

(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4月28日(28.04.2022)



(10) 国際公開番号

WO 2022/085129 A1

(51) 国際特許分類:
G06N 20/00 (2019.01)

(21) 国際出願番号: PCT/JP2020/039602

(22) 国際出願日: 2020年10月21日(21.10.2020)

(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: 内田 美尋(UCHIDA, Mihiro); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 島村 潤

(SHIMAMURA, Jun); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 安藤 慎吾(ANDO, Shingo); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 梅田 崇之(UMEDA, Takayuki); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP).

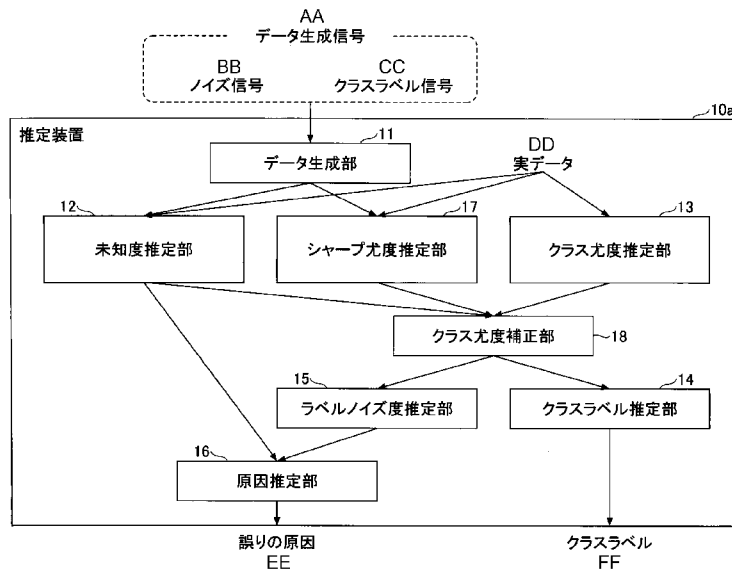
(74) 代理人: 伊東 忠重, 外(ITO, Tadashige et al.); 〒1000005 東京都千代田区丸の内二丁目1番1号 丸の内 M Y P L A Z A (明治安田生命ビル) 16階 Tokyo (JP).

(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,

(54) Title: LEARNING DEVICE, ESTIMATION DEVICE, LEARNING METHOD, ESTIMATION METHOD, AND PROGRAM

(54) 発明の名称: 学習装置、推定装置、学習方法、推定方法及びプログラム

[図5]



10a... ESTIMATION DEVICE
11... DATA GENERATING UNIT
12... UNKNOWN DEGREE ESTIMATING UNIT
13... CLASS LIKELIHOOD ESTIMATING UNIT
14... CLASS LABEL ESTIMATING UNIT
15... LABEL NOISE ESTIMATING UNIT
16... CAUSE ESTIMATING UNIT
17... SHARP LIKELIHOOD ESTIMATING UNIT
18... CLASS LIKELIHOOD CORRECTING UNIT
AA... DATA GENERATION SIGNAL
BB... NOISE SIGNAL
CC... CLASS LABEL SIGNAL
DD... ACTUAL DATA
EE... ERROR CAUSE
FF... CLASS LABEL

(57) Abstract: A learning device according to the present invention comprises: a data generating unit that learns to generate data based on a class label signal and a noise signal; an unknown degree estimating unit that uses a training set and the data generated by the data generating unit to learn to estimate the degree to which inputted data is unknown; a first class likelihood estimating unit that uses the training set to learn to estimate a first likelihood for each class label in the inputted data; a second class likelihood estimating unit that uses the training set and the data generated by the data generating unit to learn to estimate a second likelihood for each of the class labels in the inputted data; a class likelihood correcting unit that generates a third likelihood by correcting the first likelihood on the basis of the degree to which the inputted data is unknown and the second likelihood; and a class label estimating unit that, on the basis of the third likelihood, estimates

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, 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))

the class label of data related to the third likelihood. As a result, said learning device makes it possible for causes of errors attributable to a deep model to be automatically estimated.

(57) 要約: 学習装置は、クラスラベル信号及びノイズ信号に基づくデータの生成を学習するデータ生成部と、訓練セット及び前記データ生成部が生成するデータを用いて、入力データが未知である度合いの推定を学習する未知度推定部と、前記訓練セットを用いて、入力データについてクラスラベルごとの第1の尤度の推定を学習する第1のクラス尤度推定部と、前記訓練セット及び前記データ生成部が生成するデータを用いて、入力データについて前記クラスラベルごとの第2の尤度の推定を学習する第2のクラス尤度推定部と、前記未知である度合い及び前記第2の尤度に基づいて前記第1の尤度を補正することで第3の尤度を生成するクラス尤度補正部と、前記第3の尤度に基づいて、前記第3の尤度に係るデータのクラスラベルを推定するクラスラベル推定部と、を有することで、深層モデルによる誤りの原因を自動的に推定可能とする。

明 細 書

発明の名称：

学習装置、推定装置、学習方法、推定方法及びプログラム

技術分野

[0001] 本発明は、学習装置、推定装置、学習方法、推定方法及びプログラムに関する。

背景技術

[0002] 深層学習モデルは、高精度にタスクを実行できることで知られている。例えば、画像認識のタスクでは、人間を超える精度が達成されたことが報告されている。

[0003] 一方で、深層学習モデルは、未知のデータや誤ったラベル(ラベルノイズ)が付与されて学習されたデータについては意図しない挙動をすることが知られている。例えば、画像認識タスクを学習した画像認識モデルでは、未知の画像については正しいクラスラベルを推定できない可能性が有る。また、豚の画像に対して誤って「うさぎ」とラベル付けされて学習が行われた画像認識モデルは、豚の画像のクラスラベルを「うさぎ」と推定してしまう可能性が有る。実用上、このような挙動をする深層学習モデルは好ましくない。

先行技術文献

非特許文献

[0004] 非特許文献1：Odena, Augustus, Christopher Olah, and Jonathon Shlens. "Conditional image synthesis with auxiliary classifier gans." International conference on machine learning. 2017.

発明の概要

発明が解決しようとする課題

[0005] したがって、推定の誤りの原因に応じて対処が行われる必要が有る。例えば、未知データであることが原因であれば、訓練セットに対して未知データが追加される必要がある。また、ラベルノイズが原因であれば、ラベルの修

正が必要である。

[0006] しかし、人間が誤りの原因を正確に推定するのは困難である。

[0007] 本発明は、上記の点に鑑みてなされたものであって、深層モデルによる誤りの原因を自動的に推定可能とすることを目的とする。

課題を解決するための手段

[0008] そこで上記課題を解決するため、学習装置は、クラスラベル信号及びノイズ信号に基づくデータの生成を学習するデータ生成部と、訓練セット及び前記データ生成部が生成するデータを用いて、入力データが未知である度合いの推定を学習する未知度推定部と、前記訓練セットを用いて、入力データについてクラスラベルごとの第1の尤度の推定を学習する第1のクラス尤度推定部と、前記訓練セット及び前記データ生成部が生成するデータを用いて、入力データについて前記クラスラベルごとの第2の尤度の推定を学習する第2のクラス尤度推定部と、前記未知である度合い及び前記第2の尤度に基づいて前記第1の尤度を補正することで第3の尤度を生成するクラス尤度補正部と、前記第3の尤度に基づいて、前記第3の尤度に係るデータのクラスラベルを推定するクラスラベル推定部と、を有し、前記データ生成部は、前記未知である度合い、及び前記クラスラベル推定部によって推定されるクラスラベルに基づいて前記生成を学習する。

発明の効果

[0009] 深層モデルによる誤りの原因を自動的に推定可能とすることができる。

図面の簡単な説明

[0010] [図1] A C G A N を説明するための図である。

[図2] 本発明の実施の形態におけるクラスラベル推定装置10のハードウェア構成例を示す図である。

[図3] 第1の実施の形態におけるクラスラベル推定装置10の機能構成例を示す図である。

[図4] 第1の実施の形態におけるラベルノイズの検出性能を示す図である。

[図5] 第2の実施の形態におけるクラスラベル推定装置10aの機能構成例を

示す図である。

[図6]第2の実施の形態におけるクラスラベル推定装置10aの学習時の機能構成例を説明するための図である。

[図7]第2の実施の形態におけるクラスラベル推定装置10aの推論時の機能構成例を説明するための図である。

[図8]第2の実施の形態のラベルノイズの検出性能を説明するための第1の図である。

[図9]第2の実施の形態のラベルノイズの検出性能を説明するための第2の図である。

[図10]第2の実施の形態の未知データの検出性能を説明するための第1の図である。

[図11]第2の実施の形態の未知データの検出性能を説明するための第2の図である。

発明を実施するための形態

[0011] 本実施の形態ではACGAN (Auxiliary Classifier Generative Adversarial Network) をベースとするモデル (DNN (Deep Neural Network)) が開示される。そこで、まず、ACGANについて簡単に説明する。

[0012] 図1は、ACGANを説明するための図である。ACGANはccGAN (conditional GAN) の一種であり、GANにおける (Discriminator (識別器)) に補助のクラス分類器 (補助分類器) を付けることでクラスラベル (カテゴリラベル) を指定したデータ生成が可能になったGAN (Generative Adversarial Network) をいう。

