

(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2016-162276

(P2016-162276A)

(43) 公開日 平成28年9月5日(2016.9.5)

(51) Int.Cl.
G06N 5/04 (2006.01)

F I
G06N 5/04

テーマコード (参考)

審査請求 未請求 請求項の数 12 O L (全 18 頁)

(21) 出願番号 特願2015-41313 (P2015-41313)
(22) 出願日 平成27年3月3日 (2015.3.3)

(71) 出願人 504182255
国立大学法人横浜国立大学
神奈川県横浜市保土ヶ谷区常盤台79番1号
(74) 代理人 100064908
弁理士 志賀 正武
(74) 代理人 100149548
弁理士 松沼 泰史
(74) 代理人 100188558
弁理士 飯田 雅人
(74) 代理人 100196058
弁理士 佐藤 彰雄
(72) 発明者 長尾 智晴
神奈川県横浜市保土ヶ谷区常盤台79番1号 国立大学法人横浜国立大学内
最終頁に続く

(54) 【発明の名称】 説明文生成装置、説明文書作成方法およびプログラム

(57) 【要約】

【課題】 決定木による判断プロセスを人間に理解させるための説明文を生成する。

【解決手段】 説明文生成装置は、特徴量の判定条件と分岐先のノードとを関連付けられた内部ノードと、所属クラスを関連付けられた葉ノードとを有する決定木を取得する決定木取得部と、特徴量に関連付けて、所定の言語により当該特徴量を表す説明語を記憶する説明記憶部と、前記ノードの判定条件に用いられる特徴量に関連付けられた前記説明語に基づいて、前記決定木の説明文を生成する説明文生成部とを備える。

【選択図】 図1

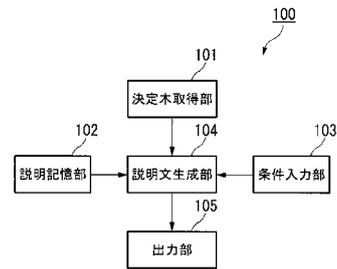


図1

【特許請求の範囲】

【請求項 1】

特徴量の判定条件と分岐先のノードとを関連付けられた内部ノードと、所属クラスを関連付けられた葉ノードとを有する決定木を取得する決定木取得部と、

特徴量に関連付けて、所定の言語により当該特徴量を表す説明語を記憶する説明記憶部と、

前記ノードの判定条件に用いられる特徴量に関連付けられた前記説明語に基づいて、前記決定木の説明文を生成する説明文生成部と

を備える説明文生成装置。

【請求項 2】

前記判定条件は、前記特徴量と所定の閾値との比較に係るものであって、

前記説明記憶部が、閾値に関連付けて当該閾値によって定義される範囲を表す述語を記憶し、

前記説明文生成部が、前記ノードの判定条件に用いられる特徴量に関連付けられた前記説明語と、当該判定条件が用いる閾値に関連付けられた前記述語とに基づいて、前記決定木の説明文を生成する

請求項 1 に記載の説明文生成装置。

【請求項 3】

説明文に表すべきノードの量の入力を受け付けるノード量入力部をさらに備え、

前記説明文生成部が、前記ノード量入力部に入力された量のノードに係る説明文を生成する

請求項 1 または請求項 2 に記載の説明文生成装置。

【請求項 4】

前記説明文生成部が、前記決定木のノードのうち深さが小さいものから順に、前記ノード量入力部に入力された量のノードに係る説明文を生成する

請求項 3 に記載の説明文生成装置。

【請求項 5】

前記説明記憶部が、前記説明語に関連付けて、当該説明語のわかりやすさを示す値を記憶し、

前記説明文生成部が、前記決定木のノードのうち判定条件に用いられる特徴量のわかりやすさが高いものから順に、前記ノード量入力部に入力された量のノードに係る説明文を生成する

請求項 3 に記載の説明文生成装置。

【請求項 6】

前記決定木による分類の対象となる対象データの入力を受け付ける対象データ入力部と、

前記決定木に基づいて前記対象データの所属クラスを分類する分類部と

をさらに備え、

前記説明文生成部が、前記決定木のうち前記分類部が分類の過程で辿った経路に係る説明文を生成する

請求項 1 から請求項 5 の何れか 1 項に記載の説明文生成装置。

【請求項 7】

前記説明文生成部が、一の所属クラスを示す葉ノードのうち、前記決定木の学習に用いた学習データの到達数が最も多いものへ到達する経路に属するノードに係る説明文を生成する

請求項 1 から請求項 5 の何れか 1 項に記載の説明文生成装置。

【請求項 8】

前記説明記憶部が、複数の特徴量の組み合わせに関連付けて当該組み合わせを表す説明語を記憶し、

前記説明文生成部が、前記決定木の経路に属するノードの中に前記特徴量の組み合わせ

10

20

30

40

50

に対応するノードの組み合わせが含まれる場合に、前記特徴量の組み合わせに関連付けられた説明語に基づいて前記決定木の説明文を生成する

請求項 1 から請求項 7 の何れか 1 項に記載の説明文生成装置。

【請求項 9】

前記決定木が、前記決定木による分類の対象となる対象データの所属クラスを決定する主決定木と、前記主決定木の前記判定条件に係る前記特徴量の範囲を決定する副決定木とを含む

請求項 1 から請求項 8 の何れか 1 項に記載の説明文生成装置。

【請求項 10】

前記説明文生成部が、前記主決定木の説明文を生成し、前記主決定木の経路に属するノードの中に前記副決定木によって前記特徴量の範囲が決定されるものが含まれる場合に、さらに当該副決定木の説明文を生成する

10

請求項 9 に記載の説明文生成装置。

【請求項 11】

説明文生成装置が、特徴量の判定条件と分岐先のノードとを関連付けられた内部ノードと、所属クラスを関連付けられた葉ノードとを有する決定木を取得するステップと、

前記説明文生成装置が、特徴量に関連付けて、所定の言語により当該特徴量を表す説明語を記憶する説明記憶部が前記ノードの判定条件に用いられる特徴量に関連付けて記憶する前記説明語に基づいて、前記決定木の説明文を生成するステップと、

