(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2020年6月18日(18.06.2020)

WIPO | PCT

(10) 国際公開番号
WO 2020/121115 A1

(51) 国際特許分類:
G06F 16/906 (2019.01) *G06N 20/00* (2019.01)

(21) 国際出願番号： PCT/IB2019/060377

(22) 国際出願日： 2019年12月3日(03.12.2019)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(30) 優先権データ：
特願 2018-233037 2018年12月13日(13.12.2018) JP

(71) 出願人: 株式会社半導体エネルギー研究所 (**SEMICONDUCTOR ENERGY LABORATORY CO., LTD.**) [JP/JP]; 〒2430036 神奈川県厚木市長谷３９８ Kanagawa (JP).

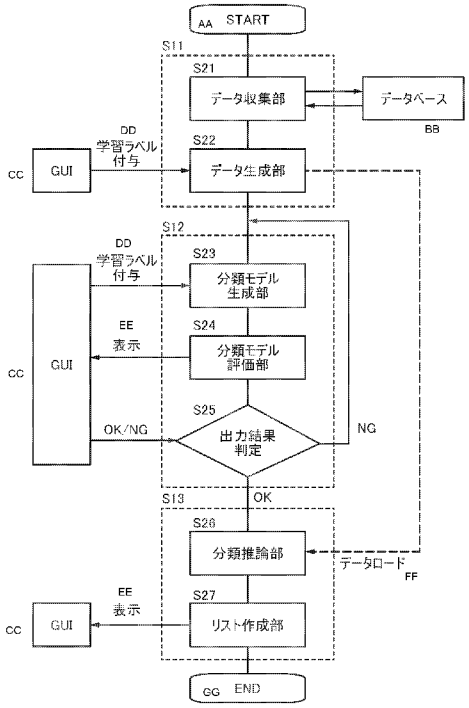
(72) 発明者: 桃純平 (**MOMO, Junpei**); 〒2430036 神奈川県厚木市長谷３９８株式会社半導体エネルギー研究所内 Kanagawa (JP). 福留貴浩 (**FUKUTOME, Takahiro**); 〒2430036 神奈川県厚木市長谷３９８株式会社半導体エネルギー研究所内 Kanagawa (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): 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,

(54) **Title:** CONTENT CLASSIFICATION METHOD AND CLASSIFICATION MODEL GENERATION METHOD

(54) 発明の名称： コンテンツの分類方法および分類モデルの生成方法

図2



S21 Data accumulation unit
S22 Data generation unit
S23 Classification model generation unit
S24 Classification model assessment unit
S25 Output result determination
S26 Classification deduction unit
S27 List generation unit
BB Database
CC GUI
DD Learning label application
EE Display
FF Data load

(57) **Abstract:** The present invention addresses the problem of providing a classification model for classifying content. Included are learning content and content. A first feature value and a learning label are applied to the learning content, and a second feature value is applied to the content. This content classification method includes: a step for using a plurality of pieces of learning content to generate a plurality of first classification models by machine learning; a step for using the plurality of first classification models to generate a second classification model; and a step for using the second

WO 2020/121115 A1

[続葉有]

QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (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), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告(条約第21条(3))

classification model to apply determination information to a plurality of pieces of content and display the same on a GUI. The determination information includes a classification label or a score. In the GUI, it is possible to specify a specific number range among scores and to list and display corresponding content. The feature value applied to the content is a management parameter (meta information).

(57) 要約:コンテンツを分類する分類モデルを提供する。 学習用コンテンツと、コンテンツと、を有し、学習用コンテンツには、第１の特徴量および学習用 ラベルが付与され、コンテンツには、第２の特徴量が付与される。複数の学習用コンテンツを用い て複数の第１の分類モデルを機械学習によって生成するステップと、複数の第１の分類モデルを用 いて第２の分類モデルを生成するステップと、第２の分類モデルを用いて複数のコンテンツに判定 情報を付与しＧＵＩに表示するステップと、を含むコンテンツの分類方法である。判定情報には、 分類ラベルまたはスコアが含まれる。なお、ＧＵＩでは、スコアのうち特定の数値範囲を指定し、 該当するコンテンツをリスト化して表示することができる。なお、コンテンツに与えられる特徴量 は、管理パラメータ（メタ情報）である。

発明の名称

コンテンツの分類方法および分類モデルの生成方法

技術分野

［０００１］

本発明の一態様は、コンピュータ装置を利用したコンテンツの分類方法、コンテンツの分類システム、分類モデルの生成方法、およびグラフィカルユーザインターフェースに関する。

［０００２］

なお、本発明の一態様は、コンピュータ装置に関する。本発明の一態様は、コンピュータ装置を利用した電子化されたコンテンツ（テキストデータ、画像データ、音声データ、または動画データ）の分類方法に関する。特に、本発明の一態様は、コンテンツの集まりから機械学習を用いて効率的に分類するコンテンツの分類システムに関する。なお、本発明の一態様は、コンピュータ装置がプログラムによって管理するグラフィカルユーザインターフェースを用いたコンテンツの分類方法、コンテンツの分類システム、および分類モデルの生成方法に関する。

背景技術

［０００３］

ユーザは、コンテンツの集まりから、ユーザの指定するトピックスに関する情報を容易に分類し、且つ抽出したいと考えている。但し、大量のコンテンツから目的の条件に合ったコンテンツを分類する場合、個人が有する知識、または経験などによりコンテンツの分類結果にばらつきが生じる。

［０００４］

近年では、個人が有する知識、経験によって分類されたコンテンツの分類結果を教師データとしてコンピュータ装置に与え、コンテンツの分類方法を機械学習させる提案がされている。例えば特許文献１では、ユーザの指定するトピックスと関連性の高いドキュメントを決定するための機械学習のアプローチが開示されている。

［先行技術文献］

［特許文献］

［０００５］

［特許文献１］特開２００９－１０４６３０号公報

発明の概要

発明が解決しようとする課題

［０００６］

あるコンテンツの集合を目的に即して分類を行う場合がある。本発明の一態様では、コンテンツが、特許の場合について説明する。特許には、それぞれ固有の特許番号が与えられている。したがって以降では、コンテンツを特許番号と言い換えて説明する場合がある。なお、本発明の一態様で扱うコンテンツの分類方法では、特許番号に与えられる複数の管理パラメータに着目する。但し、コンテンツは、特許文献に限定されない。コンテンツは、テキストデータ、画像データ、音声データ、または動画データなどの情報を扱うことができる。

［０００７］

コンテンツは、様々なメタ情報を用いて管理されている。メタ情報とは、あるコンテンツそのものではなく、そのコンテンツが属する属性、または関連する情報を記述するデータを示す。一例と

して特許番号には、コンテンツの内容として特許請求の範囲、要約書、図面、および明細書が関連付けられている。さらに特許番号には、メタ情報（評価情報、経過日数、ファミリー情報など）が与えられ、メタ情報を用いた管理がされている。特許番号は、メタ情報を用いることで重要度に応じた分類が行われる。分類の精度や効率は、対象となる文書の内容にもよるが、ユーザの経験や熟練により差が出やすく、また、大量の文書を分類する必要があり、効率化の課題があった。

［０００８］

機械学習を用いた分類モデルの生成には、大量の学習データを用意する必要があるためユーザに過度の負担を強いる、という課題がある。学習データに含まれる分類されたコンテンツ数のばらつきが分類モデルの精度に影響を与える、という課題がある。

［０００９］

上記問題に鑑み、本発明の一態様は、効率良く分類モデルを生成し、それを用いて情報を分類する方法を提供することを課題の一とする。または、本発明の一態様は、対話的に分類モデルを生成するためのグラフィカルユーザインターフェースを提供することを課題の一とする。または、本発明の一態様は、確率の高い情報を分類するプログラムを提供することを課題の一とする。

［００１０］

なお、これらの課題の記載は、他の課題の存在を妨げるものではない。なお、本発明の一態様は、これらの課題の全てを解決する必要はないものとする。なお、これら以外の課題は、明細書、図面、請求項などの記載から、自ずと明らかとなるものであり、明細書、図面、請求項などの記載から、これら以外の課題を抽出することが可能である。

課題を解決するための手段

［００１１］

コンピュータ装置が有する記憶装置には、プログラムが保存されている。プログラムは、コンピュータ装置が有する表示装置にグラフィカルユーザインターフェース（以下、ＧＵＩ）を介して様々な情報を表示することができる。なお、ユーザは、ＧＵＩを介してコンピュータ装置にプログラムの操作、情報の付与、データベースに対する応答、機械学習の指示などを行えるものとする。また、プログラムは、ＧＵＩを介して機械学習による演算処理結果、データベースから学習用コンテンツまたは未分類のコンテンツをダウンロードした内容などを表示装置に表示することができる。なお、以降において、単にコンテンツと記す場合は、学習用コンテンツ、未分類のコンテンツ、または分類済コンテンツを含む。

［００１２］

提案するコンテンツの分類システムは、機械学習を利用してコンテンツの分類モデルを生成し、生成されるコンテンツの分類モデルを用いて、未分類のコンテンツを分類する。例えば、複数のメタ情報を有するコンテンツを学習用コンテンツとする。学習用コンテンツは、さらに学習用ラベルが付与されることで学習用コンテンツから特徴ベクトルを生成する。なお特徴ベクトルを生成する場合、メタ情報または学習用ラベルは、学習用コンテンツの特徴量として扱うことができる。

［００１３］

学習用コンテンツは、教師データとして扱われる。分類モデルは、学習用コンテンツを基に機械学習を行うことで取得することができる。ここで得られる分類モデルは、複数のメタ情報を有するコンテンツの分類を行う。なお、分類の種類は、ユーザの目的により２種類でも良いし、３種類以上としても良い。ユーザは、分類モデルを利用することで、全ての文書を手作業または目視で判断

するよりも短い時間で文書全体の分類を行うことが可能となる。

［００１４］

なお、学習用コンテンツは、データベースに保存されている学習用コンテンツからダウンロードすることができる。もしくは、コンピュータ装置の記憶装置に保存されている学習用コンテンツを用いることができる。なお、学習用コンテンツには、学習用ラベルが含まれて管理されていてもよい。さらに、データベースに保存されている分類モデルをダウンロードしてもよい。もしくは、コンピュータ装置の記憶装置に保存されている分類モデルを用いてもよい。

［００１５］

本発明の一態様は、学習用コンテンツと、コンテンツと、を有し、学習用コンテンツには、第１の特徴量および学習用ラベルが付与され、コンテンツには、第２の特徴量が付与される。複数の学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成するステップと、複数の第１の分類モデルを用いて第２の分類モデルを生成するステップと、第２の分類モデルを用いて複数のコンテンツに判定情報を付与しグラフィカルユーザインターフェース内に表示するステップと、を含むコンテンツの分類方法である。

［００１６］

本発明の一態様は、学習用コンテンツと、コンテンツと、を有し、学習用コンテンツには、第１の特徴量および学習用ラベルが付与され、コンテンツには、第２の特徴量が付与される。複数の学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成するステップと、複数の第１の分類モデルの出力から平均値を算出するステップと、複数の平均値を用いて第２の分類モデルを生成するステップと、第２の分類モデルを用いて複数のコンテンツに判定情報を付与しグラフィカルユーザインターフェース内に表示するステップと、を含むコンテンツの分類方法である。

［００１７］

本発明の一態様は、学習用コンテンツと、コンテンツと、を有し、学習用コンテンツには、第１の特徴量および学習用ラベルが付与され、コンテンツには、第２の特徴量が付与される。複数の学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成するステップと、複数の第１の分類モデルがそれぞれ第一の評価基準による評価をするステップと、複数の第１の分類モデルがそれぞれ第二の評価基準による評価をするステップと、複数の第一の評価基準による評価結果と、第二の評価基準と、による評価結果から第２の分類モデルを生成するステップと、第２の分類モデルを用いて複数のコンテンツに判定情報を付与しグラフィカルユーザインターフェース内に表示するステップと、を含むコンテンツの分類方法である。

［００１８］

上記構成において、第一の評価基準は精度（Ｐｒｅｃｉｓｉｏｎ）であり、第二の評価基準は感度（Ｓｅｎｓｉｔｉｖｉｔｙ）であるコンテンツの分類方法が好ましい。

［００１９］

上記各構成において、任意の学習用コンテンツを用いて第１の分類モデルを生成するステップを含むコンテンツの分類方法が好ましい。

［００２０］

上記各構成において、学習用コンテンツには、さらに分類情報が与えられ、第２の分類モデルの出力を用いて分類ラベルが付与された複数のコンテンツから分類情報と同じ判定情報を有するコンテンツを選択してグラフィカルユーザインターフェース内に表示するステップと、を含むコンテン

