明　細　書

発明の名称 ： 疾患検出システムおよび疾患検出方法

技術分野

[0001] 本発明は、疾患検出システムおよび疾患検出方法に関し、特に、ユーザ端末からインターネット上のサーバにアクセスして疾患の可能性を検出するシステムに用いて好適なものである。

背景技術

[0002] 従来、ＩＴ（Information Technology）技術を活用して疾患の検出を行うシステムが知られている（例えば、特許文献１～３参照）。特許文献１に記載のシステムでは、医師の所見などを記載した自由文を言語解析することによって病状語句を検出するようになされている。また、特許文献２に記載のシステムでは、病理標本画像に付随した診断テキスト情報から、定型化されていない文章の集まりを自然言語解析の手法を使って単語やフレーズに分割し、それらの出現頻度や相関関係を分析して有用な情報を抽出する手法を用いて、疾患とキーワードとの関係の変換辞書を作成するようになされている。

[0003] また、特許文献３に記載のシステムでは、人間ドック入院患者の病状診断に用いるデータを記憶する患者データベースと、患者データベースの項目とその属性を条件部とし、診断結果を結論部とする因果関係を記憶する知識データベースと、患者データベースに対する医師の持つ解説知識、所見知識等および患者への指示・注意事項等コメント項目に関するデータを記憶するコメントデータベースとから病状診断を行うようになされている。

[0004] ここで、患者データベースは、血液検査や生化学検査等の検査結果を記憶する検査結果データと、心電図検査や腹部超音波検査等の検体検査を通して得られる診察結果を記憶する診察結果データと、身長・体重・血圧・脈拍等の測定結果を記憶する測定結果データと、患者の愁訴、日常の生活習慣等の問診結果を記憶する問診結果データとからなる人間ドック入院患者の病状診

断に用いるデータを各々患者ＩＤコードと共に記憶したものである。

[0005] 特許文献1：特開２００８－１０８１９９号公報

特許文献2：特開２０１２－１７９３３６号公報

特許文献3：特開平６－１７６００７号公報

発明の開示

[0006] 上記特許文献１～３に記載のシステムは何れも、患者が医師の診断を受けたときに医師によって作られた所見やコメント等のテキスト情報をもとに、疾患や病状の検出を行うようになされている。そのため、まだ医師の診断を受けていない一般ユーザについて、罹患の疑いがある疾患をあらかじめ検出することはできない。すなわち、予防医学を目的とした一般向けの疾患セルフチェックシステムとして用いることはできない。

[0007] なお、従来、インターネット上で公開された問診票の質問に対して一般ユーザが自覚症状に応じて回答をすることで、その回答内容をもとに、罹患している可能性のある疾患の検出を行うシステムも提供されている。しかしながら、この種のシステムは、あらかじめ用意した質問に対して、どのような内容の回答をすればどの疾患に罹患している可能性が疑われるかという検出アルゴリズムが固定化されている。しかし、通常、その検出アルゴリズムは特定の医師の経験則に基づいて生成され、それが常に変わらず最適のものであるとは限らない。

[0008] 本発明は、このような問題を解決するために成されたものであり、医師の診断を受けていない一般ユーザに対しても罹患の疑いのある疾患をより高精度に検出できるようにすることを目的とする。

[0009] 上記した課題を解決するために、本発明では、ユーザのプロファイルデータと、問診票に対するユーザからの回答である問診回答データと、自由文記述形式の質問に対するユーザからの回答である自由回答データを解析して得られるテキスト関連データと、所定の重み値とを用いて、ユーザが特定の疾患に罹患している可能性を表す罹患確率を分析し、当該罹患確率が閾値以上の疾患を疾患候補として抽出する。また、上述のプロファイルデータ、問診

回答データおよびテキスト関連データをデータベースに随時格納して蓄積し、データベース内のプロファイルデータ、問診回答データおよびテキスト関連データを説明変数とし、抽出された疾患候補を目的変数として回帰分析を行う。これにより、今回の分析で抽出された疾患候補と相関する因子となる説明変数について重み値を算出し、次回以降における罹患確率の分析に用いるようにしている。

[0010] 上記のように構成した本発明によれば、問診票に対する問診回答データだけでなく、また、自由文で回答可能な質問に対する自由回答データだけでもなく、その両方を用いて、ユーザが特定の疾患に罹患している可能性を表す罹患確率が算出される。しかも、その罹患確率を算出する際には、過去の分析時に罹患確率が閾値以上となって疾患候補として抽出される原因となった因子に関して算出された重み値が罹患確率の算出に反映されているので、罹患確率の算出をより正確に行うことができる。この重み値は、罹患確率の分析回数が増えるほど、統計学的に有意な値になっていくので、罹患確率の分析精度は上がっていく。これにより、医師の診断を受けていない一般ユーザに対しても罹患の可能性のある疾患をより高精度に検出することができる。