[0013] すなわち、図1において、生成器は、ノイズ信号とクラスラベル信号からデータ (画像等) を生成する。ノイズ信号は、生成対象の画像の特徴を含むデータをいう。クラスラベル信号は生成対象の画像が示す対象物のクラスラベルを示すデータをいう。識別器は、生成器が生成したデータ (以下、「生成データ」という。) が訓練セットに含まれる実データか否か (すなわち、生成データであるか否か) を識別する。補助分類器は、識別器が識別したデ

ータのクラスラベル（以下、単に「ラベル」という。）を推定する。

[0014] 以下、図面に基づいて本発明の実施の形態を説明する。図2は、本発明の実施の形態におけるクラスラベル推定装置10のハードウェア構成例を示す図である。図2のクラスラベル推定装置10は、それぞれバスBで相互に接続されているドライブ装置100、補助記憶装置102、メモリ装置103、プロセッサ104、及びインタフェース装置105等を有する。

[0015] クラスラベル推定装置10での処理を実現するプログラムは、CD-ROM等の記録媒体101によって提供される。プログラムを記憶した記録媒体101がドライブ装置100にセットされると、プログラムが記録媒体101からドライブ装置100を介して補助記憶装置102にインストールされる。但し、プログラムのインストールは必ずしも記録媒体101より行う必要はなく、ネットワークを介して他のコンピュータよりダウンロードするようにしてもよい。補助記憶装置102は、インストールされたプログラムを格納すると共に、必要なファイルやデータ等を格納する。

[0016] メモリ装置103は、プログラムの起動指示があった場合に、補助記憶装置102からプログラムを読み出して格納する。プロセッサ104は、CPU若しくはGPU（Graphics Processing Unit）、又はCPU及びGPUであり、メモリ装置103に格納されたプログラムに従ってクラスラベル推定装置10に係る機能を実行する。インタフェース装置105は、ネットワークに接続するためのインタフェースとして用いられる。

[0017] 図3は、第1の実施の形態におけるクラスラベル推定装置10の機能構成例を示す図である。図3において、クラスラベル推定装置10は、データ生成部11、未知度推定部12、クラス尤度推定部13、クラスラベル推定部14、ラベルノイズ度推定部15及び原因推定部16等を有する。これら各部は、クラスラベル推定装置10にインストールされた1以上のプログラムが、プロセッサ104に実行させる処理により実現される。なお、図3に示される機能構成は、ACGANをベースとする。

[0018] データ生成部11は、ACGANにおける生成器である。すなわち、デー

タ生成部 11 は、ノイズ信号とクラスラベル信号とを入力とし、ノイズ信号及びクラスラベル信号を用いて実データ（実際に存在するデータ）に似たデータであって、クラスラベル信号が示すラベルに対応するデータ（例えば、画像データ等）を生成する。学習時において、データ生成部 11 は、未知度推定部 12 が生成データを実データであると推定するように学習する。推論時（運用時の実データのクラスラベルの推定時）には、データ生成部 11 は用いられない。

[0019] 未知度推定部 12 は、A C G A Nにおける識別器である。すなわち、未知度推定部 12 は、データ生成部 11 が生成した生成データ又は訓練セットに含まれる実データを入力とし、入力データに関する未知度（当該データが生成データである度合いを示す連続値）を出力する。未知度推定部 12 は、当該未知度について閾値処理を行う。データ生成部 11 が生成したデータを未知度推定部 12 の学習に用いることで、訓練セット外の未知データを明示的に未知であると識別可能なように未知度推定部 12 を学習することができる。

[0020] クラス尤度推定部 13 及びクラスラベル推定部 14 は、A C G A Nにおける補助分類器を構成する。

[0021] クラス尤度推定部 13 は、未知度推定部 12 に対する入力データと同じ入力データを入力とし、当該入力データについて、ラベルごとの尤度を推定（計算）する。尤度は深層学習モデルでの `softmax` レイヤで計算される。したがって、ラベルごとの尤度は、`softmax` ベクトルによって表現される。クラス尤度推定部 13 は、生成データ及び実データの双方を用いて学習される。

[0022] クラスラベル推定部 14 は、クラス尤度推定部 13 によって推定されたラベルごとの尤度に基づいて、入力データのラベルを推定する。

[0023] ラベルノイズ度推定部 15 及び原因推定部 16 は、A C G A Nによる推定の誤りの原因を推定するために、第 1 の実施の形態において、A C G A Nに追加される機構である。

- [0024] ラベルノイズ度推定部 15 は、クラス尤度推定部 13 によって推定されるラベルごとの尤度に基づいて、ラベルノイズ（訓練セット中のラベル誤り）の影響度合いであるラベルノイズ度を推定する。
- [0025] `softmax` ベクトルは、ラベルノイズの影響が無い場合に $[1.00, 0.00, 0.00]$ のように、いずれかの 1 つのクラスの尤度が圧倒的に 1 に近いシャープなベクトルになる。一方で、ラベルノイズの影響がある場合に $[0.33, 0.33, 0.33]$ といった、いずれのクラスの尤度も似た値になるフラットなベクトルになる。したがって、`softmax` ベクトルのフラットさはラベルノイズ度を表すといえる。そこで、ラベルノイズ度推定部 15 は、例えば、`softmax` ベクトルの最大値、上位 2 つの値の差、又はエントロピー等をラベルノイズ度として出力する。
- [0026] 原因推定部 16 は、未知度推定部 12 によって推定された未知度と、ラベルノイズ度推定部 15 によって推定されたラベルノイズ度とを用いて、ラベルの推定対象のデータが未知のため誤認識する可能性があるのか、ラベルノイズのため誤認識する可能性があるのか、問題が無い場合誤認識しないのか（すなわち、誤りの原因）を推定する。例えば、原因推定部 16 は、未知度、ラベルノイズ度それぞれに対して閾値処理を行うなどして出力を決定する。
- [0027] 当該閾値処理の具体例について説明する。未知度は未知データに対してのみ大きくなる指標となり、ラベルノイズ度はラベルノイズのデータのみで大きくなる指標となることを期待されていることを前提とし、未知度に対する閾値 α 及びラベルノイズ度に対する閾値 β がそれぞれ設定される。原因推定部 16 は、未知度が閾値 α より高い場合は未知データであることを原因として推定し、ラベルノイズ度が閾値 β より高い場合はラベルノイズを原因として推定する。また、未知度が閾値 α 以下であり、かつ、ラベルノイズ度が閾値 β 以下である場合は、（ラベルの推定について）問題が無いと推定する。
- [0028] 上記ように、図 3 の構成には、ACGAN による推定の誤りの原因を推定するための機構が含まれている。

[0029] しかし、上記の構成については、本願発明者により、ラベルノイズの検出性能が低く、未知のデータについてもラベルノイズとして判定されてしまうことが確認されている。

[0030] 図4は、第1の実施の形態におけるラベルノイズの検出性能を示す図である。図4において、縦軸はラベルノイズ検出性能の指標（A U R O C）である。A U R O Cは、1に近いほど性能が良いことを表す。また、チャンスレートで正解するような当て推量で判断する検出器だとA U R O Cは0.5になる。

[0031] また、横軸における「max__prob」、「diff__prob」、「entropy」は、順番に、softmaxベクトルの最大値がラベルノイズ度がとされる場合、上位2つの値の差がラベルノイズ度がとされる場合、エントロピーがラベルノイズ度がとされる場合に対応する。図4上の各プロットは、これら3つの場合におけるデータセットごとのラベルノイズの検出性能（A U R O C）を示す。

[0032] 図4によれば、「max__prob」、「diff__prob」及び「entropy」のいずれの場合についても、多くのデータセットについてのA U R O Cが0.5付近であり、必ずしも良い性能が得られているとはいえない。このレベルの性能では、誤りの原因の推定についても高い性能は期待できない。したがって、図4の深層モデルの運用保守を行う際に適切な改善ができず、コストがかかったり効率的に不具合の修正が行えなかったりする可能性がある。

[0033] この原因として、本願発明者は、ラベルノイズ度推定部15の入力として、未知のデータ（すなわち、データ生成部11が生成したデータ）に基づくフラットなsoftmaxベクトルが含まれているためであると考察した。すなわち、ラベルノイズは、本来、既知のデータに対して定義される概念であるにも関わらず、第1の実施の形態では、既知と未知のデータを総合した評価値が用いられている。具体的には、本来、ラベルごとの尤度として取得したいsoftmaxベクトルは、 $p(y|x, D = \{\text{訓練セット}\})$ であ

るが、実際に得られる softmax ベクトルは、 $p(y|x, D = \{\text{訓練セット, 生成データ}\})$ である。

[0034] そこで、次に、上記の考察に基づいて改良された第2の実施の形態について説明する。第2の実施の形態では第1の実施の形態と異なる点について説明する。第2の実施の形態において特に言及されない点については、第1の実施の形態と同様でもよい。

[0035] 図5は、第2の実施の形態におけるクラスラベル推定装置10aの機能構成例を示す図である。図5中、図3と同一又は対応する部分には同一符号を付し、その説明は適宜省略する。

[0036] 図5において、クラスラベル推定装置10aは、図3の構成に対し、シャープ尤度推定部17及びクラス尤度補正部18を更に有する。また、クラス尤度推定部13に変更が加えられる。

