

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2017-151868

(P2017-151868A)

(43) 公開日 平成29年8月31日(2017.8.31)

(51) Int.Cl.	F I	テーマコード (参考)
<b>GO6N 99/00 (2010.01)</b>	GO6N 99/00 153	
<b>GO6F 17/30 (2006.01)</b>	GO6F 17/30 210D	
<b>GO6N 5/04 (2006.01)</b>	GO6N 5/04	

審査請求 有 請求項の数 16 O L (全 24 頁)

(21) 出願番号 特願2016-35605 (P2016-35605)  
 (22) 出願日 平成28年2月26日 (2016. 2. 26)  
 (11) 特許番号 特許第6041331号 (P6041331)  
 (45) 特許公報発行日 平成28年12月7日 (2016. 12. 7)

(71) 出願人 304020177  
 国立大学法人山口大学  
 山口県山口市吉田1677-1  
 (74) 代理人 100141173  
 弁理士 西村 啓一  
 (72) 発明者 浜本 義彦  
 山口県宇部市常盤台2丁目16-1 国立  
 大学法人山口大学工学部内  
 (72) 発明者 荻原 宏是  
 山口県宇部市常盤台2丁目16-1 国立  
 大学法人山口大学工学部内  
 (72) 発明者 飯塚 徳男  
 山口県宇部市南小串1丁目1-1 国立大  
 学法人山口大学医学部内

最終頁に続く

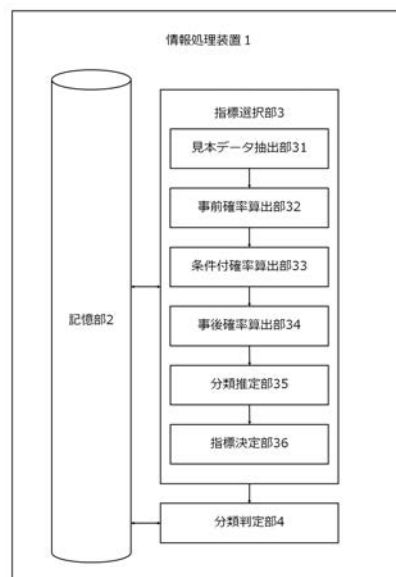
(54) 【発明の名称】 情報処理装置と情報処理プログラム並びに情報処理方法

(57) 【要約】

【課題】論理的に予測を支援することができる情報処理装置を得る。

【解決手段】入力データが、複数の分類の中の特定分類に属するか否かを、複数の指標の中から選択された選択指標に基づいて判定する情報処理装置(1)であって、複数の指標のそれぞれは、複数の区分を含み、複数の指標の中から選択指標を選択する指標選択部(3)と、入力データが該当する選択指標に含まれる指標ごとの区分に基づいて、入力データが特定分類に属するか否かを決定する分類判定部(4)と、を有してなり、選択指標の選択に用いられる複数の見本データごとの、見本データが該当する複数の指標ごとの区分を示す指標情報と、見本データが特定分類に属するか否かを示す分類情報と、が記憶される記憶部(2)、を備える、ことを特徴とする。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

入力データが、複数の分類の中の特定分類に属するか否かを、複数の指標の中から選択された選択指標に基づいて判定する情報処理装置であって、

前記複数の指標のそれぞれは、複数の区分を含み、

前記複数の指標の中から前記選択指標を選択する指標選択部と、

前記入力データが該当する前記選択指標に含まれる指標ごとの区分に基づいて、前記入力データが前記特定分類に属するか否かを決定する分類判定部と、  
を有してなり、

前記選択指標の選択に用いられる複数の見本データごとの、前記見本データが該当する前記複数の指標ごとの区分を示す指標情報と、前記見本データが前記特定分類に属するか否かを示す分類情報と、が記憶される記憶部、  
を備え、

前記指標選択部は、

(a) 前記複数の見本データのうち、一部の見本データを訓練データとして抽出し、他の一部の見本データを評価データとして抽出する、見本データ抽出部と、

(b) 前記複数の指標のいずれかの組合せで構成される指標群ごとに、

前記訓練データの前記指標情報と、前記訓練データの前記分類情報と、前記評価データの前記指標情報と、を用いて、前記評価データが前記特定分類に属するか否かを推定し、  
前記評価データの前記分類情報を用いて、前記推定の当否を判定する、  
分類推定部と、

(c) 前記指標群ごとの前記推定の当否の判定結果に基づいて、前記選択指標を決定する、指標決定部と、

を備える、

ことを特徴とする情報処理装置。

**【請求項 2】**

前記分類推定部は、

(b-1) 前記訓練データの前記指標情報と前記訓練データの前記分類情報とを用いて、前記特定分類が生起したときの前記指標群に含まれる指標の区分ごとの第 1 条件付確率と、前記非特定分類が生起したときの前記指標群に含まれる指標の区分ごとの第 2 条件付確率と、を算出し、

(b-2) 前記評価データの前記指標情報と、前記複数の分類のうち前記特定分類が生起する第 1 事前確率と、前記第 1 条件付確率と、に基づいて、前記評価データが前記特定分類に属する第 1 事後確率を算出し、

(b-3) 前記評価データの前記指標情報と、前記複数の分類のうち前記特定分類ではない非特定分類が生起する第 2 事前確率と、前記第 2 条件付確率と、に基づいて、前記評価データが前記非特定分類に属する第 2 事後確率を算出し、

(b-4) 前記第 1 事後確率と前記第 2 事後確率とを比較して、前記評価データが前記特定分類に属するか否かの推定の結果を示す推定情報を出し、

(b-5) 前記評価データの前記推定情報と前記評価データの前記分類情報とを比較して、前記推定の当否の判定結果を示す判定情報を出し、

前記指標決定部は、

(c-1) 前記指標群ごとの前記判定情報のうち、所定の条件を満たす判定情報に対応する指標群を特定し、

(c-2) 前記特定された指標群に含まれる指標を前記選択指標として決定する、

請求項 1 記載の情報処理装置。

**【請求項 3】**

前記分類推定部は、前記第 1 事前確率と、前記第 2 事前確率と、を算出する、  
請求項 2 記載の情報処理装置。

**【請求項 4】**

10

20

30

40

50

前記見本データ抽出部は、前記複数の見本データの中から複数の前記評価データを抽出し、

前記分類推定部は、前記複数の評価データごとに、前記判定情報を出し、

前記指標決定部は、前記複数の評価データごとの前記判定情報に基づいて、前記選択指標を決定する、

請求項 2 または 3 記載の情報処理装置。

【請求項 5】

前記分類推定部は、

前記複数の評価データごとに、前記第 1 条件付確率と前記第 2 条件付確率と前記第 1 事後確率と前記第 2 事後確率とを算出して、前記推定情報を出し、

前記複数の評価データごとの前記推定情報に基づいて、前記複数の評価データごとの前記判定情報を出し、

請求項 4 記載の情報処理装置。

【請求項 6】

前記第 1 事前確率は、前記第 2 事前確率と等しい、

請求項 2 乃至 5 のいずれかに記載の情報処理装置。

【請求項 7】

コンピュータを、請求項 1 乃至 6 のいずれかに記載の情報処理装置として機能させる、ことを特徴とする情報処理プログラム。

【請求項 8】

入力データが、複数の分類の中の特定分類に属するか否かを、複数の指標の中から選択された選択指標に基づいて判定する情報処理装置により実行される情報処理方法であって、

前記複数の指標のそれぞれは、複数の区分を含み、

前記情報処理装置は、

前記複数の指標の中から前記選択指標を選択する指標選択部と、

前記入力データが該当する前記選択指標に含まれる指標ごとの区分に基づいて、前記入力データが前記特定分類に属するか否かを決定する分類判定部と、

前記選択指標の選択に用いられる複数の見本データごとの、前記見本データが該当する前記複数の指標ごとの区分を示す指標情報と、前記見本データが前記特定分類に属するか否かを示す分類情報と、が記憶される記憶部と、

