

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号  
特許第6900190号  
(P6900190)

(45) 発行日 令和3年7月7日 (2021.7.7)

(24) 登録日 令和3年6月18日 (2021.6.18)

(51) Int.Cl.  
G 0 6 F 1 6 / 9 0 3 (2019.01)

F I  
G O 6 F 1 6 / 9 0 3

請求項の数 15 (全 30 頁)

(21) 出願番号	特願2016-256060 (P2016-256060)	(73) 特許権者	000001007 キヤノン株式会社 東京都大田区下丸子3丁目30番2号
(22) 出願日	平成28年12月28日 (2016.12.28)	(74) 代理人	100126240 弁理士 阿部 琢磨
(65) 公開番号	特開2017-130196 (P2017-130196A)	(74) 代理人	100124442 弁理士 黒岩 創吾
(43) 公開日	平成29年7月27日 (2017.7.27)	(72) 発明者	八谷 大岳 東京都大田区下丸子3丁目30番2号キヤ ノン株式会社内
審査請求日	令和1年12月23日 (2019.12.23)	(72) 発明者	真継 優和 東京都大田区下丸子3丁目30番2号キヤ ノン株式会社内
(31) 優先権主張番号	特願2016-5598 (P2016-5598)	審査官	鹿野 博嗣
(32) 優先日	平成28年1月14日 (2016.1.14)		最終頁に続く
(33) 優先権主張国・地域又は機関	日本国 (JP)		

(54) 【発明の名称】 認識学習装置、認識学習方法及びプログラム

(57) 【特許請求の範囲】

【請求項 1】

特定ドメインの概念構造を表す概念構造情報であって、認識対象の候補を概念情報として含む概念構造情報に基づいて、前記特定ドメインと前記認識対象の候補との関連度を生成する生成手段と、

前記生成手段により生成された関連度に基づいて、前記認識対象の候補から認識対象を選択する選択手段と、

前記選択手段により選択された認識対象に係る学習データを用いて認識器を学習する学習手段と、

を有することを特徴とする認識学習装置。

10

【請求項 2】

前記生成手段は、前記概念構造情報における前記認識対象の候補の階層に基づいて前記関連度を生成することを特徴とする請求項 1 に記載の認識学習装置。

【請求項 3】

前記生成手段は、更に前記概念構造情報における前記認識対象の候補の発生頻度に基づいて前記関連度を生成することを特徴とする請求項 2 に記載の認識学習装置。

【請求項 4】

前記生成手段は、前記概念構造情報における前記認識対象の候補より下位階層の概念情報の数に基づいて前記関連度を生成することを特徴とする請求項 1 に記載の認識学習装置。

20

## 【請求項 5】

前記選択手段により選択された前記認識対象を前記概念構造情報に重畳した可視化情報を生成する可視化手段を更に有することを特徴とする請求項 1 から 4 のいずれか 1 項に記載の認識学習装置。

## 【請求項 6】

前記可視化手段は、前記選択手段により選択された認識対象それぞれに対する前記認識器の認識精度を算出し、当該算出した認識精度を前記可視化情報として生成することを特徴とする請求項 5 に記載の認識学習装置。

## 【請求項 7】

前記可視化手段により生成された可視化情報に対するユーザの操作に応じて、前記認識対象を更新する更新手段を更に有し、

前記学習手段は、前記更新手段により更新された認識対象に係る学習データを用いて前記認識器を再度学習することを特徴とする請求項 5 または 6 に記載の認識学習装置。

## 【請求項 8】

前記可視化手段は、前記認識対象それぞれの動画像データを前記可視化情報として生成し、

前記可視化手段により生成された可視化情報に対するユーザの指示に応じて、前記認識対象に係る学習データに追加または削除を行う編集手段を更に有することを特徴とする請求項 5 から 7 のいずれか 1 項に記載の認識学習装置。

## 【請求項 9】

前記学習手段は、前記関連度に基づいて選択された前記認識対象の重要度を示す重要度情報を生成し、当該生成した前記重要度情報に基づいて学習することを特徴とする請求項 1 から請求項 8 のいずれか 1 項に記載の認識学習装置。

## 【請求項 10】

文章データから前記特定ドメインに関する前記概念構造情報を生成する概念構造生成手段を更に有し、

前記生成手段は、前記概念構造生成手段により生成された概念構造情報に基づいて、前記特定ドメインと前記認識対象の候補との前記関連度を生成することを特徴とする請求項 1 から請求項 9 のいずれか 1 項に記載の認識学習装置。

## 【請求項 11】

ユーザの入力に応じて、前記特定ドメイン、特定の利用者、特定のユースケースの少なくとも 1 つに対して構築された前記概念構造情報を選択する概念情報選択手段を更に有することを特徴とする請求項 1 から 10 のいずれか 1 項に記載の認識学習装置。

## 【請求項 12】

前記概念構造情報として、オントロジー情報を用いることを特徴とする請求項 1 から 11 のいずれか 1 項に記載の認識学習装置。

## 【請求項 13】

認識学習装置により実行される認識学習方法であって、

特定ドメインの概念構造を表す概念構造情報であって、認識対象の候補を概念情報として含む概念構造情報に基づいて、前記特定ドメインと前記認識対象の候補との関連度を生成するステップと、

前記生成された関連度に基づいて、前記認識対象の候補から認識対象を選択するステップと、

前記選択された認識対象に係る学習データを用いて認識器を学習するステップと、を有することを特徴とする認識学習方法。

## 【請求項 14】

コンピュータを、請求項 1 から 12 のいずれか 1 項に記載の認識学習装置として機能させるためのプログラム。

## 【請求項 15】

オントロジー情報である前記概念情報を記憶する第 1 の記憶手段と、

10

20

30

40

50

前記認識対象を特定するための動画データを記憶する第２の記憶手段とを更に有することを特徴とする請求項１から１２のいずれか１項に記載の認識学習装置。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、データから認識対象を認識する認識器を学習する技術に関する。

【背景技術】

【０００２】

近年、監視カメラが撮影した動画像データから、人や群衆の活動パターンを分析したり、特定の事象を検出し通報するサービスがある。該サービスを実現するためには、監視カメラが撮影した動画像データから、人か車かなどの物体の属性や、歩いているか走っているかなどの行動の種類、鞆かカゴかなどの人の所持品の種類を検出可能な機械学習の認識技術が不可欠である。該サービスは、介護施設、一般家庭、駅や市街地などの公共施設、スーパー、コンビニなどの店舗など様々な環境において活用される。また、同じ環境においても、利用者の該サービスに対するニーズは多様である。そのため、多様な環境およびユースケースに対応可能な、柔軟で高精度な機械学習の認識技術が必要とされている。

【０００３】

非特許文献１には、柔軟で高精度な機械学習の認識を実現するための技術が提案されている。非特許文献１の技術では、先ず、ImageNetなどの大規模な教師ありデータを用いて１０００カテゴリに対応可能な汎用的なConvolutional Neural Network（以下CNNと省略）を事前に学習する。そして、その学習の後、ユーザの特定のニーズに合わせて、カテゴリ数を限定して詳細に学習するようにしている。この事前の学習はプレトレーニング、詳細の学習はファインチューニングと呼ばれる。膨大なパラメータ数を要するCNNをプレトレーニングしておくことにより、ファインチューニングでは比較的短時間で、特定のニーズに合わせて高精度な認識器を獲得することができるという利点がある。また、プレトレーニングで大規模なデータを用いることにより、膨大な数のパラメータが特定の認識対象にオーバーフィットする問題を緩和することができるという期待されている。

【０００４】

また、特許文献１では、楽曲に対する人間の感性によって判断される印象の予測において、プレトレーニングした複数の階層型ニューラルネットワークから、いずれかを選択し、入力された印象度に用いてファインチューニングする方法が提案されている。

【先行技術文献】

【特許文献】

【０００５】

【特許文献１】特開２００６－３１６３７号公報

【非特許文献】

【０００６】

【非特許文献１】Rich feature hierarchies for accurate object detection and semantic segmentation, Ross Girshick, Jeff Donahue, Trevor Darrel, Jitendra Malik, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014

【非特許文献２】オントロジー構築ツールの現状、吉崎晃司、溝口理一郎、人工知能学会誌、２０（６）、７０７－７１４、２００５－１１－０１

【非特許文献３】Rich feature hierarchies for accurate object detection and semantic segmentation, Ross Girshick, Jeff Donahue, Trevor Darrel, Jitendra Malik, IEEE Conference

on Computer Vision and Pattern Recognition (CVPR), 2014

【非特許文献4】3D Convolutional Neural Networks for Human Action Recognition, S. Ji, W. Xu, M. Yang and K. Yu, Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221-231, 2012

【非特許文献5】Two-stream convolutional networks for action recognition in videos, K. Simonyan and A. Zisserman, Advances in Neural Information Processing Systems 25 (NIPS), 2014.

10

【非特許文献6】ImageNet Classification with Deep Convolutional Neural Networks, Krizhevsky, A., Sutskever, I. and Hinton, G. E. Neural Information Processing Systems (NIPS), 2012

【発明の概要】

【発明が解決しようとする課題】

【0007】

20

しかしながら、特許文献1に記載の方法では、プレトレーニングとファインチューニングで共通の階層型ニューラルネットワークの構造を用いている。そのため、利用者のニーズに合わせて認識対象を柔軟に変えることが困難である。

【0008】

一方、非特許文献1の技術によれば、CNNの出力数を変えることが出来るため、プレトレーニングとファインチューニングとで認識対象を柔軟に変えることができる。しかしながら、プレトレーニングの認識対象であるImageNetの1000カテゴリが、将来CNNを利用する利用者のニーズをカバーしているとは限らない。もし、プレトレーニングで該ニーズをカバーしていない場合は、ファインチューニングに膨大な数のパラメータを再度学習する必要が発生し、プレトレーニングによる学習時間の短縮化とオーバーフィットの回避の恩恵が受けられない。この問題を回避するために、さらにカテゴリ数を増やして、あらゆる認識対象に対してプレトレーニングを行うことも可能であるが、無数の認識対象を識別するためには、さらに膨大な数のパラメータが必要となる。しかしながら、最終的に利用者が必要とする認識対象は小規模の場合もあるので、多くの場合に不必要に複雑なCNNを学習してしまうという問題がある。一方、無数の認識対象の中から利用者のニーズを考慮して、プレトレーニングに用いる認識対象を手で選定するのは大変な労力となる。

30

【0009】

そこで、本発明は、上記問題を解決すべくなされたもので、利用者のニーズを考慮した認識器のプレトレーニングやファインチューニングなどの学習を可能にすることを目的とする。

40

【課題を解決するための手段】

【0010】

上記課題を解決するために、本発明の認識学習装置は、特定ドメインの概念構造を表す概念構造情報であって、認識対象の候補を概念情報として含む概念構造情報に基づいて、前記特定ドメインと前記認識対象の候補との関連度を生成する生成手段と、前記生成手段により生成された関連度に基づいて、前記認識対象の候補から認識対象を選択する選択手段と、前記選択手段により選択された認識対象に係る学習データを用いて認識器を学習する学習手段と、を有することを特徴とする。

