4 E は、本実施形態の場合、健康診断項目とその係数の組み合わせと、判定しきい値とから構成される。健康診断項目とその係数の組み合わせは、学習後の判別モデル 1 4 D における複数の係数のうち、値が 0 でない係数と当該係数に対応する説明変数としての健康診断項目との組み合わせを意味する。また、判定しきい値は、学習後の判別モデル 1 4 D において或る個人が基準年度の次年度に保健指導基準に該当する確

率が予め設定された閾値（例えば1/2）以上であると判定できる、上記組み合わせに含まれる係数の値の合計値の最小値を意味する。

[0023] プロセッサ15は、CPUなどのマイクロプロセッサとその周辺回路を有し、記憶部14からプログラム14Pを読み込んで実行することにより、上記ハードウェアとプログラム14Pとを協働させて各種処理部を実現する機能を有している。プロセッサ15で実現される主な処理部として、入力部15A、判別モデル学習部15B、および条件作成部15Cがある。

[0024] 入力部15Aは、通信I/F部11または操作入力部12から、健康診断データ14A、フラグ14B、保険者の希望条件14C、および学習前の判別モデル14Dを入力して、記憶部14に格納する機能を有する。

[0025] 判別モデル学習部15Bは、健康診断データ14Aとフラグ14Bと保険者の希望条件14Cと学習前の判別モデル14Dを記憶部14から読み込み、健康診断データ14Aとフラグ14Bと保険者の希望条件14Cとを用いて、判別モデル14Dを学習し、学習後の判別モデル14Dを記憶部14に保存する機能を有する。

[0026] 判別モデル14Dの学習では、判別モデル学習部15Bは、健康診断データ14Aのうち、次年度に保健指導基準に該当した個人の健康診断データを正例として用い、次年度に保健指導基準に該当しなかった個人の健康診断データを負例として用いる。

[0027] また判別モデル14Dの学習では、判別モデル学習部15Bは、判別モデルの尤度を表す項と、値が0でない係数の個数に依存する罰則項と、保険者の希望条件を満たさない個人が次年度に保健指導基準に該当することに依存する罰則項とを有する目的関数を最適化するように判別モデル14Dの上記係数を学習する。

[0028] 条件作成部15Cは、学習後の判別モデル14Dを記憶部14から読み込み、学習後の判別モデル14Dにおける値が0でない係数と当該係数に対応する説明変数としての健康診断項目との組み合わせと、上述した判定しきい値とを、選定条件14Eとして生成して記憶部14に保存する機能を有する

。また条件作成部 15 C は、記憶部 14 から選定条件 14 D を読み込み、画面表示部 13 に出力し、あるいは通信 I/F 部 11 を通じて外部に出力する機能を有する。

[0029] 次に、図 2 を参照して、本実施形態にかかる保健指導対象者選定支援装置 1 の動作について説明する。

[0030] まず入力部 15 A は、健康診断データ 14 A、フラグ 14 B、保険者の希望条件 14 C、および学習前の判別モデル 14 D を、通信 I/F 部 11 または操作入力部 12 から入力し、記憶部 14 に格納する（ステップ S 1）。

[0031] 次に、判別モデル学習部 15 B は、記憶部 14 から健康診断データ 14 A、フラグ 14 B、保険者の希望条件 14 C、および学習前の判別モデル 14 D を読み出し、健康診断データ 14 A、フラグ 14 B、および保険者の希望条件 14 C を用いて、判別モデル 14 D を学習する（ステップ S 2）。すなわち、判別モデル学習部 15 B は、健康診断データ 14 A のうち、次年度に保健指導基準に該当した個人の健康診断データを正例として用い、次年度に保健指導基準に該当しなかった個人の健康診断データを負例として用いて、個人の健康診断データと個人が次年度に所定の保健指導基準に該当するか否かとを適切に関係付けるための判別モデル 14 D における各説明変数の係数の値を、保険者の希望条件 14 C が極力満たされるように学習する。学習後の判別モデル 14 D は、記憶部 14 に保存される。