[0037] 具体的には、第2の実施の形態において、クラス尤度推定部13は、訓練セットに含まれる実データのみで学習される。

[0038] シャープ尤度推定部17は、入力データについてラベルごとの尤度を推定（計算）する。ラベルごとの尤度は深層学習モデルの softmax レイヤで計算される。クラス尤度推定部13は、生成データ及び実データの双方を用いて学習される。以上の点について、シャープ尤度推定部17は、第1の実施の形態におけるクラス尤度推定部13と同じである。但し、シャープ尤度推定部17は、シャープな softmax ベクトルを推定（出力）する。斯かる推定を行えるようにするため、シャープ尤度推定部17は、推定結果の softmax ベクトルがシャープになるように学習してもよい。このような学習の方法の一例として、 softmax ベクトルのエントロピーの項を損失関数の制約項とする方法がある。シャープなベクトルであることとエントロピーが小さいことは同義であるため、エントロピーが小さくなるように学習することでシャープなベクトルを推定することが期待される。

[0039] 又は、シャープ尤度推定部17は、第1の実施の形態におけるクラス尤度推定部13と同様の学習を行った上で、当該学習に基づく推定結果（以下、

「当初の推定結果」という。)であるsoftmaxベクトルのうち、フラットなsoftmaxベクトルに対してシャープになるような変換を行ってもよい。シャープになるような変換は、例えば、以下の(1)～(3)の手順で行われてもよい。

(1) 当初の推定結果のsoftmaxベクトルの最大値となる次元を特定する。

(2) 当初の推定結果のsoftmaxベクトルと同じサイズの $[0, \dots, 0]$ というベクトルを用意する。

(3) (2)で用意したベクトルのうち、(1)で特定した次元の値を1に変更する。

[0040] 他にも、推定結果のsoftmaxベクトルの最大値 $-\varepsilon$ (ε は 10^{-9} 等、微小な値)を閾値にして、当該softmaxベクトルの各次元を2値化する等、変換には様々な方法が考えられる。

[0041] クラス尤度補正部18は、クラス尤度推定部13によって推定された尤度を、未知度推定部12で推定された未知度とシャープ尤度推定部17によって推定された尤度とに基づいて補正する。補正方法として、例えば、以下の数1の(1)のように未知度で重みをつけて足し合わせる方法(すなわち、加重和を補正值とする方法)や、数1の(2)のように、クラス尤度推定部13による推定された尤度とシャープ尤度推定部17によって推定された尤度とを未知度に対する条件に応じて選択する方法が挙げられる。クラス尤度補正部18は、ラベルノイズ度推定部15への出力とクラスラベル推定部14への出力とで異なる方法(でアルゴリズム)を用いてクラス尤度推定部13によって推定された尤度を補正してもよい。

[0042]

[数1]

$$(1) (1 - rf) \times softmax + rf \times softmax_{sharp}$$

$$(2) \begin{cases} softmax_{sharp} & \text{if } rf > th \\ softmax & \text{otherwise} \end{cases} \quad \begin{matrix} (2-1) \\ (2-2) \end{matrix}$$

但し、 rf は、未知度である。 $softmax$ は、クラス尤度推定部13からの出力（ $softmax$ ベクトル）である。 $softmax_{sharp}$ は、シャープ尤度推定部17からの出力（ $softmax$ ベクトル）である。 th は、閾値である。

[0043] なお、数1において、(2-1)は、「実データでないと推定されたデータに対してシャープ尤度推定部17の出力を選択的に用いる（当該出力を補正後の尤度とする）」ことを示す。(2-2)は、「実データと推定されたに対しクラス尤度推定部13の出力を選択的に用いる（当該出力を補正後の尤度とする）」ことを示す。

[0044] シャープ尤度推定部17及びクラス尤度補正部18の追加により、原因推定部16による推定精度の向上も期待される。すなわち、未知度が閾値 α より高く、かつ、ラベルノイズ度が閾値 β より高いケースも論理的には考えられるが、シャープ尤度推定部17及びクラス尤度補正部18によりこのようなケースが無くなることが期待されるからである。

[0045] なお、第2の実施の形態において、クラスラベル推定部14及びラベルノイズ度推定部15は、クラス尤度推定部13からの出力ではなくクラス尤度補正部18からの出力を入力とする点において第1の実施の形態と異なる。

[0046] 図6は、第2の実施の形態におけるクラスラベル推定装置10aの学習時の機能構成例を説明するための図である。図6中、図5と同一部分には同一符号が付されている。図6に示される各部のうち、データ生成部11、未知度推定部12、シャープ尤度推定部17及びクラス尤度推定部13が学習の

対象となるニューラルネットワークである。一方、クラス尤度補正部 18 及びクラスラベル推定部 14 は、学習時においては、データ生成部 11 の学習に利用されるアルゴリズムである。

[0047] データ生成部 11 は、従来の ACGAN と同様に、未知度推定部 12 によって未知度が低く推定されるように、かつ、クラスラベル推定部 14 でクラスラベル信号と同じラベルが推定されるように学習する。

[0048] 未知度推定部 12 は、従来の ACGAN と同様に、入力データがデータ生成部 11 の出力なのか実データなのかを識別できるように学習する。

[0049] シャープ尤度推定部 17 は、生成データ及び訓練セット内の実データを入力とし、入力データのラベルの尤度が相対的に高くなるように学習する。例えば、シャープ尤度推定部 17 は、正解クラスの尤度 = 99% のように圧倒的に当該尤度が高くなるように学習する。なお、入力データのラベルとは、入力データが生成データの場合にはクラスラベル信号が示すラベルであり、入力データが訓練セット内の実データの場合には、訓練セットにおいて実データに付与されたラベルである。

[0050] クラス尤度推定部 13 は、入力データである実データに付与されたラベルの尤度が相対的に高くなるように学習する。なお、学習時において、クラス尤度推定部 13 には生成データは入力されない。

[0051] クラス尤度補正部 18 は、クラス尤度推定部 13 によって推定されたラベルごとの尤度を、未知度推定部 12 で推定された未知度とシャープ尤度推定部 17 によって推定されたラベルごとの尤度とに基づいて補正する。

[0052] クラスラベル推定部 14 は、クラス尤度補正部 18 によって補正されたラベルごとの尤度に基づいて、入力データのラベルを推定する。推定結果は、データ生成部 11 の学習に用いられる。

[0053] 図 7 は、第 2 の実施の形態におけるクラスラベル推定装置 10a の推論時の機能構成例を説明するための図である。図 7 中、図 5 と同一部分には同一符号が付されている。図 7 に示されるように、推論時において、データ生成部 11 は利用されない。また、推論時における実データは、ラベルが付与さ

れていない、ラベルの推定対象のデータ（例えば、実運用において利用されるデータ）である。

[0054] 推論時における各部の処理は、上記において説明した通りである。すなわち、未知度推定部 12 は、実データの未知度を推定する。シャープ尤度推定部 17 及びクラス尤度推定部 13 のそれぞれは、実データについてラベルごとの尤度を推定する。クラス尤度補正部 18 は、クラス尤度推定部 13 による推定結果である `softmax` ベクトルを、未知度推定部 12 によって推定された未知度とシャープ尤度推定部 17 による推定結果とに基づいて補正する。クラスラベル推定部 14 は、補正されたラベルごとの尤度に基づいて実データのラベルを推定する。ラベルノイズ度推定部 15 は、補正されたラベルごとの尤度に基づいてラベルノイズ度を推定する。原因推定部 16 は、未知度及びラベルノイズ度に対する閾値処理によって誤りの原因（未知、ラベルノイズ又は問題無し）を推定する。

[0055] 図 8 及び図 9 は、第 2 の実施の形態のラベルノイズの検出性能を説明するための図である。図 8 及び図 9 の見方は、図 4 と同様である。但し、図 8 及び図 9 の横軸において、「ベースモデル」は、第 1 の実施の形態の構成に対応する。「加重和」及び「選択」は、第 2 の実施の形態に対応する。「加重和」は、クラス尤度補正部 18 による補正が未知度による加重和によって行われるケースに対応する。「選択」は、クラス尤度補正部 18 による補正が未知度に基づくいずれか一方の尤度の選択によって行われるケースに対応する。

[0056] なお、図 8 と図 9 とは、ラベルノイズの種類が異なる。図 8 は、ラベルノイズが「`Symmetric noise`」である場合に対応し、図 9 は、ラベルノイズが「`Asymmetric noise`」である場合に対応する。「`Symmetric noise`」は、データに対して用意されたラベルのそれぞれについて等確率に誤るラベルノイズをいう。例えば、「犬、ねこ、うさぎ、サル」という 4 つのクラスがあった場合に、犬を犬以外の 3 クラスに等確率で誤り、ねこをねこ以外の 3 クラスに等確率に誤り、…とい

ったラベルノイズが「Symmetric noise」である。一方で、「Asymmetric noise」は、「Symmetric noise」とは異なり、誤る確率が等確率にならないラベルノイズをいう。例えば、「犬，ねこ，うさぎ，サル」という4つのクラスがあった場合に、犬をねことは誤るがうさぎやサルとは誤らないようなラベルノイズが「Asymmetric noise」である。

[0057] 図8及び図9のいずれにおいても、第2の実施の形態によれば、ラベルノイズの検出性能（AUROC）がチャンスレート（=0.5）以下であるデータセットが少なくなったことが分かる。したがって、第2の実施の形態によって、ラベルノイズの検出性能が向上したことが検証されたと考えられる。

