

(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1月20日(20.01.2022)



(10) 国際公開番号

WO 2022/014047 A1

(51) 国際特許分類:
G06N 20/00 (2019.01)

(21) 国際出願番号: PCT/JP2020/027875

(22) 国際出願日: 2020年7月17日(17.07.2020)

(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: 山田 真徳 (YAMADA, Masanori); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 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

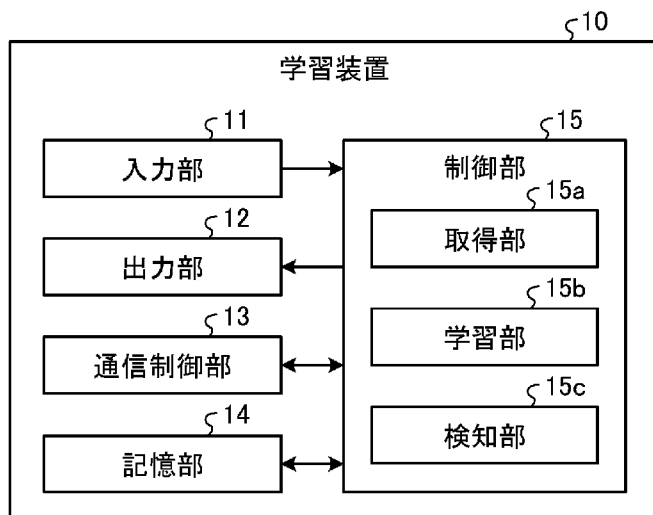
関3丁目8番1号 虎の門三井ビルディング Tokyo (JP).

(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, 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,

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

(54) 発明の名称: 学習装置、学習方法および学習プログラム



10 Learning device
11 Input unit
12 Output unit
13 Communication control unit
14 Storage unit
15 Control unit
15a Acquisition unit
15b Learning unit
15c Detection unit

(57) Abstract: In the present invention, an acquisition unit (15a) acquires data for which labels are to be predicted. With respect to a model indicating the probability distribution of labels for the acquired data, a learning unit (15b) learns the model using an eigenvector corresponding to the maximum eigenvalue of a Fisher information matrix for the data.

(57) 要約: 取得部(15a)が、ラベルを予測するデータを取得する。学習部(15b)が、取得されたデータのラベルの確率分布を表すモデルにおいて、データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、該モデルを学習する。

[続葉有]

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)）

明 細 書

発明の名称：学習装置、学習方法および学習プログラム

技術分野

[0001] 本発明は、学習装置、学習方法および学習プログラムに関する。

背景技術

[0002] 近年、機械学習は、大きな成功を収めている。特に深層学習の登場により、画像や自然言語の分野では、機械学習が主流の方法となっている。

[0003] 一方、深層学習は、悪意のあるノイズが乗せられた `Adversarial Example` による攻撃に対して脆弱であることが知られている。このような `Adversarial Example` に対する有力な対策として、代理損失を利用する TRADES (Tradeoff-inspired Adversarial Defense via Surrogate-loss minimization) と呼ばれる手法が提案されている（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：A.Madry et al., “Towards Deep Learning Models Resistant to Adversarial Attacks”, [online], arXiv:1706.06083v4 [stat.ML], 2019年9月, [2020年6月25日検索]、インターネット<URL: <https://arxiv.org/pdf/1706.06083.pdf>>

非特許文献2：H.Zhang et al., “Theoretically Principled Trade-off between Robustness and Accuracy”, [online], arXiv:1901.08573v3 [cs.LG], 2019年6月, [2020年6月25日検索]、インターネット<URL: <https://arxiv.org/pdf/1901.08573.pdf>>

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 TRADES では、`Adversarial Example` に対する汎化性能を向上させることが困難な場合がある。つまり

、代理損失で近似して最適なモデルを探索する際に、従来、微分を実行できない点を回避するように、初期値に乱数を用いているため、汎化性能を向上させることが困難な場合がある。