[0032] 次に、条件作成部 15 C は、記憶部 14 から学習後の判別モデル 14 D を読み出し、学習後の判別モデル 14 D の複数の説明変数のうち、その係数の値が 0 以外の説明変数の組み合わせと判定しきい値とを、選定条件 14 E として記憶部 14 に保存し、また、画面表示部 13 に出力し、あるいは通信 I/F 部 11 を通じて外部に出力する（ステップ S 3）。

[0033] 続いて、本実施形態の動作をより詳細に説明する。なお、以下では、下付きの添え字は、アンダーラインを付して表現する。例えば、 $A_b$  は  $A\_B$  と表記する。また、上付きの添え字は、ハットを付して表現する。例えば、 $A^b$  は  $A^{\wedge}B$  と表記する。

[0034] (1-1) ステップ S 1 の詳細

入力部 1 5 A は、健康診断データ 1 4 A と、フラグ 1 4 B と、保険者の希望条件 1 4 C と、学習前の判別モデル 1 4 D とを入力する。

[0035] 入力された健康診断データを  $X_n (n=1, 2, \dots, N)$  とする。N は、保健指導対象者候補となる個人の数を表す。 $X_n$  は、個人 n の基準年度の健康診断データとする。 $X_{nj} (j=1, \dots, M)$  は、個人 n の健康診断項目 j の測定結果を表す（健康診断項目は、問診結果をも含む）。M は健康診断項目数である。

[0036] さらに、入力されたフラグを、 $Y_n (n=1, \dots, N)$  とする。すなわち、 $Y_n$  は、個人 n が基準年度の次年度の健康診断で所定の保健指導基準に該当したか ( $Y_n=1$ )、しなかったか ( $Y_n=0$ ) を表すフラグである。

[0037] 入力部 1 5 A は、前処理として、健康診断データの 2 値化を行う。2 値化のための閾値を健康診断項目ごとに設定し、その閾値をもとに  $X_{nj} (j=1, \dots, M)$  を 0 か 1 に 2 値化する。各検査項目の閾値は、厚生労働省が定める保健指導判定基準を使用してもよい。例えば、BMI 値を、25 以上ならば 1 に、それ以外は 0 に 2 値化する。なお、入力された健康診断データが既に 2 値化されている場合には、この処理は不要である。

[0038] (1-2) ステップ S 2 の詳細

判別モデル 1 4 D で求める値は、以下の P である。判別モデル学習部 1 5 B は、この値を算出するための判別モデル 1 4 D を用いて、そのパラメータを学習する。

[0039]  $P(Y_n=1)$  は、個人 n が所定の保健指導基準に該当する確率を表す。個人 n が所定の保健指導基準に該当しない確率は、 $1 - P(Y_n=1)$  で求められる。これを、 $P(Y_n=0)$  とする。

[0040] ここで、 $Y_n=1$  の  $X_n$  のことを正例、 $Y_n=0$  の  $X_n$  のことを負例と呼ぶ。例えば、ある  $X_n$  が正例または負例である確率を出力可能なロジスティック回帰モデルを用いると良い。ロジスティック回帰モデルは、 $X_n$  から  $Y_n=1$  か  $Y_n=0$  かを判別する二値判別問題によく適用されるモデルである。以下、ロジスティック回帰の数理的構造について説明する。

[0041] Xを基準年の健康診断データに対応するM次元の説明変数とし、Yを基準年の次年度における健康診断データで所定の保健指導基準に該当したか、しなかったかを表す確率変数とする（Y=1は該当、Y=0は非該当を表す）。WをM次元の重みベクトルとすると、ロジスティック回帰モデルは、次式で表される。

$$P(Y=1|X; W)=1/(1+\exp(W^T X)) \quad \dots(1)$$

$$P(Y=0|X; W)=1-P(Y=1|X; W) \quad \dots(2)$$

ただし、 $P(\bullet|\circ;\star)$ は $\star$ をパラメータとし、 $\circ$ が与えられた場合の $\bullet$ の条件付確率を表す。また、上付きのTはベクトルの転置を表す。

