

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7496567号
(P7496567)

(45)発行日 令和6年6月7日(2024.6.7)

(24)登録日 令和6年5月30日(2024.5.30)

(51)国際特許分類		F I	
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00 3 5 0 C
G 0 1 N	21/88 (2006.01)	G 0 6 T	7/00 6 1 0 C
G 0 6 N	20/00 (2019.01)	G 0 1 N	21/88 J
		G 0 6 N	20/00 1 3 0

請求項の数 16 (全28頁)

(21)出願番号	特願2022-568080(P2022-568080)	(73)特許権者	314012076 パナソニックIPマネジメント株式会社 大阪府門真市元町2番6号
(86)(22)出願日	令和3年10月14日(2021.10.14)	(74)代理人	110002527 弁理士法人北斗特許事務所
(86)国際出願番号	PCT/JP2021/038140	(72)発明者	ナインゴラン ジェフリー 大阪府門真市大字門真1006番地 パ ナソニックインダストリー株式会社内
(87)国際公開番号	WO2022/123905	(72)発明者	菅澤 裕也 大阪府門真市大字門真1006番地 パ ナソニックインダストリー株式会社内
(87)国際公開日	令和4年6月16日(2022.6.16)	(72)発明者	村田 久治 大阪府門真市大字門真1006番地 パ ナソニックインダストリー株式会社内
審査請求日	令和5年5月24日(2023.5.24)	(72)発明者	佐藤 吉宣
(31)優先権主張番号	特願2020-202864(P2020-202864)		
(32)優先日	令和2年12月7日(2020.12.7)		
(33)優先権主張国・地域又は機関	日本国(JP)		

最終頁に続く

(54)【発明の名称】 処理システム、学習処理システム、処理方法、及びプログラム

(57)【特許請求の範囲】

【請求項1】

ラベルが付与された複数の学習データを取得する第1取得部と、
前記複数の学習データに基づき生成された学習済みモデルを取得する第2取得部と、
ラベルが付与された識別データを取得する第3取得部と、
前記学習済みモデルを用いて前記識別データを識別する識別部と、
前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々の類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する1以上の学習データを抽出する抽出部と、
前記識別データ、及び前記1以上の学習データに基づいて、誤ラベルの有無、及び、前記識別データと前記1以上の学習データとのいずれに前記誤ラベルが付与されているかを判断する判断部と、

10

を備える、

処理システム。

【請求項2】

前記判断部による判断結果に関する情報を外部に提示する提示部を更に備える、
請求項1に記載の処理システム。

【請求項3】

前記判断結果が前記誤ラベルの有ることを示す場合、前記提示部は、前記識別データと、前記1以上の学習データとのいずれに前記誤ラベルが有るかを示す情報を提示する、

20

請求項 2 に記載の処理システム。

【請求項 4】

前記判断結果が前記誤ラベルの無いことを示す場合、前記提示部は、前記識別データと前記 1 以上の学習データとの両方を提示する、

請求項 2 又は 3 に記載の処理システム。

【請求項 5】

前記判断部は、前記識別部による前記識別データの識別結果と、前記識別データに付与された前記ラベルとが不一致の場合に、前記誤ラベルの有無の判断を実行する、

請求項 1 ~ 4 のいずれか 1 項に記載の処理システム。

【請求項 6】

前記判断部は、前記識別データに付与された前記ラベルと前記 1 以上の学習データに付与された前記ラベル、及び、前記識別データと前記 1 以上の学習データとの前記類似度に関する指標、の少なくとも一方に基づき、前記誤ラベルの有無を判断する、

請求項 1 ~ 5 のいずれか 1 項に記載の処理システム。

【請求項 7】

前記判断部は、前記識別データに付与された前記ラベルと前記 1 以上の学習データに付与された前記ラベルとが不一致な割合に基づき、前記誤ラベルの有無を判断する、

請求項 6 に記載の処理システム。

【請求項 8】

前記判断部は、前記識別データに付与された前記ラベルと前記 1 以上の学習データに付与された前記ラベル、及び、前記 1 以上の学習データの前記類似度に関する指標の両方に基づき、前記誤ラベルの有無を判断する、

請求項 6 に記載の処理システム。

【請求項 9】

ラベルが付与された複数の学習データを取得する第 1 取得部と、

前記複数の学習データに基づき生成された学習済みモデルを取得する第 2 取得部と、

ラベルが付与された識別データを取得する第 3 取得部と、

前記学習済みモデルを用いて前記識別データを識別する識別部と、

前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々の類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する 2 以上の学習データを抽出する抽出部と、

前記識別データ、及び前記 2 以上の学習データに基づいて、誤ラベルの有無を判断する判断部と、

を備え、

前記判断部は、

前記 2 以上の学習データから、前記類似度に関する指標が所定条件を満たすほどに前記識別データに類似する特定の学習データを特定し、

前記特定の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと不一致であり、かつ、前記 2 以上の学習データにおける前記特定の学習データ以外の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと一致する場合、前記識別データよりも前記特定の学習データの方に前記誤ラベルが有る可能性が高いと判断する、

処理システム。

【請求項 10】

ラベルが付与された複数の学習データを取得する第 1 取得部と、

前記複数の学習データに基づき生成された学習済みモデルを取得する第 2 取得部と、

ラベルが付与された識別データを取得する第 3 取得部と、

前記学習済みモデルを用いて前記識別データを識別する識別部と、

前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々の類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する 2 以

10

20

30

40

50

上の学習データを抽出する抽出部と、

前記識別データ、及び前記2以上の学習データに基づいて、誤ラベルの有無を判断する判断部と、

を備え、

前記判断部は、

前記2以上の学習データから、前記類似度に関する指標が所定条件を満たすほどに前記識別データに類似する特定の学習データを特定し、

前記特定の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと不一致であり、かつ、前記2以上の学習データにおける前記特定の学習データ以外の学習データに付与された前記ラベルが前記特定の学習データに付与された前記ラベルと一致する場合、前記特定の学習データよりも前記識別データの方に前記誤ラベルが有る可能性が高いと判断する、

10

処理システム。

【請求項11】

前記学習済みモデルは、深層学習を適用して前記複数の学習データに基づき生成されたモデルである、

請求項1～10のいずれか1項に記載の処理システム。

【請求項12】

請求項1～11のいずれか1項に記載の処理システムと、

前記学習済みモデルを生成する学習システムと、

20

を備える、

学習処理システム。

【請求項13】

ラベルが付与された複数の学習データを取得する第1取得ステップと、

前記複数の学習データに基づき生成された学習済みモデルを取得する第2取得ステップと、

ラベルが付与された識別データを取得する第3取得ステップと、

前記学習済みモデルを用いて前記識別データを識別する識別ステップと、

前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々の類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する1以上の学習データを抽出する抽出ステップと、

30

前記識別データ、及び前記1以上の学習データに基づいて、誤ラベルの有無、及び、前記識別データと前記1以上の学習データとのいずれに前記誤ラベルが付与されているかを判断する判断ステップと、

を含む、

処理方法。

【請求項14】

ラベルが付与された複数の学習データを取得する第1取得ステップと、

前記複数の学習データに基づき生成された学習済みモデルを取得する第2取得ステップと、

ラベルが付与された識別データを取得する第3取得ステップと、

前記学習済みモデルを用いて前記識別データを識別する識別ステップと、

前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々の類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する2以上の学習データを抽出する抽出ステップと、

40

前記識別データ、及び前記2以上の学習データに基づいて、誤ラベルの有無を判断する判断ステップと、

を含み、

前記判断ステップでは、

前記2以上の学習データから、前記類似度に関する指標が所定条件を満たすほどに前記

50

識別データに類似する特定の学習データを特定し、

前記特定の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと不一致であり、かつ、前記2以上の学習データにおける前記特定の学習データ以外の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと一致する場合、前記識別データよりも前記特定の学習データの方に前記誤ラベルが有る可能性が高いと判断する、

処理方法。

【請求項15】

ラベルが付与された複数の学習データを取得する第1取得ステップと、
前記複数の学習データに基づき生成された学習済みモデルを取得する第2取得ステップと、

10

ラベルが付与された識別データを取得する第3取得ステップと、
前記学習済みモデルを用いて前記識別データを識別する識別ステップと、
前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々との類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する2以上の学習データを抽出する抽出ステップと、

前記識別データ、及び前記2以上の学習データに基づいて、誤ラベルの有無を判断する判断ステップと、
を含み、

前記判断ステップでは、

20

前記2以上の学習データから、前記類似度に関する指標が所定条件を満たすほどに前記識別データに類似する特定の学習データを特定し、

前記特定の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと不一致であり、かつ、前記2以上の学習データにおける前記特定の学習データ以外の学習データに付与された前記ラベルが前記特定の学習データに付与された前記ラベルと一致する場合、前記特定の学習データよりも前記識別データの方に前記誤ラベルが有る可能性が高いと判断する、

処理方法。

【請求項16】

1以上のプロセッサに、請求項13～15のいずれか1項に記載の処理方法を実行させるためのプログラム。

30

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、一般に、処理システム、学習処理システム、処理方法、及びプログラムに関する。より詳細には本開示は、ラベルが付与されたデータに関する処理システム、当該処理システムを備える学習処理システム、処理方法、及びプログラムに関する。

【背景技術】

【0002】

特許文献1には、データ解析装置が開示されている。データ解析装置は、ラベル付き教師データをモデル構築用データとモデル検証用データとに分割し、モデル構築用データを用いて機械学習モデルを構築し、機械学習モデルをモデル検証用データに適用してサンプルを識別するという一連の処理を規定回数繰り返す。データ解析装置は、その識別結果であるラベルと元々データに付されていたラベルとが不一致であった誤識別の回数をサンプル毎に求め、その誤識別回数又はその誤識別の確率に基づいてサンプルがミスラベル状態であるか否かを判定する。これにより、教師データに含まれる、ミスラベル状態である可能性が高いサンプルを高い確度で検出することができる。

40

【0003】

特許文献1のデータ解析装置では、上記の一連の処理を規定回数繰り返す必要があり、ミスラベル(誤ラベル)の特定に長時間を要する可能性がある。

50

【先行技術文献】

【特許文献】

【0004】

【文献】特開2018-155522号公報

【発明の概要】

【0005】

本開示は上記事由に鑑みてなされ、誤ラベルの特定に要する時間の削減を図ることができる、処理システム、学習処理システム、処理方法、及びプログラムを提供することを目的とする。

