

(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2018-200524

(P2018-200524A)

(43) 公開日 平成30年12月20日(2018.12.20)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 17/30 (2006.01)	G06F 17/30	210D
G06N 99/00 (2010.01)	G06F 17/30	220Z
	G06N 99/00	153
	G06N 99/00	156

審査請求 未請求 請求項の数 4 O L (全 11 頁)

(21) 出願番号	特願2017-104304 (P2017-104304)	(71) 出願人	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(22) 出願日	平成29年5月26日 (2017.5.26)	(74) 代理人	110002147 特許業務法人酒井国際特許事務所
		(72) 発明者	山中 友貴 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
		(72) 発明者	山田 真徳 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
		(72) 発明者	長谷川 彩子 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

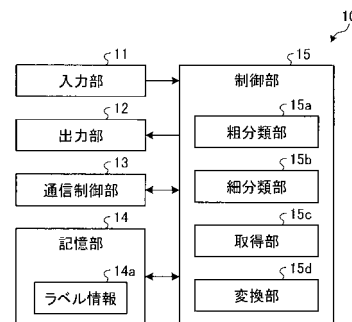
(54) 【発明の名称】 分類装置、分類方法および分類プログラム

(57) 【要約】

【課題】階層構造をもつラベルが付与された教師データを用いてセキュリティインシデントの種別を分類する。

【解決手段】取得部15cが、分類の階層を表す粒度の異なるラベルが付与された教師データを取得し、変換部15dが、取得された第2の粒度のラベルが付与されたデータのラベルを、第1の粒度のラベルに変換する。データに付与する第1の粒度のラベルを出力する粗分類部15aは、取得された第1の粒度のラベルが付与された教師データと、変換された第1の粒度のラベルが付与された教師データとを用いて、第1の粒度のラベルの付与を学習する。データに付与する粒度が第1の粒度より細かい粒度のラベルを出力し、第1の粒度のラベルから第2の粒度のラベルを出力する細分類部15bは、取得された第2の粒度のラベルが付与された教師データと、粗分類部15aが出力した第1の粒度のラベルとを用いて、第2の粒度のラベルの付与を学習する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

データに付与する分類の階層を表す第 1 の粒度のラベルを出力する粗分類部と、

データに付与する粒度が前記第 1 の粒度のラベルより細かい第 2 の粒度のラベルを出力するとともに、前記第 1 の粒度のラベルから前記第 2 の粒度のラベルを出力する細分類部と、

前記第 1 の粒度のラベルが付与された教師データまたは前記第 2 の粒度のラベルが付与された教師データを取得する取得部と、

取得された前記第 2 の粒度のラベルが付与された教師データのラベルを、前記第 1 の粒度のラベルに変換する変換部と、を備え、

前記粗分類部は、取得された前記第 1 の粒度のラベルが付与された教師データと、変換された前記第 1 の粒度のラベルが付与された教師データとを用いて、前記第 1 の粒度のラベルの付与を学習し、

前記細分類部は、取得された前記第 2 の粒度のラベルが付与された教師データと、前記粗分類部が出力した前記第 1 の粒度のラベルとを用いて、前記第 2 の粒度のラベルの付与を学習することを特徴とする分類装置。

【請求項 2】

さらに、データに付与する、前記第 1 の粒度よりさらに粗い粒度のラベルを出力する第 2 の粗分類部を備え、該第 2 の粗分類部が前記粗分類部に連結されることを特徴とする請求項 1 に記載の分類装置。

【請求項 3】

分類装置で実行される分類方法であって、

データに付与する分類の階層を表す第 1 の粒度のラベルを出力する粗分類工程と、

データに付与する粒度が前記第 1 の粒度のラベルより細かい第 2 の粒度のラベルを出力するとともに、前記第 1 の粒度のラベルから前記第 2 の粒度のラベルを出力する細分類工程と、

前記第 1 の粒度のラベルが付与された教師データまたは前記第 2 の粒度のラベルが付与された教師データを取得する取得工程と、

取得された前記第 2 の粒度のラベルが付与された教師データのラベルを、前記第 1 の粒度のラベルに変換する変換工程と、を含み、

前記粗分類工程において、取得された前記第 1 の粒度のラベルが付与された教師データと、変換された前記第 1 の粒度のラベルが付与された教師データとを用いて、前記第 1 の粒度のラベルの付与を学習し、