[0042] 学習データとして正例と負例 $\{X_n, Y_n\}$  ( $n=1, \dots, N$ )が与えられた場合に、ロジスティック回帰では、以下の目的関数を最適化することによって、パラメータWの値を算出する。ただし、 $X_n$ と $Y_n$ はそれぞれXとYの実現値とする。

$$L(W)=-\sum_{n=1}^N \log P(Y_n|X_n, W) \quad \dots(3)$$

ここで、 $\sum_{n=1}^N$ は、 $n=1$ からNまでの総和を表す。

[0043]  $L(W)$ は勾配法に準じた方法によって最大化することが可能である。 $L(W)$ を最大化するパラメータWの値を $W^*$ とする。

[0044] 上記式3で与えられる目的関数を最大化する $W^*$ の値と対応する健康診断項目の例を図3に示す。図3に示す例は、腹囲、BMI、血糖、脂質、飲酒しない、という5つの健康診断項目を説明変数とする上記目的関数について、それを最大化する各説明変数の係数の値を求めた結果の一例である。

[0045] ここで、図3を参照すると、5つ全ての説明変数の係数が0以外の値となっている。これは、選定条件の条件項目数が5項目であることを意味している。一般に被保険者を保健指導に参加させるには、保険者から被保険者に対して、選定理由を開示し、被保険者の理解を得る必要がある。条件項目数が多い複雑な選定理由を設定してしまうと、被保険者の理解を得ることが困難になる。そのため、選定条件となる条件項目数はできるだけ少ない方が望ましい。また、上記3式の目的関数では、保険者の希望条件14Cに沿う最適化が行えない。

[0046] そこで、判別モデル学習部 15 B は、上記3式による目的関数に代えて、下記4式による目的関数を最適化する  $W^*$  を学習する。

$$L(W) = -\sum_{n=1}^N \log P(Y_n | X_n, W) - (\lambda/2) * ||W|| - (\alpha/2) F^T L' F \quad \dots(4)$$

ここで、 $||W||$  は  $W$  のノルムであり、ノルム1を用いる。  $F$  は、  $N$  次元ベクトルで、  $n$  番目の要素は  $W^T X_n$  である。また  $L'$  は、正規化グラフラプラシアンである。  $\lambda$  と  $\alpha$  は、右辺の第1項と第2項と第3項のバランスを調整するパラメータである。

[0048] 上記4式の右辺の第一項は、上記3式の右辺と同じであり、判別モデルの尤度を表す。

[0049] 上記4式の右辺の第二項は、値が0でない係数の個数に依存する罰則項である。この右辺の第二項は、判別モデルの個々の説明変数の係数、つまり  $W^*$  の要素が0でない説明変数の個数を減らす効果がある。

[0050] 上記4式の右辺の第三項は、保険者の希望条件を満たさない個人が保健指導基準に該当することに依存する罰則項である。この右辺の第三項は、保険者が希望する条件に沿うように、  $W^*$  の各要素の重みを調整する効果がある。この右辺の第3項を入れることによって、保険者が希望する条件を出来る限り満たすように、  $W^*$  を学習することになるため、学習後の  $W^*$  の非ゼロの成分を抽出して、条件にすると、条件項目数のより少なく、保険者が希望する条件に沿うような、保健指導対象者選定条件が生成できるようになる。

[0051] 以下、正規化グラフラプラシアン  $L'$  について説明する。

[0052] まず初めに、グラフラプラシアンについて説明する。グラフラプラシアンは、  $N \times N$  の行列である。  $N$  は、保健指導対象者候補となる個人の数を表す。説明の便宜上、個人の数を  $N=3$  (個人1, 個人2, 個人3) として、グラフラプラシアンの例を図4に示す。また正規化グラフラプラシアンは、この例ではグラフラプラシアンと同じになる。