【0006】

本開示の一態様の処理システムは、第1取得部と、第2取得部と、第3取得部と、識別部と、抽出部と、判断部と、を備える。前記第1取得部は、ラベルが付与された複数の学習データを取得する。前記第2取得部は、前記複数の学習データに基づき生成された学習済みモデルを取得する。前記第3取得部は、ラベルが付与された識別データを取得する。前記識別部は、前記学習済みモデルを用いて前記識別データを識別する。前記抽出部は、前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々の類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する1以上の学習データを抽出する。前記判断部は、前記識別データ、及び前記1以上の学習データに基づいて、誤ラベルの有無、及び、前記識別データと前記1以上の学習データとのいずれに前記誤ラベルが付与されているかを判断する。

10

本開示の一態様の処理システムは、第1取得部と、第2取得部と、第3取得部と、識別部と、抽出部と、判断部と、を備える。前記第1取得部は、ラベルが付与された複数の学習データを取得する。前記第2取得部は、前記複数の学習データに基づき生成された学習済みモデルを取得する。前記第3取得部は、ラベルが付与された識別データを取得する。前記識別部は、前記学習済みモデルを用いて前記識別データを識別する。前記抽出部は、前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々の類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する2以上の学習データを抽出する。前記判断部は、前記識別データ、及び前記1以上の学習データに基づいて、誤ラベルの有無を判断する。前記判断部は、前記2以上の学習データから、前記類似度に関する指標が所定条件を満たすほどに前記識別データに類似する特定の学習データを特定し、前記特定の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと不一致であり、かつ、前記2以上の学習データにおける前記特定の学習データ以外の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと一致する場合、前記識別データよりも前記特定の学習データの方に前記誤ラベルが有る可能性が高いと判断する。

20

30

本開示の一態様の処理システムは、第1取得部と、第2取得部と、第3取得部と、識別部と、抽出部と、判断部と、を備える。前記第1取得部は、ラベルが付与された複数の学習データを取得する。前記第2取得部は、前記複数の学習データに基づき生成された学習済みモデルを取得する。前記第3取得部は、ラベルが付与された識別データを取得する。前記識別部は、前記学習済みモデルを用いて前記識別データを識別する。前記抽出部は、前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々の類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する2以上の学習データを抽出する。前記判断部は、前記識別データ、及び前記1以上の学習データに基づいて、誤ラベルの有無を判断する。前記判断部は、前記2以上の学習データから、前記類似度に関する指標が所定条件を満たすほどに前記識別データに類似する特定の学習データを特定し、前記特定の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと不一致であり、かつ、前記2以上の学習データにおける前記特定の学習データ以外の学習データに付与された前記ラベルが前記特定の学習データに付与された前記ラベルと一致する場合、前記特定の学習データよりも前記識別データの方に前記誤ラベルが有る可能性が高いと判断する。

40

50

【0007】

本開示の一態様の学習処理システムは、上記の処理システムと、前記学習済みモデルを生成する学習システムと、を備える。

【0008】

本開示の一態様の処理方法は、第1取得ステップと、第2取得ステップと、第3取得ステップと、識別ステップと、抽出ステップと、判断ステップと、を含む。前記第1取得ステップでは、ラベルが付与された複数の学習データを取得する。前記第2取得ステップでは、前記複数の学習データに基づき生成された学習済みモデルを取得する。前記第3取得ステップでは、ラベルが付与された識別データを取得する。前記識別ステップでは、前記学習済みモデルを用いて前記識別データを識別する。前記抽出ステップでは、前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々との類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する1以上の学習データを抽出する。前記判断ステップは、前記識別データ、及び前記1以上の学習データに基づいて、誤ラベルの有無、及び、前記識別データと前記1以上の学習データとのいずれに前記誤ラベルが付与されているかを判断する。

10

本開示の一態様の処理方法は、第1取得ステップと、第2取得ステップと、第3取得ステップと、識別ステップと、抽出ステップと、判断ステップと、を含む。前記第1取得ステップでは、ラベルが付与された複数の学習データを取得する。前記第2取得ステップでは、前記複数の学習データに基づき生成された学習済みモデルを取得する。前記第3取得ステップでは、ラベルが付与された識別データを取得する。前記識別ステップでは、前記学習済みモデルを用いて前記識別データを識別する。前記抽出ステップでは、前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々との類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する2以上の学習データを抽出する。前記判断ステップでは、前記識別データ、及び前記2以上の学習データに基づいて、誤ラベルの有無を判断する。前記判断ステップでは、前記2以上の学習データから、前記類似度に関する指標が所定条件を満たすほどに前記識別データに類似する特定の学習データを特定し、前記特定の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと不一致であり、かつ、前記2以上の学習データにおける前記特定の学習データ以外の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと一致する場合、前記識別データよりも前記特定の学習データの方に前記誤ラベルが有る可能性が高いと判断する。

20

30

本開示の一態様の処理方法は、第1取得ステップと、第2取得ステップと、第3取得ステップと、識別ステップと、抽出ステップと、判断ステップと、を含む。前記第1取得ステップでは、ラベルが付与された複数の学習データを取得する。前記第2取得ステップでは、前記複数の学習データに基づき生成された学習済みモデルを取得する。前記第3取得ステップでは、ラベルが付与された識別データを取得する。前記識別ステップでは、前記学習済みモデルを用いて前記識別データを識別する。前記抽出ステップでは、前記学習済みモデルで適用される前記識別データと前記複数の学習データの各々との類似度に関する指標に基づき、前記複数の学習データから、前記識別データと類似する2以上の学習データを抽出する。前記判断ステップでは、前記識別データ、及び前記2以上の学習データに基づいて、誤ラベルの有無を判断する。前記判断ステップでは、前記2以上の学習データから、前記類似度に関する指標が所定条件を満たすほどに前記識別データに類似する特定の学習データを特定し、前記特定の学習データに付与された前記ラベルが前記識別データに付与された前記ラベルと不一致であり、かつ、前記2以上の学習データにおける前記特定の学習データ以外の学習データに付与された前記ラベルが前記特定の学習データに付与された前記ラベルと一致する場合、前記特定の学習データよりも前記識別データの方に前記誤ラベルが有る可能性が高いと判断する。

40

【0009】

本開示の一態様のプログラムは、1以上のプロセッサに、上記の処理方法を実行させるためのプログラムである。

50

【図面の簡単な説明】

【0010】

【図1】図1は、一実施形態に係る処理システムを備える学習処理システム全体の概略ブロック構成図である。

【図2】図2A及び図2Bは、同上の処理システムにおける動作例1及び動作例2を説明するための説明図である。

【図3】図3は、同上の学習処理システムの動作を説明するためのフローチャートである。

【図4】図4は、同上の処理システムにおける動作例3を説明するための説明図である。

【図5】図5は、同上の処理システムにおける動作例4を説明するための説明図である。

【図6】図6は、同上の処理システムにおける動作例5を説明するための説明図である。

10

【発明を実施するための形態】

【0011】

(1) 概要

以下の実施形態において説明する各図は、模式的な図であり、各図中の各構成要素の大きさ及び厚さそれぞれの比が、必ずしも実際の寸法比を反映しているとは限らない。

【0012】

本実施形態に係る処理システム1は、図1に示すように、第1取得部11と、第2取得部12と、第3取得部13と、識別部14と、抽出部15とを備える。

【0013】

第1取得部11は、ラベルが付与された複数の学習データD2を取得する。第2取得部12は、複数の学習データD2に基づき生成された学習済みモデルM1を取得する。

20

【0014】

ここでいう学習データD2は、一例として画像データである。学習データD2は、例えば撮像装置4(図1参照)で撮像された画像データである。しかし、画像データは、CG等の加工されたデータでもよい。またここでは画像データは、静止画であることを想定するが、動画又はコマ送りの1コマ1コマのデータでもよい。学習データD2は、その画像データ内に写るオブジェクト5(図2A及び図2B参照:被写体)に関する学習済みモデルM1を生成するためのデータである。つまり、学習データD2は、モデルを機械学習するために用いられる学習用データである。本開示でいう「モデル」は、識別対象(オブジェクト5)に関する入力データが入力されると、識別対象がどのような状態にあるかを推定し、推定結果(識別結果)を出力するプログラムである。「学習済みモデル」は、学習用データを用いた機械学習が完了したモデルをいう。また「学習データ(セット)」は、モデルに入力される入力データ(画像データ)と、入力データに付与されたラベルと、を組み合わせたデータセットであり、いわゆる教師データである。つまり、本実施形態では、学習済みモデルM1は、教師あり学習による機械学習が完了したモデルである。

30

【0015】

なお、本開示において、「画像データ内に写るオブジェクト5」とは、「画像データによって表される画像内に写るオブジェクト5」という意味を含む。

【0016】

本実施形態では一例として、学習済みモデルM1は、深層学習(ディープラーニング)を適用して複数の学習データD2に基づき生成されたモデルである。

40

【0017】

本実施形態では一例として、識別対象であるオブジェクト5は、図2A及び図2Bに示すように、電池である。つまり、学習データD2は、電池の画像(画像データ)である。したがって、学習済みモデルM1は、電池の外観の様子を推定し、推定結果を出力する。具体的には、学習済みモデルM1は、推定結果として、電池の外観が、良(OK)であるか不良(NG)であるかを出力する、言い換えれば、電池の外観検査のために用いられる。以下では、説明を分かりやすくするために、複数の学習データD2の各々に付与されるラベルは、「OK」又は「NG」の二種類だけである場合を想定する。しかし、本開示でいう「ラベル」の種類は、「OK」、「NG」の二種類に限定されない。例えば「NG」

50

について、より詳細な内容（不良の種類等）を示すラベルが付与されてよい。

【0018】

上記の内容を言い換えると、処理システム1は、学習済みモデルM1を用いて、電池の外観の様子を推定し、推定結果を出力する。具体的には、処理システム1は、学習済みモデルM1を用いて、推定結果として、電池の外観が、良（OK）であるか不良（NG）であるかを出力する。

【0019】

本実施形態の第3取得部13は、ラベルが付与された識別データD1を取得する。本実施形態では、識別データD1は、学習データD2と同様に、一例として画像データであり、その画像データ内に写るオブジェクト5は電池である。識別データD1は、例えば機械学習が完了した学習済みモデルM1を更新する際に、再学習するために新たに入手された教師データである。より具体的には、識別データD1は、現存の学習データとは別に新しく追加する学習データ、又は現存の学習データを更新するために用いられる学習データとなる予定のデータである。識別データD1には、複数の学習データD2と同様に、「OK」又は「NG」が付与され得る。

10

【0020】

ところで、モデルの機械学習を行うためには、教師データ（識別データD1及び学習データD2）に対して、人がラベルを付ける作業（ラベリング）が発生する。しかし、人がラベルを付ける際には、単純な作業ミス、又は人による基準の曖昧さが発生し得る。その結果、ラベル付きの教師データには、適切ではないラベル（誤ラベル）が付与されたデータが含まれている可能性がある。誤ラベルは、新たに入手された識別データD1にも、学習済みモデルM1の生成に用いた学習データD2にも存在し得る。