前記細分類工程において、取得された前記第 2 の粒度のラベルが付与された教師データと、前記粗分類工程において出力された前記第 1 の粒度のラベルとを用いて、前記第 2 の粒度のラベルの付与を学習することを特徴とする分類方法。

【請求項 4】

データに付与する分類の階層を表す第 1 の粒度のラベルを出力する粗分類ステップと、

データに付与する粒度が前記第 1 の粒度のラベルより細かい第 2 の粒度のラベルを出力するとともに、前記第 1 の粒度のラベルから前記第 2 の粒度のラベルを出力する細分類ステップと、

前記第 1 の粒度のラベルが付与された教師データまたは前記第 2 の粒度のラベルが付与された教師データを取得する取得ステップと、

取得された前記第 2 の粒度のラベルが付与された教師データのラベルを、前記第 1 の粒度のラベルに変換する変換ステップと、を含み、

前記粗分類ステップにおいて、取得された前記第 1 の粒度のラベルが付与された教師データと、変換された前記第 1 の粒度のラベルが付与された教師データとを用いて、前記第 1 の粒度のラベルの付与を学習し、

前記細分類ステップにおいて、取得された前記第 2 の粒度のラベルが付与された教師データと、前記粗分類ステップにおいて出力された前記第 1 の粒度のラベルとを用いて、前

10

20

30

40

50

記第2の粒度のラベルの付与を学習する処理をコンピュータに実行させることを特徴とする分類プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、分類装置、分類方法および分類プログラムに関する。

【背景技術】

【0002】

近年、データの質の向上、データ数の増加、計算リソースの増加、あるいは機械学習のアルゴリズムの進歩に伴って、深層学習を用いた分類問題を解く事例が増加している。深層学習は、教師データとして与えられたラベル情報に基づいて、自動的にデータの特徴を抽出しながら学習を行うことにより、高い分類性能を示すことが知られている。

10

【0003】

例えば、通信データ等に基づいて深層学習を用いて、発生したセキュリティインシデントの種別を自動的に分類することができる。その場合には、通信データ等に対してインシデントの種別を示すラベルが付与された教師データが必要となる。深層学習に用いられる教師データには、同じ階層で分類されている同じ粒度のラベルが付与されている必要がある。例えば、「DOS攻撃」と「XXXマルウェアによるDOS攻撃」とは粒度が異なるラベルであり、そのままでは深層学習に用いることができない。

【0004】

20

ここで、教師データは、人手によりラベルを付与したり、ウイルスチェックのサービスサイトであるVirusTotal等からセキュリティベンダが付与したラベルの情報を取得したりすることにより用意される。人手により付与されたラベルの粒度は、ラベル付与を実施した実施者によって異なる場合がある。また、セキュリティベンダにより付与されたラベルや、SHA-1ハッシュ値等がラベルとして用いられる場合には、ラベルが異なる粒度の階層構造をもつ場合がある。

【先行技術文献】

【非特許文献】

【0005】

【非特許文献1】Richard P.Lippmann、“Pattern Classification Using Neural Networks”、[online]、1989年11月、IEEE Communications Magazine、pp.47-64、[2017年5月11日検索]、インターネット<URL:<http://ieeexplore.ieee.org/abstract/document/41401/>>

30

【非特許文献2】Tom Auld、Andrew W.Moore、Stephen F.Gull、“Bayesian Neural Networks for Internet Traffic Classification”、[online]、2007年1月、IEEE TRANSACTIONS ON NEURAL NETWORKS、VOL.18、pp.223-239、[2017年5月11日検索]、インターネット<URL:<http://ieeexplore.ieee.org/abstract/document/4049810/>>

【発明の概要】

【発明が解決しようとする課題】

40

【0006】

しかしながら、従来の技術では、階層構造をもつラベルが付与された教師データを用いてセキュリティインシデントの種別を分類することは困難であった。例えば、深層学習を用いてセキュリティインシデントの種別を分類する場合に、大量のインシデントのデータを収集し、各データに対して粒度の揃った適切なラベルを付与することは困難である。一方、粗い粒度のラベルの教師データを用いる場合には、細かい粒度の詳細なラベルが付与された教師データは学習に使用できない。また、新たなセキュリティインシデントが発見された場合にも、適切な粒度のラベルが付与されなければ教師データとして追加することができない。