20

前記説明文を含む文書を作成するステップと

を備える説明文書作成方法。

【請求項 12】

コンピュータを、

特徴量の判定条件と分岐先のノードとを関連付けられた内部ノードと、所属クラスを関連付けられた葉ノードとを有する決定木を取得する決定木取得部、

特徴量に関連付けて、所定の言語により当該特徴量を表す説明語を記憶する説明記憶部、

前記ノードの判定条件に用いられる特徴量に関連付けられた前記説明語に基づいて、前記決定木の説明文を生成する説明文生成部

として機能させるためのプログラム。

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、説明文生成装置、説明文書作成方法およびプログラムに関する。

【背景技術】

【0002】

コンピュータによる機械学習技術に基づく判断プロセスが、実生活で利用される機会が増えている。当該判断プロセスは、コンピュータによって最適化されたものであり、人間が当該判断プロセスを理解することは困難であることが多い。

【0003】

40

人間による理解が比較的容易な判断プロセスとして、決定木および決定ネットワーク（例えば、特許文献 1 を参照）が知られている。決定木とは、特徴量の判定条件と分岐先のノードとを関連付けられた内部ノードと、所属クラスを関連付けられた葉ノードと、各ノードを接続するリンクとからなる木構造の分類モデルである。決定ネットワークは、複数のデータの入力を可能とするグラフ構造の分類モデルである。なお、決定ネットワークは、分類の対象となる対象データの所属クラスを決定する主決定木と、主決定木の判定条件を決定する副決定木とを組み合わせたものであり、決定木の種類といえる。

【先行技術文献】

【特許文献】

【0004】

50

【特許文献1】特開2012-73750号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、決定木は、比較的理解しやすい判断プロセスであるとはいえ、情報工学に明るくない利用者にとっては、依然として理解することが困難である。

本発明の目的は、決定木による判断プロセスを人間に理解させるための説明文を生成する説明文生成装置、説明文書作成方法およびプログラムを提供することにある。

【課題を解決するための手段】

【0006】

10

本発明の第1の態様によれば、説明文生成装置は、特徴量の判定条件と分岐先のノードとを関連付けられた内部ノードと、所属クラスを関連付けられた葉ノードとを有する決定木を取得する決定木取得部と、特徴量に関連付けて、所定の言語により当該特徴量を表す説明語を記憶する説明記憶部と、前記ノードの判定条件に用いられる特徴量に関連付けられた前記説明語に基づいて、前記決定木の説明文を生成する説明文生成部とを備える。

【0007】

本発明の第2の態様によれば、第1の態様に係る説明文生成装置は、前記判定条件は、前記特徴量と所定の閾値との比較に係るものであって、前記説明記憶部が、閾値に関連付けて当該閾値によって定義される範囲を表す述語を記憶し、前記説明文生成部が、前記ノードの判定条件に用いられる特徴量に関連付けられた前記説明語と、当該判定条件が用い

20

【0008】

本発明の第3の態様によれば、第1または第2の態様に係る説明文生成装置は、説明文に表すべきノードの量の入力を受け付けるノード量入力部をさらに備え、前記説明文生成部が、前記ノード量入力部に入力された量のノードに係る説明文を生成する。

【0009】

本発明の第4の態様によれば、第3の態様に係る説明文生成装置は、前記説明文生成部が、前記決定木のノードのうち深さが小さいものから順に、前記ノード量入力部に入力された量のノードに係る説明文を生成する。

【0010】

30

本発明の第5の態様によれば、第3の態様に係る説明文生成装置は、前記説明記憶部が、前記説明語に関連付けて、当該説明語のわかりやすさを示す値を記憶し、前記説明文生成部が、前記決定木のノードのうち判定条件に用いられる特徴量のわかりやすさが高いものから順に、前記ノード量入力部に入力された量のノードに係る説明文を生成する。

【0011】

本発明の第6の態様によれば、第1から第5の何れかの態様に係る説明文生成装置は、前記決定木による分類の対象となる対象データの入力を受け付ける対象データ入力部と、前記決定木に基づいて前記対象データの所属クラスを分類する分類部とをさらに備え、前記説明文生成部が、前記決定木のうち前記分類部が分類の過程で辿った経路に係る説明文を生成する。

40

【0012】

本発明の第7の態様によれば、第1から第5の何れかの態様に係る説明文生成装置は、前記説明文生成部が、一の所属クラスを示す葉ノードのうち、前記決定木の学習に用いた学習データの到達数が最も多いものへ到達する経路に属するノードに係る説明文を生成する。

【0013】

本発明の第8の態様によれば、第1から第7の何れかの態様に係る説明文生成装置は、前記説明記憶部が、複数の特徴量の組み合わせに関連付けて当該組み合わせを表す説明語を記憶し、前記説明文生成部が、前記決定木の経路に属するノードの中に前記特徴量の組み合わせに対応するノードの組み合わせが含まれる場合に、前記特徴量の組み合わせに関

50

連付けられた説明語に基づいて前記決定木の説明文を生成する。

【0014】

本発明の第9の態様によれば、第1から第8の何れかの態様に係る説明文生成装置は、前記決定木が、前記決定木による分類の対象となる対象データの所属クラスを決定する主決定木と、前記主決定木の前記判定条件に係る前記特徴量の範囲を決定する副決定木とを含む。

【0015】

本発明の第10の態様によれば、第9の態様に係る説明文生成装置は、前記説明文生成部が、前記主決定木の説明文を生成し、前記主決定木の経路に属するノードの中に前記副決定木によって前記特徴量の範囲が決定されるものが含まれる場合に、さらに当該副決定木の説明文を生成する。

10

【0016】

本発明の第11の態様によれば、説明文書作成方法は、説明文生成装置が、特徴量の判定条件と分岐先のノードとを関連付けられた内部ノードと、所属クラスを関連付けられた葉ノードとを有する決定木を取得するステップと、前記説明文生成装置が、特徴量に関連付けて、所定の言語により当該特徴量を表す説明語を記憶する説明記憶部が前記ノードの判定条件に用いられる特徴量に関連付けて記憶する前記説明語に基づいて、前記決定木の説明文を生成するステップと、前記説明文を含む文書を作成するステップとを備える。

【0017】

本発明の第12の態様によれば、プログラムは、コンピュータを、特徴量の判定条件と分岐先のノードとを関連付けられた内部ノードと、所属クラスを関連付けられた葉ノードとを有する決定木を取得する決定木取得部、特徴量に関連付けて、所定の言語により当該特徴量を表す説明語を記憶する説明記憶部、前記ノードの判定条件に用いられる特徴量に関連付けられた前記説明語に基づいて、前記決定木の説明文を生成する説明文生成部として機能させる。