20

【0021】

本開示において、誤ラベルとは、データに付与されたラベルであって、適切ではないラベルをいう。誤ラベルは、例えば、OKラベルが付与されるべきデータに実際に付与されたNGラベル、NGラベルが付与されるべきデータに実際に付与されたOKラベルをいう。

【0022】

本実施形態の処理システム1では、識別部14は、学習済みモデルM1を用いて識別データD1を識別する。抽出部15は、学習済みモデルM1で適用される識別データD1と複数の学習データD2との類似度に関する指標に基づき、複数の学習データD2から、識別データD1と類似する1以上の学習データD2を抽出する。ここでいう「学習済みモデルM1で適用される類似度に関する指標」は、例えば、深層学習における出力層の直前の全結合層における指標であり、本実施形態では、ユークリッド距離を用いている。つまり、比較する2つの画像から得られる画素値等の特徴量から「距離」が求められ、2つの画像の近さが推定される。類似度の指標となる「距離」は、類似度とは反比例となる。類似度の指標となる「距離」は、ユークリッド距離以外にも、マハラノビス距離、マンハッタン距離、チェビシェフ距離、又はミンコフスキー距離でもよい。また指標は、距離に限定されず、類似度、又は（相関）係数等でもよく、例えばn次元ベクトルの類似度、コサイン類似度、ピアソンの相関係数、偏差パターン類似度、ジャカード係数、ダイス係数、又はシン普森係数でもよい。

30

40

【0023】

要するに、類似する1以上の学習データD2は、学習済みモデルM1が入力データ（識別データD1）を分類する際に用いられる類似度の指標に基づき抽出される。抽出部15は、識別データD1と類似度が高い複数（例えば上位3個）の学習データD2を抽出する。

【0024】

このように類似する1以上の学習データD2が抽出されるので、識別データD1と類似する1以上の学習データD2とを1回でも確認するだけで、誤ラベルの有無を特定し得る。結果的に、誤ラベルの特定に要する時間の削減を図ることができる。

【0025】

また本実施形態に係る学習処理システム100は、図1に示すように、処理システム1

50

と、学習済みモデルM1を生成する学習システム2とを備える。したがって、誤ラベルの特定に要する時間の削減を図ることが可能な学習処理システム100を提供できる。

【0026】

また本実施形態に係る処理方法は、第1取得ステップと、第2取得ステップと、第3取得ステップと、識別ステップと、抽出ステップと、を含む。第1取得ステップでは、ラベルが付与された複数の学習データD2を取得する。第2取得ステップでは、複数の学習データD2に基づき生成された学習済みモデルM1を取得する。第3取得ステップでは、ラベルが付与された識別データD1を取得する。識別ステップでは、学習済みモデルM1を用いて識別データD1を識別する。抽出ステップでは、学習済みモデルM1で適用される識別データD1と複数の学習データD2の各々との類似度に関する指標に基づき、複数の学習データD2から、識別データD1と類似する1以上の学習データD2を抽出する。この構成によれば、誤ラベルの特定に要する時間の削減を図ることが可能な処理方法を提供できる。この処理方法は、コンピュータシステム（処理システム1）上で用いられる。つまり、この処理方法は、プログラムでも具現化可能である。本実施形態に係るプログラムは、本実施形態に係る処理方法を、1以上のプロセッサに実行させるためのプログラムである。

10

【0027】

（2）詳細

以下、本実施形態に係る処理システム1を備えた学習処理システム100、及びその周辺構成を含んだ全体のシステムについて、図1を参照しながら詳しく説明する。なお、周辺構成の少なくとも一部が、学習処理システム100の構成に含まれてもよい。

20

【0028】

（2.1）全体構成

学習処理システム100は、図1に示すように、処理システム1と、学習システム2とを備える。また学習処理システム100の周辺構成として、推定システム3と、1又は複数台の撮像装置4（図1では1台のみ図示）とが設けられている。

【0029】

処理システム1、学習システム2、及び推定システム3は、サーバ等から構築されることを想定する。ここでいう「サーバ」は、1台のサーバ装置から構成されることを想定する。つまり、処理システム1、学習システム2、及び推定システム3の主な機能が、1台のサーバ装置に設けられていることを想定する。

30

【0030】

ただし、「サーバ」は、複数台のサーバ装置から構成されてもよい。具体的には、処理システム1、学習システム2、及び推定システム3の機能が、それぞれ個別のサーバ装置に設けられてもよいし、これらのうちの2つのシステムが1台のサーバ装置に設けられてもよい。またそのようなサーバ装置が、例えばクラウド（クラウドコンピューティング）を構築してもよい。

【0031】

またサーバ装置は、電池の外観検査を実施する工場内に設置されてもよいし、工場の外部（例えば事業本部）に設置されてもよい。処理システム1、学習システム2、及び推定システム3の機能がそれぞれ個別のサーバ装置に設けられている場合、各サーバ装置は、他のサーバ装置と通信可能に接続されていることが望ましい。

40

【0032】

学習システム2は、オブジェクト5に関する学習済みモデルM1を生成するように構成される。学習システム2は、ラベル付きの複数の学習データD2（画像データ）に基づき学習済みモデルM1を生成する。ここでいう学習済みモデルM1は、例えばニューラルネットワークを用いたモデル、又は多層ニューラルネットワークを用いた深層学習（ディープラーニング）により生成されるモデルを含むことを想定する。ニューラルネットワークは、例えばCNN（Convolutional Neural Network：畳み込みニューラルネットワーク）、又はBNN（Bayesian Neural Network：ベイズニューラルネットワーク）等を

50

含み得る。学習済みモデルM1は、ASIC (Application Specific Integrated Circuit) 又はFPGA (Field-Programmable Gate Array) 等の集積回路に、学習済みのニューラルネットワークを実装することで実現されている。学習済みモデルM1は、ディープラーニングにより生成されるモデルに限定されない。学習済みモデルM1は、サポートベクターマシン、又は決定木等により生成されるモデルでもよい。

【0033】

複数の学習データD2の各々は、画像データに対して「OK (良品)」又は「NG (不良品)」を示すラベルを付与することで生成される。ラベルの付与に関する作業 (ラベリング) は、ユーザによって操作部19等のユーザインタフェースを介して学習処理システム100に対して行われる。学習システム2は、ラベル付きの複数の学習データD2を用いて、電池の良品、及び不良品を機械学習することにより、学習済みモデルM1を生成する。

10

【0034】

また学習システム2は、新たに取得したラベル付きの学習データを識別データD1として用いて再学習を行うことで、学習済みモデルM1の性能の向上を図ることができる。例えばオブジェクト5に新しい種類の不良が見つければ、学習システム2に、新しい不良に関する再学習を行わせることが可能である。

【0035】

学習システム2で生成された学習済みモデルM1は、格納部に格納 (記憶) される。学習済みモデルM1を格納する格納部は、EEPROM (Electrically Erasable Programmable Read-Only Memory) のような書き換え可能な不揮発性メモリを含む。

20

【0036】

処理システム1は、識別データD1と類似する学習データD2を抽出する抽出処理を実行して、教師データ (識別データD1及び学習データD2) に誤ラベルが存在するか否かの確認を行いやすくする機能を有している。以下では、処理システム1を備えた学習処理システム100を利用する者を単に「ユーザ」と呼ぶことがある。ユーザは、例えば、工場内で電池 (オブジェクト5) の製造工程を監視するオペレータ、又は管理責任者等に相当し得る。

【0037】

処理システム1は、図1に示すように、処理部10と、提示部17と、通信部18と、操作部19とを備える。処理システム1は、記憶部を更に備える。

30

【0038】

処理システム1の一部の機能は、サーバと通信可能な情報端末に分散的に設けられてもよい。本開示でいう「情報端末」は、パーソナルコンピュータ (ノートパソコン又は据置型のパソコン)、スマートフォンやタブレット端末等の携帯型の端末等を含み得る。ここでは、提示部17及び操作部19の機能が、ユーザが使用する情報端末に設けられている。情報端末には、サーバと通信するための専用のアプリケーションソフトが予めインストールされる。

【0039】

処理部10は、1以上のプロセッサ (マイクロプロセッサ) と1以上のメモリを含むコンピュータシステムにより実現され得る。つまり、1以上のプロセッサが1以上のメモリに記憶された1以上のプログラム (アプリケーション) を実行することで、処理部10として機能する。プログラムは、ここでは処理部10のメモリに予め記録されているが、インターネット等の電気通信回線を通じて、又はメモリカード等の非一時的な記録媒体に記録されて提供されてもよい。

40

【0040】

処理部10は、提示部17、通信部18、及び操作部19等に関する制御処理を実行する。処理部10の機能はサーバにあることを想定する。また処理部10は、識別処理、抽出処理、及び判断処理を実行する機能を有しており、図1に示すように、第1取得部11、第2取得部12、第3取得部13、識別部14、抽出部15、及び判断部16を有する

50

。第1取得部11、第2取得部12、第3取得部13、識別部14、抽出部15、及び判断部16の詳細については、次の欄で説明する。

【0041】

提示部17は、液晶ディスプレイ又は有機EL(Electro-Luminescence)ディスプレイを構成する。提示部17は、上述の通り、情報端末に設けられている。提示部17は、タッチパネル式のディスプレイでもよい。提示部17は、後述する判断部16による判断結果に関する情報(提示情報D4)を外部に提示する。提示部17は、提示情報D4以外にも、推定システム3の推定結果等の種々の情報を表示し得る。

【0042】

通信部18は、1又は複数台の撮像装置4と直接的に、或いはユーザの情報端末又は生産管理システムの機能を有した別サーバ等を介して間接的に通信するための通信インタフェースである。通信部18の機能は、処理部10と同じサーバにあることを想定するが、例えば情報端末に設けられてもよい。通信部18は、撮像装置4又は別サーバ等から、識別データD1及び学習データD2を受信する。

10

【0043】

識別データD1及び学習データD2はいずれも、一例として撮像装置4で撮像された画像データにラベル(ここでは「OK」又は「NG」)が付与されたデータであり、画像データは、オブジェクト5を示す画素領域を含む。またオブジェクト5は、上述の通り電池であり、識別データD1及び学習データD2はいずれも電池の外観を示す画素領域を含むデータである。撮像装置4は、例えばラインセンサカメラを含む。

20

【0044】

撮像装置4で撮像されたオブジェクト5に関する大量の画像データの中からの、教師データ(学習データD2)に適用する画像データの選別は、例えば、ユーザからの指示に応じて行われる。画像データの選別作業、及び画像データへのラベリングを支援する機能が、学習処理システム100に設けられている。例えば、学習処理システム100は、選別及びラベリングに関する指示を受け付けるユーザインタフェース(例えば操作部19)を備える。