[0058] また、図10及び図11は、第2の実施の形態の未知データの検出性能を説明するための図である。図10及び図11の縦軸は、未知データの検出性能（AUROC）である。また、横軸の「rf」はベースモデルによる未知度に基づく検出性能に対応する、「ex_rf」は、第2の実施の形態による未知度に基づく検出性能に対応する。更に、図10と図11との関係は、図8と図9との関係と同じである。それ以外の横軸は、ラベルノイズ度に基づく未知データの検出性能に対応する。

[0059] 第2の実施の形態では、未知度とラベルノイズ度がそれぞれ独立に評価されるため、未知データでラベルノイズ度が低くなる保証はないが、図10及び図11によれば、第2の実施の形態において、未知データについてラベルノイズ度による検出性能が低くなっていることが分かる。すなわち、ラベルノイズが未知データに反応しなくなったため、誤りの検出結果に未知データとラベルノイズが誤りの原因として同時に推定される可能性が低いことを期待することができる。換言すれば、ラベルノイズ度に基づいて検出される誤りがラベルノイズであること（未知データではないこと）が保証されることを期待することができる。

[0060] なお、「rf」の列と「ex_rf」の列とについては、未知データの検

出性能は類似している。このことは、ラベルごとの尤度の推定方法を変化させたことによる悪影響が未知度での未知データの検出に対してほとんど無いことを示す。

[0061] 上述したように、第2の実施の形態によれば、タスク（ラベルの推定）を実行しつつ、深層モデルによる誤りの原因を自動的に推定可能とすることができる。また、ラベルノイズの評価値としてモデルの妥当性を担保することができる。更に、ラベルノイズの評価値である `softmax` のフラットさが未知のデータに反応してしまうのを防ぎ（未知データに対して `softmax` ベクトルがフラットになることを避けて）、ラベルノイズによる誤りの推定の性能を高めることができる。

[0062] なお、第2の実施の形態において、クラスラベル推定装置10aは、学習装置及びクラスラベル推定装置10の一例である。クラス尤度推定部13は、第1のクラス尤度推定部の一例である。シャープ尤度推定部17は、第2のクラス尤度推定部の一例である。

[0063] 以上、本発明の実施の形態について詳述したが、本発明は斯かる特定の実施形態に限定されるものではなく、請求の範囲に記載された本発明の要旨の範囲内において、種々の変形・変更が可能である。

符号の説明

[0064] 10、10a クラスラベル推定装置
11 データ生成部
12 未知度推定部
13 クラス尤度推定部
14 クラスラベル推定部
15 ラベルノイズ度推定部
16 原因推定部
17 シャープ尤度推定部
18 クラス尤度補正部
100 ドライブ装置

1 0 1	記録媒体
1 0 2	補助記憶装置
1 0 3	メモリ装置
1 0 4	プロセッサ
1 0 5	インタフェース装置
B	バス

請求の範囲

- [請求項1] クラスラベル信号及びノイズ信号に基づくデータの生成を学習するデータ生成部と、
- 訓練セット及び前記データ生成部が生成するデータを用いて、入力データが未知である度合いの推定を学習する未知度推定部と、
- 前記訓練セットを用いて、入力データについてクラスラベルごとの第1の尤度の推定を学習する第1のクラス尤度推定部と、
- 前記訓練セット及び前記データ生成部が生成するデータを用いて、入力データについて前記クラスラベルごとの第2の尤度の推定を学習する第2のクラス尤度推定部と、
- 前記未知である度合い及び前記第2の尤度に基づいて前記第1の尤度を補正することで第3の尤度を生成するクラス尤度補正部と、
- 前記第3の尤度に基づいて、前記第3の尤度に係るデータのクラスラベルを推定するクラスラベル推定部と、
- を有し、
- 前記データ生成部は、前記未知である度合い、及び前記クラスラベル推定部によって推定されるクラスラベルに基づいて前記生成を学習する、
- ことを特徴とする学習装置。
- [請求項2] 前記第2のクラス尤度推定部は、前記クラスラベル信号が示すクラスラベル又は前記訓練セットに付与されたクラスラベルに対する前記第2の尤度が相対的に高くなるように前記クラスラベルごとの第2の尤度の推定を学習する、
- ことを特徴とする請求項1記載の学習装置。
- [請求項3] 前記クラス尤度補正部は、前記第1の尤度と前記第2の尤度との加重和、又は前記第1の尤度若しくは前記第2の尤度を前記第3の尤度として生成する、
- ことを特徴とする請求項1又は2記載の学習装置。

- [請求項4] 入力データが未知である度合いを推定する未知度推定部と、
 訓練セットを用いた学習に基づいて、前記入力データについてクラスラベルごとの第1の尤度を推定する第1のクラス尤度推定部と、
 クラスラベル信号及びノイズ信号に基づいて生成されたデータ及び前記訓練セットを用いた学習に基づいて、前記入力データについて前記クラスラベルごとの第2の尤度を推定する第2のクラス尤度推定部と、
 前記未知である度合い及び前記第2の尤度に基づいて前記第1の尤度を補正することで第3の尤度を生成するクラス尤度補正部と、
 前記第3の尤度に基づいて、前記訓練セットにおけるラベルノイズの度合いを推定するラベルノイズ度推定部と、
 前記未知の度合い及び前記ラベルノイズの度合いに基づいて、前記入力データに関する誤りの原因を推定する原因推定部と、
 を有することを特徴とする推定装置。
- [請求項5] クラスラベル信号及びノイズ信号に基づくデータの生成を学習するデータ生成手順と、
 前記データ生成手順が生成するデータ及び訓練セットを用いて、入力データが未知である度合いの推定を学習する未知度推定手順と、
 前記訓練セットを用いて、入力データについてクラスラベルごとの第1の尤度の推定を学習する第1のクラス尤度推定手順と、
 前記データ生成手順が生成するデータ及び前記訓練セットを用いて、入力データについて前記クラスラベルごとの第2の尤度の推定を学習する第2のクラス尤度推定手順と、
 前記未知である度合い及び前記第2の尤度に基づいて前記第1の尤度を補正することで第3の尤度を生成するクラス尤度補正手順と、
 前記第3の尤度に基づいて、前記第3の尤度に係るデータのクラスラベルを推定するクラスラベル推定手順と、
 をコンピュータが実行し、

前記データ生成手順は、前記未知である度合い、及び前記クラスラベル推定手順によって推定されるクラスラベルに基づいて前記生成を学習する、

ことを特徴とする学習方法。

[請求項6]

入力データが未知である度合いを推定する未知度推定手順と、

訓練セットを用いた学習に基づいて、前記入力データについてクラスラベルごとの第1の尤度を推定する第1のクラス尤度推定手順と、

クラスラベル信号及びノイズ信号に基づいて生成されたデータ及び前記訓練セットを用いた学習に基づいて、前記入力データについて前記クラスラベルごとの第2の尤度を推定する第2のクラス尤度推定手順と、

前記未知である度合い及び前記第2の尤度に基づいて前記第1の尤度を補正することで第3の尤度を生成するクラス尤度補正手順と、

前記第3の尤度に基づいて、前記訓練セットにおけるラベルノイズの度合いを推定するラベルノイズ度推定手順と、

前記未知の度合い及び前記ラベルノイズの度合いに基づいて、前記入力データに関する誤りの原因を推定する原因推定手順と、

をコンピュータが実行することを特徴とする推定方法。

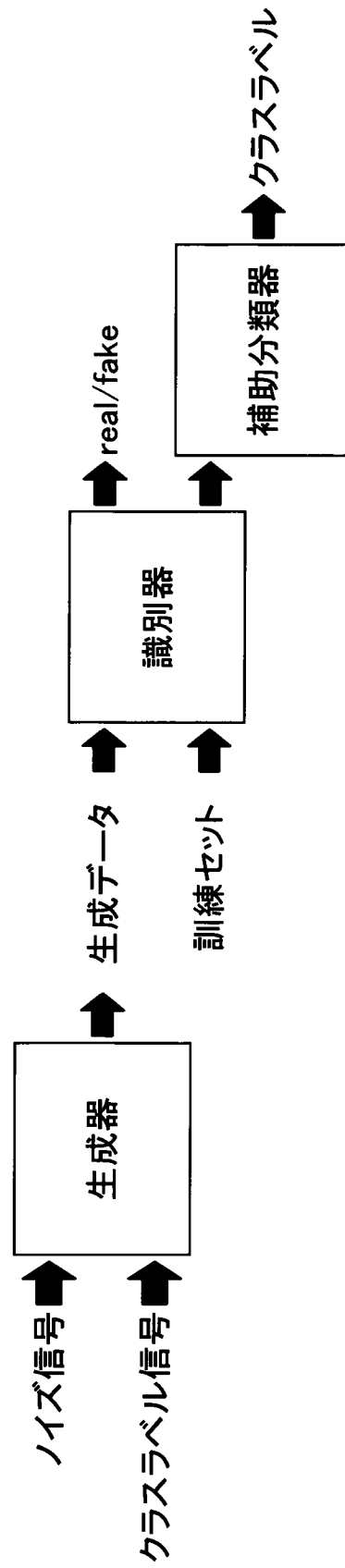
[請求項7]

請求項1乃至3いずれか一項記載の学習装置としてコンピュータを機能させることを特徴とするプログラム。

[請求項8]

