

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年10月20日(20.10.2022)



(10) 国際公開番号

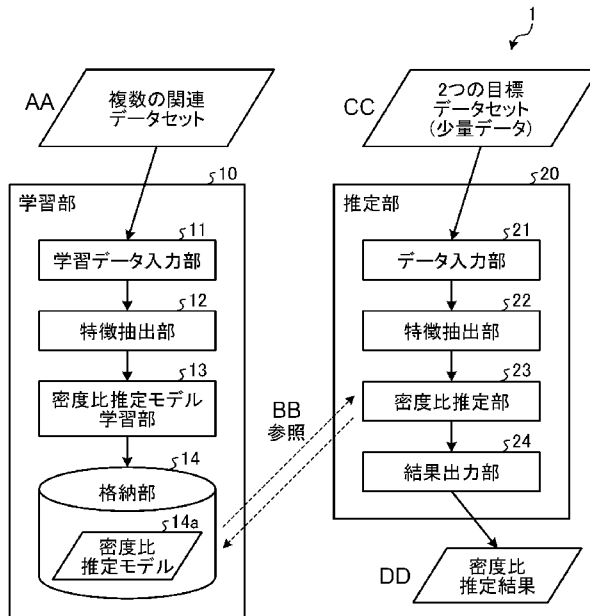
WO 2022/219686 A1

- (51) 国際特許分類:
G06N 20/00 (2019.01)
- (21) 国際出願番号: PCT/JP2021/015224
- (22) 国際出願日: 2021年4月12日(12.04.2021)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 日本電信電話株式会社 (**NIPPON TELEGRAPH AND TELEPHONE CORPORATION**) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 熊谷 充敏 (**KUMAGAI, Atsutoshi**); 〒1808585 東京都武蔵野市緑町3丁目9-1 1 N T T 知的財産センタ内 Tokyo (JP).
- (74) 代理人: 特許業務法人酒井国際特許事務所 (**SAKAI INTERNATIONAL PATENT OFFICE**); 〒1000013 東京都千代田区霞が関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH,

(54) **Title:** DENSITY RATIO ESTIMATION DEVICE, DENSITY RATIO ESTIMATION METHOD, AND DENSITY RATIO ESTIMATION PROGRAM

(54) 発明の名称: 密度比推定装置、密度比推定方法および密度比推定プログラム

[図2]



10... LEARNING UNIT
11... LEARNING DATA INPUT UNIT
12, 22... FEATURE EXTRACTION UNIT
13... DENSITY RATIO ESTIMATION MODEL LEARNING UNIT
14... STORAGE UNIT
14a... DENSITY RATIO ESTIMATION MODEL
20... ESTIMATION UNIT
21... DATA INPUT UNIT
23... DENSITY RATIO ESTIMATION UNIT
24... RESULT OUTPUT UNIT
AA... PLURALITY OF RELATED DATA SETS
BB... REFERENCE
CC... TWO TARGET DATA SETS (SMALL AMOUNT OF DATA)
DD... DENSITY RATIO ESTIMATION RESULT

(57) **Abstract:** In the present invention, a learning data input unit (11) acquires data of a density ratio estimation processing target, and a related data set consisting of a plurality of related data pieces in which the configuration of a feature amount is identical but the value of the feature amount is different. When a density ratio estimation model learning unit (13) designates related data selected from each of two of the related data sets that were acquired as pseudo learning data or pseudo test data and inputs two pieces of the pseudo learning data into a density ratio estimation model (14a), the density ratio estimation model (14a) is learned using the pseudo learning data and the pseudo test data so that the density ratio between the two sets of related data is accurately estimated.

[続葉有]

KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,
MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ,
NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

- 一 国際調査報告 (条約第21条(3))
-

(57) 要約：学習データ入力部（11）が、密度比推定の処理対象のデータと特徴量の構成が同一であって該特徴量の値が異なる複数の関連データからなる関連データセットを取得する。密度比推定モデル学習部（13）が、取得された2セットの関連データセットのそれぞれから選択した関連データを疑似学習データまたは疑似テストデータとして、2つの該疑似学習データを密度比推定モデル（14a）に入力した場合に、2セットの関連データセット間の密度比を正確に推定するように、該疑似学習データと疑似テストデータとを用いて、密度比推定モデル（14a）を学習する。

明 細 書

発明の名称：

密度比推定装置、密度比推定方法および密度比推定プログラム

技術分野

[0001] 本発明は、密度比推定装置、密度比推定方法および密度比推定プログラムに関する。

背景技術

[0002] 近年、転移学習、異常検知、2標本検定、変化点検知、特徴選択、クラスタリング、Generative Adversarial Networks (GANs) 等の様々な機械学習の応用において現れる量である密度比を精度よく推定する技術が期待されている。もっとも naïve なアプローチとして、各密度関数をそれぞれにデータから推定した後にその比をとって密度比を推定する。

[0003] しかしながら、密度関数の推定は非常に困難であることが知られており、精度よく推定することは困難である。したがって、誤差を含む密度関数の推定量の比をとるとその誤差が増長されてしまう。そこで、密度を推定した後に比をとるかわりに、密度比をモデル化して、2つのデータセットから直接に密度比を推定する技術が開示されている（非特許文献1，2参照）。

先行技術文献

非特許文献

[0004] 非特許文献1：Kanamori T., Hido S., Sugiyama M., “A Least-squares Approach to Direct Importance Estimation” The Journal of Machine Learning Research 10, 2009年, pp.1391-1445

非特許文献2：Sugiyama M., Nakajima S., Kashima H., Buenau P., Kawanabe M., “Direct Importance Estimation with Model Selection and Its Application to Covariate Shift Adaptation”, In NeurIPS, 2007年

発明の概要

発明が解決しようとする課題

[0005] しかしながら、従来の技術は大量のデータを必要とするものの、大量のデータを用意することが困難な場合がある。例えば、新規ユーザや新規システム等の新しいデータソースから得られるデータセットで密度比を推定したい場合には、直ちに大量のデータを用意することは困難である。また、治験において薬の効果を確認するための2標本検定や、衝突試験において車の構造変化の効果を確認するための2標本検定等に用いるために密度比を推定したい場合には、コストが高いために大量のデータを用意することは困難である。

[0006] 本発明は、上記に鑑みてなされたものであって、大量のデータを必要とせずに、高精度に密度比推定を行うことを目的とする。

課題を解決するための手段

[0007] 上述した課題を解決し、目的を達成するために、本発明に係る密度比推定装置は、密度比推定の処理対象のデータと特徴量の構成が同一であって該特徴量の値が異なる複数の関連データからなる関連データセットを取得する取得部と、取得された2つの前記関連データセットのそれぞれから選択した関連データを疑似学習データまたは疑似テストデータとして、2つの該疑似学習データをモデルに入力した場合に、2つの前記関連データセット間の密度比を正確に推定するように、該疑似学習データと疑似テストデータとを用いて、前記モデルを学習する学習部と、を有することを特徴とする。