【0045】

操作部19は、マウス、キーボード、及びポインティングデバイス等を含む。操作部19は、上述の通り、例えばユーザが使用する情報端末に設けられている。提示部17が、情報端末のタッチパネル式のディスプレイである場合、操作部19の機能を兼ねてもよい。

30

【0046】

推定システム3は、学習システム2で生成された学習済みモデルM1を用いて、入力される対象画像データD3に関する推定を行う(推論フェーズ)。推定システム3は、1又は複数台の撮像装置4と直接的に、或いはユーザの情報端末又は生産管理システムの機能を有した別サーバ等を介して間接的に通信可能に構成される。推定システム3は、実際に製造工程を経た電池(製品又は半製品)が撮像装置4で撮像された対象画像データD3を受信して、電池の外観検査を実行する。

【0047】

推定システム3は、学習済みモデルM1を用いて、対象画像データD3に写るオブジェクト5が「良品」又は「不良品」のどちらであるかを推定する。推定システム3は、対象画像データD3に対する識別結果(推定結果)を、ユーザが利用する情報端末、又は生産管理システム等に出力する。ユーザは、情報端末を通じて、推定結果を確認できる。また生産管理システムが、推定結果を取得して、「不良品」と推定された電池については、次の工程に搬送される前に、破棄するように設備を制御してもよい。

40

【0048】

後述する処理システム1の識別部14の機能は、推定システム3の機能と実質的に同等である。

【0049】

(2.2)誤ラベル検知

50

処理部 10 は、識別処理、抽出処理、及び判断処理を実行して誤ラベルを検知する機能を有する。具体的には、処理部 10 は、図 1 に示すように、第 1 取得部 11、第 2 取得部 12、第 3 取得部 13、識別部 14、抽出部 15、及び判断部 16 を有する。

【0050】

第 1 取得部 11 は、ラベル付きの複数の学習データ D2 を取得するように構成される。第 1 取得部 11 は、ユーザにより操作部 19 を介して実行された操作入力に応じて、学習済みモデル M1 の生成に用いられた全てのラベル付きの複数の学習データ D2 を、例えばこれらを格納する格納部から取得する。提示部 17 は、第 1 取得部 11 が取得したラベル付きの学習データ D2 をユーザが閲覧できるように、画面表示可能である。

【0051】

第 2 取得部 12 は、複数の学習データ D2 に基づき学習システム 2 で生成された学習済みモデル M1 を取得するように構成される。第 2 取得部 12 は、ユーザにより操作部 19 を介して実行された操作入力に応じて、学習済みモデル M1 を、例えばこれを格納する格納部から取得する。

【0052】

第 3 取得部 13 は、ラベル付きの識別データ D1 を取得するように構成される。第 3 取得部 13 は、ユーザにより操作部 19 を介して実行された操作入力に応じて、新たに用意されたラベル付きの識別データ D1 を、例えばこれを格納する格納部から取得する。提示部 17 は、第 3 取得部 13 が取得したラベル付きの識別データ D1 をユーザが閲覧できるように、画面表示可能である。

【0053】

識別部 14 は、学習済みモデル M1 を用いて識別データ D1 を識別するように構成される（識別処理）。識別部 14 は、第 2 取得部 12 が取得した学習済みモデル M1 を用いて、識別データ D1 内に写るオブジェクト 5（電池）が、OK 又は NG のいずれであるかの識別を行わせる。つまり、識別部 14 は、推定システム 3 で言えば対象画像データ D3（入力データ）のように、学習済みモデル M1 に、識別データ D1 が OK か NG かを分類（識別）させる。後述の通り、識別処理による結果は、識別データ D1 に実際付与されていたラベルと比較される。

【0054】

抽出部 15 は、学習済みモデル M1 で適用される識別データ D1 と複数の学習データ D2 との類似度に関する指標に基づき、複数の学習データ D2 から、識別データ D1 と類似する 1 以上の学習データ D2 を抽出するように構成される（抽出処理）。以下では、抽出した 1 以上の学習データ D2 を、「類似データ D2 1」と呼ぶことがある。ここでは、抽出部 15 は、深層学習における出力層の直前の全結合層の情報に基づき、類似データ D2 1（学習データ D2）を抽出する。抽出部 15 は、識別データ D1 の画像から得られる画素値等に関する特徴量と、各学習データ D2 の画像から得られる画素値等に関する特徴量とから、類似度の指標（一例としてユークリッド距離）を求め、これにより画像同士の近さを推定する。以下では、類似度の指標を、単に「距離」と呼ぶ。抽出部 15 は、上記指標を求めて、識別データ D1 と各学習データ D2 との類似度を推定する。

【0055】

識別データ D1 に対する類似データ D2 1 の距離が小さいほど、その類似データ D2 1 が識別データ D1 に近い画像であることを意味する。言い換えると、学習済みモデル M1 は、全結合層において、入力データから得られる特徴量と各学習データ D2 から得られる特徴量との距離を比較する。つまり、抽出部 15 は、学習済みモデル M1 を用いて、学習済みモデル M1 の全結合層において、入力データから得られる特徴量と各学習データ D2 から得られる特徴量との距離を比較する。その結果、入力データは、学習済みモデル M1 により、入力データと距離の小さい学習データ D2 のラベルに基づき、良品（OK）である可能性が高い、又は不良品（NG）である可能性が高いという結果に分類され、その分類結果が出力層より出力される。

【0056】

10

20

30

40

50

このように、抽出部 15 は、識別データ D 1 と各学習データ D 2 との距離に基づき、複数の学習データ D 2 から、識別データ D 1 と類似性の高い類似データ D 2 1 を抽出する。例えば、抽出部 15 は、上記距離が、予め定められた特定の閾値以下である学習データ D 2 を、類似データ D 2 1 として抽出する。或いは、抽出部 15 は、複数の学習データ D 2 から、類似度の高い（距離が小さい）上位の N 個（N は自然数）の学習データ D 2 を、類似データ D 2 1 として抽出してもよい。特定の閾値や N 個（個数）は、ユーザにより任意に設定可能である。本実施形態では、処理システム 1 は、特定の閾値や N 個（個数）に関する設定情報を、ユーザにより操作部 19 を介して受け付け可能に構成される。設定情報は、処理部 10 のメモリ等に保存される。以下では、識別データ D 1 に対する距離が小さい上位 3 個の類似データ D 2 1 が抽出されるものとする。

10

【0057】

判断部 16 は、識別データ D 1、及び 1 以上の学習データ D 2 に基づいて、誤ラベルの有無を判断するように構成される（判断処理）。本実施形態では、処理部 10 は、特定の条件を満たしたときに、判断部 16 に判断処理を実行させる。特定の条件は、識別処理による識別結果と識別データ D 1 のラベルとが不一致であること、である。言い換えると、判断部 16 は、識別部 14 による識別データ D 1 の識別結果と、識別データ D 1 に付与されたラベルとが不一致の場合に、誤ラベルの有無の判断を実行する。このように特定の条件を満たす場合だけ、判断処理が実施されるため、無用に判断処理が実施される可能性が低減され、処理負荷の低減を図れる。また結果的に、誤ラベルのデータ特定に要する時間を更に削減できる。ここでは、上述した抽出処理も、上記特定の条件が満たされた場合に実行されるため、処理負荷をより低減できる。

20

【0058】

要するに、判断部 16 は、識別データ D 1、及び 1 以上の類似データ D 2 1（学習データ D 2）に基づいて、識別データ D 1 に付与されたラベル及び 1 以上の類似データ D 2 1 に付与されたラベルについて誤ラベルの有無を判断する。なお、本開示において、「識別データ D 1 のラベル」とは、識別データ D 1 に付与されたラベルといい、「学習データ D 2 のラベル」とは、学習データ D 2 に付与されたラベルをいう。

【0059】

なお、判断部 16 は、識別部 14 による識別データ D 1 の識別結果と、識別データ D 1 に付与されたラベルとが一致の場合、誤ラベルの有無の判断を実行しない。

30

【0060】

また本実施形態では、判断部 16 は、識別データ D 1 のラベルと 1 以上の類似データ D 2 1（学習データ D 2）のラベル、及び、識別データ D 1 に対する 1 以上の類似データ D 2 1（学習データ D 2）の類似度に関する指標の、少なくとも一方に基づき、誤ラベルの有無を判断する。次の欄の「（2.3）動作」では、動作例 1 で「ラベル」に基づき誤ラベルの有無を判断するケースを説明し、動作例 2 で「ラベル」と「類似度の指標」の両方に基づき誤ラベルの有無を判断するケースを説明する。動作例 1 及び 2 ではいずれも、誤ラベルが識別データ D 1 に有る場合の例である。また本実施形態では、判断部 16 は、学習データ D 2 に誤ラベルが有ることを特定するための機能を更に有し、次の欄の「（2.3）動作」の動作例 3 で説明する。

40

【0061】

要するに、判断部 16 は、識別データ D 1 に付与されたラベルと 1 以上の類似データ D 2 1（学習データ D 2）に付与されたラベル、及び、識別データ D 1 と 1 以上の類似データ D 2 1（学習データ D 2）との類似度に関する指標、の少なくとも一方に基づき、誤ラベルの有無を判断する。

【0062】

処理システム 1 の記憶部は、種々の情報を記憶する。より詳細には、記憶部は、第 1 取得部 11 で取得された複数の学習データ D 2 と、第 2 取得部 12 で取得された学習済みモデル M 1 と、第 3 取得部 13 で取得された識別データ D 1 とを記憶する。また、記憶部は、抽出部 15 で抽出された 1 以上の類似データ D 2 1 を記憶する。さらに、記憶部は、判

50

断部 16 による判断結果を記憶する。

【0063】

(2.3) 動作

以下、処理システム 1 に関する動作について、動作例 1 ~ 5 にて説明する。各動作例における動作の順序は単なる一例であり、特に限定されない。

【0064】

<動作例 1：識別データに誤ラベル有り>

以下、動作例 1 について、図 2 A、図 2 B、及び図 3 を参照して説明する。

【0065】

処理システム 1 の処理部 10 は、第 1 取得部 11 ~ 第 3 取得部 13 にて、それぞれ、ラベル付きの複数の学習データ D2、学習済みモデル M1、及びラベル付きの識別データ D1 を取得する (図 3：S1 ~ S3、第 1 ~ 第 3 取得ステップ)。これらのデータの取得順に決まりはない。本動作例 (動作例 1) では、識別データ D1 には、「NG」のラベルが付与されていたとする (図 2 A 参照)。