【発明の効果】

50

## 【 0 0 1 1 】

以上の構成によれば、本発明では、利用者のニーズを考慮した認識器のプレトレーニングやファインチューニングなどの学習が可能になる。

## 【図面の簡単な説明】

## 【 0 0 1 2 】

【図 1】第 1 の実施形態に関わる認識学習システムの構成の一例を示す概略ブロック図。

【図 2】第 1 の実施形態においてオントロジー情報の一例を示す図。

【図 3】第 1 の実施形態において概念構造記憶部が記憶する情報の一例を示す図。

【図 4】第 1 の実施形態において動画像データ記憶部が記憶する情報の一例を示す図。

【図 5】第 1 の実施形態において認識器記憶部が記憶する情報の一例を示す図。

【図 6】第 1 の実施形態において認識対象可視化情報の一例を示す図。

【図 7】第 1 の実施形態において認識器のプレトレーニングの一例を示すフローチャート

。

【図 8】第 2 の実施形態に関わる認識学習システムの構成の一例を示す概略ブロック図。

【図 9】第 2 の実施形態において表示部に表示される認識対象可視化情報の一例を示す図

。

【図 10】第 2 の実施形態において認識器のプレトレーニングの一例を示すフローチャート。

【図 11】第 3 の実施形態に関わる認識学習システムの構成の一例を示す概略ブロック図

。

【図 12】第 3 の実施形態において表示部に表示される認識対象可視化情報の一例を示す図。

【図 13】第 3 の実施形態において端末装置による動画像データの追加の一例を示す図。

【図 14】第 4 の実施形態に関わる認識学習システムの構成の一例を示す概略ブロック図

。

【図 15】その他の実施形態に関わる認識学習システムの構成の一例を示す概略ブロック図。

【図 16】その他の実施形態においてオントロジー情報の選択メニューの一例を示す図。

【図 17】第 1 の実施形態において意味的ネットワークの一例を示す図。

【図 18】第 5 の実施形態に関わる認識学習システムの構成の一例を示す概略ブロック図

。

【図 19】第 5 の実施形態に関わる端末装置の表示形態の一例を示す図。

## 【発明を実施するための形態】

## 【 0 0 1 3 】

## [ 第 1 の実施形態 ]

以下、本発明の第 1 の実施形態の詳細について図面を参照しつつ説明する。本実施形態における認識学習システム 1 について、認識器の提供者が、利用者の特定のドメインに合わせて、認識器をプレトレーニングする場合について説明する。具体的には、認識学習システム 1 は、特定のドメインと認識対象の候補である概念情報との関係を表す概念情報に基づき、該概念情報の中から認識対象を選定してプレトレーニングを施す。そして、認識学習システム 1 は、プレトレーニングされた認識器の認識対象の範囲をオントロジーに基づいて可視化し、提供者に提示する。ここで、特定のドメインに対するオントロジー上の概念情報の集合が、プレトレーニングの認識対象の候補である。この概念情報とは、概念化および言語化可能な物体の状態であり、該状態を言語的に示すラベル情報により特徴付けられる。概念情報には、例えば、「人」、「車」などの物体の属性や、「歩いている」、「走っている」などの物体の行動や、「鞆」、「カゴ」などの人の所持品などが含まれる。また、概念構造情報には、例えば、図 17 で後述する意味ネットワークなどがある。

## 【 0 0 1 4 】

特定のドメインは、本システムが利用されると想定される環境であり、例えば、介護施設、一般家庭、公共施設の駅や市街、店舗などである。利用者とは、例えば、客や店員な

10

20

30

40

50

どの活動パターンの分析や緊急通報などの用途で、本システムを監視カメラとともに直接利用するエンドユーザ、または、本システムを第三者に提供するために認識器の調整を行うシステムインテグレータである。また、提供者は、本システムを開発し利用者に提供する研究開発者、または上述したシステムインテグレータである。

#### 【 0 0 1 5 】

図 1 は、本実施形態に係る認識学習装置を用いた認識学習システムの構成の一例を示す概略ブロック図である。認識学習システム 1 は、認識学習装置 1 0 と、端末装置 1 0 0 とを備えている。なお、これらの装置間は、ネットワークを介して接続されていてもよい。このネットワークには、例えば、固定電話回線網や、携帯電話回線網や、インターネットが適用できる。

10

#### 【 0 0 1 6 】

認識学習装置 1 0 は、ハードウェア構成として、不図示の表示部 D S と操作検出部 O P とを備えている。表示部 D S は、液晶パネルや有機 E L パネルなどの画像表示パネルを備えており、認識学習装置 1 0 から入力された情報を表示する。なお、表示部 D S は、後述する「介護施設」、「一般家庭」、「駅」、「市街」、および「店舗」などのドメイン名情報の一覧を表示する。また、表示部 D S は、認識学習装置 1 0 の認識対象可視化部 1 4 の説明にて後述する認識対象の範囲を示す認識対象可視化情報を表示する。

#### 【 0 0 1 7 】

操作検出部 O P は、表示部 D S の画像表示パネルに配置されたタッチセンサを備えており、ユーザの指やタッチペンの動きに基づくユーザの操作を検出するとともに、検出した操作を示す操作情報を認識学習装置 1 0 に出力する。なお、操作検出部 O P は、コントローラ、キーボード及びマウスなどの入力デバイスを備え、画像表示パネルに表示された画像に対するユーザの操作を示す操作情報を取得してもよい。この操作情報には、例えば、ドメイン名情報の候補の中から特定のドメイン名情報の選択や、認識器のプレトレーニングの実行を示す「プレトレーニングの実行」などがある。操作検出部 O P は、操作情報を検出した場合、自装置が記憶する選択されたドメイン名を識別するドメイン I D と検出した操作情報とを、認識学習装置 1 0 に出力する。

20

#### 【 0 0 1 8 】

次に、認識学習装置 1 0 のソフトウェア構成に関し、その詳細な構成について説明する。認識学習装置 1 0 は、特定のドメインに合わせて認識器をプレトレーニングする装置である。認識学習装置 1 0 は、概念構造記憶部 M 1 と、動画像データ記憶部 M 2 と、認識器記憶部 M 3 と、意味的関連度生成部 1 1 と、認識対象選定部 1 2 と、認識学習部 1 3 と、認識対象可視化部 1 4 と、を備えている。

30

#### 【 0 0 1 9 】

概念構造記憶部 M 1 は、ドメインを識別するドメイン I D と、ドメインを言語的に示すドメイン名情報と、該ドメインの概念構造を表す概念構造情報とを、ドメイン I D と関連付けて記憶する。ここで、ドメイン I D とは、提供者により予め設定された本システムが利用されるドメインを識別する情報である。ドメイン名情報は、該ドメインを言語的に説明するドメインの名称を示す情報である。ドメイン名情報は、例えば、「介護施設」、「一般家庭」、「駅」、「市街」、「店舗」である。概念構造情報とは、予め提供者によりドメインごとに定義される、ドメイン名情報の概念構造を示す情報である。例えば、提供者がオントロジーの編集ソフト（非特許文献 2）を用いて、該ドメインを分析し、ドメインを構成する概念集合の木構造を記述する。ここで、木構造上の上位と下位との概念間には、上位下位関係を表す *i s - a* 関係および部分全体関係を表す *h a s - a* 関係などが用いられる。この概念間の *i s - a* 関係および *h a s - a* 関係としては、数万種類の概念が登録されている *W o r d N e t* などの辞書を活用することができる。また、概念構造情報の各概念を示す概念情報は、概念を識別する認識対象 I D と、上位の概念を識別する上位 I D と、概念を言語的に示す概念名情報を含んでいる。なお、この認識対象 I D は、動画像データ記憶部 M 2 にて後述する認識対象 I D と共通である。

40

#### 【 0 0 2 0 】

50

図 2 は、概念構造記憶部 M 1 が記憶する概念構造の 1 つであるオントロジー情報の一例を示す図である。同図では、ドメイン名情報「店舗」20 と複数の概念情報 21 ~ 29 が木構造で接続されていることが示されている。具体的には、図 2 のオントロジー情報では、ルートのドメイン名情報「店舗」を構成する上位概念として、概念情報「人」21、概念情報「機器」22、概念情報「家具」23 などが記述されている。ここで、ドメイン名情報「店舗」20 と概念情報 21 ~ 23 の間には、has - a 関係が用いられている。つまり、「店舗」20 は「人」21、「機器」22 及び「家具」23 を含んで構成される。また、各概念情報には、概念名情報の他に、該概念を識別する認識対象 ID と、上位の概念を識別する上位 ID とが記載されている。そして、それぞれの概念情報に上位と下位の概念情報が数珠状に連結されている。例えば、概念情報「人」21 は、下位の概念情報として「店員」24、「客」25 および「強盗」26 などを持つ。ここで、概念情報「人」21 と下位の概念情報 24 ~ 26 の間には、is - a 関係が用いられている。つまり、概念情報「店員」24、「客」25 および「強盗」26 は、概念情報「人」21 の種類である。そして、概念情報「店員」24、「客」25 および「強盗」26 は、下位の概念情報 27 ~ 29 に、has - a 関係で接続されている。このように特定のドメイン名情報に対して、has - a 関係と is - a 関係とを交互に繰り返すことにより、オントロジー情報は構築されている。

#### 【0021】

図 17 は、概念構造記憶部 M 1 が記憶する概念構造情報の 1 つである意味ネットワークの一例を示す図である。同図では、ドメイン名情報「店舗」20 と複数の概念情報 21 ~ 29 が有向グラフで接続されていることが示されている。意味ネットワークは、2 つの概念とそれらの概念間を結ぶ矢印の集合で構成されている。各矢印は、2 つの概念間の関係を、主語、動詞および目的語の形式で表現している。具体的には、矢印の元の概念が主語、矢印の先の概念が目的語、そして矢印に付加された語が動詞を表している。例えば、同図では、「である」という動詞が付加された矢印 30 は、店員 31 を主語、人 32 を目的語とし、「店員は人である」という関係を表している。同図では、図 2 と同様に、ルートのドメイン名情報「店舗」と、抽象的な概念情報「人」、「機器」、「家具」および「店員」との関係を、上述した主語・動詞・目的語の形式で表している。また、「人」を中心に、更に細分化した「店員」、「客」および「強盗」、さらには「制服」、「カゴ」、「バーコードリーダ」、「鞆」、「財布」、「カゴ」、「サングラス」、「ナイフ」および「バッド」などの関係を同様の形式で表している。また、各概念情報には、概念名情報の他に、該概念を識別する認識対象 ID と、上位の概念を識別する上位 ID とが記載されている。

#### 【0022】

以上説明したように、概念構造情報には、例えば、オントロジー情報や意味ネットワークなどがある。以降は、説明を簡単化するために、概念構造記憶部 M 1 が概念構造情報としてオントロジー情報を記憶している場合について説明する。

#### 【0023】

図 3 には、本実施形態の概念構造記憶部 M 1 が記憶する情報の一例を示す表を図示する。同図が示すように、ドメイン ID は、例えば、アルファベット及び数字とから成る文字列である。例えば、2 つのドメインは、ドメイン ID 「R100」とドメイン ID 「R200」とにより識別される。そして、同図では、ドメイン ID 「R100」と、ドメイン名情報「店舗」と、オントロジー情報とが、ドメイン ID 「R100」に関連付けられて記憶されている。また、ドメイン ID 「R200」と、ドメイン名情報「駅」と、オントロジー情報とが、ドメイン ID 「R200」に関連付けられて記憶されている。つまり、該表の行が特定のドメインのオントロジー情報に対応し、該表全体がオントロジー情報の集合に対応する。なお、オントロジー情報は、図 2 では視覚的に理解しやすいように木構造を可視化しているが、概念構造記憶部 M 1 上には、UML (Unified Modeling Language) などを用いて表現されたテキストデータとして記憶される。

10

20

30

40

50

## 【 0 0 2 4 】

図 1 の説明に戻ると、動画像データ記憶部 M 2 は、認識対象を識別する認識対象 ID と、認識対象を言語的に示す認識対象名情報と、動画データと、データの種別を示すデータ種情報とを、データ ID に関連付けて記憶する。ここで、データ ID とは、動画データのの一つである動画データとを識別する情報であり、認識対象 ID とは、提供者により予め設定された認識対象を識別する情報である。なお、この認識対象 ID は、概念構造記憶部 M 1 の説明にて上述した概念情報を識別する認識対象 ID と共通である。認識対象名情報は、提供者により予め設定された各認識対象を言語化したものである。認識対象名情報は、例えば、「人」、「車」、「歩いている」、「走っている」、「鞆」および「カゴ」などである。また、この認識対象名情報には、物体の領域の動画像上の座標と大きさを 10 含む。動画像データとは、予め提供者により認識対象に属すると判断された動画像データの 1 つである。また、データ種情報とは、該動画データが、学習用なのか評価用なのかを区別する情報である。

## 【 0 0 2 5 】

図 4 には、本実施形態の動画像データ記憶部 M 2 が記憶する情報の一例を示す表を図示する。同図が示すように、データ ID と認識対象 ID とは、例えば、アルファベット及び数字とから成る文字列である。例えば、2 つのデータは、データ ID 「 D 0 0 0 1 」とデータ ID 「 D 0 0 0 2 」とにより識別される。また、2 つの認識対象は、認識対象 ID 「 C 1 0 0 0 」と認識対象 ID 「 C 1 1 0 0 」とにより識別される。また、同図が示すように、認識対象名情報は、「人」、「店員」という物体の状態を説明する言語情報に加え、 20 該物体の動画像上の座標と大きさを含んでいる。この領域は x 座標、 y 座標、高さ、幅の順番で表現される。具体的には、同図では、データ ID 「 D 0 0 0 1 」のラベル情報は、物体の状態は「人」であり、領域の座標は ( 5 0 0 、 1 0 ) 、領域の高さは 1 8 0 、および幅は 5 0 であることが示されている。そして、同図では、データ ID 「 D 0 0 0 1 」と、認識対象 ID 「 C 1 0 0 0 」と、認識対象名情報「人」、領域： ( 5 0 0 、 1 0 、 1 8 0 、 5 0 ) と、動画像データとが、データ ID 「 D 0 0 0 1 」に関連付けられていることが示されている。また、データ ID 「 D 0 0 0 2 」と、認識対象 ID 「 C 1 1 0 0 」と、ラベル情報「店員」、領域： ( 2 0 0 、 2 0 0 、 1 8 0 、 5 0 ) と、動画像データとが、データ ID 「 D 0 0 0 2 」に関連付けられていることが示されている。つまり、該表の行が動画像データに対応し、該表全体が動画像データに対応する。 30

