

(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  
2023年1月5日(05.01.2023)

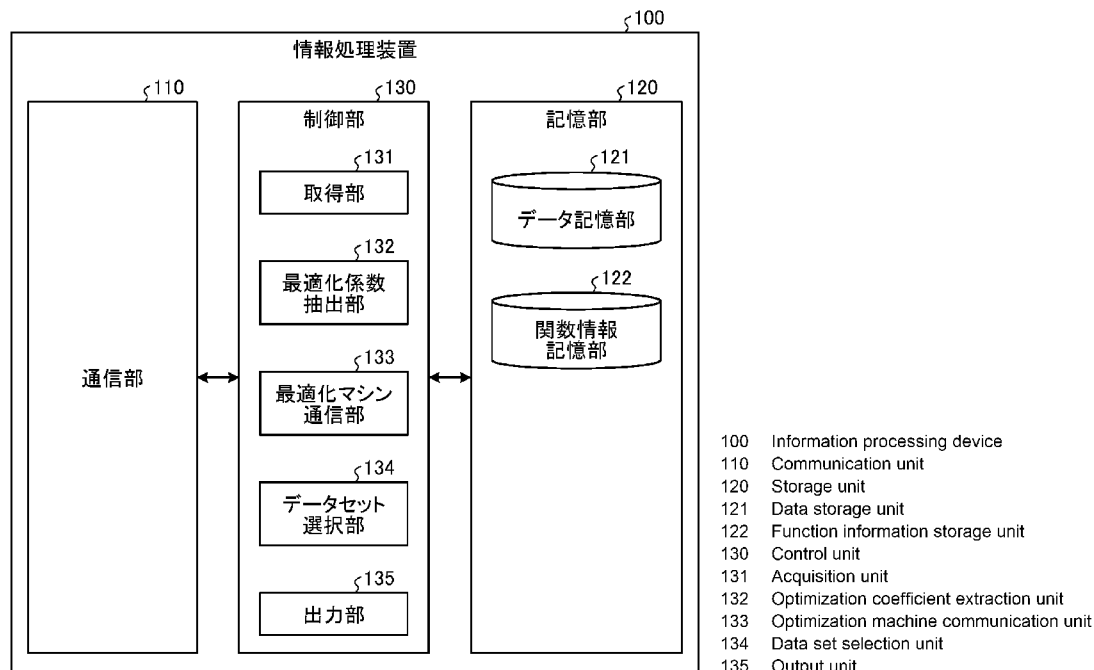


(10) 国際公開番号  
**WO 2023/276213 A1**

- (51) 国際特許分類:  
G06N 20/00 (2019.01) G06N 99/00 (2019.01)
- (21) 国際出願番号: PCT/JP2022/002805
- (22) 国際出願日: 2022年1月26日(26.01.2022)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:  
特願 2021-109634 2021年6月30日(30.06.2021) JP
- (71) 出願人: ソニーグループ株式会社(SONY GROUP CORPORATION) [JP/JP]; 〒1080075 東京都港区港南1丁目7番1号 Tokyo (JP).
- (72) 発明者: 井手 直紀(IDE, Naoki); 〒1080075 東京都港区港南1丁目7番1号 ソニーグループ株式会社内 Tokyo (JP).
- (74) 代理人: 弁理士法人 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,

(54) Title: INFORMATION PROCESSING DEVICE, INFORMATION PROCESSING METHOD, AND INFORMATION PROCESSING PROGRAM

(54) 発明の名称: 情報処理装置、情報処理方法及び情報処理プログラム



(57) Abstract: An information processing device according to the present disclosure comprises: an acquisition unit which acquires a data providing method, a model intended to be trained, and designation information pertaining to the size and category of a sample set to be used for training the model; and a selection unit which selects, from a data set, the sample set to be used for training the model on the basis of information entropy determined according to the model and the designated information.