20

【発明の効果】

【0018】

上記態様のうち少なくとも1つの態様によれば、説明文生成装置は、決定木のノードの判定条件を所定の言語で表した説明文を生成する。これにより、説明文生成装置は、当該言語を理解することができる利用者に、決定木による判断プロセスを理解させることができる。

30

【図面の簡単な説明】

【0019】

【図1】第1の実施形態に係る説明文生成装置の構成を示す概略ブロック図である。

【図2】説明語テーブルの例を示す図である。

【図3】述語テーブルの例を示す図である。

【図4】第1の実施形態に係る説明文生成装置の動作を示すフローチャートである。

【図5】決定木の一例を示す図である。

【図6】第2の実施形態に係る説明文生成装置の構成を示す概略ブロック図である。

【図7】第2の実施形態に係る説明文生成装置の動作を示すフローチャートである。

40

【図8】第3の実施形態に係る説明語テーブルの例を示す図である。

【図9】決定ネットワークの一例を示す図である。

【図10】少なくとも1つの実施形態に係るコンピュータの構成を示す概略ブロック図である。

【発明を実施するための形態】

【0020】

《第1の実施形態》

以下、図面を参照しながら実施形態について詳しく説明する。

図1は、第1の実施形態に係る説明文生成装置の構成を示す概略ブロック図である。

説明文生成装置100は、決定木の入力を受け付け、当該決定木による判断プロセスの

50

説明文を生成する。具体的には、説明文生成装置 100 は、決定木によって分類される複数の所属クラスそれぞれについて、当該所属クラスに属すると判断する判断プロセスの説明文を生成する。なお、決定木とは、入力データを複数の所属クラスの 1 つに分類する分類モデルである。具体的には、決定木は、特徴量の判定条件と分岐先のノードとを関連付けられた内部ノードと、所属クラスを関連付けられた葉ノードと、各ノードを接続するリンクとからなる木構造の分類モデルである。

説明文生成装置 100 は、決定木取得部 101、説明記憶部 102、条件入力部 103、説明文生成部 104、出力部 105 を備える。

【0021】

決定木取得部 101 は、説明対象となる決定木および当該決定木の学習時に生成されるメタデータを取得する。メタデータには、決定木の各ノードについての学習データの通過率または通過数が含まれる。

説明記憶部 102 は、説明語テーブルと述語テーブルとを記憶する。説明語テーブルには、決定木による判断対象となる特徴量と当該特徴量を表す日本語の説明語とが関連付けて格納される。

【0022】

図 2 は、説明語テーブルの例を示す図である。

図 2 に示す例によれば、説明語テーブルには、「黄色周辺の色相（色相 30 度以上 90 度未満）の画素数を全画素数で除算した値」という特徴量に関連付けて、「黄色っぽい部分が」という説明語が格納される。また、説明語テーブルには、「緑色周辺の色相（色相 90 度以上 150 度未満）の画素数を全画素数で除算した値」という特徴量に関連付けて、「緑色っぽい部分が」という説明語が格納される。また、説明語テーブルには、「水色周辺の色相（色相 150 度以上 210 度未満）の画素数を全画素数で除算した値」という特徴量に関連付けて、「水色っぽい部分が」という説明語が格納される。また、説明語テーブルには、「青色周辺の色相（色相 210 度以上 270 度未満）の画素数を全画素数で除算した値」という特徴量に関連付けて、「青色っぽい部分が」という説明語が格納される。また、説明語テーブルには、「紫色周辺の色相（色相 270 度以上 330 度未満）の画素数を全画素数で除算した値」という特徴量に関連付けて、「紫色っぽい部分が」という説明語が格納される。また、説明語テーブルには、「赤色周辺の色相（色相 330 度以上 360 度未満および色相 0 度以上 30 度未満）の画素数を全画素数で除算した値」という特徴量に関連付けて、「赤色っぽい部分が」という説明語が格納される。また、説明語テーブルには、「色相の分散」という特徴量に関連付けて、「色数が」という説明語が格納される。また、説明語テーブルには、「彩度の平均値」という特徴量に関連付けて、「鮮やかさが」という説明語が格納される。また、説明語テーブルには、「明度の平均値」という特徴量に関連付けて、「明るさが」という説明語が格納される。また、説明語テーブルには、「垂直線 ± 20 度のエッジをもつ画素数を全画素数で除算した値」という特徴量に関連付けて、「縦線が」という説明語が格納される。また、説明語テーブルには、「水平線 ± 20 度のエッジをもつ画素数を全画素数で除算した値」という特徴量に関連付けて、「横線が」という説明語が格納される。また、説明語テーブルには、「縦線・横線以外のエッジをもつ画素数を全画素数で除算した値」という特徴量に関連付けて、「斜線が」という説明語が格納される。

【0023】

述語テーブルには、1 以下の値に正規化された特徴量の閾値に関連付けて、当該閾値によって定義される範囲を表す日本語の述語が格納される。

図 3 は、述語テーブルの例を示す図である。

図 3 に示す例によれば、述語テーブルには、「0.0 以上 0.2 未満」の閾値に関連付けて、当該閾値未満の範囲を表す述語として「あまりない」という述語が格納され、当該閾値以上の範囲を表す述語として「あまりないわけではない」という述語が格納される。また、述語テーブルには、「0.2 以上 0.4 未満」の閾値に関連付けて、当該閾値未満の範囲を表す述語として「少ない」という述語が格納され、当該閾値以上の範囲を表す述

10

20

30

40

50

語として「少ないわけではない」という述語が格納される。また、述語テーブルには、「0.4以上0.6未満」の閾値に関連付けて、当該閾値未満の範囲を表す述語として「ある程度多いわけではない」という述語が格納され、当該閾値以上の範囲を表す述語として「ある程度多い」という述語が格納される。また、述語テーブルには、「0.6以上0.8未満」の閾値に関連付けて、当該閾値未満の範囲を表す述語として「多いわけではない」という述語が格納され、当該閾値以上の範囲を表す述語として「多い」という述語が格納される。また、述語テーブルには、「0.8以上1.0以下」の閾値に関連付けて、当該閾値未満の範囲を表す述語として「とても多いわけではない」という述語が格納され、当該閾値以上の範囲を表す述語として「とても多い」という述語が格納される。