[0053] 図4のグラフラプラシアンおよび正規化グラフラプラシアンにおける各行と列は、個人を表す。図4の一行目は、個人1と個人2にリンクがあるという

ことを表す。図4の二行目は、個人1と個人2にリンクがあるということを表す。図4の三行目は、リンクがないことを表す。

[0054] 保険者の希望条件14Cが、40歳代の個人を積極的に保健指導に参加させたいという条件であった場合を例にすると、リンクがあるというのは、保険者が希望する条件を満たすことを示す。図4の例では、個人1と個人2は、保険者が希望する条件（40歳代の個人）を満たし、個人3は、保険者が希望する条件（40歳代の個人）を満たさないことを示している。ここで、上記の正規化グラフラプシアン $L'$ 自体を保険者の希望条件14Cとして記憶部14に記憶しておいて、式4の目的関数に適用するようにしても良い。或いは、40歳代の個人を積極的に保健指導に参加させたいという条件自体を保険者の希望条件14Cとして記憶部14に記憶しておいて、その条件から上記の正規化グラフラプシアン $L'$ を生成して式4の目的関数に適用するようにしても良い。

[0055] 因みに、図4の正規化グラフラプシアンを使用した場合、上記4式の右辺の第三項は次式と等価である。

$$-(\alpha/2) \{ (W^{\{T\}}X_1 - W^{\{T\}}X_2)^2 + (W^{\{T\}}X_3)^2 \} \quad \dots(5)$$

上記4式の目的関数を最大化するためには、保険者の希望する条件を満たさない個人3に係る $(W^{\{T\}}X_3)^2$ の値が小さくなるように $W^*$ を学習する必要がある。また、保険者の希望する条件を満足する個人1,2については、それら個人間で $W^{\{T\}}X$ の値が等しくなるように $W^*$ を学習する必要がある。

[0056]  $W^*$ は、M次元ベクトルであるが、M次元ベクトルの各要素の内、非ゼロの要素の値とその要素に対応する健康診断項目をセットで、選定条件14Eとして記憶部14に保存する。つまり、 $(W^*_j, \text{健康診断項目}j) \{j | W^*_j \neq 0\}$ を選定条件14Eとして記憶部14に保存する。

[0057] 上記式4で与えられる目的関数を最大化する $W^*$ の値と対応する健康診断項目の例を図5に示す。図5に示す例は、腹囲、BMI、血糖、脂質、飲酒しない、という5つの健康診断項目を説明変数とする式4による目的関数について、それを最大化する各説明変数の係数の値を求めた結果である。図5を参照する

と、図3と相違し、BMIと血糖値の係数の値が0になっている。

[0058] 図6は、図5から非ゼロの要素に対応する健康診断項目と当該健康診断項目に対応する係数の値とを抽出して生成した選定条件14Eの一例を示す。図6を参照すると、図3と相違し、選定条件の条件項目数が5項目から3項目に削減されている。なお、図6の例では、係数の値はスコアとして扱っている。

[0059] さらに、上記1式において $P(Y=1|X; W)=TH$ になるときの、 $W^T X$ の値を、『次年度に所定の基準に該当』か、或いは『次年度に所定の基準に非該当』かを定める判定しきい値として算出し、選定条件14Eの一部として記憶部14に保存する。THは例えば0.5であるが、他の値であってもよい。

[0060] 条件作成部15Cは、上述のようにして生成した選定条件14Eを構成する、 $(W*_j, \text{健康診断項目 } j) \{j | W*_j \neq 0\}$ の条件と、判定しきい値とを、画面表示部13に出力し、或いは通信I/F部11から外部に出力する。