を備え、

前記情報処理装置は、

( a ) 前記複数の見本データのうち、一部の見本データを訓練データとして抽出し、他の一部の見本データを評価データとして抽出する、見本データ抽出ステップと、

( b ) 前記複数の指標のいずれかの組合せで構成される指標群ごとに、

前記訓練データの前記指標情報と、前記訓練データの前記分類情報と、前記評価データの前記指標情報と、を用いて、前記評価データが前記特定分類に属するか否かを推定し、

前記評価データの前記分類情報を用いて、前記推定の当否を判定する、

分類推定ステップと、

( c ) 前記指標群ごとの前記推定の当否の判定結果に基づいて、前記選択指標を決定する、指標決定ステップと、

を有してなる、

ことを特徴とする情報処理方法。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

## 【 0 0 0 1 】

本発明は、複数の事象の中から特定の事象が生じるか否かの予測を支援する情報処理装置と情報処理プログラム並びに情報処理方法に関するものである。

## 【 背景技術 】

## 【 0 0 0 2 】

臨床現場において、例えば、肝癌の難治性はその再発の可能性の高さにある。すなわち、たとえ全ての癌が手術により完全に摘出されたとしても、手術から1年後には約30%の確率で癌が再発する。癌が完全に摘出された手術後は、癌をCT (Computed Tomography) や超音波検査その他の検査で視認することができない。したがって、手術後の患者に対して、再発の可能性が不明のまま、癌の再発を防止するために、抗がん剤の投与や各種検査が実施される。そのため、患者の肉体的・精神的・経済的な負担は大きい。換言すれば、癌の再発を精度よく予測することができれば、患者の負担を軽減することができる。

10

## 【 0 0 0 3 】

これまでも、患者の検査データをもとに肝臓の状態をスコア式として表すことが提案がされている(例えば、非特許文献1, 2, 3, 4参照)。

## 【 先行技術文献 】

## 【 特許文献 】

## 【 0 0 0 4 】

【 非特許文献 1 】 TATEISHI R, Yoshida H, SHIINA S, et al: Proposal of a New Prognostic Model for Hepatocellular Carcinoma-an Analysis of 403 Patients. Gut 54: 419-425, 2005.

20

【 非特許文献 2 】 IKAI H, TAKAYASU K, OMATA M, et al: A modified Japan Integrated Stage Score for Prognostic Assessment in Patients with Hepatocellular Carcinoma. J Gastroenterology, 41, 884-892, 2006.

【 非特許文献 3 】 M. MINAGAWA, I. IKAI, Y. Matsuyama, Y. Yamaoka, M. MAKUUCHI, Staging of Hepatocellular Carcinoma Assessment of the Japanese TNM and AJCC/UICC TNM Systems in a Cohort of 13,772 Patients in Japan, Annals of Surgery, Vol. 245, No.6, pp. 909-922, June 2007.

【 非特許文献 4 】 JM HENDERSON, M. SHERMAN, A. TAVILL, M. ABECASSIS, G. CHEJFEC, and T. GRAMLICH, AHPBA/AJCC Consensus Conference on Staging of Hepatocellular Carcinoma: Consensus Statement, HPB, Vol. 5, No.4, pp. 243-250, 2003.

30

## 【 発明の概要 】

## 【 発明が解決しようとする課題 】

## 【 0 0 0 5 】

しかし、いずれのスコア式も臨床現場の要求に応えることができていない。その原因は、スコア式で用いられる個々のマーカーが再発予測の決め手ではないだけでなく、スコア式で用いられるマーカーの組も医師による試行錯誤の末に得られたものであって、その最適性の論理的な保証はない。また、スコア式で用いられるマーカーは、スコア式ごとに予め決まっている。そのため、スコア式で用いられるマーカーのデータに欠損があれば、そのスコア式を用いることができない。

40

## 【 0 0 0 6 】

不完全なスコア式の予測の性能を向上させる方法として、最先端の分子生命科学によって発見された癌と関連のある遺伝子変異などの分子標的マーカーを用いることが考えられる。しかし、これらの分子標的マーカーの多くは、いわゆる保険適用の対象ではない。保険適用には、治験による薬事法の承認が必要である。そのため、分子標的マーカーが患者に適用されるまでには、膨大な研究費と相当の時間とを要する。

## 【 0 0 0 7 】

このように、臨床現場においては、ある疾患の手術(治療)を受けた患者に同疾患が再発するか否かを予測する場合に、予測に必要なマーカーを論理的に特定することや、すでに検査済の患者の検査データから論理的に予測することが望まれている。

50

## 【 0 0 0 8 】

本発明は、以上のような従来技術の問題点を解消するためになされたもので、客観的なデータに基づいて論理的に特定の事象の発生の予測を支援することができる情報処理装置と情報処理プログラム並びに情報処理方法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】

## 【 0 0 0 9 】

本発明は、入力データが、複数の分類の中の特定分類に属するか否かを、複数の指標の中から選択された選択指標に基づいて判定する情報処理装置であって、複数の指標のそれぞれは、複数の区分を含み、複数の指標の中から選択指標を選択する指標選択部と、入力データが該当する選択指標に含まれる指標ごとの区分に基づいて、入力データが特定分類に属するか否かを決定する分類判定部と、を有してなり、選択指標の選択に用いられる複数の見本データごとの、見本データが該当する複数の指標ごとの区分を示す指標情報と、見本データが特定分類に属するか否かを示す分類情報と、が記憶される記憶部、を備え、指標選択部は、(a) 複数の見本データのうち、一部の見本データを訓練データとして抽出し、他の一部の見本データを評価データとして抽出する、見本データ抽出部と、(b) 複数の指標のいずれかの組合せで構成される指標群ごとに、訓練データの指標情報と、訓練データの分類情報と、評価データの指標情報と、を用いて、評価データが特定分類に属するか否かを推定し、評価データの分類情報を用いて、推定の可否を判定する、分類推定部と、(c) 指標群ごとの推定の可否の判定結果に基づいて、選択指標を決定する、指標決定部と、を備える、ことを特徴とする。

10

20

## 【発明の効果】

## 【 0 0 1 0 】

本発明によれば、客観的なデータに基づいて論理的に特定の事象の発生の予測を支援することができる。

## 【図面の簡単な説明】

## 【 0 0 1 1 】

【図1】本発明にかかる情報処理装置の実施の形態を示すブロック図である。

【図2】図1の情報処理装置に記憶される見本データの例を示す模式図である。

【図3】図1の情報処理装置に記憶される見本データの分類情報の例を示す模式図である。

30

【図4】図1の情報処理装置に記憶される見本データの指標情報の例を示す模式図である。

【図5】図1の情報処理装置に記憶される見本データから抽出された訓練データの指標ごとの区分に属するデータ数の例を示す模式図である。

【図6】本発明にかかる情報処理方法の例を示すフローチャートである。

【図7】図6の情報処理方法における指標の選択処理の例を示すフローチャートである。

【図8】図1の情報処理装置に記憶される見本データから抽出された評価データの指標ごとに該当する区分と分類情報と推定情報と判定情報との関係を示す模式図である。

【図9】図6の情報処理方法における入力データの分類判定処理の例を示すフローチャートである。

40

【図10】図1の情報処理装置が決定した選択指標が表示された画面の例を示す模式図である。

【図11】図1の情報処理装置が決定した選択指標が表示された別の画面の例を示す模式図である。

【図12】図1の情報処理装置が決定した選択指標が表示された別の画面の例を示す模式図である。

## 【発明を実施するための形態】

## 【 0 0 1 2 】

以下、図面を参照しながら、本発明にかかる情報処理装置と情報処理プログラム並びに情報処理方法の実施の形態について説明する。

50

## 【0013】

本発明は、過去の実績を示す見本データから定まる評価値を用いて、数学的基盤の明確な統計的パターン認識理論に基づき、入力データで示される予測の対象において特定の事象が生じるか否かの予測を支援するものである。

## 【0014】

以下に説明する実施の形態は、予測の対象となる肝癌の患者（以下「対象患者」という。）の検査データと、対象患者とは異なる別の患者であって肝癌の摘出手術（以下「手術」という。）を受けた患者（以下「見本患者」という。）の検査データおよび手術後の経過と、を用いて、手術を受けた対象患者の肝癌の再発の有無の予測を支援する場合を例にして、本発明の内容を説明する。

10

## 【0015】

本発明にかかる情報処理装置は、後述のとおり、見本患者の見本データと、対象患者の入力データと、を用いて、対象患者の肝癌の再発の不確実性を事後確率として数値で表す。本発明にかかる情報処理装置は、この事後確率を用いる統計的パターン認識のBayes識別則により、予測の信頼性を向上させている。

## 【0016】

また、本発明にかかる情報処理装置は、後述のとおり、見本データを用いて、複数の検査項目（マーカー）の中から肝癌の再発の予測に有効な検査項目の選択の不確実性を事後確率として数値で表す。検査項目は、本発明における指標の例である。本発明にかかる情報処理装置は、この事後確率を用いる統計的パターン認識のBayes識別則により、検査項目の選択の信頼性を向上させている。

