

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7217842号
(P7217842)

(45)発行日 令和5年2月3日(2023.2.3)

(24)登録日 令和5年1月26日(2023.1.26)

(51)国際特許分類 F I
G 0 6 N 20/00 (2019.01) G 0 6 N 20/00

請求項の数 12 (全30頁)

(21)出願番号	特願2022-566633(P2022-566633)	(73)特許権者	000006013 三菱電機株式会社 東京都千代田区丸の内二丁目7番3号
(86)(22)出願日	令和2年12月15日(2020.12.15)	(74)代理人	110002491 弁理士法人クロスボーダー特許事務所
(86)国際出願番号	PCT/JP2020/046743	(72)発明者	森 郁海 東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内
(87)国際公開番号	WO2022/130498	(72)発明者	板垣 弦矢 東京都千代田区丸の内二丁目7番3号 三菱電機株式会社内
(87)国際公開日	令和4年6月23日(2022.6.23)	審査官	山本 俊介
審査請求日	令和4年11月1日(2022.11.1)		
早期審査対象出願			

最終頁に続く

(54)【発明の名称】 類似度算出装置、類似度算出方法、及び、類似度算出プログラム

(57)【特許請求の範囲】

【請求項1】

転移元データ候補と転移先データとを用いて転移学習を行う類似度算出装置であって、
前記転移元データ候補の特徴量の分布を示す転移元データ分布が有する極値を示す転移元極値を含む転移元極値群と、前記転移先データの特徴量の分布を示す転移先データ分布が有する極値を示す転移先極値を含む転移先極値群とに基づいて、前記転移元データ分布と前記転移先データ分布との類似度を求める類似度算出部を備える類似度算出装置であって、

前記転移元データ候補と前記転移元データ分布と前記転移元極値群とのそれぞれは少なくとも1つ存在し、前記転移元データ候補それぞれは前記転移元データ分布それぞれと1対1で対応しており、前記転移元データ分布それぞれは前記転移元極値群それぞれと1対1で対応しており、

前記類似度算出部は、前記転移元データ分布それぞれに対応する転移元極値群と前記転移先極値群とに基づいて、前記転移元データ分布それぞれと前記転移先データ分布との類似度を求め、

前記転移元極値それぞれは、前記転移元極値それぞれに対応する極値を有する転移元データ分布に対応する転移元データ候補を識別する識別子と対応付けられており、前記転移元極値それぞれを含む転移元極値群が含む転移元極値の数に応じて、極値の数が示す値ごとに用意されたデータセットの少なくとも1つに属し、

前記類似度算出装置は、さらに、

前記データセットそれぞれに対応するモデルであって、与えられた座標群に対応する識別子を推定することに用いられるモデルである学習モデルを作成するモデル作成部と、

前記学習モデルから、前記転移先極値群を含む転移先極値の数に応じて学習モデルを決定モデルとして決定するモデル決定部と

を備え、

前記類似度算出部は、前記決定モデルと前記転移先極値群とを用いて前記類似度を求める類似度算出装置。

【請求項 2】

前記データセットそれぞれは、前記データセットそれぞれに対応する転移元極値の数に対応する転移元極値を含む請求項 1 に記載の類似度算出装置。

10

【請求項 3】

前記データセットそれぞれは、前記データセットそれぞれに対応する転移元極値の数の周囲の範囲を示すデータセット許容範囲内の極値の数それぞれに対応する転移元極値を含むデータセット群である請求項 1 に記載の類似度算出装置。

【請求項 4】

前記モデル決定部は、前記転移先極値群を含む転移先極値の数の周囲の範囲を示すモデル許容範囲内の転移元極値の数それぞれに対応する学習モデルから成る学習モデル群を決定モデルとして決定する請求項 1 から 3 のいずれか 1 項に記載の類似度算出装置。

【請求項 5】

前記類似度算出装置は、さらに、

20

前記転移元極値として、前記転移元データ分布それぞれが有する極値である変換前転移元極値の次元よりも次元が高い空間に前記変換前転移元極値を射影した座標を求め、

前記転移先極値として、前記転移先データ分布が有する極値である変換前転移先極値の次元よりも次元が高い空間に前記変換前転移先極値を射影した座標を求める座標変換部を備える請求項 1 から 4 のいずれか 1 項に記載の類似度算出装置。

【請求項 6】

転移元データ候補と転移先データとを用いて転移学習を行う類似度算出装置であって、

前記転移元データ候補の特徴量の分布を示す転移元データ分布が有する極値を示す転移元極値を含む転移元極値群と、前記転移先データの特徴量の分布を示す転移先データ分布が有する極値を示す転移先極値を含む転移先極値群とに基づいて、前記転移元データ分布と前記転移先データ分布との類似度を求める類似度算出部と、

30

前記転移元極値として、前記転移元データ分布それぞれが有する極値である変換前転移元極値の次元よりも次元が高い空間に前記変換前転移元極値を射影した座標を求め、

前記転移先極値として、前記転移先データ分布が有する極値である変換前転移先極値の次元よりも次元が高い空間に前記変換前転移先極値を射影した座標を求める座標変換部とを備える類似度算出装置。

【請求項 7】

前記類似度算出装置は、さらに、

決定条件を満たす類似度に対応する転移元データ候補を転移元データと決定するデータ決定部を備える請求項 1 から 6 のいずれか 1 項に記載の類似度算出装置。

40

【請求項 8】

前記類似度算出装置は、さらに、

前記転移元データ分布と前記転移先データ分布とを受け取り、前記転移元データ分布を用いて前記転移元データ分布が有する極値を求め、前記転移先データ分布を用いて前記転移先データ分布が有する極値を求める極値計算部を備える請求項 1 から 7 のいずれか 1 項に記載の類似度算出装置。

【請求項 9】

転移元データ候補と転移先データとを用いて転移学習を行う類似度算出方法であって、

類似度算出部が、前記転移元データ候補の特徴量の分布を示す転移元データ分布が有する極値を示す転移元極値を含む転移元極値群と、前記転移先データの特徴量の分布を示す

50

転移先データ分布が有する極値を示す転移先極値を含む転移先極値群とに基づいて、前記転移元データ分布と前記転移先データ分布との類似度を求める類似度算出方法であって、
前記転移元データ候補と前記転移元データ分布と前記転移元極値群とのそれぞれは少なくとも1つ存在し、前記転移元データ候補それぞれは前記転移元データ分布それぞれと1対1で対応しており、前記転移元データ分布それぞれは前記転移元極値群それぞれと1対1で対応しており、

前記類似度算出部が、前記転移元データ分布それぞれに対応する転移元極値群と前記転移先極値群とに基づいて、前記転移元データ分布それぞれと前記転移先データ分布との類似度を求め、

前記転移元極値それぞれは、前記転移元極値それぞれに対応する極値を有する転移元データ分布に対応する転移元データ候補を識別する識別子と対応付けられており、前記転移元極値それぞれを含む転移元極値群が含む転移元極値の数に応じて、極値の数が示す値ごとに用意されたデータセットの少なくとも1つに属し、

モデル作成部が、前記データセットそれぞれに対応するモデルであって、与えられた座標群に対応する識別子を推定することに用いられるモデルである学習モデルを作成し、

モデル決定部が、前記学習モデルから、前記転移先極値群が含む転移先極値の数に応じて学習モデルを決定モデルとして決定し、

前記類似度算出部は、前記決定モデルと前記転移先極値群とを用いて前記類似度を求める類似度算出方法。

【請求項10】

転移元データ候補と転移先データとを用いて転移学習を行う類似度算出方法であって、
 類似度算出部が、前記転移元データ候補の特徴量の分布を示す転移元データ分布が有する極値を示す転移元極値を含む転移元極値群と、前記転移先データの特徴量の分布を示す転移先データ分布が有する極値を示す転移先極値を含む転移先極値群とに基づいて、前記転移元データ分布と前記転移先データ分布との類似度を求め、

座標変換部が、

前記転移元極値として、前記転移元データ分布それぞれが有する極値である変換前転移元極値の次元よりも次元が高い空間に前記変換前転移元極値を射影した座標を求め、

前記転移先極値として、前記転移先データ分布が有する極値である変換前転移先極値の次元よりも次元が高い空間に前記変換前転移先極値を射影した座標を求める類似度算出方法。

【請求項11】

コンピュータに転移元データ候補と転移先データとを用いて転移学習を行わせる類似度算出プログラムであって、

前記コンピュータに、

前記転移元データ候補の特徴量の分布を示す転移元データ分布が有する極値を示す転移元極値を含む転移元極値群と、前記転移先データの特徴量の分布を示す転移先データ分布が有する極値を示す転移先極値を含む転移先極値群とに基づいて、前記転移元データ分布と前記転移先データ分布との類似度を求めさせる類似度算出プログラムであって、

前記転移元データ候補と前記転移元データ分布と前記転移元極値群とのそれぞれは少なくとも1つ存在し、前記転移元データ候補それぞれは前記転移元データ分布それぞれと1対1で対応しており、前記転移元データ分布それぞれは前記転移元極値群それぞれと1対1で対応しており、

前記コンピュータに、前記転移元データ分布それぞれに対応する転移元極値群と前記転移先極値群とに基づいて、前記転移元データ分布それぞれと前記転移先データ分布との類似度を求めさせ、

前記転移元極値それぞれは、前記転移元極値それぞれに対応する極値を有する転移元データ分布に対応する転移元データ候補を識別する識別子と対応付けられており、前記転移元極値それぞれを含む転移元極値群が含む転移元極値の数に応じて、極値の数が示す値ごとに用意されたデータセットの少なくとも1つに属し、

10

20

30

40

50

前記コンピュータに、

前記データセットそれぞれに対応するモデルであって、与えられた座標群に対応する識別子を推定することに用いられるモデルである学習モデルを作成させ、

前記学習モデルから、前記転移先極値群を含む転移先極値の数に応じて学習モデルを決定モデルとして決定させ、

前記決定モデルと前記転移先極値群とを用いて前記類似度を求めさせる類似度算出プログラム。

【請求項 1 2】

コンピュータに転移元データ候補と転移先データとを用いて転移学習を行わせる類似度算出プログラムであって、

前記コンピュータに、

前記転移元データ候補の特徴量の分布を示す転移元データ分布が有する極値を示す転移元極値を含む転移元極値群と、前記転移先データの特徴量の分布を示す転移先データ分布が有する極値を示す転移先極値を含む転移先極値群とに基づいて、前記転移元データ分布と前記転移先データ分布との類似度を求めさせ、

前記転移元極値として、前記転移元データ分布それぞれが有する極値である変換前転移元極値の次元よりも次元が高い空間に前記変換前転移元極値を射影した座標を求めさせ、

前記転移先極値として、前記転移先データ分布が有する極値である変換前転移先極値の次元よりも次元が高い空間に前記変換前転移先極値を射影した座標を求めさせる類似度算出プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、類似度算出装置、類似度算出方法、及び、類似度算出プログラムに関する。

【背景技術】

【0002】

IoT (Internet of Things) 機器において AI (Artificial Intelligence) を利用するソリューションが増加している。具体例として、(1) 空気調和機及び照明等の IoT 家電の制御と、(2) 製造装置の故障診断と、(3) 製造ラインにおける製品の画像による検品と、(4) ビル等の入門時の動画による不審者侵入検知と、(5) EMS (Energy Management System) とが挙げられる。

IoT 機器毎に AI を利用する場合、学習処理に用いる十分な数の訓練データを確保することが困難であることが多い。そこで、少ない訓練データを用いて効率的に学習処理を行う必要がある。少ない訓練データを用いて学習する方法として、転移学習と呼ばれる方法がある。転移学習において、訓練データが収集された環境とは異なる環境における訓練データ及び学習モデルを転移させる。

転移学習では、転移元を特定するために、転移元の候補の全データに対して転移元になり得るか否かを 1 つずつ評価する。転移元の候補を転移することが有効であることを示す「正の転移」であることを評価により確認することができた場合、転移元の候補を転移元として決定する。訓練データが画像データである場合、転移元の候補の画像が転移先の画像と十分に類似していることが「正の転移」となることの十分条件となることが多い。

特許文献 1 は、画像の類似度を算出する方法として、各画像のカラーヒストグラムを算出して類似度を求める方法と、輝度の勾配方向を特徴量として類似度を求める方法と、入力画像及び学習画像のそれぞれから検出した局所特徴の特徴量が類似する個数を評価値とする方法等、何らかの評価関数を用いて、検索対象画像と検索元画像との類似度を計算する方法を開示している。