ツの分類方法が好ましい。

［００２１］

上記各構成において、学習用コンテンツ又コンテンツに与えられる特徴量は、管理パラメータであるコンテンツの分類方法が好ましい。

［００２２］

上記各構成において、判定情報には、分類ラベルまたはスコアを含むコンテンツの分類方法が好ましい。

［００２３］

上記構成において、グラフィカルユーザインターフェースは、スコアのうち特定の数値範囲を指定し、該当するコンテンツをリストとして表示するステップを含むコンテンツの分類方法が好ましい。

発明の効果

［００２４］

本発明の一態様は、精度良く情報を分類する方法を提供することができる。または、本発明の一態様は、精度良く情報を分類するユーザインターフェースを提供することができる。または、本発明の一態様は、精度良く情報を分類するプログラムを提供することができる。

［００２５］

また、本発明の一態様は、機械学習を利用した分類モデルを生成するための対話的なインターフェースをユーザに提供することができ、教師データの用意や学習結果の評価といったユーザの負担を軽減することができる。

［００２６］

なお本発明の一態様の効果は、上記列挙した効果に限定されない。上記列挙した効果は、他の効果の存在を妨げるものではない。なお他の効果は、以下の記載で述べる、本項目で言及していない効果である。本項目で言及していない効果は、当業者であれば明細書または図面等の記載から導き出せるものであり、これらの記載から適宜抽出することができる。なお、本発明の一態様は、上記列挙した効果、および／または他の効果のうち、少なくとも一つの効果を有するものである。したがって本発明の一態様は、場合によっては、上記列挙した効果を有さない場合もある。

図面の簡単な説明

［００２７］

図１は、分類方法を説明するフローチャートである。

図２は、分類方法を説明するフローチャートである。

図３は、分類システム１００と、ネットワークとの接続について説明する図である。

図４は、分類システムを説明するブロック図である。

図５Ａ、図５Ｂは、グラフィカルユーザインターフェースを説明する図である。

図６は、分類モデルの生成方法を説明する図である。

図７は、分類モデルの生成方法を説明する図である。

図８は、分類モデルの生成方法を説明する図である。

図９は、グラフィカルユーザインターフェースを説明する図である。

図１０は、グラフィカルユーザインターフェースを説明する図である。

発明を実施するための形態

［００２８］

本実施の形態では、コンテンツの分類方法について図１乃至図１０を用いて説明する。

［００２９］

本実施の形態で説明するコンテンツの分類方法は、コンピュータ装置上で動作するプログラムによって制御される。プログラムは、コンピュータ装置が有するメモリ、またはストレージに保存されている。もしくは、ネットワーク（ＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）、ＷＡＮ（Ｗｉｄｅ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）、インターネットなど）を介して接続されているコンピュータ、またはデータベースを有するサーバコンピュータに保存されている。

［００３０］

なお、コンピュータ装置が有する表示装置には、ユーザがプログラムに与えるデータと、当該データをコンピュータ装置が有する演算装置によって演算した結果を表示することができる。なお、装置の構成に関しては、図４を用いて詳細に説明をする。

［００３１］

表示装置に表示されるデータは、例えばリスト化された表示形式に従うことで、ユーザが認識しやすくなり、操作性が向上する。したがってユーザが、表示装置を介してコンピュータ装置が有するプログラムと簡便にやり取りするためのインターフェースをＧＵＩとして説明する。

［００３２］

ユーザは、ＧＵＩを介してプログラムが有するコンテンツの分類方法を利用することができる。ユーザは、ＧＵＩを用いてコンテンツの分類操作を簡便にすることができる。また、ユーザは、ＧＵＩを介することでコンテンツの分類結果を視覚的に判断しやすくなる。また、ユーザは、ＧＵＩを介することでプログラムを簡便に操作することができる。なお、当該コンテンツとは、テキストデータ、画像データ、音声データ、または動画データなどの情報を示している。

［００３３］

次に、ＧＵＩを用いたコンテンツの分類方法を、ＧＵＩの操作手順に従い説明する。最初に、データ処理部について説明する。データ処理部は、データ収集部とデータ生成部を有する。例えば、データ収集部は、ＧＵＩを介してデータベースから複数のコンテンツからなるファイルを取得する。また、データ生成部は、ユーザがＧＵＩを介してコンテンツに対して学習用ラベルを付与することで学習用コンテンツを生成することができる。もしくは、学習用ラベルが付与された学習用コンテンツをデータベースから取得してもよい。なお、複数のコンテンツとは、コンピュータ装置が有するメモリまたはストレージに保存されているファイル、もしくはネットワークに接続されたデータベース、コンピュータ、またはデータサーバなどに保存されているデータである。

［００３４］

したがって、当該データベースには、複数の学習用コンテンツまたは複数の未分類のコンテンツがリスト化されて保存されていることが好ましい。なお、学習用コンテンツおよび未分類のコンテンツには、複数の特徴量と学習用ラベルが付与されている。当該学習用ラベルは、ＧＵＩを介してユーザが修正することができる。当該学習用ラベルを学習用コンテンツに付与した場合、学習用ラベルが付与された学習用コンテンツは、データベースに保存することができる。

［００３５］

学習用コンテンツには、学習用ラベルが付与されない検証用コンテンツを含むことができる。検証用コンテンツは、学習用コンテンツを用いて生成する分類モデルを検証するために用いることが

できる。

［００３６］

一例として、コンテンツが、特許番号の場合について説明する。特許番号には、特許番号の特徴量として複数のメタ情報が付与される。一例として、メタ情報には、評価情報、経過日数、ファミリーの数、ファミリーの状態、出願種別、ライフ、ファミリー内のペンディング数、ファミリー内の放棄数、費用、発明者数、分野、またはクレーム数などがある。つまり、メタ情報とは、コンテンツの管理パラメータである。なお、ファミリーには、特許ファミリー、またはパテントファミリーなどが含まれる。

［００３７］

次に、学習処理部について説明する。学習処理部では、学習用コンテンツを用いて分類モデルを生成するステップを有する。学習処理部は、分類モデル生成部、または分類モデル評価部を有する。

［００３８］

分類モデル生成部は、分類モデルを生成することができる。分類モデル生成部は、複数の学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成するステップを含み、当該複数の第１の分類モデルを用いて第２の分類モデルを生成するステップを含む。第１の分類モデルまたは第２の分類モデルの出力値をＧＵＩに表示することができる。ユーザは、当該出力値に対する第１の分類モデルの学習用ラベルを付与（修正を含む）することができる。もしくは、ユーザは、当該出力値に対する新しい学習用コンテンツを追加することができる。

［００３９］

分類モデル評価部は、検証用コンテンツを用いて分類モデル生成部で生成される分類モデルを評価する。当該分類モデルを用いて検証用コンテンツを推論する場合、当該分類モデルは、推論した結果を判定情報として出力する。ＧＵＩは、それぞれの評価用コンテンツに判定情報を付与して表示することができる。

［００４０］

なお、ユーザは、分類モデル評価部の出力結果を判定し、必要に応じて学習用ラベルを修正し、分類モデル生成部にて分類モデルを更新することができる。または、学習用コンテンツを、追加することで、分類モデル生成部にて分類モデルを更新することができる。

［００４１］

次に判定処理部について説明する。判定処理部では、分類推論部とリスト生成部を有する。例えば、分類推論部は、分類モデル生成部で生成される第１の学習モデルと第２の学習モデルを用いて複数の未分類のコンテンツを推論し分類する。当該分類モデルは、それぞれのコンテンツに対し推論結果を判定情報として付与する。

［００４２］

リスト生成部は、当該判定情報が与えられたコンテンツからユーザの求める形式のリストを生成しＧＵＩに表示することができる。例えば、それぞれのコンテンツが出願国ごとに管理されている場合、出願国を分類情報とすることができる。出願国を分類情報とする場合、生成される分類モデルは、出願国ごとに異なる分類モデルを生成することが好ましい。なお分類情報は、出願国に限定されない。例えば、コンテンツが有するメタ情報の一つを分類情報とすることができる。

［００４３］

メタ情報を分類情報とする場合について説明する。例えば、特許ファミリーの状態を分類情報と

してもよい。特許番号には、親出願の特許番号、分割出願の特許番号などがメタ情報で与えられている場合がある。親出願の特許番号から分割出願が可能な状態、親出願の特許番号から分割出願が不可能な状態、親出願の特許番号が権利持続中の状態、親出願の特許番号が権利喪失の状態、分割出願の特許番号からさらに分割出願が可能な状態、分割出願の特許番号からさらに分割出願が不可能な状態、分割出願の特許番号が権利持続中の状態、分割出願の特許番号が権利喪失の状態などによって、異なる分類モデルを生成し、当該分類モデルを用いてそれぞれ推論することができる。

［００４４］

つまり、判定処理部は、分類モデルを用いて複数の未分類のコンテンツを推論することができる。推論される結果は、それぞれのコンテンツに判定情報として付与しＧＵＩに表示するステップを含む。判定情報には、少なくとも分類ラベルと、スコア（確率）と、が含まれる。また、ＧＵＩは、スコアのうち特定の数値範囲を指定し、該当するコンテンツをリストとして表示するステップが含まれる。

［００４５］

上述した分類モデル生成部とは異なる例を示す。分類モデル生成部は、複数の学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成するステップと、当該複数の第１の分類モデルの出力から平均値を算出するステップと、複数の当該平均値を用いて第２の分類モデルを生成するステップと、を含む。なお、第１の分類モデルまたは第２の分類モデルの出力値をＧＵＩに表示することができる。ユーザは、当該出力値に対して、第１の分類モデルの学習用ラベルを修正することができる。もしくは、ユーザは、当該出力値に対して学習用コンテンツを追加することができる。なお、当該平均値とは、相加平均計算、相乗平均計算、または調和平均計算のいずれか一を用いて算出することを意味する。

［００４６］

第２の分類モデルは、複数の当該平均値を用いて生成される。第２の分類モデルでは、第１の分類モデルの出力が平均化されるため学習用コンテンツが有する外れ値などのノイズ成分の影響を低減することができる。

［００４７］

次に、上述した分類モデル生成部とは異なる例を示す。分類モデル生成部は、複数の学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成するステップと、複数の第１の分類モデルがそれぞれ第一の評価基準による評価をするステップと、複数の第１の分類モデルがそれぞれ第二の評価基準による評価をするステップと、複数の第一の評価基準による評価結果と第二の評価基準による評価結果から第２の分類モデルを生成するステップと、を含む。なお、第１の分類モデルまたは第２の分類モデルの出力値は、ＧＵＩに表示することができる。ユーザは、当該出力値に対して、第１の分類モデルの学習用ラベルを修正することができる。もしくは、ユーザは、当該出力値に対して学習用コンテンツを追加することができる。なお、第一の評価基準は混同行列の精度であり、第二の評価基準は混同行列の感度である。

［００４８］

複数の第１の分類モデルの出力を、第一の評価基準および第二の評価基準により評価した結果を用いて第２の分類モデルを生成する。なお、第一の評価基準では、混同行列の精度を学習用ラベルに対する適合率と言い換えることができる。第二の評価基準では、混同行列の感度を学習用ラベルに対する再現率と言い換えることができる。よって生成される分類モデルは、複数の第１の分類モ

デルの当該適合率および当該再現率を含むことができる。当該複数の第１の分類モデルを用いて生成する第２の分類モデルは、分類精度が向上する。

［００４９］

上述した分類モデル生成部において、例えば、生成されたｍ個（ｍは自然数を表す）の第１の分類モデルを用いて第２の分類モデルを生成してもよい。

［００５０］

なお、上述した分類モデル生成部がｋ個（ｋは自然数を表す）の学習用コンテンツを有する場合、第１の分類モデルは、ｋ個以下の任意の学習用コンテンツを生成することができる。また、学習用コンテンツがｋ個選択される場合、任意の学習用コンテンツは、異なる２つの学習用モデルを含むことができる。また、ｋ個の異なる番号が付与された学習用コンテンツは、番号によってソートされる順に従いｑ個（ｑは自然数を表す）ずつを用いて第１の分類モデルを生成することができる。

［００５１］

プログラムは、データベースから読み込んだコンテンツをＧＵＩに表示することができる。コンテンツは、リスト化されたメタ情報を有することが好ましい。ＧＵＩは、ＧＵＩが有する表示形式に従い当該コンテンツを表示する。なお、コンテンツに付与されるリスト化されたメタ情報は、レコードと呼ばれる単位で管理されることが好ましい。例えば、それぞれのレコードは、番号に紐つけられたＩＤ（Ｉｄｅｎｔｉｆｉｃａｔｉｏｎ）、コンテンツ（画像データ、音声データ、または動画データ）、またはメタ情報などによって構成される。