20

## 【0017】

## 情報処理装置

図1は、本発明にかかる情報処理装置（以下「本装置」という。）の実施の形態を示すブロック図である。本装置1は、記憶部2と、指標選択部3と、分類判定部4と、を有してなる。指標選択部3は、見本データ抽出部31と、事前確率算出部32と、条件付確率算出部33と、事後確率算出部34と、分類推定部35と、指標決定部36と、を含む。

## 【0018】

本装置1は、パーソナルコンピュータなどで実現される。本装置1では、本発明にかかる情報処理プログラム（以下「本プログラム」という。）が動作して、本プログラムが本装置1のハードウェア資源と共働して、後述する本発明にかかる情報処理方法（以下「本方法」という。）を実現する。

30

## 【0019】

なお、図示しないコンピュータに本プログラムを実行させることで、同コンピュータを本装置1と同様に機能させて、同コンピュータに本方法を実行させることができる。

## 【0020】

記憶部2は、見本データや入力データなど、本装置1が後述する本方法を実行するために必要な情報が記憶される手段である。

## 【0021】

見本データは、見本患者の検査情報と分類情報とを含む。

40

## 【0022】

検査情報は、検査項目ごとの検査結果を示す情報である。検査項目は、定性的なものや定量的なものがある。定性的な検査項目は、例えば、性別、リンパ管侵襲の有無、潰瘍の有無、などである。定量的な検査項目は、例えば、腫瘍数、腫瘍サイズ、ビリルビン値、アルブミン値、などである。本発明は、後述する検査項目ごとの区分を用いて、検査情報のすべてを離散データで統一する。すなわち、本発明におけるBayes識別則は、通常のBayes識別則とは異なり、離散データを取り扱うことができる。

## 【0023】

分類情報は、特定の事象が生じたか否かを示す情報である。

ここで、特定の事象は、「手術後1年以内に肝癌が再発した」ことをいう。すなわち、

50

分類情報は、「手術後 1 年以内に肝癌が再発したか否かを示す情報」である。

【 0 0 2 4 】

入力データは、対象患者の検査情報を含む。

【 0 0 2 5 】

図 2 は、記憶部 2 に記憶される見本データの例を示す模式図である。

記憶部 2 には、複数の見本データが記憶される。各見本データは、見本識別子（見本 ID）と関連付けて記憶される。すなわち、本装置 1 は、見本 ID を用いて、見本 ID に対応する見本データを、記憶部 2 から読み出すことができる。見本データに含まれる検査情報は、検査項目を特定する指標識別子（指標 ID）と関連付けられて、記憶部 2 に記憶される。同図は、例えば、見本 ID「 $D_1$ 」の見本データにおいて、指標 ID「 $X_1$ 」の検査データは「 $V_{11}$ 」、指標 ID「 $X_2$ 」の検査データは「 $V_{21}$ 」であることを示す。

10

【 0 0 2 6 】

図 3 は、分類情報の例を示す模式図である。

分類情報「 $1$ 」は、「手術後 1 年以内に肝癌が再発したことを示す情報」である。分類情報「 $2$ 」は、「手術後 1 年以内に肝癌が再発しなかったことを示す情報」である。同図は、例えば、見本 ID「 $D_1$ 」の見本患者の分類情報が「 $2$ 」、すなわち、同患者は手術後 1 年以内に肝癌を再発しなかった、ことを示す。各見本患者は、「手術後 1 年以内に肝癌が再発した」分類（以下「特定分類」という。）と、「手術後 1 年以内に肝癌が再発しなかった」分類（以下「非特定分類」という。）のいずれかの分類に属する。

20

【 0 0 2 7 】

指標選択部 3 は、複数の検査項目の中から肝癌の予測に適した検査項目を選択する手段である。検査項目の選択方法については、後述する。

【 0 0 2 8 】

分類判定部 4 は、対象患者の肝癌の再発を予測する手段である。肝癌の再発の予測方法については、後述する。

【 0 0 2 9 】

見本データ抽出部 3 1 は、見本データのうち、一部の見本データを訓練データとして抽出し、他の一部の見本データを評価データとして抽出する手段である。

【 0 0 3 0 】

事前確率算出部 3 2 は、特定分類が生起する確率（以下「第 1 事前確率」という。）と、非特定分類が生起する確率（以下「第 2 事前確率」という。）と、を算出する手段である。

30

【 0 0 3 1 】

条件付確率算出部 3 3 は、第 1 条件付確率と第 2 条件付確率とを算出する手段である。

【 0 0 3 2 】

第 1 条件付確率は、訓練データの指標情報と分類情報とを用いて、指標群に含まれる指標の区分ごとに算出される、特定分類が生起したときに指標情報が区分に該当する条件付確率である。指標情報と指標群とについては、後述する。

【 0 0 3 3 】

第 2 条件付確率は、訓練データの指標情報と分類情報とを用いて、指標群に含まれる指標の区分ごとに算出される、非特定分類が生起したときに指標情報が区分に該当する条件付確率である。

40

【 0 0 3 4 】

指標情報は、見本データ（訓練データ、評価データ）ごとの情報であって、見本データに含まれる検査情報が該当する「検査項目ごとの区分」を示す情報である。

【 0 0 3 5 】

検査項目ごとの区分は、検査項目で取り得る値を複数に分割したときのそれぞれを特定する情報である。検査項目の区分は、その検査項目で取り得る値が互いに排反事象となるように分割する。すなわち、患者の検査データは、検査項目ごとにいずれかの区分に属する。検査項目が性別であれば、2 つの区分「男性」「女性」に分割される。検査項目が腫

50

瘍数であれば、例えば、3つの区分「2個未満」「2個以上7個未満」「7個以上」に分割される。

【0036】

指標群は、複数の検査項目の中から選択された1または複数の検査項目の組合せである。

【0037】

図4は、記憶部2に記憶される訓練データの指標情報の例を示す模式図である。

同図は、指標ID「 $X_1$ 」の検査項目が2つの区分に分割されていて、各区分を特定する区分識別子(区分ID)が「 $X_{1(1)}$ 」「 $X_{1(2)}$ 」であることを示す。同図は、指標ID「 $X_2$ 」の検査項目が3つの区分に分割されていて、各区分を特定する区分IDが「 $X_{2(1)}$ 」「 $X_{2(2)}$ 」「 $X_{2(3)}$ 」であることを示す。同図は、見本ID「 $D_{t1}$ 」の訓練データは、指標ID「 $X_1$ 」の検査項目において区分ID「 $X_{1(2)}$ 」の区分に属し、指標ID「 $X_2$ 」の検査項目において区分ID「 $X_{2(2)}$ 」の区分に属することを示す。

10

【0038】

検査項目ごとの区分を特定する情報、例えば、検査項目が性別における区分「男性」「女性」を特定する情報や、検査項目が腫瘍数における区分「3個未満」「3個以上7個未満」「7個以上」を特定する情報は、記憶部2に予め記憶されている。

【0039】

本装置1は、記憶部2に記憶されている検査項目ごとの区分を特定する情報を参照して、各訓練データの検査情報から訓練データの指標情報を生成して記憶部2に記憶する。

20

【0040】

図5は、記憶部2に記憶される検査項目ごとの区分に属する訓練データの数(訓練データに対応する見本患者(以下「訓練患者」という。)の人数)の例を示す模式図である。同図は、指標ごとの区分のそれぞれに属する訓練患者の人数が、分類ごとに記憶されていることを示す。

【0041】

同図は、分類ID「 $C_1$ 」の分類に属する訓練患者のうち、指標ID「 $X_1$ 」の検査項目の区分ID「 $X_{1(1)}$ 」の区分に属する訓練患者の人数は「 $n^1_{1(1)}$ 」人、区分ID「 $X_{1(2)}$ 」の区分に属する訓練患者の人数は「 $n^1_{1(2)}$ 」人であることを示す。同図は、分類ID「 $C_1$ 」の分類に属する訓練患者のうち、指標ID「 $X_2$ 」の検査項目の区分ID「 $X_{2(1)}$ 」の区分に属する訓練患者の人数は「 $n^1_{2(1)}$ 」人、区分ID「 $X_{2(2)}$ 」の区分に属する訓練患者の人数は「 $n^1_{2(2)}$ 」人、区分ID「 $X_{2(3)}$ 」の区分に属する訓練患者の人数は「 $n^1_{2(3)}$ 」人、であることを示す。