発明の効果

[0008] 本発明によれば、大量のデータを必要とせずに、高精度に密度比推定を行うことが可能となる。

図面の簡単な説明

[0009] [図1]図1は、密度比推定装置の概要を説明するための図である。

[図2]図2は、密度比推定装置の概略構成を例示する模式図である。

[図3]図3は、密度比推定モデル学習部の処理を説明するための図である。

[図4]図4は、学習処理手順を示すフローチャートである。

[図5]図5は、推定処理手順を示すフローチャートである。

[図6]図6は、密度比推定プログラムを実行するコンピュータを例示する図である。

発明を実施するための形態

[0010] 以下、図面を参照して、本発明の一実施形態を詳細に説明する。なお、この実施形態により本発明が限定されるものではない。また、図面の記載において、同一部分には同一の符号を付して示している。

[0011] [密度比推定装置の概要]

まず、図1は、密度比推定装置の概要を説明するための図である。密度比推定装置は、次式(1)に示す、2つの確率密度関数 $P_{nu}(x)$ 、 $P_{de}(x)$ の比 $r(x)$ を推定する密度比推定を行う。

[0012] [数1]

$$r(x) = \frac{p_{nu}(x)}{p_{de}(x)} \quad \dots (1)$$

[0013] 本実施形態の密度比推定装置は、少量のデータを、密度比を推定したい目標データセットとして、複数の関連するデータセットを活用して、目標データセットから精度よく密度比推定を行う。具体的には、図1に示すように、密度比推定装置は、学習フェーズにおいて、関連データセットのデータのみを用いて、少量のデータから成る2つのデータセットから、精度よく密度比推定を行うモデルを学習する。すなわち、密度比推定装置は、少量のデータからなる2つのデータセットを、密度比推定を行うモデルに入力した場合に、それ以外のデータでも精度よく密度比推定をできるように、明示的に学習を行う。

[0014] また、密度比推定装置は、テスト（推定）フェーズにおいて、学習されたモデルに目標データセットを入力することにより、目標データセットの（相対）密度比推定を行う。

[0015] なお、関連データセットとは、例えば、同一の被写体についての色味の異

なる画像等、目標データセットと構成する特徴量（名）が同一であって、条件が異なり各特徴量の値の分布が異なるデータセットを意味する。

[0016] [密度比推定装置の構成]

次に、図2は、密度比推定装置の概略構成を例示する模式図である。本実施形態に係る密度比推定装置1は、ワークステーションやパソコン等の汎用コンピュータで実現され、後述する密度比推定処理を実行する。

[0017] 本実施形態の密度比推定装置1は、図2に示すように、学習処理を行う学習部10と、推定処理を行う推定部20とを有する。学習部10は、複数の関連データセットを用いて、密度比推定モデル14aを学習する。推定部20は、学習部10により学習された密度比推定モデル14aを用いて、少量のデータからなる2つの目標データセットから、適切な密度比推定を行う。推定部20は、学習部10と同一のハードウェアに実装されてもよいし、異なるハードウェアに実装されてもよい。

[0018] [学習部]

学習部10は、学習データ入力部11、特徴抽出部12、密度比推定モデル学習部13、および格納部14を有する。

[0019] 学習データ入力部11は、キーボードやマウス等の入力デバイスを用いて実現され、操作者による入力操作に対応して、制御部に対して各種指示情報を入力する。本実施形態において、学習データ入力部11は、取得部として機能して、密度比推定の処理対象の目標データと特徴量の構成が同一であって特徴量の値が異なる複数の関連データからなる関連データセットを取得する。

[0020] なお、関連データセットは、NIC (Network Interface Card) 等で実現される図示しない通信制御部を介して、外部のサーバ装置等から学習部10に入力されてもよい。

[0021] 制御部は、処理プログラムを実行するCPU (Central Processing Unit) 等を用いて実現され、特徴抽出部12、密度比推定モデル学習部13として機能する。

[0022] 特徴抽出部 12 は、後述する密度比推定モデル学習部 13 における処理の準備として、取得された関連データセットの各サンプルを特徴ベクトルに変換する。ここで、特徴ベクトルとは、必要なデータの特徴を n 次元の数ベクトルで表記したものである。特徴抽出部 12 は、機械学習で一般的に用いられている手法を利用して、特徴ベクトルへの変換を行う。例えば、特徴抽出部 12 は、データがテキストである場合には、形態素解析による手法、 n -gram による手法、区切り文字による手法等を適用可能である。

[0023] 密度比推定モデル学習部 13 は、学習部として機能する。つまり、密度比推定モデル学習部 13 は、取得された 2 セットの関連データセットのそれぞれから選択した関連データを疑似学習データまたは疑似テストデータとして、2 つの疑似学習データを密度比推定モデル 14 a に入力した場合に、2 セットの関連データセット間の密度比を正確に推定するように、該疑似学習データと疑似テストデータとを用いて、密度比推定モデル 14 a を学習する。

[0024] 具体的には、密度比推定モデル学習部 13 は、特徴抽出部 12 が特徴を抽出した後のデータを用いて、2 つのデータセットの部分集合である少量のデータから、当該データセットのペアに適切な密度比推定を行う密度比推定モデル 14 a を学習する。つまり、密度比推定モデル学習部 13 は、ランダムに選択された 2 つのデータセットのそれぞれから、少量の疑似学習データと、疑似テストデータとを選択する。そして、密度比推定モデル学習部 13 は、疑似学習データで密度比推定を行った際に、疑似テストデータでも正確に密度比推定を行うように学習を行う。

[0025] 密度比推定モデル 14 a としては、目的関数が微分可能な形式で表される任意の密度比推定手法が適用される。

[0026] ここで、図 3 は、密度比推定モデル学習部の処理を説明するための図である。図 3 には、密度比推定モデル学習部 13 の処理の疑似コードが例示されている。

[0027] まず、 X_d を d 番目のデータセットとする。ここで、 $x_{d,n} \in X_d$ は d 番目のデータセットの n 番目のサンプルの M 次元特徴ベクトルを表す。また、すべ

てのデータセットで特徴ベクトルの次元Mは同一と仮定する。

[0028] 密度比推定装置 1 は、学習フェーズにおいて、次式 (2) に示す D 個の関連データセットが与えられた場合に、テスト (推定) フェーズに与えられた次式 (3) に示す 2 つの目標データセットの密度比を精度よく推定するように、密度比推定モデル 1 4 a の学習を行う。

[0029] [数2]

$$\text{関連データセット} \{X_d\}_{d=1}^D \quad \dots (2)$$

[数3]

$$2 \text{ つの目標データセット } S_{d'}, S_{d''} \ (d', d'' \notin \{1, \dots, D\}) \quad \dots (3)$$

[0030] 具体的には、密度比推定モデル 1 4 a は、少量のデータから成る 2 つのデータセット S_{nu} と S_{de} との密度比を推定する。

[0031] 本実施形態では、密度比推定モデル 1 4 a は、密度比を拡張した相対密度比の推定を行う。すなわち、密度比推定モデル学習部 1 3 は、2 つの関連データセット間の相対密度比を推定する密度比推定モデル 1 4 a を学習する。