［００５２］

本明細書では、メタ情報に着目して機械学習し、当該機械学習によって分類モデルを生成する。分類モデルは、メタ情報を解析して特徴ベクトル化されたコンテンツを分類する。

［００５３］

また、上述したコンテンツの分類方法では、学習用ラベルを教師データとして用いない機械学習による分類を行うことができる。例えば、分類モデルには、Ｋ－ｍｅａｎｓ、またはＤＢＳＣＡＮ（ｄｅｎｓｉｔｙ－ｂａｓｅｄ ｓｐａｔｉａｌ ｃｌｕｓｔｅｒｉｎｇ ｏｆ ａｐｐｌｉｃａｔｉｏｎｓ ｗｉｔｈ ｎｏｉｓｅ）などのアルゴリズムを用いることができる。

［００５４］

また、プログラムは、複数のメタ情報および学習用ラベルが付与された学習用コンテンツを用いて機械学習を用いて分類モデルを生成することができる。分類モデルには、決定木、ナイーブベイズ、ＫＮＮ（ｋ Ｎｅａｒｅｓｔ Ｎｅｉｇｈｂｏｒ）、ＳＶＭ（Ｓｕｐｐｏｒｔ Ｖｅｃｔｏｒ Ｍａｃｈｉｎｅｓ）、パーセプトロン、ロジスティック回帰、またはニューラルネットワークなどのアルゴリズムを用いることができる。

［００５５］

さらに、プログラムは、学習用コンテンツの数に応じて分類モデルを切り替えることができる。例えば、学習用コンテンツの数が少ないときは決定木、ナイーブベイズ、ロジスティック回帰、学習用コンテンツの数が一定量以上あればＳＶＭ、ランダムフォレスト、ニューラルネットワークを用いるとしてもよい。なお、本実施の形態で使用する分類モデルは、決定木のアルゴリズムの一つであるランダムフォレストを用いている。さらに、メタ情報の選択方法、学習用コンテンツの選択方法、または第１の分類モデルの選択方法としてランダムサンプリングまたはクロスバリエーションを用いることができる。もしくは、付与された番号のソートされる順に従いｑ個ずつを選択する

ことができる。

［００５６］

続いて、図面を用いてコンテンツの分類方法を説明する。図１は、本実施の一態様であるコンテンツの分類方法を説明するフローチャートである。なお、コンテンツの分類方法は、コンピュータ装置上で動作するプログラムによって制御される。したがって、プログラムが、データ処理部、学習処理部、または判定処理部を有することでコンテンツを分類することができる。プログラムは、ＧＵＩを介してユーザの求めるコンテンツの分類をすることができる。つまり上述したそれぞれの処理部で処理される内容は、プログラムのステップに相当する。

［００５７］

ステップＳ１１では、ユーザがＧＵＩを介してコンテンツが含まれるファイルをロードする指示をすることができる。当該ファイルは、データ処理部が有するデータベースに保存されている。なお、ファイルには、学習用コンテンツ、未分類のコンテンツなどが含まれる。

［００５８］

したがって、当該データベースには、複数の学習用コンテンツ、または複数の未分類のコンテンツがリスト化されて保存されていることが好ましい。ユーザは、ＧＵＩに表示されている学習用コンテンツの学習用ラベルを付与もしくは修正することができる。なお、ファイルには、学習用ラベルが付与されていない検証用コンテンツを含むことができる。

［００５９］

ステップＳ１２は、ロードしたファイルを用いて分類モデルを生成する学習処理部である。生成した分類モデルは、検証用コンテンツを評価し、評価結果をＧＵＩに表示することができる。ユーザは、当該評価結果に対し、学習用ラベルの修正、学習用コンテンツの追加などを指示することができる。

［００６０］

なお、ユーザは、学習用コンテンツの有するメタ情報の経時的変化を予測し、メタ情報を更新することができる。ユーザがメタ情報を更新する場合、分類モデルは、分類モデルの経時的な変化を含むことができる。したがって、ユーザは、コンテンツの経時的な分類の変化を得ることができる。分類モデルは、価値が増大していくと予想されるコンテンツ群、または価値が減少していくと予想されるコンテンツ群を分類することができる。

［００６１］

ステップＳ１３は、判定処理部である。ステップＳ１２で生成される分類モデルを用いて未分類のコンテンツを推論する。当該分類モデルは、未分類のコンテンツに対し推論結果をもとに判定情報を付与することができる。判定処理部では、当該判定情報が与えられたコンテンツをユーザの求める形式でＧＵＩに表示することができる。当該判定情報には、少なくとも分類ラベルと、スコアと、が含まれる。また、ＧＵＩは、スコアのうち特定の数値範囲を指定し、該当するコンテンツを表示することができる。

［００６２］

続いて、図２では、図１のフローチャートをより詳細に説明する。まず、ステップＳ１１の詳細について説明する。ステップＳ１１のデータ処理部は、ステップＳ２１のデータ収集部およびステップＳ２２のデータ生成部を有する。

［００６３］

ステップＳ２１のデータ収集部について説明する。ステップＳ２１のデータ収集部は、データベースからファイルをロードすることができる。なお、メタ情報またはコンテンツなどは、異なるデータベースによって管理することができる。メタ情報は、コンテンツを扱う会社、組織、またはユーザにおいて異なる場合がある。したがって、データ収集部は、異なるデータベースからコンテンツに関するメタ情報を集める機能を有する。なお、それぞれのデータベースは、異なる建物、異なる地域、または異なる国に設置することができる。

［００６４］

次にステップＳ２２のデータ生成部について説明する。データ生成部は、レコードと呼ばれる単位でコンテンツとメタ情報を管理することができる。例えば、それぞれのレコードは、番号に紐つけられたＩＤ、コンテンツ（画像データ、音声データ、または動画データ）、またはメタ情報などによって構成される。また、ユーザは、ＧＵＩに表示されるコンテンツに対して学習用ラベルを付与することで、学習用コンテンツを生成することができる。

［００６５］

続いて、ステップＳ１２の詳細について説明する。ステップＳ１２の学習処理部は、ステップＳ２３の分類モデル生成部、ステップＳ２４の分類モデル評価部、およびステップＳ２５の出力結果判定処理を有する。

［００６６］

ステップＳ２３の分類モデル生成部について説明する。分類モデル生成部は、コンテンツの分類モデルを生成することができる。分類モデル生成部は、複数の学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成することができる。当該複数の第１の分類モデルを用いて第２の分類モデルを生成することができる。ＧＵＩは、第１の分類モデルまたは第２の分類モデルの出力値を表示することができる。

［００６７］

なお、後述するステップＳ２５の出力結果判定処理後、ユーザは、当該出力値に対する第１の分類モデルの学習用ラベルを付与（修正を含む）することができる。もしくは、ユーザは、当該出力値に対する新しい学習用コンテンツを追加することができる。学習用コンテンツが有するメタ情報の経時的変化を予測し、メタ情報を更新することができる。なお、ユーザによるメタ情報を更新することで得られる効果は、ステップＳ１２の説明を参酌することができる。

［００６８］

次にステップＳ２４の分類モデル評価部について説明する。分類モデル評価部は、検証用コンテンツを用いることで、分類モデル生成部で生成される分類モデルを評価することができる。当該分類モデルは、検証用コンテンツを推論した結果を判定情報として出力する。ＧＵＩは、それぞれの評価用コンテンツに判定情報を付与して表示することができる。

［００６９］

次にステップＳ２５の出力結果判定処理について説明する。例えば、ユーザは、ステップＳ２４の分類モデル評価部の出力結果を判断し、コンテンツの分類モデルが十分に学習したと判定することができる。ユーザは、ＧＵＩに対し分類モデルの生成完了（ＯＫ）の指示を与える。例えば、ユーザがコンテンツの分類モデルの学習が不十分（ＮＧ）と判定することができる。ユーザは、ステップＳ２３に戻り学習用ラベルの変更、学習用コンテンツの追加、またはメタ情報の更新などにより分類モデルを更新する。

［００７０］

続いて、ステップＳ１３の詳細について説明する。ステップＳ１３の判定処理部は、ステップＳ２６の分類推論部、およびステップＳ２７のリスト作成部を有する。

［００７１］

ステップＳ２６の分類推論部について説明する。分類推論部は、分類モデル生成部で生成される第１の学習モデルと第２の学習モデルを用いて複数の未分類のコンテンツを推論し分類する。なお、分類推論部にはステップＳ２２のデータ生成部で生成された未分類のコンテンツが与えられる。当該分類モデルは、それぞれのコンテンツに対し推論結果を判定情報として付与する。

［００７２］

ステップＳ２７のリスト作成部について説明する。リスト生成部は、当該判定情報が与えられたコンテンツをユーザの求める形式にリスト化しＧＵＩに表示することができる。なお、それぞれのコンテンツには、メタ情報とは異なる分類情報が与えられてもよい。例えば、分類情報が学習用コンテンツに与えられている場合、生成される分類モデルは、分類情報ごとに異なる分類モデルを生成することができる。もしくは、コンテンツが有するメタ情報の一つを分類情報とすることができる。

［００７３］

なお、当該判定情報には、少なくとも分類ラベルと、スコアとが含まれる。また、ＧＵＩは、スコアのうち特定の数値範囲を指定し、該当するコンテンツをリスト化してＧＵＩに表示することができる。

［００７４］

図３は、上述したコンテンツの分類方法を有する分類システム１００と、ネットワーク（Ｎｅｔ Ｗｏｒｋ）との接続について説明する図である。

［００７５］

分類システム１００は、通信網ＬＡＮ１と接続される。通信網ＬＡＮ１には、データベースＤＢ１、またはクライアントコンピュータＣＬ１乃至ＣＬｎ（ｎは自然数である）などが接続されている。また、通信網ＬＡＮ１は、ネットワークを介して通信網ＬＡＮ２と接続することができる。なお、ネットワークは、インターネット、通信網ＷＡＮ、もしくは衛星通信を用いることができる。通信網ＬＡＮ２には、データベースＤＢ２、またはクライアントコンピュータＣＬ１１乃至ＣＬ１ｎなどが接続されている。

［００７６］

分類システム１００は、データベースＤＢ１、データベースＤＢ２、クライアントコンピュータＣＬ１乃至ＣＬｎ、またはクライアントコンピュータＣＬ１１乃至ＣＬ１ｎに記憶されているコンテンツを含むファイルを用いて、コンテンツの生成、コンテンツの分類、モデル生成、且つ未分類のコンテンツを分類することができる。

［００７７］

また、ユーザは、分類システム１００で動作するプログラムから、ＧＵＩに対して指示を出すことができる。例えば、ユーザは、インターネットを介して異なる国に設置されたデータベースの情報を用いて上述した分類モデルを生成し、未分類のコンテンツを分類することができる。つまり、コンテンツまたはメタ情報は、異なるデータベースまたはクライアントコンピュータに記憶されていてもよい。

［００７８］

なお、ＧＵＩは、データベースＤＢ１、データベースＤＢ２、クライアントコンピュータＣＬ１乃至ＣＬｎ、またはクライアントコンピュータＣＬ１１乃至ＣＬ１ｎのコンピュータ装置の記憶装置に記憶された分類システム１００によって分類された分類結果を表示することができる。

［００７９］

図４は、図３で説明した分類システム１００を説明するブロック図である。分類システム１００は、ＧＵＩ（Ｇｒａｐｈｉｃａｌ　Ｕｓｅｒ　Ｉｎｔｅｒｆａｃｅ）１１０、演算部１２０、および記憶部１３０を有している。ＧＵＩ１１０は、入力部１１１および出力部１１２を有している。入力部１１１は、コンテンツのロード元を選択する機能と、学習用ラベルを入力する機能とを有する。出力部１１２は、データベースなどからロードしたコンテンツリストを表示する機能と、分類モデルが出力する判定情報を表示する機能とを有する。なお表示されるコンテンツに含まれるメタ情報は、ＧＵＩを介してユーザが修正することができる。

［００８０］

演算部１２０は、データ処理部１２１、学習処理部１２２、および判定処理部１２３を有している。データ処理部１２１は、データ収集部およびデータ生成部を有している。学習処理部１２２は、分類モデルを作成する分類モデル生成部、および分類モデルを分類する分類モデル評価部を有している。なお、分類モデル評価部の出力結果は、ユーザによる判定が行われる評価結果判定処理の機能を有している。判定処理部１２３は、分類推論部と、分類推論部によって分類された結果をリスト化する出力リスト作成部を有する。演算部１２０は、コンピュータ装置が有する記憶部に記憶されたプログラムがマイクロプロセッサを用いて演算処理する。ただし、プログラムは、ＤＳＰ（Ｄｉｇｔａｌ　ｓｉｇｎａｌ　Ｐｒｏｃｅｓｓｏｒ）、またはＧＰＵ（Ｇｒａｐｈｉｃｓ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）を用いて演算処理することができる。