## 【 0 0 2 6 】

なお、動画像データが外部の記憶装置に記憶される場合は、動画像データには、該記憶装置上の該動画像データの場所を示すアドレスが記憶される。ここで外部の記憶装置には、例えば、インターネットを介して接続したクラウド上のサーバなどが適用できる。また、アドレスには、例えば、IP ( Internet Protocol ) アドレスや、URL ( Uniform Resource Locator ) などが適用できる。

## 【 0 0 2 7 】

再び、図 1 を参照して、認識学習装置 1 0 の構成について説明する。意味的関連度生成部 1 1 は、オントロジー情報に基づき特定ドメインと概念情報との意味的関連度を生成する。具体的には、端末装置 1 0 0 からドメイン ID と、操作情報「プレトレーニングの実行」を入力、指示したことに応じて、入力したドメイン ID に関連付けられたドメイン ID と、ドメイン名情報とオントロジー情報とを概念構造記憶部 M 1 から読み込む。そして、意味的関連度生成部 1 1 は、読み込んだオントロジー情報に含まれる概念情報の一つ一つについて、該ドメインに対する意味的関連度を、読み込んだオントロジー情報に基づいて生成する。ここで、意味的関連度の生成方法としては、例えば、次の 3 つの方法がある。 40

## 【 0 0 2 8 】

第 1 の意味的関連度生成方法として、意味的関連度生成部 1 1 は、読み込んだオントロジー情報の木構造における各概念情報の深さ ( 階層 ) に反比例するように、各概念情報の意味的関連度を生成する。例えば、図 2 のドメイン名情報「店舗」 2 0 に関するオントロ 50



ジー情報の木構造において意味的関連度は、概念情報「人」21、「機器」22、および「家具」23が最も高く、次に概念情報「店員」24、「客」25および「強盗」26が高い。具体的には、ドメイン*i*に対して概念情報*c<sub>j</sub>*の意味的関連度  $R_i(c_j)$  は、木構造の下位層の概念情報ほど低くなるように、例えば、次の数式1のように定義される。

【0029】

【数1】

$$R_i(c_j) = \frac{1}{h_i(c_j)} \quad \dots \quad (\text{数式1})$$

10

ここで、 $h_i(c_j)$  は、ドメイン*i*のオントロジー情報における*c<sub>j</sub>*の階層を表し、 $R_i(c_j)$ の最大値は1である。

【0030】

第2の意味的関連度生成方法として、意味的関連度生成部11は、第1の意味的関連度生成方法による深さに反比例する値に加えて、各概念情報のオントロジー情報の木構造における発生頻度に比例するように、各概念情報の意味的関連度を生成する。例えば、図2のドメイン名情報「店舗」20に関するオントロジー情報の木構造において、概念情報「カゴ」は27および28に2回出現するため、下位層にあるが高い意味的関連度を持つ。具体的には、ドメイン*i*に対して概念情報*c<sub>j</sub>*の意味的関連度  $R_i(c_j)$  は、例えば、次の数式2のように定義される。

20

【0031】

【数2】

$$R_i(c_j) = \frac{1}{h_i(c_j)} + \frac{N_i(c_j)}{\max_l(N_i(c_l))} \quad \dots \quad (\text{数式2})$$

30

ここで、 $N_i(x_j)$  は、ドメイン*i*のオントロジー情報における*c<sub>j</sub>*の出現回数であり、 $R_i(c_j)$ の最大値は2である。

【0032】

第3の意味的関連度生成方法として、意味的関連度生成部11は、各概念情報のオントロジー情報の木構造における子孫の数（すなわち、その候補より下位階層の概念情報の数）を、意味的関連度として生成する。例えば、図2のドメイン名情報「店舗」20に関するオントロジー情報の木構造において、概念情報「人」21は、子孫24～29の概念情報の数に相当する。

【0033】

40

そして、意味的関連度生成部11は、入力したドメインIDとともに、読み込んだオントロジー情報に含まれる概念情報と生成した意味的関連度情報との組み合わせの集合と、読み込んだオントロジー情報と、を認識対象選定部12に出力する。

【0034】

認識対象選定部12は、意味的関連度情報に基づき、概念情報の中から認識対象を選択する。具体的には、認識対象選定部12は、意味的関連度生成部11からドメインIDと、意味的関連度情報と、概念情報との組みの集合と、オントロジー情報とを入力したことに応じて、意味的関連度情報に基づき、概念情報の集合から認識対象を選定する。つまり、入力した概念情報は認識対象の候補である。ここで、認識対象の選定方法としては、例えば、次の2つの方法がある。

50

## 【 0 0 3 5 】

第1の認識対象選定方法として、認識対象選定部12は、所定の閾値以上の意味的関連度と同じ組の概念情報を認識対象として選定する。この閾値は、例えば、0から1の値を取り、意味的関連度は1以下になるように正規化される。具体的には、意味的関連度生成部11が第1の意味的関連度生成方法を用いている場合は、意味的関連度の最大値が既に1なので正規化は行わない。一方、意味的関連度生成部11が第2の意味的関連度生成方法を用いている場合は、意味的関連度の最大値が2なので、最大値が1になるように意味的関連度を2で割ることにより正規化を行う。

## 【 0 0 3 6 】

第2の認識対象選定方法として、認識対象選定部12は、入力した概念情報の数に対して所定の割合の概念情報を認識対象として選定する。具体的には、入力した概念情報を、意味的関連度の降順にソートし、上から順に所定の割合の概念情報を、認識対象として選定する。なお、詳細な説明は省くが、認識対象選定方法で用いられる所定の閾値または所定の割合は、端末装置100の表示部DSに表示された数値情報を人が調整することでもできる。その際、操作検出部OPは、人による該数値情報の変更を示す操作を検出し、該数値情報と操作情報とを認識学習装置10に出力する。認識学習装置10は、端末装置100から該数値と操作情報とを入力したことに応じて、該数値情報を所定の閾値または所定の割合として、自装置内に備える記憶部に記憶させる。

## 【 0 0 3 7 】

次に、認識対象選定部12は、選定した認識対象を識別する認識対象IDと該認識対象と同じ組の意味的関連度とを、それぞれ入力した概念情報と意味的関連度との組みの集合から抽出する。そして、認識対象選定部12は、入力したドメインIDと、オントロジー情報とともに、抽出した認識対象IDと意味的関連度の組の集合とを、認識学習部13に出力する。

## 【 0 0 3 8 】

認識学習部13は、選択された認識対象に係る学習データを用いて認識器を学習する。具体的には、認識学習部13は、認識対象選定部12からドメインIDと、オントロジー情報と、認識対象IDと意味的関連度との組みの集合とを入力したことに応じて、入力した認識対象IDと一致する認識対象IDを保持する。また、認識学習部13は、データ種情報が「学習」である行を動画像データ記憶部M2から読み込む。ここで、読み込んだ行には、認識対象名情報と、認識対象IDと、動画データムとが含まれる。そして、認識学習部13は、読み込んだ動画像データムを入力、読み込んだ認識対象IDを出力とする認識器を学習する。この認識器には、動画像データが静止画であり、認識対象が物体の種類の場合、例えば、Region CNN (R-CNN) (非特許文献3)などが適用できる。また、動画像データが動画で認識対象が物体の行動の場合、3D CNN (非特許文献4)やTwo-stream CNN (非特許文献5)などが適用できる。また、認識器は、静止画や動画像に対応した所定の特徴量抽出とサポートベクトルマシンなどの識別器との組み合わせでもよい。

## 【 0 0 3 9 】

ここでは、認識器としてR-CNNを用いた場合について、認識学習部13の処理を具体的に説明する。認識学習部13は、読み込んだ動画像データの各行に対して順次以下の処理を加える。まず、認識学習部13は、各行の動画像データムである静止画像から複数の物体の領域の候補を抽出し、該行の認識対象名情報が保持する物体の領域とオーバーラップしている割合を計算する。そして、認識学習部13は、該割合が所定の閾値より大きい場合は、該物体の領域の候補が、該行の認識対象名情報であると判定し、該物体の領域の候補を、該動画データムから切り出したパッチ画像を生成する。そして、認識学習部13は、生成した1つまたは複数のパッチ画像と、該行の認識対象IDとを、R-CNNの入力と出力との組みとして、学習データ集合に追加する。そして、該処理が読み込んだ全ての行に適用した後、学習データ集合を用いて、CNNを学習する。なお、認識学習部13は、CNNのパラメータの初期値をランダムに決定する。

## 【 0 0 4 0 】

なお、認識学習部 1 3 は、入力した意味的関連度の集合に基づき、各認識対象の重要度情報を生成する。具体的には、ドメイン  $i$  における認識対象  $c_j$  の重要度情報  $I_j(c_j)$  は、例えば、次の数式 3 のように意味的関連度に比例するように定義される。

## 【 0 0 4 1 】

$$I_j(c_j) = R_j(c_j) \cdots (\text{数式 3})$$

ここで、 $R_j(c_j)$  は比例定数である。そして、認識学習部 1 3 は、該重要度情報が高い認識対象の認識精度を優先するように、R - CNN の学習を施す。具体的には、次の数式 4 のように R - CNN の最小化するドメイン  $i$  の識別誤差に、重要度情報  $I_i(c_j)$  が重みとして適用される。

## 【 0 0 4 2 】

## 【数 3】

$$E_i \equiv \frac{1}{2} \sum_{n=1}^N \sum_{j=1}^C I_i(c_j) (y_j^n - t_j(x^n))^2 \cdots (\text{数式 4})$$

ここで、 $N$  は学習データの数、 $C$  は学習データが含む認識対象の数、 $y_n$  は  $n$  番目の学習データの出力に対応する認識対象の数の大きさのベクトルである。 $n$  番目の  $y$  の各要素は、学習データの出力に対応する場合は 1、それ以外 0 の値をとる。そして、 $x_n$  は、 $n$  番目の学習データの入力に対応する。この入力は上述した方法で生成されたパッチ画像である。そして  $t_i$  は R - CNN が入力データ  $x_n$  に対して予測した出力値であり、認識対象の数の大きさのベクトルである。

## 【 0 0 4 3 】

次に、認識学習部 1 3 は、学習した認識器のパラメータと、入力した認識対象 ID と、入力した意味的関連度との組みの集合とを、入力したドメイン ID に関連付けて、認識器記憶部 M 3 に記憶させる。この認識器のパラメータは、例えば、R - CNN のモデルパラメータである。また、認識学習部 1 3 は、入力したドメイン ID と、オントロジー情報と、認識対象 ID の集合とを、認識対象可視化部 1 4 に出力する。また、後述する認識器記憶部 M 3 にドメイン ID と、関連付けてドメイン ID と認識器のパラメータと、入力した認識対象 ID の集合と、入力した意味的関連度の集合とを記憶させる。認識器記憶部 M 3 は、認識器のパラメータを記憶する。具体的には、認識器記憶部 M 3 は、ドメイン ID と、認識器のパラメータと、認識対象 ID の集合と、意味的関連度の集合とを、ドメイン ID に関連づけて記憶する。

## 【 0 0 4 4 】

図 5 には、本実施形態の認識器記憶部 M 3 が記憶する情報の一例を示す表を図示する。認識器記憶部 M 3 には、認識器のパラメータと、認識対象選定部 1 2 により意味的関連度に基づき選定された認識対象 ID の集合と、意味的関連度生成部 1 1 により生成された意味的関連度の集合とが、ドメイン ID と関連づけられて各行に記憶される。

## 【 0 0 4 5 】

認識対象可視化部 1 4 は、選定された認識対象を示す認識対象情報をオントロジー情報上に重畳し表示する。なお、認識対象可視化部 1 4 は、認識対象選定部により選定された各認識対象に対する認識器の認識精度を評価用データから計算し、認識対象可視化情報として生成する。具体的には、認識対象可視化部 1 4 は、認識学習部 1 3 からドメイン ID と、オントロジー情報と、認識対象 ID の集合とを入力したことに応じて、ドメイン ID に関連付けられた認識器のパラメータを認識器記憶部 M 3 から読み込む。また、認識対象可視化部 1 4 は、入力した認識対象 ID と認識対象 ID とが一致し、かつデータ種情報が「評価」である行を、動画データ記憶部 M 2 から読み込む。そして、認識対象可視化部