[0006] 本発明は、上記に鑑みてなされたものであって、Adversarial Exampleに対して頑健なモデルを学習することを目的とする。

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、本発明に係る学習装置は、ラベルを予測するデータを取得する取得部と、取得された前記データのラベルの確率分布を表すモデルにおいて、前記データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、該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、Adversarial Exampleに対して頑健なモデルを学習することが可能となる。

図面の簡単な説明

[0009] [図1]図1は、学習装置の概略構成を例示する模式図である。

[図2]図2は、学習処理手順を示すフローチャートである。

[図3]図3は、検知処理手順を示すフローチャートである。

[図4]図4は、実施例を説明するための図である。

[図5]図5は、実施例を説明するための図である。

[図6]図6は、学習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] [学習装置の構成]

図1は、学習装置の概略構成を例示する模式図である。図1に例示するよ

うに、学習装置 10 は、パソコン等の汎用コンピュータで実現され、入力部 11、出力部 12、通信制御部 13、記憶部 14、および制御部 15 を備える。

[0012] 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 15 に対して処理開始などの各種指示情報を入力する。出力部 12 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等によって実現される。

[0013] 通信制御部 13 は、NIC (Network Interface Card) 等で実現され、ネットワークを介したサーバ等の外部の装置と制御部 15 との通信を制御する。例えば、通信制御部 13 は、学習対象のデータを管理する管理装置等と制御部 15 との通信を制御する。

[0014] 記憶部 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、後述する学習処理により学習されたモデルのパラメータ等が記憶される。なお、記憶部 14 は、通信制御部 13 を介して制御部 15 と通信する構成でもよい。

[0015] 制御部 15 は、CPU (Central Processing Unit) 等を用いて実現され、メモリに記憶された処理プログラムを実行する。これにより、制御部 15 は、図 1 に例示するように、取得部 15a、学習部 15b および検知部 15c として機能する。なお、これらの機能部は、それぞれあるいは一部が異なるハードウェアに実装されてもよい。例えば、学習部 15b と検知部 15c とは、別々の装置として実装されてもよい。あるいは、取得部 15a は、学習部 15b および検知部 15c とは別の装置に実装されてもよい。また、制御部 15 は、その他の機能部を備えてもよい。

[0016] 取得部 15a は、ラベルを予測するデータを取得する。例えば、取得部 15a は、後述する学習処理および検知処理に用いるデータを、入力部 11 あるいは通信制御部 13 を介して取得する。また、取得部 15a は、取得したデータを記憶部 14 に記憶させてもよい。なお、取得部 15a は、これらの

情報を記憶部 14 に記憶させずに、学習部 15 b または検知部 15 c に転送してもよい。

[0017] 学習部 15 b は、取得されたデータのラベルの確率分布を表すモデルにおいて、データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、該モデルを学習する。具体的には、学習部 15 b は、損失関数においてデータに対して乗せるノイズの初期値として、データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、損失関数を最小化するモデルを探索することにより、モデルを学習する。

[0018] ここで、データ x のラベル y の確率分布を表すモデルは、パラメータ θ を用いて次式 (1) で表される。 f は、モデルが出力するラベルを表すベクトルである。

[0019] [数1]

$$p_{\theta}(y_k|x) = \frac{\exp f_k(x; \theta)}{\sum_i \exp f_i(x; \theta)} \quad \dots (1)$$

[0020] 学習部 15 b は、次式 (2) で表される損失関数が小さくなるように、モデルのパラメータ θ を決定することにより、モデルの学習を行う。ここで、 $p(y|x)$ は、真の確率を表す。

[0021] [数2]

$$l(x, y; \theta) = p(y|x) \log p_{\theta}(y|x) \quad \dots (2)$$

[0022] また、学習部 15 b は、データ x にノイズ η が乗せられた、次式 (3) に示す **Adversarial Example** に対して正しくラベルを予測できるように、モデルの学習を行う。

[0023] [数3]

$$\max_{\eta} E_{x, y \sim p(x, y)} [l(x + \eta, y; \theta)] \quad \dots (3)$$

[0024] 学習部 15 b は、次式 (4) に示す損失関数を最小化する θ を探索して決定することにより、**Adversarial Example** に頑強なモデ

