

(19) 日本国特許庁(JP)

再公表特許(A1)

(11) 国際公開番号

W02019/189249

発行日 令和3年2月12日(2021.2.12)

(43) 国際公開日 令和1年10月3日(2019.10.3)

(51) Int.Cl.

G06N 20/00 (2019.01)

F I

G06N 20/00 160

テーマコード(参考)

審査請求 有 予備審査請求 未請求 (全 27 頁)

出願番号 特願2020-510958 (P2020-510958)
 (21) 国際出願番号 PCT/JP2019/012984
 (22) 国際出願日 平成31年3月26日(2019.3.26)
 (31) 優先権主張番号 特願2018-66057 (P2018-66057)
 (32) 優先日 平成30年3月29日(2018.3.29)
 (33) 優先権主張国・地域又は機関
 日本国(JP)

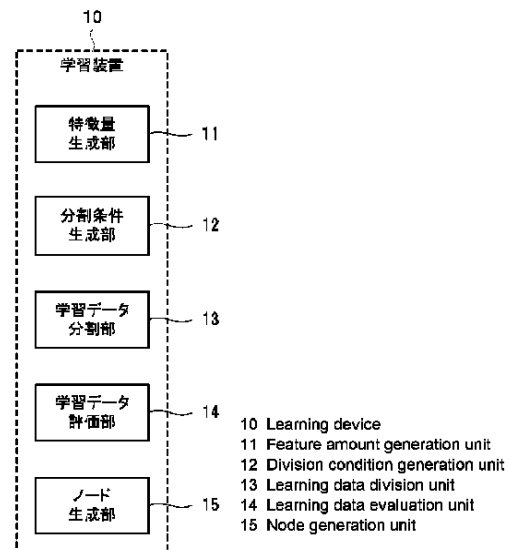
(71) 出願人 000004237
 日本電気株式会社
 東京都港区芝五丁目7番1号
 (74) 代理人 110002044
 特許業務法人ブライタス
 (72) 発明者 中野谷 学
 東京都港区芝五丁目7番1号 日本電気株
 式会社内

最終頁に続く

(54) 【発明の名称】 学習装置、学習方法、及びプログラム

(57) 【要約】

学習データに基づいて特徴量を生成する、特徴量生成部11と、特徴量と特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、分割条件生成部12と、学習データを分割条件に基づいてグループに分割する、学習データ分割部13と、分割前のグループと分割後のグループとを用いて、分割条件ごとに有意性を評価する、学習データ評価部14と、分割後のグループにおいて、分割条件に有意性がある場合、分割条件に関連する分割条件決定木のノードを生成する、ノード生成部15と、を有する学習装置10である。



【特許請求の範囲】

【請求項 1】

学習データに基づいて特徴量を生成する、特徴量生成部と、
 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、分割条件生成部と、
 前記学習データを前記分割条件に基づいてグループに分割する、学習データ分割部と、
 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、学習データ評価部と、
 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ノード生成部と、
 を有することを特徴とする学習装置。

10

【請求項 2】

請求項 1 に記載の学習装置であって、
 前記分割前後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件生成部に前記分割条件を追加させる、分割条件追加部と、
 を有することを特徴とする学習装置。

【請求項 3】

請求項 1 又は 2 に記載の学習装置であって、
 前記分割条件生成部は、前記特徴量間の関係性を表す論理演算子を用いて、前記分割条件を生成する
 ことを特徴とする学習装置。

20

【請求項 4】

請求項 3 に記載の学習装置であって、
 前記分割条件生成部は、前記複雑さ要件が示す、前記分割条件で使用する前記特徴量の数 (F 1 、 F 2) が二つである場合、前記分割条件を、以下の条件を用いて生成する

F 1 a n d F 2
 n o t F 1 a n d F 2
 F 1 o r F 2
 F 1 a n d n o t F 2
 F 1 x o r F 2

30

ことを特徴とする学習装置。

【請求項 5】

(a) 学習データに基づいて特徴量を生成する、ステップと、
 (b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ステップと、
 (c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、
 (d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと、
 、
 (e) 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、
 を有することを特徴とする学習方法。

40

【請求項 6】

請求項 5 に記載の学習方法であって、
 (f) 前記分割前後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件を追加させる、ステップと、
 、
 を有することを特徴とする学習方法。

【請求項 7】

請求項 5 又は 6 に記載の学習方法であって、

50

前記 (b) のステップにおいて、前記特徴量間の関係性を表す論理演算子を用いて、前記分割条件を生成する

ことを特徴とする学習方法。

【請求項 8】

請求項 7 に記載の学習方法であって、

前記 (b) のステップにおいて、前記複雑さ要件が示す、前記分割条件で使用する前記特徴量 (F 1 、 F 2) の数が二つである場合、前記分割条件を、以下の条件を用いて生成する

F 1 a n d F 2
 n o t F 1 a n d F 2
 F 1 o r F 2
 F 1 a n d n o t F 2
 F 1 x o r F 2

10

ことを特徴とする学習方法。

【請求項 9】

コンピュータに、

(a) 学習データに基づいて特徴量を生成する、ステップと、
 (b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ステップと、
 (c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、
 (d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと、
 (e) 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、

20

を実行させる命令を含む、プログラムを記録しているコンピュータ読み取り可能な記録媒体。

【請求項 10】

請求項 9 に記載のコンピュータ読み取り可能な記録媒体であって、

前記プログラムが、前記コンピュータに、

(f) 前記分割後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件を追加させる、ステップを実行させる命令を更に含む、

30

プログラムを記録しているコンピュータ読み取り可能な記録媒体。

【請求項 11】

請求項 9 又は 10 に記載のコンピュータ読み取り可能な記録媒体であって、

前記 (b) のステップにおいて、前記特徴量間の関係性を表す論理演算子を用いて、前記分割条件を生成する

ことを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項 12】

請求項 11 に記載のコンピュータ読み取り可能な記録媒体であって、

前記 (b) のステップにおいて、前記複雑さ要件が示す、前記分割条件で使用する前記特徴量 (F 1 、 F 2) の数が二つである場合、前記分割条件を、以下の条件を用いて生成する

40

F 1 a n d F 2
 n o t F 1 a n d F 2
 F 1 o r F 2
 F 1 a n d n o t F 2
 F 1 x o r F 2

ことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

50

【技術分野】

【0001】

本発明は、決定木による学習をする学習装置、学習方法に関し、更には、これらを実現するためのプログラムを記録したしているコンピュータ読み取り可能な記録媒体に関する。

【背景技術】

【0002】

IT (Information Technology) システムにおいて、システムの構成管理及び構成変更は、大きく三つのフェーズに分けられる。そして、システムの構成管理及び構成変更は、三つのフェーズそれぞれにおいて行われ、次に示す(1)(2)(3)の作業を反復することにより実現される。

10

【0003】

(1) システム構成を把握する作業。(2) 変更要件の定義をする作業。(3) 現在運用中のシステム構成を、(1)(2)から導かれるシステムへ変更する操作手順の生成、及び、生成した操作手順を実行する作業。

【0004】

ところが、これら三つの作業のうち、(3)に示した作業には、多大な作業工数を費やしている。そこで、作業工数を削減するための技術が提案されている。

【0005】

関連する技術として、特許文献1には、システムを構成する要素の動作状態と、その動作状態間の制約とを定義することで、システムの変更用いる操作手順を生成する技術が開示されている。

20

【0006】

特許文献2には、部品の状態と制約関係とを状態遷移図により表現する技術が開示されている。

【0007】

特許文献3には、決定木の学習を実行する前に、パラメータ間の相互作用を検証することで、依存関係がありそうなパラメータと、そうでないパラメータとを識別し、分割条件候補となるパラメータのセットを絞り込む手法が開示されている。