［００８１］

記憶部１３０には、データベースなどからロードして生成したコンテンツおよびメタ情報がリスト化されて一時的に記憶される。

［００８２］

記憶部１３０は、例えば、１Ｔ（トランジスタ）１Ｃ（容量）型のメモリセルを備えたＤＲＡＭ（ダイナミックランダムアクセスメモリ）を用いることができる。なお、ＤＲＡＭのメモリセルに用いられるトランジスタは、ＯＳトランジスタを用いてもよい。ＯＳトランジスタは、半導体層に金属酸化物を有するトランジスタである。メモリセルにＯＳトランジスタが用いられるメモリ装置を「ＯＳメモリ」と呼ぶ。ここでは、ＯＳメモリの一例として、１Ｔ１Ｃ型のメモリセルを有するＲＡＭのことを、「ＤＯＳＲＡＭ（Ｄｙｎａｍｉｃ　Ｏｘｉｄｅ　Ｓｅｍｉｃｏｎｄｕｃｔｏｒ　ＲＡＭ）」と呼ぶ。

［００８３］

ＯＳトランジスタのオフ電流は非常に小さい。したがってＤＯＳＲＡＭは、リフレッシュの頻度を低減できるためリフレッシュ動作に要する電力を削減できる。ここでいう、オフ電流とは、トランジスタがオフ状態の場合、ソースとドレインとの間に流れる電流をいう。トランジスタがｎチャネル型である場合、例えば、しきい値電圧が０Ｖ乃至２Ｖ程度であれば、ゲートとソース間の電圧が負の電圧であるときのソースとドレインとの間に流れる電流をオフ電流と呼ぶことができる。

［００８４］

図５Ａは、ＧＵＩ３０の構成を説明する図である。ＧＵＩ３０は、一例として、ｐ個の学習用コンテンツをリスト表示する管理画面を示す。学習用コンテンツは、レコード単位で管理される。当該レコードには、番号（Ｎｏ）３１、コンテンツ（ＩＤ）３２、特徴量を示すメタ情報（Ｆｅａｔｕｒｅ）３３（メタ情報（Ｆ１）３３ａ乃至メタ情報（Ｆｍ）３３ｍ）、分類情報（Ｃａｓｅ）３４（分類情報（Ｃ１）３４ａ乃至分類情報（Ｃｑ）３４ｑ）、学習用ラベル（Ｊ－Ｌａｂｅｌ）３５などが含まれる。一例として、図５Ａでは、学習用ラベル３５が、“Ｙｅｓ”、“Ｎｏ”の２値のいずれかの値を与えるが、学習用ラベル３５は、２値に限定されず、３値以上でもよい。

［００８５］

図５Ｂは、ＧＵＩ３０Ａの構成を説明する図である。ＧＵＩ３０Ａは、評価推論部にてｎ個の未分類のコンテンツを推論し、当該推論した判定情報をリスト表示する管理画面を示す。未分類のコンテンツは、学習用コンテンツと同じく、番号３１、コンテンツ３２、メタ情報３３、分類情報３４を有する。さらに、それぞれのレコードには、分類ラベル（Ａ－Ｌａｂｅｌ）３６、スコア（Ｓｃｏｒｅ）３７が判定情報として付与される。

［００８６］

なお、ＧＵＩ３０およびＧＵＩ３０Ａは、同じ表示画面で管理することができる。後述する図９または図１０では、学習用コンテンツと判定情報とを同じ管理画面に表示することができるＧＵＩの表示例を示す。

［００８７］

図６は、機械学習によって上述した学習用コンテンツＳａｍｐｌｅに関連付けられた複数の特徴Ｆｅａｔｕｒｅを用いて分類モデルを生成する方法を説明する図である。特徴Ｆｅａｔｕｒｅは、それぞれがメタ情報のいずれか一を示し、コンテンツを管理するための管理パラメータに相当する。なお、本実施の一態様では、演算部Ｆ、演算部Ｓ、演算部Ｖ、第１の分類モデル、および第２の分類モデルを用いて分類モデルの生成方法を説明する。

［００８８］

学習用コンテンツＳａｍｐｌｅ（１）乃至学習用コンテンツＳａｍｐｌｅ（ｋ）は、それぞれｊ個の特徴Ｆｅａｔｕｒｅ、および学習用ラベルＬａｂｅｌが付与されている。一例として、演算部Ｆ１は、学習用コンテンツＳａｍｐｌｅ（１）からコンピュータが処理できる形式の特徴ベクトルＶｌａｂｅｌ１（１）を生成することができる。また、演算部Ｆｋは、学習用コンテンツＳａｍｐｌｅ（ｋ）からコンピュータが処理できる形式の特徴ベクトルＶｌａｂｅｌ１（ｋ）を生成することができる。なお、特徴ベクトルＶｌａｂｅｌ１（１）は、演算部Ｆ１がそれぞれの特徴に対して異なる重み係数を与えることで生成することができる。また、特徴ベクトルＶｌａｂｅｌ１（１）は、ランダムに選択されるｊ個以下の特徴Ｆｅａｔｕｒｅを用いて生成することができる。

［００８９］

次に、複数の第１の分類モデルを生成する。演算部Ｓ１乃至演算部Ｓｍは、それぞれ異なる第１の分類モデルに相当する。一例として、演算部Ｓ１が特徴ベクトルＶｌａｂｅｌ１（１）乃至特徴ベクトルＶｌａｂｅｌ１（ｋ）を用いて第１の分類モデルを生成することができる。なお、演算部Ｓ１に与えられる特徴ベクトルＶｌａｂｅｌ１はｋ個以下であればよい。異なる例として、演算部Ｓｍが異なる特徴ベクトルＶｌａｂｅｌ１（１）乃至特徴ベクトルＶｌａｂｅｌ１（ｋ）を用いて第１の分類モデルを生成することができる。よって、異なる２つの第１の分類モデルは、それぞれｋ個以下の特徴ベクトルＶｌａｂｅｌ１を含み、且つ、特徴ベクトルＶｌａｂｅｌ１のいずれか一

が同じ特徴ベクトルを含むことができる。

［００９０］

第１の分類モデルを生成するために選択されるｋ個の学習用コンテンツＳａｍｐｌｅは、ランダムに選択されてもよいし、学習用コンテンツに付与された番号でソートされた順番に選択してもよい。学習用コンテンツがランダムに選択される場合、第１の分類モデルは学習用コンテンツのばらつきを含むことができる。また、学習用コンテンツに付与された番号でソートされた順番に選択される場合、時系列、またはメタ情報のいずれか一の特徴に基づいて付与された番号に沿った傾向を含むことができる。

［００９１］

したがって、第１の分類モデルは、学習用コンテンツＳａｍｐｌｅ（１）乃至学習用コンテンツＳａｍｐｌｅ（ｋ）から生成される特徴ベクトルＶｌａｂｅｌ１を用いて特徴ベクトルＶｌａｂｅｌ２を生成することができる。

［００９２］

第２の分類モデルは、演算部Ｖ１によって生成される。一例として、演算部Ｖ１は、ｍ個の特徴ベクトルＶｌａｂｅｌ２を用いて第２の分類モデルを生成するステップを有する。なお、第２の分類モデルは、特徴ベクトルＶｌａｂｅｌ２（１）乃至特徴ベクトルＶｌａｂｅｌ２（ｍ）を用いることで異なる特徴を有する分類モデルを生成することができる。

［００９３］

したがって、第２の分類モデルは、学習用コンテンツＳａｍｐｌｅ（１）乃至学習用コンテンツＳａｍｐｌｅ（ｋ）から生成される特徴ベクトルＶｌａｂｅｌ１を用いて出力値ＰＯＵＴを出力することができる。ＧＵＩは、出力値ＰＯＵＴを表示することができる。なお、出力値ＰＯＵＴには、判定情報である分類ラベルと、スコアとが含まれる。したがって、第２の分類モデルは、コンテンツの分類を行うことができる。また第２の分類モデルは、それぞれのコンテンツに判定情報を付与することができる。

［００９４］

図６で説明した当該分類モデルを用いて推論するには、当該分類モデルの学習用コンテンツＳａｍｐｌｅに未分類のコンテンツを与えることで判定結果を得る。ただし、未分類のコンテンツには、学習用コンテンツと異なり学習用ラベルが付与されていない。

［００９５］

図７は、図６と異なる分類モデルを生成する方法を説明する図である。図７では、図６と異なる点について説明し、発明の構成（または実施例の構成）において、同一部分または同様な機能を有する部分には同一の符号を異なる図面間で共通して用い、その繰り返しの説明は省略する。

［００９６］

図７では、ｍ個の特徴ベクトルＶｌａｂｅｌ２の平均値Ａｖを算出し、特徴ベクトルＶｌａｂｅｌ＿ａを生成する。第２の分類モデルは、ｐ個の特徴ベクトルＶｌａｂｅｌ＿ａを用いて生成することができる。第２の分類モデルは、ｍ個の特徴ベクトルＶｌａｂｅｌ２の平均値Ａｖを算出することで、異なる特徴を有する分類モデルを生成することができる。生成される分類モデルは、コンテンツの分類を正確にすることができる。

［００９７］

図８は、図７とは異なる分類モデルを生成する方法を説明する図である。図８では、図７と異な

る点について説明し、発明の構成（または実施例の構成）において、同一部分または同様な機能を有する部分には同一の符号を異なる図面間で共通して用い、その繰り返しの説明は省略する。

［００９８］

図８では、ｍ個の特徴ベクトルＶｌａｂｅｌ２を評価するための評価基準が評価判定部ＪＧに与えられる。例えば、評価判定部ＪＧ１には、第一の評価基準として精度（Ｐｒｅｃｉｓｉｏｎ）が与えられそれぞれの特徴ベクトルＶｌａｂｅｌ２（１）を評価することができる。次に、評価判定部ＪＧ１は、第二の評価基準として感度（Ｓｅｎｓｉｔｉｖｉｔｙ）が与えられそれぞれの特徴ベクトルＶｌａｂｅｌ２（１）を評価することができる。評価判定部ＪＧ１は、評価結果Ｖｌａｂｅｌ＿ｂ（１）を出力する

［００９９］

第２の分類モデルは、評価結果Ｖｌａｂｅｌ＿ｂ（１）乃至評価結果Ｖｌａｂｅｌ＿ｂ（ｐ）を用いて生成される。例えば、複数の特徴ベクトルＶｌａｂｅｌ２は、異なる第１の評価基準および第２の評価基準によって評価されてもよいし、同じ評価基準で評価されてもよい。なお、図８では記載していないが、図７と同様に、第１の評価基準、および第２の評価基準による評価結果Ｖｌａｂｅｌ＿ｂの平均値を算出することができる。

［０１００］

第２の分類モデルは、ｍ個の特徴ベクトルＶｌａｂｅｌ２の評価結果を用いることで、異なる特徴を有する分類モデルを生成することができる。生成される分類モデルは、コンテンツの分類を正確にすることができる。

［０１０１］

図９は、ＧＵＩ５０を説明する図である。ＧＵＩ５０は、コンテンツ（学習用コンテンツ、未分類のコンテンツ、分類済のコンテンツ）の表示領域、コンテンツを含むファイルのダウンロード元を選択するアイコン５８ａ、選択されたファイルが保存されるアドレス情報を表示するテキストボックス５８ｂ、機械学習を実行するアイコン（Ｌｅａｒｎｉｎｇ　Ｓｔａｒｔ）５９を有している。

［０１０２］

表示領域には、一例として８個のレコードがロードされている例を示している。それぞれのレコードは、番号（Ｎｏ）５１、ＩＤ（Ｉｎｄｅｘ）５２、特徴量（Ｆｅａｔｕｒｅ）５３、分類情報（Ｃａｓｅ）５４、学習用ラベル（ＪＬ）５５、分類ラベル（ＡＬ）５６、およびスコア（Ｐｒｏｂ）５７の構成要素を有する。特徴量５３は、詳細な情報として特徴量Ｆ（１）５３ａ乃至特徴量Ｆ（ｊ）５３ｊを表示することができる。なお、ｊは自然数である。また、分類情報５４は、詳細な情報として分類情報Ｃ（１）５４ａ乃至分類情報Ｃ（４）５４ｄを表示することができる。なお、分類情報は、自然数で表すことができる種類を有することができる。

［０１０３］

なお、図９は、分類モデルによって学習用コンテンツおよび未分類のコンテンツが分類された結果をＧＵＩ５０に表示する例を示している。

［０１０４］

一例として、レコード番号のＮｏ１乃至Ｎｏ３は、学習用コンテンツに相当する。学習用コンテンツには、学習用ラベルが与えられ、且つ、レコード番号Ｎｏ１乃至Ｎｏ３には、分類情報が与えられている。

［０１０５］

レコード番号のＮｏ４乃至Ｎｏ８は、分類済のコンテンツに相当する。分類済のコンテンツには、分類ラベル５６、スコア５７が与えられている。なお、図９では、レコード番号Ｎｏ１およびレコード番号Ｎｏ３を学習することで得られる分類モデルを用いて、レコード番号Ｎｏ４乃至レコード番号Ｎｏ７を分類した結果を表示している。一例として、レコード番号Ｎｏ２を学習することで得られた分類モデルを用いて、レコード番号Ｎｏ８を分類した結果を表示している。図９ではスペースの関係上レコード数は８までしか表示していないが、レコード数は複数の種類を扱うことができる。