ルを学習する。ここで、 β は定数である。

[0025] [数4]

$$\text{loss} = p(y|x) \log p_{\theta}(y|x) + \beta \max_{\eta} (D_{\text{KL}}(p_{\theta}(y|x) || p_{\theta}(y|x + \eta))) \quad \dots (4)$$

[0026] 上記式(4)の損失関数を最小化するためには、次式(5)に示すように、上記式(4)の第2項を微分して探索を行う。

[0027] [数5]

$$\frac{\partial}{\partial x'} D_{\text{KL}}(p_{\theta}(y|x) || p_{\theta}(y|x')) \quad \dots (5)$$

ここで

$$x' = x + \eta_0$$

[0028] ここで、式(4)の第2項においてノイズ η を変えながらその最大値を探索する際に、 η の初期値 η_0 を0とすると、 $x' = x$ となるため、式(4)の第2項の微分を実行することができない。

[0029] そこで、従来のTRADESでは、ノイズ η の初期値 η_0 を乱数 η_{rand} としている。そうすると、Adversarial Exampleに対する汎化性能を十分に向上させることが困難な場合がある。

[0030] ここで、上記式(4)の損失関数は、フィッシャー情報行列 G とその固有値 λ とを用いて、次式(6)のように変形することができる。

[0031] [数6]

$$\left. \begin{aligned} \text{loss} &= p(y|x) \log p_{\theta}(y|x) + \beta \eta G_{ij} \eta \\ &= p(y|x) \log p_{\theta}(y|x) + \beta \lambda \|\eta\|_2 \end{aligned} \right\} \dots (6)$$

ここでフィッシャー情報行列 G は

$$G_{ij} \equiv E_{p(y|x)} \left[\left(\frac{\partial}{\partial x'_i} \log p(y|x') \right) \left(\frac{\partial}{\partial x'_j} \log p(y|x') \right) \right]$$

[0032] そこで、本実施形態の学習部15bは、データ x に対するフィッシャー情報行列 G の最大固有値に対応する固有ベクトルを用いて、該モデルを学習する。具体的には、学習部15bは、上記式(5)において、次式(7)に示

すように、データ x に対して乗せるノイズ η の初期値 η_0 として、データ x に対するフィッシャー情報行列 G の最大固有値に対応する固有ベクトル η_{eig} を用いる。そして、上記式 (4) に示した損失関数を最小化する θ を探索することにより、モデルを学習する。

[0033] [数7]

$$x' = x + \eta_{\text{eig}} \quad \dots (7)$$

[0034] これにより、学習部 15 b は、損失関数を最小化するパラメータ θ の探索を精度高く行うことが可能となる。したがって、学習部 15 b は、Adversarial Example に対して頑健なモデルの学習が可能となる。

[0035] 検知部 15 c は、学習されたモデルを用いて、取得されたデータのラベルを予測する。この場合に、検知部 15 c は、学習されたパラメータ θ を上記式 (1) に適用することより、新たに取得されたデータの各ラベルの確率を算出し、最も確率が高いラベルを出力する。これにより、例えば、データが Adversarial Example であった場合にも、正しいラベルを出力することができる。このように、検知部 15 c は、blind spot attack に耐えて、Adversarial Example に正しいラベルを予測することが可能となる。

[0036] [学習処理]

次に、図 2 を参照して、本実施形態に係る学習装置 10 による学習処理について説明する。図 2 は、学習処理手順を示すフローチャートである。図 2 のフローチャートは、例えば、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。

[0037] まず、取得部 15 a が、ラベルを予測するデータを取得する（ステップ S 1）。

[0038] 次に、学習部 15 b が、取得されたデータのラベルの確率分布を表すモデルを学習する（ステップ S 1）。その際に、学習部 15 b は、モデルにおいて、データに対するフィッシャー情報行列の最大固有値に対応する固有ベク