WO 2023/276213 A1

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：本開示に係る情報処理装置は、データ供給方法と、学習したいモデルと、前記モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とを取得する取得部と、前記モデルに応じて決定される情報エントロピーと、前記指定情報とに基づいて、データセットから、前記モデルの学習に用いるサンプルセットを選択する選択部と、を備える。

## 明 細 書

発明の名称：

情報処理装置、情報処理方法及び情報処理プログラム

### 技術分野

[0001] 本開示は、情報処理装置、情報処理方法及び情報処理プログラムに関する。

### 背景技術

[0002] 従来、深層学習では、予測モデル等のモデルの構造や学習方法の開拓による性能向上が行われていた。このような性能向上においては、開発者等のユーザの仕事は、独自のモデルの構造や学習方法を考案し、その性能を、例えば一般物体画像分類等のタスクと、例えばイメージネット等のデータセットとにより、ベンチマーク評価することが中心であった（例えば非特許文献1）。

[0003] また、深層学習等の機械学習においては、実世界での活用へとフェーズが変遷している。このため、ユーザは、実現したい独自の予測機能等のタスクを考案し、これを実現するための独自のデータを収集することが必要となっている。

### 先行技術文献

#### 非特許文献

[0004] 非特許文献1：“Deep Residual Learning for Image Recognition”，  
Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun<インターネット><https://arxiv.org/abs/1512.03385>（令和3年6月23日検索）

### 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、ユーザ自身がタスクの実現のために、学習に用いるデータを選択することは難しいという課題がある。例えば、大量のデータが含まれるデータセットの中から、ユーザが自身のタスクを実現するモデルを学習す

るために適切なデータセット等のサンプルセットを選択することは困難である。そのため、データセットから適切なサンプルセットを選択することが望まれている。

[0006] そこで、本開示では、データセットから適切なサンプルセットを選択することができる情報処理装置、情報処理方法及び情報処理プログラムを提案する。

### 課題を解決するための手段

[0007] 上記の課題を解決するために、本開示に係る一形態の情報処理装置は、データ供給方法と、学習したいモデルと、前記モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とを取得する取得部と、前記モデルに応じて決定される情報エントロピーと、前記指定情報とに基づいて、データセットから、前記モデルの学習に用いるサンプルセットを選択する選択部と、を備える。

### 図面の簡単な説明

- [0008] [図1]本開示の情報処理システムの構成例を示す図である。  
[図2]情報処理システムによる処理手順を示すフローチャートである。  
[図3]情報処理システムのブロック構成の一例を示す図である。  
[図4]ユーザインターフェースの一例を示す図である。  
[図5]情報処理システムのブロック構成の一部を示す図である。  
[図6]トンネル効果を模式的に示す図である。  
[図7]情報処理のメインプログラムの一例を示す図である。  
[図8]データ選択最適化の処理の手順を示すフローチャートである。  
[図9]データ選択最適化の処理プログラムの一例を示す図である。  
[図10]本開示の情報処理装置の構成例を示す図である。  
[図11]本開示の最適化マシンの構成例を示す図である。  
[図12]クラスのリストの一例を示す図である。  
[図13]ユーザが学習したいクラスの選択の一例を示す図である。  
[図14]ユーザからのデータの受けの一例を示す図である。

[図15]情報処理システムを適用した第5の実施例を示す図である。

[図16]第5の実施例における処理の手順を示すフローチャートである。

[図17]画像判定処理の流れを示す図である。

[図18]情報処理装置の機能を実現するコンピュータの一例を示すハードウェア構成図である。

## 発明を実施するための形態

[0009] 以下に、本開示の実施形態について図面に基づいて詳細に説明する。なお、この実施形態により本願にかかる情報処理装置、情報処理方法及び情報処理プログラムが限定されるものではない。また、以下の各実施形態において、同一の部位には同一の符号を付することにより重複する説明を省略する。