【0008】

非特許文献1及び特許文献2には、操作手順を自動化するソフトウェアツールが開示されている。そのソフトウェアツールによれば、システム変更後の状態、又は操作手順を定義情報として入力し、自動的にシステムの変更、及び設定を行う。

30

【0009】

非特許文献3及び4には、サーバ装置のリソース(CPU (Central Processing Unit) やメモリ割当て量など)、アプリケーションの組み合わせを実際に様々なパターンで試行し評価、学習することで最適な変更手順、又は変更パラメータを導出する強化学習を用いた技術が開示されている。

【先行技術文献】

【特許文献】

40

【0010】

【特許文献1】特開2015-215885号公報

【特許文献2】特開2015-215887号公報

【特許文献3】特開2005-063353号公報

【非特許文献】

【0011】

【非特許文献1】"Puppet"、[online]、[平成29年1月19日検索]、インターネット<URL: <https://puppet.com/>>

【非特許文献2】"Ansible"、[online]、[平成29年1月19日検索]、インターネット<URL: <https://www.ansible.com/>>

50

【非特許文献3】J. Rao, X. Bu, C. Z. Xu and K. Wang, "A Distributed Self-Learning Approach for Elastic Provisioning of Virtualized Cloud Resources,"、[online]、[平成29年1月19日検索]平成23年8月30日、IEEE Xplore、[平成29年1月19日検索]、インターネット<URL:http://ieeexplore.ieee.org/abstract/document/6005367/>

【非特許文献4】I. J. Jureta, S. Faulkner, Y. Achbany and M. Saerens, "Dynamic Web Service Composition within a Service-Oriented Architecture,"、[online]、[平成29年1月19日検索]平成19年7月30日、IEEE Xplore、[平成29年1月19日検索]、インターネット<URL:http://ieeexplore.ieee.org/document/4279613/>

【発明の概要】

【発明が解決しようとする課題】

10

【0012】

しかしながら、非特許文献1及び非特許文献2に開示されている、操作手順を自動化するソフトウェアツールが自動化できるのは、操作手順の実行のみで、操作手順の作成は対象外である。

【0013】

そこで、非特許文献1又は非特許文献2に、操作手順の生成を自動化する、特許文献1又は特許文献2に開示の技術を適用することが考えられる。すなわち、特許文献1又は特許文献2に開示の技術を用いて、操作手順の実行を自動化するソフトウェアツールの入力の形式に合わせた、システムの構成を変更する操作手順を示す情報を生成する。そして、生成した操作手順を、非特許文献1又は非特許文献2に開示の技術に適用することで、操作手順の作成から実行までを自動化する。

20

【0014】

ところが、特許文献1及び特許文献2に開示された技術では、事前に(1)システム構成を把握する作業、(2)変更要件の定義をする作業を、手作業で行う必要があるため、多大な作業工数を費やすことになるという問題がある。

【0015】

そこで、上述した問題に対して、非特許文献3又は非特許文献4に開示された技術を用いることが考えられる。つまり、サーバ装置のリソース(例えば、CPU、メモリ割当て量など)、又はアプリケーションの組み合わせを、実際に様々なパターンで試行し評価、学習することで、操作手順、及びパラメータを導出することが考えられる。

30

【0016】

しかし、上述した非特許文献3、4に開示された強化学習を用いた自動化は、特許文献1、2に開示されているような、システム内の構成要素間の依存関係を直接扱うアプローチと異なり、評価、学習の対象は、あるシステムの状態における特定の制御内容の好ましさである。制御内容は、例えば、システムの応答速度などの観測可能な値によって規定される。

【0017】

そのため、強化学習では、システムの状態を観測する手段と、実行可能な制御集合とを入力するだけで学習が実行可能となるため、適用が比較的容易である。ところが、強化学習では、学習結果から、依存性をはじめとした構成要素間の挙動に関する関係性を、一般には読み取ることができない。そのため、学習結果を別の制御タスクへ再利用することが困難である。

40

【0018】

そこで、こうした問題への対処方法として、強化学習にいわゆる関数近似を適用することが考えられる。強化学習における関数近似とは、学習の結果得られた具体的な制御に関する好ましさを示す情報を、より抽象的な条件から予測できる近似関数を求めることである。すなわち、抽象的な条件から予測できる近似関数を学習することである。

【0019】

上述した対処方法は、元々はロボット制御などの分野において、連続量(無限に選択肢が存在する)の制御を取り扱う際、すべての制御パターンをコンピュータの記憶領域で管

50

理することが不可能なため、無限集合を有限集合へマッピングし、その有限集合における制御パターンを取り扱うために発展してきた手法である。また、上述した対処方法は、記憶領域の問題だけでなく、広大かつ雑多な選択肢を適切に抽象化することで学習結果の汎用性を高めることもできる。

【 0 0 2 0 】

関数近似で用いる近似関数は、近似対象の性質や近似の目的に合わせて選択する必要がある。代表的な関数としては、線形多項式、又はニューラルネットワーク、又は決定木などが挙げられる。

【 0 0 2 1 】

しかし、システムの設計、又は制御内容からその設計、制御の良し悪しを予測する問題においては、決定木による関数近似が有力な近似手法の一つと考えられる。その理由として、まず、パラメータ間に依存関係があることが挙げられる。すなわち、あるパラメータの最適値は、他のパラメータの値に依存して異なる値となるからである。また、非線形な挙動への対応ができることが挙げられる。設定値の些細な違いが好ましさに大きく影響するためである。更に、生成された関数の解釈性の高さが優れているからである。すなわち、その関数が制御特性をより良く表現できているかを、人間が評価できるためである。

10

【 0 0 2 2 】

決定木の代表的な学習として C 4 . 5、C A R T (Classification And Regression Trees)、C H A I D (Chi-squared Automatic Interaction Detection) などがある。これらは、木の分割条件を選択する際、使用する指標が異なることが特徴である。例えば、C 4 . 5 では、分割条件で分割したデータが、分割前のデータと比較してよりエントロピーを減少させる分割条件を採用する。

20

【 0 0 2 3 】

決定木の学習により生成される分割条件は、設計や制御に関する単一のパラメータにより規定される論理式で表される。具体的に説明する。通信帯域、及び C P U コア数といった二つのパラメータを調整することで、アプリケーションサーバのスループットを最適化するというタスクの場合、学習された決定木の節 (ノード) に関連する分割条件は、例えば、「通信帯域 < 1 0 M b p s」、及び「C P U 数 > 1」などが考えられる。

【 0 0 2 4 】

また、パラメータが他のパラメータに依存する場合、分割条件の分割先で依存するパラメータに関する分割条件を採用することになる。例えば、「通信帯域 1 0 M b p s」の場合、C P U コア数がボトルネックとなる。また、C P U コア数がスループットに影響しないシステムの場合、決定木の頂点ノードで「通信帯域 < 1 0 M b p s」の分割条件を設定し、分割先のノードで C P U コア数に関する分割条件を定義する。

30

【 0 0 2 5 】

しかし、決定木の学習では、単一のパラメータごとにどれだけ学習データをうまく分類できるかを評価して分割条件を決定するため、複数のパラメータ間に依存関係がある場合、分割条件を適切に設定できないことがある。例えば、上述した通信帯域、及び C P U コア数といったパラメータに加え、更にメモリサイズなどの単一のパラメータが制御の対象である場合、分割条件を適切に設定できない。具体的には、メモリサイズが、見かけ上最もスループットへ影響を与えるパラメータだったとすると、分割条件としてメモリサイズに関するものが採用されてしまう。