【先行技術文献】

【特許文献】

【0003】

【文献】特開 2019 - 087229 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

特許文献 1 が開示する手法によれば、計算量が比較的多いという課題がある。

【0005】

本開示は、転移学習において、比較的少ない計算量で転移元を求めることを目的とする。

【課題を解決するための手段】

【0006】

本開示に係る類似度算出装置は、

転移元データ候補と転移先データとを用いて転移学習を行う類似度算出装置であって、前記転移元データ候補の特徴量の分布を示す転移元データ分布が有する極値を示す転移元極値を含む転移元極値群と、前記転移先データの特徴量の分布を示す転移先データ分布が有する極値を示す転移先極値を含む転移先極値群とに基づいて、前記転移元データ分布と前記転移先データ分布との類似度を求める類似度算出部を備える。

【発明の効果】

【0007】

本開示によれば、転移学習において、比較的少ない計算量で転移元を求めることができる。

【図面の簡単な説明】

【0008】

【図 1】実施の形態 1 に係る類似度算出装置 1 の構成例。

【図 2】実施の形態 1 に係る特徴データ分布 1025 の具体例。

【図 3】実施の形態 1 に係るデータセット 105 の具体例。

【図 4】実施の形態 1 に係る類似度算出装置 1 のハードウェア構成例。

【図 5】実施の形態 1 に係る類似度算出装置 1 の動作の概要を説明する図。

【図 6】実施の形態 1 に係る類似度算出装置 1 の学習フェーズに係る動作を示すフローチャート。

【図 7】実施の形態 1 に係る類似度算出装置 1 の学習フェーズに係る動作を説明する図であり、(a) は転移元データ候補を示す画像、(b) は(a) をグレースケール化した画像、(c) は(b) に対応する特徴データ分布。

【図 8】実施の形態 1 に係る類似度算出装置 1 の推論フェーズに係る動作を示すフローチャート。

【図 9】実施の形態 1 に係る類似度算出装置 1 の動作を説明する図。

【図 10】実施の形態 1 の変形例に係る類似度算出装置 1 のハードウェア構成例。

【図 11】実施の形態 2 に係る類似度算出装置 1 の動作の概要を説明する図であり、(a) は座標系を変換する前を説明する図、(b) は座標系を変換した後を説明する図。

【図 12】実施の形態 2 に係る類似度算出装置 1 の構成例。

【図 13】実施の形態 2 に係る類似度算出装置 1 の学習フェーズに係る動作を示すフローチャート。

【図 14】実施の形態 2 に係る類似度算出装置 1 の推論フェーズに係る動作を示すフローチャート。

【図 15】実施の形態 3 に係る類似度算出装置 1 の動作の概要を説明する図であり、(a) は許容範囲がない場合を説明する図、(b) は許容範囲がある場合を説明する図。

【図 16】実施の形態 3 に係る類似度算出装置 1 の構成例。

【図 17】実施の形態 3 に係る類似度算出装置 1 の動作を説明する図。

【図 18】実施の形態 4 に係る類似度算出装置 1 の動作の概要を説明する図であり、(a) は許容範囲がない場合を説明する図、(b) は許容範囲がある場合を説明する図。

【図 19】実施の形態 4 に係る類似度算出装置 1 の構成例。

【発明を実施するための形態】

10

20

30

40

50

【 0 0 0 9 】

実施の形態の説明及び図面において、同じ要素及び対応する要素には同じ符号を付している。同じ符号が付された要素の説明は、適宜に省略又は簡略化する。図中の矢印はデータの流れ又は処理の流れを主に示している。また、「部」を、「回路」、「工程」、「手順」、「処理」又は「サーキットリー」に適宜読み替えてもよい。

【 0 0 1 0 】

実施の形態 1 .

以下、本実施の形態について、図面を参照しながら詳細に説明する。

【 0 0 1 1 】

*** 構成の説明 ***

図 1 は、本実施の形態に係る類似度算出装置 1 の構成例を示している。類似度算出装置 1 は、本図に示すように、特徴抽出部 1 1 と、データ分布計算部 1 2 と、極値計算部 1 3 と、グルーピング部 1 4 と、データセット記憶部 1 5 と、モデル作成部 1 6 と、分類器記憶部 1 7 と、モデル決定部 1 8 と、類似度算出部 1 9 と、データ決定部 2 0 とを備える。

なお、類似度算出装置 1 は、転移元データ候補 2 と転移先データ 3 とを用いて転移学習を行う。また、符号の末尾の T 及び上付きの t は転移先データ分布に関するものであることを示し、符号の末尾の S 及び上付きの s は転移元データ分布に関するものであることを示す。

【 0 0 1 2 】

特徴抽出部 1 1 は、転移元データ候補 2 を用いて転移元データ候補 2 の特徴量を示す特徴ベクトル 1 0 1 S を生成し、転移先データ 3 を用いて転移先データ 3 の特徴量を示す特徴ベクトル 1 0 1 T を生成する。転移元データ候補 2 及び転移先データ 3 それぞれは、他の装置から受信したデータであってもよく、類似度算出装置 1 が記憶しているデータであってもよい。なお、転移元データ候補 2 は、転移元データの候補であり、少なくとも 1 つ存在する。転移先データ 3 は 1 つ存在するものとする。特徴抽出部 1 1 が求める特徴量のデータ形式は、ベクトル形式でなくてもよい。なお、特徴ベクトル 1 0 1 を特徴ベクトル 1 0 1 S 及び特徴ベクトル 1 0 1 T の総称とする。特徴ベクトル 1 0 1 は具体例として画素値の配列である。

特徴抽出部 1 1 は、特徴量として、具体例として、主成分分析により転移元データ候補 2 又は転移先データ 3 を要約したデータと、転移元データ候補 2 又は転移先データ 3 に関する平均値、分散、又は相関係数等の統計量と、ヒストグラム又はフーリエ変換等の(同種)写像を用いて転移元データ候補 2 又は転移先データ 3 を変換した値と、回帰分析等を利用して転移元データ候補 2 又は転移先データ 3 に対する関数近似を行いその係数又はその変化量である勾配ベクトルとの少なくともいずれかを求める。ヒストグラムという用語は、ヒストグラムにより表現することができるデータを指すこともある。なお、特徴抽出部 1 1 が求める特徴量は、転移元データ候補 2 又は転移先データ 3 の特徴を十分に反映しているものとする。

【 0 0 1 3 】

データ分布計算部 1 2 は、特徴ベクトル 1 0 1 S を用いて特徴データ分布 1 0 2 S を生成し、特徴ベクトル 1 0 1 T を用いて特徴データ分布 1 0 2 T を生成する。特徴データ分布 1 0 2 S は、転移元データ候補 2 の特徴量の分布を示し、転移元データ分布とも呼ばれ、少なくとも 1 つ存在する。転移元データ候補 2 それぞれは特徴データ分布 1 0 2 S それぞれと 1 対 1 で対応している。特徴データ分布 1 0 2 T は、転移先データ 3 の特徴量の分布を示し、転移先データ分布とも呼ばれる。なお、特徴データ分布 1 0 2 を特徴データ分布 1 0 2 S 及び特徴データ分布 1 0 2 T の総称とする。特徴データ分布 1 0 2 は、典型的には関数であり、離散分布であっても、連続分布であってもよい。

【 0 0 1 4 】

極値計算部 1 3 は、特徴データ分布 1 0 2 S を用いて極値データ 1 0 3 S を求める。極値データ 1 0 3 S は、データ ID (I d e n t i f i c a t i o n) ^S と、極値座標群 G ^S と、極値の数 E ^S とから成り、計算結果とも呼ばれる。データ ID ^S は、転移元データ候

10

20

30

40

50

補 2 を識別する識別子であり、転移元データ ID とも呼ばれる。極値座標群 G^s は、特徴データ分布 102S が有する極値の座標から成る集合であり、転移元極値座標群又は転移元極値群とも呼ばれ、少なくとも 1 つ存在する。特徴データ分布 102S それぞれは極値座標群 G^s それぞれと 1 対 1 で対応している。極値座標群 G^s が含む各極値は転移元極値とも呼ばれる。転移元極値それぞれは、転移元極値それぞれに対応する極値を有する特徴データ分布 102S に対応する転移元データ候補 2 を識別する識別子と対応付けられている。極値の数 E^s は極値座標群 G^s が含む極値の数である。

また、極値計算部 13 は、特徴データ分布 102T を用いて極値データ 103T を求める。極値データ 103T は極値座標群 G^t と、極値の数 E^t とから成り、計算結果とも呼ばれる。極値座標群 G^t は、特徴データ分布 102T が有する極値の座標から成る集合であり、転移先極値座標群又は転移先極値群とも呼ばれる。極値の数 E^t は極値座標群 G^t が含む極値の数である。極値データ 103T は転移先データ 3 を識別する識別子を含んでもよい。極値の数を極値の数 E^s と極値の数 E^t との総称とする。

10

極値計算部 13 は、転移元データ分布と転移先データ分布とを受け取り、転移元データ分布を用いて転移元データ分布が有する極値を求め、転移先データ分布を用いて転移先データ分布が有する極値を求める。

【0015】

特徴データ分布 102S が $y = f^s(x)$ であり、極値の数 E^s が n である場合、極値座標群 G^s が含む各座標 (x_i^s, y_i^s) は [数 1] のように定義される。[数 1] は極値座標群 G^s の具体例でもある。なお、極値座標群 G^s に含まれる各座標の次元数は 2 に限られないが、説明の便宜上、当該各座標の次元を 2 としている。極値座標群 G^t は [数 1] と同様に定義される。なお、変数 s を与えられた特徴データ分布 102S 全てから成る集合 S の各元としている。即ち、 $s \in S$ である。

20

【0016】

【数 1】

$$\left\{ (x_i^s, y_i^s) \in \mathbb{R}^2 \left| \begin{array}{l} i = 1, \dots, n, s \in S, \\ y_i^s = f^s(x_i^s) \geq f^s(x_{\varepsilon_i}^s) \vee y_i^s = f^s(x_i^s) \leq f^s(x_{\varepsilon_i}^s), \\ x_{\varepsilon_i}^s \in [x_i^s - \varepsilon, x_i^s + \varepsilon] \end{array} \right. \right\}$$

30

【0017】

図 2 は、特徴データ分布 102S の具体例を示している。本図は、[数 1] に示す n が 6 である場合における、極値座標群 G^s に含まれる各極値の具体例を示している。

【0018】

グルーピング部 14 は、極値データ 103S を用いて学習データ 104 を生成し、生成した学習データ 104 をデータセット記憶部 15 に保存する。学習データ 104 は訓練データと教師データとを含む。学習データ 104 において、訓練データは極値座標群 G^s が含む各極値座標であり、教師データは訓練データである極値座標を有する特徴データ分布 102S に対応するデータ ID^s である。教師データはラベルでもある。ラベルは、転移元データ候補 2 を識別する識別子でもある。

40

グルーピング部 14 は、学習データ 104 に対応する極値の数 E^s により学習データ 104 を分類した上で学習データ 104 をデータセット記憶部 15 に保存する。学習データ 104 に対応する極値の数 E^s は、学習データ 104 が含む極値座標を含む極値座標群 G^s に含まれる極値の数 E^s である。

【0019】

データセット記憶部 15 は、データセット 105 - 1 からデータセット 105 - N を記憶し、また、データセット 1 ~ N 記憶部とも呼ばれる。N は、極値の数 E^s の最大値を示す定数であり、動的に変更されてもよい。データセット 105 - n ($1 \leq n \leq N$, n は整数) は、極値の数 E^s が n である特徴データ分布 102S それぞれに対応する学習データ

50

104を元とする集合である。即ち、 n は極値の数 E^S に対応する変数である。なお、データセット105-1からデータセット105-Nの少なくとも一部は空のデータであってもよい。

転移元極値それぞれは、転移元極値それぞれに対応する極値を有する特徴データ分布102Sに対応する転移元データ候補2を識別する識別子と対応付けられている。転移元極値それぞれは、転移元極値それぞれを含む転移元極値群が含む転移元極値の数に応じて、データセット105の少なくとも1つに属する。データセット105は、極値の数が示す値ごとに用意されている。本実施の形態において、データセット105それぞれは、データセット105それぞれに対応する転移元極値の数に対応する転移元極値を含む。データセット105-nに対応する転移元極値の数は n である。転移元極値の数が n である場合において、転移元極値の数に対応する転移元極値は、 n 個の転移元極値を有する特徴データ分布102Sが有する転移元極値である。

10

【0020】

[数2]は、極値の数 E^S が $n(=|\{(x_i^S, y_i^S)\}|)$ である特徴データ分布102S全てから生成されたデータセット105-nの具体例を示している。

【0021】

【数2】

$$g_n = \{(x_i^S, y_i^S), c^S \mid |\{(x_i^S, y_i^S)\}| = n, i = 1, \dots, n, S \in \mathcal{S}\}$$

20

【0022】

ここで、データセット105-nを集合 g_n と表現し、極値座標群 G^S が含む各座標を (x_i^S, y_i^S) と表現し、データID S を c^S と表現している。