図面の簡単な説明

[0011] [図1]本実施形態による疾患検出システムの全体構成例を示す図である。

[図2]本実施形態による疾患検出サーバの機能構成例を示すブロック図である。

[図3]本実施形態による疾患検出サーバの動作例を示すフローチャートである。

[図4]本実施形態による疾患検出システムの他の構成例を示す図である。

発明を実施するための最良の形態

[0012] 以下、本発明の一実施形態を図面に基づいて説明する。図１は、本実施形態による疾患検出システムの全体構成例を示す図である。図１に示すように、本実施形態の疾患検出システムは、ユーザ端末１００と疾患検出サーバ２００とを備え、当該ユーザ端末１００と疾患検出サーバ２００とがインター

ネット３００等の通信ネットワークを介して接続可能に構成されている。ユーザ端末１００は、オンラインによる疾患の検出を受ける一般ユーザが使用するものである。なお、図１では図示を簡略化しているが、実際にはユーザ端末１００は複数存在する。

[0013] 図２は、本実施形態による疾患検出サーバ２００の機能構成例を示すブロック図である。図２に示すように、本実施形態の疾患検出サーバ２００は、その機能構成として、通信インタフェース部１１、プロファイルデータ取得部１２、問診回答データ取得部１３、自由回答データ取得部１４、自由文解析部１５、データベース記憶部１６、データベース１７、罹患確率分析部１８、疾患データベース１９、重み値記憶部２０、疾患候補抽出部２１、検出結果提示部２２および重み値算出部２３を備えている。

[0014] 上記各機能ブロック１１～２３は、ハードウェア、ＤＳＰ（Digital Signal Processor）、ソフトウェアの何れによっても構成することが可能である。例えばソフトウェアによって構成する場合、上記各機能ブロック１１～２３は、実際にはコンピュータのＣＰＵ、ＲＡＭ、ＲＯＭなどを備えて構成され、ＲＡＭやＲＯＭ、ハードディスクまたは半導体メモリ等の記録媒体に記憶されたプログラムが動作することによって実現される。

[0015] 通信インタフェース部１１は、インターネット３００を介してユーザ端末１００との間で双方向の通信を行うことにより、情報の送信および受信を行う。プロファイルデータ取得部１２は、ユーザのプロファイル（年齢、性別、住所、職業、生活習慣など）を表すプロファイルデータを取得する。例えば、ユーザ端末１００から疾患検出サーバ２００に対するアクセスが行われたとき、プロファイルデータ取得部１２が所定のプロファイル入力画面をユーザ端末１００に提示し、そのプロファイル入力画面から入力されたユーザのプロファイルデータをプロファイルデータ取得部１２が取得する。

[0016] また、疾患検出サーバ２００が、所定のＩＤに関連付けてユーザのプロファイルデータをあらかじめユーザ情報データベース（図示せず）として記憶している場合、あるいは、そのようなユーザ情報データベースと疾患検出サ

ーバ２００とが連携している場合、プロファイルデータ取得部１２は、ユーザ端末１００から送信されるＩＤに基づいて、当該ＩＤに対応するプロファイルデータをユーザ情報データベースから取得するようにしてもよい。

[0017] 問診回答データ取得部１３は、問診票のデータをユーザ端末１００に提示し、問診票に対する回答である問診回答データを取得する。ここで提示する問診票のデータは、選択肢で回答可能なオープン質問に関する質問項目や、１日の睡眠時間や１回の平均飲酒量を問うようなクローズ質問に関する質問項目を含む問診票の形式で表わされた入力画面のデータである。ユーザは、この入力画面に表示された質問項目の１つ１つに対して回答を入力していく。そして、この回答によって生成される問診回答データがユーザ端末１００から疾患検出サーバ２００に送信され、問診回答データ取得部１３がこれを取得する。

[0018] 既知の主要な疾患に対しては、通常は集団ごとの有病率が疫学研究によって明らかにされ、また、適当な問診票が用意されている。本実施形態では、この問診票を利用する。代表的な問診票は、各質問に対する疾患の感度（特定の病気に罹患している集団に対して検査を行ったときに陽性（異常値）を示す割合）および特異度（特定の病気に罹患していない集団に対して検査を行ったときに陰性（正常値）を示す割合）が知られており、回答後の疾病罹患確率を算出することが可能である。疾患毎の有病率、疾患毎に用意された問診票に含まれる各質問に対する感度・特異度が、疾病罹患確率を算出する際に用いるパラメータとして疾患データベース１９にあらかじめ記憶されている。

[0019] 自由回答データ取得部１４は、自由文で回答可能な質問のデータをユーザ端末１００に提示し、質問に対する回答である自由回答データを取得する。ここで提示する質問のデータは、自由文記述形式で表わされた入力画面のデータである。ユーザは、この入力画面に表示された質問に対して、自覚症状などを自由文により記述する形式で回答していく。そして、この回答によって生成されるテキスト文による自由回答データがユーザ端末１００から疾患