[0010] 以下に示す項目順序に従って本開示を説明する。

### 1. 概要

#### 1-1. 背景（課題等）

#### 1-2. 本開示の概要（処理・効果等）

### 2. 第1の実施例

#### 2-1. 情報処理システムの装置構成

#### 2-2. 情報処理システムの全体処理フロー

#### 2-3. 情報処理システムのブロック全体像

##### 2-3-1. 最適化マシン

##### 2-3-2. ユーザーインターフェース

##### 2-3-2-1. ユーザーインターフェース部（データ入力部）

##### 2-3-2-2. ユーザーインターフェース部（マシン選択部）

##### 2-3-2-3. ユーザーインターフェース部（データ出力部）

##### 2-3-3. 情報処理装置（メインブロック）

### 3. 第2の実施例

#### 3-1. 量子アニーリングと組合せ最適化問題

#### 3-2. データセットからのサンプル抽出最適化

#### 3-3. リサンプリング最適化の量子アニーリング等への実装

4. 第3の実施例
  - 4-1. 回帰モデル
  - 4-2. ニクラス分類モデル
  - 4-3. 多クラス分類モデル
5. 第4の実施例
  - 5-1. 情報処理のメインプログラム例
  - 5-2. データ選択最適化の処理フロー例
  - 5-3. データ選択最適化の処理プログラム例
6. 情報処理装置の構成
7. 最適化マシンの構成
8. ユーザへの提示例
  - 8-1. クラスのリスト
  - 8-2. クラスの選択
  - 8-3. データの受け付け
9. 適用例
  - 9-1. 第5の実施例（画像分類）
  - 9-2. 第6の実施例（生体分子）
10. その他の構成例等
  - 10-1. その他の構成例
  - 10-2. プログラムやパラメータ等の処理に用いる情報の生成方法
  - 10-3. その他
11. 本開示に係る効果
12. ハードウェア構成

[0011] [1. 概要]

本開示の詳細を説明する前に、本開示に係る技術的な背景の概要及び本開示の概要について説明する。なお、以下の技術的な説明の中で従来技術に関する詳細な説明は適宜省略する。

[0012] [1-1. 背景（課題等）]

まず、本開示に関連する課題について以下詳述する。深層学習等の機械学習においては、モデルの学習に用いるデータセットの収集は容易ではなく、データセットを収集するコストの軽減は課題の一つである。データセットを収集するコストの削減を、学習方法で解決しようとする試みはいくつか知られており、例えば以下の三つの学習方法が挙げられる。

・半教師学習： ラベリングされていないデータをそのまま活用する学習方法

・能動学習： ラベリングされていないデータをラベリングしながら学習する方法

・転移学習： 別のデータセットで学習済みのモデルを活用する学習方法

[0013] 上記の三つの学習方法のうち、前者二つ、すなわち半教師学習及び能動学習は、データセット収集（構築）の中でも、ラベリングのコストに対する対策である。ラベリングとは、データに情報を追加する作業で、通常人手でなされるためコストが高い。

[0014] また、ラベリングには、科学実験に基づいた物理量の計測や、病理画像の診断等、高度な知識、作業がなければ実現が難しい処理も含まれる。たとえば、半導体のバンドギャップの計測や、有機・無機分子の光学特性、反応速度、平衡乖離度の計測情報の計測等も、これらをデータとして機械学習の予測モデルを学習するという観点からはラベリングとみなすことができる。

[0015] また、能動学習は、データセットから一ないし複数のデータを適切に選んでラベリングし、これを予測モデル（以下単に「モデル」ともいう）の更新に用いる学習方法である。モデルに応じて、ラベリングすべきデータを効率よく選べば、全てのデータにラベリングする必要がなくなり、ラベリングのコスト削減につながる。

[0016] 一方、上記の三つの学習方法のうち、後者一つ、すなわち転移学習は、ラベリングのコスト削減よりも別のデータセットの情報の流用に主眼をおく技術である。転移学習では、ユーザは、システムから別のデータセットの情報を埋め込んだモデルをダウンロードして、このモデルのパラメータ（の一部）を手元のデータセットで更新する。