【0023】

図3は、データセット105の具体例を示している。本図の上部に極値座標群 G^S が示されており、 $G^{S^1}, G^{S^2}, \dots, G^{S^n}$ は各極値座標群 G^S である。ここで、上付きの添え字は、各極値座標群 G^S が含む極値座標を有する特徴データ分布102Sに対応するデータID S を示すラベルである。また、本図の下部に上部に示される極値座標群 G^S に対応するデータセット105が示されている。 G^{S^1} 及び G^{S^n} それぞれに対応する極値の数 E^S は3であるため G^{S^1} 及び G^{S^n} の各要素はデータセット105-3が含む各要素に含まれ、 G^{S^2} に対応する極値の数 E^S は1であるため G^{S^2} の要素はデータセット105-1が含む要素に含まれる。

30

【0024】

モデル作成部16は、データセット105を用いて学習を行うことにより学習モデルである分類器106を作成する。具体的には、モデル作成部16は、データセット105-nをデータセット記憶部15から取得し、取得したデータセット105-nを用いて分類器106-nを作成する。即ち、モデル作成部16は、極値の数 E^S ごとに分類器106を作成する。なお、データセット105-nが空である場合、モデル作成部16は分類器106-nを作成しなくてもよい。学習モデルである分類器106は、データセット105それぞれに対応するモデルであって、与えられた座標群に対応する識別子を推定することに用いられるモデルであり、特徴データ分布102Tの各極値を分類することに用いられるモデルである。具体的には、分類器106-nは、極値の数 E^S が n である特徴データ分布102Sそれぞれに対応するクラスに、極値の数 E^T が n である特徴データ分布102Tの各極値が所属する確率(クラス所属確率)を求める。なお、学習モデルを単にモデルと表現することもある。

40

モデル作成部16は、学習に用いる学習アルゴリズムとして、具体例として、KNN(k-nearest neighbor, k近傍法)又はSVC(Support Vector Classification)等を用いる。

【0025】

分類器記憶部17は、分類器106-1から分類器106-Nを記憶し、また、分類器

50

1 ~ N 記憶部とも呼ばれる。分類器 106 - n は、極値の数 E^s が n である特徴データ分布 102S に対応する。

【0026】

モデル決定部 18 は、類似度算出部 19 が用いる分類器 106 を、極値座標群 G^t を用いて決定する。具体例として、極値の数 E^t が $m (= |\{(x_i^t, y_i^t)\}|)$ である場合、モデル決定部 18 は、分類器記憶部 17 から分類器 106 - m を取得し、類似度算出部 19 が用いる分類器 106 として取得した分類器 106 - m を決定する。

モデル決定部 18 は、決定した分類器 106 を含む推論データ 108 を生成する。推論データ 108 は、極値座標群 G^t と、極値座標群 G^t に対応する分類器 106 とを含む。推論データ 108 - m は、極値の数 E^t が m である極値座標群 G^t を含む推論データ 108 である。

10

モデル決定部 18 は、少なくとも 1 つの学習モデルから、極値座標群 G^t が含む極値の数に応じて学習モデルを決定モデルとして決定する。

【0027】

類似度算出部 19 は、推論データ 108 が含む分類器 106 と極値座標群 G^t とを用いて少なくとも 1 つの特徴データ分布 102S それぞれと特徴データ分布 102T との類似度を求め、求めた類似度を含む類似度データ 109 を出力する。類似度データ 109 は、データ ID^s のラベルと、データ ID^s に対応する類似度とを含む。

類似度算出部 19 は、決定モデルと極値座標群 G^t とを用いて類似度を求める。ここで、決定モデルは極値座標群 G^s に基づいて生成されている。よって、類似度算出部 19 は、特徴データ分布 102S それぞれに対応する極値座標群 G^s と極値座標群 G^t とに基づいて、特徴データ分布 102S それぞれと特徴データ分布 102T との類似度を求める。

20

【0028】

以下、類似度の具体例を説明する。[数3] は、各 (x_i^t, y_i^t) のクラス所属確率を推定クラス c_j ごとに示している。ここで、出力は「推定クラス(クラス所属確率)」という形式で表記されている。また、1 つの入力に対する出力の数は、極値の数 E^s が m である特徴データ分布 102S の総数である。

【0029】

【数3】

入力 : $(x_1^t, y_1^t) \rightarrow$ 出力 : $c_1(p_{c_1}^1), \dots, c_j(p_{c_j}^1), \dots$
 ...
 入力 : $(x_m^t, y_m^t) \rightarrow$ 出力 : $c_1(p_{c_1}^m), \dots, c_j(p_{c_j}^m), \dots$

30

【0030】

特徴データ分布 102T と推定クラス c_j との類似度を示す $score(c_j)$ は、具体例として [数4] のように示される。なお、 $score(c_j)$ は、転移先データ 3 の各転移元データ候補 2 との類似度でもある。

【0031】

40

【数4】

$$score(c_j) = \sum_{i=1}^m (w_i \cdot p_{c_j}^i)$$

【0032】

ここで、 w_i は各極値 (x_i^s, y_i^s) を重視する程度を示す重みパラメータである。

【0033】

データ決定部 20 は、転移元データ決定部とも呼ばれ、類似度データ 109 及び決定条

50

件に基づいて少なくとも1つの転移元データ候補2から転移元データを決定し、決定した転移元データを含む決定データ4を出力する。決定データ4は、転移元データ候補2を示すラベルと、当該転移元データ候補2に対応する類似度とを含む。決定条件は、データ決定部20が転移元データ候補2を転移元データと決定する条件である。決定条件は、具体例として、類似度がある閾値以上であることである。データ決定部20は、決定条件を満たす類似度に対応する転移元データ候補2を転移元データと決定する。

データ決定部20は、具体例として、決定条件を満たす類似度に対応するデータID^Sを特定し、特定したデータID^Sに対応する転移元データ候補2及び類似度を出力する。

【0034】

図4は、本実施の形態に係る類似度算出装置1のハードウェア構成例を示している。類似度算出装置1は、コンピュータから成る。類似度算出装置1は、複数のコンピュータから成ってもよい。

10

また、類似度算出装置1は、クラウドコンピューティングにおけるサーバ(コンピュータ)において動作させてもよく、エッジコンピューティングにおけるサーバ(コンピュータ)において動作させてもよい。

【0035】

類似度算出装置1は、本図に示すように、プロセッサ51と、メモリ52と、補助記憶装置53と、入出力IF(Interface)54と、通信装置55と等のハードウェアを備えるコンピュータである。これらのハードウェアは、信号線59を介して互いに接続されている。

20

【0036】

プロセッサ51は、演算処理を行うIC(Integrated Circuit)であり、かつ、コンピュータが備えるハードウェアを制御する。プロセッサ51は、具体例として、CPU(Central Processing Unit)、DSP(Digital Signal Processor)、又はGPU(Graphics Processing Unit)である。

類似度算出装置1は、プロセッサ51を代替する複数のプロセッサを備えてもよい。複数のプロセッサは、プロセッサ51の役割を分担する。

【0037】

メモリ52は、典型的には、揮発性の記憶装置である。メモリ52は、主記憶装置又はメインメモリとも呼ばれる。メモリ52は、具体例として、RAM(Random Access Memory)である。メモリ52に記憶されたデータは、必要に応じて補助記憶装置53に保存される。

30

【0038】

補助記憶装置53は、典型的には、不揮発性の記憶装置である。補助記憶装置53は、具体例として、ROM(Read Only Memory)、HDD(Hard Disk Drive)、又はフラッシュメモリである。補助記憶装置53に記憶されたデータは、必要に応じてメモリ52にロードされる。

メモリ52と、補助記憶装置53とは、一体的に構成されていてもよい。

【0039】

40

入出力IF54は、入力装置及び出力装置が接続されるポートである。入出力IF54は、具体例として、USB(Universal Serial Bus)端子である。入力装置は、具体例として、キーボード及びマウスである。出力装置は、具体例として、ディスプレイである。

【0040】

通信装置55は、レシーバ及びトランスミッタである。通信装置55は、具体例として、通信チップ又はNIC(Network Interface Card)である。

【0041】

類似度算出装置1の各部は、他の装置等と通信する際に、通信装置55を適宜用いてもよい。類似度算出装置1の各部は、入出力IF54を介してデータを受け付けてもよく、

50

また、通信装置 5 5 を介してデータを受け付けてもよい。

【 0 0 4 2 】

補助記憶装置 5 3 は、類似度算出プログラムを記憶している。類似度算出プログラムは、類似度算出装置 1 が備える各部の機能をコンピュータに実現させるプログラムである。類似度算出プログラムは、複数のファイルから成ってもよい。類似度算出プログラムは、メモリ 5 2 にロードされて、プロセッサ 5 1 によって実行される。類似度算出装置 1 が備える各部の機能は、ソフトウェアにより実現される。

【 0 0 4 3 】

類似度算出プログラムを実行する際に用いられるデータと、類似度算出プログラムを実行することによって得られるデータと等は、記憶装置に適宜記憶される。類似度算出装置 1 の各部は、適宜記憶装置を利用する。記憶装置は、具体例として、メモリ 5 2 と、補助記憶装置 5 3 と、プロセッサ 5 1 内のレジスタと、プロセッサ 5 1 内のキャッシュメモリとの少なくとも 1 つから成る。なお、データと、情報とは、同等の意味を有することもある。記憶装置は、コンピュータと独立したものであってもよい。データセット記憶部 1 5 と分類器記憶部 1 7 とのそれぞれは記憶装置から成る。

メモリ 5 2 及び補助記憶装置 5 3 の機能は他の記憶装置によって実現されてもよい。

【 0 0 4 4 】

類似度算出プログラムは、コンピュータが読み取り可能な不揮発性の記録媒体に記録されていてもよい。不揮発性の記録媒体は、具体例として、光ディスク又はフラッシュメモリである。類似度算出プログラムは、プログラムプロダクトとして提供されてもよい。

【 0 0 4 5 】

*** 動作の説明 ***

類似度算出装置 1 の動作手順は、類似度算出方法に相当する。また、類似度算出装置 1 の動作を実現するプログラムは、類似度算出プログラムに相当する。

【 0 0 4 6 】

類似度算出装置 1 は、以下の 2 つの基本的な方針に従って、特徴データ分布 1 0 2 S の特徴データ分布 1 0 2 T との類似度を評価する。

方針 (1) : 類似度算出装置 1 は、特徴データ分布 1 0 2 S のある極値が特徴データ分布 1 0 2 T のある極値と近接している場合に、特徴データ分布 1 0 2 S の当該ある極値において特徴データ分布 1 0 2 S は特徴データ分布 1 0 2 T と部分的に類似しているものとみなす。なお、類似度算出装置 1 は、特徴データ分布 1 0 2 T のある極値と、特徴データ分布 1 0 2 S のある極値とがより近い場合に、特徴データ分布 1 0 2 S は特徴データ分布 1 0 2 T と部分的により類似しているとみなしてもよい。

方針 (2) : 類似度算出装置 1 は、部分的に類似している極値をより多く持つ特徴データ分布 1 0 2 S ほど類似度を高く算出する。

【 0 0 4 7 】

図 5 は、前述の方針 (1) 及び方針 (2) に示す方針を具体的に説明する図である。本図において、2 つの特徴データ分布 1 0 2 S (転移元データ分布 D 2 及び転移元データ分布 D 3) と、1 つの転移先の特徴データ分布 (転移先データ分布 D 1) とが示されている。また、丸を用いて極大値が示されており、四角形を用いて極小値が示されており、転移先データ分布 D 1 の各極値を中心とした円により、各極値に近接している領域が示されている。各近接している領域を近傍領域 R 1 から近傍領域 R 4 としている。また、転移元データ分布のある極値が転移先データ分布 D 1 のある極値と近接している場合に、双方の極値を破線で結んでいる。

本図より、転移元データ分布 D 2 の極値は、近傍領域 R 1 から近傍領域 R 4 のいずれにおいても転移先データ分布 D 1 の極値と近接している。一方、転移元データ分布 D 3 の極値は、近傍領域 R 1 と、近傍領域 R 2 と、近傍領域 R 4 とにおいてのみ転移先データ分布 D 1 の極値と近接している。つまり、転移元データ分布 D 2 は部分的に類似している極値を 4 つ持っており、転移元データ分布 D 3 は部分的に類似している極値を 3 つ持っている。従って、転移元データ分布 D 2 は転移元データ分布 D 3 と比較して部分的に類似してい

10

20

30

40

50

る極値をより多く持っている。そのため、類似度算出装置 1 は、転移元データ分布 D 2 に対応する類似度が転移元データ分布 D 3 に対応する類似度より高くなるように各々の類似度を算出する。

【 0 0 4 8 】

< 学習フェーズ >

図 6 は、類似度算出装置 1 の学習フェーズにおける動作の一例を示すフローチャートである。図 1 及び本図を用いて学習フェーズを説明する。

【 0 0 4 9 】

(ステップ S 1 0 1 : 特徴抽出処理)