40

【 0 0 2 6 】

そうすると、分割後の学習データは、メモリサイズによる分割条件で分断されてしまい、分断後の学習データそれぞれにおいて、上述したような通信帯域及び C P U コア数の依存関係に従った分割条件が導かれる保証はない。このような問題は、パラメータ間の依存関係の実体が排他的論理和であるときに顕著である。

【 0 0 2 7 】

図 1 は、学習データの一例を示す図である。図 1 に示す「A」「B」「C」「D」は、パラメータ (真 : 1、偽 : 0 のバイナリ値) を示している。また、「Y」は、近似したい

50

値（予測値）を示している。具体的には、予測値 Y は、パラメータ A 、 B の排他的論理和（真：1、偽：0）を10倍した実数値に $[0, 1]$ 区間の一様乱数を加えた値である。なお、パラメータ C 、 D は、実際には予測には全く影響しないパラメータである。なお、「id」「1」から「8」は、パラメータ A から D と予測値 Y とを有する行それぞれに付された識別番号である。

【0028】

従って、図1に示す学習データを用いて生成される決定木は、パラメータ C 、 D を分割条件に含めない、図2のような決定木になるのが理想的である。図2は、理想的な決定木の一例を示す図である。ところが、既存の決定木の学習を用いて生成した決定木は、図3のような決定木になる。図3は、既存の決定木の学習により生成された決定木の一例を示す図である。

10

【0029】

既存の決定木の学習では、単一のパラメータで評価するため、図2に示した決定木と比較すると、図3に示した決定木には不要な分割条件を含むため、予測精度の低い決定木が生成されてしまう。言い換えれば、本質的な分割条件が全体に適用されていない、複雑な決定木が生成されてしまう。

【0030】

具体的には、パラメータ C は予測値 Y に影響を与えないパラメータであるにもかかわらず、最も予測値と相関があるパラメータであるため、最上位の分割条件となる。そのため、図3に示す左側（False： $C = 1$ ）の部分木では、パラメータ A 、 B による排他的論理和を表す決定木が生成されたものの、図3に示す右側（True： $C = 1$ ）の部分木では、パラメータ A 、 B による排他的論理和を表す部分木が生成されない。

20

【0031】

そこで、特許文献3を用いることが考えられる。特許文献3では、決定木の学習を実行する前に、パラメータ間の相互作用を検証することで、依存関係がありそうなパラメータと、そうでないパラメータとを識別し、分割条件候補となるパラメータセットを絞り込んでいる。しかし、特許文献3は、決定木の学習前において、パラメータの品質を安定化させることを目的としており、上述した問題を解決するものではない。

【0032】

本発明の目的の一例は、決定木の予測精度を向上させる学習装置、学習方法、及びコンピュータ読み取り可能な記録媒体を提供することにある。

30

【課題を解決するための手段】

【0033】

上記目的を達成するため、本発明の一側面における学習装置は、
学習データに基づいて特徴量を生成する、特徴量生成部と、
前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、分割条件生成部と、

前記学習データを前記分割条件に基づいてグループに分割する、学習データ分割部と、
分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、学習データ評価部と、

40

分割前後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ノード生成部と、

を有することを特徴とする。

【0034】

また、上記目的を達成するため、本発明の一側面における学習方法は、

- (a) 学習データに基づいて特徴量を生成する、ステップと、
- (b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ステップと、
- (c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、
- (d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと

50

、
 (e) 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、
 を有することを特徴とする。

【0035】

更に、上記目的を達成するため、本発明の一側面におけるプログラムをコンピュータ読み取り可能な記録媒体は、

コンピュータに、

- (a) 学習データに基づいて特徴量を生成する、ステップと、
- (b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ステップと、
- (c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、
- (d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと

10

、
 (e) 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、
 を実行させる命令を含むプログラムを記録していることを特徴とする。

【発明の効果】

【0036】

以上のように本発明によれば、決定木の予測精度を向上させることができる。

20

【図面の簡単な説明】

【0037】

【図1】図1は、学習データの一例を示す図である。

【図2】図2は、理想的な決定木の一例を示す図である。

【図3】図3は、既存の決定木の学習により生成された決定木の一例を示す図である。

【図4】図4は、学習装置の一例を示す図である。

【図5】図5は、学習装置を有するシステムの一例を示す図である。

【図6】図6は、複雑さ要件に対する分割条件の一例を示す図である。

【図7】図7は、分割結果の一例を示す図である。

【図8】図8は、評価結果の一例を示す図である。

30

【図9】図9は、評価結果の一例を示す図である。

【図10】図10は、学習データの動作の一例を示す図である。

【図11】図11は、学習装置を実現するコンピュータの一例を示す図である。

【発明を実施するための形態】

【0038】

(実施の形態)

以下、本発明の実施の形態について、図1から図11を参照しながら説明する。

【0039】

[装置構成]

最初に、図4を用いて、本実施の形態における学習装置10の構成について説明する。
 図4は、学習装置の一例を示す図である。

40

【0040】

図4に示すように、学習装置10は、決定木の予測精度を向上させるための装置である。学習装置10は、特徴量生成部11と、分割条件生成部12と、学習データ分割部13と、学習データ評価部14と、ノード生成部15とを有する。

【0041】

このうち、特徴量生成部11は、学習データに基づいて特徴量を生成する。分割条件生成部12は、特徴量と特徴量の数を示す複雑さ要件とに応じて分割条件を生成する。学習データ分割部13は、学習データを分割条件に基づいてグループに分割する。学習データ評価部14は、分割前のグループと分割後のグループとを用いて、分割条件ごとに有意性

50

を評価する。ノード生成部 15 は、分割前後のグループにおいて、分割条件に有意性がある場合、分割条件に関連する分割条件決定木のノードを生成する。

【0042】

このように、本実施の形態では、特徴量と複雑さ要件とに応じて生成した分割条件を用いて、学習データをグループに分割して、分割前のグループと分割後のグループとを用いて、分割条件ごとに有意性を評価する。そして、分割前後のグループにおいて、分割条件に有意性がある場合、分割条件に関連する分割条件決定木のノードを生成する。そうすることで、決定木には不要な分割条件を含まない、予測精度の高い決定木を生成することができる。言い換えれば、本質的な分割条件を適用した決定木が生成できる。

【0043】

続いて、図5を用いて、本実施の形態における学習装置10の構成をより具体的に説明する。図5は、学習装置を有する学習システムの一例を示す図である。

【0044】

図5に示すように、本実施の形態における学習装置10は、特徴量生成部11と、分割条件生成部12と、学習データ分割部13と、学習データ評価部14と、ノード生成部15と、分割条件追加部16とを有する。

【0045】

また、図5においては、システムは、学習装置10に加え、学習装置10へ学習データ20を入力するための入力装置30と、学習装置10が生成した決定木データ50を出力するための出力装置40とを有する。学習データ20は、決定木を生成するシステムに入力する、設計ルールを表現したデータである。

【0046】

特徴量生成部11は、入力装置30を介して学習データ20を取得した後、学習データ20に基づいて、分割条件の要素となる特徴量(抽象的な特徴量)を生成する。その後、特徴量生成部11は、生成した特徴量に基づいて、学習データ20を変換する。

【0047】

具体的には、図1に示した学習データを変換後の学習データとした場合、パラメータA、B、C、Dが特徴量(抽象的な特徴量)となり、A列からD列までの各行の値は、その特徴量に関する元の学習データの評価値を表している。そして、図1において、一行目の学習データに該当する、変換前の学習データが「サーバ装置MのCPU数:1」「サーバ装置NのCPU数:3」「サーバ装置Mの通信帯域:2」「サーバ装置Nの通信帯域:1」で、抽象的な特徴量Aが「サーバ装置MのCPU数>サーバ装置NのCPU数」であるとする。その場合において、当該学習データは特徴量Aの評価値として、特徴量Aで示される論理式を満たさない(1<3)ためFalse(0)という評価値を得る。なお、上述したサーバ装置Mの通信帯域「2」、及びサーバ装置Nの通信帯域「1」は、例えば、通信帯域に割り付けられた番号を示す。