【0007】

50

本発明は、上記に鑑みてなされたものであって、階層構造をもつラベルが付与された教師データを用いてセキュリティインシデントの種別を分類することを目的とする。

【課題を解決するための手段】

【0008】

上述した課題を解決し、目的を達成するために、本発明に係る分類装置は、データに付与する分類の階層を表す第1の粒度のラベルを出力する粗分類部と、データに付与する粒度が前記第1の粒度のラベルより細かい第2の粒度のラベルを出力するとともに、前記第1の粒度のラベルから前記第2の粒度のラベルを出力する細分類部と、前記第1の粒度のラベルが付与された教師データまたは前記第2の粒度のラベルが付与された教師データを取得する取得部と、取得された前記第2の粒度のラベルが付与された教師データのラベルを、前記第1の粒度のラベルに変換する変換部と、を備え、前記粗分類部は、取得された前記第1の粒度のラベルが付与された教師データと、変換された前記第1の粒度のラベルが付与された教師データとを用いて、前記第1の粒度のラベルの付与を学習し、前記細分類部は、取得された前記第2の粒度のラベルが付与された教師データと、前記粗分類部が出力した前記第1の粒度のラベルとを用いて、前記第2の粒度のラベルの付与を学習することを特徴とする。

10

【発明の効果】

【0009】

本発明によれば、階層構造をもつラベルが付与された教師データを用いてセキュリティインシデントの種別を分類することができる。

20

【図面の簡単な説明】

【0010】

【図1】図1は、本発明の一実施形態に係る分類装置の概略構成を例示する模式図である。

【図2】図2は、ラベル情報のデータ構成を例示する図である。

【図3】図3は、分類装置の処理概要を説明するための説明図である。

【図4】図4は、ラベルの変換について説明するための説明図である。

【図5】図5は、他の実施形態に係る分類装置の構成を説明するための説明図である。

【図6】図6は、分類処理手順を示すフローチャートである。

【図7】図7は、分類プログラムを実行するコンピュータを例示する図である。

30

【発明を実施するための形態】

【0011】

以下、図面を参照して、本発明の一実施形態を詳細に説明する。なお、この実施形態により本発明が限定されるものではない。また、図面の記載において、同一部分には同一の符号を付して示している。

【0012】

[分類装置の構成]

図1は、分類装置の概略構成を例示する模式図である。図1に例示するように、分類装置10は、パソコン等の汎用コンピュータで実現され、入力部11、出力部12、通信制御部13、記憶部14、および制御部15を備える。

40

【0013】

入力部11は、キーボードやマウス等の入力デバイスを用いて実現され、操作者による入力操作に対応して、制御部15に対して処理開始などの各種指示情報を入力する。出力部12は、液晶ディスプレイなどの表示装置、プリンター等の印刷装置等によって実現される。

【0014】

通信制御部13は、NIC (Network Interface Card) 等で実現され、LAN (Local Area Network) やインターネットなどの電気通信回線を介したIoT機器2等の外部の装置と制御部15との通信を制御する。

【0015】

50

記憶部 14 は、R A M (Random Access Memory)、フラッシュメモリ (Flash Memory) 等の半導体メモリ素子、または、ハードディスク、光ディスク等の記憶装置によって実現され、後述する分類処理により特定される正常な状態のモデルのパラメータ等が記憶される。なお、記憶部 14 は、通信制御部 13 を介して制御部 15 と通信する構成でもよい。

【0016】

本実施形態において、記憶部 14 には、ラベル情報 14 a が格納される。ラベル情報 14 a は、分類装置 10 で分類されたデータに付与されるラベルと、各ラベルの分類の階層を表す粒度とを示す情報である。ラベル情報 14 a は、例えば、米国政府が推進しているサイバー攻撃対策において、検知に有効なサイバー攻撃を特徴付ける指標等を記述するための技術仕様である S T I X (Structured Threat Information eXpression) を参照して構築される。

10

【0017】

図 2 は、ラベル情報 14 a のデータ構成を例示する図である。図 2 には、マルウェアが行った通信のネットワークログを分類する場合の階層型のラベルが例示されている。図 2 に示す例では、D o S、ランサムウェア、およびトロイの木馬は、階層 1 の粒度の粗いラベルである。また、例えば D o S マルウェア 1 および D o S マルウェア 2 は、階層 1 の D o S より粒度の細かい階層 2 のラベルである。