トルを用いて、該モデルを学習する。具体的には、学習部 15 b は、損失関数においてデータに対して乗せるノイズの初期値として、データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、損失関数を最小化するモデルを探索することにより、モデルを学習する。これにより、一連の学習処理が終了する。

[0039] [検知処理]

次に、図 3 を参照して、本実施形態に係る学習装置 10 による検知処理について説明する。図 3 は、検知処理手順を示すフローチャートである。図 3 のフローチャートは、例えば、検知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。

[0040] まず、取得部 15 a が、上記した図 2 のステップ S 1 の処理と同様に、ラベルを予測する新たなデータを取得する（ステップ S 11）。

[0041] 次に、検知部 15 c が、学習されたモデルを用いて、取得されたデータのラベルを予測する（ステップ S 12）。この場合に、検知部 15 c は、学習されたパラメータ θ を上記式（1）に適用することにより、新たに取得されたデータ x' の $p(x')$ を算出し、最も確率が高いラベルを出力する。例えば、データ x' が *Adversarial Example* であった場合にも、正しいラベルを出力することができる。これにより、一連の検知処理が終了する。

[0042] 以上、説明したように、取得部 15 a が、ラベルを予測するデータを取得する。学習部 15 b が、取得されたデータのラベルの確率分布を表すモデルにおいて、データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、該モデルを学習する。具体的には、学習部 15 b は、損失関数においてデータに対して乗せるノイズの初期値として、データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、損失関数を最小化するモデルを探索する。

[0043] これにより、学習装置 10 は、*Adversarial Example* に対して頑健なモデルの学習が可能となる。

[0044] また、検知部15cが、学習されたモデルを用いて、取得されたデータのラベルを予測する。これにより、検知部15cは、`blind spot attack`に耐えて、`Adversarial Example`に対しても正しいラベルを予測することが可能となる。

[0045] [実施例]

図4および図5は、本発明の実施例について説明するための図である。本実施例では、画像のデータセット：`Cifar10`、深層学習モデル：`Resnet18`を用いて、上記実施形態のモデルの正確性の評価を行った。具体的には、テストデータと、PGDと呼ばれる手法でテストデータから生成した`Adversarial Example`とを用いて、上記式(4)に示した損失関数の β を変えて学習した、上記実施形態のモデルと従来手法のモデルとの評価を行った。

[0046] PGDのパラメータとして、`esp=8/255`、`train_iter=7`、`eval_iter=20`、`eps_iter=0.01`、`rand_init=True`、`clip_min=0.0`、`clip_max=1.0`を用いた。

[0047] そして、テストデータに対する`top1`の正解率（以下、`natural acc`と記す）と、テストデータから生成した`Adversarial Example`に対する`top1`の正解率（以下、`robust acc`と記す）とを算出した。

[0048] 図4には、`robust acc`と β との関係が例示されている。また、図5には、`natural acc`と β との関係が例示されている。図4に示すように、本発明（実施形態）のモデル、従来手法のモデルとも、`Adversarial Example`に対する予測精度は、 β に依存しないことがわかる。これに対し、図5に示すように、 β が大きくなるほど、本発明のモデル、従来手法のモデルとも、通常のデータに対する予測精度が低下する。これは、上記式(4)の第1項が通常のデータに対する損失関数を表す部分であり、第2項が`Adversarial Example`に対する損

失関数を表す部分であることに起因して、 β が大きいほど、第2項の影響が大きくなるためである。

[0049] そこで、robust accが高い場合の β を採用して、各モデルの正確性を比較する。その結果、従来手法のモデルでは、 $\beta = 20$ Robust Acc = 56.87、Natural Acc = 95.75であった。また、本発明のモデルでは、 $\beta = 10$ 、Robust Acc = 61.62、Natural Acc = 95.84であった。このように、 β によらず、本発明のモデルが従来手法のモデルより高いことがわかる。このように、実施形態のモデルは、上記式(4)の第2項に対応して、Adversarial Exampleに対して頑強なモデルを学習できることが確認された。

[0050] [プログラム]