10

20

30

40

50

14は、認識学習部13にて説明した処理と同様に、読み込んだ各行の情報とに基づき、画像パッチを生成し、入力と出力のペアの集合である評価データを生成する。そして、認識対象可視化部14は、読み込んだ認識器パラメータから構築したR-CNNのモデルに対して、評価データを適用し、各認識対象に対する認識精度を計算する。この認識精度の計算方法として、認識対象可視化部14は、例えば、n番目の評価データの入力 $x_n$ に対するR-CNNの予測結果 $t_j(x_n)$ の最大値を取る要素がn番目の評価データの出力に一致する割合を計算する。つまり、認識対象可視化部14は、各認識対象に対するPrecisionを計算する。

#### 【0046】

そして、認識対象可視化部14は、入力した認識対象IDの集合とオントロジー情報とに基づき、認識器の認識対象を視覚的に表す認識対象可視化情報を生成する。具体的な認識対象可視化情報の生成方法として、認識対象可視化部14は、入力したオントロジー情報が保持する複数の概念情報の概念名情報と上位IDとに基づき、複数の概念名情報のテキスト間をエッジで結んだ木構造を、認識対象可視化情報として生成する。ここで、各概念情報が保持する認識対象IDが、入力した認識対象IDの集合に含まれる場合は、該概念情報が認識対象であることを示す認識対象情報を、認識対象可視化情報に重畳する。なお、認識対象可視化部14は、計算した各認識対象の認識精度を、入力した認識対象IDと一致する認識対象IDを持つ概念情報とともに認識対象情報として、認識対象可視化情報に重畳してもよい。そして、認識対象可視化部14は、生成した認識対象可視化情報を端末装置100に出力する。

#### 【0047】

図6は、認識対象可視化部14が生成した認識対象可視化情報の一例を示す図である。同図では、図2と同様にドメイン名情報「店舗」と複数の概念情報が接続された木構造において、概念情報「店員」が、認識対象であることを示す認識対象情報の矩形60上に記載されている。また、該認識対象に対する認識精度61が、認識対象情報として記載されている。

#### 【0048】

次に、図7を参照して、認識学習システム1における認識学習装置10の動作について説明する。図7は、本実施形態の認識学習システム1の認識学習装置の認識器のプレトレーニングの一例を示すフローチャートである。まず、ステップP101において、端末装置100は、ドメイン名情報一覧を表示する。具体的には、端末装置100は、予め記憶しておいた所定のドメイン名情報の一覧を表示部DSに表示する。

#### 【0049】

次に、ステップP102において、端末装置100は、ドメインIDを認識学習装置10に出力する。具体的には、まず、端末装置100の操作部OPは、表示部DSに表示されたドメイン名情報の一覧に対する、人の選択および「プレトレーニングの実行」の操作情報を検知する。そして、検知したことに応じて、端末装置100は、選択されたドメイン名情報を識別する、自装置内に記憶されていたドメインIDを取得する。そして、端末装置100は、取得したドメインIDを認識学習装置10に出力する。

#### 【0050】

以下、ステップP103以降の処理が認識学習装置10における処理となる。まず、ステップP103において、意味的関連度生成部11は、オントロジー情報を読み込む。具体的には、意味的関連度生成部11は、端末装置100からドメインIDが入力されたことに応じて、ドメインIDに関連付けられた読み込んだオントロジー情報を概念構造記憶部M1から読み込む。

#### 【0051】

次に、ステップP104において、意味的関連度生成部11は、意味的関連度を生成する。具体的には、意味的関連度生成部11は、読み込んだオントロジー情報に含まれる全ての概念情報について、読み込んだドメイン名情報との意味的関連度を、上述した意味的関連度の生成方法を用いて生成する。そして、意味的関連度生成部11は、入力したドメ

10

20

30

40

50

インIDと、読み込んだオントロジー情報に含まれる概念情報と、生成した意味的関連度との組みの集合とを認識対象選定部12に出力する。

【0052】

次に、ステップP105において、認識対象選定部12は、認識対象を選択する。具体的には、意味的関連度生成部11から、ドメインIDと、概念情報と、意味的関連度との組みの集合とを入力したことに応じて、認識対象選定部12は、上述した認識対象の選定方法を用いて、入力した概念情報の集合から認識対象を選定する。そして、認識対象選定部12は、選定した認識対象を識別する認識対象IDと意味的関連度との組みの集合を、入力した概念情報と意味的関連度との組みの集合から抽出し、入力したドメインIDと、オントロジー情報とともに、認識学習部13に出力する。

10

【0053】

次に、ステップP106において、認識学習部13は、重要度情報を計算する。具体的には、認識対象選定部12から、ドメインIDと、オントロジー情報と、認識対象IDと、意味的関連度とを入力したことに応じて、上述した重要度情報の生成方法を用いて、意味的関連度に基づき、各認識対象IDの重要度情報を計算する。

【0054】

次に、ステップP107において、認識学習部13は、認識器をプレトレーニングする。具体的には、認識学習部13は、動画像データ記憶部M2から、入力した認識対象IDと同一の認識対象IDを持ち、データ種情報が「学習」の行を読み込む。そして、認識学習部13は、読み込んだ各行が保持する情報から、入力と出力の組みの集合である学習データを生成する。そして、認識学習部13は、学習データと算出した重要度情報とに基づき、認識器を学習する。そして、認識学習部13は、認識器のパラメータを記憶させる。具体的には、入力したドメインIDと学習した認識器のパラメータとを、該ドメインIDの集合に関連付けて、認識器記憶部M3に記憶させる。また、認識学習部13は、入力したドメインIDと、オントロジー情報と、認識対象IDの集合とを、認識対象可視化部14に出力する。

20

【0055】

次に、ステップP108において、認識対象可視化部14は、認識対象の精度を測定する。具体的には、認識対象可視化部14は、認識学習部13からドメインIDと、オントロジー情報と、認識対象IDの集合とを入力したことに応じて、ドメインIDに関連付けられた認識器のパラメータを、認識器記憶部M3から読み込む。また、認識対象可視化部14は、入力した認識対象IDと同一の認識対象IDを持ち、データ種情報が「評価」の行を、動画像データ記憶部M2から読み込む。そして、読み込んだ各行が保持する情報から入力と出力の組みの集合である評価データを生成し、読み込んだ認識器のパラメータにより構成される認識器の各認識対象に対するPrecisionなどの認識精度を計算する。

30

【0056】

次に、ステップP109において、認識対象可視化部14は、認識対象可視化情報を生成する。具体的には、入力したオントロジー情報が保持する概念情報と、計算した認識精度と、入力した認識対象IDの集合とに基づいて、上述した認識対象可視化情報の生成方法を用いて、認識対象可視化情報を生成する。また、認識対象可視化部14は、生成した認識対象可視化情報を、端末装置100に出力する。ここまでの処理が、認識学習装置10における処理である。

40

【0057】

次に、ステップP110において、端末装置100は、認識対象可視化情報を表示する。具体的には、端末装置100は、認識学習装置10の認識対象可視化部14から、認識対象可視化情報を入力したことに応じて、端末装置100は、入力した認識対象可視化情報を表示部DSに表示する。そして、端末装置100は処理を終了する。

【0058】

なお、本実施形態では、認識学習装置が認識学習部13と認識対象可視化部14とを含

50

む場合について説明したが、認識学習部 13 と認識対象可視化部 14 とは別の装置に備えられていてもよい。その場合、認識学習装置 10 は、概念構造記憶部 M1 と、意味的関連度生成部 11 と、認識対象選定部 12 と保持する。そして、認識学習装置 10 は、ステップ P101 から処理を進めステップ P105 にて、認識対象 ID と意味的関連度との組みの集合と概念構造情報を、別の装置に出力して処理を終了する。

#### 【0059】

以上説明したように、本実施形態の認識学習装置は、認識器が利用されるドメインを概念的に表現するオントロジー情報に基づき、該ドメインに意味的に関連する認識対象を選定する。そして、これらの認識対象に係る学習データを用いて、認識器をプレトレーニングする。これにより、認識器の提供者が膨大な数の認識対象の候補の中からプレトレーニング用の認識対象を選定する負荷を大幅に軽減することができる。また、特定のドメインに関連した認識対象に限定してプレトレーニングすることができるため、認識器の複雑度を抑えることができ、プレトレーニングにおけるオーバーフィットを回避できると期待される。

#### 【0060】

また、認識学習装置の認識対象可視化部 14 は、選定された認識対象をオントロジー情報に重畳して表示するため、ドメインに意味的に関連する網羅的な概念集合に対して、プレトレーニング済みの認識器の認識対象を可視化することができる。これにより、認識器の提供者および利用者は、プレトレーニング済みの認識器のドメインに対する適用範囲を直感的に把握することができる。また、提供者および利用者は、特定のドメインに対して共通の概念構造の理解を持つことができるため、認識器の授受を円滑に行うことができる。

#### 【0061】

また、認識学習装置の認識対象可視化部は、各認識対象に対するプレトレーニング済みの認識器の認識精度を生成し、オントロジー情報に重畳して表示する。これにより、認識器の提供者および利用者は、プレトレーニング済みの認識器の、該ドメインに対する適用範囲を直感的だけではなく、定量的にも把握することができる。

#### 【0062】

また、認識学習装置の認識学習部は、意味的関連度に基づいて選択された認識対象の重要度情報を生成し、選択された認識対象に対して重要度情報に基づいて重みづけしてプレトレーニングを施す。これにより、特定のドメインにおいて、より多くの利用者が必要とする認識対象に対して認識器の精度を優先的にプレトレーニングすることができる。

#### 【0063】

##### [ 第 2 の実施形態 ]

次に、本発明に係る第 2 の実施形態について説明する。なお、上述した第 1 の実施形態における各構成と同一の構成については、同一の符号を付し、その説明を省略する。本実施形態における認識学習システム 1a について、プレトレーニング済みの認識器をファインチューニングする場合を例に説明する。つまり、本実施形態の認識学習装置 10 は、第 1 の実施形態により認識器のプレトレーニングが済み、認識対象可視化情報が端末装置 100 に表示された状態から処理を開始することを前提にしている。そして、利用者からの認識対象可視化情報に対するフィードバックを示す操作情報に基づいて、認識学習装置 10a が適応的に認識器を学習する点において、第 1 の実施形態と異なる。

#### 【0064】

図 8 は、本発明の第 2 の実施形態に係る認識学習システム 1a の構成の一例を示す構成図である。認識学習システム 1a は、認識学習装置 10a と、端末装置 100 とを備えている。端末装置 100 の操作検出部 OP は、第 1 の実施形態と同様に人の表示部 DS に対する操作情報を検知し、選択されたドメイン名情報に対応するドメイン ID と、検出した操作情報とを認識学習装置 10a に出力する。この操作情報は、第 1 の実施形態の操作情報に加え、表示部 DS に表示された認識対象可視化情報に対する、人からの認識対象の「追加」および「削除」や、「ファインチューニングの実行」などがある。なお、表示部 D

10

20

30

40

50

Sには、これら人からの操作情報を人から取得するためのボタンを表示する。

【0065】

図9は、端末装置100の表示部DSに表示された認識対象可視化情報と、操作情報を取得するためのボタンの一例を示している。同図が示すように、プレトレーニングにて認識対象となっている概念情報の隣に「削除」ボタン90が表示される。また、プレトレーニングの認識対象となっていない概念情報の隣には「追加」ボタン91が表示される。さらに、「ファインチューニングの実行」ボタン92が表示される。操作検出部OPは、人による「削除」ボタン90、「追加」ボタン91または「ファインチューニングの実行」ボタン92の押下を示す操作情報を検知する。操作情報が「削除」の場合、端末装置100は、認識対象可視化情報が保持する認識対象IDの集合から、「削除」対象の認識対象IDを削除する。一方、操作情報が「追加」の場合、端末装置100は、認識対象可視化情報が保持する認識対象IDの集合に、「追加」の対象の認識対象IDを追加する。そして、表示部DSは、更新された認識対象可視化情報を再表示する。また、操作情報が「ファインチューニング実行」の場合、端末装置100は、操作情報を、認識対象可視化情報が保持するドメインIDと認識対象IDの集合とともに認識学習装置10aに出力する。

10

【0066】

図8の説明に戻ると、認識学習装置10aは、特定のドメインに対する認識器のファインチューニングを施す装置である。認識学習装置10aは、概念構造記憶部M1と、画像データ記憶部M2と、認識器記憶部M3と、意味的関連度生成部11と、認識対象選定部12と、認識学習部13aと、認識対象可視化部14と、認識対象更新部15とを備えている。