[0032] 2 つのデータセットの確率密度関数 $p_{nu}(x)$ 、 $p_{de}(x)$ の相対密度比は次式 (4) で定義される。

[0033] [数4]

$$r_\alpha(x) = \frac{p_{nu}(x)}{\alpha p_{nu}(x) + (1-\alpha)p_{de}(x)} \ (0 \leq \alpha < 1) \quad \dots (4)$$

[0034] ここで、 α はハイパーパラメータであり、 $\alpha = 0$ のときに、相対密度比が密度比に一致する。すなわち、本発明は通常密度比推定に用いることも可能である。相対密度比は、次式 (5) に示すように、密度比とは異なり上に有界であるため、密度比推定モデル 1 4 a の推定が容易に可能となる。

[0035] [数5]

$$r_\alpha(x) \leq \frac{1}{\alpha} \quad \dots (5)$$

[0036] まず、2 つのデータセット S_{nu} 、 S_{de} を、次式 (6)、(7) に示す置換不変ニューラルネットワークを用いて、固定長ベクトル z_{nu} 、 z_{de} に変換

する。

[0037] [数6]

$$\mathbf{z}_{\text{nu}} := g \left(\frac{1}{|\mathcal{S}_{\text{nu}}|} \sum_{\mathbf{x} \in \mathcal{S}_{\text{nu}}} f(\mathbf{x}) \right) \in \mathbb{R}^K, \quad \dots (6)$$

[数7]

$$\mathbf{z}_{\text{de}} := g \left(\frac{1}{|\mathcal{S}_{\text{de}}|} \sum_{\mathbf{x} \in \mathcal{S}_{\text{de}}} f(\mathbf{x}) \right) \in \mathbb{R}^K, \quad \dots (7)$$

[0038] ここで、 f 、 g は任意のフィードフォワードニューラルネットワークである。式（6）、（7）は、途中平均をとる操作を挟むことにより、各データセット内のサンプルの順番が変わっても出力は変わらない。すなわち、この置換不変ニューラルネットワークは、データセットを入力として受け取ることができる。

[0039] ここで得られた2つのデータセットの相対密度比の推定量は、当該データセットを表すベクトル $\mathbf{z}_{\text{nu}}$ 、 $\mathbf{z}_{\text{de}}$ を用いて、次式（8）に示すように定義する。

[0040] [数8]

$$\hat{r}_{\alpha}(\mathbf{x}; \mathcal{S}) := \mathbf{w}^{\top} h([\mathbf{x}, \mathbf{z}_{\text{nu}}, \mathbf{z}_{\text{de}}]) \quad \dots (8)$$

ここで、 $\mathbf{w} \in \mathbb{R}_{>0}^T$

[0041] ここで、 h は非負のアウトプットを返す任意のフィードフォワードニューラルネットワークであり、 $\mathbf{w}$ は非負のパラメタベクトルである。上記式（8）において、 h と $\mathbf{w}$ とが非負であることから、相対密度比の非負性が保証される。この相対密度比の推定量は、データセットベクトル $\mathbf{z}_{\text{nu}}$ 、 $\mathbf{z}_{\text{de}}$ に依存するため、ニューラルネットワーク f 、 g 、 h を適切に学習することで、2つのデータセット $\mathcal{S}_{\text{nu}}$ 、 $\mathcal{S}_{\text{de}}$ に適した密度比推定モデル14aを得られるものと期待される。

[0042] 密度比推定モデル学習部13は、真の相対密度比と推定した相対密度比との2乗誤差を最小化するように、密度比推定モデル14aを学習する。まず、密度比推定モデル学習部13は、次式（9）に示す真の相対密度比と推定

した相対密度比との2乗誤差を最小化することで、パラメタベクトル w を推定する。

[0043] [数9]

$$\begin{aligned} J_{\alpha}(\mathbf{w}) &:= \frac{1}{2} \mathbb{E}_{q_{\alpha}(\mathbf{x})} \left[(r_{\alpha}(\mathbf{x}) - \hat{r}_{\alpha}(\mathbf{x}; \mathcal{S}))^2 \right] \\ &= \frac{\alpha}{2} \mathbb{E}_{p_{\text{nu}}(x)} [\hat{r}_{\alpha}(\mathbf{x}; \mathcal{S})^2] + \frac{1-\alpha}{2} \mathbb{E}_{p_{\text{de}}(x)} [\hat{r}_{\alpha}(\mathbf{x}; \mathcal{S})^2] \\ &\quad - \mathbb{E}_{p_{\text{nu}}(x)} [\hat{r}_{\alpha}(\mathbf{x}; \mathcal{S})] + \text{Const.}, \end{aligned} \quad \dots (9)$$

ここで、 $q_{\alpha}(x) = \alpha p_{\text{nu}}(x) + (1-\alpha)p_{\text{de}}(x)$
 $\mathbb{E}$ は期待値、
 Const. は $\hat{r}_{\alpha}(x)$ によらない定数項

[0044] この2乗誤差をデータセット S_{nu} 、 S_{de} で近似して、いったん w の非負制約を外すことで、次式(10)に示す目的関数が得られる。

[0045] [数10]

$$\tilde{\mathbf{w}} := \arg \min_{\mathbf{w} \in \mathbb{R}^T} \left[\frac{1}{2} \mathbf{w}^{\top} \mathbf{K} \mathbf{w} - \mathbf{k}^{\top} \mathbf{w} + \frac{\lambda}{2} \mathbf{w}^{\top} \mathbf{w} \right] \quad \dots (10)$$

[0046] ここで、上記式(10)の λ を含む項は、オーバーフィットを防ぐための正則化項である。また、 $\mathbf{K}$ は次式(11)に示すように定義され、 $\mathbf{k}$ は次式(12)に示すように定義される。

[0047] [数11]

$$\begin{aligned} \mathbf{K} &:= \frac{\alpha}{|S_{\text{nu}}|} \sum_{\mathbf{x} \in S_{\text{nu}}} h([\mathbf{x}, \mathbf{z}_{\text{nu}}, \mathbf{z}_{\text{de}}]) h([\mathbf{x}, \mathbf{z}_{\text{nu}}, \mathbf{z}_{\text{de}}])^{\top} \\ &\quad + \frac{1-\alpha}{|S_{\text{de}}|} \sum_{\mathbf{x} \in S_{\text{de}}} h([\mathbf{x}, \mathbf{z}_{\text{nu}}, \mathbf{z}_{\text{de}}]) h([\mathbf{x}, \mathbf{z}_{\text{nu}}, \mathbf{z}_{\text{de}}])^{\top}, \end{aligned} \quad \dots (11)$$

[数12]

$$\mathbf{k} := \frac{1}{|S_{\text{nu}}|} \sum_{\mathbf{x} \in S_{\text{nu}}} h([\mathbf{x}, \mathbf{z}_{\text{nu}}, \mathbf{z}_{\text{de}}]), \quad \dots (12)$$