[0017] また、転移学習では、システムから配布されるモデルは、一般にサイズ（フットプリント）が大きい傾向があるが、これは様々な情報に対応できるサイズにしたことが原因として挙げられる。しかし、転移先のユーザからすると、予測対象以外の情報処理のために時間とエネルギーを消費していることとなる。つまり、ユーザ側からするとモデルに対する情報の効率が悪い。

[0018] また、データセットの情報がモデルパラメータに変換されているため、どのようなデータを学習しているのかわからない点も課題として挙げられる。昨今、人工知能活用の倫理性が問われており、より、公明な学習には、データそのものが供給される方が望ましい。

[0019] 以上を鑑みると、ユーザが持っている予測モデル等、ユーザが所望するモデルに応じて、データセットが有する情報を過不足なく集約したサブセットを効率よく探索して提供することが望まれている。

[0020] [ 1 - 2. 本開示の概要（処理・効果等） ]

そこで、以下で詳細を説明する情報処理システム 1（図 1 参照）は、例えばユーザからモデルが提示された場合、情報処理システム 1 がアクセスできるデータセットから、ユーザのモデルにとって最も効率よく情報をもたらすサブセットを提供する。

[0021] 例えば、情報処理システム 1 は、データセットの情報を学習したモデルを共通に配布する代わりに、ユーザの持っているモデルにカスタマイズされたモデルを更新するために最適な（サンプル）サブセットを選択し、選択したサブセットを提供する。情報処理システム 1 が提供する情報の一例としては、データセットやデータシミュレータからサンプルされたサブセットがある。なお、情報処理システム 1 は、必要に応じてそのサブセットで学習したモデル等の様々な情報を提供してもよいが、この点についての詳細は後述する。

[0022] 情報処理システム 1 が選択するサブセットは、ユーザが持っているモデルに対して、最もモデルの曖昧さを減らすデータセット、言い換えると、最も情報抽出の効率が大きくなるよう選択されたサブセットである。情報処理シ

ステム 1 によるサブセットの選択の最適化は、最適化処理の高精度化のため、量子アニーリングマシン、または二値二次形式組合せ最適化の高速化に特化した組合せ最適化アクセラレータ等の任意の組合せ最適化マシン（以下「最適化マシン」ともいう）を用いて実現される。

[0023] 情報処理システム 1 によれば、ユーザは、現時点での情報で生成したモデルの更新に使うデータを、既存のデータセット等から効率よく収集できる。このため、情報処理システム 1 によれば、開発期間と開発経費の削減と試行回数の増加の効果を得られる。

[0024] 情報処理システム 1 によれば、ユーザは、モデルの更新に最も寄与の大きいサブセットを取得できる。このため、情報処理システム 1 によれば、モデルは、予測したい情報だけに最適化され、機能の性能向上が実現される。

[0025] また、情報処理システム 1 によれば、ユーザは、ユーザが持っているモデルの更新に最も寄与の大きいラベルなしデータのサブセットを得ることができ。このため、情報処理システム 1 によれば、知見が最も多いと予想される分析対象のデータから順にラベリング、あるいは、ラベリングに相当する分析作業を実現できる。

[0026] また、情報処理システム 1 によれば、サブセットの最適化は、量子アニーリングマシン等の最適化マシンを用いて実現される。量子アニーリングは漸近的に大域最適解に到達することが理論的に知られているので、従来の組合せ最適化の実装アルゴリズムの代表例である貪欲法の近似よりも最適なサブセットを提供することができる。まず、以下では、情報処理システム 1 全体のシステム構成例等を第 1 の実施例として説明し、その後に各処理や適用例について説明する。

[0027] [2. 第 1 の実施例]

[2-1. 情報処理システムの装置構成]