10

【0066】

次に、処理部 10 は、識別部 14 にて、学習済みモデル M1 を用いて識別データ D1 を識別する (図 3：S4、識別ステップ)。ここでは、識別結果が「OK」だったとする (図 2 A 参照)。処理部 10 は、識別結果と識別データ D1 のラベルとを比較し、不一致であれば (図 3：S5 の Yes)、抽出処理及び判断処理に進む。一方、識別結果と識別データ D1 のラベルとが一致すれば (図 3：S5 の No)、処理部 10 は、抽出処理及び判断処理に進まずに、提示部 17 から、例えば「エラーなし」といったメッセージを提示させて、処理を終える。本動作例では、識別結果が「OK」で、ラベルが「NG」のため、抽出処理及び判断処理に進む。

20

【0067】

処理部 10 は、抽出部 15 にて、複数の学習データ D2 から類似データ D21 を抽出する (図 3：S7、抽出ステップ)。この例では、距離の小さい上位 3 個の類似データ D21 が抽出されている (図 2 A 及び図 2 B 参照)。またこの例では、3 個の類似データ D21 の距離 (識別データ D1 と類似データ D21 との間の距離) は、左から順に、0.79、0.81、0.83 であり、距離が 0 (ゼロ) に近いほど識別データ D1 に近い画像であると、学習済みモデル M1 により識別される。またこの例では、3 個の類似データ D21 のラベルは、全て「OK」である。

30

【0068】

次に処理部 10 は、判断部 16 にて、識別データ D1 と 3 個の類似データ D21 とに基づいて、誤ラベルの有無を判断する (図 3：S8)。本開示では、判断部 16 は、誤ラベル度合いを計算し、誤ラベル度合いが高ければ (例えば 90% 以上であれば)、識別データ D1 に誤ラベルが有る可能性が高いと判断する。具体的には、本動作例では、判断部 16 が、識別データ D1 のラベルと 1 以上の類似データ D21 (学習データ D2) のラベルとが不一致な割合 (誤ラベル度合い) に基づき、誤ラベルの有無を判断するように構成される。図 2 A の例では、識別データ D1 のラベルが「NG」に対して、3 個の類似データ D21 の全てのラベルが「OK」である。その結果、不一致割合は、100% である。したがって、本動作例では、判断部 16 は、教師データに誤ラベルが有ると判断し、特に、識別データ D1 に誤ラベルが有る可能性が高いと判断する。なお、不一致割合が 90% 未満のケースについては、後述する動作例 5 で説明する。

40

【0069】

処理部 10 は、提示部 17 から、判断部 16 の判断結果を含む提示情報 D4 を提示する (図 3：S9)。本動作例では、図 2 B に示すように、誤ラベルが有る可能性が高いと判断された識別データ D1 は、その画像上に「誤ラベル」という文字データを重ね、その画像の周囲を枠で囲むように提示される。つまり、判断結果が誤ラベルの有ることを示す場合、提示部 17 は、識別データ D1 と、1 以上の類似データ D21 (学習データ D2) とのどちらに誤ラベルが有るかを示す情報を提示する。ここでは提示部 17 の同じ画面上に

50

、識別データD1の画像とセットで、3個の類似データD21も参考用に提示される(図2B参照)。また識別データD1のラベルの情報と識別結果、類似データD21のラベルの情報と距離の情報も、画像と共に提示される。したがって、ユーザは、提示部17に提示された情報を確認すれば、識別データD1に付与されていた「NG」ラベルは誤りで、正しいラベルは「OK」であることを容易に理解できる。

【0070】

<動作例2：識別データに誤ラベル有り>

以下、動作例2について、動作例1の図2Bを参照しながら説明する。上述した動作例1と実質的に共通する動作については、詳細な説明を省略する場合がある。

【0071】

動作例1では、図3のS8の判断処理にて、ラベルに基づき、つまり誤ラベル度合いとしてラベルの不一致割合に基づき、誤ラベルの有無が判断される。本動作例(動作例2)では、判断部16は、識別データD1のラベルと1以上の類似データD21(学習データD2)のラベル、及び、1以上の類似データD21(学習データD2)の類似度に関する指標の両方に基づき、誤ラベルの有無を判断する。つまり、本動作例の判断方法は、動作例1で例示した判断方法と異なる。

【0072】

具体的には、判断部16は、誤ラベル度合いをFとすると、以下の式(1)から、誤ラベル度合いFを計算する。

【0073】

【数1】

$$F = \frac{1}{N} \sum_{i=0}^N p_i \quad \dots \text{式(1)}$$

【0074】

式(1)中のNは、類似データD21の個数(ここではN=3)である。Piは、類似データiのラベルと識別データD1のラベルとが一致する場合、0(ゼロ)とし、不一致の場合、以下の式(2)から計算される。ここではK=0.001とする。

【0075】

【数2】

$$P_i = e^{-K \times L_i} \quad \dots \text{式(2)}$$

【0076】

式(2)のPiは、距離i(Li)が小さいほど、1に近づく値となる。式(2)のPiが1に近い値ということは、類似データiと識別データD1は、ラベルが互いに不一致にも関わらず、それらの画像の類似度が高いことを意味する。したがって、誤ラベル度合いF×100(確率)は、100%に近くなるほど、判断部16は、誤ラベルが有り、特に、識別データD1に誤ラベルが有る可能性が高いと判断する。

【0077】

図2Bの例で言えば、3個の類似データD21の距離は、左から順に、0.79、0.81、0.83であり、これらのラベルは全て識別データD1のラベルと不一致であるため、式(2)から、類似データiのPiが計算される。実際に各距離を式(2)代入してF×100を求めると、識別データD1に誤ラベルが有る確率は、{(0.99921+0.99919+0.99917)/3}×100=99.9%となる。

【0078】

処理システム1は、ユーザからの操作部19等への操作入力によって、動作例1の「ラ

10

20

30

40

50

ベルの割合」による判断方法、及び本動作例の「ラベルと類似度の指標の両方」による判断方法のいずれかを選択できるように構成されてもよい。

【0079】

本動作例で説明したように、ラベル及び類似度の指標の両方に基づき、誤ラベルの有無を判断することで、動作例1のように不一致割合により誤ラベルの有無を判断する場合に比べて、誤ラベルの判断に関する信頼性が容易に向上する。特に、抽出された類似データD21間で距離のばらつきが大きい場合に、動作例1の不一致割合に比べて、精度がより高くなり得る。

【0080】

<動作例3：学習データに誤ラベル有り>

以下、動作例3について、図3及び図4を参照しながら説明する。上述した動作例1と実質的に共通する動作については、詳細な説明を省略する場合がある。

【0081】

動作例1及び2の説明で参照した図2Bでは、識別データD1に誤ラベルが有る一例を示していた。本動作例（動作例3）では、学習データD2に誤ラベルが有る一例について説明する。

【0082】

処理システム1の処理部10は、ラベル付きの複数の学習データD2、学習済みモデルM1、及びラベル付きの識別データD1を取得する（図3：S1～S3）。本動作例では、識別データD1には、「OK」のラベルが付与されている（図4参照）。

【0083】

次に、処理部10は、学習済みモデルM1を用いて識別データD1を識別する（図3：S4）。ここでは、識別結果が「NG」だったとする（図4参照）。処理部10は、識別結果と識別データD1のラベルとを比較する（図3：S5）。本動作例では、識別結果が「NG」で、ラベルが「OK」のため、抽出処理及び判断処理に進む。

【0084】

処理部10は、複数の学習データD2から類似データD21を抽出する（図3：S7）。この例では、3個の類似データD21の距離は、左から順に、0（ゼロ）、1.82、1.95である。またこの例では、3個の類似データD21のラベルは、左から順に、「NG」、「OK」、「OK」である。

【0085】

次に処理部10は、識別データD1と3個の類似データD21とに基づいて、誤ラベルの有無を判断する（図3：S8）。

【0086】

ここで本実施形態の判断部16は、上述の通り、学習データD2に誤ラベルが有ることを特定するための機能を更に有している。具体的には、判断部16は、1以上の類似データD21（学習データD2）から、類似度に関する指標が所定条件を満たすほどに（満たす程度に）識別データD1に類似する特定の学習データD22を特定する。判断部16は、特定の学習データD22のラベルが識別データD1のラベルと不一致であり、かつ、1以上の類似データD21における特定の学習データD22以外の学習データD23のラベルが識別データD1のラベルと一致する場合、識別データD1よりも特定の学習データD22の方に誤ラベルが有る可能性が高いと判断する。

【0087】

本実施形態では、類似度に関する指標が「距離」であるため、判断部16は、「距離が所定の距離（閾値）以下」という所定条件を満たす特定の学習データD22を特定する。ここでは一例として、所定の距離（閾値）を0.001とするが、所定の距離は特に限定されない。類似度に関する指標が、n次元ベクトルの類似度、又はコサイン類似度等の「類似度」である場合、判断部16は、「類似度が所定の類似度（閾値）以上」という所定条件を満たす特定の学習データD22を特定する。所定の距離（閾値）や所定の類似度（閾値）は、ユーザにより任意に設定可能である。処理システム1は、所定の距離（閾値）

10

20

30

40

50

や所定の類似度（閾値）に関する設定情報を、ユーザにより操作部 19 を介して受け付け可能に構成される。設定情報は、処理部 10 のメモリ等に保存される。

【0088】

図 4 の例では、3 個の類似データ D 2 1 のうち左端の類似データ D 2 1 の「距離」が所定の距離（0.001）以下であることから、判断部 16 は、左端の類似データ D 2 1 が、識別データ D 1 に非常に似ている特定の学習データ D 2 2 に該当すると判定する。この特定の学習データ D 2 2 のラベル（NG）は、識別データ D 1 のラベル（OK）と不一致であり、また特定の学習データ D 2 2 以外の 2 個の学習データ D 2 3 のラベル（OK）は、識別データ D 1 のラベル（OK）と一致する。そのため、判断部 16 は、識別データ D 1 よりも、この特定の学習データ D 2 2 の方に誤ラベルが有る可能性が高いと判断する。

10

【0089】

ここでは特定の学習データ D 2 2 の数が、特定の学習データ D 2 2 以外の、識別データ D 1 のラベルと一致する学習データ D 2 3 の数の $1/2$ 以下である場合に、判断部 16 は、この特定の学習データ D 2 2 の方に誤ラベルが有る可能性が高いと判断する。図 4 の例では、特定の学習データ D 2 2 の数が 1 個で、学習データ D 2 3 の数（2 個）の $1/2 = 1$ 個であるため、特定の学習データ D 2 2 に誤ラベルが有る可能性が高いと判断される。