検出サーバ２００に送信され、自由回答データ取得部１４がこれを取得する。

[0020] なお、ここでは問診票形式の入力画面と自由文記述形式の入力画面とを別々にユーザ端末１００に提示する例について説明したが、本発明はこれに限定されない。例えば、問診票と自由文記述欄との両方が含まれた１つの入力画面をユーザ端末１００に提示するようにしてもよい。

[0021] 自由文解析部１５は、自由回答データ取得部１４により取得された自由回答データの自然文をテキストマイニングのために解析する。これにより、自由文から構成単語、分かち書き、セグメント（主語述語を含む一文）、単語間距離情報、文章係り受け情報の少なくとも１つを抽出し、これをテキスト関連データとして生成する。

[0022] データベース記憶部１６は、プロファイルデータ取得部１２により取得されたプロファイルデータ、問診回答データ取得部１３により取得された問診回答データ、および、自由文解析部１５により生成されたテキスト関連データを互いに関連付けて実績データベース１７に格納する。データベース記憶部１６は、複数のユーザ端末１００から疾患検出サーバ２００にプロファイルデータ、問診回答データおよび自由回答データが送信されてくるたびに、プロファイルデータ、問診回答データおよびテキスト関連データを実績データベース１７に随時格納する。

[0023] 罹患確率分析部１８は、プロファイルデータ取得部１２により取得されたプロファイルデータ、問診回答データ取得部１３により取得された問診回答データ、自由文解析部１５により生成されたテキスト関連データ、疾患データベース１９に疾患毎に記憶されている各種パラメータ、および、重み値記憶部２０に記憶されている所定の重み値（以下に説明するリスク比）を用いて、ユーザが特定の疾患に罹患している可能性を表す罹患確率を分析する。

[0024] 具体的には、罹患確率分析部１８は、疾患データベース１９に登録されている全ての疾患について、以下の（式１）に基づいて統計学的に罹患確率（検査後確率）を算出する。

罹患確率（検査後確率）Ｐ＝検査前確率×累積リスク比・・・（式１）

[0025] まず、検査前確率の算出について説明する。検査前確率は、プロファイルデータ取得部１２により取得されたプロファイルデータと、疾患データベース１９に疾患毎に記憶されている各種パラメータとに基づいて算出する。すなわち、疾患データベース１９に登録されている全ての疾患について、統計学的有意性をもってリスク比（病気の要因を持つ人が要因のない人と比べてどれだけ病気にかかりやすいかを示す指標）が明らかになっているプロファイル要素を疾患データベース１９から参照し、その疾患の既知の有病率とプロファイルに応じたリスク比とを乗算し、これを検査前確率とする。

[0026] 具体的には、例えば全体の有病率が３％というパラメータが疾患データベース１９に記憶された或る疾患に関して、男性が対女性比で２倍リスク比が高いことが同じく疾患データベース１９にパラメータとして記憶されている場合、プロファイルデータ取得部１２により取得された回答者の性別が男性であった場合の検査前確率は４％と算出できる。

[0027] また、回答者と同じプロファイルを有する者（以下、対象プロファイル保有者という）の数が母集団（一般人口）の数に比して充分に小さければ、検査前確率は単純にそのプロファイルを有することのリスク比の乗算で近似できる。例えば、全体の有病率が１％というパラメータが疾患データベース１９に記憶された疾患に対して、Ａ県Ｂ地区在住者は一般母集団に対して２倍リスクが高いことが同じく疾患データベース１９にパラメータとして記憶されている場合、プロファイルデータ取得部１２により取得された回答者の住所がＡ県Ｂ地区であった場合の検査前確率は２％であると算出できる。

[0028] なお、この場合における母集団の数は、疾患データベース１９に格納されている。一方、対象プロファイル保有者の数は、実績データベース１７に蓄積されている複数ユーザのプロファイルデータから得ることができる。罹患確率分析部１８は、実績データベース１７および疾患データベース１９を参照して、母集団に占める対象プロファイル保有者の割合が所定値以下か否かを判定し、所定値以下の場合には上述のような近似演算を行う。

[0029] このように、ユーザのプロファイルに関していくつかのリスクが該当する場合は、それらのリスク比を乗算して検査前確率とする。ここで、対象プロファイル保有者の数の母集団の数に対する割合が所定値以下の場合、複数のリスク比を乗算して算出される累積リスク比は、次の（式２）で表わすことができる。

[0030] [数1]

$$\prod_{n=0}^{N}(RR(X_n)) \cdots \text{（式２）}$$

[0031] 上記（式２）において、X_nは統計学的有意性をもってリスク比が明らかになっているプロファイル項目を示す。ここで、ｎは各プロファイル項目を識別する番号であり、Ｎはプロファイル項目の数を示す。$RR(X_n)$はその母集団に対するリスク比を表しており、その値は疾患データベース１９に格納されている。ただし、Ｎ＝０の場合（統計的有意性を持ってリスク比が明らかになっているプロファイルがない疾患の場合）、（式２）の値は１である。