特徴抽出部 1 1 は、転移元データ候補 2 を入力として受け取り、受け取った転移元データ候補 2 を用いて特徴ベクトル 1 0 1 S を生成し、生成した特徴ベクトル 1 0 1 S を出力する。

10

【 0 0 5 0 】

(ステップ S 1 0 2 : データ分布計算処理)

データ分布計算部 1 2 は、特徴ベクトル 1 0 1 S を入力として受け取り、受け取った特徴ベクトル 1 0 1 S を用いて特徴データ分布 1 0 2 S を生成し、生成した特徴データ分布 1 0 2 S を出力する。

【 0 0 5 1 】

図 7 は、ステップ S 1 0 1 及びステップ S 1 0 2 の具体例を模式的に説明する図である。本図を用いてステップ S 1 0 1 及びステップ S 1 0 2 の具体例を説明する。なお、本図において、転移元データ候補 2 を示す (a) は白黒により表現されているものの、実際にはトラが写っているカラー画像のデータである。

20

転移元データ候補 2 が画像であり、当該画像中の柄又は模様等に特徴がある場合において、特徴抽出部 1 1 は、特徴量として画素値の出現頻度を示すデータを出力してもよい。出現頻度を示すデータは、ヒストグラムによって表すこともできるデータである。特徴抽出部 1 1 は、1 枚の画像中に画素値 x が占める割合の確率を特徴量とする。特徴抽出部 1 1 は、具体例として、カラー画像をグレースケール化した後、カーネル密度推定によって画素値 x を確率変数とする確率質量関数 $p(x)$ を求め、 $(x, p(x))$ を特徴ベクトル 1 0 1 S とする。(b) は (a) をグレースケール化したデータを示している。

データ分布計算部 1 2 は、 $(x, p(x))$ を受け取り、 $(x, p(x))$ をスケール変換して特徴データ分布 1 0 2 S を求める。(c) は、データ分布計算部 1 2 が求めた特徴データ分布 1 0 2 S を示している。

30

ここで、特徴ベクトル 1 0 1 S である $(x, p(x))$ は特徴データ分布 1 0 2 S と捉えることができる。特徴ベクトル 1 0 1 S を特徴データ分布 1 0 2 S と捉えることができる場合において、データ分布計算部 1 2 は、受け取った特徴ベクトル 1 0 1 S をそのまま特徴データ分布 1 0 2 S としてもよく、受け取った特徴ベクトル 1 0 1 S をスケール変換したデータを特徴データ分布 1 0 2 S としてもよい。特徴ベクトル 1 0 1 S をスケール変換する目的として、後続の処理において特徴データ分布 1 0 2 S 同士を比較しやすくすることが挙げられる。スケール変換の具体例として、データ分布計算部 1 2 は、データ分布の最小値から最大値までの範囲を $[0, 1]$ の範囲にスケール変換する。

40

【 0 0 5 2 】

(ステップ S 1 0 3 : 極値計算処理)

極値計算部 1 3 は、特徴データ分布 1 0 2 S を入力として受け取り、受け取った特徴データ分布 1 0 2 S を用いて極値データ 1 0 3 S を生成し、生成した極値データ 1 0 3 S を出力する。具体的には、極値計算部 1 3 は、特徴データ分布 1 0 2 S から極大値及び極小値それぞれの数と、極大値及び極小値それぞれの座標を示す極値座標群 G^S とを求め、求めたデータを極値データ 1 0 3 S として出力する。

【 0 0 5 3 】

(ステップ S 1 0 4 : グルーピング処理)

グルーピング部 1 4 は、極値データ 1 0 3 S を入力として受け取り、受け取った極値デ

50

ータ103Sが示す極値の数と同数の学習データ104を、データセット記憶部15に保存する。

【0054】

類似度算出装置1は、用意された転移元データ候補2の数分、ステップS101からステップS104までを繰り返す。

【0055】

(ステップS105:モデル作成処理)

モデル作成部16は、データセット105を入力として受け取り、受け取ったデータセット105を用いて学習を行うことにより分類器106を作成し、作成した分類器106を分類器記憶部17に保存する。

10

【0056】

図8は、類似度算出装置1の推論フェーズにおける動作の一例を示すフローチャートである。図1及び本図を用いて推論フェーズを説明する。

【0057】

<推論フェーズ>

(ステップS111:特徴抽出処理)

特徴抽出部11は、転移先データ3を入力として受け取り、受け取った転移先データ3を用いて特徴ベクトル101Tを生成し、生成した特徴ベクトル101Tを出力する。

【0058】

(ステップS112:データ分布計算処理)

データ分布計算部12は、特徴ベクトル101Tを入力として受け取り、受け取った特徴ベクトル101Tを用いて特徴データ分布102Tを生成し、生成した特徴データ分布102Tを出力する。

20

【0059】

(ステップS113:極値計算処理)

極値計算部13は、特徴データ分布102Tを受け取り、受け取った特徴データ分布102Tを用いて極値データ103Tを生成し、生成した極値データ103Tを出力する。

【0060】

(ステップS114:モデル決定処理)

以下、極値の数 E^t が m であるものとして説明する。

30

モデル決定部18は、極値データ103Tを入力として受け取り、受け取った極値データ103Tに対応する分類器106である分類器106-mを分類器記憶部17から取得し、受け取った極値データ103Tが含む極値座標群 G^t と、取得した分類器106-mとを用いて推論データ108-mを生成し、生成した推論データ108-mを出力する。

【0061】

(ステップS115:類似度算出処理)

類似度算出部19は、推論データ108-mを入力として受け取り、受け取った推論データ108-mが含む分類器106-mと極値座標群 G^t とを用いて類似度を求め、類似度データ109を出力する。

具体例として、まず、類似度算出部19は、受け取った極値座標群 G^t から各極値を取り出す。この際、類似度算出部19は、合計で m 個の極値座標の組 (x_i^t, y_i^t) ($i = 1, \dots, m$)を作成する。なお、各極値はテストデータでもある。

40

次に、類似度算出部19は、 m 個のテストデータを一つずつ分類器106-mに入力し、推定クラスごと、クラス所属確率を求める。推定クラスは、テストデータが所属すると推定されるクラスである。類似度算出部19は、候補となる各クラスに各テストデータが所属すると推定し、各クラスにおける各テストデータのクラス所属確率を求める。以降、 (x_i^t, y_i^t) の推定クラスのラベルの集合を C^t とし、 C^t の元を c_j とする。

【0062】

次に、 $c_j \in C^t$ に対して $score(c_j)$ を計算し、データID s のラベルと、データID s に対応する類似度との組である $(c_j, score(c_j))$ を出力する。

50

【 0 0 6 3 】

(ステップ S 1 1 6 : データ決定処理)

データ決定部 2 0 は、類似度データ 1 0 9 を入力として受け取り、受け取った類似度データ 1 0 9 を用いて決定データ 4 を生成し、生成した決定データ 4 を出力する。

【 0 0 6 4 】

図 9 は、データ決定部 2 0 が転移元データ候補 2 を特定する具体例を説明する図である。本図において、転移先データ 3 はトラの画像であり、「転移元データ候補 2 のラベル(類似度)」という形式により転移元データ候補 2 の各画像に対応する転移元データ候補 2 のラベルと類似度とを記載している。また、各画像は実際にはカラー写真である。本例において、データ決定部 2 0 は、0.5 を閾値とし、類似度が閾値よりも大きいことを決定条件としている。このとき、決定データ 4 は、(leopard 1, 0.972) と、(cat 1, 0.891) と、(tiger 1, 0.779) と、(tiger 2, 0.697) と、(cheetah 2, 0.594) と、(cat 2, 0.567) とである。他の画像は類似度が閾値以下であるため、データ決定部 2 0 は他の画像を選択しない。

10

【 0 0 6 5 】

以上のように、類似度算出装置 1 は、1 つ以上の転移元データ候補 2 に基づいて学習を行うことによって分類器 1 0 6 を作成し、さらに、1 つの転移先データ 3 に基づいて、作成した分類器 1 0 6 を用いて推論を行い、推論の結果に基づいて、転移元データと当該転移元データに対応する類似度とを出力する。

【 0 0 6 6 】

*** 実施の形態 1 の効果の説明 ***

以上のように、本実施の形態に係る類似度算出装置 1 は、具体例として、まず、データ分布計算部 1 2 が、特徴ベクトル 1 0 1 S から算出した $(x, p(x))$ を 2 次元の特徴データ分布 1 0 2 S として出力する。次に、極値計算部 1 3 が、特徴データ分布 1 0 2 S が有する極小値及び極大値を算出する。次に、極値座標群 G^s の各極値と極値座標群 G^t の各極値とがどれだけ近い距離関係であるかを判定する。

20

そのため、本実施の形態によれば、特徴ベクトル 1 0 1 をそのまま用いる場合と比較して、特徴データ分布 1 0 2 を用いるために特徴を示す情報の次元を低くすることができ、さらに、転移先の特徴を有する極値に計算対象を絞るために類似度の計算負荷が低くなる。従って、本実施の形態によれば、比較的短い処理時間で、特徴データ分布 1 0 2 そのものを用いた場合に近い類似度を求めることができる。

30

【 0 0 6 7 】

また、本実施の形態に係る類似度算出装置 1 は、特徴データ分布 1 0 2 T の極値ごとに、転移元データ候補 2 に対応するクラスのいずれに所属すると考えられるかを示すクラス所属確率を算出し、各々のクラス所属確率に各々の重みパラメータ w_i を乗じた後に結合して類似度を求める。そのため、本実施の形態によれば、特徴データ分布 1 0 2 T の特定の極値に対する w_i を高く設定(具体例として、 $w_i > 1$)することにより特定の極値の影響を大きくすることができ、また、特定の極値の影響を小さくするために w_i を低く設定(具体例として、 $w_i < 1$)することができる。

従って、本実施の形態によれば、類似度の算出において、ユーザが指定する極値(特徴)を重視した類似度を算出することができる。

40

【 0 0 6 8 】

特許文献 1 が開示する技術では、画像の類似度を評価する際、既存の類似度算出方法を用いており、ユーザは処理負荷等を考慮して所望の類似性を検出することができる適切な方法を選択する必要がある。また、本技術によれば、類似の度合いを制御することが困難である。具体的には、本技術によれば、指紋認証のように環境ノイズのみを考慮する完全一致に近い動作をさせることと、画像セグメンテーション又は行動認識のように個人差及び個体差等を無視してラベル付けするあいまい検索に近い動作をさせること等、単一の評価関数を用いて複数種類の類似性を評価する処理を実施することができない。

しかしながら、前述の通り、本実施の形態に係る類似度算出装置 1 によれば、類似度に

50

係る計算負荷は低く、また、重みを活用することにより複数種類の類似性を評価することができる。

【0069】

他の構成

<変形例1>

特徴データ分布102である $(x, p(x))$ は、転移元データ候補2及び転移先データ3の特徴を十分に反映しており、かつ、波形状に近似されていればどのようなものでも構わない。

具体例として、 $p(x)$ に非特許文献1で示されているBoF (Bag-of-Features)を用いる方法がある。BoFは、画像等の元データから抽出した特徴ベクトル群をクラスタリングし、クラスタリングに係る各クラスIDと出現頻度(ヒストグラム)とを生成する手法である。BoFを用いる場合の具体例として、特徴抽出部11は、画像を受け取ると、非特許文献2に示されるHOG (Histograms of Oriented Gradients)又は非特許文献3に示されるSIFT (Scaled Invariance Feature Transform)を特徴ベクトル101として出力する。その後、データ分布計算部12は、BoFを用いて x にクラスIDを設定し、 $p(x)$ に当該クラスIDに対応するクラスの出現頻度を設定して、特徴データ分布102として $(x, p(x))$ を出力する。

10

【0070】

[非特許文献1]

Csurka, G., Dance, C.R., Fan, L., Willamowski, J. and Bray, C.: Visual Categorization with Bags of Keypoints, ECCV (European Conference on Computer Vision) International Workshop on Statistical Learning in Computer Vision, pp. 1-22 (2004).

20

[非特許文献2]

Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, 2005 IEEE (Institute of Electrical and Electronics Engineers) Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886-893, doi: 10.1109/CVPR.2005.177 (2005).

30

[非特許文献3]

Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints, Int. J. Comput. Vision, Vol. 60, No. 2, pp. 91-110 (2004).