［０１０６］

但し、レコード数を大量に扱う場合、表示上の課題がある。したがって、分類ラベル５６またはスコア５７はソート機能を備えることが好ましい。ソート条件の一例として、ＧＵＩは、分類ラベル５６が“Ｙｅｓ”の判定結果を選択して表示することができる。また、ＧＵＩは、スコア５７の数値範囲を指定して表示することができる。ＧＵＩに上述したソート条件を与える場合、ＧＵＩは、教師データを与えた学習コンテンツと同じような特徴を有するコンテンツを分類し表示することができる。

［０１０７］

例えば、コンテンツが特許番号の場合について説明する。特許番号には、複数のメタ情報が付与されている。特許の権利が維持されている特許番号の場合、学習用ラベルには、“Ｙｅｓ”を与える。特許の権利が放棄されている特許番号の場合、学習用ラベルには、“Ｎｏ”を与える。続いて機械学習を実行し分類モデルを生成する。

［０１０８］

上述した分類モデルは、未分類のコンテンツに対し、判定情報を付与することができる。判定情報には、分類ラベル５６と、スコア５７が表示される。一例として、ユーザは、ソート機能を用いて分類ラベル５６に“Ｎｏ”を与える。さらに、スコア５７には、“０．８”乃至“１．０”を設定する。ＧＵＩに上述したソート条件を与えることで、ＧＵＩは、特許番号が放棄された学習用コンテンツと同じような特徴を有するレコードを選択して表示することができる。

［０１０９］

図１０は、図９とは異なるＧＵＩ５０Ａを説明する図である。図１０は、レコード数を大量に扱う場合の効率的なＧＵＩの表示例である。なお、図１０では、図９と異なる点について説明し、発明の構成（または実施例の構成）において、同一部分または同様な機能を有する部分には同一の符号を異なる図面間で共通して用い、その繰り返しの説明は省略する。

［０１１０］

図１０では、任意の選択した分類情報に関するレコードごと分類して表示することができる点が図９と異なっている。図１０では、分類情報Ｃ（１）乃至分類情報Ｃ（４）の種類に応じて表示を切替ることができる。

［０１１１］

ユーザが、レコードに付与された複数の特徴量５３、および分類モデルの判定情報を見て、十分な分類精度が得られていれば分類モデルの更新を終了する。ユーザが、レコードに付与された判定情報を見て、分類精度が十分でないと判断した場合は、ユーザ指定ラベルが付与されていないレコードに対して学習用ラベルを付与し、アイコン５９を押すことで、分類モデルの更新を行うことができる。なお、学習用コンテンツが有するメタ情報の経時的変化を予測し、特徴量５３を更新して

もよい。ユーザが特徴量５３を更新する場合、分類モデルは、分類モデルの経時的な変化を含むことができる。したがって、ユーザは、コンテンツの経時的な分類の変化を得ることができる。分類モデルは、価値が増大していくと予想されるコンテンツ群、または価値が減少していくと予想されるコンテンツ群を分類していくことができるようになる。

［０１１２］

図示していないが、特徴量５３、分類情報５４ａ乃至分類情報５４ｄ、学習用ラベル５５、分類ラベル５６、またはスコア５７に含まれる数値やラベル情報は、表示させる並び順を変更することができる、もしくは、選択した数値やラベル情報は、フィルタ機能を用いて必要な並び順にソートし表示することができる。これにより、ユーザは、分類モデルの判定結果を効率良く評価することができる。

［０１１３］

図１乃至図１０を用いて説明したコンテンツの分類方法は、確率の高い情報を分類する方法を提供することができる。例えば、ＧＵＩは、確率の高い情報を分類するのに適している。プログラムは、分類モデルに新しい教師データ（学習用ラベル）が与えられることで分類モデルを更新することができる。プログラムは、分類モデルが更新されることで確率の高い情報を分類することができる。

［０１１４］

さらに、生成した分類モデルは電子機器本体または外部のメモリに保存することができ、新たなファイルの分類の際に呼び出して使うことができる。さらに、新たな教師データを追加しながら上記で説明した方法に従って分類モデルを更新することができる。

［０１１５］

以上、本実施の形態で示す構成、方法は、他の実施の形態で示す構成、方法と適宜組み合わせて用いることができる。

［符号の説明］

［０１１６］

ＣＬ１：クライアントコンピュータ、ＣＬ１ｎ：クライアントコンピュータ、ＣＬ１１：クライアントコンピュータ、ＣＬｎ：クライアントコンピュータ、ＤＢ１：データベース、ＤＢ２：データベース、ＬＡＮ１：通信網、ＬＡＮ２：通信網、Ｖｌａｂｅｌ１：特徴ベクトル、Ｖｌａｂｅｌ２：特徴ベクトル、３１：番号、３２：コンテンツ、３３：メタ情報、３４：分類情報、３５：学習用ラベル、５０：ＧＵＩ、５０Ａ：ＧＵＩ、５１：番号、５３：特徴量、５４：分類情報、５６：分類ラベル、５７：スコア、５８ａ：アイコン、５８ｂ：テキストボックス、５９：アイコン、１００：分類システム、１１０：ＧＵＩ、１１１：入力部、１１２：出力部、１２０：演算部、１２１：データ処理部、１２２：学習処理部、１２３：判定処理部、１３０：記憶部

請求の範囲

［請求項１］

学習用コンテンツと、コンテンツと、を有し、

前記学習用コンテンツには、第１の特徴量および学習用ラベルが付与され、

前記コンテンツには、第２の特徴量が付与され、

複数の前記学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成するステップと、

前記複数の第１の分類モデルを用いて第２の分類モデルを生成するステップと、

前記第２の分類モデルを用いて複数の前記コンテンツに判定情報を付与しグラフィカルユーザインターフェース内に表示するステップと、

を含むコンテンツの分類方法。

［請求項２］

学習用コンテンツと、コンテンツと、を有し、

前記学習用コンテンツには、第１の特徴量および学習用ラベルが付与され、

前記コンテンツには、第２の特徴量が付与され、

複数の前記学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成するステップと、

前記複数の第１の分類モデルの出力から平均値を算出するステップと、

複数の前記平均値を用いて第２の分類モデルを生成するステップと、

前記第２の分類モデルを用いて複数の前記コンテンツに判定情報を付与しグラフィカルユーザインターフェース内に表示するステップと、

を含むコンテンツの分類方法。

［請求項３］

学習用コンテンツと、コンテンツと、を有し、

前記学習用コンテンツには、第１の特徴量および学習用ラベルが付与され、

前記コンテンツには、第２の特徴量が付与され、

複数の前記学習用コンテンツを用いて複数の第１の分類モデルを機械学習によって生成するステップと、

前記複数の第１の分類モデルがそれぞれ第一の評価基準による評価をするステップと、

前記複数の第１の分類モデルがそれぞれ第二の評価基準による評価をするステップと、

複数の前記第一の評価基準による評価結果と、前記第二の評価基準と、による評価結果から第２の分類モデルを生成するステップと、

前記第２の分類モデルを用いて複数の前記コンテンツに判定情報を付与しグラフィカルユーザインターフェース内に表示するステップと、

を含むコンテンツの分類方法。

［請求項４］

請求項３において、

前記第一の評価基準は精度であり、

前記第二の評価基準は感度であるコンテンツの分類方法。

［請求項５］

請求項１乃至３のいずれかにおいて、

任意の前記学習用コンテンツを用いて前記第１の分類モデルを生成するステップを含むコンテンツの分類方法。

［請求項６］

請求項１乃至３のいずれかにおいて、

前記学習用コンテンツには、さらに分類情報が与えられ、

前記第２の分類モデルの出力を用いて分類ラベルが付与された複数の前記コンテンツから前記分類情報と同じ前記判定情報を有するコンテンツを選択して前記グラフィカルユーザインターフェース内に表示するステップと、

を含むコンテンツの分類方法。

［請求項７］

請求項１乃至３のいずれかにおいて、

前記学習用コンテンツ又前記コンテンツに与えられる特徴量は、管理パラメータであるコンテンツの分類方法。

［請求項８］

請求項１乃至３のいずれかにおいて、

前記判定情報には、分類ラベルまたはスコアを含むコンテンツの分類方法。

［請求項９］

請求項８において、

前記グラフィカルユーザインターフェースは、前記スコアのうち特定の数値範囲を指定し、該当するコンテンツをリストとして表示するステップを含むコンテンツの分類方法。

図1

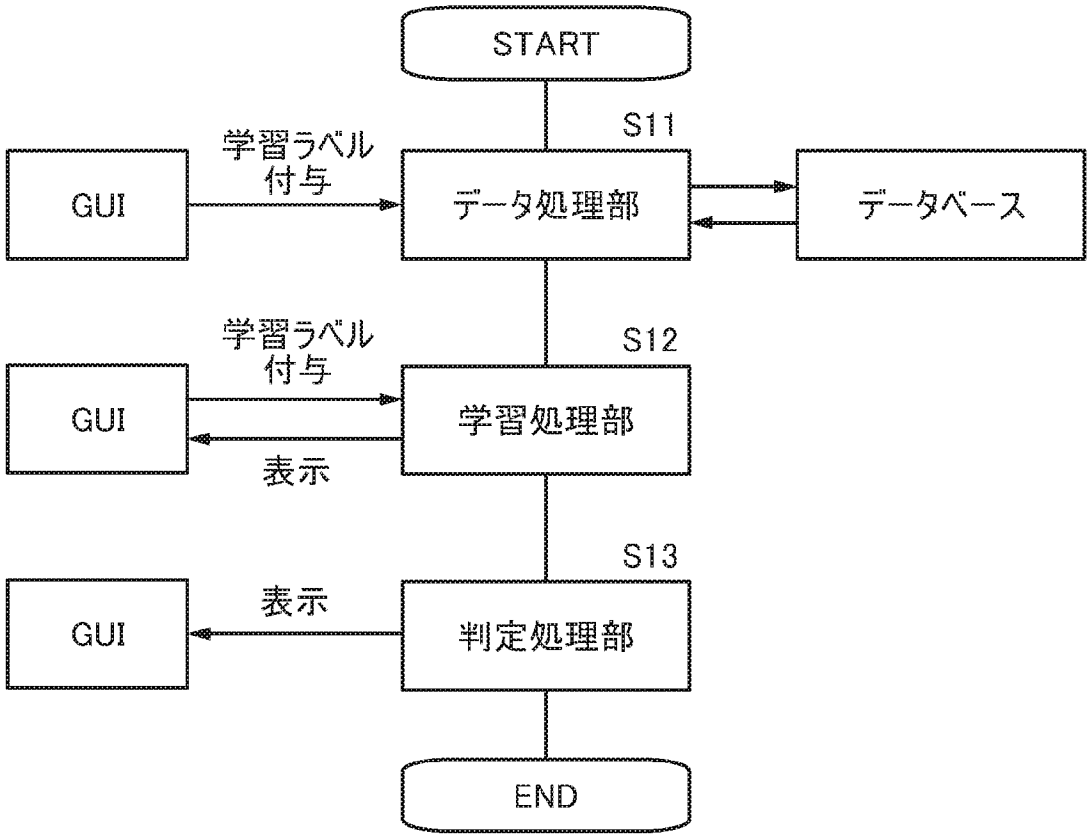

START
S11
学習ラベル
付与
GUI
データ処理部
データベース
学習ラベル
付与
S12
GUI
学習処理部
表示
S13
GUI
表示
判定処理部
END