[0032] 一方、性別や年齢のように、対象プロファイル保有者の数が母集団の数に比して充分に小さくない場合、（式２）の$RR(X_n)$の部分は、次の（式３）を使用する。ただし、（式３）において、POP_nは母集団に占める対象プロファイル保有者の割合とする。

$$RR(X_n) / (POP_n \cdot RR(X_n) + 1 - POP_n) \quad \cdots \text{（式３）}$$

[0033] 次に、上記（式１）における累積リスクの算出について説明する。累積リスクの算出は、問診回答データ取得部１３により取得された問診回答データから算出するパターンと、自由文解析部１５により生成されたテキスト関連データから算出するパターンとがある。

[0034] まず、問診回答データから累積リスクを算出する例について説明する。例えば、ある疾患に関して互いに相関がなく独立していると考えられる複数の既知のスクリーニング指標が疾患データベース１９にパラメータとして記憶されていて、そのスクリーニング指標に関する質問項目に対して陽性回答あるいは陰性回答がなされていることが問診回答データ取得部１３により取得

された問診回答データにより示されている場合は、疾患データベース１９にパラメータとして記憶されている既知であるところの感度・特異度をもとに、その疾患の累積リスクを算出することができる。

[0035] すなわち、prevを検査前確率、seを感度、spを特異度で表わすものとして、陽性反応的中率PPVが次の（式４）により表されるので、問診票に対して陽性回答をした質問項目のリスク比は、（式４）を検査前確率prevで除して算出することができる。

PPV＝prev・se／prev・se・(１－prev)(１－sp)・・・（式４）

[0036] 具体的には、罹患確率分析部１８は、（式２）または（式３）で算出された母集団中の疾患Ａの罹患確率を$prev_A$、感度・特異度が知られている質問項目X_nの感度をse(X_n)、特異度をsp(X_n)としたとき（ただし、ｎは各質問項目を識別する番号であり、Ｎは質問項目の数を示す）、このＮ個の質問項目X_nに陽性回答をした場合の、疾患が陽性である可能性を示す累積リスク比は、次の（式５）の通り表すことができる。

[0037] [数2]

$$\prod_{n=0}^{N}\left(\frac{se(X_n)}{prev_A \cdot se(X_n)+(1-prev_A)(1-sp(X_n))}\right) \cdots (式5)$$

[0038] 一方、このＮ個の質問項目X_nに陰性回答をした場合の、疾患が陽性である可能性を示す累積リスクは、次の（式６）の通り表すことができる。

[0039] [数3]

$$\prod_{n=0}^{N}\left(\frac{1-se(X_n)}{prev_A \cdot (1-se(X_n))+(1-prev_A) \cdot sp(X_n)}\right) \cdots (式6)$$

[0040] なお、ここでは問診票の質問項目X_nから累積リスクを算出する例について説明したが、回答者が事前に健康診断等の臨床検査を受けている場合、その臨床検査によって測定された値を問診票の回答として利用するようにしてもよい。この場合、（式５）および（式６）の質問項目X_nは、臨床検査の検査項目に置き換えればよい。また、（式５）および（式６）で求められる累積リスク比は、Ｎ個の検査項目X_nで陽性または陰性となった場合の、疾患が陽

性である可能性を示す累積リスク比を示すものとなる。なお、以下では説明の簡略化のため、問診票の質問項目は臨床検査の検査項目を含む概念であるものとする。

[0041] また、感度・特異度が知られているオープン質問に関する質問項目の形ではなく、回帰分析等によってリスク比が知られているクローズ質問に関する累積リスクは、質問項目Y_mのリスク比を$RR(Y_m)$としたとき（ただし、mは各質問項目を識別する番号であり、Mは質問項目の数を示す）、疾患が陽性である可能性を示す累積リスク比は、次の（式７）の通り表すことができる。

[0042] [数4]

$$\prod_{m=0}^{M}(RR(Y_m)) \quad \cdots \text{（式７）}$$

[0043] ただし、性別や年齢のように、回答者の数が母集団の数に比して充分に小さくない場合、（式７）の$RR(Y_m)$の部分は、次の（式８）を使用する。なお、（式８）において、POP_mは母集団に占める回答者の割合とする。

$$RR(Y_m)／(POP_m \cdot RR(Y_m)＋１－POP_m) \quad \cdots \text{（式８）}$$

[0044] 以上に示した（式５）～（式８）により、質問項目X_nについての陽性回答を$X_n(aff)$、陰性回答を$X_n(neg)$として、感度・特異度が既知である質問項目X_nと、リスク比が既知である質問項目Y_mとを用いた際の累積リスク比ΠＲは、次の（式９）のように表せる。ただし、Ｎ＝０、Ｍ＝０の場合、当該累積リスク比の値は１となる。

[0045] [数5]

$$\Pi R = \prod_{n=0}^{N}\left(\frac{se(X_{n(aff)})}{prev_A \cdot se(X_{n(aff)})+(1-prev_A)(1-sp(X_{n(aff)}))}\right) \cdot \prod_{n=0}^{N}\left(\frac{1-se(X_{n(neg)})}{prev_A \cdot (1-se(X_{n(neg)}))+(1-prev_A)\cdot sp(X_{n(neg)})}\right) \cdot \prod_{m=0}^{M}(RR(Y_m)) \quad \cdots \text{（式９）}$$