20

【0067】

認識対象更新部15は、端末装置100の表示部DSに表示された認識可視化情報に対する人の操作を示す操作情報に基づき、認識対象を更新する。具体的には、認識対象更新部15は、端末装置100からドメインIDと、操作情報「ファインチューニングの実行」と、認識対象IDの集合とを入力したことを検知する。そして、その入力に応じて、ドメインIDに関連付けられた認識器のパラメータと、認識対象IDの集合と、意味的関連度情報の集合とを、認識器記憶部M3から読み込む。そして、認識対象更新部15は、読み込んだ認識対象IDの集合と認識器のパラメータとを、入力した認識対象IDの集合に基づいて更新する。具体的には、認識対象更新部15は、読み込んだ認識対象IDの集合を、入力した認識対象IDの集合に置き換える。また、認識対象更新部15は、入力した認識対象IDの集合に基づき、読み込んだ認識器のパラメータを更新する。このパラメータ更新方法には、次の2つの方法がある。

30

【0068】

第1のパラメータ更新方法として、認識対象更新部15は、読み込んだ認識対象IDの集合を入力した認識対象IDの集合に置き換える際に、読み込んだ認識対象IDの集合から削除された認識対象IDに関連するパラメータを認識器のパラメータから削除する。具体的には、R-CNNの出力層の全結合のネットワークにおいて、該削除された認識対象IDに対応する出力ノードと隠れ層の全ノードとの結合に用いられる重みパラメータを削除する。

40

【0069】

第2のパラメータ更新方法として、認識対象更新部15は、読み込んだ認識対象IDを入力した認識対象IDの集合に置き換える際に、読み込んだ認識対象IDの集合に追加された認識対象IDに関連するパラメータを認識器のパラメータに追加する。具体的には、R-CNNの出力層に、追加された認識対象IDに対応する新しい出力ノードを追加する。そして、該出力ノードと隠れ層の全ノードとの間の結合する重みパラメータをランダムに設定する。

【0070】

そして、認識対象更新部15は、読み込んだ認識対象IDの集合と、入力した認識対象IDの集合とに基づき、認識学習部13aが用いる認識器の学習関連のパラメータを調整

50

する。この学習関連パラメータ調整方法には、例えば、次の２つの方法がある。

【００７１】

第１の学習関連パラメータ調整方法として、入力した認識対象ＩＤの集合により置き換えられた読み込んだ認識対象ＩＤの数が、所定の閾値以下の場合、認識学習部１３は、Ｒ－ＣＮＮの上位層の学習率を、下位層の学習率に対して大幅に大きな値設定する。例えば、Ｒ－ＣＮＮの出力層の全結合の重みパラメータを、下位層の畳み込み層およびプーリング層の重みパラメータに対して、１０倍や１００倍の値に設定する。つまり、プレトレーニング済みの認識器の認識対象に対して大きな変更は無いため、低レベルなフィルタに相当する下位層にはファインチューニングにて大きな更新を加えないようにする。一方、新たに追加された認識対象を識別するのに、直接的に寄与する上位の全結合ネットワークには大きな更新を加えるようにする。

10

【００７２】

第２の学習関連パラメータ調整方法として、入力した認識対象ＩＤの集合により置き換えられた読み込んだ認識対象ＩＤの数が、所定の閾値より大きい場合、認識学習部１３は、Ｒ－ＣＮＮが学習率を、全体的に高めの値に設定する。例えば、Ｒ－ＣＮＮの出力層の全結合の重みパラメータと、下位層の畳み込み層およびプーリング層の重みパラメータは、同程度の値に設定する。つまり、プレトレーニング済みの認識器の認識対象に対して大きな変更があるため、上位の全結合ネットワークだけでなく、低レベルなフィルタも大きく更新が加わるようにする。

【００７３】

20

そして、認識対象更新部１５は、入力したドメインＩＤと、更新した認識対象ＩＤの集合と認識器パラメータと、調整した学習関連パラメータと、読み込んだ意味的関連度の集合とを、認識学習部１３ａに出力する。

【００７４】

認識学習部１３ａは、認識器をファインチューニングする。具体的には、認識学習部１３ａは、認識対象更新部１５から、ドメインＩＤと、認識器パラメータと、認識対象ＩＤと、学習関連パラメータと、意味的関連度情報とを入力する。そして、第１の実施形態の認識学習部１３と同様に、入力したドメインＩＤの認識器を再度学習する。ただし、認識器のパラメータの初期値をランダムに決定するプレトレーニングによる学習を行う第１の実施形態とは異なり、認識学習部１３ａは、初期値に入力した認識器パラメータに設定し、学習関連のパラメータに入力した学習関連パラメータを設定する。認識学習部１３ａは、学習した認識器のパラメータと、入力した認識対象ＩＤの集合と、意味的関連度の集合とを、入力したドメインＩＤに関連付けて認識器記憶部Ｍ３に記憶させる。

30

【００７５】

次に、図１０を参照して、本実施形態の認識学習システム１ａにおけるファインチューニングの動作について説明する。図１０は、本実施形態の認識学習システム１ａの認識学習装置１０ａのファインチューニングの動作の一例を示すフローチャートである。なお、第１の実施形態と同一の動作については同一の符号を付して、その説明を省略する。

【００７６】

まず、ステップＦ１０１において、端末装置１００は操作情報を取得する。具体的には、端末装置１００は、表示部ＤＳに表示された認識対象可視化情報に対する人の操作を示す操作情報「ファインチューニングの実行」を取得する。そして、端末装置１００は、取得した操作情報と、認識対象可視化情報が保持するドメインＩＤと認識対象ＩＤの集合とを認識学習装置１０ａの認識対象更新部１５に出力する。

40

【００７７】

以下に説明する各工程は、認識学習装置１０ａにおける処理となる。まず、ステップＦ１０２において、認識対象更新部１５は、認識器パラメータを更新する。具体的には、認識対象更新部１５は、ドメインＩＤと認識対象ＩＤの集合とを入力したことに応じて、入力したドメインＩＤに関連付けられた認識器パラメータと、認識対象ＩＤの集合と、意味的関連度情報の集合とを、認識器記憶部Ｍ３から読み込む。そして、認識対象更新部１５

50



は、読み込んだ認識対象IDの集合を、入力した認識対象IDの集合に置き換える。そして、認識対象更新部15は、読み込んだ認識対象IDの集合と、入力した認識対象IDとに基づいて、読み込んだ認識器パラメータを、前述したパラメータ更新方法により更新する。

#### 【0078】

次に、ステップF103において、認識対象更新部15は、学習関連パラメータを調整する。具体的には、認識対象更新部15は、読み込んだ認識対象IDの集合と、入力した認識対象IDの集合とに基づき、前述した学習関連パラメータ調整方法を用いて認識器の学習関連のパラメータを調整する。そして、認識対象更新部15は、更新した認識対象IDの集合および認識器パラメータと、読み込んだ意味的関連度の集合とを、認識学習部13aに出力する。

10

#### 【0079】

次に、ステップF104において、認識学習部13aは、初期パラメータと学習関連パラメータとを設定する。具体的には、認識対象更新部15から、認識対象IDの集合と、学習関連パラメータと、認識器パラメータと、意味的関連度の集合とを入力したことに応じて、認識器の初期パラメータを、入力した認識器パラメータに設定する。そして、認識器の学習に用いる学習関連のパラメータを、入力した学習関連パラメータに設定する。そして、認識学習部13aは、処理をステップP107に移す。

#### 【0080】

以上説明したように、本実施形態に係る認識学習装置は、オントロジーとともに表示された認識器の認識対象に対する人の操作に基づき認識対象を変更し、ファインチューニングを施す。これにより、認識器の提供者及び利用者は、ドメインに必要とされる網羅的な認識対象を把握しながら、直感的な操作により、該ドメイン上の特定のユースケースに合わせて認識器の対象範囲の編集およびファインチューニングを行うことが出来る。

20

#### 【0081】

##### [第3の実施形態]

次に、本発明を実施するための第3の実施形態について、図面を参照して説明する。なお、上述した第1、第2の実施形態における各構成と同一の構成については、同一の符号を付して説明を省略する。本実施形態における認識学習システム1bは、利用者が所有している独自の動画像データを追加し、認識器をファインチューニングする場合に適用できるものである。

30

#### 【0082】

図11は、本発明の第3の実施形態に係る認識学習システム1bの構成の一例を示す構成図である。認識学習システム1bは、認識学習装置10bと、端末装置100とを備えている。

#### 【0083】

端末装置100の表示部DSは、第1の実施形態と同様に認識対象可視化情報を表示するとともに、各認識対象の学習に用いられた動画像データを表示する。具体的には、操作検出部OPが、認識対象可視化情報上の概念情報に対する人のクリックが示す「画像の表示」操作情報を検知する。そして、端末装置100は、操作情報が「画像の表示」の場合、クリック対象の認識対象を識別する認識対象IDを、認識学習装置10bの動画像データ編集部16に出力する。また、操作検出部OPが、認識対象可視化情報に対する動画像データのドロップアンドドロップが示す操作情報「画像の追加」を検知する。そして、端末装置100は、操作情報が「画像を追加」の場合、追加対象の概念情報を識別する認識対象IDと、追加された動画像データと認識対象名情報とを、認識学習装置10bの動画像データ編集部16に出力する。

40

#### 【0084】

図12は、端末装置100の表示部DSに表示された認識対象可視化情報と、認識器の学習に用いられた動画像データの表示の一例を示している。同図が示すように、人が認識対象可視化情報上の、概念情報をクリック120すると、操作検出部OPは、操作情報「

50

画像を表示」を検出する。そして、端末装置 100 は、操作情報とクリック対象の認識対象 ID を認識学習装置 10b に出力したことに応じて、認識学習装置 10b から、動画像データを取得する。そして、表示部 DS は、取得した動画像データの一覧 121 を表示する。

#### 【0085】

図 13 は、端末装置 100 の動画像データの追加の一例を示している。まず、図 12 と同様に、人が認識対象可視化情報上の概念情報をクリック 130 すると、表示部 DS は、取得した動画像データの一覧 131 を表示する。そして、人が新規に追加したい動画像データを、動画像データの一覧 131 にドラッグアンドドロップ 132 すると、操作検出部 OP は、操作情報@画像を追加」を検出する。そして、表示部 DS は、該動画像データ 133 を追加先の一覧 131 を表示する。それとともに、端末装置 100 は、追加された動画像データと追加先の認識対象 ID と認識対象名情報とを認識学習装置 10b に出力する。

#### 【0086】

再び、図 11 を参照して、認識学習装置 10b の詳細な構成について説明する。認識学習装置 10b は、特定のドメインに対する認識器のプレトレーニングおよびファインチューニングを施す装置である。認識学習装置 10b は、概念構造記憶部 M1 と、動画像データ記憶部 M2 と、認識器記憶部 M3 と、意味的関連度生成部 11 と、認識対象選定部 12 と、認識学習部 13b と、認識対象可視化部 14 と、動画像データ編集部 16 とを備えている。

#### 【0087】

動画像データ編集部 16 は、端末装置 100 の表示部 DS に表示された認識可視化情報に対する人の操作を示す操作情報に基づき、認識器の学習に用いる動画像データを編集する。具体的には、端末装置 100 から操作情報と、認識対象 ID と、認識対象名情報とを入力したことに応じて、動画像データ編集部 16 は、操作情報の内容に基づいて処理を行う。操作情報が「画像を表示」の場合、動画像データ編集部 16 は、入力した認識対象 ID と一致する認識対象 ID を保持する行を、動画像データ記憶部 M2 から読み込む。そして、動画像データ編集部 16 は、読み込んだ行が保持する動画像データを、端末装置 100 に出力する。一方、操作情報が「画像を追加」の場合、動画像データ編集部 16 は、端末装置 100 から新たに動画像データを入力する。そして、新たな割り振った動画像データ ID に関連付けて、入力した認識対象 ID と、認識対象名情報と、動画像データと、「学習」に設定したデータ種情報とを、動画像データ記憶部 M2 に記憶させる。

#### 【0088】

なお、本実施形態では、動画像データ編集部 16 が動画像データを動画像データ記憶部 M2 に追加する場合の説明をしたが、同様に、動画像データ編集部 16 は、動画像データ記憶部 M2 から動画像データを削除することができる。また、認識学習システム 1b における動画像データの追加の動作は、実施形態 1 の識別装置の動作と基本的には同じであるため、説明を省略する。