[0048] この目的関数の大域的最適解は、次式(13)に示す閉形式で得られる。

[0049] [数13]

$$\tilde{\mathbf{w}} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{k}, \quad \dots (13)$$

[0050] w の次元 T が大きい場合には、この計算は高速に実行可能である。な

お、上記式（１３）は、非負制約を満たさない可能性がある。そのため、相対密度比は、最終的には次式（１４）によって推定される。

[0051] [数14]

$$\hat{\mathbf{w}} = \max(0, \tilde{\mathbf{w}}) \quad \hat{r}_\alpha^*(\mathbf{x}; \mathcal{S}) := \hat{\mathbf{w}}^\top h([\mathbf{x}, \mathbf{z}_{\text{nu}}, \mathbf{z}_{\text{de}}]). \quad \dots(14)$$

[0052] 密度比推定モデル学習部１３は、次式（１５）に示す目的関数を用いて、ニューラルネットワーク f 、 g 、 h および正則化パラメタ λ を学習対象とする。

[0053] [数15]

$$\mathbb{E}_{d, d' \sim \{1, \dots, D\}} \left[\mathbb{E}_{(S_{\text{nu}}, S_{\text{de}}), (Q_{\text{nu}}, Q_{\text{de}}) \sim X_d \times X_{d'}} \left[\tilde{J}_\alpha(Q; \mathcal{S}) \right] \right], \quad \dots(15)$$

[0054] ここで、 S 、 Q は関連データセット $\{1, \dots, D\}$ からランダムに選んだ２つのデータセット X_{nu} 、 X_{de} から選択されるデータセットであり、 $S = (S_{\text{nu}}, S_{\text{de}})$ 、 $Q = (Q_{\text{nu}}, Q_{\text{de}})$ である。また、 S を使って推定した相対密度比と真の相対密度比との２乗誤差を、データ Q を使って評価した場合の誤差は、次式（１６）で表される。

[0055] [数16]

$$\begin{aligned} \tilde{J}_\alpha(Q; \mathcal{S}) = & \frac{\alpha}{2|Q_{\text{nu}}|} \sum_{\mathbf{x} \in Q_{\text{nu}}} \hat{r}_\alpha^*(\mathbf{x}; \mathcal{S})^2 \\ & + \frac{1-\alpha}{2|Q_{\text{de}}|} \sum_{\mathbf{x} \in Q_{\text{de}}} \hat{r}_\alpha^*(\mathbf{x}; \mathcal{S})^2 - \frac{1}{|Q_{\text{nu}}|} \sum_{\mathbf{x} \in Q_{\text{nu}}} \hat{r}_\alpha^*(\mathbf{x}; \mathcal{S}). \quad \dots(16) \end{aligned}$$

[0056] 上記式（１６）で表された量を最小化することで、少量のデータセット S から相対密度比を推定するように学習された密度比推定モデル１４aが、その他の大量のデータ Q においても正確に相対密度比を推定できるようになる。

[0057] このように、目標データセットと関連がある様々な関連データセットに対して学習することにより、関連データセットの学習により得た有用な知識を転用して、目標データセットからも正確に密度比推定を行えることが期待される。

[0058] なお、誤差関数は、上記の２乗誤差に限定されず、例えば、KLダイバー

ジェンス等でもよい。

[0059] 図2の説明に戻る。格納部14は、RAM(Random Access Memory)、フラッシュメモリ(Flash Memory)等の半導体メモリ素子、または、ハードディスク、光ディスク等の記憶装置によって実現される。本実施形態の格納部14には、学習された密度比推定モデル14aが格納される。

[0060] [推定部]

推定部20は、データ入力部21、特徴抽出部22、密度比推定部23、および結果出力部24を有する。

[0061] データ入力部21は、キーボードやマウス等の入力デバイスを用いて実現され、操作者による入力操作に対応して、制御部に対して各種指示情報を入力したり、密度比推定処理の対象の目標データセットを受け付けたりする。

[0062] なお、目標データセットは、NIC等で実現される図示しない通信制御部を介して、外部のサーバ装置等から推定部20に入力されてもよい。また、データ入力部21は、学習データ入力部11と同一のハードウェアでもよい。

[0063] 制御部は、処理プログラムを実行するCPU等を用いて実現され、特徴抽出部22と密度比推定部23とを有する。

[0064] 特徴抽出部22は、学習部10の特徴抽出部12と同様に、密度比推定部23における処理の準備として、取得された目標データセットの各サンプルを特徴ベクトルに変換する。

[0065] 密度比推定部23は、推定部として機能して、学習された密度比推定モデル14aを用いて、密度比推定の処理対象の2つの目標データセット間の密度比を推定する。

[0066] 結果出力部24は、液晶ディスプレイなどの表示装置、プリンター等の印刷装置、情報通信装置等によって実現され、密度比推定処理の結果を操作者に対して出力する。例えば、入力された目標データセットから推定された(相対)密度比を出力する。

[0067] [密度比推定処理]

次に、図4および図5を参照して、密度比推定装置1による密度比推定処理について説明する。密度比推定装置1の密度比推定処理は、学習部10による学習処理と、推定部20による推定処理とを含む。

[0068] [学習処理]

図4は、学習処理手順を例示するフローチャートである。図4のフローチャートは、例えば、ユーザによる学習処理の開始を指示する操作入力があったタイミングで開始される。

[0069] まず、学習データ入力部11が、密度比推定の処理対象の目標データと特徴量の構成が同一であって特徴量の値が異なる複数の関連データからなる関連データセットの入力を受け付ける（ステップS1）。次に、特徴抽出部12が、受け付けた関連データセットの各サンプルを特徴ベクトルに変換する（ステップS2）。

[0070] 次に、密度比推定モデル学習部13が、入力された2セットの関連データセットのそれぞれから選択した関連データを疑似学習データまたは疑似テストデータとして、2つの該疑似学習データを密度比推定モデル14aに入力した場合に、2セットの関連データセット間の密度比を正確に推定するように、該疑似学習データと疑似テストデータとを用いて、密度比推定モデル14aを学習する（ステップS3）。例えば、密度比推定モデル学習部13は、2つの関連データセット間の相対密度比を推定する密度比推定モデル14aを学習する。

[0071] また、密度比推定モデル学習部13は、学習した密度比推定モデル14aを格納部14に格納する。

[0072] [推定処理]

次に図5は、推定処理手順を例示するフローチャートである。図5のフローチャートは、例えば、ユーザによる推定処理の開始を指示する操作入力があったタイミングで開始される。

[0073] まず、データ入力部21が、処理対象の2つの少量の目標データセットを受け付け（ステップS11）、特徴抽出部22が、受け付けた目標データセ

ットの各サンプルを特徴ベクトルに変換する（ステップS 1 2）。

[0074] 次に、密度比推定部 2 3 が、学習された密度比推定モデル 1 4 a を用いて 2 つの目標データセット間の密度比を推定する（ステップS 1 3）。そして、結果出力部 2 4 が、密度比推定結果の出力すなわち推定された密度比の出力を行う（ステップS 1 4）。