上記実施形態に係る学習装置10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、学習装置10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学習処理を実行する学習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、上記の学習プログラムを情報処理装置に実行させることにより、情報処理装置を学習装置10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 Handyphone 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、学習装置10の機能を、クラウドサーバに実装してもよい。

[0051] 図6は、学習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

これらの各部は、バス1080によって接続される。

[0052] 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。

[0053] ここで、ハードディスクドライブ1031は、例えば、OS 1091、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。

[0054] また、学習プログラムは、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学習装置10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。

[0055] また、学習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、CPU 1020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RAM 1012に読み出して、上述した各手順を実行する。

[0056] なお、学習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10

41等を介してCPU1020によって読み出されてもよい。あるいは、学習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。

[0057] 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。

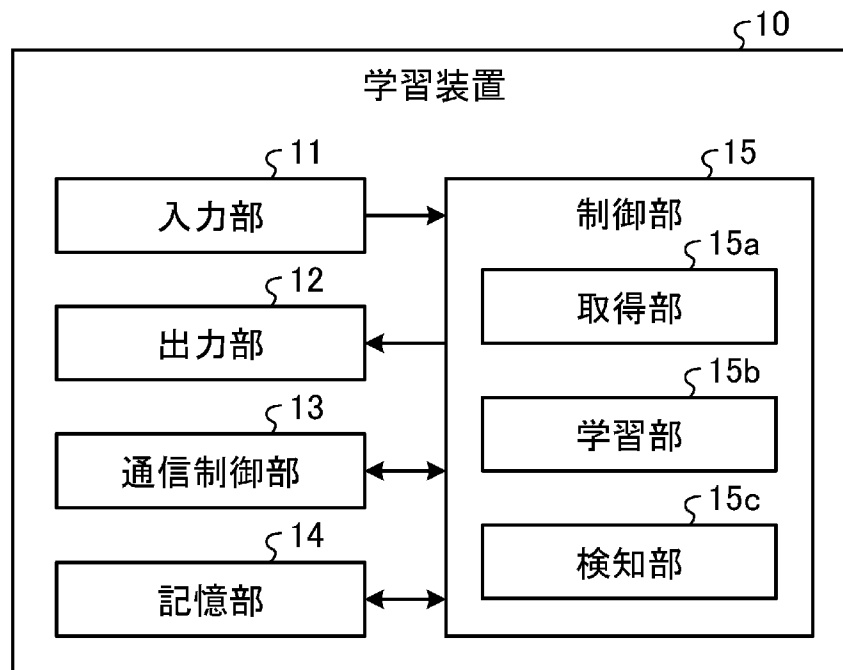
符号の説明

- [0058] 10 学習装置
- 11 入力部
- 12 出力部
- 13 通信制御部
- 14 記憶部
- 15 制御部
- 15 a 取得部
- 15 b 学習部
- 15 c 検知部