【0048】

このように、サーバ装置間のCPU数の大小比較という特徴量Aは、具体的な設計値ではなく、パラメータの相対的な関係性を示した例である。そのため、この考え方に基づいて、CPU数に限定されずにIPアドレス、通信帯域、メモリ割り当て数など様々な設計、パラメータを相対的な関係性で評価できる。なお、予測値Yは、元の学習データと同じとし、変換は行わない。

【0049】

分割条件生成部12は、学習データに基づいて生成された特徴量と、指定された複雑さ要件とに応じて、分割条件(具体的な分割条件)を生成する。複雑さ要件は、一つの分割条件に使用する特徴量の数を示し、初期値は1である。また、段階的に複雑さを増やす場合、複雑な条件に上限値も設定する。例えば、上限値は2とすることが考えられる。

【0050】

また、具体的な分割条件は、複雑さ要件が1の場合、図1の学習データの分割条件は、「A=True(1)/B=True(1)/C=True(1)/D=True(1)

10

20

30

40

50

」の四つとなる。また、複雑さ要件が2の場合、分割条件は特徴量を二つ含む論理式となる。

【0051】

図6は、複雑さ要件に対する分割条件を示す図である。図6には、複雑さ要件が2の場合(図6の分割条件60)に、図1の学習データに対して生成される、分割条件61が示されている。すなわち、図1に示した特徴量A、B、C、Dから二つを選び、分割条件60に示した五つの条件(F1 and F2、not F1 and F2、F1 or F2、F1 and not F2、F1 xor F2)を適用して、図6に示す30通り(4C2×5通り)の分割条件61を生成する。

【0052】

また、複雑さ要件が3以上の場合、複雑さ要件の数だけ、特徴量を含んだ論理式を生成する。なお、初期動作では複雑さ要件の初期値に従い、上述した四つの分割条件「A = True(1) / B = True(1) / C = True(1) / D = True(1)」が生成される。

【0053】

学習データ分割部13は、学習データと分割条件とを取得した後、学習データを分割条件に従い分割する。学習データの分割は、例えば、図1に示す学習データを、複雑さ要件が1の分割条件「A = True(1) / B = True(1) / C = True(1) / D = True(1)」に従い分割した場合、図7に示すような分割結果70となる。図7は、分割結果の一例を示す図である。

【0054】

学習データ評価部14は、分割結果を取得した後、分割結果がどの程度学習データをよりよく分割できるかを評価する。評価は、分割前後のグループ間の予測値の分散に統計的な有意差があるか否かを評価する。すなわち、分割前後のグループに対して等分散検定を行い、予め設定された基準となる有意水準を用いて算出される有意水準において、分割前後のグループの分散が等しいという帰無仮説が棄却できる場合、その分割条件を有効な分割条件と見做して決定木の枝の分割条件とする。

【0055】

なお、上述したような単一論式による二分木の場合、分割後のグループは二つ生成されるため、実施される等分散検定は、分割前対分割後の二つのグループがあり、どちらかの検定が有意であれば、その分割条件は有効と見做す。

【0056】

また、有効な分割条件が複数検出された場合、等分散検定におけるp値が、最小の分割条件を実際の決定木の分割条件として採用する。等分散検定は、予測値の確率分布に関する仮定の違いなどによりいくつかの手法が存在する。例えば、予測値に特定の確率分布を仮定しない場合にはCrown-Forsythe検定を用いる。なお、学習対象のデータの性質に応じて検定手法を選択してもよい。

【0057】

図8に、図7における分割結果に基づく評価結果を示す。図8は、評価結果の一例を示す図である。有意水準は、予め設定された基準となる有意水準を、検定実施回数で割った値とする。すなわち、これは等分散検定の繰り返しによる偽陽性の発生確率の増大に対応するための措置である。図8においては、基準となる有意水準を0.01とし、検定実施回数を4×2としているので、有意水準は0.01 / (4×2) = 0.00125となる。なお、この有意水準の設定は一例であり限定されない。

【0058】

分割条件追加部16は、評価結果を取得した後、すべての分割条件について有意性がない場合(p値が有意水準以上である場合)、より複雑な分割条件で再評価を行うため、複雑さ要件を増加する。

【0059】

具体的には、図8に示した評価結果80の場合、分割条件追加部16は、すべての分割

10

20

30

40

50

条件について有意性がないので、現状の複雑さ要件を増加させる。例えば、現状の複雑さ要件は1であるので、複雑さ要件を2とする。

【0060】

その後、分割条件生成部12は、更新された複雑さ要件に従い分割条件を再生成する。続いて、分割条件生成部12は、複雑さ要件が2であるので、図6に示した分割条件を生成する。以降、学習データ分割部13、学習データ評価部14は、新たな分割条件に対して、分割と評価とを実施する。

【0061】

図9は、評価結果の一例を示す図である。図9では、有意性が認められる分割条件が複数検出されているが、p値が最小の分割条件であるAとBとの排他的論理和「 $A \times o r B$ 」が最適な分割条件として採用される。

10

【0062】

また、学習データ評価部14は、最適な分割条件が検出された場合、その最適な分割条件はノード生成部15へ送られる。

【0063】

ノード生成部15は、その最適な分割条件に関連付けられた決定木のノードを一つ生成する。また、ノード生成部15は、そのノードにおける分割条件により分割されたグループを分割条件生成部12へ送る。なお、二分木の場合、二つのグループとなる。続いて、分割条件生成部12は、分割されたグループを受けた場合、複雑さ要件を初期値である1に設定する。そして、分割条件生成部12は、受け取ったグループを新たに分割前のグループと見做して上述した処理を継続する。

20

【0064】

更に、ノード生成部15は、有効な分割条件が検出されず、複雑さ要件を増加させていきその上限に達しても有効な分割条件が見つからない場合、その分割できなかったグループを末端ノードとしてノード生成の対象にする。図9に示した評価結果90の場合、分割後のグループ1(真)(5, 6, 7, 8)、グループ0(偽)(1, 2, 3, 4)それぞれは複雑さ要件の上限値である2まで、分割条件を評価しても有意な分割条件が検出されない。その場合、分割条件の生成を停止して、ノード生成部15は、当該グループを決定木の最下層ノード(葉)とする。

30

【0065】

そして、すべてのグループに関して最下層ノードの生成が完了した場合、ノード生成部15は、生成した決定木データ50を、出力装置40を介して出力する。その結果、図2に示した決定木が出力される。

【0066】

[装置動作]

次に、本発明の実施の形態における学習装置の動作について図10を用いて説明する。図10は、学習装置の動作の一例を示す図である。以下の説明においては、適宜図1から図9を参照する。また、本実施の形態では、学習装置を動作させることによって、学習方法が実施される。よって、本実施の形態における学習方法の説明は、以下の学習装置の動作説明に代える。

40

【0067】

ステップA1において、特徴量生成部11は、取得した学習データ20に基づいて、分割条件の要素となる特徴量(抽象的な特徴量)を生成する。その後、特徴量生成部11は、生成した特徴量に基づいて、学習データ20を変換する。

【0068】

ステップA2において、分割条件生成部12は、変換した学習データに含まれる特徴量と、指定された分割条件の複雑さ要件とに応じて、分割条件(具体的な分割条件)を生成する。ステップA3において、学習データ分割部13は、学習データと分割条件とを取得した後、学習データを分割条件に従い分割する。