【0071】

<変形例2>

転移元データ候補2がセンサ等から取得した多次元時系列データである場合、特徴抽出部11は、具体例として、複数のサンプリングデータを時系列順に入力し、主成分分析により各サンプリングデータを要約(次元削減)した値を特徴ベクトル101Sとする。

本変形例において、データ分布計算部12は、特徴ベクトル101Sを時系列順に結合することにより特徴データ分布102Sを生成してもよい。なお、当該特徴データ分布102Sは極値が算出可能なデータ分布であるものとする。

40

【0072】

<変形例3>

図10は、本変形例に係る類似度算出装置1のハードウェア構成例を示している。

50

類似度算出装置 1 は、本図に示すように、プロセッサ 5 1 とメモリ 5 2 と補助記憶装置 5 3 との少なくとも 1 つに代えて、処理回路 5 8 を備える。

処理回路 5 8 は、類似度算出装置 1 が備える各部の少なくとも一部を実現するハードウェアである。

処理回路 5 8 は、専用のハードウェアであってもよく、また、メモリ 5 2 に格納されるプログラムを実行するプロセッサであってもよい。

【0073】

処理回路 5 8 が専用のハードウェアである場合、処理回路 5 8 は、具体例として、単一回路、複合回路、プログラム化したプロセッサ、並列プログラム化したプロセッサ、ASIC (ASIC は Application Specific Integrated Circuit)、FPGA (Field Programmable Gate Array) 又はこれらの組み合わせである。

類似度算出装置 1 は、処理回路 5 8 を代替する複数の処理回路を備えてもよい。複数の処理回路は、処理回路 5 8 の役割を分担する。

【0074】

類似度算出装置 1 において、一部の機能が専用のハードウェアによって実現されて、残りの機能がソフトウェア又はファームウェアによって実現されてもよい。

【0075】

処理回路 5 8 は、具体例として、ハードウェア、ソフトウェア、ファームウェア、又はこれらの組み合わせにより実現される。

プロセッサ 5 1 とメモリ 5 2 と補助記憶装置 5 3 と処理回路 5 8 とを、総称して「プロセッシングサーキットリー」という。つまり、類似度算出装置 1 の各機能構成要素の機能は、プロセッシングサーキットリーにより実現される。

他の実施の形態に係る類似度算出装置 1 についても、本変形例と同様の構成であってもよい。

【0076】

実施の形態 2 .

以下、主に前述した実施の形態と異なる点について、図面を参照しながら説明する。

本実施の形態の概要を説明する。本実施の形態に係る類似度算出装置 1 は、学習モデルを作成する際に、特徴データ分布 102 S の極小値及び特徴データ分布 102 T の極大値、又は、特徴データ分布 102 S の極大値及び特徴データ分布 102 T の極小値が誤ってマッチングすることを防ぐことを目的として極値を表現する座標系を変換する。これにより、学習モデルが極大値と極小値とを識別しやすくなる。

【0077】

図 11 は、極値を表現する座標系を変換するイメージの具体例を示している。本図の (a) は座標系を変換する前のイメージを示しており、(b) は、(a) に示される極値を高次元空間に射影することにより、極値の座標系を変換した後のイメージを示している。なお、当該高次元空間の次元数は 3 に限られない。(b) の極大集合は (a) に示される極大値が射影された点の集合を示し、極小集合は (a) に示される極小値が射影された点の集合を示す。(a) における誤マッチングは、転移先データ分布 D1 の 1 つの極小値と転移元データ分布 D3 の 1 つの極大値とが近接していることを示す。誤マッチングに係る極小値と極大値とは、(b) において分離超平面を境界にして明確に分離している。

【0078】

*** 構成の説明 ***

図 12 は、本実施の形態に係る類似度算出装置 1 の構成例を示している。類似度算出装置 1 は、本図に示すように、実施の形態 1 に係る類似度算出装置 1 が備える構成要素に加え、座標変換部 21 を備える。

【0079】

座標変換部 21 は、極値データ 103 S を用いて変換座標群 C^S を生成し、生成した変換座標群 C^S を含む変換極値データ 107 S を出力する。変換極値データ 107 S は、変

10

20

30

40

50

換座標群 C^S と、データ ID^S と、極値の数 E^S とを含み、転移元変換座標群とも呼ばれる。変換座標群 C^S は、極値座標群 G^S が含む各座標の座標系を変換した座標から成る集合である。極値座標群 G^S が含む各座標は変換前転移元極値とも呼ばれる。座標変換部 21 は、転移元極値として、変換前転移元極値の次元よりも次元が高い空間に変換前転移元極値を射影した座標を求める。変換座標群 C^S は極値座標群 G^S の下位概念でもある。変換座標群 C^S が含む各座標は転移元極値でもある。

また、座標変換部 21 は、極値データ 103T を用いて変換座標群 C^t を生成し、変換した変換座標群 C^t を含む変換極値データ 107T を出力する。変換極値データ 107T は、変換座標群 C^t と、極値の数 E^t とを含み、転移先変換座標群とも呼ばれる。変換座標群 C^t は、極値座標群 G^t が含む各座標の座標系を変換した座標から成る集合である。極値座標群 G^t が含む各座標は変換前転移先極値とも呼ばれる。座標変換部 21 は、転移先極値として、変換前転移先極値の次元よりも次元が高い空間に変換前転移先極値を射影した座標を求める。変換座標群 C^t は極値座標群 G^t の下位概念でもある。変換座標群 C^t が含む各座標は転移先極値でもある。

10

【0080】

変換座標群 C^S が含む各座標は、[数5]のように定義される。[数5]は変換座標群 C^S の具体例でもある。座標変換部 21 は、本例において、極値座標群 G^S が含む各極大値を 1 に変換し、極値座標群 G^S が含む各極小値を -1 に変換し、変換した座標を極値座標群 G^S が含む各座標に追加することによって変換座標群 C^S を得る。また、本例において変換座標群 C^S が含む各極値は (x_i^S, y_i^S, z_i) と表される。変換座標群 C^t が含む各座標は、[数5]と同様に定義される。

20

【0081】

【数5】

$$\left\{ (x_i^S, y_i^S, z_i) \in \mathbb{R}^3 \left| \begin{array}{l} |\{(x_i^S, y_i^S)\}| = n, i = 1, \dots, n, S \in \mathcal{S}, \\ z_i \in \{+1, -1\}, \\ z_i = \begin{cases} +1, & (x_i^S, y_i^S) \text{が極大} \\ -1, & (x_i^S, y_i^S) \text{が極小} \end{cases} \end{array} \right. \right\}$$

30

【0082】

動作の説明

実施の形態 1 に係る類似度算出装置 1 の動作との差異を主に説明する。

【0083】

<学習フェーズ>

図 13 は、類似度算出装置 1 の学習フェーズにおける動作の一例を示すフローチャートである。図 12 及び本図を用いて学習フェーズを説明する。

【0084】

(ステップ S201: 座標変換処理)

座標変換部 21 は、極値データ 103S を入力として受け取り、受け取った極値データ 103S を用いて変換極値データ 107S を生成し、生成した変換極値データ 107S を出力する。

40

【0085】

(ステップ S104: グルーピング処理)

本ステップの処理は、実施の形態 1 の学習フェーズに係るグルーピング処理における極値データ 103S を変換極値データ 107S に読み替えた処理である。

【0086】

<推論フェーズ>

図 14 は、類似度算出装置 1 の推論フェーズにおける動作の一例を示すフローチャートである。図 12 及び本図を用いて推論フェーズを説明する。

50

【 0 0 8 7 】

(ステップ S 2 1 1 : 座標変換処理)

座標変換部 2 1 は、極値データ 1 0 3 T を入力として受け取り、受け取った極値データ 1 0 3 T を用いて変換極値データ 1 0 7 T を生成し、生成した変換極値データ 1 0 7 T を出力する。

【 0 0 8 8 】

(ステップ S 1 1 4 : モデル決定処理)

本ステップの処理は、実施の形態 1 の推論フェーズに係るモデル決定処理における極値データ 1 0 3 T を変換極値データ 1 0 7 T に読み替えた処理である。

【 0 0 8 9 】

*** 実施の形態 2 の効果の説明 ***

以上のように、本実施の形態に係る類似度算出装置 1 は、具体例として、特徴データ分布 1 0 2 の各極値に対して、各極値が極大値と極小値とのいずれであるかを表すパラメータ z を導入して各極値の座標系を変換することにより、極大値の集合と極小値の集合との間の距離を離す。そのため、本実施の形態に係る類似度算出装置 1 によれば、転移先の極大値と転移元の極小値、あるいは転移先の極小値と転移元の極大値が誤ってマッチングされるリスクが下がる。

従って、本実施の形態に係る類似度算出装置 1 によれば、実施の形態 1 に係る類似度算出装置 1 と比較してより高い精度の類似度を求めることができる。

【 0 0 9 0 】

*** 他の構成 ***

< 変形例 4 >

変換座標群 C^s の各座標は、[数 6] に示すように、x_i^s と y_i^s とのそれぞれに z_i を乗算した形式でもよい。

【 0 0 9 1 】

【 数 6 】

$$\left\{ \begin{array}{l} (z_i \cdot x_i^s, z_i \cdot y_i^s, z_i) \in \mathbb{R}^3 \\ \left. \begin{array}{l} |\{(x_i^s, y_i^s)\}| = n, i = 1, \dots, n, s \in S, \\ z_i \in \{+1, -1\}, \\ z_i = \begin{cases} +1, & (x_i^s, y_i^s) \text{が極大} \\ -1, & (x_i^s, y_i^s) \text{が極小} \end{cases} \end{array} \right\} \end{array} \right.$$

【 0 0 9 2 】

[数 5] に示すように座標系を変換した場合においても、極大値及び極小値が 3 次元ユークリッド空間座標系 (R³) におけるユークリッド距離が少なくとも 2 の距離をもって配置されることが保証される。

本変形例では、座標系の変換後において極大値及び極小値がさらに大きい距離をもって配置されやすくするために、x 及び y それぞれに z を乗算する。極大値 (x_i^s, y_i^s) と極小値 (x_j^s, y_j^s) とが与えられた場合において、座標系の変換後における両者間の距離は [数 7] が示す通りになる。

【 0 0 9 3 】

【 数 7 】

$$\begin{aligned} & \sqrt{(x_i^s - (-x_j^s))^2 + (y_i^s - (-y_j^s))^2 + (+1 - (-1))^2} \\ & = \sqrt{(x_i^s + x_j^s)^2 + (y_i^s + y_j^s)^2 + 4} \end{aligned}$$

【 0 0 9 4 】

10

20

30

40

50

[数7]について考える。特徴データ分布102の変数 x が $[0, 1]$ により正規化されている場合、 $x \geq 0$ である。 y はヒストグラムであるので $y \geq 0$ である。よって、 $x_i^S - x_j^S < x_i^S + x_j^S$ 及び $y_i^S - y_j^S < y_i^S + y_j^S$ が成り立つ。そのため、極大値及び極小値が少しでも離れていれば、変換後の極大値及び極小値が少なくとも2より離れやすくなることが保証される。

よって、推論にSVM (Support Vector Machine) のようなアルゴリズムを用いる場合、分離超平面により、極大値と極小値とを識別しやすくなる。

従って、本変形例によれば、特徴データ分布102Tの極大値を入力したときに特徴データ分布102Sの極小値が誤って推論される確率と、特徴データ分布102Tの極小値を入力したときに特徴データ分布102Sの極大値が誤って推論される確率とを下げる

10

【0095】

実施の形態3.