#### 【0089】

以上説明したように、認識学習装置の動画像データ編集部は、認識器のプレトレーニングおよびファインチューニングに利用された動画像データを、オントロジー情報上の概念情報ごとに表示することができる。また、動画像データ編集部は、各概念情報に動画像データを追加したり、既にある動画像データを削除したりすることができる。これにより、認識器の提供者および利用者は、認識器により対応可能な各認識対象の多様性を直感的に確認することができるとともに、データの追加と削除により各認識対象の多様性を調整することができる。

#### 【0090】

##### [ 第 4 の実施形態 ]

次に、本発明を実施するための第 4 の実施形態について、図面を参照して説明する。なお、上述した第 1 ~ 第 3 の実施形態における各構成と同一の構成については、同一の符号

10

20

30

40

50

を付して説明を省略する。本実施形態における認識学習システム1cは、文章データから自動的に生成したオントロジー情報に基づきプレトレーニング用の認識対象を選定する場合に適用できるものである。

#### 【0091】

図14は、本発明の第4の実施形態に係る認識学習システム1cの構成の一例を示す構成図である。認識学習システム1cは、認識学習装置10cと、端末装置100とを備えている。

#### 【0092】

認識学習装置10cは、特定のドメインに対する認識器のプレトレーニングおよびファインチューニングを施す装置である。認識学習装置10cは、概念構造記憶部M1と、動画像データ記憶部M2と、認識器記憶部M3とを備える。また更に、文章データ記憶部M4と、意味的関連度生成部11と、認識対象選定部12と、認識学習部13と、認識対象可視化部14と、オントロジー生成部17とを備えている。

#### 【0093】

文章データ記憶部M4は、図示は省略するが、文章を識別する文章IDと、テキスト形式の文章データを示す文書情報とを、文章IDに関連付けて記憶する。この文章データには、例えば、Wikipediaなどの辞書データベースや、インターネット上で公開されているニュースが適用できる。

#### 【0094】

なお、本実施形態では、文章データ記憶部M4が記憶する文章情報は、テキスト形式の文章データである場合の説明をしたが、文章情報は、外部の記憶装置が記憶する文章データのアドレスを示す情報でもよい。ここで、外部の記憶装置には、例えば、インターネットを介して接続したウェブサーバーやストレージサーバなどが適用できる。また、アドレスには、例えば、IP(Internet Protocol)アドレスや、URL(Uniform Resource Locator)などが適用できる。また、本実施形態では、認識学習装置が文章データ記憶部M4を備える場合について説明したが、文章データ記憶部M4は、外部の記憶装置が備えてもよい。

#### 【0095】

端末装置100の表示部DSは、第1の実施形態と同様にドメイン名情報の一覧を表示するとともに、「オントロジー情報の自動生成」ボタンを表示する。そして、操作検出部OPは、人によるドメイン名情報の選択と、該ボタンの押下とを表す操作情報を検出する。そして、端末装置100は、選択されたドメイン名情報と、該ドメイン名情報を識別するドメインIDと、操作検出部OPが検出した操作情報「オントロジー情報の自動生成」とを、認識学習装置10cに出力する。

#### 【0096】

また、端末装置100は、自装置がドメイン名情報と、ドメインIDと、操作情報「オントロジー情報の自動生成」を認識学習装置10cに出力したことに応じて、認識学習装置10cからオントロジー情報を入力する。そして、端末装置100の表示部DSは、入力したオントロジー情報を可視化した木構造を、表示する。

#### 【0097】

オントロジー生成部17は、概念構造を生成する概念構造生成部として機能し、具体的には、予め記憶しておいた文章データから前記特定のドメインに関する前記オントロジー情報を生成する。具体的には、オントロジー生成部17は、端末装置100からドメイン情報と、ドメインIDと、操作情報「オントロジー情報の自動生成」を入力したことに応じて、文章データ記憶部M4からドメイン名情報を含む文章情報を読み込む。そして、オントロジー生成部17は、読み込んだ文章情報から、所定のオントロジー情報生成方法を用いて、オントロジー情報を生成する。このオントロジー情報の生成方法としては、日本語Wikipediaからis-a関係およびhas-a関係を抽出した日本語Wikipediaオントロジーや、日本語Wikipediaと日本語WordNetを統合したオントロジーなどが適用できる。

10

20

30

40

50

## 【0098】

そして、オントロジー生成部17は、生成したオントロジー情報と、入力したドメインIDと、ドメイン情報とを、ドメインIDに関連付けて、概念構造記憶部M1に記憶させるとともに、端末装置100に生成したオントロジー情報を出力する。

## 【0099】

なお、オントロジー生成部17は、予め細かい粒度の概念のオントロジー情報を記憶しておき、文章情報から生成された粗い粒度の概念構造の一部として活用してもよい。例えば、図2が示す「店舗」ドメインの1層目と2層目との概念構造20～26を、文章情報から自動的に生成し、3層目以降を、予め記憶しておいた「店員」、「客」および「強盗」のオントロジー情報を用いることにより生成してもよい。

10

## 【0100】

なお、オントロジー生成部17は、文章データ記憶部M4の文章情報を予め定められた期間ごとに更新し、概念構造記憶部M1がドメインIDに関連付けて記憶するオントロジー情報を更新してもよい。また、認識学習システム1bにおける動画像データの追加の動作は、第1の実施形態の識別装置の動作と基本的には同じであるため、説明を省略する。

## 【0101】

以上説明したように、認識学習装置のオントロジー生成部17は、文章データから自動的にオントロジー情報を生成することができる。これにより、提供者がオントロジー情報を構築する負荷を軽減することができる。また、オントロジー生成部17は、定期的にオントロジー情報を更新することができるので、認識器の提供者および利用者は、時代の移り変わりや新しい技術の誕生や流行などによる概念構造の変化に適応したオントロジー情報を利用することができる。

20

## 【0102】

## [第5の実施形態]

次に、本発明を実施するための第4の実施形態について、図面を参照して説明する。なお、上述した第1～第3の実施形態における各構成と同一の構成については、同一の符号を付して説明を省略する。本実施形態の認識学習システム1eについて、認識オンラインサービスとして提供する場合について説明する。ここで、認識オンラインサービスとは、インターネットまたはLAN(Local Area Network)などに接続されたサーバ端末上で動作するディープラーニングなどの認識器を、ユーザが自身のデータに合わせて調整し活用できるサービスである。例えば、ユーザはウェブブラウザ上で動作するユーザインタフェースを操作し、自身のデータをサーバ端末にアップロードし、認識器をファインチューニングすることができる。この認識オンラインサービスには、例えば、Google Cloud Platformなどがある。

30

## 【0103】

図18に示すように、本実施の形態に係るオンラインサービスとして実行可能な認識学習システム1eは、認識学習装置10eと、当該認識学習装置10eと通信可能に接続される端末装置100とを含む複数の端末装置とを備えている。これらの装置は、通信ネットワークとしてのインターネット200を介して接続されている。

## 【0104】

なお、本実施の形態の通信ネットワークは、インターネットに限定されるものではなく、端末装置100と認識学習装置10eとの間を通信可能に相互に接続できるものであれば、例えば、専用回線、公衆回線、LAN等であってもよい。或いは、これらを組み合わせたものであってもよい。また、端末装置100と認識学習装置10eとの間の通信は、例えば、TCP/IP上で動作するHTTP(Hyper Text Transfer Protocol)を使用する。なお、TCP/IPは、Transmission Control Protocol/Internet Protocolである。

40

## 【0105】

次に、認識学習装置10eの詳細な構成について説明する。認識学習装置10eは、実施形態1から4にて説明した認識学習装置と同様に、特定のドメインに合わせて認識器を

50

プレトレーニングおよびファインチューニングする装置である。図 17 に示すように、認識学習装置 10e は、概念構造記憶部 M1 と、動画像データ記憶部 M2 と、認識器記憶部 M3 と、意味的関連度生成部 11 と、認識対象選定部 12 と、認識器学習部 13 と、認識対象可視化部 14 と、サーバ通信部 19 とを備えている。つまり、本実施形態の認識学習装置 10e はサーバ通信部 19 を備える点において、第 1 の実施形態の認識学習装置と異なる。

#### 【0106】

サーバ通信部 19 は、ネットワークインタフェースカード等を備え、端末装置 100 との間で、インターネット 200 を介して各種データの送受信を行う。このデータには、例えば、端末装置 100 から認識学習装置 10e に送信される、ユーザの操作を示す操作情報や、認識学習装置 10e から端末装置 100 に送信される、端末装置 100 に表示する認識対象可視化情報などの表示情報が含まれる。ここで表示情報は、インターネットブラウザ上でのユーザインタフェースの実装に必要なユーザインタフェース情報を含む。このユーザインタフェース情報には、例えば、HTML (Hyper Text Markup Language)、CSS (Cascading Style Sheets)、JavaScript (登録商標) 等のプログラムコード、画像、テキスト等である。つまり、認識学習装置 10e は、認識オンラインサービスとして、認識器のプレトレーニングとファインチューニングの機能をインターネット 200 と端末装置 100 とを介してユーザに適用する。

#### 【0107】

図 19 は、認識器として代表的な CNN である Alexnet (非特許文献 6) を用いた場合を例に、ユーザが認識オンラインサービスにて操作するユーザインタフェースの一例を示している。ここで、Alexnet は、ImageNet の 1000 カテゴリのデータを用いてプレトレーニングされた認識器である。同図が示すように、表示部 DS 上に、ウェブブラウザ 180 が表示されていて、該ウェブブラウザ 180 がアクセスしている認識オンラインサービスを提供する認識学習装置 10e の URL 181 を表示している。また、182 には、該ウェブブラウザ 180 が認識学習装置 10e から受信した表示情報が表示されている。ここで、表示情報には、実施形態 1 から 4 にて上述した認識対象可視化情報およびユーザの操作情報を取得するためのボタンなどが HTML などのユーザインタフェースを情報を用いて記述されたものである。例えば、183 と 184 とは、それぞれ HTML の INPUT タグを記述された認識対象の追加と削除のボタンである。

#### 【0108】

ここで、認識対象可視化情報は、AlexNet のプレトレーニングに用いられた ImageNet の 1000 カテゴリの概念構造を表すオントロジー情報である。同図では、AlexNet のプレトレーニングに、Artifact -> Instrumentality -> Equipment -> Electronic equipment に属する CD Player が用いられていることが示されている。つまり、ユーザが AlexNet のプレトレーニングに用いられた各カテゴリの ImageNet データベース全体に対する位置づけを確認できることが示されている。

#### 【0109】

また、同図では、ユーザの操作情報を取得するためのボタンとしては、認識対象の追加 183、削除 184 およびファインチューニング 185 などがあることが示されている。つまり、ユーザが追加、削除ボタンを用いて、認識対象のカテゴリの追加および削除ができ、さらにファインチューニングの実行ボタンにより、AlexNet の認識器を更新することができることが示されている。

#### 【0110】

以上説明したように、本実施形態の認識学習装置は、HTML などを用いたユーザインタフェースの実装およびインターネットとの接続により、ユーザが遠隔地から認識器のプレトレーニングおよびファインチューニングが可能になる。これにより、認識オンラインサービスの提供者は、ユーザにプレトレーニング済みの認識器の網羅性や得意としている

分野を、視覚的に提示することができる。また、ユーザは、ブラックボックス化しているオンラインサービスの認識器の特徴を把握でき、さらには自身のデータや自身の目的に合わせて、直感的な操作により認識器を編集し更新することができる。

#### 【0111】

##### [その他の実施形態]

なお、上記の各実施形態では、オントロジー情報は、特定ドメインに関連する網羅的な概念情報を含む場合について説明したが、該オントロジー情報は、ドメイン内の特定のユースケースや特定のユーザ層ごとに構築されてもよい。例えば、「店舗」ドメイン内の特定のユースケースとしては、「レジ前用」、「商品棚用」、「強盗検知用」、「万引き検知用」および「客層分析用」などがある。また、「店舗」ドメイン内の特定のユーザ層の例としては、「店員用」、「店長用」、および「スーパーバイザー用」などがある。そして、端末装置に表示されたメニューから特定のユースケースや特定のユーザ層を選ぶことにより、自動的に該当するオントロジー情報を読み込んでよい。

10

#### 【0112】

図15は、オントロジー情報の選択を可能にする実施形態の認識学習システム1dの構成の一例を示す構成図である。同図が示すように、認識学習システム1dは、認識学習装置10dと、端末装置100とを備えている。認識学習装置10dは、第1の実施形態の認識装置に加えオントロジー選択部18を備える。