【0090】

処理部 10 は、提示部 17 から、判断部 16 の判断結果を含む提示情報 D 4 を提示する（図 3：S9）。本動作例では、図 4 に示すように、誤ラベルが有る可能性が高いと判断された特定の学習データ D 2 2 は、その画像上に「誤ラベル」という文字データを重ね、その画像の周囲を枠で囲むように提示される。また識別データ D 1 のラベルの情報と識別結果、類似データ D 2 1 のラベルの情報と距離の情報も、画像と共に提示される。したがって、ユーザは、提示部 17 に提示された情報を確認すれば、特定の学習データ D 2 2 に付与されていた「NG」ラベルは誤りで、正しいラベルは「OK」であることを容易に理解できる。

20

【0091】

特定の学習データ D 2 2 の数が学習データ D 2 3 の数の $1/2$ より大きい場合、判断部 16 は、誤ラベルは無いと判断する。処理部 10 は、例えば「目視確認してください。」といったメッセージと共に、識別データ D 1 の画像と 3 個の類似データ D 2 1 の画像を、提示部 17 に提示させる。言い換えると、判断結果が誤ラベルの無いことを示す場合、提示部 17 は、識別データ D 1 と、1 以上の類似データ D 2 1（学習データ D 2）の両方を提示する。つまり、誤ラベルの有無について処理システム 1 で自動判断しにくい場合には、ユーザの目視確認を促す。

30

【0092】

<動作例 4：動作例 3 の派生>

以下、動作例 4 について、図 5 を参照しながら説明する。上述した動作例 1 と実質的に共通する動作については、詳細な説明を省略する場合がある。

【0093】

本動作例（動作例 4）では、上述した動作例 3 の派生であり、動作例 3 と同様に識別データ D 1 に非常に似ている特定の学習データ D 2 2 が存在する一方で、識別データ D 1 に誤ラベルが有る点で動作例 3 と相違する。

40

【0094】

図 5 の例では、識別データ D 1 について、識別結果が「OK」で、ラベルが「NG」である。図 5 の例では、図 4 と同様に、3 個の類似データ D 2 1 の距離は、左から順に、0（ゼロ）、1.82、1.95 である。ただし、図 5 の例では、図 4 と異なり、3 個の類似データ D 2 1 のラベルは全て「OK」である。

【0095】

本変形例でも、判断部 16 は、1 以上の類似データ D 2 1（学習データ D 2）から、類似度に関する指標が所定条件（ここでは距離が所定の距離（閾値）以下）を満たすほどに識別データ D 1 に類似する特定の学習データ D 2 2 を特定する。ここで、判断部 16 は、

50

特定の学習データD 2 2のラベルが識別データD 1のラベルと不一致であり、かつ、1以上の類似データD 2 1における特定の学習データD 2 2以外の学習データD 2 3のラベルが特定の学習データD 2 2のラベルと一致する場合、特定の学習データD 2 2よりも識別データD 1の方に誤ラベルが有る可能性が高いと判断する。

【0096】

図5の例では、3個の類似データD 2 1のうち左端の類似データD 2 1の「距離」が所定の距離(0.001)以下であることから、判断部16は、左端の類似データD 2 1が、識別データD 1に非常に似ている特定の学習データD 2 2に該当すると判定する。この特定の学習データD 2 2のラベル(OK)は、識別データD 1のラベル(NG)と不一致であり、また特定の学習データD 2 2以外の2個の学習データD 2 3のラベル(OK)は、特定の学習データD 2 2のラベル(OK)と一致する。そのため、判断部16は、この特定の学習データD 2 2よりも、識別データD 1の方に誤ラベルが有る可能性が高いと判断する。

10

【0097】

ここでは特定の学習データD 2 2のラベルと一致する学習データD 2 3の数が、特定の学習データD 2 2のラベルと不一致の学習データD 2 3の数より大きい場合に、判断部16は、識別データD 1の方に誤ラベルが有る可能性が高いと判断する。図5の例では、特定の学習データD 2 2のラベルと一致する学習データD 2 3が2個で、特定の学習データD 2 2のラベルと不一致の学習データD 2 3の数(0個)より大きいため、識別データD 1に誤ラベルが有る可能性が高いと判断される。

20

【0098】

本動作例では、図5に示すように、誤ラベルが有る可能性が高いと判断された識別データD 1は、その画像上に「誤ラベル」という文字データを重ね、その画像の周囲を枠で囲むように提示される。また識別データD 1のラベルの情報と識別結果、各類似データD 2 1のラベルの情報と距離の情報も、画像と共に提示される。したがって、ユーザは、提示部17に提示された情報を確認すれば、識別データD 1に付与されていた「NG」ラベルは誤りで、正しいラベルは「OK」であることを容易に理解できる。

【0099】

特定の学習データD 2 2のラベルと一致する学習データD 2 3の数が、特定の学習データD 2 2のラベルと不一致の学習データD 2 3の数以下の場合、判断部16は、誤ラベルは無いと判断する。処理部10は、例えば「目視確認してください。」といったメッセージと共に、識別データD 1の画像と3個の類似データD 2 1の画像を、提示部17に提示させる。言い換えると、判断結果が誤ラベルの無いことを示す場合、提示部17は、識別データD 1と、1以上の類似データD 2 1(学習データD 2)の両方を提示する。つまり、誤ラベルの有無について処理システム1で自動判断しにくい場合には、ユーザの目視確認を促す。

30

【0100】

<動作例5：類似データにOK、NG混在>

以下、動作例5について、図6を参照しながら説明する。上述した動作例1と実質的に共通する動作については、詳細な説明を省略する場合がある。

40

【0101】

動作例1及び2の説明で参照した図2Bでは、抽出された3個の類似データD 2 1のラベルが全てOKである。本動作例(動作例5)では、動作例3の説明で参照した図4と同様に、抽出された3個の類似データD 2 1のラベルに、OKとNGが混在した一例を示す図6を用いて説明する。ただし、動作例3の説明で参照した図4とは違って、図6では、距離が所定の距離(0.001)以下の、識別データD 1に非常に似た類似データD 2 1は無い。

【0102】

処理システム1の処理部10は、ラベル付きの複数の学習データD 2、学習済みモデルM 1、及びラベル付きの識別データD 1を取得する(図3：S 1～S 3)。本動作例では

50

、識別データD1には、「NG」のラベルが付与されている（図6参照）。

【0103】

次に、処理部10は、学習済みモデルM1を用いて識別データD1を識別する（図3：S4）。ここでは、識別結果が「OK」だったとする（図6参照）。処理部10は、識別結果と識別データD1のラベルとを比較する（図3：S5）。本動作例では、識別結果が「OK」で、ラベルが「NG」のため、抽出処理及び判断処理に進む。

【0104】

処理部10は、複数の学習データD2から類似データD21を抽出する（図3：S7）。この例では、3個の類似データD21の距離は、左から順に、1.86、1.93、2.01である。またこの例では、3個の類似データD21のラベルは、左から順に、「OK」、「OK」、「NG」である。要するに、図6の3個の類似データD21では、識別データD1に対する距離が互いに概ね同じにも関わらず、OKとNGのラベルが混在している。

10

【0105】

次に処理部10は、識別データD1と3個の類似データD21とに基づいて、誤ラベルの有無を判断する（図3：S8）。

【0106】

本動作例では、動作例1と同様に、例えば、判断部16が、識別データD1のラベルと3個の類似データD21のラベルとが不一致な割合（誤ラベル度合い）に基づき、誤ラベルの有無を判断するように構成される。図6の例では、識別データD1のラベルが「NG」に対して、3個の類似データD21のラベル中、不一致のラベルは2個である。その結果、不一致割合（誤ラベル度合い）は、約67%である。したがって、本動作例では、判断部16は、誤ラベル度合いが閾値（例えば90%）未満であるため、誤ラベルは無いと判断する。

20

【0107】

この場合、処理部10は、例えば「類似データにOKの画像とNGの画像とが混在。目視確認してください。」といったメッセージと共に、識別データD1の画像と3個の類似データD21の画像とを、提示部17に提示させる。言い換えると、判断結果が誤ラベルの無いことを示す場合、提示部17は、識別データD1と、1以上の類似データD21（学習データD2）との両方を提示する。つまり、動作例3と同様に、誤ラベルの有無について処理システム1で自動判断しにくい場合には、ユーザの目視確認を促す。

30

【0108】

<利点>

モデルの機械学習を行うためには、教師データ（識別データD1及び学習データD2）に対して、人によるラベリングが発生する。しかし、人がラベルを付ける際には、単純な作業ミス、又は人による基準の曖昧さが発生し得る。特にオブジェクト5の種類によっては、OKラベルを付けるべき画像とNGラベルを付けるべき画像とで、熟練度が低い人が一見するだけだと同じような画像に見える可能性がある。その結果、ラベル付きの教師データには、誤ラベルのデータが含まれている可能性がある。例えば、OKラベルを付けるべき画像にNGラベルが誤ラベルとして付けられたり、NGラベルを付けるべき画像にOKラベルが誤ラベルとして付けられたりする。誤ラベルは、新たに入手された識別データD1にも、学習済みモデルM1の生成に用いた多数の学習データD2にも存在し得る。

40

【0109】

動作例1～5で説明したように、本実施形態に係る処理システム1では、識別データD1と類似する1以上の類似データD21が（自動的に）抽出される。識別データD1と類似データD21とを、ユーザが提示部17を通じて1回でも目視確認するだけで、誤ラベルの有無を特定しやすい。したがって、処理システム1は、誤ラベルの特定に関する作業を支援できる。結果的に、誤ラベルの特定に要する時間の削減を図ることができる。また誤ラベルが解消された教師データを用いて学習を行うので、学習済みモデルM1を用いた推論フェーズの精度も向上する。

50

【 0 1 1 0 】

処理システム 1 には、誤ラベルを自動的に検知する機能、つまり誤ラベルの有無を判断する判断部 1 6 が設けられているが、判断部 1 6 は、処理システム 1 の必須の構成要素ではない。ただし、本実施形態のように、判断部 1 6 が設けられていることで、誤ラベルの特定に要する時間を更に削減できる。

【 0 1 1 1 】

また処理システム 1 には、判断部 1 6 による判断結果に関する情報（提示情報 D 4 ）を外部に提示する提示部 1 7 が設けられているため、ユーザによる目視確認がより容易となる。

【 0 1 1 2 】

さらに判断部 1 6 による判断結果が誤ラベルの有ることを示す場合、提示部 1 7 は、識別データ D 1 と、類似データ D 2 1 とのどちらに誤ラベルが有るかを示す情報を提示する。そのため、ユーザは、どちらのデータに誤ラベルが有るかを容易に目視確認できる。