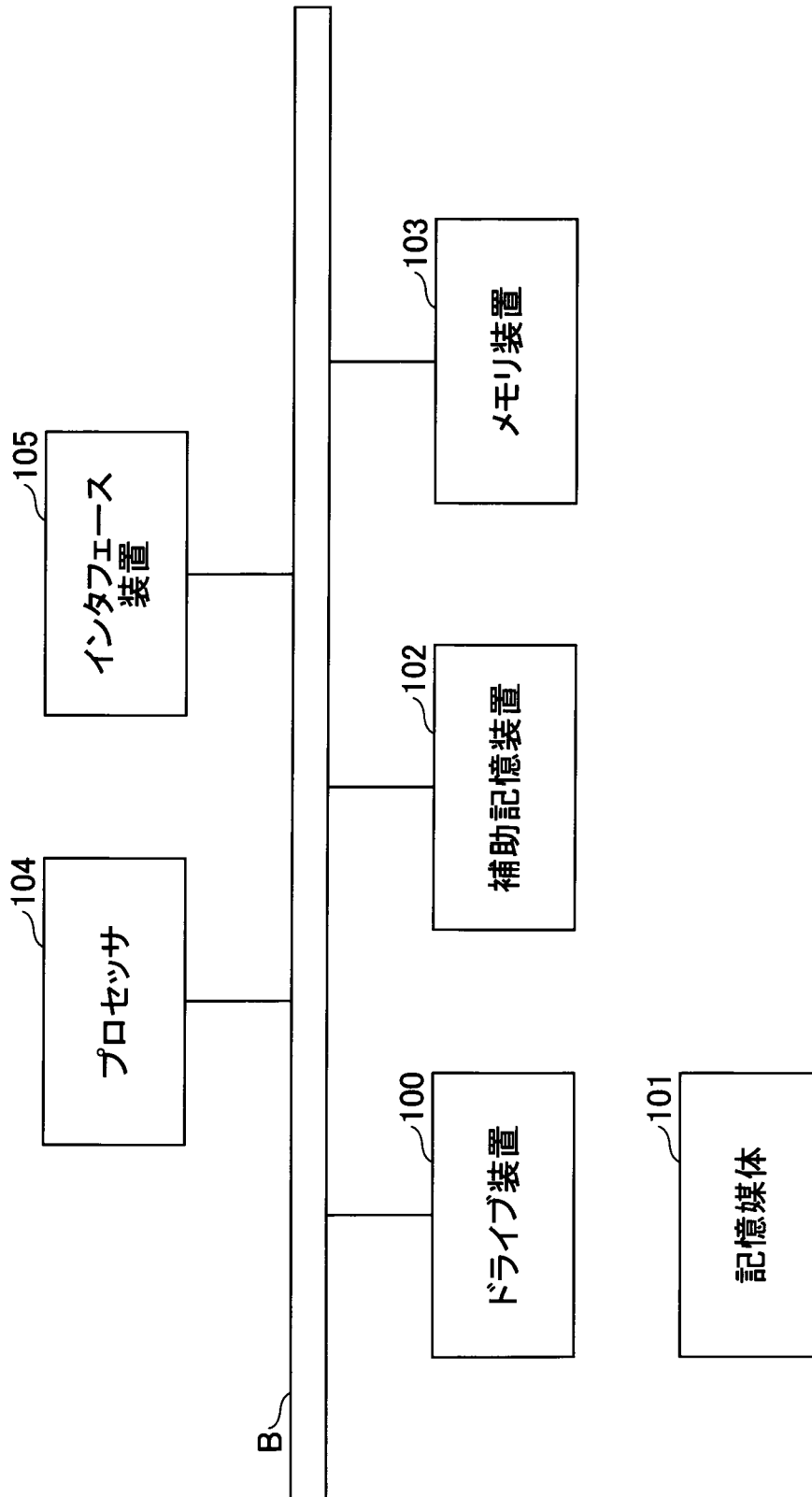
請求項4記載の推定装置としてコンピュータを機能させることを特徴とするプログラム。

[図1]

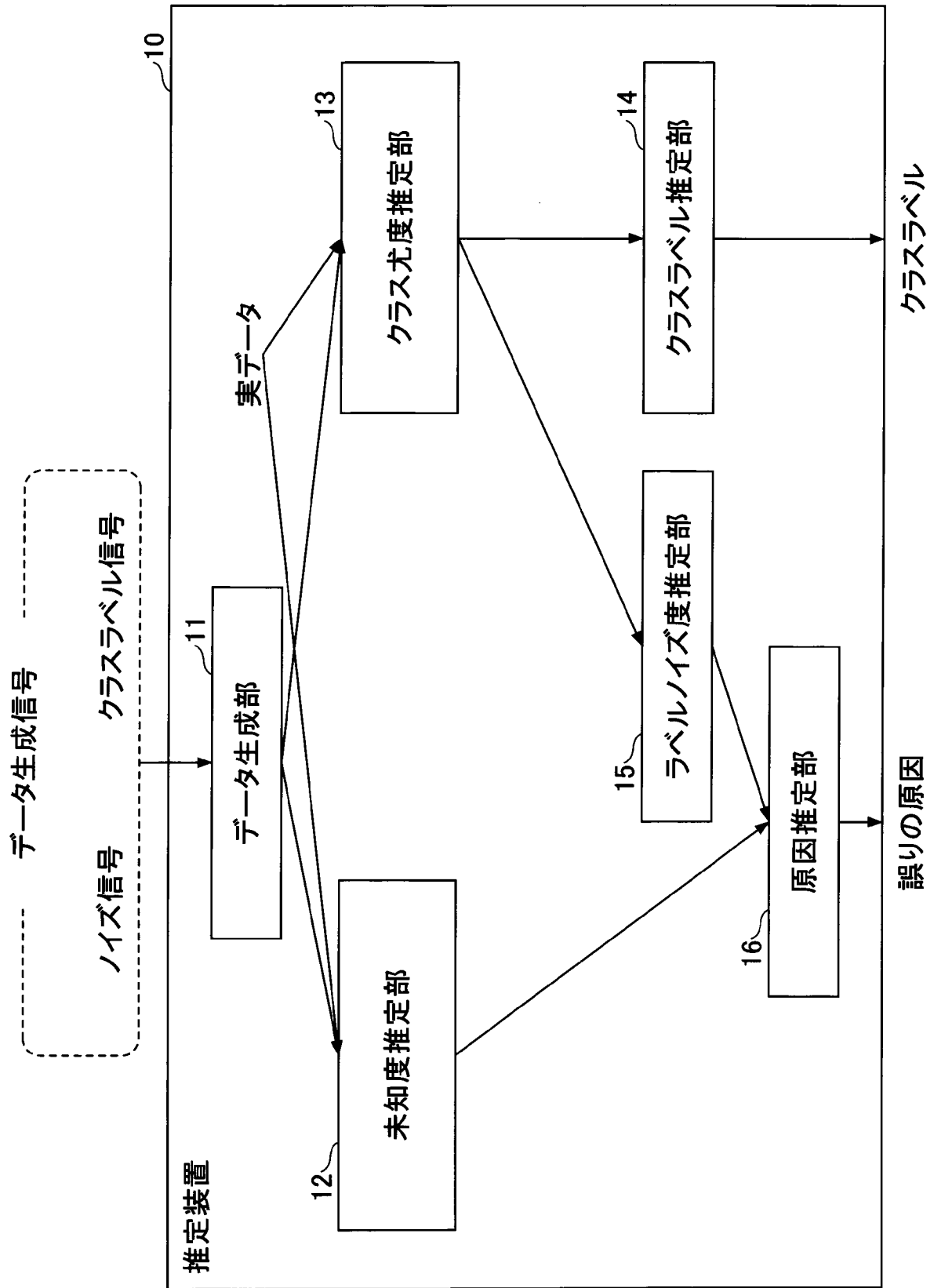


[図2]