30

【0042】

ここで、分類ID「 $C_1$ 」の分類に属する検査項目ごとの訓練患者の人数は、検査項目に関わらず一定である。すなわち、以下の関係が成り立つ。

$$n^1_{1(1)} + n^1_{1(2)} = n^1_{2(1)} + n^1_{2(2)} + n^1_{2(3)}$$

【0043】

本装置1は、訓練データの指標情報と分類情報とを用いて、検査項目ごとの各区分に属する訓練患者の人数を、分類ごとに集計して記憶部2に記憶する。

40

【0044】

本装置1は、分類ごとに集計された、指標ごとの区分のそれぞれに属する訓練患者の人数を用いて、各検査項目の区分ごとの条件付確率(第1条件付確率と第2条件付確率)を算出する。

【0045】



ここで、検査項目 $x_{t_j}$ の $r_{t_j}$ 番目の区分を $x_{t_j(r_{t_j})}$ と表わす。

添字 $t_j$ は、検査項目の識別番号(指標 I D)であり、 $t_j \in \{1, 2, \dots, M\}$ である。

このとき患者は、パターン $\mathbf{X} = [x_{t_1(r_{t_1})}, x_{t_2(r_{t_2})}, \dots, x_{t_d(r_{t_d})}]^T$ と表わされる。

$A^T$ は、行列 $A$ の転置を表す。

区分 $x_{t_j(r_{t_j})}$ の条件付確率 $P(x_{t_j(r_{t_j})} | \omega_i)$ は、式 1 のとおり定義される。

【 0 0 4 6 】

10

(式 1)

$$P(x_{t_j(r_{t_j})} | \omega_i) = \frac{n_{t_j(r_{t_j})}^i}{\sum_{k=1}^d n_{t_k(r_{t_k})}^i} \quad j = 1, 2, \dots, d$$

【 0 0 4 7 】

ここで、 $n_{t_j(r_{t_j})}^i$  は分類 $\omega_i$ の患者 $n^i$ 人の中で、区分 $x_{t_j(r_{t_j})}$ に属する患者の人数を表す。

20

【 0 0 4 8 】

一般に、検査項目に対する検査結果が複数の区分のいずれかに属するという事象が互いに独立であるとき、以下の式 2 が成り立つ。

【 0 0 4 9 】

(式 2)

$$\begin{aligned} P(\mathbf{X} | \omega_i) &= P(x_{t_1(r_{t_1})}, x_{t_2(r_{t_2})}, \dots, x_{t_d(r_{t_d})} | \omega_i) \\ &= P(x_{t_1(r_{t_1})} | \omega_i) P(x_{t_2(r_{t_2})} | \omega_i) \dots P(x_{t_d(r_{t_d})} | \omega_i) \\ &= \prod_{k=1}^d P(x_{t_k(r_{t_k})} | \omega_i) \end{aligned}$$

30

【 0 0 5 0 】

事象が 2 つの分類  $\omega_1$  と  $\omega_2$  のいずれかに属するという 2 クラス問題において、事後確率  $P(\omega_i | \mathbf{X})$  は、Bayes の定理により、式 3 となる。

【 0 0 5 1 】

(式 3)

$$P(\omega_i | \mathbf{X}) = \frac{P(\omega_i)P(\mathbf{X} | \omega_i)}{P(\omega_1)P(\mathbf{X} | \omega_1) + P(\omega_2)P(\mathbf{X} | \omega_2)}$$

40

【 0 0 5 2 】

事前確率  $P(\omega_i)$  が等確率の 0.5 とすると、式 3 に示した事後確率  $P(\omega_i | \mathbf{X})$  は、式 4 となる。

【 0 0 5 3 】

(式 4)

$$P(\omega_i | \mathbf{x}) = \frac{P(\mathbf{x} | \omega_i)}{P(\mathbf{x} | \omega_1) + P(\mathbf{x} | \omega_2)}$$

【 0 0 5 4 】

式 4 に式 2 を代入すると、事後確率  $P(\omega_i | \mathbf{x})$  は式 5 で算出される。

【 0 0 5 5 】

( 式 5 )

$$P(\omega_i | \mathbf{x}) = \frac{\prod_{k=1}^d P(x_{t_k(r_{t_k})} | \omega_i)}{\prod_{k=1}^d P(x_{t_k(r_{t_k})} | \omega_1) + \prod_{s=1}^d P(x_{t_s(r_{t_s})} | \omega_2)} \quad 10$$

【 0 0 5 6 】

ここで、マーカーが  $X_1$  と  $X_2$  のとき、つまり、 $d = 2$  のときに、訓練患者の検査データが  $X_{1(1)}$  と  $X_{2(3)}$  に属したとする。このとき、式 6 が成り立つ。

【 0 0 5 7 】

( 式 6 )

$$P(x_{1(1)} | \omega_1) = \frac{n_{1(1)}^1}{n_{1(1)}^1 + n_{2(3)}^1} \quad 20$$

$$P(x_{2(3)} | \omega_1) = \frac{n_{2(3)}^1}{n_{1(1)}^1 + n_{2(3)}^1}$$

【 0 0 5 8 】

よって、式 5 中の  $P(X_{1(1)}, X_{2(3)} | \omega_1)$  は、式 7 で算出される。

【 0 0 5 9 】

( 式 7 )

$$P(x_{1(1)}, x_{2(3)} | \omega_1) = P(x_{1(1)} | \omega_1) P(x_{2(3)} | \omega_1) \quad 30$$

【 0 0 6 0 】

同様にして、 $P(X_{1(1)}, X_{2(3)} | \omega_2)$  を算出した上で、式 5 により、分類  $\omega_1$  と  $\omega_2$  の事後確率を算出する。算出された分類ごとの事後確率を比較して、事後確率が高い分類に同患者を識別する。

【 0 0 6 1 】

情報処理方法

次に、本装置 1 が実行する本方法の実施の形態について説明する。

【 0 0 6 2 】

図 6 は、本方法の実施の形態を示すフローチャートである。

40

まず、本装置 1 は、指標選択部 3 を用いて、複数の検査項目の中から、肝癌の再発の予測に適した検査項目を、選択指標として選択する ( S 1 ) 。

次いで、本装置 1 は、分類判定部 4 を用いて、選択指標を用いて入力データの分類を判定、つまり、入力データに対応する対象患者の手術後 1 年以内の肝癌の再発の有無を予測する ( S 2 ) 。

【 0 0 6 3 】

指標の選択

図 7 は、本装置 1 が指標選択部 3 を用いて実行する指標の選択処理 ( S 1 ) の例を示すフローチャートである。

【 0 0 6 4 】

50

まず、本装置 1 は、見本データ抽出部 3 1 を用いて、記憶部 2 に記憶されている見本データを取得し ( S 1 1 )、取得した見本データのうち、一部の見本データを訓練データとして抽出し、他の一部の見本データを評価データとして抽出する ( S 1 2 )。本装置 1 は、例えば、取得した 1 0 0 件の見本データのうち、8 0 件の見本データを訓練データとして抽出し、残り 2 0 件の見本データを評価データとして抽出する。本装置 1 は、見本データの内容 ( 検査結果 ) を参照することなく、ランダムに見本データから訓練データと評価データとを抽出する。本装置 1 により抽出される訓練データと評価データの数や、訓練データと評価データの数の比率は、予め、記憶部 2 に記憶されていて、本装置 1 は、記憶部 2 に記憶されているこれらの数や比率に基づいて、見本データを抽出する。

【 0 0 6 5 】

次いで、本装置 1 は、複数の検査項目のいずれかの組合せで構成される指標群を特定する ( S 1 3 )。複数の検査項目が、例えば、 $X_1$ 、 $X_2$ 、 $X_3$  の 3 つのとき、本装置 1 により特定される指標群は、 $(X_1)$ 、 $(X_2)$ 、 $(X_3)$ 、 $(X_1, X_2)$ 、 $(X_1, X_3)$ 、 $(X_2, X_3)$ 、 $(X_1, X_2, X_3)$  の 7 つである。

【 0 0 6 6 】

なお、本装置 1 により特定される指標群に含まれる指標の数が予め決められていて、この指標の数は本装置 1 が指標群を特定する際に参照できるように記憶部 2 に記憶されていてもよい。すなわち、例えば、指標群に含まれる指標の数が「2」であり、見本データに含まれる指標が  $X_1$ 、 $X_2$ 、 $X_3$  の 3 つのとき、本装置 1 は、指標群として  $(X_1, X_2)$  と  $(X_1, X_3)$  の 2 つを特定する。