【0024】

条件入力部103は、利用者から、生成すべき説明文の詳細度および所属クラス1つあたりの説明文の文章数の入力を受け付ける。説明文の詳細度とは、決定木による判断プロセスの説明の詳細さを示す値である。具体的には、詳細度は、説明文に含まれる説明語と述語の数の多さを表す。つまり、詳細度は、説明対象となる所属クラスに関連付けられた葉ノードに到達する経路に含まれるノードのうち、説明文に表すべきノードの割合を示す。説明文に含まれる説明語と述語の数は、詳細度が高いほど多くなる。したがって、詳細度が高いほど、説明文の正確さが向上する一方、説明文のわかりやすさは低下する。

【0025】

説明文生成部104は、条件入力部103に入力された詳細度および文章数と、説明記憶部102が記憶する情報とに基づいて、決定木取得部101が取得した決定木によって分類される所属クラスごとの説明文を生成する。具体的には、説明文生成部104は、分類儀のノードの判定条件に用いられる特徴量に関連付けられた説明語と、当該判定条件が示す範囲に関連付けられた述語とを組み合わせることで、決定木の説明文を生成する。

出力部105は、説明文生成部104が生成した説明文を含むデータを出力する。本実施形態に係る出力部105は、説明文を含む文書データをプリンタに出力することで、説明文を含む文書を作成する。なお、他の実施形態に係る出力部105は、説明文を含むデータを、外部の記録媒体、ディスプレイ、またはその他の装置に出力しても良い。

【0026】

次に、本実施形態に係る説明文生成装置100による説明文書作成方法について説明する。

図4は、第1の実施形態に係る説明文生成装置の動作を示すフローチャートである。

まず決定木取得部101は、説明文の生成対象となる決定木および当該決定木のメタデータを取得する(ステップS1)。次に、条件入力部103は、利用者から説明文の詳細度および所属クラス1つあたりの説明文の文章数の入力を受け付ける(ステップS2)。

【0027】

次に、説明文生成部104は、決定木取得部101が取得した決定木の葉ノードを参照し、当該決定木により分類される所属クラスを特定する(ステップS3)。次に、説明文生成部104は、特定した所属クラスを1つずつ選択し、選択した各所属クラスについて、以下に示すステップS5~ステップS16の処理を実行する(ステップS4)。

【0028】

説明文生成部104は、決定木取得部101が取得した決定木のメタデータを参照し、選択した所属クラスに関連付けられた葉ノードのうち、学習データの到達率または到達数が最も高い葉ノードを特定する(ステップS5)。なお、既に当該所属クラスの説明文を1つ以上生成している場合、説明文生成部104は、説明文の生成に用いられていない葉ノードのうち、学習データの到達率または到達数が最も高い葉ノードを特定する。

【0029】

次に、説明文生成部104は、根ノードから特定した葉ノードに到達する経路を特定する(ステップS6)。次に、説明文生成部104は、特定した経路に含まれる内部ノードの数と、入力された説明文の詳細度とに基づいて、説明文の生成に用いる内部ノードの数を特定する(ステップS7)。具体的には、説明文生成部104は、経路に含まれる内部

10

20

30

40

50

ノードの数に詳細度を表す割合を乗算し、得られた値の小数点以下を切り上げることで、説明文の生成に用いる内部ノードの数を特定する。

次に、説明文生成部104は、ステップS6で特定した経路に含まれる内部ノードのうち、ステップS7で特定した数の内部ノードを、深さが小さいものから順に選択し、選択した各ノードについて、以下に示すステップS9～ステップS13の処理を実行する(ステップS8)。

【0030】

説明文生成部104は、選択された内部ノードに関連付けられた判断条件に用いられる特徴量を特定する(ステップS9)。説明文生成部104は、説明記憶部102が記憶する説明語テーブルから特定した特徴量に関連付けられた説明語を読み出す(ステップS10)。次に、説明文生成部104は、選択された内部ノードに関連付けられた判断条件に用いられる閾値を特定する(ステップS11)。説明文生成部104は、説明記憶部102が記憶する述語テーブルから特定した閾値とステップS6で特定した経路に属する範囲(閾値以上か否か)に関連付けられた述語を読み出す(ステップS12)。次に、説明文生成部104は、読み出した説明語および述語を接続することで、ステップS8で選択された内部ノードの説明文を生成する(ステップS13)。

10

【0031】

説明文生成部104は、ステップS7で特定した数の内部ノードのそれぞれの説明文を生成すると、生成した説明文を組み合わせることで、ステップS5で特定した葉ノードの説明文を生成する(ステップS14)。

20

説明文生成部104は、ステップS14で生成した葉ノードの説明文の数がステップS4で入力された文章数に達したか否かを判定する(ステップS15)。説明文の数が入力された文章数に達していない場合(ステップS15:NO)、ステップS5に戻り、学習データの到達率または到達数が次に高い葉ノードの説明文を生成する。説明文の数が入力された文章数に達した場合(ステップS15:YES)、ステップS14で生成された葉ノードの説明文を組み合わせることで、ステップS4で選択した所属クラスの説明文を生成する(ステップS16)。このとき、説明文生成部104は、決定木取得部101が取得した決定木のメタデータを参照し、ステップS5で特定した各葉ノードの学習データの到達率の和を、説明文が示す説明の網羅率として、当該説明文に追加しても良い。当該説明文の文章数は入力された文章数に等しく、当該説明文に含まれる説明語および述語の数は、入力された詳細度に応じたものとなる。

30

【0032】

説明文生成部104が、決定木により分類され得る全ての所属クラスの説明文を生成すると、出力部105は、当該説明文を含む文書データを生成し、当該文書データをプリンタに出力する(ステップS17)。これにより、説明文生成装置100は、決定木の説明文を含む文書を作成することができる。

【0033】

以下、第1の実施形態に係る説明文の生成方法の具体例を説明する。

図5は、決定木の一例を示す図である。図5に示す決定木は、入力された画像データがハスが写った画像データであるか、ひまわりが写った画像データであるかを判断するものである。当該決定木は、ハスが写った画像50枚とひまわりが写った画像50枚とを学習データとして学習されたものである。