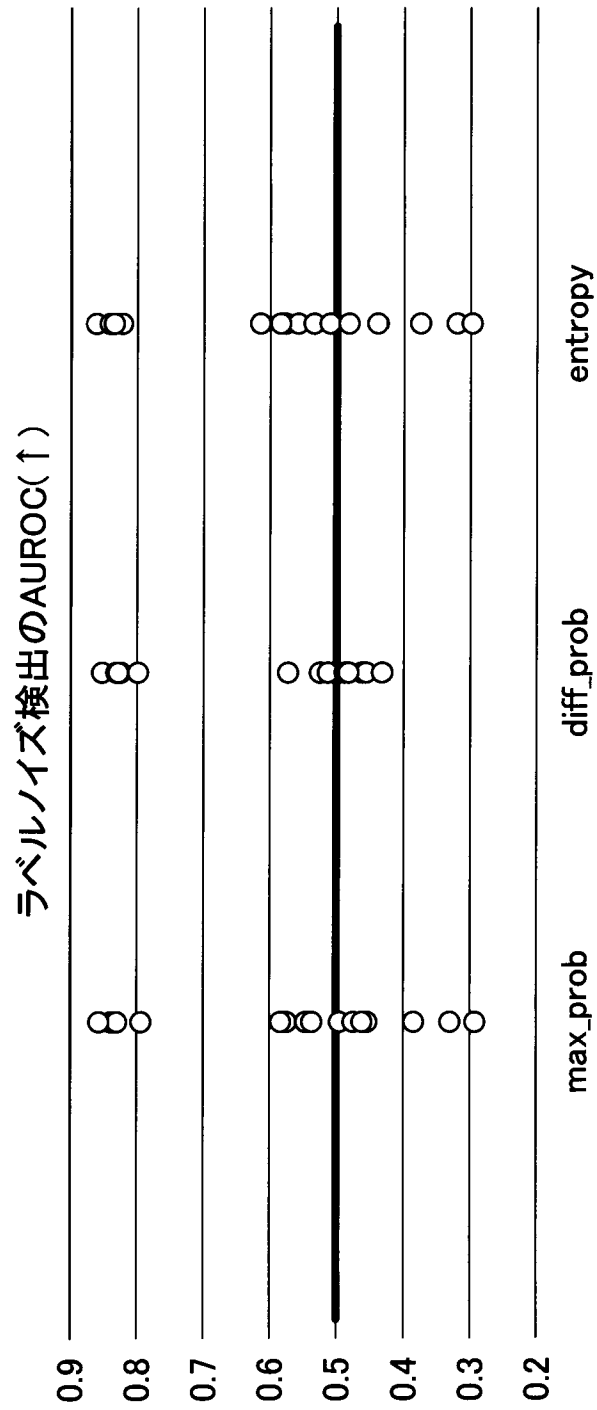
10



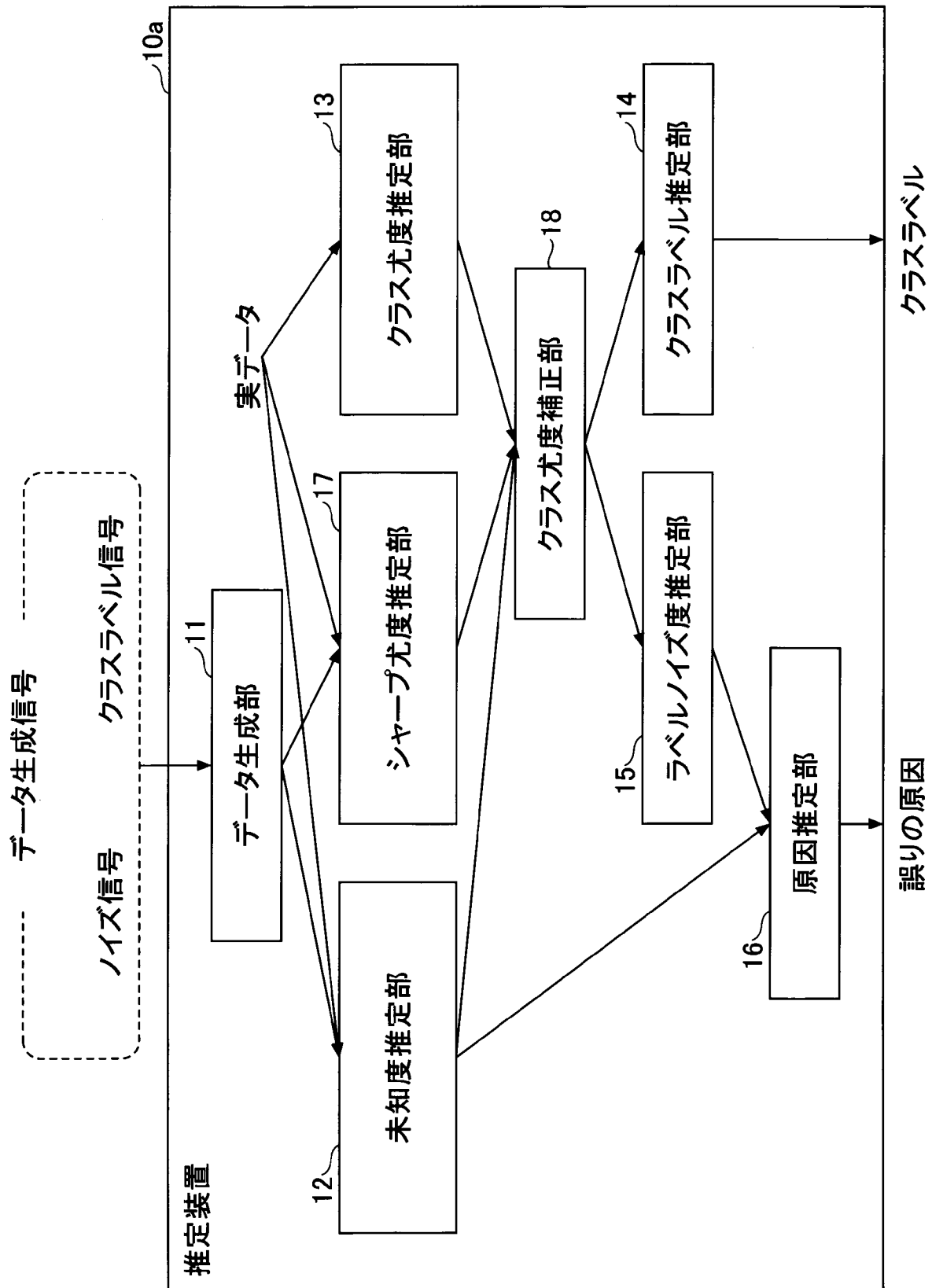
[図3]



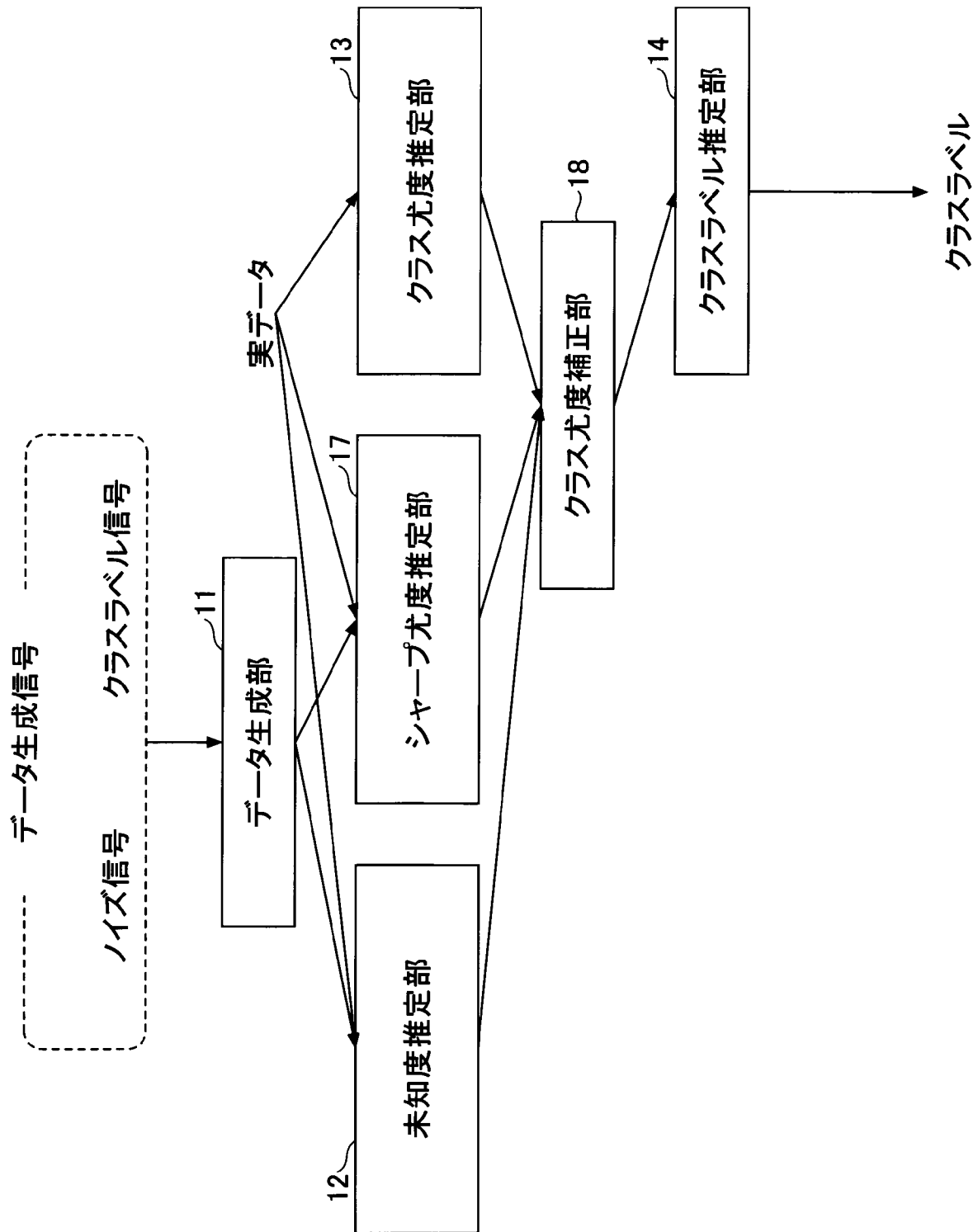
[図4]