まず、図 1 に示す情報処理システム 1 の構成について説明する。図 1 は、本開示の情報処理システムの構成例を示す図である。図 1 に示すように、情報処理システム 1 は、最適化マシン 10 と、端末装置 20 と、データベース

50と、情報処理装置100とが含まれる。例えば、情報処理システム1は、大量データを格納したデータベース50、データセットからサブセットを生成するサーバである情報処理装置100、ユーザが情報処理装置100にリクエストを出し、サブセットを受け取るための端末装置20、及び、情報処理装置100がサブセットを生成するために利用する最適化マシン10で構成される。

[0028] 情報処理装置100と、最適化マシン10と、端末装置20と、データベース50とは所定の通信網（ネットワークNT）を介して、有線または無線により通信可能に接続される。なお、情報処理システム1には、複数の情報処理装置100や、複数の最適化マシン10や、複数の端末装置20や、複数のデータベース50が含まれてもよい。

[0029] 最適化マシン10は、組合せ最適化問題を解くために用いられるコンピュータ（最適化計算専用マシン）である。例えば、最適化マシン10は、最適なサブセットを作るときに用いる最適化処理のアクセラレータである。最適化マシン10としては、量子アニーリングマシンを筆頭にイジングマシン等の組合せ最適化マシン（離散最適化専用マシン）等がある。なお、最適化マシン10の具体例については後述する。

[0030] 端末装置20は、ユーザに利用されるコンピュータである。端末装置20は、情報処理装置100にサブセットをリクエストし、情報処理装置100からサブセットを受信する。例えば、このサブセットは、モデルの学習に利用されてもよく、実際にモデルを端末装置20、または、端末装置20の後段で学習する際に利用されてもよい。例えば、サブセットは、モデルのサイズを集約する際にリファレンスのセットとして利用される。なお、情報処理システム1では、サブセットで学習済みのパラメータを供給してもよい。

[0031] 端末装置20からのサブセットの要求の態様は、モデル自体の入力（送信）であってもよい。また、ユーザが事前にモデルを作成する代わりに、情報処理システム1側から提示されるガイダンスに沿ってモデルを決めるようにしても良い。さらに、サブセットの仕様に関する要求として、例えば、端末

装置 20 は、サブセットのサイズ（データの数等）の指定を受け付けてもよい。ユーザによるサブセットのデータ数の指定を受け付けた場合、端末装置 20 は、ユーザが指定したサブセットのデータ数を示す指定情報を情報処理装置 100 に送信する。端末装置 20 は、ユーザが収集したデータセットを情報処理装置 100 に送信してもよい。

[0032] 端末装置 20 は、ユーザから各種操作が入力される。端末装置 20 は、ユーザによる入力を受け付ける。端末装置 20 は、ユーザによる最適化マシン 10 の選択を受け付ける。端末装置 20 は、表示したコンテンツに対するユーザの操作を入力として受け付ける。なお、端末装置 20 は、上述した処理を実現可能であれば、どのような装置であってもよい。例えば、端末装置 20 は、スマートフォンや、タブレット型端末や、ノート型 PC (Personal Computer) や、デスクトップ PC や、携帯電話機や、PDA (Personal Digital Assistant) 等の装置であってもよい。

[0033] データベース 50 は、大規模データセットを記憶するデータベースである。データベース 50 に記憶されるデータセットは、様々な用途に転移できるモデルを学習できる規模のデータセット（大規模データセット）である。

[0034] 情報処理装置 100 は、データベース 50 に記憶されたデータセットから、モデルの学習に用いるサブセットやシミュレータの出力等のサンプルセットを選択するサーバ装置（コンピュータ）である。情報処理装置 100 は、データセットからのデータ供給やシミュレーション等のデータ供給方法を基に、サンプルセットを選択する。例えば、データ供給方法は、データセットからのデータ供給である。また、例えば、サンプルセットはデータセットのサブセットである。情報処理装置 100 は、学習したいモデルに応じて決定される情報エントロピーと、モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とに基づいて、データセットから、モデルの学習に用いるサンプルセットを選択する。例えば、学習したいモデルは、学習パラメータ付きの予測モデルである。また、例えば、モデルのタスクとは、モデルの入力に対応するモデルの出力の種別である。また、例えば、情報エ