[0075] 以上、説明したように、密度比推定装置 1 において、学習データ入力部 1 1 が、密度比推定の処理対象のデータと特徴量の構成が同一であって該特徴量の値が異なる複数の関連データからなる関連データセットを取得する。密度比推定モデル学習部 1 3 が、取得された 2 セットの関連データセットのそれぞれから選択した関連データを疑似学習データまたは疑似テストデータとして、2 つの該疑似学習データを密度比推定モデル 1 4 a に入力した場合に、2 セットの関連データセット間の密度比を正確に推定するように、該疑似学習データと疑似テストデータとを用いて、密度比推定モデル 1 4 a を学習する。

[0076] このように、密度比推定装置 1 は、関連データセットを用いて学習することにより、任意の目標データセットに対して高コストな計算を要する再学習を行わなくても、高精度に密度比を推定することが可能となる。つまり、密度比推定装置 1 は、関連データセットの有用な情報を活用して、少ない目標データセットの密度比推定を高精度に実現することが可能となる。したがって、データが少量しか得られない場合にも、低コストに精度よく密度比の推定が可能となる。そのため、少量のデータを対象とした転移学習、異常検知、2 標本検定、変化点検知、特徴選択、クラスタリング、GANs 等の性能向上が可能となる。

[0077] また、密度比推定モデル学習部 1 3 は、2 つの関連データセット間の相対密度比を推定する密度比推定モデル 1 4 a を学習する。このように、上に有界である相対密度比を適用することにより、密度比推定装置 1 は、計算コストを抑えて密度比推定を行うことが可能となる。

[0078] また、密度比推定モデル学習部 1 3 は、真の相対密度比と推定した相対密

度比との2乗誤差を最小化するように、密度比推定モデル14aを学習する。これにより、密度比推定装置1は、高精度に密度比を推定することが可能となる。

[0079] また、密度比推定部23が、学習された密度比推定モデル14aを用いて、密度比推定の処理対象の2つの目標データセット間の密度比を推定する。これにより、データが少量しか得られない場合にも、低コストに精度よく密度比推定が可能となる。

[0080] [プログラム]

上記実施形態に係る密度比推定装置1が実行する処理をコンピュータが実行可能な言語で記述したプログラムを作成することもできる。一実施形態として、密度比推定装置1は、パッケージソフトウェアやオンラインソフトウェアとして上記の密度比推定処理を実行する密度比推定プログラムを所望のコンピュータにインストールさせることによって実装できる。例えば、上記の密度比推定プログラムを情報処理装置に実行させることにより、情報処理装置を密度比推定装置1として機能させることができる。また、その他にも、情報処理装置にはスマートフォン、携帯電話機やPHS (Personal Handy phone System) 等の移動体通信端末、さらには、PDA (Personal Digital Assistant) 等のスレート端末等がその範疇に含まれる。また、密度比推定装置1の機能を、クラウドサーバに実装してもよい。

[0081] 図6は、密度比推定プログラムを実行するコンピュータの一例を示す図である。コンピュータ1000は、例えば、メモリ1010と、CPU1020と、ハードディスクドライブインタフェース1030と、ディスクドライブインタフェース1040と、シリアルポートインタフェース1050と、ビデオアダプタ1060と、ネットワークインタフェース1070とを有する。これらの各部は、バス1080によって接続される。

[0082] メモリ1010は、ROM (Read Only Memory) 1011およびRAM 1012を含む。ROM 1011は、例えば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライブイ

インタフェース１０３０は、ハードディスクドライブ１０３１に接続される。ディスクドライブインタフェース１０４０は、ディスクドライブ１０４１に接続される。ディスクドライブ１０４１には、例えば、磁気ディスクや光ディスク等の着脱可能な記憶媒体が挿入される。シリアルポートインタフェース１０５０には、例えば、マウス１０５１およびキーボード１０５２が接続される。ビデオアダプタ１０６０には、例えば、ディスプレイ１０６１が接続される。

[0083] ここで、ハードディスクドライブ１０３１は、例えば、OS１０９１、アプリケーションプログラム１０９２、プログラムモジュール１０９３およびプログラムデータ１０９４を記憶する。上記実施形態で説明した各情報は、例えばハードディスクドライブ１０３１やメモリ１０１０に記憶される。

[0084] また、密度比推定プログラムは、例えば、コンピュータ１０００によって実行される指令が記述されたプログラムモジュール１０９３として、ハードディスクドライブ１０３１に記憶される。具体的には、上記実施形態で説明した密度比推定装置１が実行する各処理が記述されたプログラムモジュール１０９３が、ハードディスクドライブ１０３１に記憶される。

[0085] また、密度比推定プログラムによる情報処理に用いられるデータは、プログラムデータ１０９４として、例えば、ハードディスクドライブ１０３１に記憶される。そして、CPU１０２０が、ハードディスクドライブ１０３１に記憶されたプログラムモジュール１０９３やプログラムデータ１０９４を必要に応じてRAM１０１２に読み出して、上述した各手順を実行する。

[0086] なお、密度比推定プログラムに係るプログラムモジュール１０９３やプログラムデータ１０９４は、ハードディスクドライブ１０３１に記憶される場合に限られず、例えば、着脱可能な記憶媒体に記憶されて、ディスクドライブ１０４１等を介してCPU１０２０によって読み出されてもよい。あるいは、密度比推定プログラムに係るプログラムモジュール１０９３やプログラムデータ１０９４は、LAN (Local Area Network) やWAN (Wide Area Network) 等のネットワークを介して接続された他のコンピュータに記憶

され、ネットワークインタフェース1070を介してCPU1020によって読み出されてもよい。

[0087] 以上、本発明者によってなされた発明を適用した実施形態について説明したが、本実施形態による本発明の開示の一部をなす記述および図面により本発明は限定されることはない。すなわち、本実施形態に基づいて当業者等によりなされる他の実施形態、実施例および運用技術等は全て本発明の範疇に含まれる。

符号の説明

- [0088] 1 密度比推定装置
- 10 学習部
- 11 学習データ入力部
- 12 特徴抽出部
- 13 密度比推定モデル学習部
- 14 格納部
- 14a 密度比推定モデル
- 20 推定部
- 21 データ入力部
- 22 特徴抽出部
- 23 密度比推定部
- 24 結果出力部