【0069】

50

ステップ A 4 において、学習データ評価部 1 4 は、分割結果を取得した後、分割結果がどの程度学習データをよりよく分割できているか否かを評価する。例えば、分割前後のグループ間の予測値の分散に統計的な有意差があるか否かを評価する。

【 0 0 7 0 】

ステップ A 5 において、学習データ評価部 1 4 は、すべての分割条件について有意性があるか否かを判定する。有意性がない場合（ステップ A 5 : N o ）、ステップ A 7 において、分割条件追加部 1 6 は、複雑さ要件が上限値であるか否かを判定する。

【 0 0 7 1 】

有意性がある場合（ステップ A 5 : Y e s ）、又は、有意性がなく複雑さ要件が上限値である場合（ステップ A 7 : N o ）、ステップ A 6 において、ノード生成部 1 5 は、有意性がある分割条件に関連付けられた決定木のノードを生成する。

10

【 0 0 7 2 】

ステップ A 8 において、複雑さ要件が上限値でない場合（ステップ A 7 : N o ）、分割条件追加部 1 6 は、より複雑な分割条件で再評価を行うため、複雑さ要件を増加する。そして、増加した複雑さ要件において、再度ステップ A 2 から A 5 の処理を行う。なお、現状の複雑さ要件が 1 であれば、複雑さ要件を 2 とする。

【 0 0 7 3 】

ステップ A 9 において、ノード生成部 1 5 は、すべてのグループに対して最下層ノードを生成したか否かを判定する。すべてのグループの最下層ノードを生成した場合（ステップ A 9 : Y e s ）、この処理を終了する。すべてのグループの最下層ノードを生成していない場合（ステップ A 9 : N o ）、ステップ A 1 0 において、分割条件生成部 1 2 は、複雑さ要件を初期値である 1 に設定する。そして、分割条件生成部 1 2 は、分割されたグループに対して新たに処理を実行する。

20

【 0 0 7 4 】

[本実施の形態の効果]

以上のように本実施の形態によれば、特徴量と複雑さ要件とに応じて生成した分割条件を用いて、学習データをグループに分割する。そして、分割前のグループと分割後のグループとを用いて、分割条件ごとに有意性を評価する。その結果、分割前後のグループにおいて、分割条件に有意性がある場合、分割条件に関連する分割条件決定木のノードを生成する。そうすることで、決定木には不要な分割条件を含まない、予測精度の高い決定木を生成することができる。言い換えれば、本質的な分割条件を適用した決定木が生成できる。

30

【 0 0 7 5 】

[プログラム]

本発明の実施の形態におけるプログラムは、コンピュータに、図 1 0 に示すステップ A 1 から A 1 0 を実行させるプログラムであればよい。このプログラムをコンピュータにインストールし、実行することによって、本実施の形態における学習装置と学習方法とを実現することができる。この場合、コンピュータのプロセッサは、特徴量生成部 1 1、分割条件生成部 1 2、学習データ分割部 1 3、学習データ評価部 1 4、ノード生成部 1 5、分割条件追加部 1 6 として機能し、処理を行なう。

40

【 0 0 7 6 】

また、本実施の形態におけるプログラムは、複数のコンピュータによって構築されたコンピュータシステムによって実行されてもよい。この場合は、例えば、各コンピュータが、それぞれ、特徴量生成部 1 1、分割条件生成部 1 2、学習データ分割部 1 3、学習データ評価部 1 4、ノード生成部 1 5、分割条件追加部 1 6 のいずれかとして機能してもよい。

【 0 0 7 7 】

[物理構成]

ここで、実施の形態におけるプログラムを実行することによって、学習装置を実現するコンピュータについて図 1 1 を用いて説明する。図 1 1 は、学習装置を実現するコンピュ

50

ータの一例を示す図である。

【0078】

図11に示すように、コンピュータ110は、CPU111と、メインメモリ112と、記憶装置113と、入力インターフェイス114と、表示コントローラ115と、データリーダー/ライター116と、通信インターフェイス117とを備える。これらの各部分は、バス121を介して、互いにデータ通信可能に接続される。なお、コンピュータ110は、CPU111に加えて、又はCPU111に代えて、GPU (Graphics Processing Unit)、又はFPGA (Field-Programmable Gate Array) を備えていてもよい。

【0079】

CPU111は、記憶装置113に格納された、本実施の形態におけるプログラム(コード)をメインメモリ112に展開し、これらを所定順序で実行することにより、各種の演算を実施する。メインメモリ112は、典型的には、DRAM (Dynamic Random Access Memory) 等の揮発性の記憶装置である。また、本実施の形態におけるプログラムは、コンピュータ読み取り可能な記録媒体120に格納された状態で提供される。なお、本実施の形態におけるプログラムは、通信インターフェイス117を介して接続されたインターネット上で流通するものであってもよい。

10

【0080】

また、記憶装置113の具体例としては、ハードディスクドライブの他、フラッシュメモリ等の半導体記憶装置があげられる。入力インターフェイス114は、CPU111と、キーボード及びマウスといった入力機器118との間のデータ伝送を仲介する。表示コントローラ115は、ディスプレイ装置119と接続され、ディスプレイ装置119での表示を制御する。

20

【0081】

データリーダー/ライター116は、CPU111と記録媒体120との間のデータ伝送を仲介し、記録媒体120からのプログラムの読み出し、及びコンピュータ110における処理結果の記録媒体120への書き込みを実行する。通信インターフェイス117は、CPU111と、他のコンピュータとの間のデータ伝送を仲介する。

【0082】

また、記録媒体120の具体例としては、CF (Compact Flash (登録商標)) 及びSD (Secure Digital) 等の汎用的な半導体記憶デバイス、フレキシブルディスク (Flexible Disk) 等の磁気記録媒体、又はCD-ROM (Compact Disk Read Only Memory) などの光学記録媒体があげられる。

30

【0083】

なお、本実施の形態における学習装置1は、プログラムがインストールされたコンピュータではなく、各部に対応したハードウェアを用いることによっても実現可能である。更に、学習装置1は、一部がプログラムで実現され、残りの部分がハードウェアで実現されていてもよい。

【0084】

[付記]

以上の実施の形態に関し、更に以下の付記を開示する。上述した実施の形態の一部又は全部は、以下に記載する(付記1)から(付記12)により表現することができるが、以下の記載に限定されるものではない。

40

【0085】

(付記1)

学習データに基づいて特徴量を生成する、特徴量生成部と、
前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、分割条件生成部と、

前記学習データを前記分割条件に基づいてグループに分割する、学習データ分割部と、
分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、学習データ評価部と、

50

分割前後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ノード生成部と、
を有することを特徴とする学習装置。

【0086】

(付記2)

付記1に記載の学習装置であって、
前記分割前後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件生成部に前記分割条件を追加させる、分割条件追加部と、
を有することを特徴とする学習装置。

10

【0087】

(付記3)

付記1又は2に記載の学習装置であって、
前記分割条件生成部は、前記特徴量間の関係性を表す論理演算子を用いて、前記分割条件を生成することを特徴とする学習装置。

【0088】

(付記4)

付記3に記載の学習装置であって、
前記分割条件生成部は、前記複雑さ要件が示す、前記分割条件で使用する前記特徴量の数(F1、F2)が二つである場合、前記分割条件を、以下の条件を用いて生成する

20

F1 and F2
not F1 and F2
F1 or F2
F1 and not F2
F1 xor F2

ことを特徴とする学習装置。

【0089】

(付記5)

(a) 学習データに基づいて特徴量を生成する、ステップと、
(b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ステップと、
(c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、
(d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと、
(e) 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、
を有することを特徴とする学習方法。

30

【0090】

(付記6)

付記5に記載の学習方法であって、
(f) 前記分割前後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件を追加させる、ステップと、
を有することを特徴とする学習方法。