ントロピーは、カルバックライブラー情報量、または、フィッシャー情報量を用いて算出される情報エントロピーであるが詳細は後述する。例えば、情報処理装置100は、最適化マシン10、端末装置20、及びデータベース50との間の通信により処理に必要な情報を受信し、受信した情報を用いてモデルの学習に用いるサブセットを選択する。例えば、情報処理装置100は、端末装置20のリクエストに応じて、データベース50からサブセットを選択し、選択したサブセットを端末装置20に送信する。

[0035] なお、情報処理システム1におけるサブセットの選択においては、最適化マシン10による最適化がなされる。情報処理システム1は、最適と判定されたサブセットを作る際に、端末装置20からのデータセットをデータベース50に記憶されたデータセット等の大規模データセットに追加してもよい。

[0036] [2-2. 情報処理システムの全体処理フロー]

次に、図2を用いて情報処理システムの全体処理フローを説明する。図2は、情報処理システムによる処理手順を示すフローチャートである。

[0037] 図2に示すように、まず、情報処理装置100は、端末装置20からデータセットの要求を受け付ける（ステップS1）。情報処理装置100は、端末装置20からデータセットの提供のリクエストを受けとる。例えば、リクエストは、モデルそのものが特定のフォーマットであってもよく、端末装置20から情報処理装置100に送信される。他には、リクエストとして、ユーザが扱いたいドメインを構成するカテゴリのクラス名等が、端末装置20から情報処理装置100に送信されてもよい。また、リクエストとして、ユーザが収集したデータセットそのものが、端末装置20から情報処理装置100に送信されてもよい。

[0038] そして、情報処理装置100は、最適化マシン10に、データの最適な組合せ情報を要求する（ステップS2）。情報処理装置100は、データベース50に記憶されたデータセット等の大規模データからユーザのリクエストに対して最適なデータを選択するため、最適化マシン10に必要な情報を送

信する。この情報は、端末装置 20 に提供するデータサイズの制約と、データの組合せが持つ情報量に関するものであるが、詳細は後述する。

[0039] そして、情報処理装置 100 は、最適化に応じて、端末装置 20 へデータを提供する（ステップ S3）。情報処理装置 100 は、最適化マシン 10 からの結果に基づいて、データベース 50 に記憶されたデータセットのサブセットを生成し、端末装置 20 へ生成したデータセットを提供する。

[0040] そして、情報処理装置 100 は、次の要求があるまで待機し（ステップ S4）、次の要求があった場合、ステップ S1 に戻って処理を繰り返す。また、情報処理装置 100 は、次の要求がなく所定の期間が経過した場合、処理を終了する。

[0041] [2-3. 情報処理システムのブロック全体像]

次に、図 3 を用いて情報処理システムのブロック全体像を説明する。図 3 は、情報処理システムのブロック構成の一例を示す図である。

[0042] 図 3 に示すように、情報処理システム 1 には、ユーザインターフェース部 21、22、23、メインブロックでありデータセット生成部として機能する情報処理装置 100、メインブロックに接続されるデータセットのストレージであるデータベース 50、さらに、最適化処理に特化した複数の最適化マシン 10 で構成されている。

[0043] [2-3-1. 最適化マシン]