【 0 1 1 3 】

特に、判断結果が誤ラベルの無いことを示す場合、提示部 1 7 は、識別データ D 1 と、類似データ D 2 1 の両方を提示する。そのため、ユーザが、識別データ D 1 と類似データ D 2 1 の両方の目視確認を行いやすくなり、結果的に、実際にはどちらかのデータに誤ラベルが有る場合にその誤ラベルを見つけやすくなる。また誤ラベルとは別の不具合（例えば、学習不足又は過学習等）が有る場合も見つけやすくなる。

【 0 1 1 4 】

例えばユーザが提示部 1 7 を確認すると、類似度が高い（距離が小さい）上位の類似データ D 2 1 が、識別データ D 1 とあまり類似していない場合には、ユーザは、学習済みモデル M 1 の学習不足の可能性が高いと判断できる。

【 0 1 1 5 】

なお、処理システム 1 の処理部 1 0 が、抽出した上位の類似データ D 2 1 の距離から、学習不足か否かを自動的に判断してもよい。図 3 で言えば、例えば抽出処理（S 7 ）の後に、抽出した各類似データ D 2 1 の距離をチェックし、距離が一定値以上であれば、学習不足と判断し、次の判断処理（S 8 ）に進まずに、「学習不足」のメッセージを提示部 1 7 から提示して処理を終えてもよい。

【 0 1 1 6 】

（ 3 ）変形例

上記実施形態は、本開示の様々な実施形態の一つに過ぎない。上記実施形態は、本開示の目的を達成できれば、設計等に応じて種々の変更が可能である。また、上記実施形態に係る処理システム 1 と同様の機能は、処理方法、コンピュータプログラム、又はコンピュータプログラムを記録した非一時的記録媒体等で具現化されてもよい。

【 0 1 1 7 】

以下、上記実施形態の変形例を列挙する。以下に説明する変形例は、適宜組み合わせて適用可能である。以下では、上記実施形態を「基本例」と呼ぶこともある。

【 0 1 1 8 】

本開示における処理システム 1 は、コンピュータシステムを含んでいる。コンピュータシステムは、ハードウェアとしてのプロセッサ及びメモリを主構成とする。コンピュータシステムのメモリに記録されたプログラムをプロセッサが実行することによって、本開示における処理システム 1 としての機能が実現される。プログラムは、コンピュータシステムのメモリに予め記録されてもよく、電気通信回線を通じて提供されてもよく、コンピュータシステムで読み取り可能なメモリカード、光学ディスク、ハードディスクドライブ等の非一時的記録媒体に記録されて提供されてもよい。コンピュータシステムのプロセッサは、半導体集積回路（IC）又は大規模集積回路（LSI）を含む 1 ないし複数の電子回路で構成される。ここでいう IC 又は LSI 等の集積回路は、集積の度合いによって呼び方が異なっており、システム LSI、VLSI（Very Large Scale Integration）、又は ULSI（Ultra Large Scale Integration）と呼ばれる集積回路を含む。さらに、L

10

20

30

40

50

S Iの製造後にプログラムされる、F P G A (Field-Programmable Gate Array)、又はL S I内部の接合関係の再構成若しくはL S I内部の回路区画の再構成が可能な論理デバイスについても、プロセッサとして採用することができる。複数の電子回路は、1つのチップに集約されていてもよいし、複数のチップに分散して設けられていてもよい。複数のチップは、1つの装置に集約されていてもよいし、複数の装置に分散して設けられていてもよい。ここでいうコンピュータシステムは、1以上のプロセッサ及び1以上のメモリを有するマイクロコントローラを含む。したがって、マイクロコントローラについても、半導体集積回路又は大規模集積回路を含む1ないし複数の電子回路で構成される。

【0119】

また、処理システム1における複数の機能が、1つのハウジング内に集約されていることは必須の構成ではない。例えば、処理システム1の構成要素は、複数のハウジングに分散して設けられていてもよい。

10

【0120】

反対に、処理システム1における複数の機能が、1つのハウジング内に集約されてもよい。さらに、処理システム1の少なくとも一部の機能、例えば、処理システム1の一部の機能がクラウド(クラウドコンピューティング)等によって実現されてもよい。

【0121】

基本例では、識別データD1は、再学習するために新たに入手された教師データである。しかし、識別データD1は、学習済みモデルM1の生成に使用した学習データD2でもよい。例えば学習済みモデルM1を生成した後に、学習済みモデルM1の精度が100%とは言えない場合がある。そのような場合には、学習済みモデルM1の正確性を確認、評価するために、学習済みモデルM1の生成に用いた学習データD2の一部、又は全部を、識別データD1として、処理システム1に入力してもよい。

20

【0122】

識別データD1は、モデルの機械学習を行う際に用意される複数の教師データの一部であってもよい。つまり、モデルを学習する際に用意される複数の教師データは、複数の学習データD2と識別データD1とに分割される。この場合、処理システム1は、複数の教師データを分割して学習済みモデルM1を評価する交差検証を行うと共に、識別データD1に付与されたラベル、及び複数の学習データD2の各々に付与されたラベルについて誤ラベルの有無を検知することが可能である。

30

【0123】

また、処理システム1は、複数の教師データに対して学習データD2と識別データD1との分割を複数回行い、k-分割交差検証を行い、さらに、識別データD1に付与されたラベル、及び複数の学習データD2の各々に付与されたラベルについて誤ラベルの有無を検知してもよい。

【0124】

基本例では、誤ラベルが識別データD1(又は類似データD21)に有ると判断された場合であっても、提示部17は、識別データD1と類似データD21との両方を提示する。しかし、提示部17は、誤ラベルが有ると判断されたデータのみを提示してもよい。

【0125】

撮像装置4は、ラインセンサカメラに限定されず、エリアセンサカメラを含んでもよい。

40

【0126】

基本例では、教師データ(識別データD1及び学習データD2)は、画像データにラベルが付与されたデータである。しかし、教師データは、画像データに限定されず、テキストデータ、又は音声データにラベルが付与されたデータでもよい。つまり、学習済みモデルM1は、画像の識別(画像認識)に限定されず、例えばテキストの識別(テキスト認識)、又は音声の識別(音声認識)に適用されてもよい。

【0127】

基本例では、学習システム2で生成される学習済みモデルM1は、ディープラーニングにより生成されるモデルである。ただし、学習済みモデルM1は、ディープラーニングに

50

限定されない。学習済みモデルM1は、いかなるタイプの人工知能又はシステムとして実装されてもよい。

【0128】

基本例では、機械学習のアルゴリズムは、ニューラルネットワーク（ディープラーニングを含む）である。ただし、機械学習のアルゴリズムは、ニューラルネットワークに限定されず、他の教師あり学習のアルゴリズムであってもよい。機械学習のアルゴリズムは、例えば、線形回帰（Linear Regression）、ロジスティック回帰（Logistic Regression）、サポートベクターマシン（Support Vector Machine：SVM）、決定木（Decision Tree）、ランダムフォレスト（Random Forest）、勾配ブースティング（Gradient Boosting）、ナイーブベイズ（Naive Bayes）分類器、又はk近傍法（k-Nearest Neighbors：k-NN）であってもよい。

10

【0129】

（4）まとめ

以上説明したように、第1の態様に係る処理システム（1）は、第1取得部（11）と、第2取得部（12）と、第3取得部（13）と、識別部（14）と、抽出部（15）と、を備える。第1取得部（11）は、ラベルが付与された複数の学習データ（D2）を取得する。第2取得部（12）は、複数の学習データ（D2）に基づき生成された学習済みモデル（M1）を取得する。第3取得部（13）は、ラベルが付与された識別データ（D1）を取得する。識別部（14）は、学習済みモデル（M1）を用いて識別データ（D1）を識別する。抽出部（15）は、識別データ（D1）と複数の学習データ（D2）の各々との類似度に関する指標に基づき、複数の学習データ（D2）から、識別データ（D1）と類似する1以上の学習データ（類似データD21）を抽出する。上記指標は、学習済みモデル（M1）で適用される指標である。

20

【0130】

この態様によれば、識別データ（D1）と類似する1以上の学習データ（D2）が抽出されるので、識別データ（D1）と、識別データ（D1）と類似する1以上の学習データ（類似データD21）とを（例えば1回）確認するだけで、誤ラベルの有無を特定し得る。結果的に、誤ラベルの特定に要する時間の削減を図ることができる。

【0131】

第2の態様に係る処理システム（1）は、第1の態様において、識別データ（D1）、及び1以上の学習データ（類似データD21）に基づいて、誤ラベルの有無を判断する判断部（16）を更に備える。

30

【0132】

この態様によれば、誤ラベルの有無が自動的に判断されるので、誤ラベルの特定に要する時間を更に削減できる。

【0133】

第3の態様に係る処理システム（1）は、第2の態様において、判断部（16）による判断結果に関する情報を外部に提示する提示部（17）を更に備える。

【0134】

この態様によれば、判断部（16）による判断結果に関する情報が提示されるので、ユーザによる目視確認が容易となる。

40

【0135】

第4の態様に係る処理システム（1）に関して、第3の態様において、判断結果が誤ラベルの有ることを示す場合、提示部（17）は、識別データ（D1）と、1以上の学習データ（類似データD21）とのいずれに誤ラベルが有るかを示す情報を提示する。

【0136】

この態様によれば、識別データ（D1）と1以上の学習データ（類似データD21）とのどちらに誤ラベルが有るかを容易に目視確認できる。

【0137】

第5の態様に係る処理システム（1）に関して、第3又は第4の態様において、判断結

50

果が誤ラベルの無いことを示す場合、提示部（１７）は、識別データ（Ｄ１）と、１以上の学習データ（類似データＤ２１）との両方を提示する。

【０１３８】

この態様によれば、ユーザが識別データ（Ｄ１）と１以上の学習データ（類似データＤ２１）との両方の目視確認を行いやすくなり、結果的に、実際にはどちらかのデータに誤ラベルが有る場合にその誤ラベルを見つけやすくなる。また誤ラベルとは別の不具合が有る場合も見つけやすくなる。

【０１３９】

第６の態様に係る処理システム（１）に関して、第２～第５の態様のいずれか１つにおいて、判断部（１６）は、識別部（１４）による識別データ（Ｄ１）の識別結果と識別データ（Ｄ１）に付与されたラベルとが不一致の場合に、誤ラベルの有無の判断を実行する。

10

【０１４０】

この態様によれば、処理負荷の低減を図れる。また誤ラベルの特定に要する時間を更に削減できる。

【０１４１】

第７の態様に係る処理システム（１）に関して、第２～第６の態様のいずれか１つにおいて、判断部（１６）は、識別データ（Ｄ１）に付与されたラベルと１以上の学習データ（類似データＤ２１）に付与されたラベル、及び、識別データ（Ｄ１）と１以上の学習データ（類似データＤ２１）との類似度に関する指標、の少なくとも一方に基づき、誤ラベルの有無を判断する。