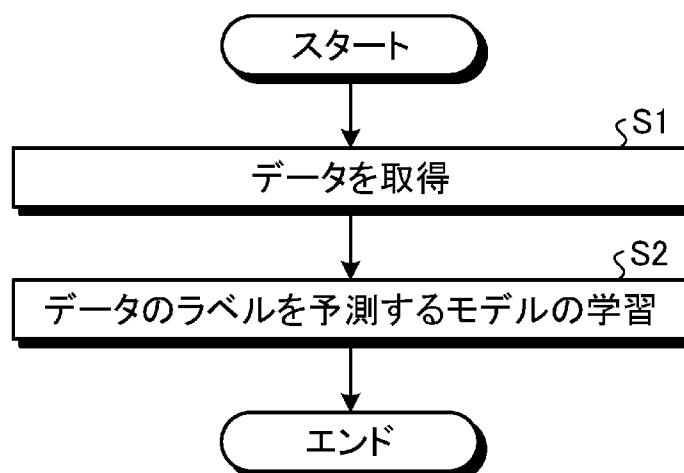
請求の範囲

- [請求項1] ラベルを予測するデータを取得する取得部と、
 取得された前記データのラベルの確率分布を表すモデルにおいて、
 前記データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、該モデルを学習する学習部と、
 を有することを特徴とする学習装置。
- [請求項2] 前記学習部は、損失関数において前記データに対して乗せるノイズの初期値として、前記固有ベクトルを用いることを特徴とする請求項1に記載の学習装置。
- [請求項3] 学習された前記モデルを用いて、取得された前記データのラベルを予測する検知部を、さらに備えることを特徴とする請求項1に記載の学習装置。
- [請求項4] 学習装置で実行される学習方法であって、
 ラベルを予測するデータを取得する取得工程と、
 取得された前記データのラベルの確率分布を表すモデルにおいて、
 前記データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、該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学習方法。
- [請求項5] ラベルを予測するデータを取得する取得ステップと、
 取得された前記データのラベルの確率分布を表すモデルにおいて、
 前記データに対するフィッシャー情報行列の最大固有値に対応する固有ベクトルを用いて、該モデルを学習する学習ステップと、
 をコンピュータに実行させるための学習プログラム。

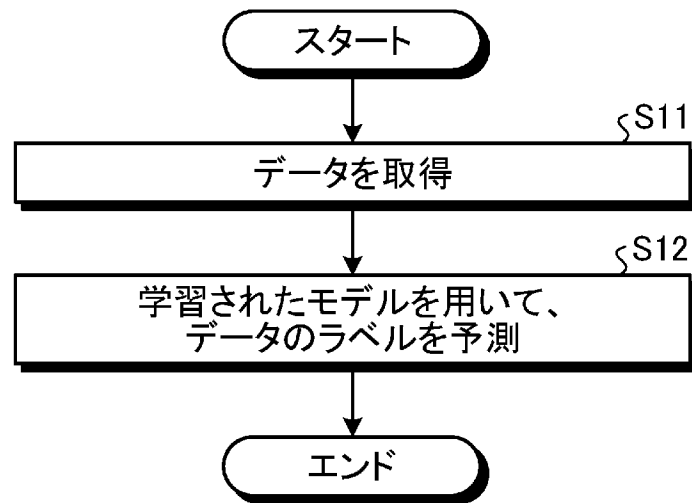
[図1]



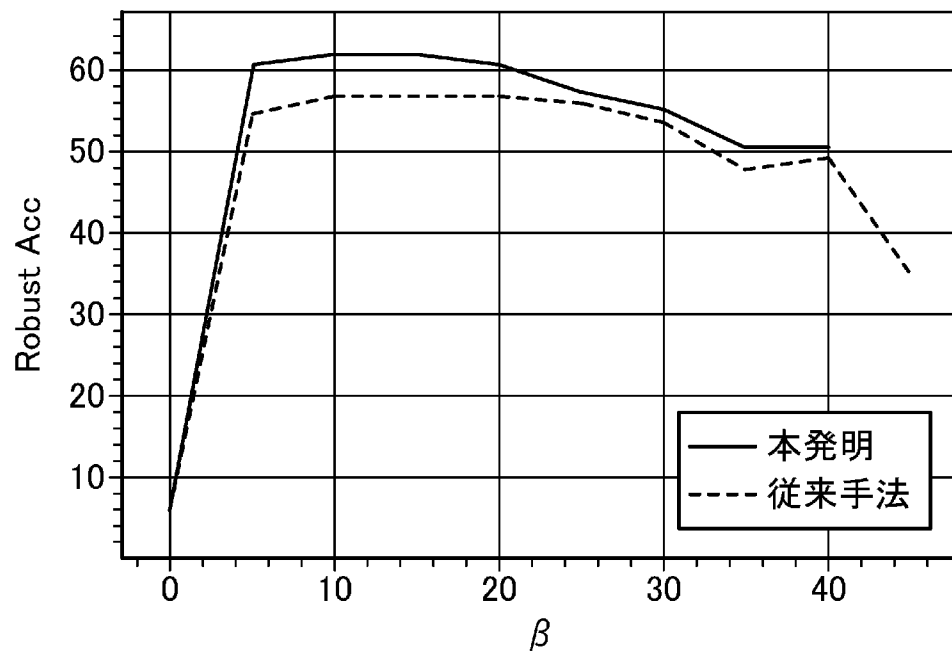
[図2]