図2

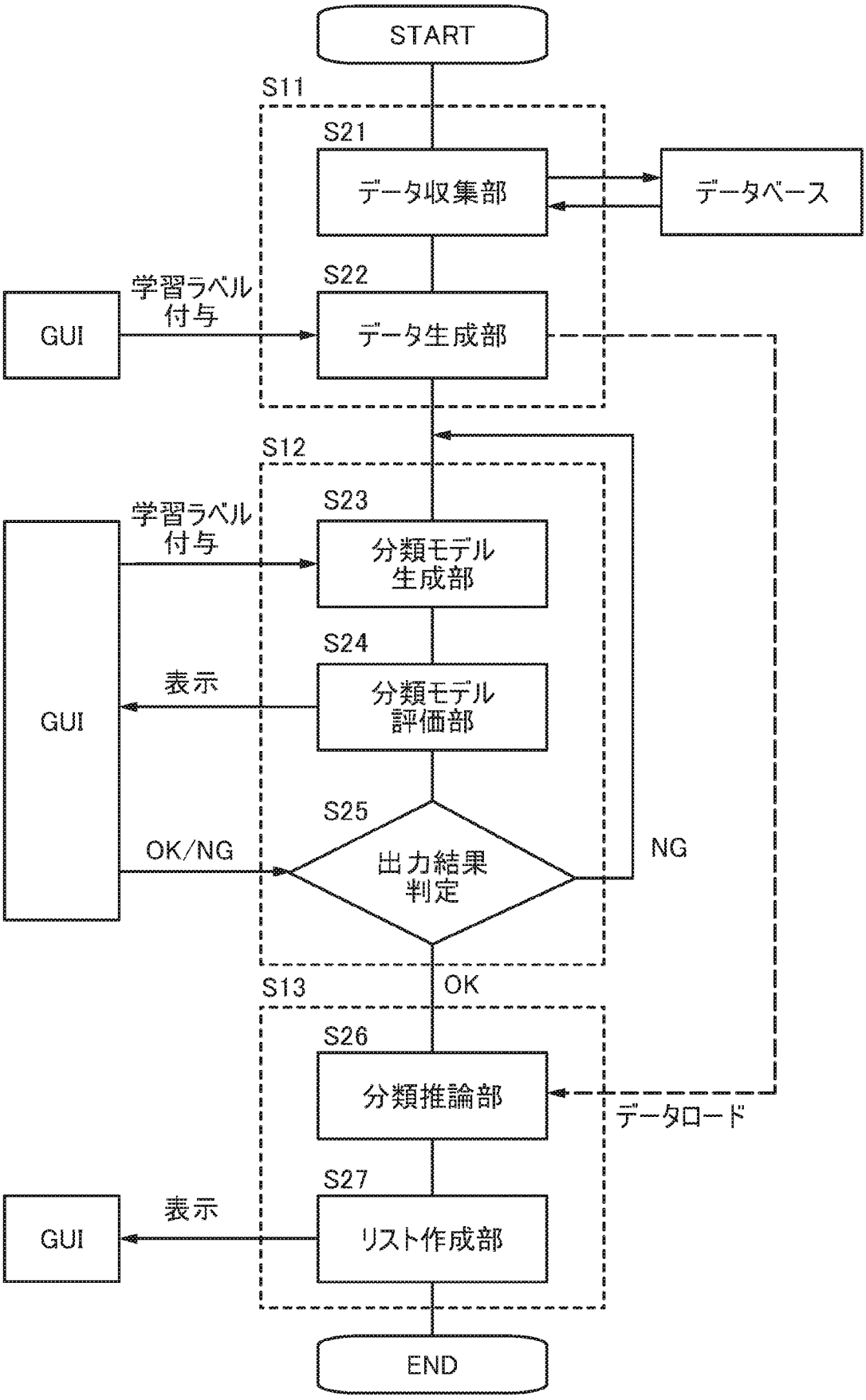

START
S11
S21
データ収集部
データベース
学習ラベル
付与
GUI
S22
データ生成部
S12
S23
学習ラベル
付与
分類モデル
生成部
S24
表示
GUI
分類モデル
評価部
S25
OK/NG
出力結果
判定
NG
S13
OK
S26
分類推論部
データロード
S27
表示
GUI
リスト作成部
END

図3

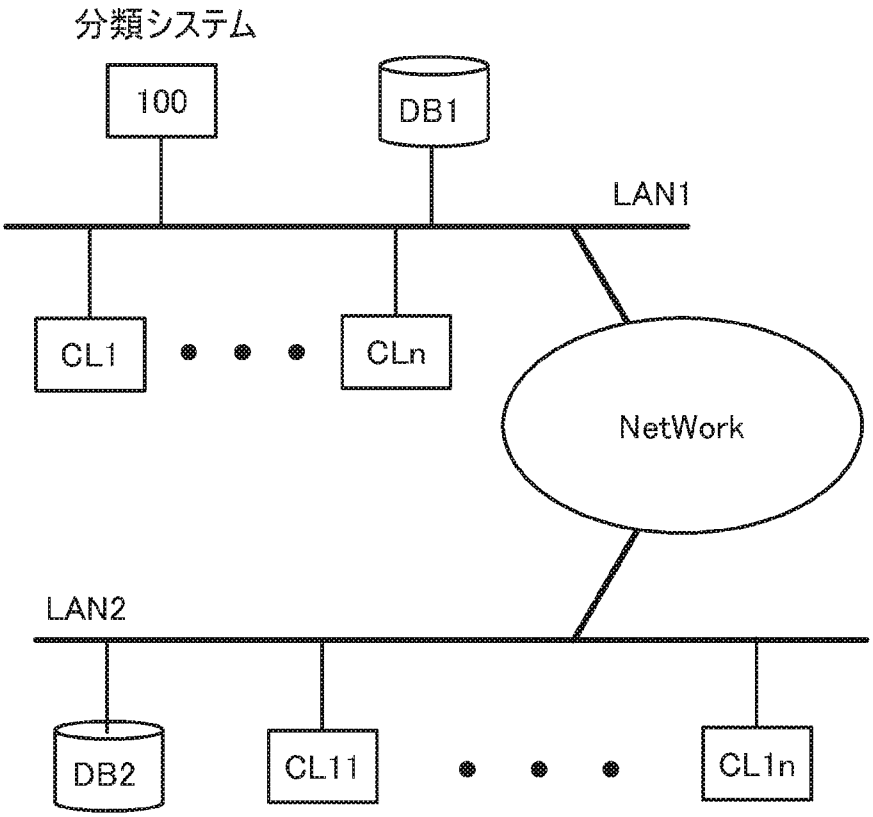


図4

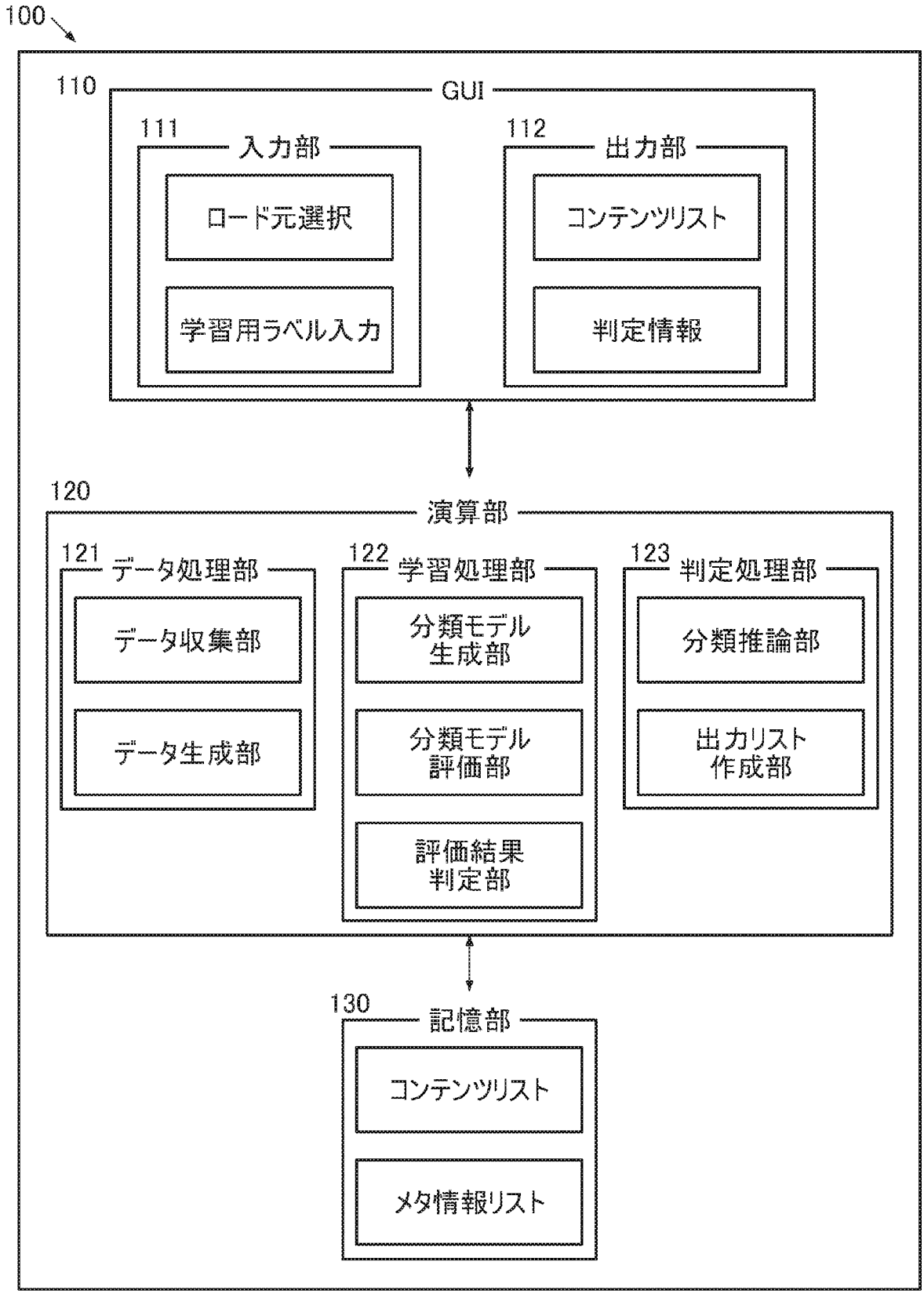

100
110
GUI
111
入力部
ロード元選択
学習用ラベル入力
112
出力部
コンテンツリスト
判定情報
120
演算部
121
データ処理部
データ収集部
データ生成部
122
学習処理部
分類モデル生成部
分類モデル評価部
評価結果判定部
123
判定処理部
分類推論部
出力リスト作成部
130
記憶部
コンテンツリスト
メタ情報リスト

図5A

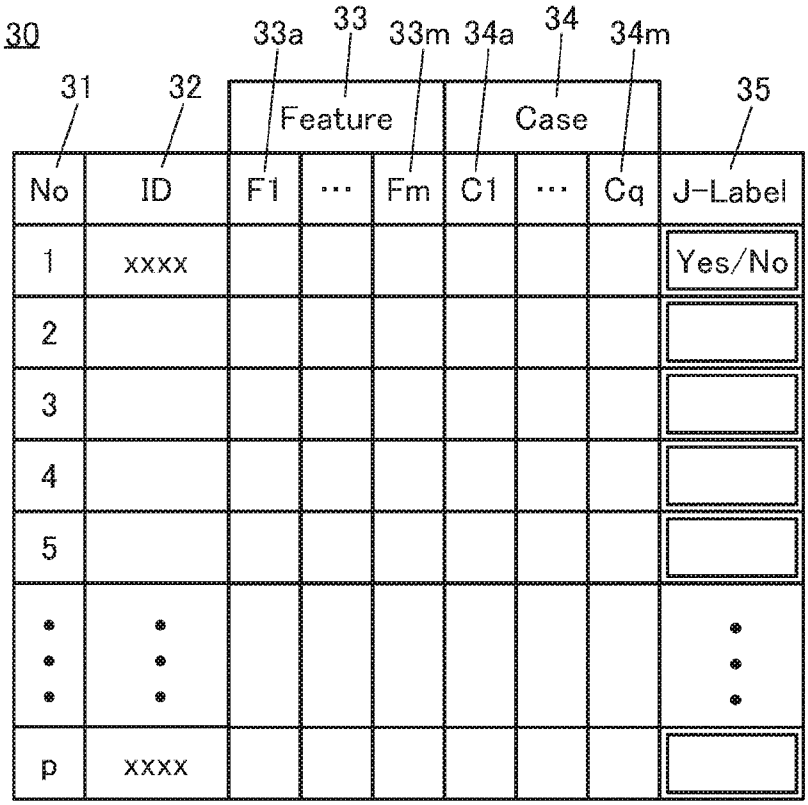

30
31
32
33a
33
33m
34a
34
34m
35
Feature
Case
No
ID
F1
…
Fm
C1
…
Cq
J-Label
1
xxxx
Yes/No
2
3
4
5
p
xxxx