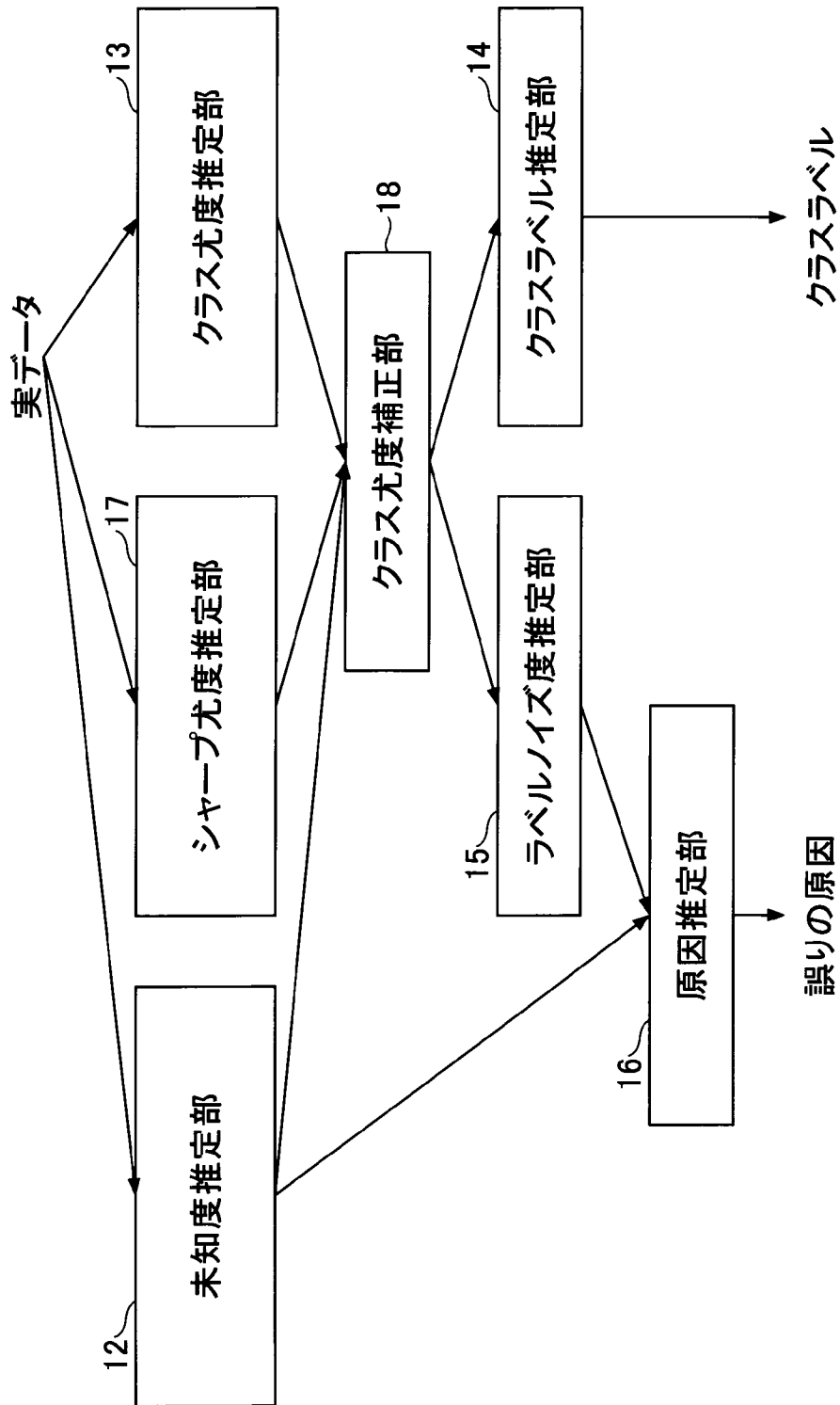
[図5]



[図6]



[図7]



[図8]

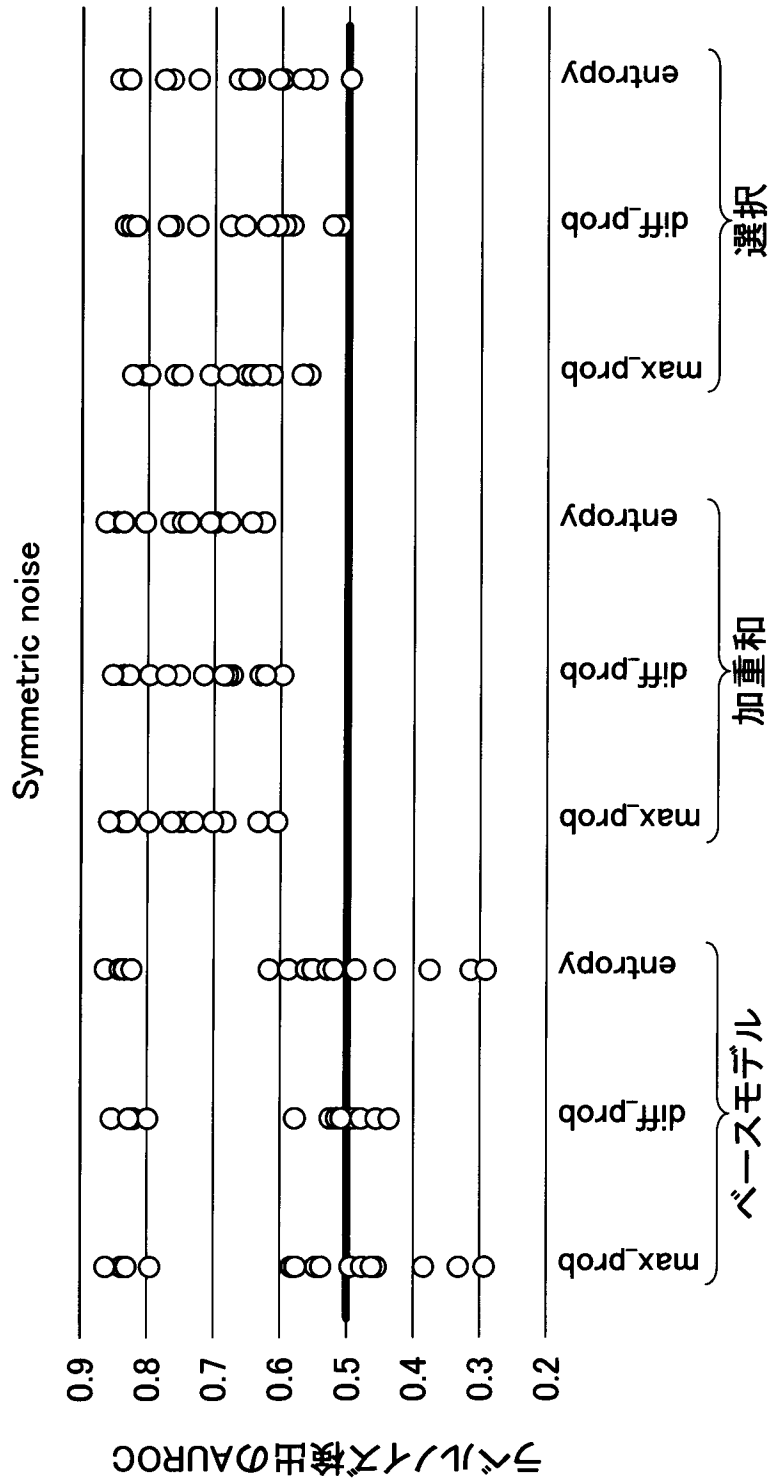
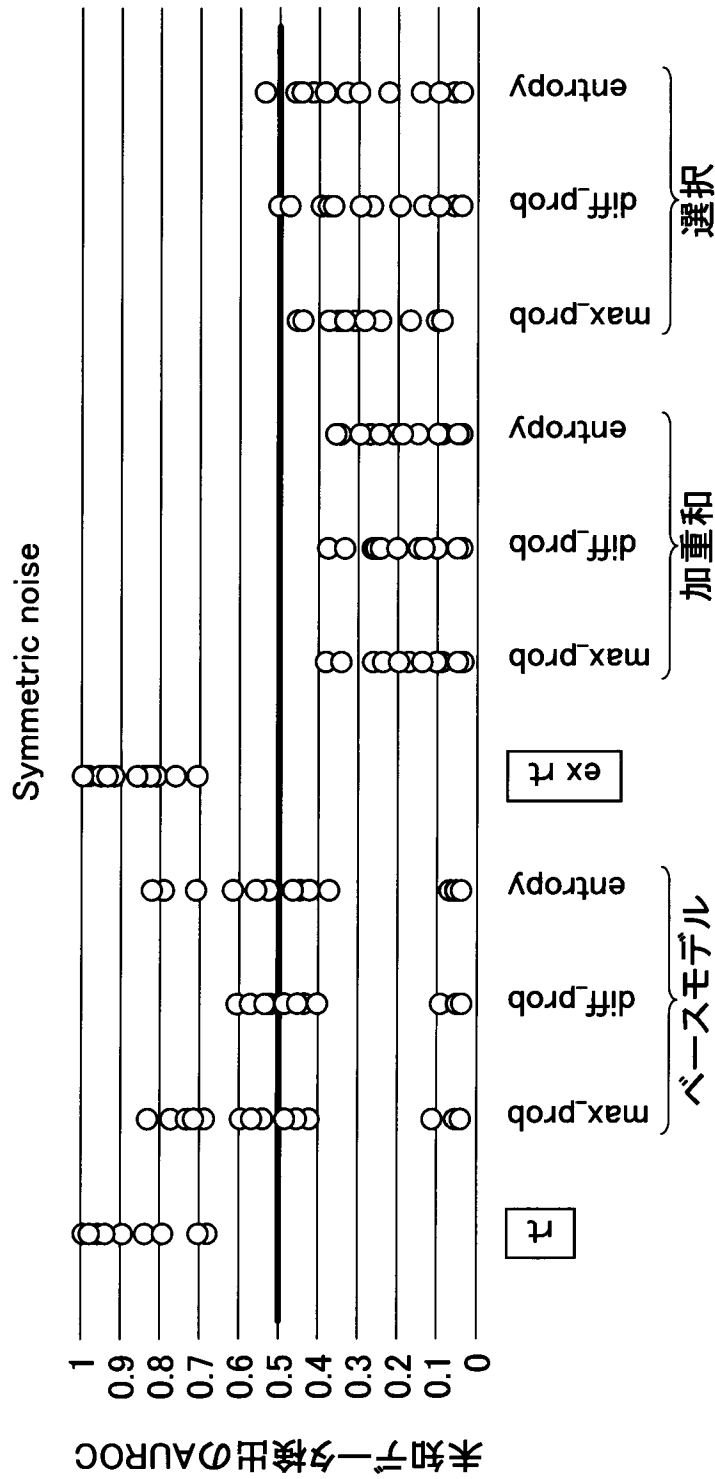
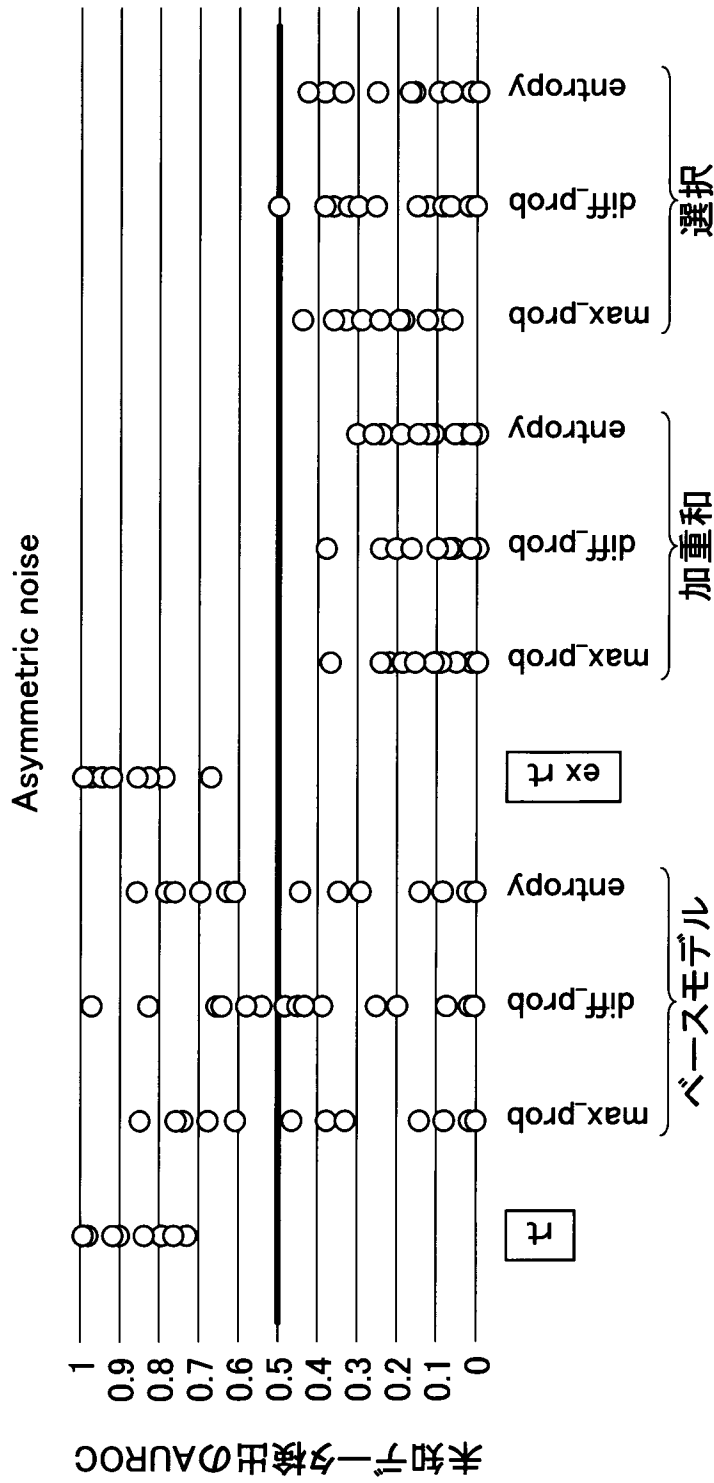