請求の範囲

- [請求項1] 密度比推定の処理対象のデータと特徴量の構成が同一であって該特徴量の値が異なる複数の関連データからなる関連データセットを取得する取得部と、
- 取得された2つの前記関連データセットのそれぞれから選択した関連データを疑似学習データまたは疑似テストデータとして、2つの該疑似学習データをモデルに入力した場合に、2つの前記関連データセット間の密度比を正確に推定するように、該疑似学習データと疑似テストデータとを用いて、前記モデルを学習する学習部と、
- を有することを特徴とする密度比推定装置。
- [請求項2] 前記学習部は、2つの前記関連データセット間の相対密度比を推定する前記モデルを学習することを特徴とする請求項1に記載の密度比推定装置。
- [請求項3] 前記学習部は、真の相対密度比と推定した相対密度比との2乗誤差を最小化するように、前記モデルを学習することを特徴とする請求項2に記載の密度比推定装置。
- [請求項4] 学習された前記モデルを用いて、密度比推定の処理対象の複数のデータからなるデータセット間の密度比を推定する推定部を、さらに有することを特徴とする請求項1に記載の密度比推定装置。
- [請求項5] 密度比推定装置が実行する密度比推定方法であって、
- 密度比推定の処理対象のデータと特徴量の構成が同一であって該特徴量の値が異なる複数の関連データからなる関連データセットを取得する取得工程と、
- 取得された2つの前記関連データセットのそれぞれから選択した2つの関連データを疑似学習データおよび疑似テストデータとして、2つの該疑似学習データをモデルに入力した場合に、2つの前記関連データセット間の密度比を正確に推定するように、該疑似学習データと疑似テストデータとを用いて、前記モデルを学習する学習工程と、

を含んだことを特徴とする密度比推定方法。

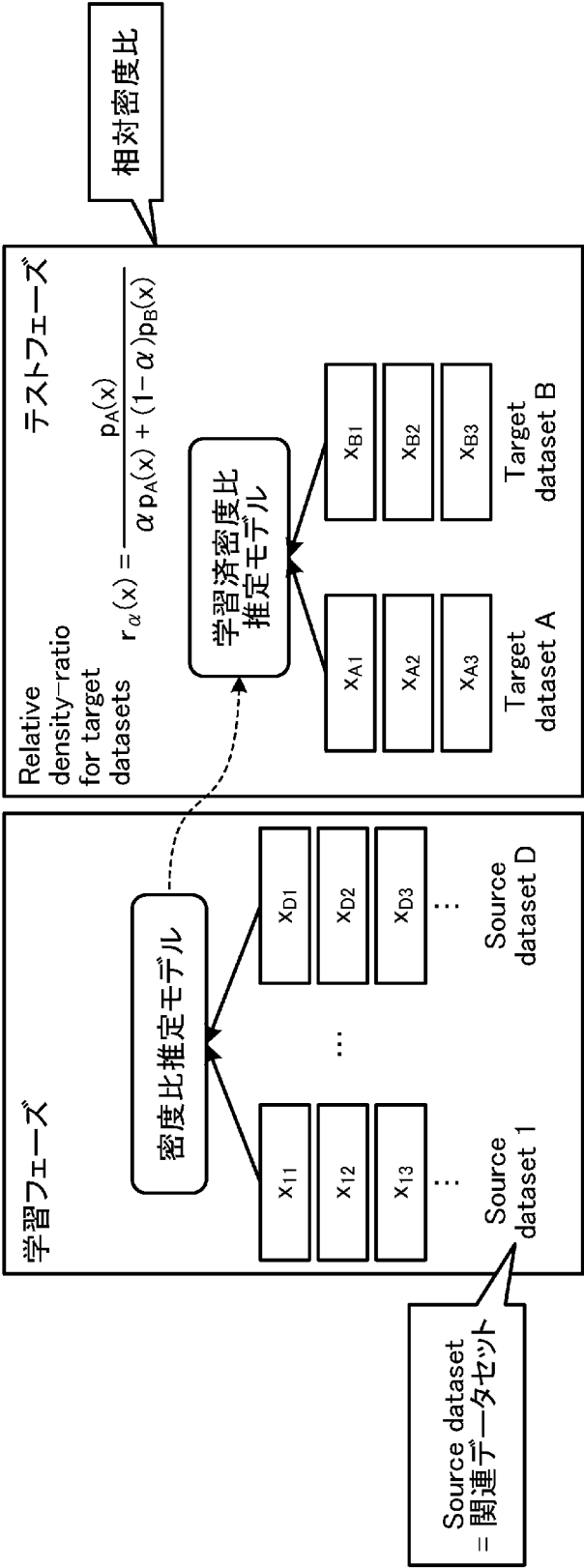
[請求項6]

コンピュータに、

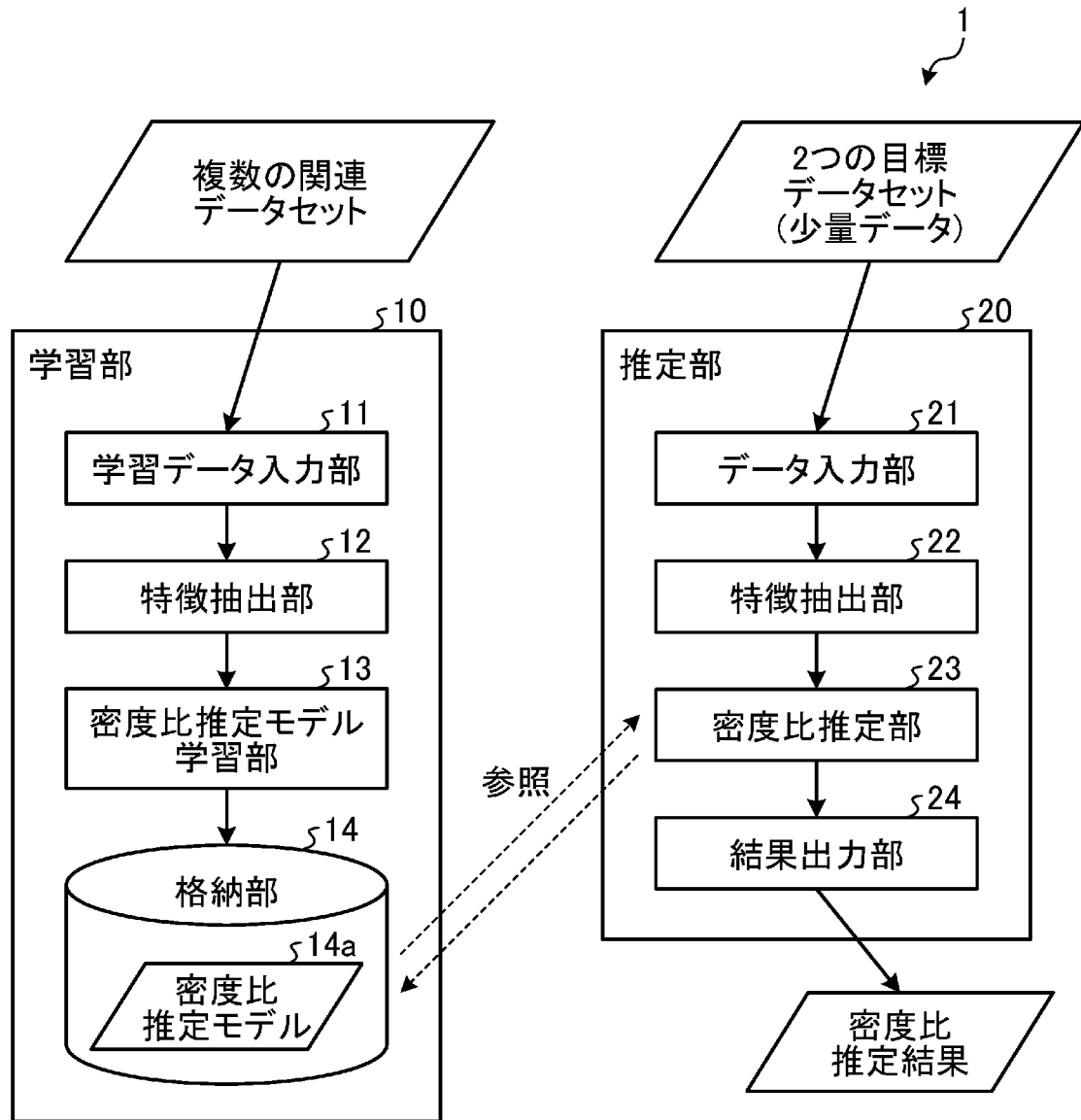
密度比推定の処理対象のデータと特徴量の構成が同一であって該特徴量の値が異なる複数の関連データからなる関連データセットを取得する取得ステップと、