[0061] このように本実施形態によれば、保険者は、次期間に保健指導基準に該当する見込みがあるかないかの観点から保健指導対象者を選定する条件を作成することが可能になる。その理由は、健康診断データの複数の健康診断項目を複数の説明変数とし、この説明変数とこの説明変数毎の係数とから構成される多項式で表現され、個人が基準年度の次年度の健康診断データで保健指導基準に該当するか否かを判別するための判別モデル14Dを、基準年度の健康診断データ14Aと各個人が基準年度の次年度の健康診断データで保健指導基準に該当したか否かを表すフラグ14Bとを用いて学習し、学習後の判別モデル14Dにおける複数の説明変数としての複数の健康診断項目と係数の値との組み合わせを選定条件14Eとして生成するためである。

[0062] また本実施形態によれば、選定条件としての健康診断項目の数を削減することが可能になる。その理由は、判別モデル14Dの学習では、値が0でない係数の個数に依存する罰則項を有する式4による目的関数を最適化するように判別モデル14Dの係数の値を学習するためである。

[0063] また本実施形態によれば、保険者が希望する条件に沿うような選定条件を

生成することが可能になる。その理由は、判別モデル 1 4 D の学習では、保険者の希望する条件を満たさない個人が保健指導基準に該当することに依存する罰則項を有する式 4 による目的関数を最適化するように判別モデル 1 4 D の係数の値を学習するためである。

[0064] また本実施形態によれば、個人が次年度に所定の基準に該当するか否かを判定するためのしきい値を、選定条件 1 4 E の一部として生成することができる。その理由は、上記 1 式において  $P(Y=1|X; W)=TH$  になるときの、 $W^T X$  の値を算出して、選定条件 1 4 E の一部として記憶部 1 4 に保存するためである。

[0065] [第 2 の実施形態]

図 7 を参照すると、本発明の第 2 の実施形態にかかる保健指導対象者選定条件作成支援装置 2 は、保健指導基準に該当する見込みがあるかないかの観点から保健指導対象者を選定する条件を作成する機能と、作成された選定条件に従って保健指導対象者を選定する機能とを有している。

[0066] 保健指導対象者選定条件作成支援装置 2 は、ハードウェアとして、通信インターフェース部（通信 I / F 部）2 1、操作入力部 2 2、画面表示部 2 3、記憶部 2 4、およびプロセッサ 2 5 を有する。

[0067] 通信 I / F 部 2 1、操作入力部 2 2、および画面表示部 2 3 は、第 1 の実施形態における通信 I / F 部 1 1、操作入力部 1 2、および画面表示部 1 3 と同様の機能を有している。

[0068] 記憶部 2 4 は、ハードディスクや半導体メモリなどの記憶装置からなり、プロセッサ 2 5 での各種処理に必要な処理情報やプログラム 2 4 P を記憶する機能を有している。プログラム 2 4 P は、プロセッサ 2 5 に読み込まれて実行されることにより各種処理部を実現するプログラムであり、通信 I / F 部 2 1 などのデータ入出力機能を介して外部装置（図示せず）やコンピュータ読取可能な記憶媒体（図示せず）から予め読み込まれて記憶部 2 4 に保存される。記憶部 2 4 に記憶される主な処理情報として、健康診断データ 2 4 A と、フラグ 2 4 B と、保険者の希望条件 2 4 C と、判別モデル 2 4 D と、

選定条件 2 4 E と、選定用の健康診断データ 2 4 F と、選定者 2 4 G とがある。

[0069] 健康診断データ 2 4 A、フラグ 2 4 B、保険者の希望条件 2 4 C、判別モデル 2 4 D、および選定条件 2 4 E は、第 1 の実施形態における健康診断データ 1 4 A、フラグ 1 4 B、保険者の希望条件 1 4 C、判別モデル 1 4 D、および選定条件 1 4 E と同じである。