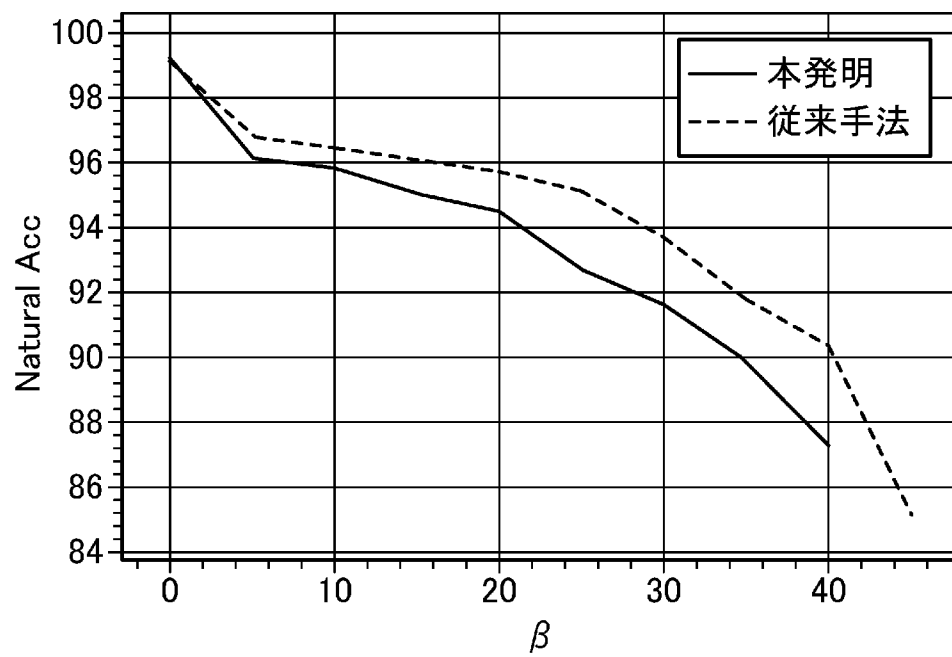
[図3]