以下、主に前述した実施の形態と異なる点について、図面を参照しながら説明する。

実施の形態1と本実施の形態との主な差異は、類似度算出の対象となる転移元データ候補2の範囲を広げるために、モデル作成部16が許容範囲内の極値の数それぞれに対応するデータセット105を用いる点である。当該許容範囲はデータセット許容範囲とも呼ばれる。データセット許容範囲は、データセット105それぞれに対応する転移元極値の数の周囲の範囲を示す。

【0096】

20

図15は、モデル作成部16がモデルを作成する際、許容範囲内の極値の数それぞれに対応するデータセット105をモデル作成部16が選択するイメージの具体例を示している。モデル作成部16は、データセット105に対応する極値の数を基準としてデータセット105を選択する。なお、説明の便宜上本図は特徴データ分布102に基づいて学習モデルが決定される様子を示しているが、実際には、特徴データ分布102Tに対応する極値データ103Tと、特徴データ分布102Sそれぞれに対応するデータセット105とに基づいて学習モデルは決定される。

本図において、転移先データ分布D1の極値の数は4であり、転移元データ分布D2の極値の数は3であり、転移元データ分布D3の極値の数は5である。(a)は許容範囲を考慮せずにモデル作成部16がデータセット105を選択するイメージを示している。(a)において転移先データ分布D1の極値の数と同じ極値の数を有する転移元データ分布がないので、モデル作成部16はデータセット105を選択することができない。一方、(b)は許容範囲を考慮してモデル作成部16がデータセット105を選択するイメージを示している。モデル作成部16は±1の範囲において極値の数を許容する。転移元データ分布D2の極値の数及び転移元データ分布D3の極値の数は共に転移先データ分布D1の極値の数±1の範囲内であるため、モデル作成部16は転移元データ分布D2及び転移元データ分布D3それぞれに対応するデータセット105を選択する。

30

【0097】

構成の説明

図16は、本実施の形態に係る類似度算出装置1の構成例を示している。本実施の形態に係る類似度算出装置1の実施の形態1に係る類似度算出装置1に対する主な差異は、モデル作成部16が、データセット105の代わりにデータセット群111を受け取る点である。データセット群111は、複数のデータセット105から成る集合である。モデル作成部16は、複数のデータセット105を受け取り、受け取った複数のデータセット105を用いてデータセット群111を生成してもよい。

40

【0098】

データセット群111は、具体例としてデータセット群111-(n±b)である。データセット群111-(n±b)は、n-bからn+bまでの範囲内の極値の数それぞれに対応するデータセット105から成る集合である。ここで、bは許容範囲に対応するパラメータであり、b ≥ 0かつb ∈ Zである。なお、n-bからn+bまでの範囲内のい

50

れか1つ以上の極値の数に対応するデータセット105はなくてもよい。

【0099】

[数8]は、データセット群111 - (n ± b)の具体例を示している。各記号の意味は、[数2]と同様である。なお、データセット群111 - (n ± b)を集合 $g_{n \pm b}$ と表現している。

データセット群111は、データセット許容範囲内の極値の数それぞれに対応する転移元極値を含み、また、データセット105の下位概念でもある。データセット群111 - (n ± b)において、データセット105に対応する転移元極値の数はnであり、データセット許容範囲はn - bからn + bまでである。

【0100】

【数8】

$g_{n \pm b}$

$$= \{((x_i^s, y_i^s), c^s) | n - b \leq |\{(x_i^s, y_i^s)\}| \leq n + b, i = 1, \dots, n, s \in S, b \geq 0, b \in \mathbb{Z}\}$$

10

【0101】

モデル作成部16は、データセット105 - nの代わりにデータセット群111をデータセット記憶部15から取得し、取得したにデータセット群111を用いて分類器106を作成する。モデル作成部16は、具体例として、データセット群111 - (n ± b)を用いて分類器106 - nを作成する。

20

【0102】

動作の説明

本実施の形態に係る類似度算出装置1の動作を示すフローチャートは、実施の形態1に係る類似度算出装置1の動作を示すフローチャートと同じである。以下、実施の形態1に係る類似度算出装置1の動作との差異を主に説明する。

【0103】

<学習フェーズ>

(ステップS105:モデル作成処理)

本処理は、実施の形態1の学習フェーズに係るモデル作成処理におけるデータセット105をデータセット群111 - (n ± b)に読み替えた処理である。即ち、モデル作成部16は、分類器106 - nを作成する際にデータセット群111 - (n ± b)を用いる。

30

【0104】

以降の処理は実施の形態1と同様である。なお、モデル作成部16は、データセット群111 - (n ± b)を用いて学習した学習モデルを、分類器106 - nとして分類器記憶部17に保存する。

【0105】

なお、訓練データをデータセット群111 - (n ± b)とした場合、分類器106 - (n - b)から分類器106 - (n + b)までの学習において同じ極値座標群 $G^s \{(x_i^{s1}, y_i^{s1})\}$ (但し、 $|(x_i^{s1}, y_i^{s1})| = n, i = 1, \dots, n, s1 \in S$)が使われる。そのため、 $\{(x_i^{s1}, y_i^{s1})\}$ と $\{(x_i^t, y_i^t)\}$ との類似度が高い場合、 $\{(x_i^{s1}, y_i^{s1})\}$ を用いて生成されたいずれの分類器106においても、 $\{(x_i^t, y_i^t)\}$ に対応するクラス所属確率であって、s1に対応するクラスへのクラス所属確率が高くなる。その結果、s1に対応するクラスをc1としたとき、類似度score(c1)の値が高くなる。そのため、学習モデルの学習に使用する訓練データを選定する基準である極値の数に許容範囲を設けても、転移先データ3の各極値を適切に分類することができるものと考えられる。

40

【0106】

実施の形態3の効果の説明

以上のように、本実施の形態に係る類似度算出装置1は、類似度の算出に用いられる学習モデルの学習に使用する訓練データを選定する基準である極値の数に対してデータセッ

50

ト許容範囲を設ける。具体例として、類似度算出装置 1 は、パラメータ b を導入することによって類似度算出の対象となる転移元データ候補 2 の候補範囲を広げる。

そのため、類似度算出装置 1 は、極値の数は異なるものの特徴データ分布 102S と特徴データ分布 102T とが互いに大局的に類似している場合において、両データ分布の類似性を判定することができる。

【0107】

図 17 は、特徴データ分布 102S と特徴データ分布 102T とが大局的に類似している場合における類似度算出装置 1 の動作を具体的に説明する図である。本図において、転移先データ分布 D1 と転移元データ分布 D2 とのそれぞれの極値の数は異なるが、 x の値がいずれであっても両データ分布間で x の値に対応する y の値は近接している。即ち、両データ分布は大局的に類似している。また、近傍領域 R5 及び近傍領域 R6 は転移先データ分布 D1 の極値に近接する領域であり、両領域において、転移先データ分布 D1 の極値と、転移元データ分布 D2 の極値とが近接している。

10

類似度算出装置 1 は、転移先データ分布 D1 の極値の近傍に位置する転移元データ分布 D2 の極値が多いほど、転移先データ分布 D1 に対応するクラスを転移元データ分布 D2 に対応するクラスにより高い確率で分類する。

【0108】

また、類似度算出装置 1 は、パラメータ b を用いて類似の度合いを制御することができる。具体例として、類似度算出装置 1 は、パラメータ b の値が小さいほど完全一致検索に近い動作をし、パラメータ b の値が大きいほどあいまい検索に近い動作をする。

20

【0109】

他の構成

<変形例 5>

データセット許容範囲には対称性がなくてもよい。具体例として、モデル作成部 16 は、分類器 106 - n を生成する際に用いるデータセット群 111 として、データセット群 111 - ($n \pm b$) の代わりにデータセット群 111 - ($k = n, n + 1, \dots, n + b$) を用いる。データセット 111 - ($k = n, n + 1, \dots, n + b$) は、 n から $n + b$ までの範囲内の極値の数に対応するデータセット 105 である。

【0110】

<変形例 6>

データセット許容範囲は動的に設定されても構わない。

本変形例の具体例として、データセット 105 として、データセット 105 - 1 とデータセット 105 - 3 とデータセット 105 - 4 のみが求められている場合において、各データセット 105 と、各データセット 105 の両隣それぞれにおいて最も近接している 1 つずつのデータセット 105 とを選択して各データセット 105 に対応するデータセット群 111 を生成することを考える。このとき、データセット 105 - 3 を基準とすると、データセット 105 - 1 とデータセット 105 - 4 とが最も近接しているデータセット 105 として選択される。このように、モデル作成部 16 は、各極値の数に対応するデータセット 105 が存在するか否かを考慮してデータセット 105 を選定してデータセット群 111 を生成してもよい。

30

40

【0111】

実施の形態 4 .

以下、主に前述した実施の形態と異なる点について、図面を参照しながら説明する。

実施の形態 1 と本実施の形態との主な差異は、類似度算出の対象となる転移元データ候補 2 の範囲を広げるために、モデル決定部 18 が許容範囲内である極値の数それぞれに対応する分類器 106 を用いる点である。当該許容範囲はモデル許容範囲とも呼ばれる。モデル許容範囲は、転移先極値群が含む転移先極値の数の周囲の範囲を示す。

【0112】

図 18 は、許容範囲内の極値の数に対応する分類器 106 をモデル作成部 16 が選択するイメージの具体例を示している。本図の見方は図 15 の見方と同様である。モデル作成

50

部 1 6 は、ある極値の数を基準として少なくとも 1 つの分類器 1 0 6 を選択する。

本図において、(a) は許容範囲を考慮せずにモデル決定部 1 8 が分類器 1 0 6 を選択するイメージを示している。(a) において、転移先データ分布 D 1 の極値の数と同じ数の極値の数に対応する分類器 1 0 6 がいないため、モデル決定部 1 8 は分類器 1 0 6 を選択することができない。一方、(b) は許容範囲を考慮してモデル決定部 1 8 が分類器 1 0 6 を選択するイメージを示している。(b) において、許容範囲は 3 以上 5 以下であり、転移元データ分布 D 2 及び転移元データ分布 D 3 それぞれの極値の数は許容範囲内である。そのため、モデル作成部 1 6 は分類器 1 0 6 - 3 及び分類器 1 0 6 - 5 を選択し、特徴データ分布 1 0 2 - 3 に対応する分類器群 1 1 2 として、分類器 1 0 6 - 3 と分類器 1 0 6 - 4 と分類器 1 0 6 - 5 とから成る分類器群 1 1 2 を生成する。

10

【 0 1 1 3 】

*** 構成の説明 ***

図 1 9 は、本実施の形態に係る類似度算出装置 1 の構成例を示している。本実施の形態に係る類似度算出装置 1 の実施の形態 1 に係る類似度算出装置 1 に対する主な差異は、モデル決定部 1 8 が、分類器 1 0 6 - m の代わりに分類器群 1 1 2 を出力する点である。

【 0 1 1 4 】

分類器群 1 1 2 は、具体例として分類器群 1 1 2 - (m ± d) である。ここで、d は許容範囲に対応するパラメータであり、d > 0 かつ d ≤ Z である。分類器群 1 1 2 - (m ± d) は、分類器 1 0 6 - (m - d) から分類器 1 0 6 - (m + d) までの分類器 1 0 6 から成る集合である。なお、m - d から m + d まで範囲内のいずれか 1 つ以上の極値の数に対応する分類器 1 0 6 はなくてもよい。

20

【 0 1 1 5 】

モデル決定部 1 8 は、極値座標群 G^t に対応する分類器 1 0 6 の代わりに極値座標群 G^t に対応する分類器群 1 1 2 を生成し、推論データ 1 0 8 として、極値座標群 G^t と、極値座標群 G^t に対応する分類器群 1 1 2 とを含むデータを生成する。

モデル決定部 1 8 は、モデル許容範囲内の転移元極値の数それぞれに対応する学習モデルから成る学習モデル群を決定モデルとして決定する。

【 0 1 1 6 】

類似度算出部 1 9 は、分類器 1 0 6 の代わりに分類器群 1 1 2 が含む各分類器 1 0 6 を用いて少なくとも 1 つの特徴データ分布 1 0 2 S それぞれと特徴データ分布 1 0 2 T との類似度を求める。

30

【 0 1 1 7 】

*** 動作の説明 ***

本実施の形態に係る類似度算出装置 1 の動作を示すフローチャートは、実施の形態 1 に係る類似度算出装置 1 の動作を示すフローチャートと同じである。以下、実施の形態 1 に係る類似度算出装置 1 の動作との差異を主に説明する。

【 0 1 1 8 】

< 推論フェーズ >

(ステップ S 1 1 4 : モデル決定処理)

本処理は実施の形態 1 の推論フェーズに係るモデル決定処理における分類器 1 0 6 - m を分類器群 1 1 2 - (m ± d) に読み替えた処理である。即ち、モデル決定部 1 8 は、分類器 1 0 6 - m の代わりに分類器群 1 1 2 - (m ± d) を特定し、特定した分類器群 1 1 2 - (m ± d) と極値座標群 G^t とを含む推論データ 1 0 8 を出力する。

40

【 0 1 1 9 】

(ステップ S 1 1 5 : 類似度算出処理)

類似度算出部 1 9 は、分類器 1 0 6 - m の代わりに分類器群 1 1 2 - (m ± d) を用いて類似度を算出する。

【 0 1 2 0 】

以下、本実施の形態と実施の形態 3 とを組み合わせた場合における類似度算出処理を具体的に説明する。類似度算出部 1 9 は、クラス所属確率が分類器 1 0 6 において重複して

50

出現する可能性があることを考慮し、類似度 $score(c_j)$ を算出する方法を変更する。分類器 106 - k ($k = m - d, m - d + 1, \dots, m + d$) を用いた場合における特徴データ分布 102 T のラベル c_j との類似度サブスコア $score_k(c_j)$ を [数 9] に示すように定義する。[数 9] の記号の説明は、[数 4] の記号の説明と同じである。

【0121】

【数 9】

$$score_k(c_j) = \sum_{i=1}^k (w_i \cdot p_{c_j}^i)$$

10

【0122】

類似度算出部 19 は、複数の分類器 106 - k が同一のラベルに分類することができる場合、当該複数の分類器 106 - k それぞれについて、k の値が当該ラベルに対応する極値の数に近いほど重視する。即ち、類似度算出部 19 は、k の値が当該極値の数に近いサブスコアほど重視して類似度を算出する。