【0018】

図 1 の説明に戻る。制御部 15 は、C P U (Central Processing Unit) 等を用いて実現され、メモリに記憶された処理プログラムを実行する。これにより、制御部 15 は、図 1 に例示するように、粗分類部 15 a、細分類部 15 b、取得部 15 c、および変換部 15 d として機能する。なお、これらの機能部は、それぞれ、あるいは一部が異なるハードウェアに実装されてもよい。

20

【0019】

粗分類部 15 a は、データに付与する分類の階層を表す第 1 の粒度のラベルを出力する。また、細分類部 15 b は、データに付与する粒度が第 1 の粒度のラベルより細かい第 2 の粒度のラベルを出力するとともに、第 1 の粒度のラベルから第 2 の細かい粒度のラベルを出力する。第 1 の粒度のラベルとは、例えば、図 2 に例示した階層 1 のラベルであり、第 2 の粒度のラベルとは、階層 2 のラベルである。以下、第 1 の粒度のラベルを粒度の粗いラベルとも記し、第 2 の粒度のラベルを粒度の細かいラベルまたは詳細ラベルとも記す。

30

【0020】

ここで、図 3 を参照して粗分類部 15 a および細分類部 15 b について説明する。図 3 は、分類装置 10 の処理概要を説明するための説明図である。図 3 に例示するように、本実施形態の分類装置 10 は、2 階層の階層型のラベルのうち、粒度の粗いラベルを出力する粗分類部 15 a と、粒度の細かいラベルを出力する細分類部 15 b とが組み合わせられて構成される。

【0021】

粗分類部 15 a は、単一の階層のラベルを取り扱う従来の分類装置と同様に構成され、データ x が入力されると、隠れ層 (hidden) を介して、データ x に付与する「D o S」等の粒度の粗いラベル y を出力する。細分類部 15 b は、hidden を介して、入力されたデータ x に付与する「D o S マルウェア 1」等の粒度の細かい詳細ラベル z を出力する。また、細分類部 15 b は、hidden を介して、入力された粒度の粗いラベル y から粒度の細かい詳細ラベル z を出力する。

40

【0022】

粗分類部 15 a および細分類部 15 b は、後述するように、階層型の粒度の異なるラベルが付与されたデータを教師データとして深層学習を行う。これにより、分類装置 10 は入力されたデータ x に付与する詳細ラベル z を高精度に出力することが可能となる。このように、深層学習で最適化される誤差関数を拡張することにより、分類装置 10 の分類性

50

能を向上させることが可能となる。

【0023】

なお、粗分類部15aおよび細分類部15bの深層学習の方式は限定されず、例えば、変分オートエンコーダ (Variational auto encoder) 等を用いた高度な半教師あり学習が適用されてもよい。

【0024】

図1の説明に戻る。取得部15cは、第1の粒度のラベルが付与された教師データまたは第2の粒度のラベルが付与された教師データを取得する。例えば、取得部15cは、入力部11あるいは通信制御部13を介して、階層型のラベルが付与されたデータを教師データとして取得する。

10

【0025】

変換部15dは、取得された第2の粒度のラベルが付与された教師データのラベルを、第1の粒度のラベルに変換する。具体的には、変換部15dは、階層型のラベルを階層の数の要素をもつベクトル形式に変換する。

【0026】

図4は、ラベルの変換について説明するための説明図である。図4に例示するように、変換部15dは、ラベル情報14aを参照し、例えば人手により付与された粒度のそろっていない2階層の階層型のラベルを、2つの要素をもつベクトルに変換する。例えば、変換部15dは、「DoSマルウェア1」を(DoS, DoSマルウェア1)に変換する。

20

【0027】

また、粗いラベルのみが付与されていて詳細ラベルが不明である場合に、変換部15dは、(DoS, *)あるいは(DoS, null)のように、不明な詳細ラベルの値を*またはnull等とする。図4に示す例では、例えば、階層1の粒度の粗いラベルである「トロイの木馬」が(トロイの木馬, *)というベクトルに変換されている。これにより、変換部15dは、取得部15cが取得した教師データのラベルが階層2の詳細ラベルであった場合に、ラベル情報14aを参照し、階層1の粗いラベルに変換する。