【 0 0 6 7 】

次いで、本装置 1 は、条件付確率算出部 3 3 を用いて、特定された指標群ごとに、指標群に含まれる検査項目の区分ごとの条件付確率 ( 第 1 条件付確率と第 2 条件付確率 ) を、訓練データに基づいて式 1 により算出する ( S 1 4 )。

【 0 0 6 8 】

次いで、本装置 1 は、評価データごとに、事後確率算出部 3 4 を用いて、分類ごとの事後確率を、式 5 により算出する。

【 0 0 6 9 】

なお、本実施の形態は、特定分類と非特定分類とが生起する確率は等確率 ( 0 . 5 ) としているため、本装置 1 は、事前確率 ( 第 1 事前確率と第 2 事前確率 ) を算出することなく、事後確率を式 5 で算出する。

【 0 0 7 0 】

次いで、本装置 1 は、分類推定部 3 5 を用いて、評価データの分類を、算出された分類ごとの事後確率に基づいて推定して ( S 1 5 )、推定の当否を判定する ( S 1 6 )。

【 0 0 7 1 】

本装置 1 は、例えば、算出された分類ごとの事後確率の大小を比較して、事後確率の大きい分類を評価データの分類として推定する。

【 0 0 7 2 】

本装置 1 は、推定された分類と、記憶部 2 に見本データとして記憶されている評価データの分類情報と、を比較して、両者が一致するか否かで、評価データの分類の推定の当否を判定する。すなわち、両者が一致すれば評価データの分類の推定は妥当と判定され、両者が不一致であれば評価データの分類の推定は不当と判定される。

【 0 0 7 3 】

図 8 は、特定された指標群に含まれる検査項目が「 $X_1$ 、 $X_2$ 、 $\dots$ 」であり、見本 ID「 $D_{e_1}$ 、 $D_{e_2}$ 、 $\dots$ 」の各評価データが属する検査項目ごとの区分と、各評価データの分類情報と、本装置 1 が推定した各評価データの分類 ( 推定情報 ) と、本装置 1 が判定した推定の当否の判定 ( 判定情報 ) と、の関係を示す模式図である。

【 0 0 7 4 】

同図は、見本 ID「 $D_{e_1}$ 」の評価データの分類情報が分類 ID「 $C_1$ 」の分類で、推定情報が分類 ID「 $C_1$ 」の分類で、判定情報が妥当であることを示す「T」であること

10

20

30

40

50

を示す。すなわち、本装置 1 は、見本 I D 「 D<sub>e1</sub> 」 の評価データに対応する患者が手術後 1 年以内に肝癌の再発があったことを言い当てている。

【 0 0 7 5 】

一方、同図は、見本 I D 「 D<sub>e2</sub> 」 の評価データの分類情報が分類 I D 「<sub>2</sub>」の分類で、推定情報が分類 I D 「<sub>1</sub>」の分類で、判定情報が不当であることを示す「 F 」であることを示す。すなわち、本装置 1 は、見本 I D 「 D<sub>e2</sub> 」 の評価データに対応する患者が手術後 1 年以内に肝癌の再発が無かったことを言い当てていない。

【 0 0 7 6 】

本装置 1 は、すべての指標群に対して、処理 S 1 4 から S 1 6 を実行する ( S 1 7 ) 。

【 0 0 7 7 】

次いで、本装置 1 は、指標決定部 3 6 を用いて、指標群ごとの推定の当否の判定結果を比較して ( S 1 8 ) 、選択指標を決定する ( S 1 9 ) 。

【 0 0 7 8 】

本装置 1 は、例えば、指標群ごとに評価データの判定情報を集計して、推定が妥当と判定された評価データの数が最大の指標群に含まれる検査項目を、選択指標として決定する。

【 0 0 7 9 】

入力データの分類判定

図 9 は、本装置 1 が分類判定部 4 を用いて実行する入力データの分類判定処理 ( S 2 ) の例を示すフローチャートである。

【 0 0 8 0 】

まず、本装置 1 は、入力データの指標情報を取得する ( S 2 1 ) 。入力データの指標情報は、前述の見本データの指標情報と同様に、入力データに含まれる検査情報が該当する「検査項目ごとの区分」を示す情報である。本装置 1 は、入力データと、選択指標に対応する検査項目ごとの区分を特定する情報と、を記憶部 2 から読み出す。本装置 1 は、選択指標に対応する検査項目ごとに、入力データの検査情報が該当する区分を特定して、入力データの指標情報を取得する。

【 0 0 8 1 】

次いで、本装置 1 は、見本データを用いて決定された選択指標に含まれる指標の区分ごとの条件付確率 ( 処理 S 1 4 と同様に算出 ) を用いて、入力データの分類ごとの事後確率を、式 5 により算出する ( S 2 2 ) 。

【 0 0 8 2 】

次いで、本装置 1 は、算出された分類ごとの事後確率を比較する ( S 2 3 ) 。

【 0 0 8 3 】

次いで、本装置 1 は、入力データの分類を特定、すなわち、入力データの分類を特定分類または非特定分類のいずれかに識別する ( S 2 4 ) 。

【 0 0 8 4 】

本装置 1 は、例えば、分類ごとの事後確率の大小を比較して、最大の事後確率に対応する分類を、入力データの分類として識別する。

【 0 0 8 5 】

特定された入力データの分類は、入力データと関連付けて記憶部 2 に記憶される。特定された入力データの分類は、本装置 1 のディスプレイ ( 不図示 ) などに、例えば、「患者さんは、手術後 1 年以内に肝癌を再発する可能性は低い」などと表示して出力される。

【 0 0 8 6 】

図 1 0 は、本装置 1 が決定した選択指標が表示された画面の例を示す模式図である。同図は、検査項目 X<sub>1</sub>、X<sub>2</sub>、・・・、X<sub>n</sub> の中から選択指標として選択された検査項目が X<sub>1</sub> と X<sub>2</sub> であることを示す。

【 0 0 8 7 】

図 1 1 は、本装置 1 が決定した選択指標が表示された別の画面の例を示す模式図である

10

20

30

40

50

。同図は、検査項目  $Y_1$ 、 $Y_2$ 、 $Y_3$ 、 $\dots$ 、 $Y_n$  の中から選択された検査項目の組合せと、同組合せの事後確率とが、事後確率の降順に表示されていることを示す。本装置 1 は、例えば、本装置 1 の使用者にこの画面を閲覧させることで、検査項目の組合せの違いを事後確率の大小で確認させることができる。

【0088】

図 12 は、本装置 1 が決定した選択指標が表示されたさらに別の画面の例を示す模式図である。同図は、検査項目  $Z_1$  と  $Z_2$  の検査を受けているある患者が、検査項目  $Z_3$  の検査を受けることで、ある疾患の再発の予測精度が高まることを示す。本装置 1 は、例えば、図 11 に示したような検査項目の組合せと事後確率との対応関係と、ある患者の検査済の検査項目の組合せよりも事後確率が大きい検査項目の組合せを特定して、特定された検査項目の組合せに含まれる検査項目のうち患者の検査済の検査項目以外の検査項目を抽出することができる。

10

【0089】

まとめ

以上説明した実施の形態によれば、肝癌の再発の予測に用いる検査項目の適否を、事後確率という数値で比較することができる。その結果、本発明によれば、数値に基づいて論理的に肝癌の再発の予測を支援することができる。

【0090】

なお、以上説明した実施の形態は、特定分類と非特定分類の 2 つの分類を対象とするものであったが、本発明は 3 つ以上の分類へも適用可能である。

20

【0091】

分類  $1, 2, \dots, m$  ( $m \geq 3$ ) のそれぞれが生起する事象が互いに排反事象で、かつ、それらの和集合が全集合であるとき、入力データは、 $m$  個の分類のうちのいずれか一の分類に属する。このとき、マーカー  $X_{t_1}(r_{t_1}), X_{t_2}(r_{t_2}), \dots, X_{t_d}(r_{t_d})$  が用いられると、本装置は、分類ごとの条件付確率  $P(X_{t_1}(r_{t_1}), X_{t_2}(r_{t_2}), \dots, X_{t_d}(r_{t_d}) | i)$  ( $i = 1, 2, \dots, m$ ) を算出した上で、事後確率  $P(i | X_{t_1}(r_{t_1}), X_{t_2}(r_{t_2}), \dots, X_{t_d}(r_{t_d}))$  を式 8 により算出する。

【0092】

(式 8)