[0046] 次に、自由文解析部１５により生成されたテキスト関連データから累積リスク比を算出する例について説明する。罹患確率分析部１８は、重み値記憶

部２０を参照して、自由文解析部１５により生成されたテキスト関連データの中に、統計学的に有意にリスク比が算出されているテキスト関連データが含まれているか否かを判定する。含まれている場合は、そのテキスト関連データについても（式７）に使用して累積リスクを算出する。なお、統計学的に有意なリスク比は、後述するように重み値算出部２３により算出され、テキスト関連データと関連付けて重み値記憶部２０に記憶されている。

[0047] 疾患候補抽出部２１は、罹患確率分析部１８により求められた罹患確率Ｐが所定の閾値Ｐ(dif)以上の疾患を、疾患の検出を行っているユーザ（回答者）の疾患候補として抽出する。閾値Ｐ(dif)の値は疾患の重症度・緊急度によって個別に設定すべきであるが、例えばＰ(dif)＝７５％とし、罹患確率Ｐが７５％以上の疾患を「疾患陽性」の疾患候補として抽出する。

[0048] なお、疾患候補抽出部２１は、罹患確率分析部１８により求められた罹患確率Ｐが第１の閾値Ｐ(dif)（例えば、Ｐ(dif)＝７５％）以上の疾患を「疾患陽性」の疾患候補として抽出する一方、罹患確率Ｐが第１の閾値Ｐ(dif)より小さく第２の閾値Ｐ(sus)（例えば、Ｐ(sus)＝１０％）以上の疾患を「疾患疑い」の疾患候補として抽出するようにしてもよい。

[0049] 検出結果提示部２２は、疾患候補抽出部２１により抽出された疾患候補をユーザ端末１００に提示する。なお、疾患候補と共に、受診勧告や健康上の助言などの付帯情報を提示するようにしてもよい。この付帯情報は、例えば、疾患毎にあらかじめ疾患データベース１９に記憶しておく。「疾患陽性」と「疾患疑い」とに分けて疾患候補を抽出した場合、付帯情報の内容は、「疾患陽性」の場合と「疾患疑い」の場合とで異ならせるのが好ましい。この検出結果提示部２２による疾患候補の提示により、１つの疾患検出に関するセッションが終了する。

[0050] 重み値算出部２３は、セッションの終了後に、実績データベース１７に新規に格納されたデータを含めて、実績データベース１７に格納されたプロファイルデータ、問診回答データおよびテキスト関連データを説明変数とし、疾患候補抽出部２１により抽出された疾患候補を目的変数として回帰分析を

行う。そして、疾患候補抽出部２１により抽出された疾患候補と統計学的に有意に相関する（リスク要因となる）因子となる説明変数があった場合、その説明変数について重み値（リスク比）を算出し、当該重み値を重み値記憶部２０に記憶させる。または、その説明変数について既に重み値が重み値記憶部２０に記憶されていた場合は、重み値を更新記憶させる。

[0051] 例えば、疾患候補を虫垂炎とし、その虫垂炎の有無を目的変数、問診票に対する回答内容や、「お腹の右下が痛い」「熱がある」「昨日食べ過ぎた」という自由記述の内容、ユーザのプロファイルなどを説明変数として回帰分析をすると、例えば「お腹の右下が痛い」という語句の存在は虫垂炎のリスクを統計学的有意に上昇させる言葉であり、そのリスク比はたとえば１５倍である、などのように算出できる。

[0052] このようにして重み値記憶部２０に記憶された重み値（リスク比）は、次回セッションでの疾患確率の算出に用いられる。これにより、セッションを繰り返すごとに疾患候補の検出能力が自己進化する。すなわち、疾患検出の対象者ごとに以上の処理を繰り返すことで、疾患と結び付けられるデータ量と、その統計学的精度とを向上させることができる。

[0053] 図３は、上記のように構成した疾患検出サーバ２００の動作例を示すフローチャートである。図３（ａ）に示すフローチャートは、疾患の検出のためにユーザ端末１００から疾患検出サーバ２００にアクセスが行われ、セッションが確立されたときに開始する。

[0054] まず、プロファイルデータ取得部１２は、ユーザのプロファイルデータをユーザ端末１００から取得する（ステップＳ１）。次に、問診回答データ取得部１３は、問診票形式の入力画面をユーザ端末１００に提示し、問診票に対するユーザの回答である問診回答データを取得する（ステップＳ２）。また、自由回答データ取得部１４は、自由文記述形式の入力画面をユーザ端末１００に提示し、質問に対するユーザの回答である自由回答データを取得する（ステップＳ３）。