#### 【0113】

端末装置100は、自装置内に記憶しておいたドメイン名情報と特定の利用者層を示す利用者層情報の一覧を表示部DSに表示する。そして、利用者が特定の利用者層情報を選択したことに応じて、端末装置100は、自装置内に記憶しておいた該特定の利用者層情報に対応するドメインIDを、認識学習装置10dに出力する。端末装置100は、ドメインIDを認識学習装置10dに出力したことに応じて、認識学習装置10dから取得したオントロジー情報を表示部DSに表示する。

20

#### 【0114】

オントロジー選択部18は、概念情報を読み込む概念情報選択部として機能し、端末装置100からドメインIDを入力したことに応じて、入力したドメインIDに関連付けて記憶されているオントロジー情報を、概念構造記憶部M1から読み込む。そして、オントロジー選択部18は、読み込んだオントロジー情報を、端末装置100に出力する。

30

#### 【0115】

図16は、端末装置100の表示部DSに表示されたオントロジー情報の選択メニューの一例を示している。同図で示すように、利用者が「オントロジー選定」ボタン150をクリックすると、ドメイン名情報「店舗」151に関連する特定の利用者層情報「店員用」、「店長用」および「スーパーバイザー用」152がプルダウンメニューとして表示される。また更には、ドメイン名情報「駅」153に関連する特定の利用者層情報「運転手用」、「駅員用」および「駅長用」154がプルダウンメニューとして表示される。

#### 【0116】

上述した各実施形態によれば、ドメイン単位で利用者のニーズに関連した認識対象に対してプレトレーニングを行うことにより、ファインチューニングにおいて学習時間の短縮化とオーバーフィットの回避を期待できる。そのため、利用者の満足度が向上することができる。また、認識器の提供者と利用者の間で、特定のドメインに対する共通の概念構造を用いるため、特定のドメインの網羅性を考慮した上で、認識器の適用範囲および精度を共有することができる。また、利用者は特定のドメインに関連する網羅的な認識対象を考慮した上で、自らのユースケースに合わせて、認識対象を直感的に選定して、ファインチューニングすることができる。このため、利用者の満足度を大きく向上することができる。

40

#### 【0117】

なお、本発明に関し、その具体的な構成は上述の実施形態に限られるものではなく、本発明の要旨を逸脱しない範囲の設計等も含まれる。また、各実施形態は、上記の各実施形態を組み合わせて実施してもよい。また、上記の各実施形態においては、複数の状態を識別

50

する問題を例に、本発明の実施形態について説明したが、本発明の装置は、この発明の要旨を逸脱しない範囲内において一般的な識別問題に適用することが可能である。例えば、本発明の装置は、正常と異常を識別する異常検知の問題に適用することができる。

【0118】

また、上記の実施形態においては、動画像データに対する認識器を学習する場合について説明したが、本発明の装置は、この発明の要旨を逸脱しない範囲内において一般的なデータに適用することが可能である。例えば、動画像データ以外の、音声データ、センサーデータ、およびログデータなどに適用することができる。また、言語から生成したオントロジー情報に基づいて選定した認識対象を認識する認識器を学習することから、本発明の装置は、言語と、動画像、音声データ、センサーデータおよびログデータとのマルチモーダル情報を活用しているとも解釈できる。

10

【0119】

また、上記の各実施形態には、認識器をプレトレーニングした後、利用者が個々の目的に合わせ、該認識器をファインチューニングする場合について説明した。しかし、本発明の装置は、この発明の要旨を逸脱しない範囲内において一般的な認識器の学習に適用できる。例えば、プレトレーニングの段階で、人により認識対象をオントロジー情報に基づき選定してもよい。また、動画像データを逐次的に追加し、認識器を更新してもよい。

【0120】

また、上記の各実施形態では、ドメインが店舗の場合を例に説明したが、本発明の装置は、店舗以外の任意のドメインに適用してもよい。店舗以外には、例えば、介護施設、一般家庭、交差点、駅、空港、および市街などがある。

20

【0121】

また、上記の各実施形態では、監視カメラを用いた監視の場合を例に説明したが、本発明の装置は、監視以外の目的にも適用できる。例えば、スポーツのスタッツ分析や一般のカメラにおけるシーンの認識や審美判定などに適用できる。

【0122】

また、上記の各実施形態において、認識学習装置は概念構造記憶部M1、動画像データ記憶部M2、認識器記憶部M3、および文章データ記憶部M4を備えていると説明したが、ネットワークを介したサーバ上や、他の装置がこれらの構成を備えてもよい。また、各装置が備える各部の機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することで、サーバ装置が備える各部による処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、OSや周辺機器等のハードウェアを含むものとする。

30

【0123】

なお、この各装置が備える各部は、専用のハードウェアにより実現されるものであってもよい。または、このサーバ装置が備える各部はメモリおよびCPU（中央演算装置）により構成され、サーバ装置が備える各部の機能を実現するためのプログラムをメモリにロードして実行することによりその機能を実現させるものであってもよい。

【符号の説明】

40

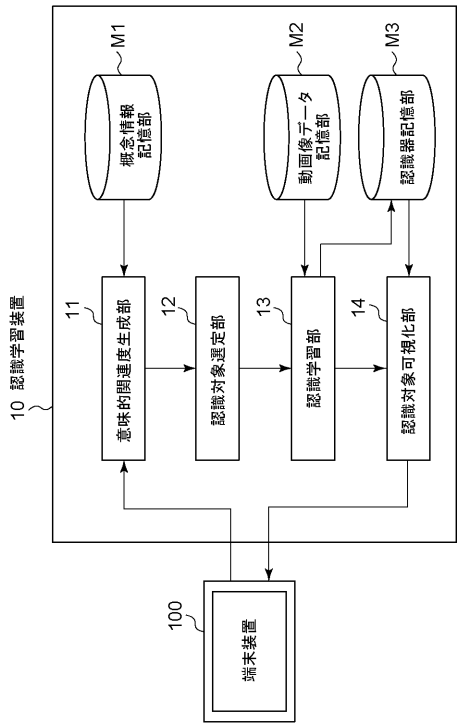
【0124】

- 10 認識学習装置
- 11 意味的関連度生成部
- 12 認識対象生成部
- 13 認識学習部
- 14 認識対象可視化部
- 15 認識対象更新部
- 16 動画像データ編集部
- 17 オントロジー生成部
- 18 オントロジー選択部

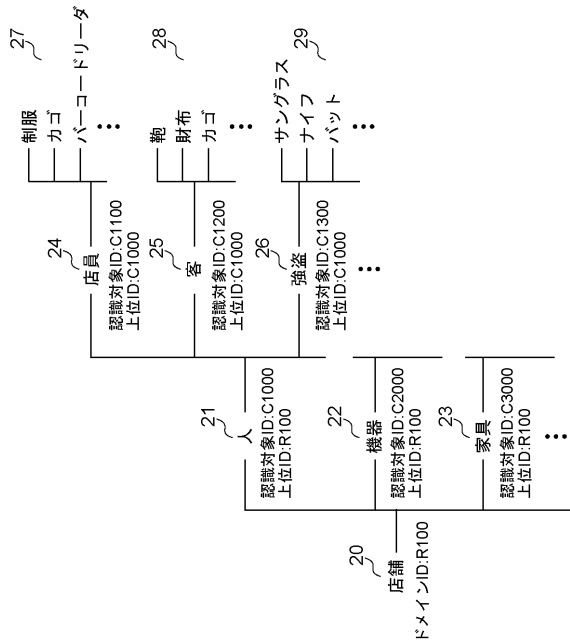
50

- M 1 概念構造記憶部
- M 2 動画像データ記憶部
- M 3 認識器記憶部
- M 4 文章データ記憶部

【 図 1 】



【 図 2 】





【図 3】

M1

ドメインID	ドメイン名情報	オントロジー情報
R100	店舗	
R200	駅	
⋮	⋮	⋮

【図 4】

M2

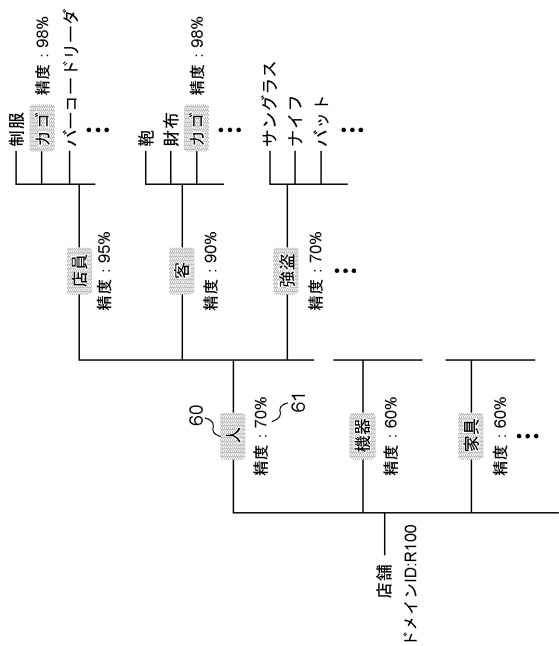
データID	認識対象ID	認識対象名	動画像データム	データ種
D0001	C1000	人 領域 : (500, 10, 180, 50)		学習
D0002	C1100	店員 領域 : (200, 200, 180, 50)		評価
⋮	⋮	⋮		

【図 5】

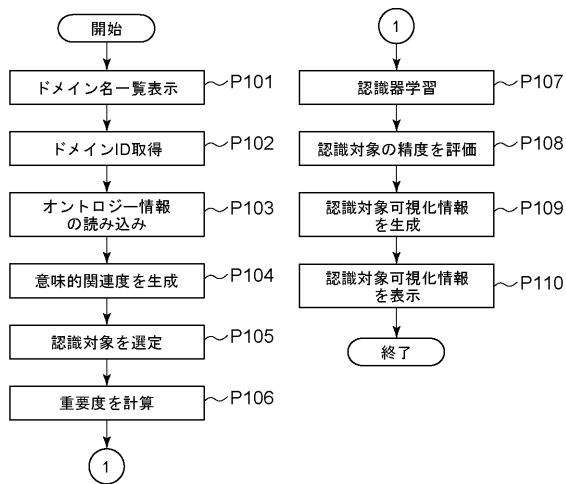
M3

ドメインID	認識器パラメータ	認識対象IDの集合	意味的関連度の集合
R100		C1000, C2000, C3000, C1100, C1200, C1300, ⋮	1, 1, 1, 0.8, 0.8, 0.8, ⋮
⋮	⋮	⋮	⋮