30

$$P(\omega_i | x_{t_1}(r_{t_1}), x_{t_2}(r_{t_2}), \dots, x_{t_d}(r_{t_d})) = \frac{\prod_{j=1}^d P(x_{t_j}(r_{t_j}) | \omega_i)}{\sum_{k=1}^m \prod_{j=1}^d P(x_{t_j}(r_{t_j}) | \omega_k)}$$

【0093】

入力データ  $X = [X_{t_1}(r_{t_1}), X_{t_2}(r_{t_2}), \dots, X_{t_d}(r_{t_d})]$  の事後確率  $P(i | X_{t_1}(r_{t_1}), X_{t_2}(r_{t_2}), \dots, X_{t_d}(r_{t_d}))$  において、式 9 が成立するとき、本装置は、入力データ  $X$  の分類を、分類  $k$  と判定する。

40

【0094】

(式 9)

$$\begin{aligned} \max_i P(\omega_i | x_{t_1}(r_{t_1}), x_{t_2}(r_{t_2}), \dots, x_{t_d}(r_{t_d})) \\ = P(\omega_k | x_{t_1}(r_{t_1}), x_{t_2}(r_{t_2}), \dots, x_{t_d}(r_{t_d})) \end{aligned}$$

【0095】

以下、これまで説明した本装置の特徴を、まとめて記載しておく。

【0096】

50

## (特徴1)

入力データが、複数の分類の中の特定分類に属するか否かを、複数の指標の中から選択された選択指標に基づいて判定する情報処理装置であって、

前記複数の指標のそれぞれは、複数の区分を含み、

前記複数の指標の中から前記選択指標を選択する指標選択部と、

前記入力データが該当する前記選択指標に含まれる指標ごとの区分に基づいて、前記入力データが前記特定分類に属するか否かを決定する分類判定部と、  
を有してなり、

前記選択指標の選択に用いられる複数の見本データごとの、前記見本データが該当する前記複数の指標ごとの区分を示す指標情報と、前記見本データが前記特定分類に属するか否かを示す分類情報と、が記憶される記憶部、  
を備え、

前記指標選択部は、

(a) 前記複数の見本データのうち、一部の見本データを訓練データとして抽出し、他の一部の見本データを評価データとして抽出する、見本データ抽出部と、

(b) 前記複数の指標のいずれかの組合せで構成される指標群ごとに、

前記訓練データの前記指標情報と、前記訓練データの前記分類情報と、前記評価データの前記指標情報と、を用いて、前記評価データが前記特定分類に属するか否かを推定し、  
前記評価データの前記分類情報を用いて、前記推定の当否を判定する、  
分類推定部と、

(c) 前記指標群ごとの前記推定の当否の判定結果に基づいて、前記選択指標を決定する、指標決定部と、  
を備える、

ことを特徴とする情報処理装置。

## (特徴2)

前記分類推定部は、

(b-1) 前記訓練データの前記指標情報と前記訓練データの前記分類情報とを用いて、前記特定分類が生起したときの前記指標群に含まれる指標の区分ごとの第1条件付確率と、前記非特定分類が生起したときの前記指標群に含まれる指標の区分ごとの第2条件付確率と、を算出し、

(b-2) 前記評価データの前記指標情報と、前記複数の分類のうち前記特定分類が生起する第1事前確率と、前記第1条件付確率と、に基づいて、前記評価データが前記特定分類に属する第1事後確率を算出し、

(b-3) 前記評価データの前記指標情報と、前記複数の分類のうち前記特定分類ではない非特定分類が生起する第2事前確率と、前記第2条件付確率と、に基づいて、前記評価データが前記非特定分類に属する第2事後確率を算出し、

(b-4) 前記第1事後確率と前記第2事後確率とを比較して、前記評価データが前記特定分類に属するか否かの推定の結果を示す推定情報を出し、

(b-5) 前記評価データの前記推定情報と前記評価データの前記分類情報とを比較して、前記推定の当否の判定結果を示す判定情報を出し、

前記指標決定部は、

(c-1) 前記指標群ごとの前記判定情報のうち、所定の条件を満たす判定情報に対応する指標群を特定し、

(c-2) 前記特定された指標群に含まれる指標を前記選択指標として決定する、  
特徴1記載の情報処理装置。

## (特徴3)

前記分類推定部は、前記第1事前確率と、前記第2事前確率と、を算出する、  
特徴2記載の情報処理装置。

## (特徴4)

前記見本データ抽出部は、前記複数の見本データの中から複数の前記評価データを抽出

10

20

30

40

50

し、

前記分類推定部は、前記複数の評価データごとに、前記判定情報を出し、  
 前記指標決定部は、前記複数の評価データごとの前記判定情報に基づいて、前記選択指標を決定する、  
 特徴 2 または 3 記載の情報処理装置。

( 特徴 5 )

前記分類推定部は、  
 前記複数の評価データごとに、前記第 1 条件付確率と前記第 2 条件付確率と前記第 1 事後確率と前記第 2 事後確率とを算出して、前記推定情報を出し、  
 前記複数の評価データごとの前記推定情報に基づいて、前記複数の評価データごとの前記判定情報を出し、  
 特徴 4 記載の情報処理装置。

10

( 特徴 6 )

前記第 1 事前確率は、前記第 2 事前確率と等しい、  
 特徴 2 乃至 5 のいずれかに記載の情報処理装置。

【符号の説明】

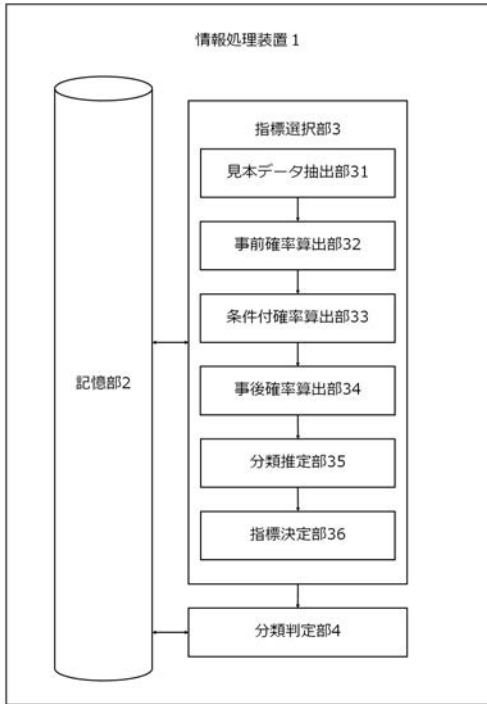
【 0 0 9 7 】

- 1 情報処理装置
- 2 記憶部
- 3 指標選択部
- 3 1 見本データ抽出部
- 3 2 事前確率算出部
- 3 3 条件付確率算出部
- 3 4 事後確率算出部
- 3 5 分類推定部
- 3 6 指標決定部
- 4 分類判定部

20

30

【 図 1 】



【 図 2 】

指標ID	見本ID		
	D <sub>1</sub>	D <sub>2</sub>	...
x <sub>1</sub>	V <sub>11</sub>	V <sub>21</sub>	...
x <sub>2</sub>	V <sub>21</sub>	V <sub>22</sub>	...
...	...	...	...

【 図 3 】

見本ID	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	...
分類ID	ω <sub>2</sub>	ω <sub>1</sub>	ω <sub>1</sub>	ω <sub>2</sub>	...

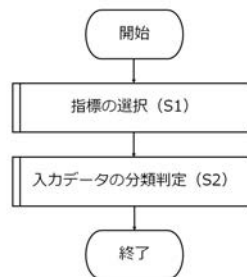
【 図 4 】

指標ID	区分ID	見本ID (訓練データ)		
		D <sub>t1</sub>	D <sub>t2</sub>	...
x <sub>1</sub>	x <sub>1</sub> (1)	-	○	...
	x <sub>1</sub> (2)	○	-	...
x <sub>2</sub>	x <sub>2</sub> (1)	-	-	...
	x <sub>2</sub> (2)	○	○	...
	x <sub>2</sub> (3)	-	-	...
...	...	...	...	...