【0028】

なお、マルウェアが行った通信のネットワークログを、SHA-1ハッシュ値等で表されるマルウェア検体IDをラベルとして用いて分類する場合には、変換部15dは、マルウェア検体IDを元にして、最適な階層の階層型ラベルに変換する。その際、変換部15dは、必要以上に詳細な階層の値、あるいは信頼度の低い階層の値を*またはnull等とする。

30

【0029】

そして、粗分類部15aは、取得された第1の粒度のラベルが付与された教師データと、変換された第1の粒度のラベルが付与された教師データとを用いて、第1の粒度のラベルの付与を学習する。また、細分類部15bは、取得された第2の粒度の詳細ラベルが付与された教師データと、粗分類部15aが出力した第1の粒度のラベルとを用いて、第2の粒度の詳細ラベルの付与を学習する。

【0030】

言い換えれば、粒度の粗いラベルが付与された教師データが入力された場合に、粗分類部15aが粒度の粗いラベルの付与を学習する。また、粒度の細かい詳細ラベルが付与された教師データが入力された場合に、変換部15dが詳細ラベルを粒度の粗いラベルに変換することにより、粗分類部15aが粒度の粗いラベルの付与を学習する。また、細分類部15bが、粒度の細かい詳細ラベルが付与された教師データと、粗分類部15aが出力した粒度の粗いラベルとを用いて、詳細ラベルの付与を学習する。このように、分類装置10は、粒度の粗いラベルが付与された教師データおよび粒度の細かい詳細ラベルが付与された教師データのすべての教師データを用いて、詳細ラベルの付与を学習することが可能となる。

40

【0031】

なお、分類装置10が取り扱うラベルの階層は3以上でもよい。ここで、図5を参照し

50

て、その場合の分類装置の構成を説明する。その場合に、分類装置 10 は、さらに、データに付与する、第 1 の粒度よりさらに粗い粒度のラベルを出力する粗分類部 15 を備え、該粗分類部 15 が粗分類部 15 a に連結される

【0032】

図 5 には、分類装置 10 が 3 階層のラベルを取り扱う場合について例示されている。図 5 に示すように、階層 1 の粒度の粗いラベルを出力する粗分類部 15 a に、粗分類部 15 a と同様に構成され階層 1 よりさらに粒度の粗いラベルを出力する粗分類部 15 が連結されている。これにより、分類装置 10 が、詳細ラベル、粗いラベル、および、さらに粗いラベルの 3 階層のラベルを取り扱うことが可能となる。

【0033】

分類装置 10 が 4 階層のラベルを取り扱う場合には、粗分類部 15 a および粗分類部 15 と同様に構成され粒度が最も粗いラベルを出力する新たな粗分類部が、粗分類部 15 に連結される。このように、分類装置 10 が取り扱うラベルの階層の数が増えた場合には、粗分類部 15 a と同様に構成され粒度が最も粗いラベルを出力する新たな粗分類部を分類装置 10 に連結することにより実現可能となる。これにより、分類装置 10 は、多様な教師データを用いて最も粒度の細かい詳細ラベルの付与を学習することが可能となる。

【0034】

[分類処理]

次に、図 6 を参照して、本実施形態に係る分類装置 10 による分類処理について説明する。図 6 は、分類処理手順を示すフローチャートである。図 6 のフローチャートは、例えば、分類処理の開始を指示する操作入力があったタイミングで開始される。

【0035】

まず、取得部 15 c が、入力部 11 あるいは通信制御部 13 を介して、粒度の粗いラベルが付与された教師データまたは粒度の細かいラベルが付与された教師データを取得する（ステップ S1）。

【0036】

次に、変換部 15 d が、教師データに付与されている階層型のラベルを階層の数の要素をもつベクトル形式に変換することにより、取得された教師データに詳細ラベルが付与されている場合には、詳細ラベルを粒度の粗いラベルに変換する（ステップ S2）。

【0037】

取得部 15 c が、粒度の粗いラベルが付与された教師データを取得した場合に（ステップ S3、Yes）、粗分類部 15 a が粗いラベルの付与を学習する（ステップ S4）。また、取得部 15 c が、詳細ラベルが付与された教師データを取得した場合に（ステップ S3、No）、粗分類部 15 a が粗いラベルの付与を学習し、細分類部 15 b が詳細ラベルの付与を学習する（ステップ S5）。これにより、一連の分類処理が終了する。