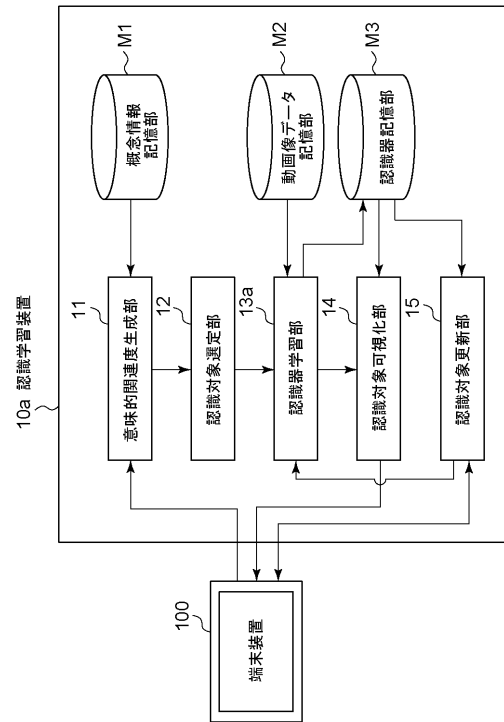
【図 6】



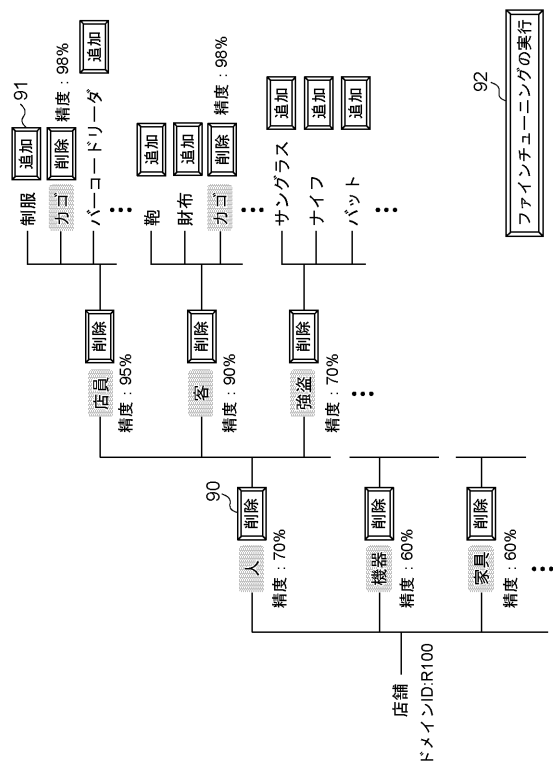
【 図 7 】



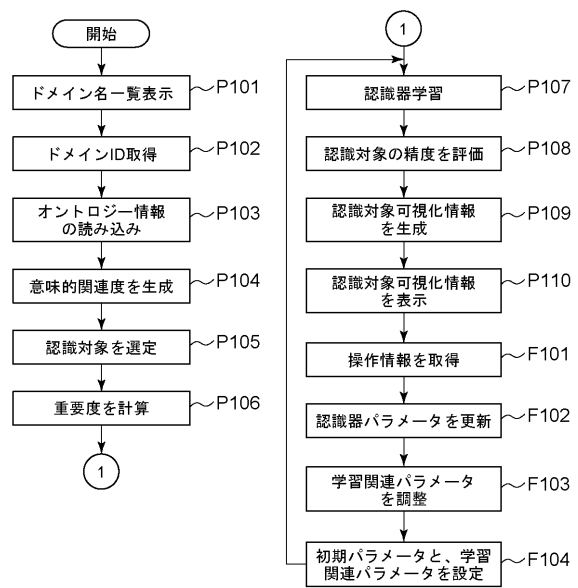
【 図 8 】



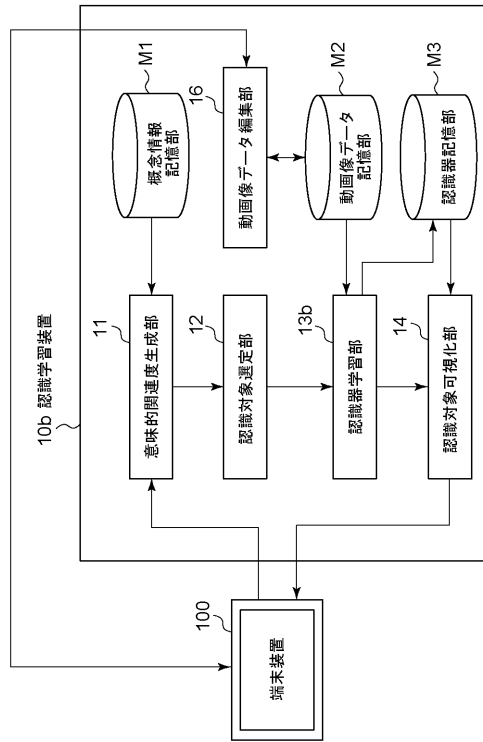
【圖 9】



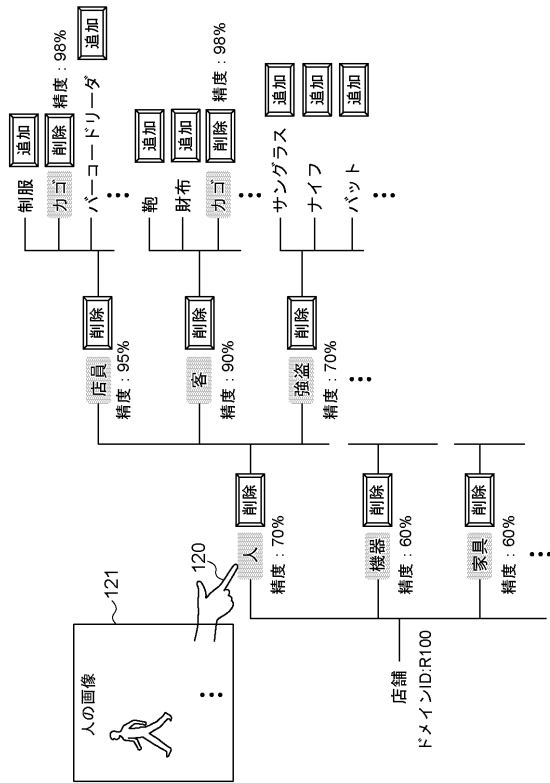
【 図 1 0 】



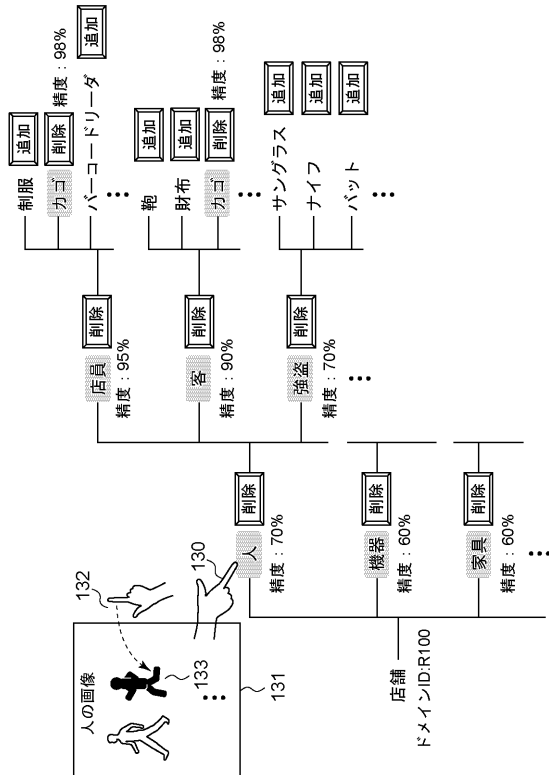
【図 1 1】



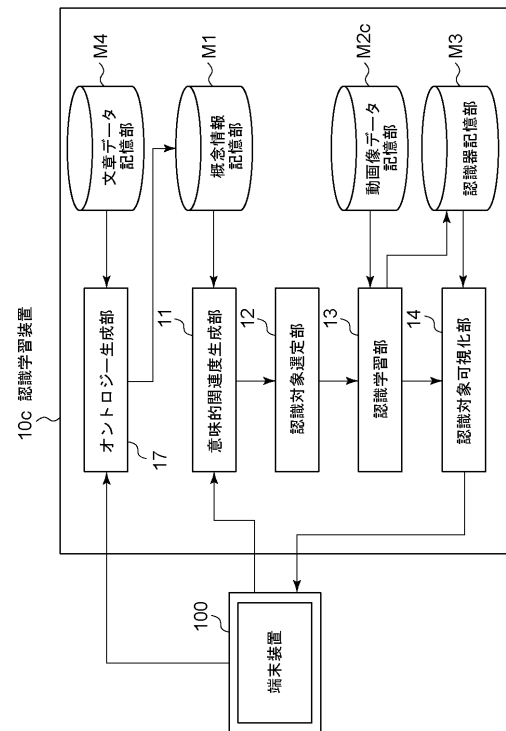
【図 1 2】



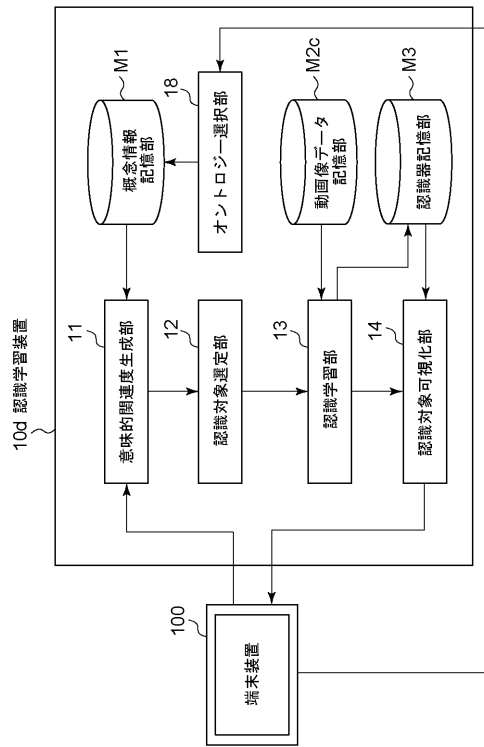
【図 1 3】



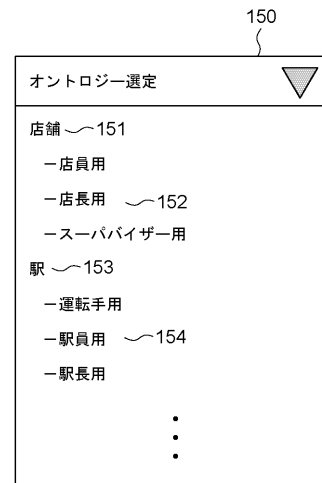
【図 1 4】



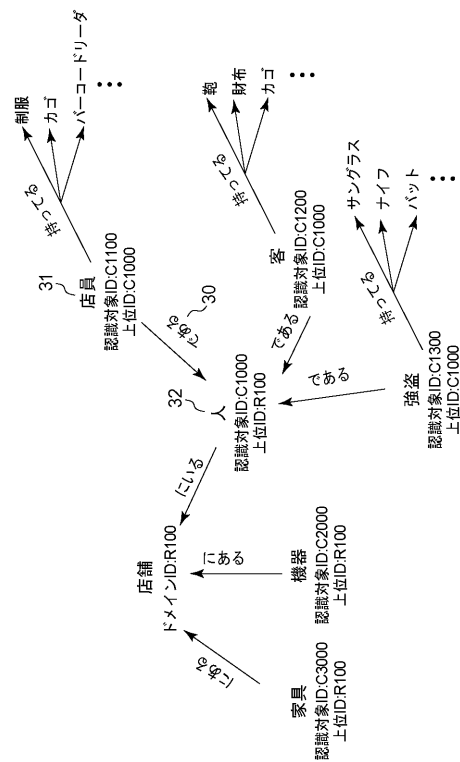
【 図 1 5 】



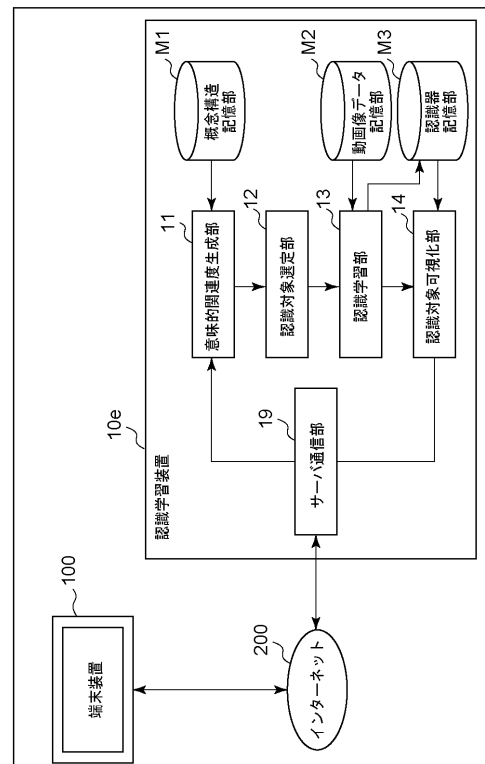
【 図 1 6 】



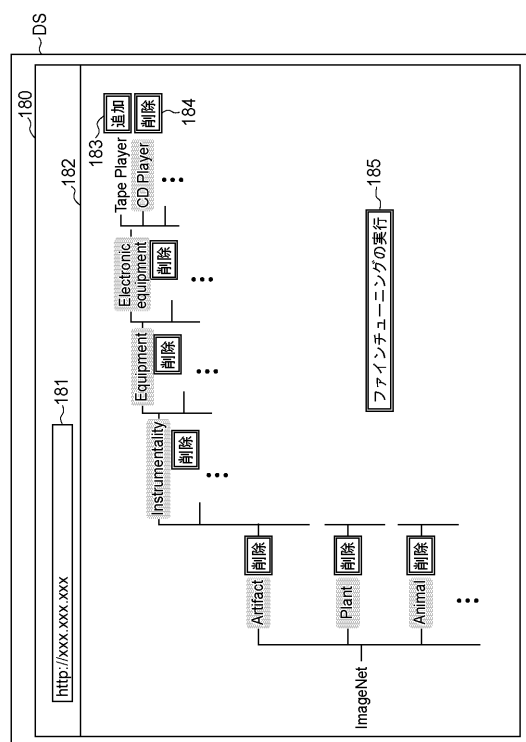
【 図 1 7 】



【 図 1 8 】



【 図 19 】



---

フロントページの続き

(56)参考文献 特開 2 0 1 1 - 1 0 8 0 8 5 ( J P , A )

米国特許出願公開第 2 0 0 4 / 0 2 2 0 9 0 5 ( U S , A 1 )

(58)調査した分野(Int.Cl. , D B 名)

G 0 6 F 1 6 / 9 0 3