[0055] 自由文解析部１５は、自由回答データ取得部１４により取得された自由回

答データの自然文を解析することにより、構成単語、分かち書き、セグメント、単語間距離情報、文章係り受け情報の少なくとも１つに関するデータをテキスト関連データとして生成する（ステップＳ４）。そして、データベース記憶部１６は、ユーザのプロファイルデータ、問診回答データ、および、自由回答データを解析して生成したテキスト関連データを実績データベース１７に格納する（ステップＳ５）。

[0056] また、罹患確率分析部１８は、今回取得したユーザのプロファイルデータ、問診回答データ、テキスト関連データ、疾患データベース１９に疾患毎に記憶されている各種パラメータ、および、重み値記憶部２０に記憶されている所定の重み値（リスク比）を用いて、疾患データベース１９に登録されている疾患毎にユーザの罹患確率Ｐを分析する（ステップＳ６）。

[0057] 疾患候補抽出部２１は、罹患確率分析部１８により求められた罹患確率Ｐが第１の閾値Ｐ(dif)以上の疾患を「疾患陽性」の疾患候補として抽出するとともに、第１の閾値Ｐ(dif)より小さく第２の閾値Ｐ(sus)以上の疾患を「疾患疑い」の疾患候補として抽出する（ステップＳ７）。そして、検出結果提示部２２は、疾患候補抽出部２１により抽出された疾患候補をユーザ端末１００に提示し、セッションを終了する（ステップＳ８）。

[0058] 図３（ｂ）に示すフローチャートは、セッションの終了後に開始する。上述のように、ステップＳ１～Ｓ８の処理によって１つのセッションが終了した後、重み値算出部２３は、実績データベース１７に格納されたプロファイルデータ、問診回答データおよびテキスト関連データを説明変数とし、抽出された疾患候補を目的変数として回帰分析を行う。

[0059] そして、疾患候補抽出部２１により抽出された疾患候補と統計学的に有意に相関する因子となる説明変数について重み値（リスク比）を算出し（ステップＳ９）、それを重み値記憶部２０に更新記憶させる（ステップＳ１０）。これにより、図３に示すフローチャートの処理を終了する。

[0060] 以上詳しく説明したように、本実施形態では、ユーザのプロファイルデータと、問診票形式の質問に対するユーザからの回答である問診回答データと

、自由文記述形式の質問に対するユーザからの回答である自由回答データを解析して得られるテキスト関連データと、疾患の検出実績に基づく回帰分析によって得られた重み値（リスク比）とを用いて、ユーザが特定の疾患に罹患している可能性を表す罹患確率を分析し、当該罹患確率が閾値以上の疾患を疾患候補として抽出するようにしている。

[0061] また、本実施形態では、ユーザからの回答を得るたびに、上述のプロファイルデータ、問診回答データおよびテキスト関連データを実績データベース１７に随時格納して蓄積する。そして、今回の分析時に取得したデータも含めて実績データベース１７内のプロファイルデータ、問診回答データおよびテキスト関連データを説明変数とし、疾患候補抽出部２１により抽出された疾患候補を目的変数として回帰分析を行う。これにより、今回の分析で抽出された疾患候補と相関する因子となる説明変数について重み値を算出し、次回以降における罹患確率の分析に用いるようにしている。

[0062] このように構成した本実施形態によれば、問診回答データおよび自由回答データの両方を用いて、ユーザが特定の疾患に罹患している可能性を表す罹患確率が算出される。しかも、その罹患確率を算出する際には、過去の分析時に罹患確率が閾値以上となって疾患候補として抽出される原因となった因子に関して算出されたリスク比が罹患確率の算出に反映されているので、罹患確率の算出をより正確に行うことができる。このリスク比は、罹患確率の分析回数が増えるほど、統計学的に有意な値になっていくので、罹患確率の分析精度は上がっていく。これにより、医師の診断を受けていない一般ユーザに対しても罹患の疑いのある疾患をより高精度に検出することができる。

[0063] なお、上記実施形態において、以下のような機能をさらに追加してもよい。例えば、自由回答データ取得部１４は、疾患候補抽出部２１により「疾患陽性」の疾患候補が抽出された場合、当該疾患候補として抽出された疾患に関する自由文による詳細な回答を求める追加質問のデータをユーザ端末１００に提示し、追加質問に対する自由回答データを取得する。この場合、抽出された疾患の特徴や症状をユーザ端末１００に提示し、それについての自覚

症状や背景などの詳細な追加回答を求めるのが好ましい。

[0064] 自由文解析部１５は、自由回答データ取得部１４により取得された追加の自由回答データをさらに解析して、構成単語、分かち書き、セグメント、単語間距離情報、文章係り受け情報の少なくとも１つに関するデータをテキスト関連データとして生成する。そして、データベース記憶部１６は、自由文解析部１５により追加で生成されたテキスト関連データを実績データベース１７に追加して格納する。このような処理を、「疾患陽性」として抽出された疾患候補の数だけ繰り返す。