40

【0091】

(付記7)

付記5又は6に記載の学習方法であって、
前記(b)のステップにおいて、前記特徴量間の関係性を表す論理演算子を用いて、前記分割条件を生成する

50

ことを特徴とする学習方法。

【0092】

(付記8)

付記7に記載の学習方法であって、

前記(b)のステップにおいて、前記複雑さ要件が示す、前記分割条件で使用する前記特徴量(F1、F2)の数が二つである場合、前記分割条件を、以下の条件を用いて生成する

F 1 a n d F 2
 n o t F 1 a n d F 2
 F 1 o r F 2
 F 1 a n d n o t F 2
 F 1 x o r F 2

10

ことを特徴とする学習方法。

【0093】

(付記9)

コンピュータに、

(a) 学習データに基づいて特徴量を生成する、ステップと、

(b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ステップと、

(c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、

20

(d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと、

(e) 分割前後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、

を実行させる命令を含む、プログラムを記録しているコンピュータ読み取り可能な記録媒体。

【0094】

(付記10)

付記9に記載のコンピュータ読み取り可能な記録媒体であって、

前記プログラムが、前記コンピュータは、

30

(f) 前記分割前後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件を追加させる、ステップを実行させる命令を更に含む、

プログラムを記録しているコンピュータ読み取り可能な記録媒体。

【0095】

(付記11)

付記9又は10に記載のコンピュータ読み取り可能な記録媒体であって、

前記(b)のステップにおいて、前記特徴量間の関係性を表す論理演算子を用いて、前記分割条件を生成する

ことを特徴とするコンピュータ読み取り可能な記録媒体。

40

【0096】

(付記12)

付記11に記載のコンピュータ読み取り可能な記録媒体であって、

前記(b)のステップにおいて、前記複雑さ要件が示す、前記分割条件で使用する前記特徴量(F1、F2)の数が二つである場合、前記分割条件を、以下の条件を用いて生成する

F 1 a n d F 2
 n o t F 1 a n d F 2
 F 1 o r F 2
 F 1 a n d n o t F 2

50

F 1 x o r F 2

ことを特徴とするコンピュータ読み取り可能な記録媒体。

【0097】

以上、実施の形態を参照して本願発明を説明したが、本願発明は上記実施の形態に限定されるものではない。本願発明の構成や詳細には、本願発明のスコープ内で当業者が理解し得る様々な変更をすることができる。

【0098】

この出願は、2018年3月29日に出願された日本出願特願2018-066057を基礎とする優先権を主張し、その開示の全てをここに取り込む。

【産業上の利用可能性】

10

【0099】

以上のように本発明によれば、決定木の予測精度を向上させることができる。本発明は、決定木の予測精度を向上させることが必要な分野において有用である。

【符号の説明】

【0100】

- 10 学習装置
- 11 特徴量生成部
- 12 分割条件生成部
- 13 学習データ分割部
- 14 学習データ評価部
- 15 ノード生成部
- 16 分割条件追加部
- 20 学習データ
- 30 入力装置
- 40 出力装置
- 50 決定木データ
- 110 コンピュータ
- 111 CPU
- 112 メインメモリ
- 113 記憶装置
- 114 入力インターフェイス
- 115 表示コントローラ
- 116 データリーダー/ライター
- 117 通信インターフェイス
- 118 入力機器
- 119 ディスプレイ装置
- 120 記録媒体
- 121 バス

20

30

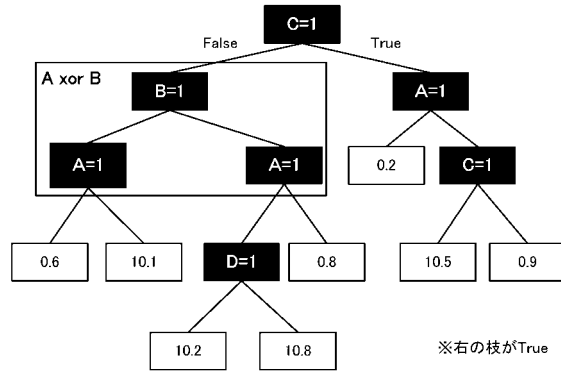
【 図 1 】

学習データ

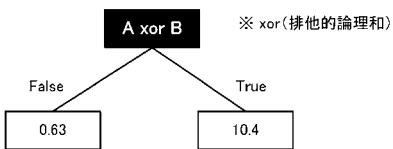
id	A	B	C	D	Y
1	0	0	1	0	0.2
2	0	0	0	0	0.6
3	1	1	0	1	0.8
4	1	1	1	1	0.9
5	1	0	1	0	10.5
6	1	0	0	1	10.1
7	0	1	0	1	10.8
8	0	1	0	0	10.2

※ 1: True(真)、0: False(偽)

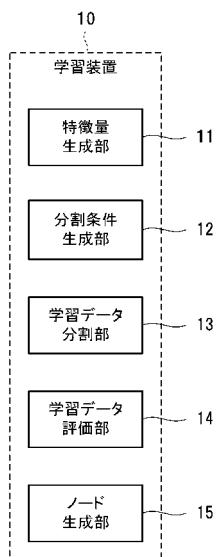
【 図 3 】



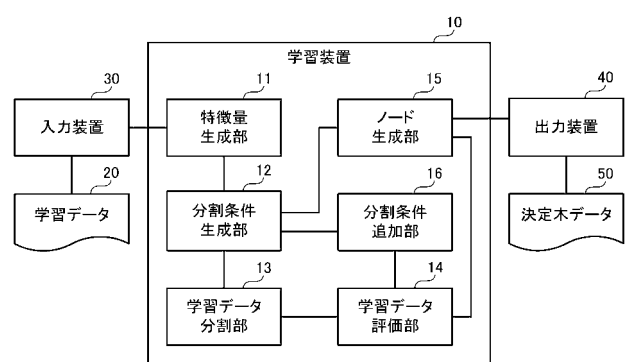
【 図 2 】



【 図 4 】



【 図 5 】



【 図 6 】

60:二つの特徴量(F1, F2)に依存して決定される分割条件

F2 \ F1		<F1 and F2>		<not F1 and F2>		<F1 or F2>		<F1 and not F2>		<F1 xor F2>	
		真	偽	真	偽	真	偽	真	偽	真	偽
真	偽	真	偽	真	偽	真	偽	真	偽	真	偽
偽	真	偽	真	偽	真	偽	真	偽	真	偽	真

61

A and B	not A and B	A or B	A and not B	A xor B
A and C	not A and C	A or C	A and not C	A xor C
A and D	not A and D	A or D	A and not D	A xor D
B and C	not B and C	B or C	B and not C	B xor C
B and D	not B and D	B or D	B and not D	B xor D
C and D	not C and D	C or D	C and not D	C xor D

【 図 7 】

70

分割条件	id	
	グループ 1(真)	グループ 0(偽)
A=1	3, 4, 5, 6	1, 2, 7, 8
B=1	3, 4, 7, 8	1, 2, 5, 6
C=1	1, 4, 5	2, 3, 6, 7, 8
D=1	3, 4, 6, 7	1, 2, 5, 8

【 図 8 】

80

分割条件	id		p値	有意性
	グループ 1(真)	グループ 0(偽)		
A=1	3, 4, 5, 6	1, 2, 7, 8	0.338	なし
B=1	3, 4, 7, 8	1, 2, 5, 6	0.717	なし
C=1	1, 4, 5	2, 3, 6, 7, 8	0.419	なし
D=1	3, 4, 6, 7	1, 2, 5, 8	0.605	なし