【 図 5 】

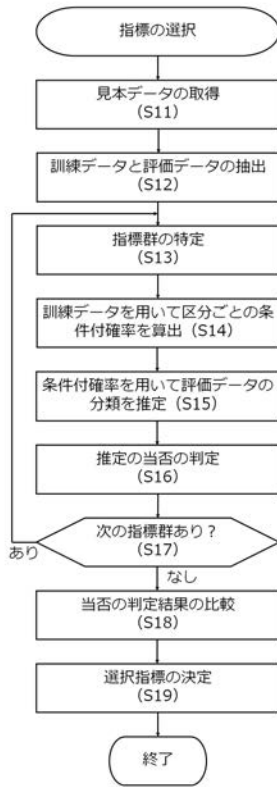
指標ID	区分ID	分類ID	
		ω <sub>1</sub>	ω <sub>2</sub>
x <sub>1</sub>	x <sub>1</sub> (1)	n <sub>1</sub> <sup>1</sup> (1)	n <sub>1</sub> <sup>2</sup> (1)
	x <sub>1</sub> (2)	n <sub>1</sub> <sup>1</sup> (2)	n <sub>1</sub> <sup>2</sup> (2)
x <sub>2</sub>	x <sub>2</sub> (1)	n <sub>2</sub> <sup>1</sup> (1)	n <sub>2</sub> <sup>2</sup> (1)
	x <sub>2</sub> (2)	n <sub>2</sub> <sup>1</sup> (2)	n <sub>2</sub> <sup>2</sup> (2)
	x <sub>2</sub> (3)	n <sub>2</sub> <sup>1</sup> (3)	n <sub>2</sub> <sup>2</sup> (3)
...	...	...	...

【 図 6 】





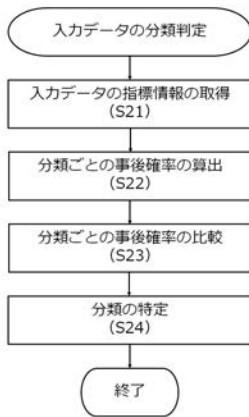
【 図 7 】



【 図 8 】

指標ID	区分ID	見本ID (評価データ)		
		D <sub>e1</sub>	D <sub>e2</sub>	...
x <sub>1</sub>	x <sub>1</sub> (1)	○	○	...
	x <sub>1</sub> (2)	-	-	...
x <sub>2</sub>	x <sub>2</sub> (1)	○	-	...
	x <sub>2</sub> (2)	-	-	...
	x <sub>2</sub> (3)	-	○	...
...	...	...	...	...
分類情報		ω <sub>1</sub>	ω <sub>2</sub>	...
推定情報		ω <sub>1</sub>	ω <sub>1</sub>	...
判定情報		T	F	...

【 図 9 】



【 図 10 】

検査項目  $X_1, X_2, \dots, X_n$  の中で、  
 ○○の再発を予測するのに必要な検査項目は、  
 検査項目  $X_1$  と  $X_2$  です

【 図 1 1 】

検査項目  $Y_1, Y_2, Y_3, \dots, Y_n$  の中で、  
 ○○の再発を予測するのに必要な検査項目は、以下のとおりです

優先順位	検査項目の組	事後確率
1	$Y_1, Y_2, Y_3$	P1
2	$Y_1, Y_2$	P2
3	$Y_1, Y_3$	P3
...	...	...

【 図 1 2 】

患者○○さんが検査済の検査項目  $Z_1$  と  $Z_2$  に  
 検査項目  $Z_3$  の検査結果を追加すると  
 ○○の再発の予測精度を高めることができます

【 手続補正書 】

【 提出日 】平成28年9月16日(2016.9.16)

【 手続補正 1 】

【 補正対象書類名 】特許請求の範囲

【 補正対象項目名 】全文

【 補正方法 】変更

【 補正の内容 】

【 特許請求の範囲 】

【 請求項 1 】

入力データに対応する判定対象が、複数の分類の中の特定分類に属するか否かを、複数の指標の中から選択された選択指標に基づいて判定する情報処理装置であって、  
 前記複数の指標のそれぞれは、複数の区分を含み、  
 前記複数の指標の中から前記選択指標を選択する指標選択部と、  
 前記入力データが該当する前記選択指標に含まれる指標ごとの区分に基づいて、前記入力データに対応する前記判定対象が前記特定分類に属するか否かを決定する分類判定部と、  
 を有してなり、

前記選択指標の選択に用いられる複数の見本対象のそれぞれに対応する複数の見本データごとの、前記見本データが該当する前記複数の指標ごとの区分を示す指標情報と、前記見本データに対応する前記見本対象が前記特定分類に属するか否かを示す分類情報と、が記憶される記憶部、

を備え、

前記指標選択部は、

( a ) 前記複数の見本データのうち、一部の見本データを訓練データとして抽出し、他の一部の見本データを評価データとして抽出する、見本データ抽出部と、

(b) 前記複数の指標のいずれかの組合せで構成される指標群ごとに、前記訓練データの前記指標情報と、前記訓練データの前記分類情報と、前記評価データの前記指標情報と、を用いて、前記評価データに対応する前記見本対象が前記特定分類に属するか否かを推定し、

前記評価データの前記分類情報を用いて、前記推定の当否を判定する、  
分類推定部と、

(c) 前記指標群ごとの前記推定の当否の判定結果に基づいて、前記選択指標を決定する、  
指標決定部と、  
を備え、

前記分類推定部は、

(b-1) 前記訓練データの前記指標情報と前記訓練データの前記分類情報とを用いて前記複数の分類ごとに集計された前記指標ごとの各区分に属する見本データの数に基づいて、前記特定分類が生起したときの前記指標群に含まれる指標の区分ごとの第1条件付確率と、前記非特定分類が生起したときの前記指標群に含まれる指標の区分ごとの第2条件付確率と、を算出し、

(b-2) 前記評価データの前記指標情報と、前記複数の分類のうち前記特定分類が生起する第1事前確率と、前記第1条件付確率と、に基づいて、前記評価データに対応する前記見本対象が前記特定分類に属する第1事後確率を算出し、

(b-3) 前記評価データの前記指標情報と、前記複数の分類のうち前記特定分類ではない非特定分類が生起する第2事前確率と、前記第2条件付確率と、に基づいて、前記評価データに対応する前記見本対象が前記非特定分類に属する第2事後確率を算出し、

(b-4) 前記第1事後確率と前記第2事後確率とを比較して、前記評価データに対応する前記見本対象が前記特定分類に属するか否かの推定の結果を示す推定情報を出し、

(b-5) 前記評価データの前記推定情報と前記評価データの前記分類情報とを比較して、前記推定の当否の判定結果を示す判定情報を出し、

前記指標決定部は、

(c-1) 前記指標群ごとの前記判定情報のうち、所定の条件を満たす判定情報に対応する指標群を特定し、

(c-2) 前記特定された指標群に含まれる指標を前記選択指標として決定する、  
ことを特徴とする情報処理装置。

【請求項2】

前記分類推定部は、前記第1事前確率と、前記第2事前確率と、を算出する、  
請求項1記載の情報処理装置。

【請求項3】

前記見本データ抽出部は、前記複数の見本データの中から複数の前記評価データを抽出し、

前記分類推定部は、前記複数の評価データごとに、前記判定情報を出し、

前記指標決定部は、前記複数の評価データごとの前記判定情報に基づいて、前記選択指標を決定する、

請求項1または2記載の情報処理装置。

【請求項4】

前記分類推定部は、

前記複数の評価データごとに、前記第1条件付確率と前記第2条件付確率と前記第1事後確率と前記第2事後確率とを算出して、前記推定情報を出し、

前記複数の評価データごとの前記推定情報に基づいて、前記複数の評価データごとの前記判定情報を出し、

請求項3記載の情報処理装置。

【請求項5】

前記第1事前確率は、前記第2事前確率と等しい、  
請求項1乃至4のいずれかに記載の情報処理装置。

## 【請求項6】

コンピュータを、請求項1乃至5のいずれかに記載の情報処理装置として機能させる、ことを特徴とする情報処理プログラム。

## 【請求項7】

入力データに対応する判定対象が、複数の分類の中の特定分類に属するか否かを、複数の指標の中から選択された選択指標に基づいて判定する情報処理装置により実行される情報処理方法であって、

前記複数の指標のそれぞれは、複数の区分を含み、

前記情報処理装置は、

前記複数の指標の中から前記選択指標を選択する指標選択部と、

前記入力データが該当する前記選択指標に含まれる指標ごとの区分に基づいて、前記入力データに対応する前記判定対象が前記特定分類に属するか否かを決定する分類判定部と、

前記選択指標の選択に用いられる複数の見本対象のそれぞれに対応する複数の見本データごとの、前記見本データが該当する前記複数の指標ごとの区分を示す指標情報と、前記見本データに対応する前記見本対象が前記特定分類に属するか否かを示す分類情報と、が記憶される記憶部と、