[0065] このような追加の機能を設けることにより、陽性とされた疾患に強く関連するテキスト関連データを実績データベース１７に格納することができる。その結果、重み値算出部２３による回帰分析によって、疾患との関連性が強いテキスト関連データについてリスク比を求めて重み値記憶部２０に記憶させ、次以降の罹患確率の分析に利用することができる。これにより、疾患確率の算出をより正確に行うことができるようになる。

[0066] また、以下のような機能をさらに追加してもよい。例えば、問診回答データ取得部１３は、疾患候補抽出部２１により「疾患疑い」の疾患候補が抽出された場合、当該疾患候補として抽出された疾患に関する詳細な回答を求める追加問診票のデータをユーザ端末１００に提示し、追加問診票に対する問診回答データを取得する。

[0067] この場合、罹患確率分析部１８は、問診回答データにより追加で取得された問診回答データを含めて、罹患確率の分析を再度行う。そして、疾患候補抽出部２１は、罹患確率分析部１８による再分析により求められた罹患確率が第１の閾値以上となった場合、「疾患疑い」として抽出されていた疾患を「疾患陽性」に変更する。このような処理を、「疾患疑い」として抽出された疾患候補の数だけ繰り返す。

[0068] このような追加の機能を設けることにより、「疾患陽性」となる疾患候補の見落としリスクを減じることができる。

[0069] また、上記実施形態では、罹患確率Ｐの算出にリスク比を用いる例につい

て説明したが、オッズ比（病気の人が病気の要因を、要因がない人と比べてどれだけ多く持っているかを示す指標）を用いるようにしてもよい。この場合、重み値算出部２３が算出する重み値もオッズ比となる。

[0070] また、上記実施形態では、ユーザ端末１００と疾患検出サーバ２００とをインターネット３００で接続する構成について説明したが、本発明はこれに限定されない。例えば、図４に示すように、産業医や企業の担当者が使用する担当者端末４００を更に備え、企業の従業員等がユーザ端末１００を用いて入力した回答に基づいて疾患検出サーバ２００で疾患の検出をした結果を、検出結果提示部２２が担当者端末４００に提示するようにしてもよい。

[0071] また、上記実施形態では、疾患検出サーバ２００の自由回答データ取得部１４が、自由文で回答可能な質問のデータをユーザ端末１００に提示し、質問に対するユーザの回答をテキスト文による自由回答データとして取得する例について説明したが、本発明はこれに限定されない。例えば、音声を用いた対話形式のやり取りによって自由回答データを取得するようにしてもよい。

[0072] 例えば、自由に回答可能な質問の音声データを疾患検出サーバ２００からユーザ端末１００に送信し、ユーザ端末１００のスピーカから質問を音声にて出力する。そして、質問に対してユーザが回答した話者音声をマイクから入力して音声データとして疾患検出サーバ２００に送信し、それを音声認識してテキスト文の自由回答データに変換するようにしてもよい。

[0073] この場合、例えば疾患検出サーバ２００からユーザ端末１００に対して「何がつらいのですか」などと音声にて質問をし、ユーザが「心配事が多くて夜眠れません」などと回答したら、疾患検出サーバ２００がその回答音声から「心配事」や「夜眠れない」といった構成単語を抽出して、これらから神経症、適応障害、うつ病などの疾患候補を抽出する。さらに、抽出した疾患候補に応じて、「それは大変ですね。何か生活上の悩みがあるんですか。気分は落ち込みますか。」などと次の追加質問を続けて行い、それに対する回答内容に応じて疾患候補の絞り込みを行うようにすることが可能である。ま

た、一連の対話を通じて入力されたデータを実績データベース１７に格納し、セッションの終了後に重み値算出部２３によって算出される重み値（リスク比）をより精度の高いものにすることも可能である。

[0074] その他、上記実施形態は、何れも本発明を実施するにあたっての具体化の一例を示したものに過ぎず、これによって本発明の技術的範囲が限定的に解釈されてはならないものである。すなわち、本発明はその要旨、またはその主要な特徴から逸脱することなく、様々な形で実施することができる。

符号の説明

[0075] １２　プロファイルデータ取得部

１３　問診回答データ取得部

１４　自由回答データ取得部

１５　自由文解析部

１６　データベース記憶部

１７　実績データベース

１８　罹患確率分析部

１９　疾患データベース

２０　重み値記憶部

２１　疾患候補抽出部

２２　検出結果提示部

２３　重み値算出部

１００　ユーザ端末

２００　疾患検出サーバ

請求の範囲

[請求項1] ユーザのプロファイルデータを取得するプロファイルデータ取得部と、

問診票のデータを上記ユーザに提示し、上記問診票に対する回答である問診回答データを取得する問診回答データ取得部と、

自由文で回答可能な質問のデータを上記ユーザに提示し、上記質問に対する回答である自由回答データを取得する自由回答データ取得部と、

上記自由回答データ取得部により取得された上記自由回答データを解析し、構成単語、分かち書き、セグメント、単語間距離情報、文章係り受け情報の少なくとも１つに関するデータをテキスト関連データとして生成する自由文解析部と、