【0123】

[数 10] は、類似度 $score(c_j)$ の具体例を示している。ここで、右辺の上段において、k は、ラベル c_j に分類することができる分類器 106 の中で、類似度算出部 19 が最も重視する 1 つの分類器 106 を特定する変数である。k' は、k 以外の値を示し、ラベル c_j に分類することができる分類器 106 を特定する変数である。ラベル c_j に分類することができる分類器 106 が分類器 106 - k を除いて複数存在する場合に、k' は複数の値から成る。fp(d) は、具体例として、d = 0 で最大となり、かつ、d と反比例する値を返す関数である。具体例として、 $fp(d) = 1 / (1 + d)^2$ が挙げられる。ここで、d はラベル c_j に対応する極値の数から k' の各要素までの距離を示す変数である。なお、k' が複数の値から成る場合に、k' の各要素に対応する fp(d) が用意される。

20

具体例として、ラベル c_j に対応する極値の数が m であり、かつ、分類器 106 - k ($k = m - d, m - d + 1, \dots, m + d$) のいずれもラベル c_j に分類することができる場合を考える。この場合において、 $k = m$ であり、 $k' = \{m - d, \dots, m - 1, m + 1, \dots, m + d\}$ である。即ち、 $score_k(c_j)$ は分類器 106 - m を用いて求めた類似度であり、 $score_{k'}(c_j)$ は分類器 106 - m 以外の分類器 106 を用いて求めた類似度を用いて求めた値である。

30

【0124】

【数 10】

$score(c_j)$

$$= \begin{cases} score_k(c_j) + fp(d) \cdot score_{k'}(c_j) & , \quad (score_k(c_j) = score_{k'}(c_j)) \\ score_k(c_j) & , \quad (otherwise) \end{cases}$$

40

【0125】

*** 実施の形態 4 の効果の説明 ***

以上のように、本実施の形態に係る類似度算出装置 1 は、類似度の算出に用いられる学習モデルを選定する基準である極値の数に対してモデル許容範囲を設ける。具体例として、類似度算出装置 1 は、パラメータ d を導入することによって類似度算出の対象となる転移元データ候補 2 の範囲を広げる。

そのため、類似度算出装置 1 は、実施の形態 3 に係る類似度算出装置 1 と同様に、図 17 に示すような場合において、両データ分布の類似性を判定することができる。

50

また、類似度算出装置 1 は、パラメータ d を用いて類似の度合いを制御することができる。具体例として、類似度算出装置 1 は、パラメータ d の値が小さいほど完全一致検索に近い動作をし、パラメータ d の値が大きいほどあいまい検索に近い動作をする。

【0126】

他の構成

<変形例 7>

モデル許容範囲には対称性がなくてもよい。具体例として、モデル決定部 18 は、極値の数が m である特徴データ分布 102T に対応する分類器群 112 として、分類器群 112 - ($m \pm d$) の代わりに分類器群 112 - ($k = m, m + 1, \dots, m + d$) を決定してもよい。分類器群 112 - ($k = m, m + 1, \dots, m + d$) は、 m から $m + d$ までの範囲内の極値の数に対応する分類器 106 から成る分類器群 112 である。

10

【0127】

<変形例 8>

モデル許容範囲は動的に設定されてもよい。

本変形例の具体例として、データセット 105 として、データセット 105 - 1 とデータセット 105 - 3 とデータセット 105 - 4 のみが求められている場合において、分類器群 112 として、各データセット 105 に対応する各分類器 106 と、各データセット 105 の両隣それぞれにおいて最も近接している 1 つずつのデータセット 105 に対応する分類器 106 とを選択する場合を考える。このとき、極値の数として 3 を基準とすると、分類器 106 - 1 と分類器 106 - 3 と分類器 106 - 4 とが選択される。このように、モデル決定部 18 は、各極値の数に対応する分類器 106 が存在するか否かを考慮して分類器 106 を選定してもよい。

20

【0128】

他の実施の形態

前述した各実施の形態の自由な組み合わせ、あるいは各実施の形態の任意の構成要素の変形、もしくは各実施の形態において任意の構成要素の省略が可能である。

【0129】

また、実施の形態は、実施の形態 1 から 4 で示したものに限定されるものではなく、必要に応じて種々の変更が可能である。フローチャート等を用いて説明した手順は、適宜変更されてもよい。

30

【符号の説明】

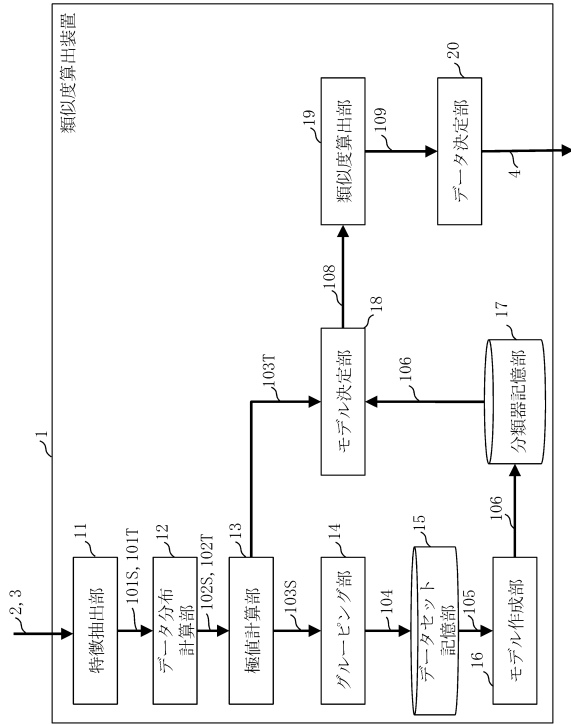
【0130】

1 類似度算出装置、2 転移元データ候補、3 転移先データ、4 決定データ、11 特徴抽出部、12 データ分布計算部、13 極値計算部、14 グルーピング部、15 データセット記憶部、16 モデル作成部、17 分類器記憶部、18 モデル決定部、19 類似度算出部、20 データ決定部、21 座標変換部、51 プロセッサ、52 メモリ、53 補助記憶装置、54 入出力 I/F、55 通信装置、58 処理回路、59 信号線、101, 101S, 101T 特徴ベクトル、102, 102S, 102T 特徴データ分布、103S, 103T 極値データ、104 学習データ、105 データセット、106 分類器、107S, 107T 変換極値データ、108 推論データ、109 類似度データ、111 データセット群、112 分類器群、D1 転移先データ分布、D2, D3 転移元データ分布、R1, R2, R3, R4, R5, R6 近傍領域。