[0070] 選定用の健康診断データ 2 4 F は、保健指導者を選定する年度（以下、選定年度と記す）における個人の健康診断データである。健康診断データ 2 4 F は、個人別に分けられている。1 個人の当該年度分の健康診断データは、基準年の健康診断データ 2 4 A と同じ項目を有する。すなわち、1 個人の健康診断データ 2 4 F は、その個人を一意に識別する個人 ID、受診年度、受診年齢、性別等の個人情報と、身長、体重、腹囲、最低血圧、最高血圧、血糖値、中性脂肪などの各種検査値と、「飲酒量が 1 日あたり 5 0 0 m l 以上」に該当するか否か等の各種問診結果とを有している。なお、選定用の健康診断データ 2 4 F は、所定の保健指導基準に該当しない個人の健康診断データのみで構成されていても良いし、所定の保健指導基準に該当する個人および該当しない個人の健康診断データが混在していても良い。

[0071] 選定者 2 4 G は、保健指導対象者として選定された個人を特定する情報、例えば個人 ID のリストである。

[0072] プロセッサ 2 5 は、CPU などのマイクロプロセッサとその周辺回路を有し、記憶部 2 4 からプログラム 2 4 P を読み込んで実行することにより、上記ハードウェアとプログラム 2 4 P とを協働させて各種処理部を実現する機能を有している。プロセッサ 2 5 で実現される主な処理部として、入力部 2 5 A、判別モデル学習部 2 5 B、条件作成部 2 5 C、および保健指導対象者選定部 2 5 D がある。

[0073] 入力部 2 5 A、判別モデル学習部 2 5 B、および条件作成部 2 5 C は、第 1 の実施形態における入力部 1 5 A、判別モデル学習部 1 5 B、および条件作成部 1 5 C と同様の機能を有する。

[0074] 保健指導対象者選定部 25 D は、選定条件 24 E と選定用の健康診断データ 24 F を記憶部 24 から読み込み、健康診断データ 24 F から、選定条件 24 E に適合する健康診断データを有する個人を保健指導対象者として決定し、選定者 24 G として記憶部 24 に記憶する機能を有する。また、保健指導対象者選定部 25 D は、記憶部 24 から選定者 24 G を読み込み、画面表示部 23 に出力し、あるいは通信 I / F 部 21 を通じて外部に出力する機能を有する。

[0075] 次に、図 8 を参照して、本実施形態にかかる保健指導対象者選定支援装置 2 の動作について説明する。

[0076] まず入力部 25 A は、健康診断データ 24 A、フラグ 24 B、保険者の希望条件 24 C、学習前の判別モデル 24 D、選定用の健康診断データ 24 F を、通信 I / F 部 21 または操作入力部 22 から入力し、記憶部 24 に格納する（ステップ S 11）。

[0077] 次に、判別モデル学習部 25 B は、記憶部 24 から健康診断データ 24 A、フラグ 24 B、保険者の希望条件 24 C、および学習前の判別モデル 24 D を読み出し、健康診断データ 24 A、フラグ 24 B、および保険者の希望条件 24 C を用いて、第 1 の実施形態における判別モデル学習部 15 B と同様に、判別モデル 24 D を学習する（ステップ S 12）。学習後の判別モデル 24 D は、記憶部 24 に保存される。

[0078] 次に、条件作成部 25 C は、記憶部 24 から学習後の判別モデル 24 D を読み出し、第 1 の実施形態における条件作成部 15 C と同様に、学習後の判別モデル 24 D の複数の説明変数のうち、その係数の値が 0 以外の健康診断項目とその係数値との組み合わせ、および判定しきい値を、選定条件 24 E として作成して記憶部 24 に保存し、また、画面表示部 23 に出力し、あるいは通信 I / F 部 21 を通じて外部に出力する（ステップ S 13）。

[0079] 次に、保健指導対象者選定部 25 D は、選定条件 24 E と選定用の健康診断データ 24 F を記憶部 24 から読み込み、健康診断データ 24 F から、選定条件 24 E に適合する健康診断データを有する個人を保健指導対象者とし

て決定し、選定者 2 4 G として記憶部 2 4 に記憶すると共に、画面表示部 2 3 に出力し、あるいは通信 I / F 部 2 1 を通じて外部に出力する（ステップ S 1 4）。