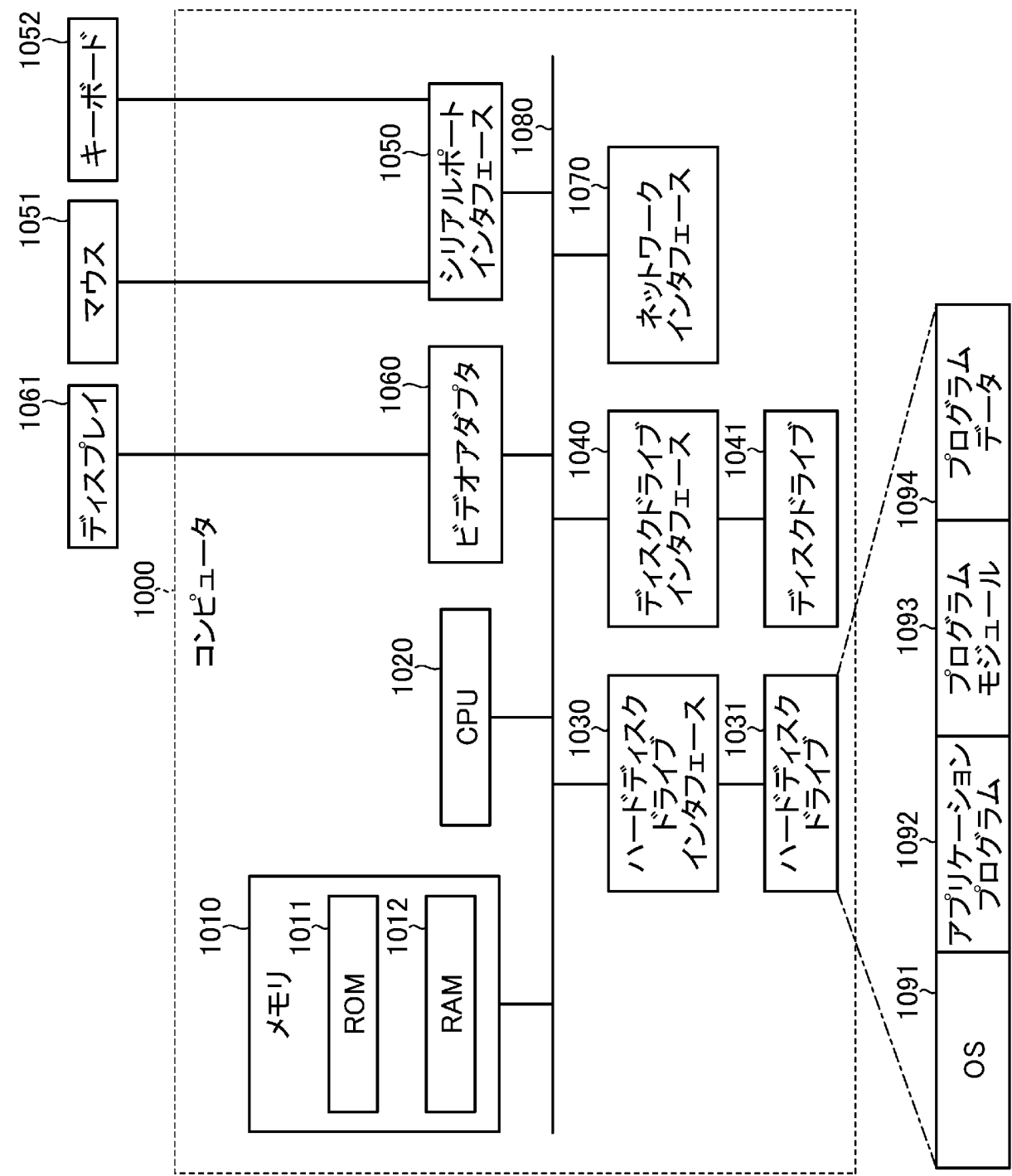
[図4]



[図5]



[図6]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/027875

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06N20/00 (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)

Int.Cl. G06N3/00-99/00

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)

IEEE Xplore

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	SHEN, C. M. et al., Defending against adversarial attacks by suppressing the largest eigenvalue of fisher information matrix, arXiv:1909.06137v1, [online], arXiv (CORNELL UNIVERSITY), 13 September 2019, version v1, pp. 1-11, [retrieved on 28 August 2020], Retrieved from the Internet: <URL:https://arxiv.org/abs/1909.06137v1> and <URL:https://arxiv.org/pdf/1909.06137v1.pdf>, abstract, p. 2, ll. 3-13, sections 3.1, 3.2, 5	1, 3-5 2

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

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 28 August 2020	Date of mailing of the international search report 08 September 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.

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）） G06N3/00-99/00 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> 国際調査で使用了電子データベース（データベースの名称、調査に使用した用語） IEEE Xplore			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												
X	SHEN, Chaomin, et al., DEFENDING AGAINST ADVERSARIAL ATTACKS BY SUPPRESSING THE LARGEST EIGENVALUE OF FISHER INFORMATION MATRIX, arXiv:1909.06137v1, [online], arXiv (Cornell University), 2019.09.13, version v1, Pages 1-11, [retrieved on 2020.08.28], Retrieved from the Internet: <URL: https://arxiv.org/abs/1909.06137v1> and <URL: https://arxiv.org/pdf/1909.06137v1.pdf>. ABSTRACT; Page 2, lines 3-13; Sections 3.1, 3.2 and 5.	1,3-5												
A		2												
☐ 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 28.08.2020	国際調査報告の発送日 08.09.2020													
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） 久保 光宏 5C 9189 電話番号 03-3581-1101 内線 3539													