上記プロファイルデータ取得部により取得された上記プロファイルデータ、上記問診回答データ取得部により取得された上記問診回答データ、および、上記自由文解析部により生成された上記テキスト関連データをデータベースに格納するデータベース記憶部と、

上記プロファイルデータ取得部により取得された上記プロファイルデータ、上記問診回答データ取得部により取得された上記問診回答データ、上記自由文解析部により生成された上記テキスト関連データ、および所定の重み値を用いて、上記ユーザが特定の疾患に罹患している可能性を表す罹患確率を分析する罹患確率分析部と、

上記罹患確率分析部により求められた上記罹患確率が閾値以上の疾患を疾患候補として抽出する疾患候補抽出部と、

上記データベースに格納された上記プロファイルデータ、上記問診回答データおよび上記テキスト関連データを説明変数とし、上記疾患候補抽出部により抽出された疾患候補を目的変数として回帰分析を行うことにより、上記疾患候補抽出部により抽出された疾患候補と相関する因子となる説明変数について上記重み値を算出し、重み値記憶部

に更新記憶させる重み値算出部と備えたことを特徴とする疾患検出システム。

[請求項2] 上記疾患候補抽出部は、上記罹患確率分析部により求められた上記罹患確率が第１の閾値以上の疾患を疾患陽性の疾患候補として抽出する一方、上記罹患確率が上記第１の閾値より小さく第２の閾値以上の疾患を疾患疑いの疾患候補として抽出することを特徴とする請求項１に記載の疾患検出システム。

[請求項3] 上記自由回答データ取得部は、上記疾患候補抽出部により上記疾患陽性の疾患候補が抽出された場合、当該疾患候補として抽出された疾患に関する自由文による詳細な回答を求める追加質問のデータを上記ユーザに提示し、上記追加質問に対する自由回答データを取得し、

上記自由文解析部は、上記自由回答データ取得部により取得された追加の上記自由回答データをさらに解析して上記テキスト関連データを生成し、

上記データベース記憶部は、上記自由文解析部により追加で生成された上記テキスト関連データを上記データベースにさらに格納する

ことを特徴とする請求項２に記載の疾患検出システム。

[請求項4] 上記問診回答データ取得部は、上記疾患候補抽出部により上記疾患疑いの疾患候補が抽出された場合、当該疾患候補として抽出された疾患に関する詳細な回答を求める追加問診票のデータを上記ユーザに提示し、上記追加問診票に対する問診回答データを取得し、

上記罹患確率分析部は、上記問診回答データにより追加で取得された問診回答データを含めて、上記罹患確率を再分析し、

上記疾患候補抽出部は、上記罹患確率分析部の再分析により求められた上記罹患確率が上記第１の閾値以上となった場合、上記疾患疑いとして抽出されていた疾患を上記疾患陽性に変更する

ことを特徴とする請求項２に記載の疾患検出システム。

[請求項5] 疾患検出システムのプロファイルデータ取得部が、ユーザのプロフ

ァイルデータを取得する第１のステップと、

上記疾患検出システムの問診回答データ取得部が、問診票のデータを上記ユーザに提示し、上記問診票に対する回答である問診回答データを取得する第２のステップと、

上記疾患検出システムの自由回答データ取得部が、自由文で回答可能な質問のデータを上記ユーザに提示し、上記質問に対する回答である自由回答データを取得する第３のステップと、

上記疾患検出システムの自由文解析部が、上記自由回答データ取得部により取得された上記自由回答データを解析し、構成単語、分かち書き、セグメント、単語間距離情報、文章係り受け情報の少なくとも１つに関するデータをテキスト関連データとして生成する第４のステップと、

上記疾患検出システムのデータベース記憶部が、上記プロファイルデータ取得部により取得された上記プロファイルデータ、上記問診回答データ取得部により取得された上記問診回答データ、および、上記自由文解析部により生成された上記テキスト関連データをデータベースに格納する第５のステップと、

上記疾患検出システムの罹患確率分析部が、上記プロファイルデータ取得部により取得された上記プロファイルデータ、上記問診回答データ取得部により取得された上記問診回答データ、上記自由文解析部により生成された上記テキスト関連データ、および所定の重み値を用いて、上記ユーザが特定の疾患に罹患している可能性を表す罹患確率を分析する第６のステップと、

上記疾患検出システムの疾患候補抽出部が、上記罹患確率分析部により求められた上記罹患確率が閾値以上の疾患を疾患候補として抽出する第７のステップと、

上記疾患検出システムの重み値算出部が、上記データベースに格納された上記プロファイルデータ、上記問診回答データおよび上記テキ

スト関連データを説明変数とし、上記疾患候補抽出部により抽出された疾患候補を目的変数として回帰分析を行うことにより、上記疾患候補抽出部により抽出された疾患候補と相関する因子となる説明変数について上記重み値を算出し、重み値記憶部に更新記憶させる第８のステップと有することを特徴とする疾患検出方法。