40

図5(A)は、決定木のノードの接続関係を示す図である。図5(B)は、各内部ノードに関連付けられた判断条件の特徴量および閾値を示す図である。図5(C)は、各葉ノードに関連付けられた所属クラスおよび当該葉ノードへの学習データの到達数を示す図である。

【0034】

決定木取得部101は、ステップS1で、図5に示す決定木およびメタデータを取得する。次に条件入力部103は、ステップS2で、詳細度および文章数の入力を受け付ける。本例においては、詳細度50%、文章数2が入力されたものとする。次に説明文生成部

50

る各所属クラスの判定条件を所定の言語で表した説明文を生成する。これにより、説明文生成装置 100 は、当該言語を理解することができる利用者に、決定木による判断プロセスを理解させることができる。

【0040】

なお、本実施形態に係る説明文生成装置 100 は、全ての所属クラスに係る説明文を生成するが、これに限られない。例えば、他の実施形態に係る説明文生成装置 100 は、利用者から説明文を生成すべき所属クラスの入力を受け付け、当該所属クラスについての説明文を生成しても良い。

【0041】

《第2の実施形態》

第2の実施形態について詳しく説明する。

図6は、第2の実施形態に係る説明文生成装置の構成を示す概略ブロック図である。

第2の実施形態に係る説明文生成装置 100 は、利用者から分類の対象となる対象データの入力を受け付け、決定木によって当該対象データがある所属クラスに分類されるまでの判断プロセスの説明文を生成する。

第2の実施形態に係る説明文生成装置 100 は、第1の実施形態の構成に加え、さらに対象データ入力部 106 と分類部 107 とを備える。第2の実施形態に係る説明文生成装置 100 は、第1の実施形態と説明文生成部 104 の動作が異なる。

【0042】

対象データ入力部 106 は、決定木取得部 101 が取得した決定木による分類の対象となる対象データの入力を受け付ける。

分類部 107 は、決定木取得部 101 が取得した決定木に基づいて、入力された対象データの所属クラスを分類する。

説明文生成部 104 は、決定木取得部 101 が取得した決定木のうち、分類部 107 が分類の過程で辿った経路に係る説明文を生成する。

【0043】

次に、本実施形態に係る説明文生成装置 100 による説明文書作成方法について説明する。

図7は、第2の実施形態に係る説明文生成装置の動作を示すフローチャートである。

まず決定木取得部 101 は、説明文の生成対象となる決定木および当該決定木のメタデータを取得する（ステップ S101）。次に、対象データ入力部 106 は、利用者から対象データの入力を受け付ける（ステップ S102）。次に、分類部 107 は、入力された対象データの所属クラスを、決定木取得部が取得した決定木に基づいて分類する（ステップ S103）。分類部 107 は分類の過程で辿った根ノードから葉ノードまでの経路を特定する（ステップ S104）。次に、条件入力部 103 は、利用者から説明文の詳細度の入力を受け付ける（ステップ S105）。なお、対象データの分類の過程で辿る経路の数は1なので、文書数の入力は不要である。

【0044】

次に、説明文生成部 104 は、ステップ S103 で特定した経路に含まれる内部ノードの数と、入力された説明文の詳細度とに基づいて、説明文の生成に用いる内部ノードの数を特定する（ステップ S106）。具体的には、説明文生成部 104 は、経路に含まれる内部ノードの数に詳細度を表す割合を乗算し、得られた値の小数点以下を切り上げることで、説明文の生成に用いる内部ノードの数を特定する。

次に、説明文生成部 104 は、ステップ S103 で特定した経路に含まれる内部ノードのうち、ステップ S106 で特定した数の内部ノードを、深さが小さいものから順に選択し、選択した各ノードについて、以下に示すステップ S108 ~ ステップ S112 の処理を実行する（ステップ S107）。

【0045】

説明文生成部 104 は、選択された内部ノードに関連付けられた判断条件に用いられる特徴量を特定する（ステップ S108）。説明文生成部 104 は、説明記憶部 102 が記

10

20

30

40

50

憶する説明語テーブルから特定した特徴量に関連付けられた説明語を読み出す（ステップ S 1 0 9）。次に、説明文生成部 1 0 4 は、選択された内部ノードに関連付けられた判断条件に用いられる閾値を特定する（ステップ S 1 1 0）。説明文生成部 1 0 4 は、説明記憶部 1 0 2 が記憶する述語テーブルから特定した閾値とステップ S 1 0 4 で特定した経路に属する範囲に関連付けられた述語を読み出す（ステップ S 1 1 1）。次に、説明文生成部 1 0 4 は、読み出した説明語および述語を接続することで、ステップ S 8 で選択された内部ノードの説明文を生成する（ステップ S 1 1 2）。

【 0 0 4 6 】

説明文生成部 1 0 4 は、ステップ S 1 0 6 で特定した数の内部ノードのそれぞれの説明文を生成すると、生成した説明文を組み合わせることで、ステップ S 1 0 4 で特定した経路の説明文を生成する（ステップ S 1 1 3）。そして出力部 1 0 5 は、当該説明文を含む文書データを生成し、当該文書データをプリンタに出力する（ステップ S 1 1 4）。これにより、説明文生成装置 1 0 0 は、対象データの分類に係る決定木の説明文を含む文書を作成することができる。

10

【 0 0 4 7 】

このように、本実施形態によれば、説明文生成装置 1 0 0 は、入力された対象データが分類結果に示される所属クラスに分類された理由を表した説明文を生成する。これにより、説明文生成装置 1 0 0 は、当該言語を理解することができる利用者に、決定木による対象データの分類に係る判断プロセスを理解させることができる。

【 0 0 4 8 】

《 第 3 の実施形態 》

決定木の判断対象となる特徴量の中には、所定の言語による説明が容易なものも存在すれば、所定の言語による説明が困難なものも存在する。例えば、図 2 に例示した色相の分散、彩度の平均値、明度の平均値などの特徴量は、日本語での説明が比較的容易である。他方、例えば、色相の平均値、H O G 特徴量、S I F T 特徴量などの特徴量は、日本語での説明が困難である。

20

第 3 の実施形態に係る説明文生成装置 1 0 0 は、人間にとって理解しにくい表現を抑えた説明文を生成する。

【 0 0 4 9 】

第 3 の実施形態に係る説明文生成装置 1 0 0 は、第 1 の実施形態と説明記憶部 1 0 2 が記憶する説明語テーブルの構成が異なる。また第 3 の実施形態に係る説明文生成装置 1 0 0 は、第 1 の実施形態と説明文生成部 1 0 4 の動作が異なる。