【0038】

以上、説明したように、本実施形態の分類装置 10 において、粗分類部 15 a が、データに付与する分類の階層を表す第 1 の粒度の粗いラベルを出力する。また、細分類部 15 b が、データに付与する第 1 の粒度より細かい第 2 の粒度の詳細ラベルを出力するとともに、第 1 の粒度の粗いラベルから第 2 の粒度の詳細ラベルを出力する。また、取得部 15 c が、第 1 の粒度の粗いラベルが付与された教師データまたは第 2 の粒度の詳細ラベルが付与された教師データを取得する。

【0039】

また、変換部 15 d が、取得された第 2 の粒度の詳細ラベルが付与された教師データのラベルを、第 1 の粒度の粗いラベルに変換する。粗分類部 15 a は、取得された第 1 の粒度の粗いラベルが付与された教師データと、変換された第 1 の粒度の粗いラベルが付与された教師データとを用いて、第 1 の粒度の粗いラベルの付与を学習する。また、細分類部 15 b は、取得された第 2 の粒度の詳細ラベルが付与された教師データと、粗分類部 15 a が出力した第 1 の粒度の粗いラベルとを用いて、第 2 の粒度の詳細ラベルの付与を学習する。

10

20

30

40

50

【0040】

この場合に、粗分類部15aは、大量の教師データを用いて学習するので、精度が向上する。また、細分類部15bは、分類対象のデータに加え、粒度の粗いラベルを用いて分類するので、精度が向上する。また、粒度の粗いラベルが付与された教師データが無駄にされることなく詳細ラベルの付与の学習に利用される。

【0041】

このように、分類装置10は、階層型の粒度の異なるラベルが付与されたすべての教師データを用いて、詳細ラベルの付与を学習することが可能となる。これにより、ラベルの粒度に関わらず、幅広くより多くのラベルの情報を教師データとして学習できるので、分類性能が向上する。

10

【0042】

また、セキュリティインシデントの原因を絞り込みながら、その時点でわかった範囲で抽象的なラベルが付与された教師データを用いて学習することが可能となる。同一のマルウェアに対して、セキュリティベンダごとに異なるラベルが付与されていても、異なるラベルの教師データを用いて学習することができる。このように、本実施形態の分類装置10の分類処理により、階層構造をもつラベルが付与された教師データを用いてセキュリティインシデントの種別を分類することができる。

【0043】

さらに、データに付与するさらに粒度の粗いラベルを出力する粗分類部15を備え、該粗分類部15が粗分類部15aに連結されてもよい。これにより、分類装置10が3階層のラベルを取り扱うことが可能となる。このように、分類装置10は、多数階層構造の多様なラベルが付与された教師データを用いて、最も粒度の細かい詳細ラベルの付与を学習することが可能となる。

20

【0044】

[実施例]

【0045】

次に、本実施形態の分類装置10と、単一の階層のラベルを取り扱う従来の分類装置について、以下の条件で実験を行って、分類精度を測定した。ここで、従来の分類装置は、hiddenを介して入力されたデータに付与する粒度の細かい詳細ラベルを出力する。入力されるデータは、手書き数字の画像データとし、詳細ラベルは、1, 2, ...の10種の数字のラベルとした。また、本実施形態の分類装置10で取り扱われる粒度の粗いラベルは、偶数または奇数の2種類のラベルとした。

30

【0046】

本実施形態の分類装置10および従来の分類装置のそれぞれに、詳細ラベルが付与された教師データ300件を入力し、粗いラベルが付与された教師データ9700件を本実施形態の分類装置10に入力し、詳細ラベルの付与の学習を行った。その後、本実施形態の分類装置10および従来の分類装置のそれぞれに、学習されていない未知のデータ10000件を入力した。その結果、従来の分類装置の分類精度は82%であった。これに対し、本実施形態の分類装置10の分類精度は87%であり、約5%の分類精度の向上が確認された。

40

【0047】

[プログラム]