40

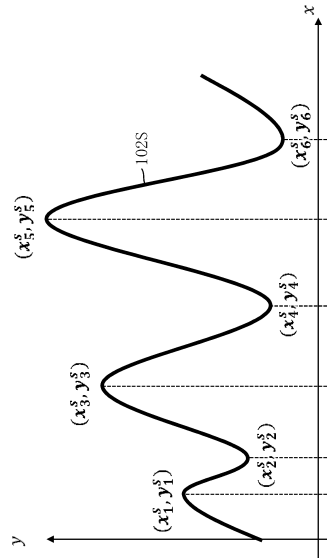
【図面】
【図 1】

図1



【図 2】

図2

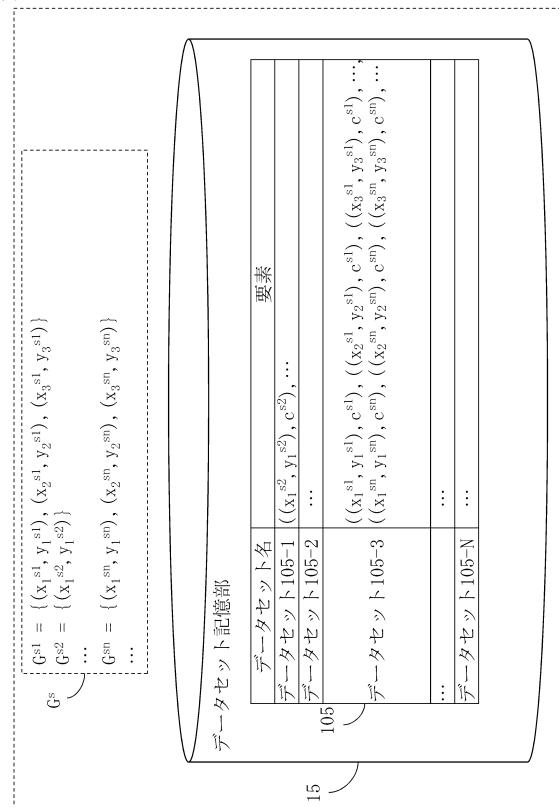


10

20

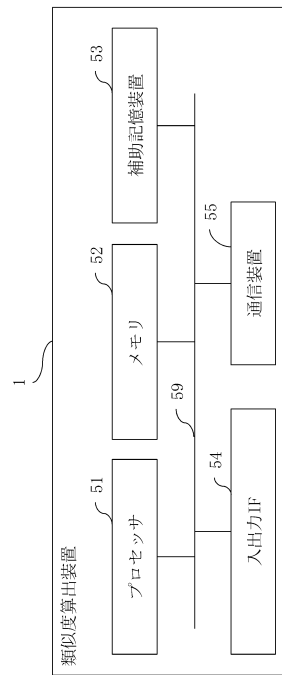
【図 3】

図3



【図 4】

図4



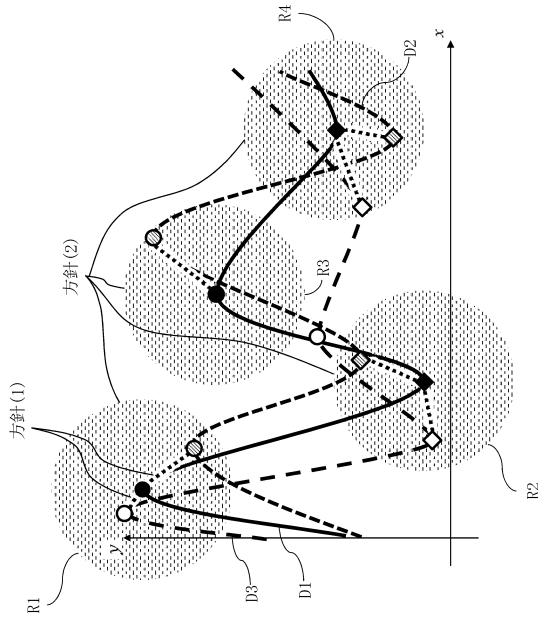
30

40

50

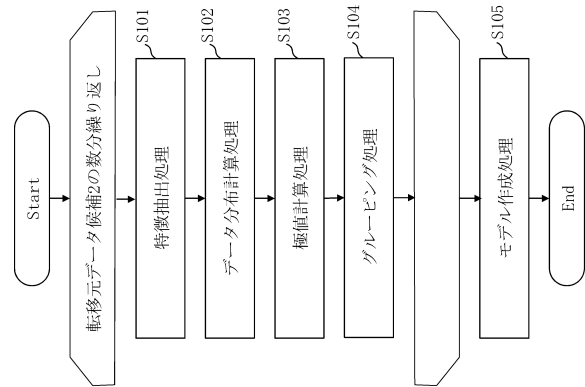
【 図 5 】

図5



【 図 6 】

図6

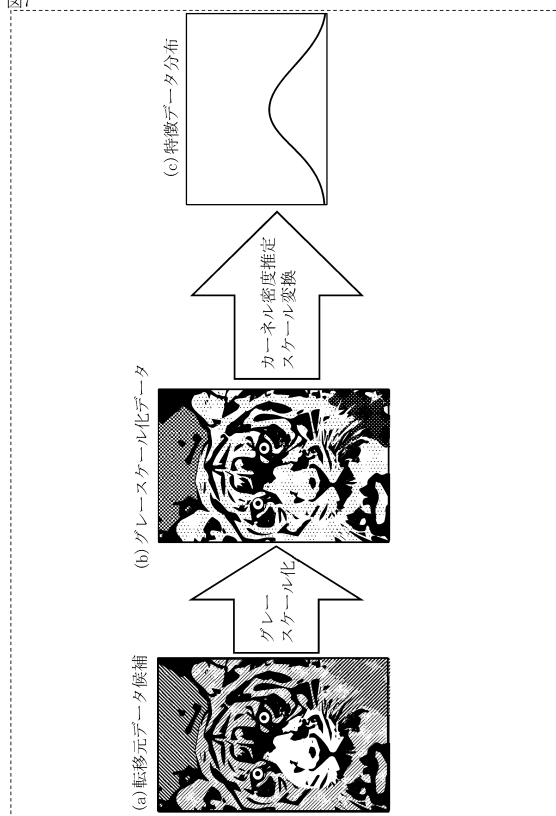


10

20

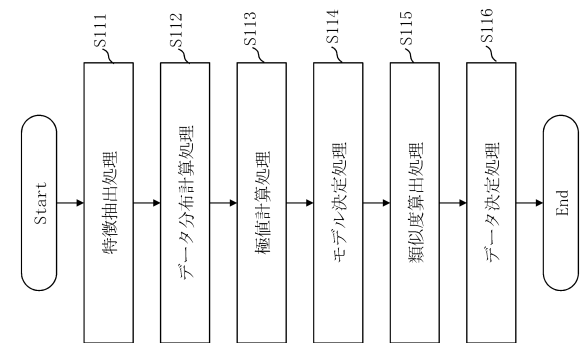
【 図 7 】

図7



【 図 8 】

図8



30

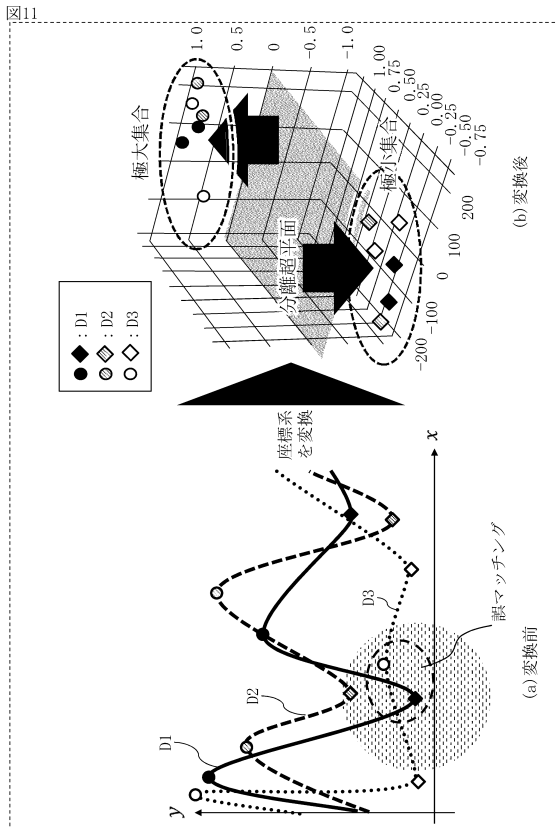
40

50

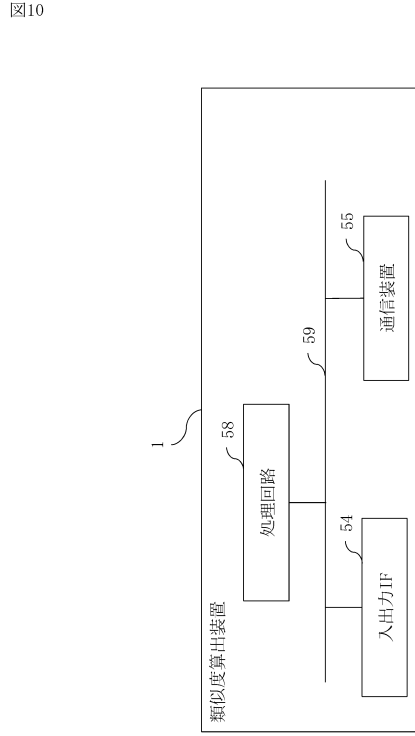
【 図 9 】



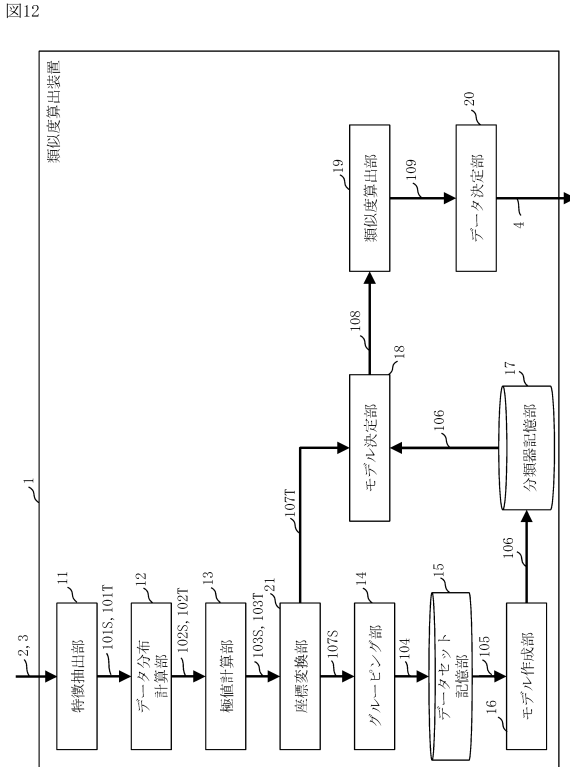
【 図 11 】



【 図 10 】



【 図 12 】



10

20

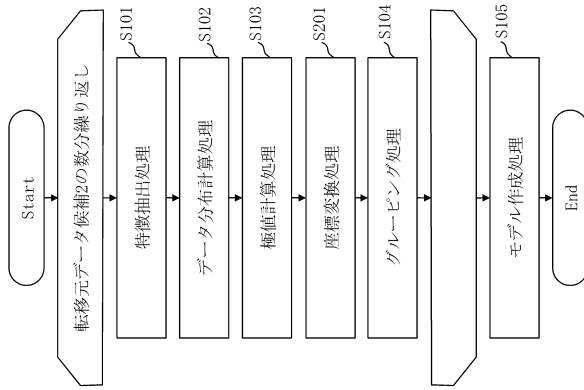
30

40

50

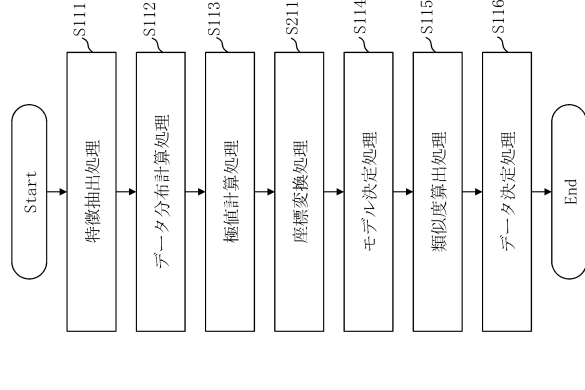
【 図 1 3 】

図13



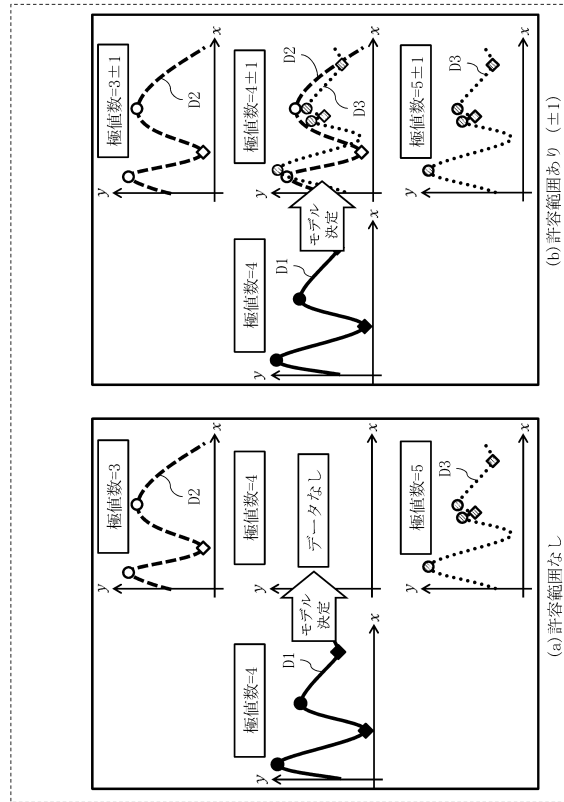
【 図 1 4 】

図14



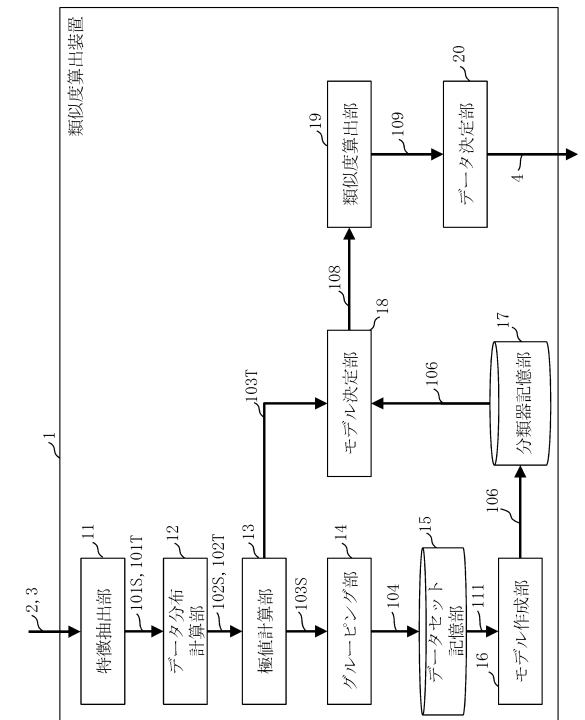
【 図 1 5 】

図15



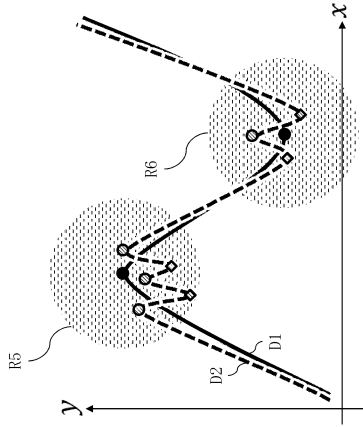
【 図 1 6 】

図16



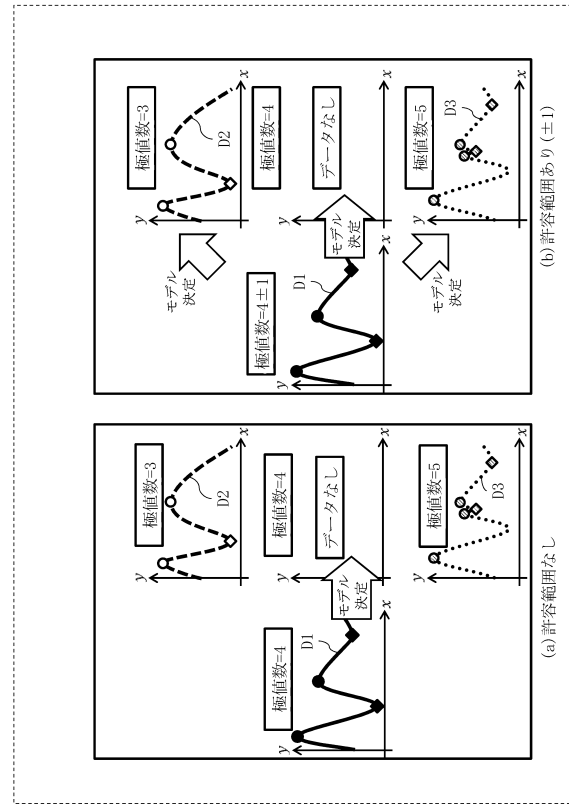
【 図 17 】

図17



【 図 18 】

図18

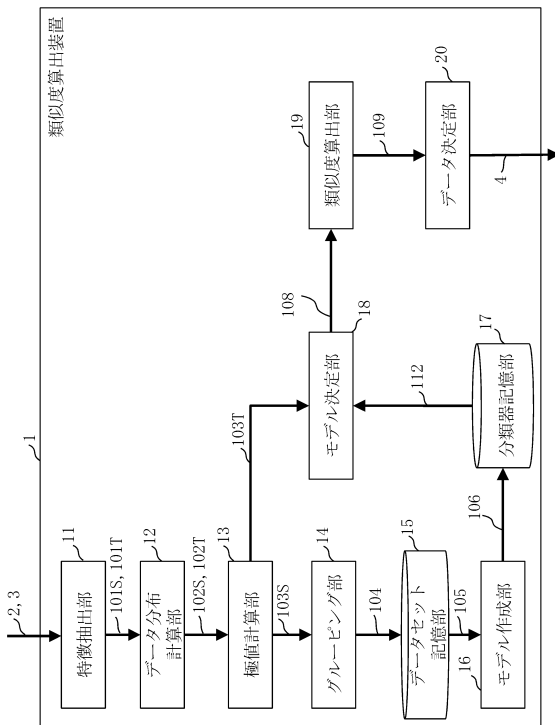


10

20

【 図 19 】

図19



30

40

50

フロントページの続き

- (56)参考文献 特開 2017 - 142739 (JP, A)
特開 2019 - 86979 (JP, A)
米国特許出願公開第 2003 / 0202692 (US, A1)
特開 2017 - 11523 (JP, A)
特開 2005 - 44033 (JP, A)
- (58)調査した分野 (Int.Cl., DB名)
G06N 20 / 00