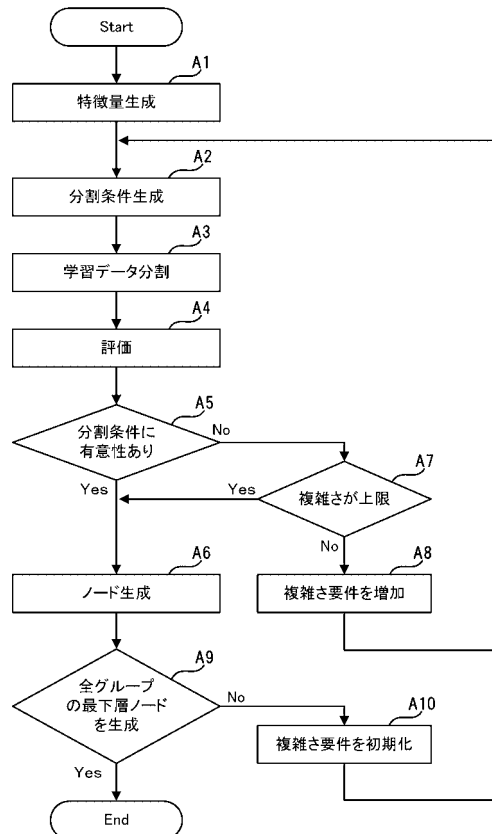
【 図 9 】

90

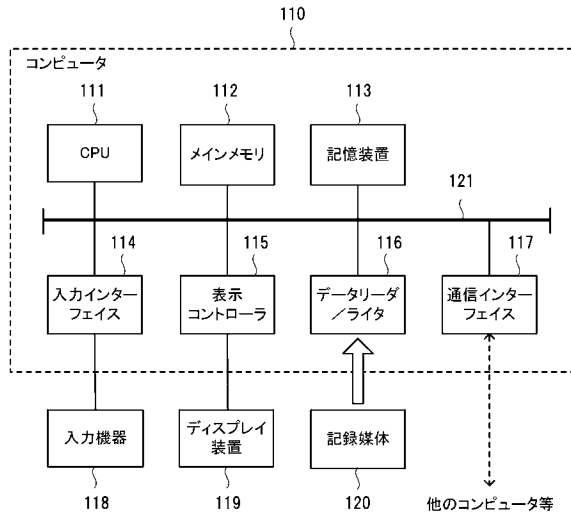
分割条件	id		p値	有意性
	グループ 1(真)	グループ 0(偽)		
A & B	3, 4	1,2,5,6,7,8	1.58E-8	あり
¬A & B	7,8	1,2,3,4,5,6	2.40E-8	あり
A & ¬B	5,6	1,2,3,4,7,8	2.03E-8	あり
A B	3,4,5,6,7,8	1,2	2.03E-8	あり
A xor B	5,6,7,8	1,2,3,4	5.82E-11	あり
A & C	4,5	1,2,3,6,7,8	0.694	なし
¬A & C	1	2,3,4,5,6,7,8	9.49E-7	あり
A & ¬C	3,6	1,2,4,5,7,8	0.300	なし
A or C	1,3,4,5,6	2,7,8	0.411	なし
A xor C	1,3,6	2,4,5,7,8	0.367	なし
A & D	3,4,6	1,2,5,7,8	0.319	なし
¬A & D	7	1,2,3,4,5,6,8	9.49E-7	あり
A & ¬D	5	1,2,3,4,6,7,8	9.49E-7	あり
A or D	3,4,5,6,7	1,2,8	0.395	なし
A xor D	5,7	1,2,3,4,6,8	1.86E-8	あり
B & C	4	1,2,3,5,6,7,8	9.49E-7	あり
¬B & C	1,5	2,3,4,6,7,8	0.256	なし
B & ¬C	3,7,8	1,2,4,5,6	0.381	なし
B or C	1,3,4,5,7,8	2,6	0.539	なし
B xor C	1,3,5,7,8	2,4,6	0.332	なし
B & D	3,4,7	1,2,5,6,8	0.416	なし
¬B & D	6	1,2,3,4,5,7,8	9.49E-7	あり
B & ¬D	8	1,2,3,4,5,6,7	9.49E-7	あり
B or D	2,4	1,3,5,6,7,8	0.439	なし
B xor D	6,8	1,2,3,4,5,7	1.58E-8	あり
C & D	4	1,2,3,5,6,7,8	9.49E-07	あり
¬C & D	3,6,7	1,2,4,5,8	0.374	なし
C & ¬D	1,5	2,3,4,6,7,8	0.256	なし
C or D	1,3,4,5,6,7	2,8	0.694	なし
C xor D	1,3,5,6,7	2,4,8	0.346	なし

※Eは指数表現(左の数値に10のn乗を掛ける)

【 図 10 】



【図 1 1】



【手続補正書】

【提出日】令和2年8月27日(2020.8.27)

【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】

学習データに基づいて特徴量を生成する、特徴量生成部と、
 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、分割条件生成部と、
 前記学習データを前記分割条件に基づいてグループに分割する、学習データ分割部と、
 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、学習データ評価部と、
 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ノード生成部と、
 を有することを特徴とする学習装置。

【請求項 2】

請求項 1 に記載の学習装置であって、
 前記分割前後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件生成部に前記分割条件を追加させる、分割条件追加部と、
 を有することを特徴とする学習装置。

【請求項 3】

請求項 1 又は 2 に記載の学習装置であって、
前記分割条件生成部は、前記特徴量間の関係性を表す論理演算子を用いて、前記分割条件を生成する

ことを特徴とする学習装置。

【請求項 4】

請求項 3 に記載の学習装置であって、
前記分割条件生成部は、前記複雑さ要件が示す、前記分割条件で使用する前記特徴量の数 (F 1、 F 2) が二つである場合、前記分割条件を、以下の条件を用いて生成する

F 1 and F 2
not F 1 and F 2
F 1 or F 2
F 1 and not F 2
F 1 xor F 2

ことを特徴とする学習装置。

【請求項 5】

(a) 学習データに基づいて特徴量を生成する、ステップと、
(b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ステップと、
(c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、
(d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと、
(e) 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、
を有することを特徴とする学習方法。

【請求項 6】

請求項 5 に記載の学習方法であって、
(f) 前記分割前後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件を追加させる、ステップと、
を有することを特徴とする学習方法。

【請求項 7】

請求項 5 又は 6 に記載の学習方法であって、
前記 (b) のステップにおいて、前記特徴量間の関係性を表す論理演算子を用いて、前記分割条件を生成する
ことを特徴とする学習方法。

【請求項 8】

請求項 7 に記載の学習方法であって、
前記 (b) のステップにおいて、前記複雑さ要件が示す、前記分割条件で使用する前記特徴量 (F 1、 F 2) の数が二つである場合、前記分割条件を、以下の条件を用いて生成する

F 1 and F 2
not F 1 and F 2
F 1 or F 2
F 1 and not F 2
F 1 xor F 2

ことを特徴とする学習方法。

【請求項 9】

コンピュータに、
(a) 学習データに基づいて特徴量を生成する、ステップと、
(b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ス

テップと、

(c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、

(d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと

、

(e) 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、

を実行させるプログラム。

【請求項 10】

請求項 9 に記載の プログラム であって、

前記プログラムが、前記コンピュータに、

(f) 前記分割前後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件を追加させる、ステップを実行させる プログラム。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】0001

【補正方法】変更

【補正の内容】

【0001】

本発明は、決定木による学習をする学習装置、学習方法に関し、更には、これらを実現するための プログラム に関する。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0008

【補正方法】変更

【補正の内容】

【0008】

非特許文献 1 及び 非特許文献 2 には、操作手順を自動化するソフトウェアツールが開示されている。そのソフトウェアツールによれば、システム変更後の状態、又は操作手順を定義情報として入力し、自動的にシステムの変更、及び設定を行う。