30

【 0 0 5 0 】

説明記憶部 1 0 2 が記憶する説明語テーブルは、特徴量に関連付けて、当該特徴量を表す説明語と、当該説明語のわかりやすさを示す説明適性値とを格納する。説明適性値は、値が高いほど説明語が分かりやすい表現であることを示す。例えば、説明適性値は、特徴量と説明語とを関連付けた管理者によって設定される。

図 8 は、第 3 の実施形態に係る説明語テーブルの例を示す図である。

図 2 に示す例によれば、説明語テーブルには、「色相の分散」という特徴量に関連付けて、「色数が」という説明語および説明適性値「0.7」が格納される。また、説明語テーブルには、「彩度の平均値」という特徴量に関連付けて、「鮮やかさが」という説明語および説明適性値「0.9」が格納される。また、説明語テーブルには、「明度の平均値」という特徴量に関連付けて、「明るさが」という説明語および説明適性値「0.9」が格納される。また、説明語テーブルには、「色相の平均値」という特徴量に関連付けて、「色相の平均値が」という説明語および説明適性値「0.1」が格納される。また、説明語テーブルには、「縦軸の空間周波数の平均値」という特徴量に関連付けて、「縦方向の明るさの変化の頻度が」という説明語および説明適性値「0.3」が格納される。また、説明語テーブルには、「横軸の空間周波数の平均値」という特徴量に関連付けて、「縦方向の明るさの変化の頻度が」という説明語および説明適性値「0.3」が格納される。

40

このように、「鮮やかさが」、「明るさが」などの人間にとって理解しやすい説明語に

50

は、高い説明適性値が関連付けられ、「色相の平均値が」、「縦方向の明るさの変化の頻度が」などの人間にとって理解しにくい説明語には、低い説明適性値が関連付けられる。

【0051】

説明文生成部104は、決定木のノードのうち判定条件に用いられる特徴量の説明適性値が大きいものから順に、条件入力部103に入力された詳細度に応じたノード数のノードに係る説明文を生成する。

これにより、説明文生成装置100は、人間にとって理解しにくい表現を抑えた説明文を生成することができる。

【0052】

《第4の実施形態》

決定木の判断対象となる特徴量の中には、複数の特徴量の組み合わせを1つの説明語で説明可能なものも存在する。例えば、RGB値のR値は「赤成分が」、G値は「緑成分が」という説明語で説明することができ、さらにR値とG値の組み合わせは、「黄色成分が」という説明語で説明することもできる。

第4の実施形態に係る説明文生成装置100は、複数の特徴量の組み合わせを1つの説明語で説明する説明文を生成する。

【0053】

第4の実施形態に係る説明文生成装置100は、第1の実施形態と説明記憶部102が記憶する説明語テーブルの構成が異なる。また第4の実施形態に係る説明文生成装置100は、第1の実施形態と説明記憶部102の動作が異なる。

説明記憶部102が記憶する説明語テーブルは、特徴量および特徴量の組み合わせに関連付けて、当該特徴量または当該特徴量の組み合わせを表す説明語を格納する。

説明文生成部104は、決定木の経路に属するノードの中に特徴量の組み合わせに対応するノードの組み合わせが含まれる場合に、当該特徴量の組み合わせに関連付けられた説明語に基づいて決定木の説明文を生成する。具体的には、図4に示すステップS9で特定した特徴量を含む組み合わせが説明語テーブルに格納されており、かつ当該組み合わせに係る他の特徴量に係る中間ノードがステップS6で特定した経路上に存在する場合に、説明文生成部104は、ステップS10で当該組み合わせに関連付けられた説明語を読み出す。

このように、説明文生成装置100は、複数の特徴量の組み合わせをまとめて1つの説明語で説明する説明文を生成することで、説明文の長さを変えずに説明文の詳細度を高めることができる。

【0054】

《第5の実施形態》

第5の実施形態に係る説明文生成装置100は、特許文献1に係る決定ネットワークの説明文を生成する。決定ネットワークは、複数のデータの入力を可能とするDAG(Directed acyclic graph)の分類モデルである。なお、決定ネットワークは、分類の対象となる対象データの所属クラスを決定する主決定木と、主決定木の判定条件を決定する副決定木とを組み合わせたものであり、決定木の種類といえる。なお、主決定木と副決定木とは、一部のノードを共通にする。

【0055】

図9は、決定ネットワークの一例を示す図である。

図9に示す決定ネットワークは、入力された画像データが、傷がついた革製品が写った画像データであるか、傷がない革製品が写った画像データであるかを判断するものである。

図9(A)は、決定ネットワークのノードの接続関係を示す図である。図9(B)は、各内部ノードに関連付けられた判断条件の特徴量および閾値を示す図である。図9(C)は、各葉ノードに関連付けられた所属クラスおよび当該葉ノードへの学習データの到達数を示す図である。

【0056】

10

20

30

40

50

図9(A)によれば、中間ノードM11には、中間ノードM12に向くリンクと葉ノードL12に向くリンクとが接続される。中間ノードM12には、葉ノードM11に向くリンクと葉ノードL12に向くリンクとが接続される。中間ノードM13には、中間ノードM11に向くリンクと葉ノードL12に向くリンクとが接続される。中間ノードM11には、分類の対象となる画像データである対象データが入力され、中間ノードM13には、対象データの近傍の画像データである参照データが入力される。当該決定ネットワークは、対象データが入力される中間ノードM11を根ノードとする、中間ノードM11およびM12ならびに葉ノードL11およびL12からなる主決定木と、中間ノードM13を根ノードとする、中間ノードM11、M12およびM13ならびに葉ノードL11およびL12からなる副決定木とを内包する。つまり対象データについては、主決定木により判断がなされる。また参照データについては、副決定木により判断がなされる。

なお、主決定木と副決定木とは、中間ノードM11およびM12を共通して備える。主決定木の中間ノードが示す特徴量の範囲は、参照データの判定の経路が当該中間ノードを含むか否かによって異なる。具体的には、参照データの判定の経路に中間ノードM11が含まれない場合、主決定木の中間ノードM11が示す中央値の範囲の閾値は、0.547となる。他方、参照データの判定の経路に中間ノードM11が含まれる場合、主決定木の中間ノードM11が示す中央値の範囲の閾値は、0.508(図9(B)の括弧内の数値)となる。