上記実施形態に係る分類装置10が実行する処理をコンピュータが実行可能な言語で記述したプログラムを作成することもできる。一実施形態として、分類装置10は、パッケージソフトウェアやオンラインソフトウェアとして上記の分類処理を実行する分類プログラムを所望のコンピュータにインストールさせることにより実装できる。例えば、上記の分類プログラムを情報処理装置に実行させることにより、情報処理装置を分類装置10として機能させることができる。ここで言う情報処理装置には、デスクトップ型またはノート型のパーソナルコンピュータが含まれる。また、その他にも、情報処理装置にはスマートフォン、携帯電話機やPHS(Personal Handyphone System)などの移動体通信端

50

未、さらには、PDA (Personal Digital Assistants) などのスレート端末などがその範疇に含まれる。

【0048】

また、分類装置10は、ユーザが使用する端末装置をクライアントとし、当該クライアントに上記の分類処理に関するサービスを提供するサーバ装置として実装することもできる。例えば、分類装置10は、教師データおよび未知のデータを入力とし、未知のデータに付与するラベルを出力する分類処理サービスを提供するサーバ装置として実装される。この場合、分類装置10は、Webサーバとして実装することとしてもよいし、アウトソーシングによって上記の分類処理に関するサービスを提供するクラウドとして実装することとしてもかまわない。以下に、分類装置10と同様の機能を実現する分類プログラムを実行するコンピュータの一例を説明する。

10

【0049】

図7は、分類プログラムを実行するコンピュータの一例を示す図である。コンピュータ1000は、例えば、メモリ1010と、CPU1020と、ハードディスクドライブインタフェース1030と、ディスクドライブインタフェース1040と、シリアルポートインタフェース1050と、ビデオアダプタ1060と、ネットワークインタフェース1070とを有する。これらの各部は、バス1080によって接続される。

【0050】

メモリ1010は、ROM (Read Only Memory) 1011およびRAM 1012を含む。ROM 1011は、例えば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライブインタフェース1030は、ハードディスクドライブ1031に接続される。ディスクドライブインタフェース1040は、ディスクドライブ1041に接続される。ディスクドライブ1041には、例えば、磁気ディスクや光ディスク等の着脱可能な記憶媒体が挿入される。シリアルポートインタフェース1050には、例えば、マウス1051およびキーボード1052が接続される。ビデオアダプタ1060には、例えば、ディスプレイ1061が接続される。

20

【0051】

ここで、ハードディスクドライブ1031は、例えば、OS 1091、アプリケーションプログラム1092、プログラムモジュール1093およびプログラムデータ1094を記憶する。上記実施形態で説明した各テーブルは、例えばハードディスクドライブ1031やメモリ1010に記憶される。

30

【0052】

また、分類プログラムは、例えば、コンピュータ1000によって実行される指令が記述されたプログラムモジュール1093として、ハードディスクドライブ1031に記憶される。具体的には、上記実施形態で説明した分類装置10が実行する各処理が記述されたプログラムモジュール1093が、ハードディスクドライブ1031に記憶される。

【0053】

また、分類プログラムによる情報処理に用いられるデータは、プログラムデータ1094として、例えば、ハードディスクドライブ1031に記憶される。そして、CPU 1020が、ハードディスクドライブ1031に記憶されたプログラムモジュール1093やプログラムデータ1094を必要に応じてRAM 1012に読み出して、上述した各手順を実行する。

40

【0054】

なお、分類プログラムに係るプログラムモジュール1093やプログラムデータ1094は、ハードディスクドライブ1031に記憶される場合に限られず、例えば、着脱可能な記憶媒体に記憶されて、ディスクドライブ1041等を介してCPU 1020によって読み出されてもよい。あるいは、分類プログラムに係るプログラムモジュール1093やプログラムデータ1094は、LANやWAN (Wide Area Network) 等のネットワークを介して接続された他のコンピュータに記憶され、ネットワークインタフェース1070を介してCPU 1020によって読み出されてもよい。