取得された2つの前記関連データセットのそれぞれから選択した関連データを疑似学習データまたは疑似テストデータとして、2つの該疑似学習データをモデルに入力した場合に、2つの前記関連データセット間の密度比を正確に推定するように、該疑似学習データと疑似テストデータとを用いて、前記モデルを学習する学習ステップと、
を実行させることを特徴とする密度比推定プログラム。

[図1]



[図2]



[図3]

f,g,h, λ のパラメタ

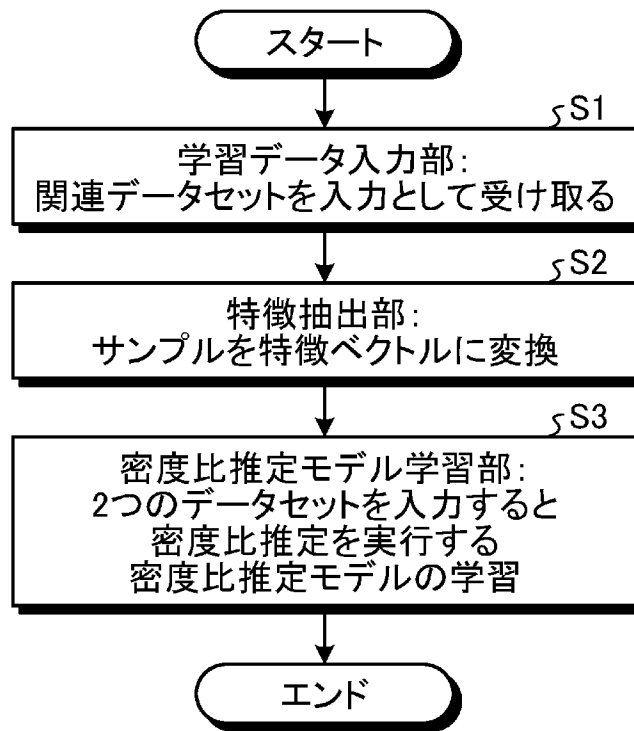
Algorithm 1 Training procedure of our model.

Require: Source datasets X , support sample size N_S , query sample size N_Q , relative parameter α

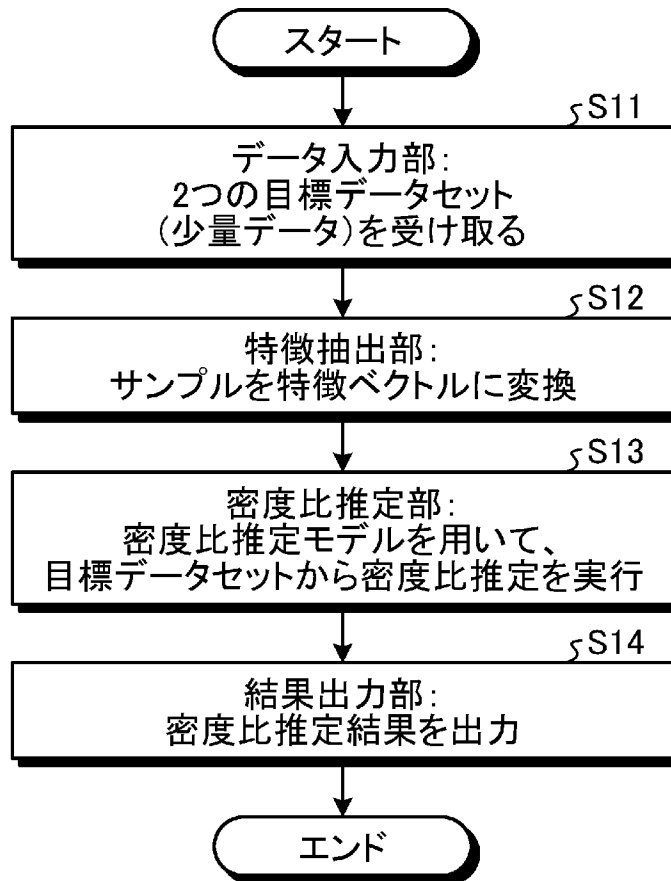
Ensure: Parameters of our model Θ

- 1: **repeat**
 - 2: Sample two datasets d and d' from $\{1, \dots, D\}$ with replacement
 - 3: Select support samples S_{nu} and S_{de} with size N_S from X_d and $X_{d'}$, respectively
 - 4: Select query samples Q_{nu} and Q_{de} with size N_Q from X_d and $X_{d'}$, respectively
 - 5: Calculate linear weights $\tilde{w}$ with the support samples by Eq.(13) to obtain the relative density-ratio Eq. (14)
 - 6: Calculate the loss $\tilde{J}_\alpha(Q; \mathcal{S})$ in Eq. (16) with the query samples
 - 7: Update parameters with the gradients of the loss $\tilde{J}_\alpha(Q; \mathcal{S})$
 - 8: **until** End condition is satisfied;
-