を備え、

前記情報処理装置が、

(a) 前記複数の見本データのうち、一部の見本データを訓練データとして抽出し、他の一部の見本データを評価データとして抽出する、見本データ抽出ステップと、

(b) 前記複数の指標のいずれかの組合せで構成される指標群ごとに、

前記訓練データの前記指標情報と、前記訓練データの前記分類情報と、前記評価データの前記指標情報と、を用いて、前記評価データに対応する前記見本対象が前記特定分類に属するか否かを推定し、

前記評価データの前記分類情報を用いて、前記推定の当否を判定する、分類推定ステップと、

(c) 前記指標群ごとの前記推定の当否の判定結果に基づいて、前記選択指標を決定する、指標決定ステップと、

を有してなり、

前記分類推定ステップは、

(b-1) 前記訓練データの前記指標情報と前記訓練データの前記分類情報とを用いて前記複数の分類ごとに集計された前記指標ごとの各区分に属する見本データの数に基づいて、前記特定分類が生起したときの前記指標群に含まれる指標の区分ごとの第1条件付確率と、前記非特定分類が生起したときの前記指標群に含まれる指標の区分ごとの第2条件付確率と、を算出するステップと、

(b-2) 前記評価データの前記指標情報と、前記複数の分類のうち前記特定分類が生起する第1事前確率と、前記第1条件付確率と、に基づいて、前記評価データに対応する前記見本対象が前記特定分類に属する第1事後確率を算出するステップと、

(b-3) 前記評価データの前記指標情報と、前記複数の分類のうち前記特定分類ではない非特定分類が生起する第2事前確率と、前記第2条件付確率と、に基づいて、前記評価データに対応する前記見本対象が前記非特定分類に属する第2事後確率を算出するステップと、

(b-4) 前記第1事後確率と前記第2事後確率とを比較して、前記評価データに対応する前記見本対象が前記特定分類に属するか否かの推定の結果を示す推定情報を出力するステップと、

(b-5) 前記評価データの前記推定情報と前記評価データの前記分類情報とを比較して、前記推定の当否の判定結果を示す判定情報を出力するステップと、を備え、

前記指標決定ステップは、

(c - 1) 前記指標群ごとの前記判定情報のうち、所定の条件を満たす判定情報に対応する指標群を特定するステップと、

(c - 2) 前記特定された指標群に含まれる指標を前記選択指標として決定するステップと、  
を備える、

ことを特徴とする情報処理方法。

【請求項 8】

入力データに対応する判定対象が、複数の分類の中の特定分類に属するか否かを、複数の指標に基づいて判定する情報処理装置であって、

前記複数の指標のそれぞれは、複数の区分を含み、

前記判定対象に対応する入力データと、複数の見本対象のそれぞれに対応する複数の見本データと、が記憶される記憶部と、

前記入力データが該当する前記指標ごとの区分に基づいて、前記判定対象が前記特定分類に属するか否かを決定する分類判定部と、

を有してなり、

前記入力データは、前記入力データが該当する前記複数の指標ごとの区分を示す指標情報、を含み、

前記見本データは、前記見本データが該当する前記複数の指標ごとの区分を示す指標情報と、前記見本対象が前記特定分類に属するか否かを示す分類情報と、を含み、

前記分類判定部は、

前記見本データの前記指標情報と前記見本データの前記分類情報とを用いて前記複数の分類ごとに集計された前記指標ごとの各区分に属する見本データの数に基づいて、前記特定分類が生起したときの前記指標の区分ごとの第 1 条件付確率と、前記複数の分類のうち前記特定分類ではない非特定分類が生起したときの前記指標の区分ごとの第 2 条件付確率と、を算出し、

前記入力データの前記指標情報と、前記複数の分類のうち前記特定分類が生起する第 1 事前確率と、前記第 1 条件付確率と、に基づいて、前記判定対象が前記特定分類に属する第 1 事後確率を算出し、

前記入力データの前記指標情報と、前記複数の分類のうち前記非特定分類が生起する第 2 事前確率と、前記第 2 条件付確率と、に基づいて、前記判定対象が前記非特定分類に属する第 2 事後確率を算出し、

前記第 1 事後確率と前記第 2 事後確率とを比較して、前記判定対象が前記特定分類に属するか否かを決定する、

ことを特徴とする情報処理装置。

【請求項 9】

前記複数の指標には、定性的な指標が含まれる、  
請求項 8 記載の情報処理装置。

【請求項 10】

前記複数の指標には、定量的な指標が含まれる、  
請求項 9 記載の情報処理装置。

【請求項 11】

前記複数の指標の中から前記選択指標を選択する指標選択部、  
を備え、

前記分類判定部は、前記入力データが該当する前記選択指標ごとの区分に基づいて、前記判定対象が前記特定分類に属するか否かを決定する、

請求項 8 乃至 10 のいずれかに記載の情報処理装置。

【請求項 12】

前記指標選択部は、前記第 1 条件付確率と前記第 2 条件付確率とを用いて、前記選択指標を選択する、

請求項 11 記載の情報処理装置。

## 【請求項 13】

前記指標選択部は、

(a) 前記複数の見本データのうち、一部の見本データを訓練データとして抽出し、他の一部の見本データを評価データとして抽出する、見本データ抽出部と、

(b) 前記複数の指標のいずれかの組合せで構成される指標群ごとに、

前記訓練データの前記指標情報と、前記訓練データの前記分類情報と、前記評価データの前記指標情報と、を用いて、前記評価データに対応する前記見本対象が前記特定分類に属するか否かを推定し、

前記評価データの前記分類情報を用いて、前記推定の当否を判定する、  
分類推定部と、

(c) 前記指標群ごとの前記推定の当否の判定結果に基づいて、前記選択指標を決定する、  
指標決定部と、

を備え、

前記分類推定部は、前記第1条件付確率と前記第2条件付確率とを用いて、前記評価データに対応する前記見本対象が前記特定分類に属するか否かを推定する、

請求項12記載の情報処理装置。

## 【請求項 14】

前記見本データ抽出部は、前記複数の見本データの中から複数の前記評価データを抽出し、

前記分類推定部は、前記複数の評価データごとに、前記第1条件付確率と前記第2条件付確率とを用いて、前記推定の当否の判定結果を示す判定情報を出し、

前記指標決定部は、前記複数の評価データごとの前記判定情報に基づいて、前記選択指標を決定する、

請求項13記載の情報処理装置。

## 【請求項 15】

コンピュータを、請求項8乃至14のいずれかに記載の情報処理装置として機能させる、  
ことを特徴とする情報処理プログラム。

## 【請求項 16】

判定対象が、複数の分類の中の特定分類に属するか否かを、複数の指標に基づいて判定する情報処理装置により実行される情報処理方法であって、

前記複数の指標のそれぞれは、複数の区分を含み、

前記情報処理装置は、

前記判定対象に対応する入力データと、複数の見本対象のそれぞれに対応する見本データと、が記憶される記憶部と、

前記入力データが該当する前記指標ごとの区分に基づいて、前記判定対象が前記特定分類に属するか否かを決定する分類判定部と、

を備え、

前記入力データは、前記入力データが該当する前記複数の指標ごとの区分を示す指標情報、を含み、

前記見本データは、前記見本データが該当する前記複数の指標ごとの区分を示す指標情報と、前記見本対象が前記特定分類に属するか否かを示す分類情報と、を含み、

前記情報処理装置が、

前記見本データの前記指標情報と前記見本データの前記分類情報とを用いて前記複数の分類ごとに集計された前記指標ごとの各区分に属する見本データの数に基づいて、前記特定分類が生起したときの前記指標の区分ごとの第1条件付確率と、前記複数の分類のうち前記特定分類ではない非特定分類が生起したときの前記指標の区分ごとの第2条件付確率と、を算出するステップと、

前記入力データの前記指標情報と、前記複数の分類のうち前記特定分類が生起する第1事前確率と、前記第1条件付確率と、に基づいて、前記判定対象が前記特定分類に属する

第 1 事後確率を算出するステップと、

前記入力データの前記指標情報と、前記複数の分類のうち前記非特定分類が生起する第 2 事前確率と、前記第 2 条件付確率と、に基づいて、前記判定対象が前記非特定分類に属する第 2 事後確率を算出するステップと、

前記第 1 事後確率と前記第 2 事後確率とを比較して、前記判定対象が前記特定分類に属するか否かを決定するステップと、

を有してなることを特徴とする情報処理方法。

フロントページの続き

(72)発明者 爲佐 卓夫

山口県宇部市南小串1丁目1-1 国立大学法人山口大学医学部内

(72)発明者 岡 正朗

山口県山口市吉田1677-1 国立大学法人山口大学内