20

【０１４２】

この態様によれば、誤ラベルの判断に関する信頼性が向上する。

【０１４３】

第８の態様に係る処理システム（１）に関して、第７の態様において、判断部（１６）は、識別データ（Ｄ１）に付与されたラベルと１以上の学習データ（類似データＤ２１）に付与されたラベルとが不一致な割合に基づき、誤ラベルの有無を判断する。

【０１４４】

この態様によれば、誤ラベルの判断に関する信頼性が容易に向上する。

【０１４５】

第９の態様に係る処理システム（１）に関して、第７の態様において、判断部（１６）は、識別データ（Ｄ１）に付与されたラベルと１以上の学習データ（類似データＤ２１）に付与されたラベル、及び、１以上の学習データ（類似データＤ２１）の類似度に関する指標の両方に基づき、誤ラベルの有無を判断する。

30

【０１４６】

この態様によれば、誤ラベルの判断に関する信頼性が更に向上する。

【０１４７】

第１０の態様に係る処理システム（１）に関して、第９の態様において、抽出部（１５）は、複数の学習データ（Ｄ２）から、１以上の学習データ（類似データＤ２１）として、２以上の学習データ（類似データＤ２１）を抽出する。判断部（１６）は、２以上の学習データ（類似データＤ２１）から、類似度に関する指標が所定条件を満たすほどに識別データ（Ｄ１）に類似する特定の学習データ（Ｄ２２）を特定する。判断部（１６）は、特定の学習データ（Ｄ２２）に付与されたラベルが識別データ（Ｄ１）に付与されたラベルと不一致であり、かつ、２以上の学習データ（類似データＤ２１）における特定の学習データ（Ｄ２２）以外の学習データ（Ｄ２３）に付与されたラベルが識別データ（Ｄ１）に付与されたラベルと一致する場合、識別データ（Ｄ１）よりも特定の学習データ（Ｄ２２）の方に誤ラベルが有る可能性が高いと判断する。

40

【０１４８】

この態様によれば、誤ラベルの判断に関する信頼性が更に向上する。

【０１４９】

第１１の態様に係る処理システム（１）に関して、第９の態様において、抽出部（１５

50

は、複数の学習データ(D 2)から、1以上の学習データ(類似データD 2 1)として、2以上の学習データ(類似データD 2 1)を抽出する。判断部(16)は、2以上の学習データ(類似データD 2 1)から、類似度に関する指標が所定条件を満たすほどに識別データ(D 1)に類似する特定の学習データ(D 2 2)を特定する。判断部(16)は、特定の学習データ(D 2 2)に付与されたラベルが識別データ(D 1)に付与されたラベルと不一致であり、かつ、2以上の学習データ(類似データD 2 1)における特定の学習データ(D 2 2)以外の学習データ(D 2 3)に付与されたラベルが特定の学習データ(D 2 2)に付与されたラベルと一致する場合、特定の学習データ(D 2 2)よりも識別データ(D 1)の方に誤ラベルが有る可能性が高いと判断する。

【0150】

この態様によれば、誤ラベルの判断に関する信頼性が更に向上する。

【0151】

第12の態様に係る処理システム(1)に関して、第1~第11の態様のいずれか1つにおいて、学習済みモデル(M 1)は、深層学習を適用して複数の学習データ(D 2)に基づき生成されたモデルである。

【0152】

この態様によれば、学習済みモデル(M 1)の信頼性、及び誤ラベルの判断に関する信頼性が更に向上する。

【0153】

第13の態様に係る学習処理システム(100)は、第1~第12の態様のいずれか1つにおける処理システム(1)と、学習済みモデル(M 1)を生成する学習システム(2)と、を備える。

【0154】

この態様によれば、誤ラベルの特定に要する時間の削減を図ることが可能な学習処理システム(100)を提供できる。

【0155】

第14の態様に係る処理方法は、第1取得ステップと、第2取得ステップと、第3取得ステップと、識別ステップと、抽出ステップと、を含む。第1取得ステップでは、ラベルが付与された複数の学習データ(D 2)を取得する。第2取得ステップでは、複数の学習データ(D 2)に基づき生成された学習済みモデル(M 1)を取得する。第3取得ステップでは、ラベルが付与された識別データ(D 1)を取得する。識別ステップでは、学習済みモデル(M 1)を用いて識別データ(D 1)を識別する。抽出ステップでは、学習済みモデル(M 1)で適用される識別データ(D 1)と複数の学習データ(D 2)の各々の類似度に関する指標に基づき、複数の学習データ(D 2)から、識別データ(D 1)と類似する1以上の学習データ(類似データD 2 1)を抽出する。

【0156】

この態様によれば、誤ラベルの特定に要する時間の削減を図ることが可能な処理方法を提供できる。

【0157】

第15の態様に係るプログラムは、1以上のプロセッサに、第14の態様における処理方法を実行させるためのプログラムである。

【0158】

この態様によれば、誤ラベルの特定に要する時間の削減を図ることが可能な機能を提供できる。

【0159】

第16の態様に係る処理システム(1)に関して、第1~第12の態様のいずれか1つにおいて、抽出部(15)は、識別部(14)による識別データ(D 1)の識別結果と、識別データ(D 1)に付与されたラベルとが不一致の場合に、複数の学習データ(D 2)から1以上の学習データ(類似データD 2 1)を抽出する。

【0160】

10

20

30

40

50

第 2 ~ 1 2 の態様に係る構成については、処理システム (1) に必須の構成ではなく、適宜省略可能である。同様に、第 1 6 の態様に係る構成についても、処理システム (1) に必須の構成ではなく、適宜省略可能である。

【符号の説明】

【 0 1 6 1 】

1 0 0 学習処理システム

1 処理システム

1 1 第 1 取得部

1 2 第 2 取得部

1 3 第 3 取得部

1 4 識別部

1 5 抽出部

1 6 判断部

1 7 提示部

2 学習システム

D 1 識別データ

D 2 学習データ

D 2 1 1 以上の類似データ (1 以上の学習データ)

D 2 2 特定の学習データ

M 1 学習済みモデル

10

20

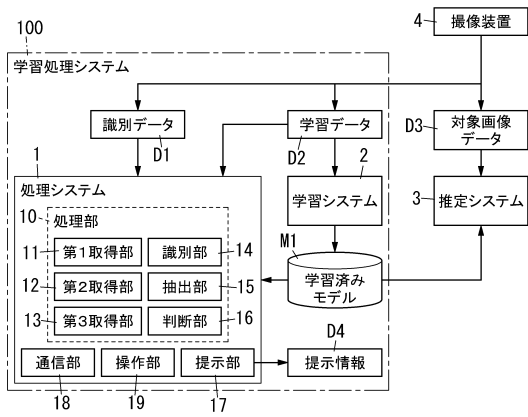
30

40

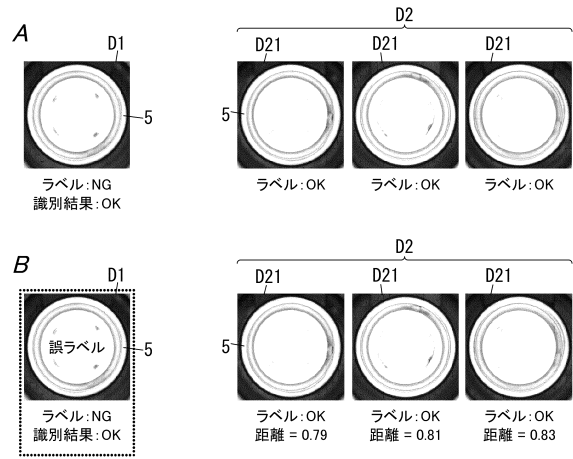
50

【図面】

【図 1】

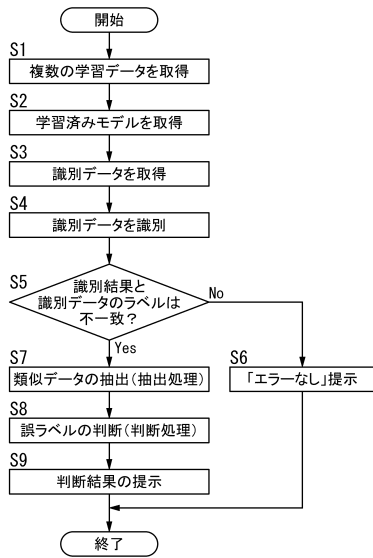


【図 2】

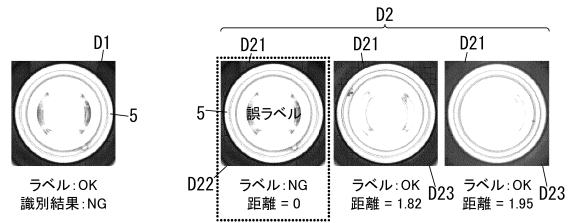


10

【図 3】



【図 4】



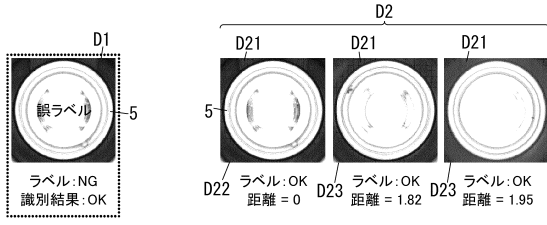
20

30

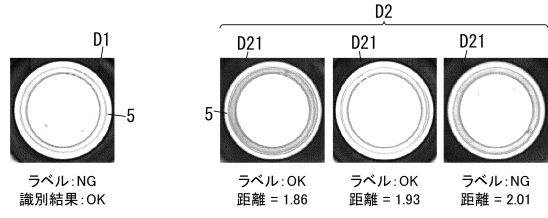
40

50

【 図 5 】



【 図 6 】



10

20

30

40

50

フロントページの続き

- 大阪府門真市大字門真 1 0 0 6 番地 パナソニックインダストリー株式会社内
(72)発明者 相川 恒
大阪府門真市大字門真 1 0 0 6 番地 パナソニックインダストリー株式会社内
審査官 山田 辰美
(56)参考文献 特開 2 0 2 0 - 0 3 0 6 9 2 (J P , A)
特開 2 0 0 9 - 2 8 2 6 8 6 (J P , A)
特開 2 0 1 8 - 1 0 6 6 6 2 (J P , A)
(58)調査した分野 (Int.Cl. , D B 名)
G 0 6 T 7 / 0 0
G 0 1 N 2 1 / 8 8
G 0 6 N 2 0 / 0 0