Figure 1 is a dot plot showing the AUROC values for three models (max_prob, diff_prob, entropy) across three noise conditions (Asymmetric noise, 加重和, ベースモデル). The y-axis represents the AUROC value from 0.2 to 0.9. The x-axis lists the models and noise conditions. A thick horizontal line is drawn at 0.5. In the 'Asymmetric noise' condition, all models have AUROC values around 0.5. In the '加重和' condition, all models have AUROC values around 0.6. In the 'ベースモデル' condition, all models have AUROC values around 0.3.

[図10]



[図11]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/039602

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06N20/00 (2019.01) i
FI: G06N20/00 130

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)

Int. Cl. 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 1971-2020

Registered utility model specifications of Japan 1996-2020

Published registered utility model applications of Japan 1994-2020

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.
A	JP 2014-85948 A (NIPPON TELEGRAPH AND TELEPHONE CORP.) 12 May 2014 (2014-05-12), entire text, all drawings	1-8
A	JP 2019-91440 A (PALO ALTO RESEARCH CENTER INC.) 13 June 2019 (2019-06-13), entire text, all drawings	1-8

☒ 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
17.12.2020

Date of mailing of the international search report
28.12.2020

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/039602

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ODENA, Augustus, OLAH, Christopher and SHEENS, Jonathon. Conditional Image Synthesis with Auxiliary Classifier GANs, Proceedings of the 34th International Conference on Machine Learning [Online], 06 August 2017 [retrieved on 17 December 2020], vol. 70, pp. 2642-2651, Internet: <URL: http://proceedings.mlr.press/v70/odena17a.html > entire text, all drawings	1-8
A	KANEKO, Takuhiro, USHIKU, Yoshitaka and HARADA, Tatsuya. Label-Noise Robust Generative Adversarial Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [Online], 15 June 2019 [retrieved on 17 December 2020], pp. 2462-2471, Internet: <URL: https://ieeexplore.ieee.org/abstract/document/8954304 >, <DOI: 10.1109/CVPR.2019.00257> entire text, all drawings	1-8
A	THEKUMPARAMPIL, Kiran K., KHETAN, Ashish, LIN, Zinan and OH, Sewoong. Robustness of conditional GANs to noisy labels. Advances in Neural Information Processing Systems, 02 December 2018 [retrieved on: 17 December 2020], vol. 31, pp. 10271-10282, Internet: <URL: https://proceedings.neurips.cc/paper/2018/file/565e8a413d0562degee4378402d2b481-Paper.pdf > entire text, all drawings	1-8

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/JP2020/039602

Patent Documents referred to in the Report	Publication Date	Patent Family	Publication Date
JP 2014-85948 A	12.05.2014	(Family: none)	
JP 2019-91440 A	13.06.2019	US 2019/0147333 A1 entire text, all drawings EP 3486838 A1	

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: G06N20/00 130														
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 - 2020年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2020年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2020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 - 2020年	日本国実用新案登録公報	1996 - 2020年	日本国登録実用新案公報	1994 - 2020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 - 2020年													
日本国実用新案登録公報	1996 - 2020年													
日本国登録実用新案公報	1994 - 2020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												
A	JP 2014-85948 A（日本電信電話株式会社）12.05.2014（2014 - 05 - 12） 全文，全図	1-8												
A	JP 2019-91440 A（パロ アルト リサーチ センター インコーポレイテッド） 13.06.2019（2019 - 06 - 13） 全文，全図	1-8												
A	ODENA, Augustus, OLAH, Christopher and SHLENS, Jonathon, Conditional Image Synthesis with Auxiliary Classifier GANs, Proceedings of the 34th International Conference on Machine Learning [Online], 06.08.2017 [検索日17.12.2020], Vol. 70, pp. 2642-2651, インターネット: <URL: http://proceedings.mlr.press/v70/odenal7a.html> 全文，全図	1-8												
A	KANEKO, Takuhiro, USHIKU, Yoshitaka and HARADA, Tatsuya, Label-Noise Robust Generative Adversarial Networks, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [Online], 06.15.2019 [検索日 17.12.2020], pp. 2462-2471, インターネット: <URL: https://ieeexplore.ieee.org/abstract/document/8954304>, <DOI: 10.1109/CVPR.2019.00257> 全文，全図	1-8												
☒ 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、一般的な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、一般的な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、一般的な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 17.12.2020	国際調査報告の発送日 28.12.2020													
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													

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
A	THEKUMPARAMPIL, Kiran K, KHETAN, Ashish, LIN, Zinan and OH, Sewoong, Robustness of conditional GANs to noisy labels, Advances in Neural Information Processing Systems, 02.12.2018 [検索日: 17.12.2020], Vol. 31, pp. 10271-10282, インターネット: <URL: https://proceedings.neurips.cc/ paper/2018/file/565e8a413d0562de9ee4378402d2b481-Paper.pdf > 全文, 全図	1-8

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/JP2020/039602

引用文献			公表日	パテントファミリー文献	公表日
JP	2014-85948	A	12.05.2014	(ファミリーなし)	
JP	2019-91440	A	13.06.2019	US 2019/0147333 A1	
				全文, 全図	
				EP 3486838 A1	