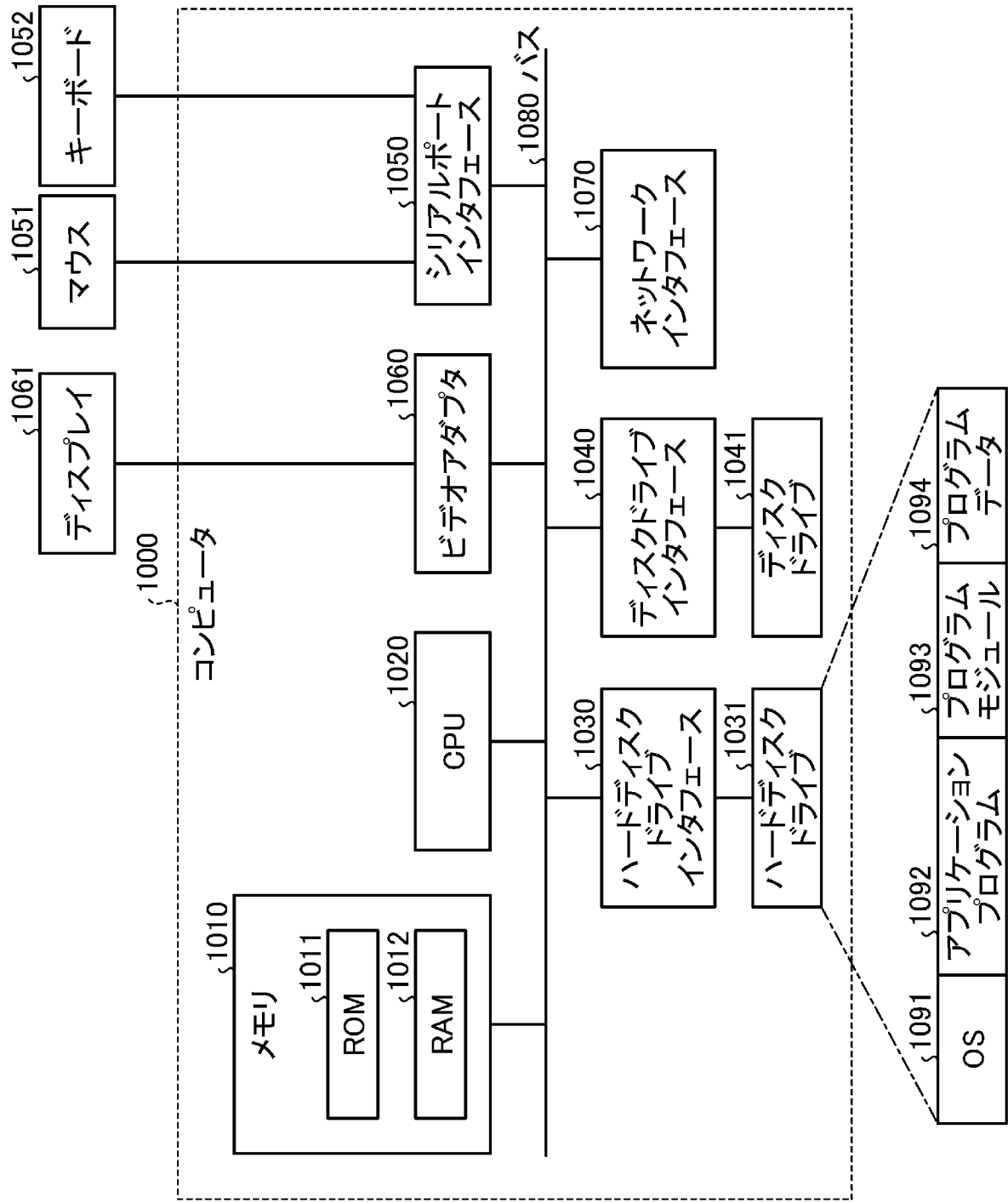
[図4]



[図5]



[図6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/015224

A. CLASSIFICATION OF SUBJECT MATTER**G06N 20/00**(2019.01)i

FI: G06N20/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N20/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996
 Published unexamined utility model applications of Japan 1971-2021
 Registered utility model specifications of Japan 1996-2021
 Published registered utility model applications of Japan 1994-2021

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2017-076289 A (CANON INC.) 20 April 2017 (2017-04-20) paragraphs [0024]-[0041], [0047]-[0053]	1-6
Y	JP 2013-160937 A (MITSUBISHI ELECTRIC CORP.) 19 August 2013 (2013-08-19) paragraphs [0019]-[0025], fig. 3	1-6
Y	YAMADA, Makoto et al. Relative Density-Ratio Estimation for Robust Distribution Comparison. Neural Computation. 01 March 2013, vol. 25, no. 5, pp. 1324-1370, Internet:<URL:https://direct.mit.edu/neco/article/25/5/1324/7871/Relative-Density-Ratio-Estimation-for-Robust>, [retrieved on 25 June 2021], <DOI:10.1162/NECO_a_00442> particularly, 2.1, 2.2	2-3

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

25 June 2021

Date of mailing of the international search report

06 July 2021

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)
 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915
 Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2021/015224

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2017-076289	A	20 April 2017	(Family: none)	
JP	2013-160937	A	19 August 2013	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 20/00(2019.01)i FI: G06N20/00														
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） G06N20/00 最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2021年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2021年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2021年</td> </tr> </table> 国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2021年	日本国実用新案登録公報	1996 - 2021年	日本国登録実用新案公報	1994 - 2021年				
日本国実用新案公報	1922 - 1996年													
日本国公開実用新案公報	1971 - 2021年													
日本国実用新案登録公報	1996 - 2021年													
日本国登録実用新案公報	1994 - 2021年													
C. 関連すると認められる文献														
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号												
Y	JP 2017-076289 A（キヤノン株式会社）20.04.2017（2017 - 04 - 20） [0024]-[0041], [0047]-[0053]	1-6												
Y	JP 2013-160937 A（三菱電機株式会社）19.08.2013（2013 - 08 - 19） [0019]-[0025], 図3	1-6												
Y	YAMADA, Makoto et al., Relative Density-Ratio Estimation for Robust Distribution Comparison, Neural Computation, 2013.03.01, Vol. 25, No. 5, pp.1324-1370, インターネット: <URL: https://direct.mit.edu/neco/article/25/5/1324/7871/Relative-Density-Ratio-Estimation-for-Robust>, [検索日 2021.06.25] <DOI: 10.1162/NECO_a_00442> 特に2.1-2.2	2-3												
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。														
<table border="0"> <tr> <td>* 引用文献のカテゴリー</td> <td>“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの</td> </tr> <tr> <td>“A” 特に関連のある文献ではなく、一般的技術水準を示すもの</td> <td>“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</td> </tr> <tr> <td>“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</td> <td>“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</td> </tr> <tr> <td>“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）</td> <td>“&” 同一パテントファミリー文献</td> </tr> <tr> <td>“O” 口頭による開示、使用、展示等に言及する文献</td> <td></td> </tr> <tr> <td>“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献</td> <td></td> </tr> </table>			* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの	“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの	“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの	“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献	“O” 口頭による開示、使用、展示等に言及する文献		“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	
* 引用文献のカテゴリー	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの													
“A” 特に関連のある文献ではなく、一般的技術水準を示すもの	“X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの													
“E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの	“Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの													
“L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）	“&” 同一パテントファミリー文献													
“O” 口頭による開示、使用、展示等に言及する文献														
“P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献														
国際調査を完了した日 25.06.2021	国際調査報告の発送日 06.07.2021													
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 金田 孝之 5B 3144 電話番号 03-3581-1101 内線 3545													

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2021/015224

引用文献			公表日	パテントファミリー文献	公表日
JP	2017-076289	A	20.04.2017	(ファミリーなし)	
JP	2013-160937	A	19.08.2013	(ファミリーなし)	