50

【0055】

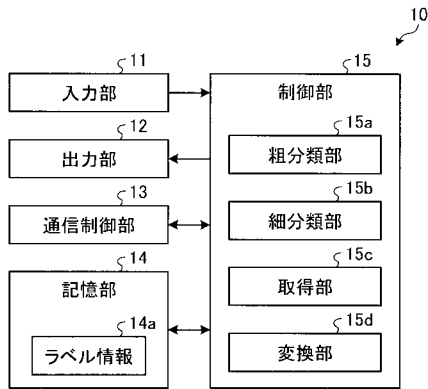
以上、本発明者によってなされた発明を適用した実施形態について説明したが、本実施形態による本発明の開示の一部をなす記述および図面により本発明は限定されることはない。すなわち、本実施形態に基づいて当業者等によりなされる他の実施形態、実施例および運用技術等は全て本発明の範疇に含まれる。

【符号の説明】

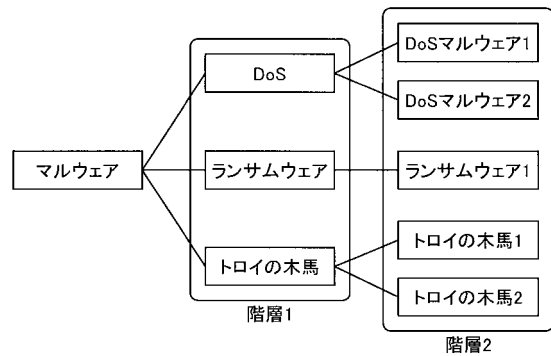
【0056】

- 10 分類装置
- 11 入力部
- 12 出力部
- 13 通信制御部
- 14 記憶部
- 14a ラベル情報
- 15 制御部
- 15a 粗分類部
- 15b 細分類部
- 15c 取得部
- 15d 変換部

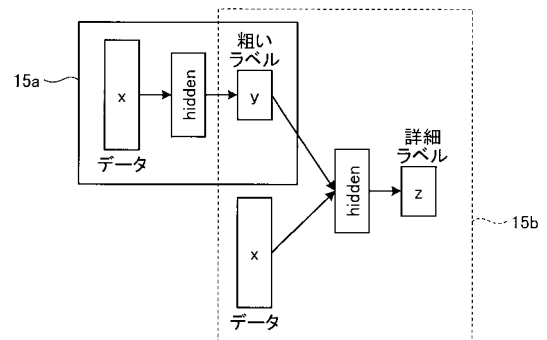
【図1】



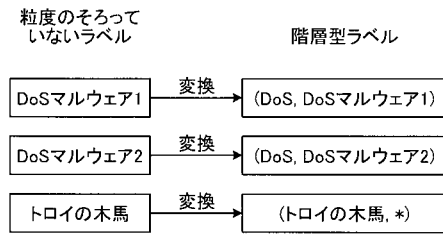
【図2】



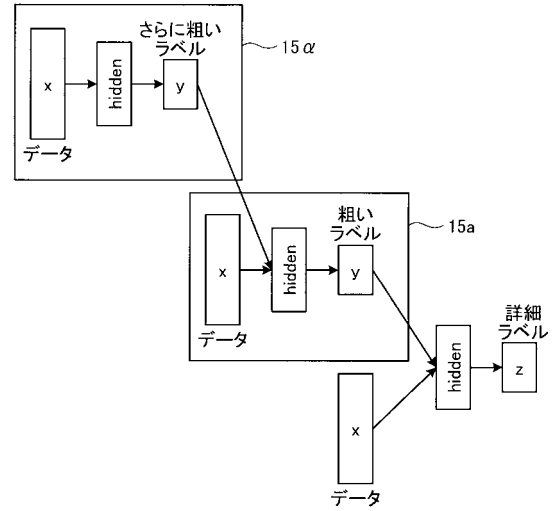
【図3】



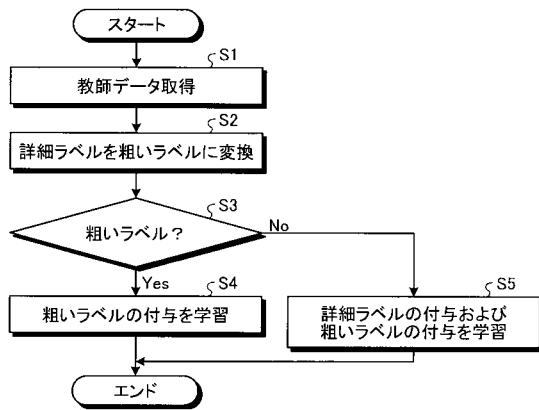
【 図 4 】



【 図 5 】



【 図 6 】



【 図 7 】