[0080] より具体的には、保健指導対象者選定部 2 5 D は、選定用の健康診断データ 2 4 F に含まれる各個人の健康診断データ毎に、選定条件 2 4 E 中の各健康診断項目のうちの該当する項目に対するスコアの総和を計算する。例えば、選定条件 2 4 E が図 7 に示すものであり、或る個人が、「腹囲」と「脂質」の項目に該当し、「飲酒しない」の項目に該当しないならば、 $3 + 3 = 6$  を当該個人のスコアとする。そして、保健指導対象者選定部 2 5 D は、個人の上記スコアを選定条件 2 4 中の判定しきい値と比較し、個人のスコア > 判定しきい値であれば、当該個人を保健指導対象者に選定する。

[0081] このように本実施形態によれば、第 1 の実施形態と同様の効果が得られると共に、次年度に保健指導基準に該当する見込みがあるかないかの観点から保健指導対象者を選定することが可能になる。

[0082] [その他の実施形態]

以上、本発明を幾つかの実施形態を挙げて説明したが、本発明は以上の実施形態のみに限定されず、その他各種の付加変更が可能である。例えば、以下のような実施形態も本発明に含まれる。

[0083] 前述した実施形態では、保険者の希望条件に沿うような保健指導対象者の選定条件を生成したが、保険者の希望条件を考慮せずに選定条件を生成するようにしてもよい。この場合、上記 4 式における右辺の第三項を省略した目的関数を使用すれば良い。

[0084] 前述した実施形態では、選定条件としての健康診断項目の数が極力少なくなるようにしたが、その必要性がない場合には、上記 4 式における右辺の第二項を省略した目的関数を使用すれば良い。

[0085] 前述した第 2 の実施形態では、判定しきい値を算出し、判定しきい値以上のスコアを有する個人を保健指導対象者に選定したが、判定しきい値を使用せずに保健指導対象者を選定するようにしても良い。例えば、上記 1 式を用い

て各個人毎に次年度に保健指導対象者に該当する確率を計算し、その確率の上位N人を保健指導対象者に選定しても良い。

[0086] なお、本発明は、日本国にて2012年4月26日に特許出願された特願2012-100937の特許出願に基づく優先権主張の利益を享受するものであり、当該特許出願に記載された内容は、全て本明細書に含まれるものとする。

## 符号の説明

[0087] 1、2…保健指導対象者選定条件作成支援装置

11、21…通信1／F部

12、22…操作入力部

13、23…画面表示部

14、24…記憶部

15、25…プロセッサ

## 請求の範囲

### [請求項1]

第1の期間における個人の健康診断データである第1の健康診断データと、前記個人が前記第1の期間の次の期間である第2の期間の健康診断データで予め定められた保健指導基準に該当したか否かを表すラベル値とを記憶するメモリと、前記メモリに接続されたプロセッサとを備え、

前記プロセッサは、

前記健康診断データの複数の健康診断項目を複数の説明変数とし、前記説明変数と前記説明変数毎の係数とから構成される多項式で表現され、前記個人が前記第2の期間の健康診断データで前記保健指導基準に該当するか否かを判別するための判別モデルを、前記第1の健康診断データと前記ラベル値とを用いて学習し、

前記学習後の前記判別モデルにおける前記複数の説明変数としての前記複数の健康診断項目と前記係数の値との組み合わせを保健指導対象者選定条件として生成する

ようにプログラムされている保健指導対象者選定条件作成支援装置。

### [請求項2]

前記プロセッサは、

前記判別モデルの学習では、前記判別モデルの尤度を表す項と、値が0でない前記係数の個数に依存する罰則項とを有する目的関数を最適化するように前記判別モデルの前記係数の値を学習する

請求項1に記載の保健指導対象者選定条件作成支援装置。

### [請求項3]

前記メモリは、さらに、保険者の希望する条件を記憶し、

前記プロセッサは、

前記判別モデルの学習では、前記判別モデルの尤度を表す項と、値が0でない前記係数の個数に依存する罰則項と、前記保険者の希望する条件を満たさない前記個人が前記保健指導基準に該当することに依存する罰則項とを有する目的関数を最適化するように前記判別モデルの前記係数の値を学習する