【0057】

第5の実施形態に係る説明文生成装置100は、第2の実施形態と説明文生成部104の動作が異なる。説明文生成部104は、第2の実施形態と同様に、分類部107によって特定される主決定木の経路に基づいて、主決定木の説明文を生成する。説明文生成装置100は、分類部107によって特定される主決定木の経路と副決定木の経路とに、共通の中間ノードが含まれる場合に、さらに当該副決定木の説明文を生成する。なお、当該副決定木の説明文は、最大でも根ノードから上記共通の中間ノードまでの経路についての説明文である。

つまり、説明文生成部104は、主決定木の経路に属するノードの中に副決定木によって特徴量の範囲が決定されるものが含まれる場合に、さらに当該副決定木の説明文を生成する。

これにより、説明文生成装置100は、複数の決定木を内包する決定ネットワークの説明文を適切に作成することができる。

【0058】

以上、図面を参照して一実施形態について詳しく説明してきたが、具体的な構成は上述のものに限られることはなく、様々な設計変更等を行うことが可能である。

例えば、上述した実施形態では、説明文生成装置100が一般的な決定木および決定ネットワークの説明文を作成する場合について説明したが、これに限られない。例えば、他の実施形態では、説明文生成装置100がファジイ決定木の説明文を生成しても良い。

【0059】

また、上述した実施形態では、説明文生成装置100が日本語の説明文を生成する場合について説明したが、これに限られない。例えば、他の実施形態では、説明文生成装置100が英語、中国語、ロシア語、またはその他の言語の説明文を生成しても良い。

【0060】

また、上述した実施形態では、説明文生成装置100が詳細度および文章数の入力を受け付ける場合について説明したが、これに限られない。例えば、他の実施形態では、説明文生成装置100は、予め定められた詳細度および文章数の説明文を作成しても良い。また他の実施形態では、説明文生成装置100は、詳細度に代えて説明対象のノード数の入力を受け付けても良い。また他の実施形態では、説明文生成装置100は、文章数に代えて必要な網羅率の入力を受け付けても良い。この場合、説明文生成装置100は、入力された網羅率を満たすまで、ステップS8~ステップS13のループを繰り返し実行する。

【0061】

10

20

30

40

50

図10は、少なくとも1つの実施形態に係るコンピュータの構成を示す概略ブロック図である。

コンピュータ901は、CPU902、主記憶装置903、補助記憶装置904、インタフェース905を備える。

上述の説明文生成装置100は、コンピュータ901に実装される。そして、上述した各処理部の動作は、プログラムの形式で補助記憶装置904に記憶されている。CPU902は、プログラムを補助記憶装置904から読み出して主記憶装置903に展開し、当該プログラムに従って上記処理を実行する。また、CPU902は、プログラムに従って、上述した各記憶部に対応する記憶領域を主記憶装置903に確保する。

【0062】

10

なお、少なくとも1つの実施形態において、補助記憶装置904は、一時的でない有形の媒体の一例である。一時的でない有形の媒体の他の例としては、インタフェース905を介して接続される磁気ディスク、光磁気ディスク、CD-ROM、DVD-ROM、半導体メモリ等が挙げられる。また、このプログラムが通信回線によってコンピュータ901に配信される場合、配信を受けたコンピュータ901が当該プログラムを主記憶装置903に展開し、上記処理を実行しても良い。

【0063】

また、当該プログラムは、前述した機能の一部を実現するためのものであっても良い。さらに、当該プログラムは、前述した機能を補助記憶装置904に既に記憶されている他のプログラムとの組み合わせで実現するもの、いわゆる差分ファイル（差分プログラム）であってても良い。

20

【符号の説明】

【0064】

- 100 説明文生成装置
- 101 決定木取得部
- 102 説明記憶部
- 103 条件入力部
- 104 説明文生成部
- 105 出力部
- 106 対象データ入力部
- 107 分類部

30

【 図 1 】

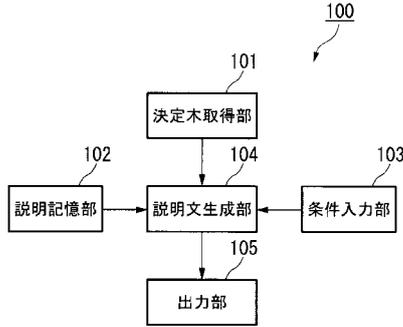


図 1

【 図 2 】

特徴量	単語
黄色周辺の色相の画素数/全画素数	黄色っぽい部分が
緑色周辺の色相の画素数/全画素数	緑色っぽい部分が
水色周辺の色相の画素数/全画素数	水色っぽい部分が
青色周辺の色相の画素数/全画素数	青色っぽい部分が
紫色周辺の色相の画素数/全画素数	紫色っぽい部分が
赤色周辺の色相の画素数/全画素数	赤色っぽい部分が
色相の分散	色数が
彩度の平均値	鮮やかさが
明度の平均値	明るさが
垂直線±20度のエッジをもつ画素数/全画素数	縦線が
水平線±20度のエッジをもつ画素数/全画素数	横線が
縦線・横線以外のエッジをもつ画素数/全画素数	斜線が