図5B

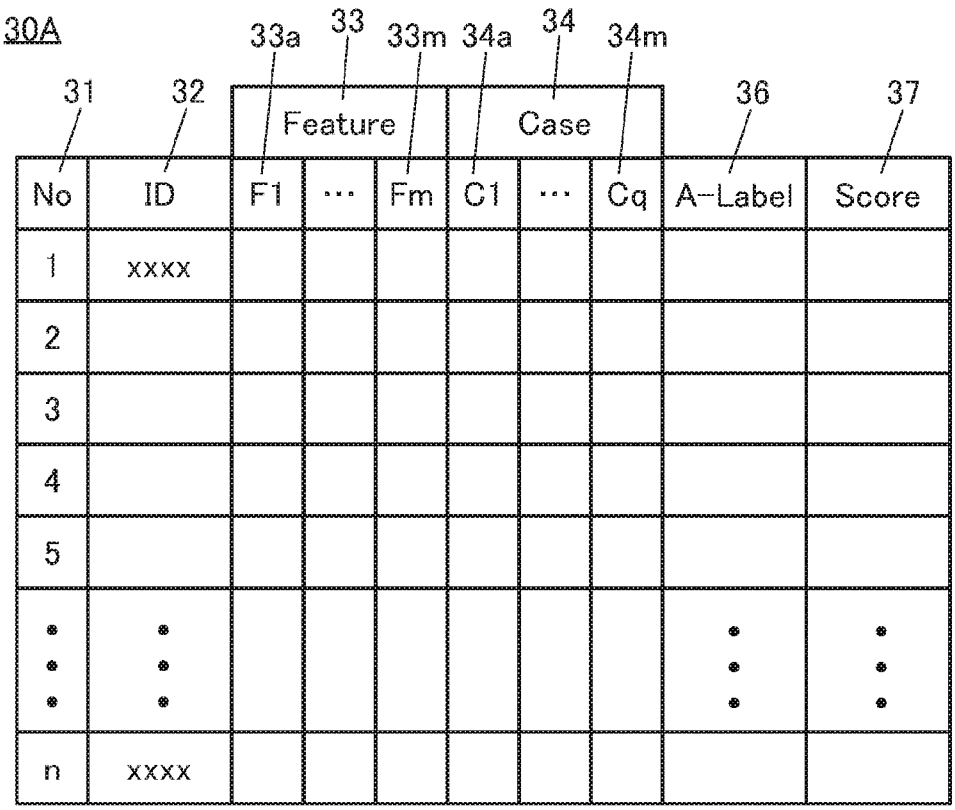

30A
31
32
33a
33
33m
34a
34
34m
36
37
Feature
Case
No
ID
F1
…
Fm
C1
…
Cq
A-Label
Score
1
xxxx
2
3
4
5
n
xxxx

図6

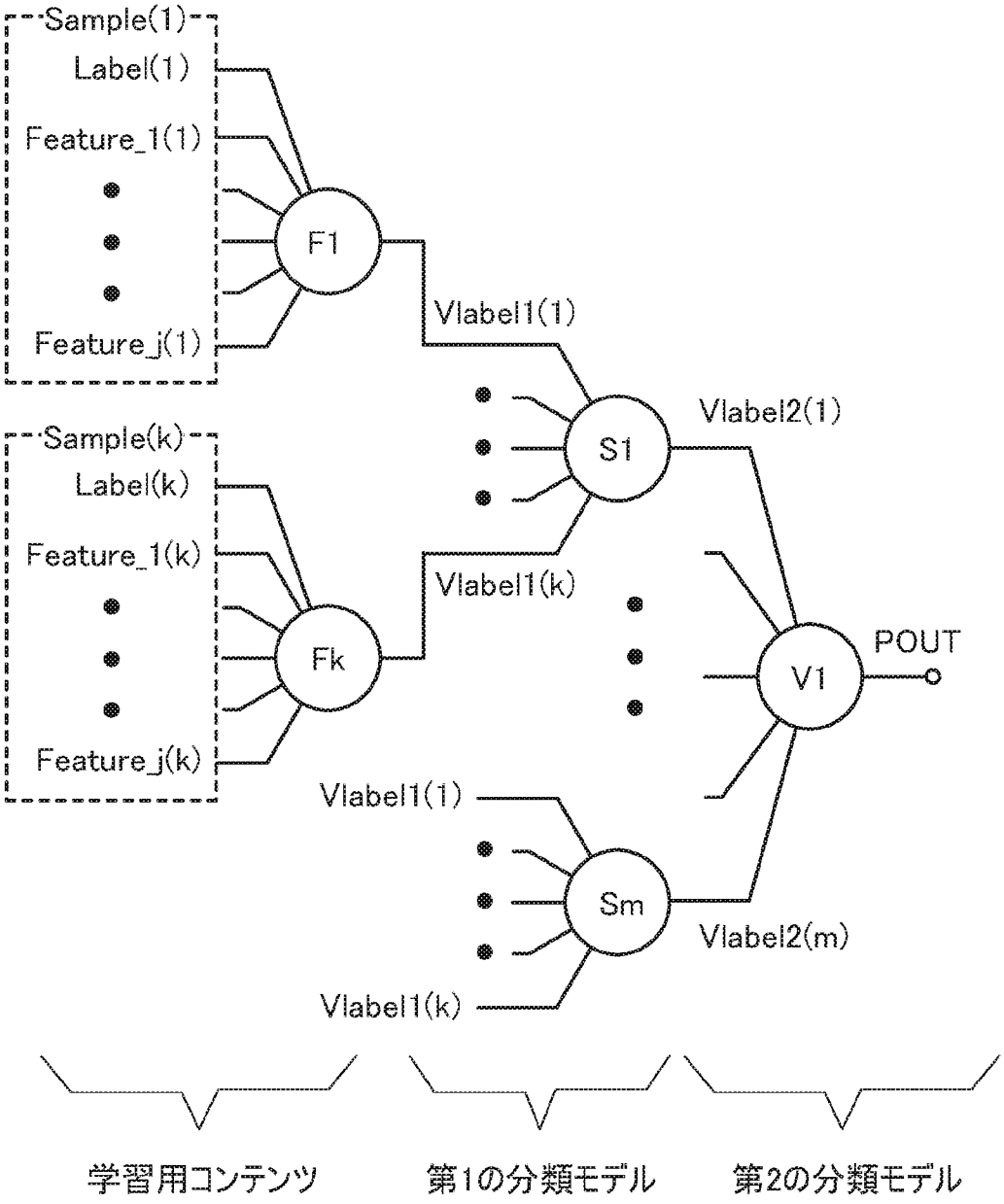

Sample(1)
Label(1)
Feature_1(1)
Feature_j(1)
F1
Vlabel1(1)
Sample(k)
Label(k)
Feature_1(k)
Feature_j(k)
Fk
Vlabel1(k)
S1
Vlabel2(1)
V1
POUT
Vlabel1(1)
Sm
Vlabel2(m)
Vlabel1(k)
学習用コンテンツ
第1の分類モデル
第2の分類モデル

図7

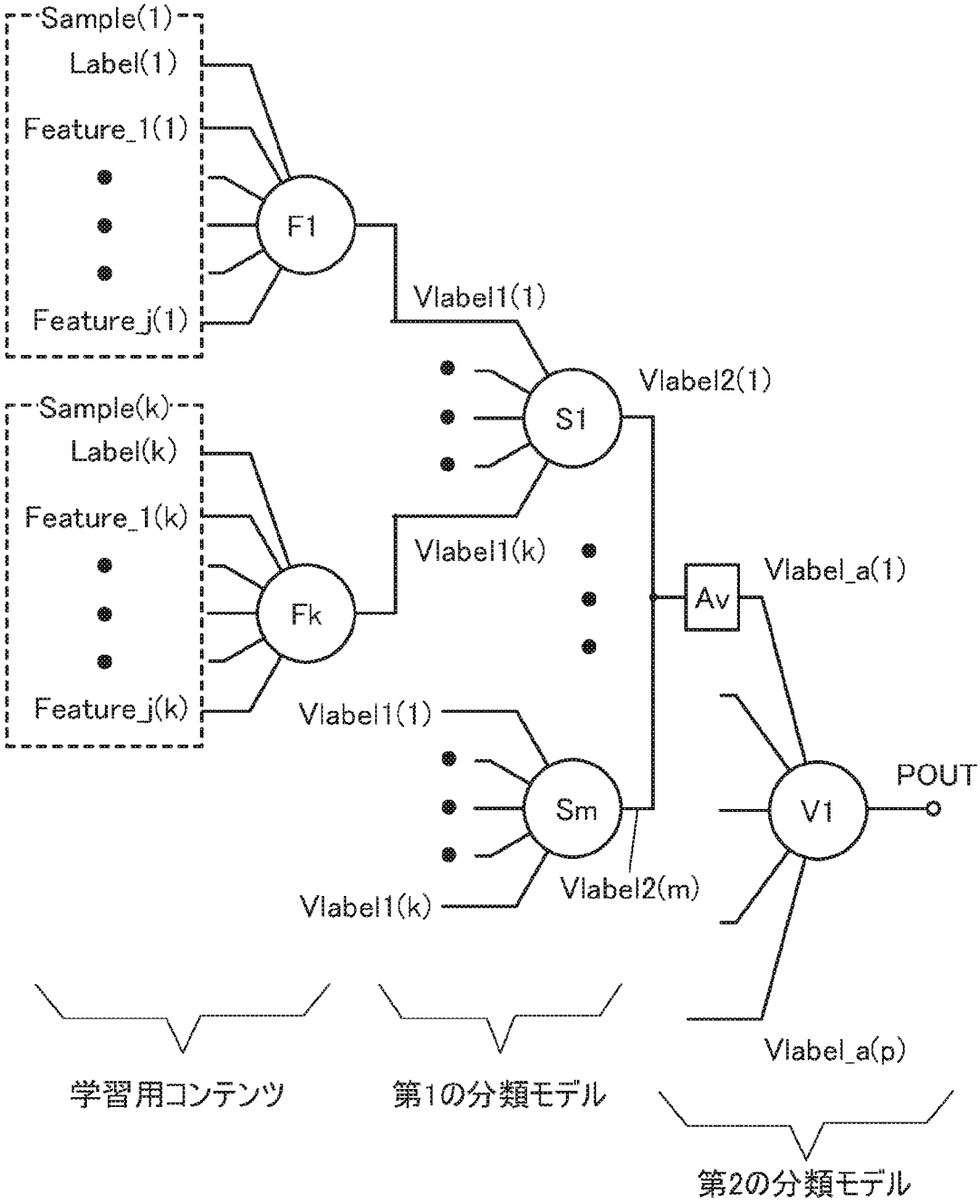

Sample(1)
Label(1)
Feature_1(1)
Feature_j(1)
F1
Vlabel1(1)
Sample(k)
Label(k)
Feature_1(k)
Feature_j(k)
Fk
Vlabel1(k)
S1
Vlabel2(1)
Av
Vlabel_a(1)
Vlabel1(1)
Vlabel1(k)
Sm
Vlabel2(m)
V1
POUT
Vlabel_a(p)
学習用コンテンツ
第1の分類モデル
第2の分類モデル

図8

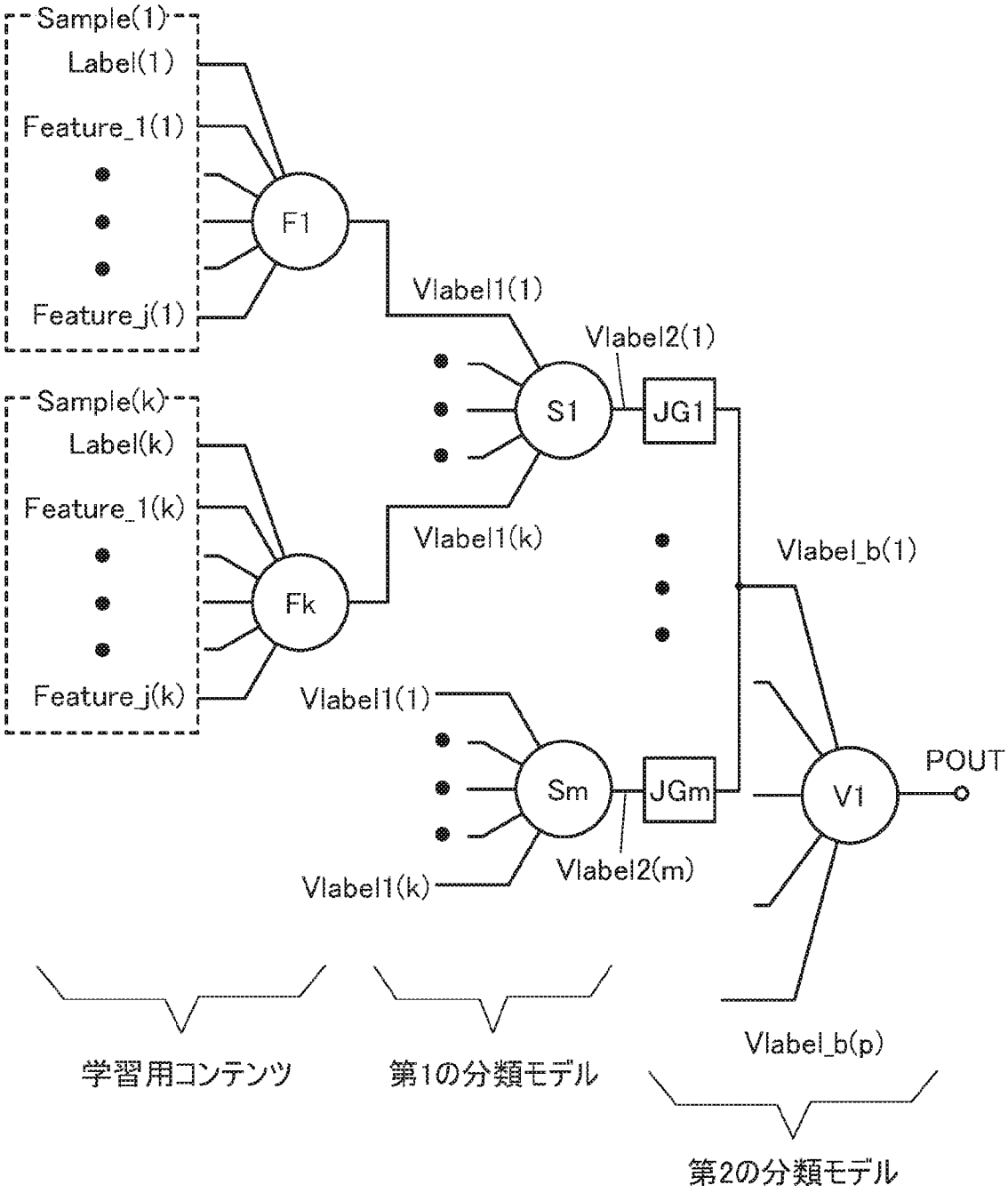

Sample(1)
Label(1)
Feature_1(1)
Feature_j(1)
F1
Vlabel1(1)
Vlabel2(1)
S1
JG1
Sample(k)
Label(k)
Feature_1(k)
Feature_j(k)
Fk
Vlabel1(k)
Vlabel_b(1)
Vlabel1(1)
Sm
JGm
V1
POUT
Vlabel1(k)
Vlabel2(m)
Vlabel_b(p)
学習用コンテンツ
第1の分類モデル
第2の分類モデル