請求項 1 に記載の保健指導対象者選定条件作成支援装置。

[請求項4]

前記プロセッサは、

前記保健指導対象者選定条件の生成では、前記学習後の判別モデルにおける前記複数の係数のうち、値が 0 でない係数と当該係数に対応する説明変数としての前記健康診断項目との組み合わせを、前記保健指導対象者選定条件として生成する

請求項 2 または 3 に記載の保健指導対象者選定条件作成支援装置。

[請求項5]

前記プロセッサは、

前記保健指導対象者選定条件の生成では、前記学習後の判別モデルにおける前記複数の係数のうち、値が 0 でない係数と当該係数に対応する説明変数としての前記健康診断項目との組み合わせと、前記学習後の判別モデルにおいて前記個人が前記第 2 の期間の健康診断データで前記保健指導基準に該当する確率が所定値以上であると判定できる、前記組み合わせに含まれる前記係数の値の合計値の最小値である判定しきい値とを、前記保健指導対象者選定条件として生成する

請求項 2 または 3 に記載の保健指導対象者選定条件作成支援装置。

[請求項6]

前記メモリは、さらに、保健指導対象者候補である個人の健康診断データである第 2 の健康診断データを記憶し、

前記プロセッサは、さらに、

前記第 2 の健康診断データから、前記保健指導対象者選定条件に適合する前記個人を決定する

請求項 1 乃至 5 に記載の保健指導対象者選定条件作成支援装置。

[請求項7]

前記メモリは、さらに、保健指導対象者候補である個人の健康診断データである第 2 の健康診断データを記憶し、

前記プロセッサは、さらに、

前記第 2 の健康診断データから、前記保健指導対象者選定条件に適合する前記個人を決定し、前記保健指導対象者選定条件に適合する前記個人の決定では、前記個人の前記第 2 の健康診断データ毎に、前記

保健指導対象者選定条件中の前記健康診断項目のうちの該当する項目に対するスコアの総和を計算して前記判定しきい値と比較する  
請求項5に記載の保健指導対象者選定条件作成支援装置。

[請求項8]

第1の期間における個人の健康診断データである第1の健康診断データと、前記個人が前記第1の期間の次の期間である第2の期間の健康診断データで予め定められた保健指導基準に該当したか否かを表すラベル値とを記憶するメモリと、前記メモリに接続されたプロセッサとを備えた装置が実行する保健指導対象者選定条件作成支援方法であって、

前記プロセッサが、

前記健康診断データの複数の健康診断項目を複数の説明変数とし、前記説明変数と前記説明変数毎の係数とから構成される多項式で表現され、前記個人が前記第2の期間の健康診断データで前記保健指導基準に該当するか否かを判別するための判別モデルを、前記第1の健康診断データと前記ラベル値とを用いて学習し、

前記学習後の前記判別モデルにおける前記複数の説明変数としての前記複数の健康診断項目と前記係数の値との組み合わせを保健指導対象者選定条件として生成する

保健指導対象者選定条件作成支援方法。

[請求項9]

第1の期間における個人の健康診断データである第1の健康診断データと、前記個人が前記第1の期間の次の期間である第2の期間の健康診断データで予め定められた保健指導基準に該当したか否かを表すラベル値とを記憶するメモリに接続されたプロセッサに、

前記健康診断データの複数の健康診断項目を複数の説明変数とし、前記説明変数と前記説明変数毎の係数とから構成される多項式で表現され、前記個人が前記第2の期間の健康診断データで前記保健指導基準に該当するか否かを判別するための判別モデルを、前記第1の健康診断データと前記ラベル値とを用いて学習するステップと、

前記学習後の前記判別モデルにおける前記複数の説明変数としての前記複数の健康診断項目と前記係数の値との組み合わせを保健指導対象者選定条件として生成するステップと、  
を行わせるためのプログラム。