図 2

【 図 4 】

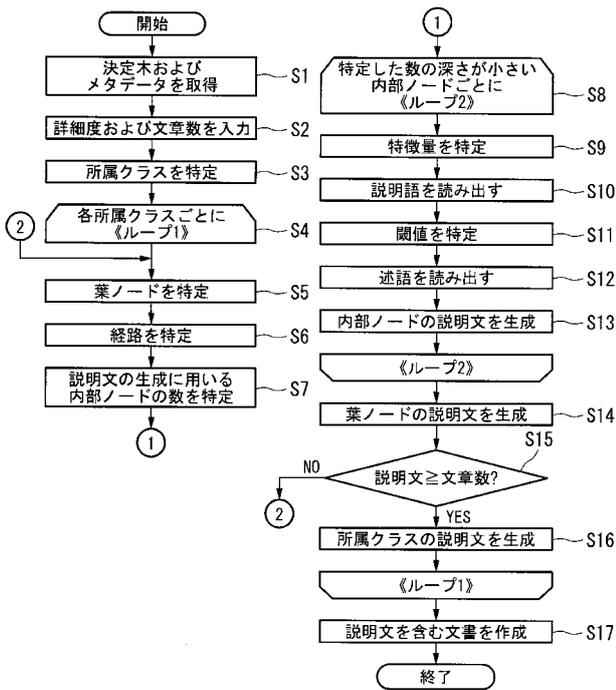


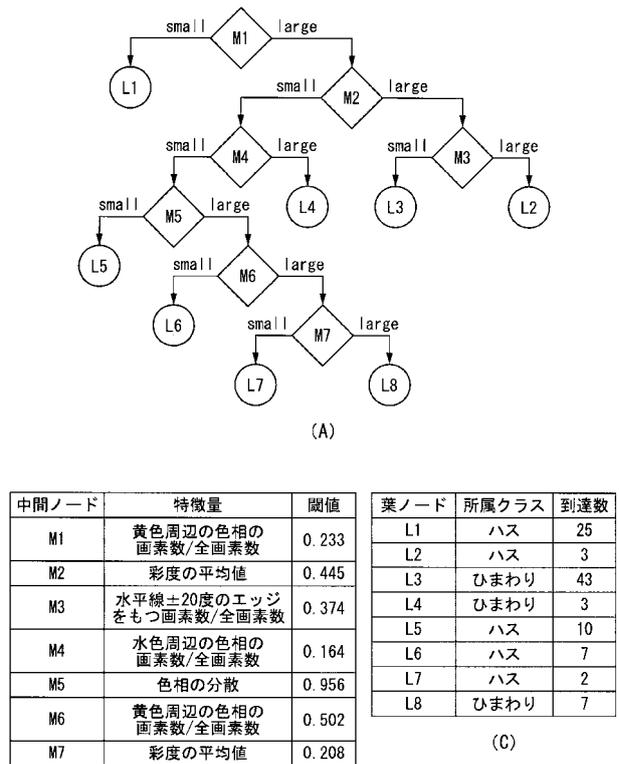
図 4

【 図 3 】

閾値	閾値未満の範囲の述語	閾値以上の範囲の述語
$0.0 \leq th < 0.2$	あまりない	あまりないわけではない
$0.2 \leq th < 0.4$	少ない	少ないわけではない
$0.4 \leq th < 0.6$	ある程度多いわけではない	ある程度多い
$0.6 \leq th < 0.8$	多いわけではない	多い
$0.8 \leq th \leq 1.0$	とても多いわけではない	とても多い

図 3

【 図 5 】



(B)

図 5

【 図 6 】

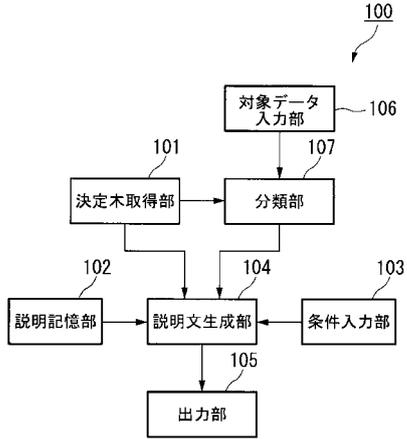


図 6

【 図 7 】

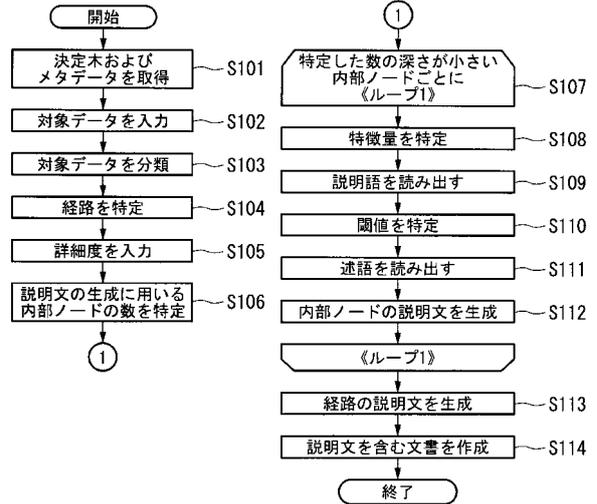


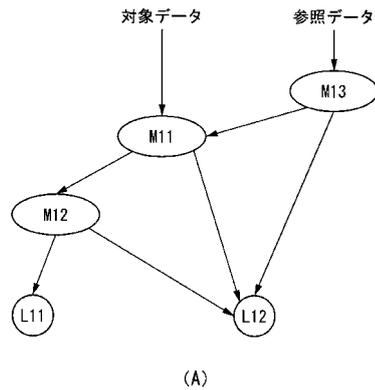
図 7

【 図 8 】

特徴量	単語	説明適性値
色相の分散	色数が	0.7
彩度の平均値	鮮やかさが	0.9
明度の平均値	明るさが	0.9
色相の平均値	色相の平均値が	0.1
縦軸の空間周波数の平均値	縦方向の明るさの変化の頻度が	0.3
横軸の空間周波数の平均値	横方向の明るさの変化の頻度が	0.3

図 8

【 図 9 】



(A)

中間ノード	特徴量	閾値
M11	中央値	0.547 (0.508)
M12	最頻値	0.664
M13	平均値	0.508

(B)

葉ノード	所属クラス
L11	傷なし
L12	傷あり

(C)

図 9

【 図 1 0 】

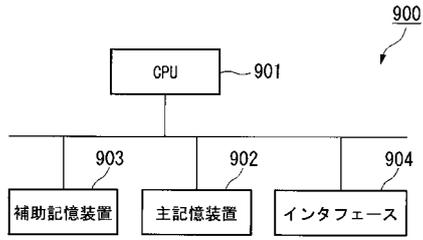


図 1 0

フロントページの続き

- (72)発明者 齊藤 航太
神奈川県横浜市保土ヶ谷区常盤台7-9番1号 国立大学法人横浜国立大学内
- (72)発明者 菅沼 雅徳
神奈川県横浜市保土ヶ谷区常盤台7-9番1号 国立大学法人横浜国立大学内
- (72)発明者 土屋 大樹
神奈川県横浜市保土ヶ谷区常盤台7-9番1号 国立大学法人横浜国立大学内
- (72)発明者 崎津 実穂
神奈川県横浜市保土ヶ谷区常盤台7-9番1号 国立大学法人横浜国立大学内