図9

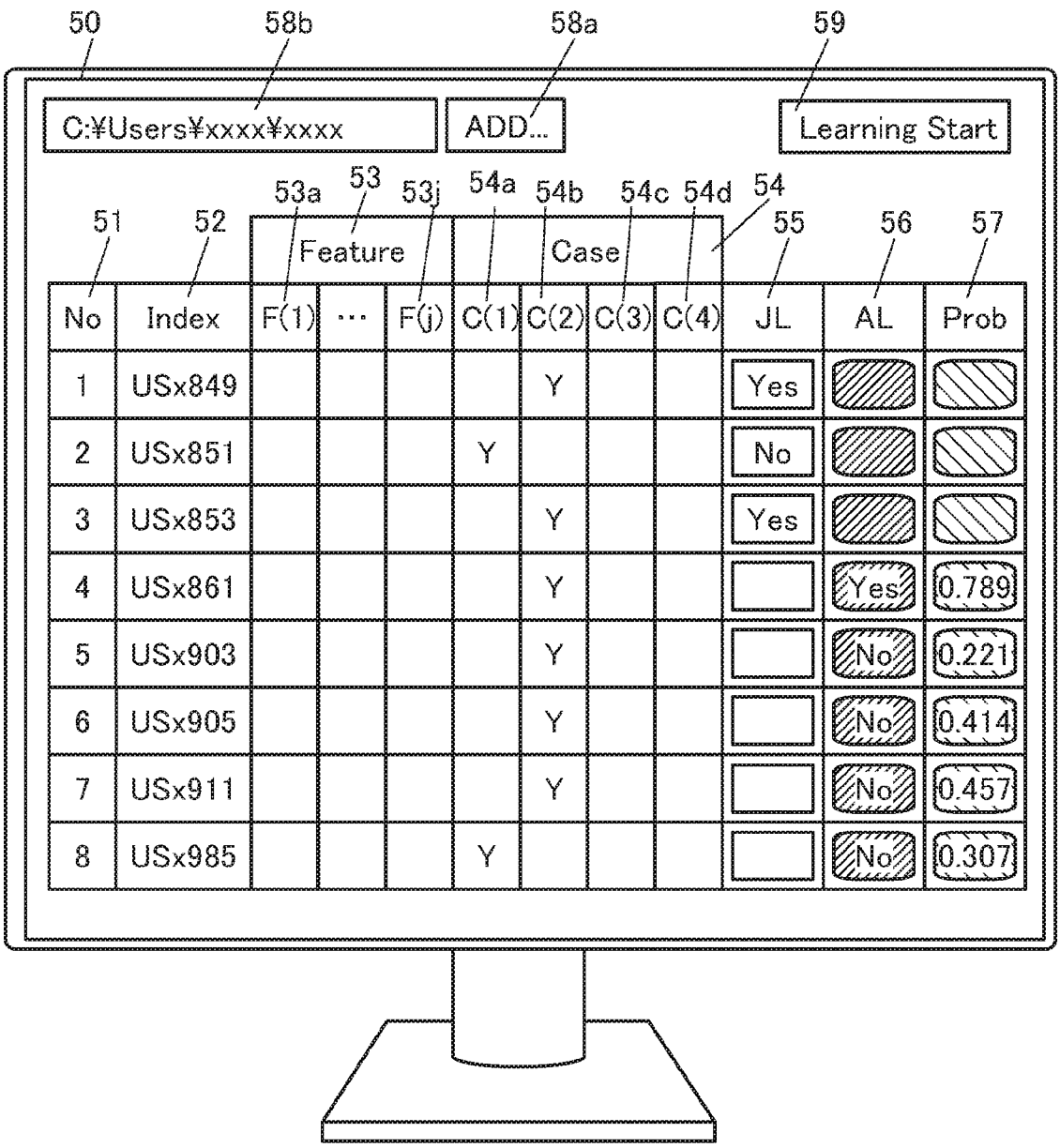
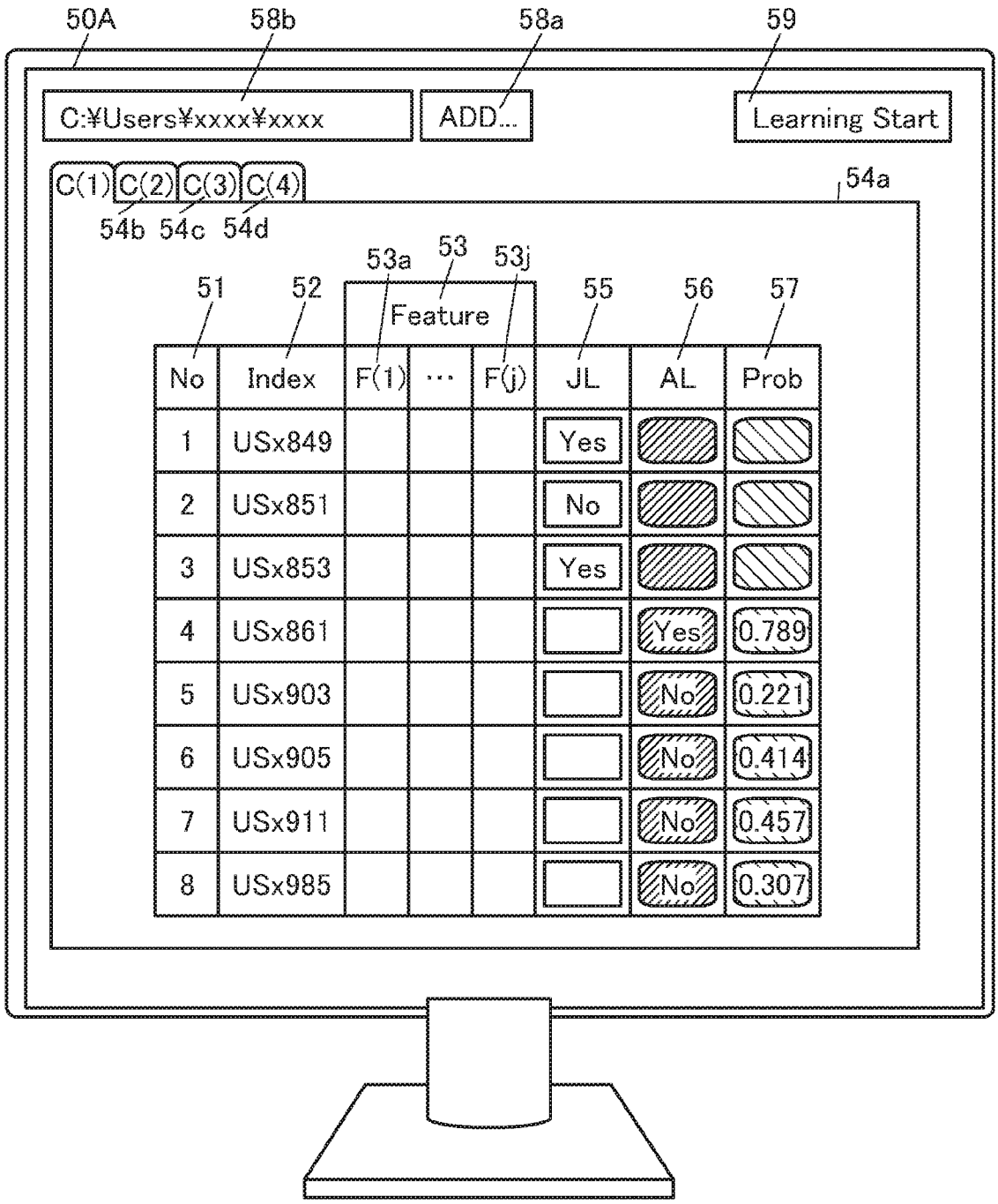

50
58b
58a
59
C:¥Users¥xxxx¥xxxx
ADD...
Learning Start
53a
53
53j
54a
54b
54c
54d
54
51
52
Feature
Case
55
56
57
No
Index
F(1)
…
F(j)
C(1)
C(2)
C(3)
C(4)
JL
AL
Prob
1 USx849 Y Yes
2 USx851 Y No
3 USx853 Y Yes
4 USx861 Y Yes 0.789
5 USx903 Y No 0.221
6 USx905 Y No 0.414
7 USx911 Y No 0.457
8 USx985 Y No 0.307

図10


50A
58b
58a
59
C:¥Users¥xxxx¥xxxx
ADD...
Learning Start
C(1)
C(2)
C(3)
C(4)
54b
54c
54d
54a
53a
53
53j
51
52
Feature
55
56
57
No
Index
F(1)
…
F(j)
JL
AL
Prob
1 USx849 Yes
2 USx851 No
3 USx853 Yes
4 USx861 Yes 0.789
5 USx903 No 0.221
6 USx905 No 0.414
7 USx911 No 0.457
8 USx985 No 0.307

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2019/060377

A. CLASSIFICATION OF SUBJECT MATTER
Int. Cl. G06F16/906(2019.01)i, G06N20/00(2019.01)i
FI: G06N20/00 130, G06F16/906

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. G06F16/906, 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 1971-2020
Registered utility model specifications of Japan 1996-2020
Published registered utility model applications of Japan 1994-2020

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.
Y	WO 2014/203328 A1 (HITACHI, LTD.) 24 December 2014, paragraphs [0020], [0069], [0070], fig. 12	1-9
Y	JP 2013-161298 A (NIPPON STEEL & SUMITOMO METAL CORP.) 19 August 2013, paragraphs [0040], [0082]-[0084]	1-9
Y	JP 2008-242880 A (KENWOOD CORP.) 09 October 2008, paragraphs [0036], [0037]	6

☐ Further documents are listed in the continuation of Box C.
☒ 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
"T" 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	Date of mailing of the international search report
29.01.2020	10.02.2020
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IB2019/060377

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
WO 2014/203328 A1	24.12.2014	(Family: none)	
JP 2013-161298 A	19.08.2013	(Family: none)	
JP 2008-242880 A	09.10.2008	(Family: none)	

国際調査報告

国際出願番号
PCT/IB2019/060377

A.　発明の属する分野の分類（国際特許分類（ＩＰＣ））

G06F 16/906(2019.01)i; G06N 20/00(2019.01)i
FI: G06N20/00 130; G06F16/906

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））

G06F16/906; 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	１９７１－２０２０年
日本国実用新案登録公報	１９９６－２０２０年
日本国登録実用新案公報	１９９４－２０２０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
Y	WO 2014/203328 A1 (株式会社日立製作所) 24.12.2014 (2014-12-24) 段落［００２０］、［００６９］～［００７０］、図１２	1-9
Y	JP 2013-161298 A (新日鐵住金株式会社) 19.08.2013 (2013-08-19) 段落［００４０］、［００８２］～［００８４］	1-9
Y	JP 2008-242880 A (株式会社ケンウッド) 09.10.2008 (2008-10-09) 段落［００３６］～［００３７］	6

☐ Ｃ欄の続きにも文献が列挙されている。

☑ 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の日の後に公表された文献
“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
“Y” 特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
“&” 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
29.01.2020	10.02.2020

名称及びあて先

日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）

福西　章人 5B 4687

電話番号 03-3581-1101 内線 3545

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/IB2019/060377

引用文献	公表日	パテントファミリー文献	公表日
WO 2014/203328 A1	24.12.2014	（ファミリーなし）	
JP 2013-161298 A	19.08.2013	（ファミリーなし）	
JP 2008-242880 A	09.10.2008	（ファミリーなし）	