【手続補正 5】

【補正対象書類名】明細書

【補正対象項目名】0032

【補正方法】変更

【補正の内容】

【0032】

本発明の目的の一例は、決定木の予測精度を向上させる学習装置、学習方法、及び プログラム を提供することにある。

【手続補正 6】

【補正対象書類名】明細書

【補正対象項目名】0035

【補正方法】変更

【補正の内容】

【0035】

更に、上記目的を達成するため、本発明の一側面における プログラム は、コンピュータに、

- (a) 学習データに基づいて特徴量を生成する、ステップと、
- (b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ステップと、
- (c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、
- (d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと、
- (e) 分割後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、
を実行させることを特徴とする。

【手続補正 7】

【補正対象書類名】明細書

【補正対象項目名】0037

【補正方法】変更

【補正の内容】

【0037】

【図1】図1は、学習データの一例を示す図である。

【図2】図2は、理想的な決定木の一例を示す図である。

【図3】図3は、既存の決定木の学習により生成された決定木の一例を示す図である。

【図4】図4は、学習装置の一例を示す図である。

【図5】図5は、学習装置を有するシステムの一例を示す図である。

【図6】図6は、複雑さ要件に対する分割条件の一例を示す図である。

【図7】図7は、分割結果の一例を示す図である。

【図8】図8は、評価結果の一例を示す図である。

【図9】図9は、評価結果の一例を示す図である。

【図10】図10は、学習装置の動作の一例を示す図である。

【図11】図11は、学習装置を実現するコンピュータの一例を示す図である。

【手続補正 8】

【補正対象書類名】明細書

【補正対象項目名】0049

【補正方法】変更

【補正の内容】

【0049】

分割条件生成部12は、学習データに基づいて生成された特徴量と、指定された複雑さ要件とに応じて、分割条件(具体的な分割条件)を生成する。複雑さ要件は、一つの分割条件に使用する特徴量の数を示し、初期値は1である。また、段階的に複雑さを増やす場合、複雑さ要件に上限値も設定する。例えば、上限値は2とすることが考えられる。

【手続補正 9】

【補正対象書類名】明細書

【補正対象項目名】0056

【補正方法】変更

【補正の内容】

【0056】

また、有効な分割条件が複数検出された場合、等分散検定におけるp値が、最小の分割条件を実際の決定木の分割条件として採用する。等分散検定は、予測値の確率分布に関する仮定の違いなどによりいくつかの手法が存在する。例えば、予測値に特定の確率分布を仮定しない場合にはBrown-Forsythe検定を用いる。なお、学習対象のデータの性質に応じて検定手法を選択してもよい。

【手続補正 10】

【補正対象書類名】明細書

【補正対象項目名】0071

【補正方法】変更

【補正の内容】

【0071】

有意性がある場合（ステップ A5：Yes）、又は、有意性がなく複雑さ要件が上限値である場合（ステップ A7：Yes）、ステップ A6 において、ノード生成部 15 は、有意性がある分割条件に関連付けられた決定木のノードを生成する。

【手続補正 11】

【補正対象書類名】明細書

【補正対象項目名】0083

【補正方法】変更

【補正の内容】

【0083】

なお、本実施の形態における学習装置 10 は、プログラムがインストールされたコンピュータではなく、各部に対応したハードウェアを用いることによっても実現可能である。更に、学習装置 10 は、一部がプログラムで実現され、残りの部分がハードウェアで実現されている。

【手続補正 12】

【補正対象書類名】明細書

【補正対象項目名】0093

【補正方法】変更

【補正の内容】

【0093】

(付記 9)

コンピュータに、

- (a) 学習データに基づいて特徴量を生成する、ステップと、
- (b) 前記特徴量と前記特徴量の数を示す複雑さ要件とに応じて分割条件を生成する、ステップと、
- (c) 前記学習データを前記分割条件に基づいてグループに分割する、ステップと、
- (d) 分割前後のグループを用いて、前記分割条件ごとに有意性を評価する、ステップと、
- (e) 分割前後のグループにおいて、前記分割条件に有意性がある場合、前記分割条件に関連する決定木のノードを生成する、ステップと、
を実行させるプログラム。

【手続補正 13】

【補正対象書類名】明細書

【補正対象項目名】0094

【補正方法】変更

【補正の内容】

【0094】

(付記 10)

付記 9 に記載のプログラムであって、

前記プログラムが、前記コンピュータは、

- (f) 前記分割前後のグループにおいて、すべての前記分割条件に有意性がない場合、前記複雑さ要件が示す前記特徴量の数を増加させ、前記分割条件を追加させる、ステップを

実行させるプログラム。

【手続補正 14】

【補正対象書類名】明細書

【補正対象項目名】0095

【補正方法】変更

【補正の内容】

【0095】

(付記 11)

付記 9 又は 10 に記載のプログラムであって、

前記 (b) のステップにおいて、前記特徴量間の関係性を表す論理演算子を用いて、前記分割条件を生成する

ことを特徴とするプログラム。

【手続補正 15】

【補正対象書類名】明細書

【補正対象項目名】0096

【補正方法】変更

【補正の内容】

【0096】

(付記 12)

付記 11 に記載のプログラムであって、

前記 (b) のステップにおいて、前記複雑さ要件が示す、前記分割条件で使用する前記特徴量 (F1、F2) の数が二つである場合、前記分割条件を、以下の条件を用いて生成する

F1 and F2

not F1 and F2

F1 or F2

F1 and not F2

F1 xor F2

ことを特徴とするプログラム。

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2019/012984

A. CLASSIFICATION OF SUBJECT MATTER Int.Cl. G06N20/00 (2019.01) i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) Int.Cl. G06N20/00		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Published examined utility model applications of Japan	1922-1996	
Published unexamined utility model applications of Japan	1971-2019	
Registered utility model specifications of Japan	1996-2019	
Published registered utility model applications of Japan	1994-2019	
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-157814 A (FUJI ELECTRIC HOLDINGS CO., LTD.) 03 June 2004 (Family: none)	1-12
A	JP 2011-28519 A (JAPAN BROADCASTING CORPORATION) 10 February 2011 (Family: none)	1-12
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed		"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
Date of the actual completion of the international search 17 June 2019 (17.06.2019)		Date of mailing of the international search report 25 June 2019 (25.06.2019)
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan		Authorized officer Telephone No.

国際調査報告		国際出願番号 PCT/J P 2 0 1 9 / 0 1 2 9 8 4	
A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int.Cl. G06N20/00(2019,01)i			
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int.Cl. G06N20/00			
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2019年 日本国実用新案登録公報 1996-2019年 日本国登録実用新案公報 1994-2019年			
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)			
C. 関連すると認められる文献			
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号	
A	JP 2004-157814 A (富士電機ホールディングス株式会社) 2004.06.03, (ファミリーなし)	1-12	
A	JP 2011-28519 A (日本放送協会) 2011.02.10, (ファミリーなし)	1-12	
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。			
* 引用文献のカテゴリー 「A」 特に関連のある文献ではなく、一般的技術水準を示すもの 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」 口頭による開示、使用、展示等に言及する文献 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願			
国際調査を完了した日 17.06.2019		国際調査報告の発送日 25.06.2019	
国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 塚田 肇	5B 3652 電話番号 03-3581-1101 内線 3545

フロントページの続き

(81)指定国・地域 AP(BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, RU, TJ, TM), EP(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT

(注) この公表は、国際事務局(WIPO)により国際公開された公報を基に作成したものである。なおこの公表に係る日本語特許出願(日本語実用新案登録出願)の国際公開の効果は、特許法第184条の10第1項(実用新案法第48条の13第2項)により生ずるものであり、本掲載とは関係ありません。