情報処理システム 1 には、最適化マシン 10 a、最適化マシン 10 b、最適化マシン 10 c 等の複数の最適化マシン 10 が含まれる。なお、最適化マシン 10 a、最適化マシン 10 b、最適化マシン 10 c 等について、特に区別せずに説明する場合は、「最適化マシン 10」と記載する。なお、図 3 では、最適化マシン 10 a、最適化マシン 10 b、最適化マシン 10 c の 3 台の最適化マシン 10 を示すが、情報処理システム 1 には、4 台以上の最適化マシン 10 が含まれてもよい。最適化マシン 10 は、イジングモデルを用いるコンピュータ（計算機）である。最適化マシン 10 は、イジングモデルを用いて、組合せ最適化問題を解くコンピュータ（組合せ最適化マシン）であ

る。

[0044] 例えば、最適化マシン10aは、量子アニーリングを用いて問題を解く量子コンピュータ（量子計算機）である。最適化マシン10aは、量子アニーリング方式や量子アニーラにより実現することができる。例えば、最適化マシン10bは、CMOS（Complementary metal-oxide-semiconductor）等を用いた組合せ最適化アクセラレータである。例えば、最適化マシン10cは、GPU（Graphics Processing Unit）等のプロセッサやFPGA（Field Programmable Gate Array）等の集積回路を用いた組合せ最適化アクセラレータである。なお、最適化マシン10は、上記の最適化マシン10aや最適化マシン10bや最適化マシン10c等のハードウェア構成に限らず、どのようなハードウェア構成であってもよい。

[0045] [2-3-2. ユーザインターフェース]

情報処理システム1には、データ入力部であるユーザインターフェース部21、マシン選択部であるユーザインターフェース部22、及びデータ提示部であるユーザインターフェース部23が含まれる。なお、ユーザインターフェース部22はオプションであり、ユーザによる最適化マシン10の選択を受け付けない場合は、情報処理システム1に含まれてなくてもよい。

[0046] 例えば、ユーザインターフェース部21、22、23は、端末装置20により実現される。すなわち、端末装置20は、ユーザインターフェース部21、22、23に対応する機能を有し、ユーザによるデータ入力の受け付け、ユーザによる最適化マシン10の選択の受け付け、及び提供された情報の提示等の各種の処理を行う。

[0047] [2-3-2-1. ユーザインターフェース部（データ入力部）]

まず、データ入力部であるユーザインターフェース部21に対応する機能について説明する。ユーザインターフェース部21は、ユーザが欲しいデータセットの情報や、カスタマイズしてほしい対象となるモデル、実際にユーザが集めたデータセットを入力するためのインターフェースである。ユーザインターフェース部21は、コマンドラインインターフェースでも、グラフ

ックインターフェースでもよい。

[0048] 例えば、最もシンプルな入力、モデルだけの入力である。例えば、ユーザインターフェース部 21 は、深層学習で学習するニューラルネットワークモデルを、フォーマットに従って生成したファイルで入力する。なお、より詳しい情報を獲得できるようにするため、以下の4つの情報等をユーザに要求してもよい。例えば、ユーザインターフェース部 21 は、以下の4つの情報等の入力を受け付ける。

[0049] ・タスク（実現したい機能）はなにか（回帰かクラス判別か分類か。クラス数は一つか複数か）

・データはラベリング済みを使うか自分でラベリングするか

・データのドメインはなにか（画像、音声、言語、項目データ・・・）

・ドメインやクラス名はシステムが保持している既存のクラスか未知のクラスか

例えば、情報処理システム 1 においては、上記の4つの情報等のアンケートの答えに従って内部で処理が分岐していくものとする。

[0050] [2-3-2-2. ユーザインターフェース部（マシン選択部）]

次に、マシン選択部であるユーザインターフェース部 22 に対応する機能について説明する。ユーザインターフェース部 22 は、データセットからデータのサンプリングする際の最適化を実現するとき、最適化方法としてどのような方法を使うかを選択するためのインターフェースである。組合せ最適化を高速に実現するための最適化計算特化型計算機にはすでに様々なマシン（最適化マシン 10）があり、クラウド等を介してユーザが利用することもできる。

[0051] なお、最適化マシン 10 の各々には、メリット・デメリットや得意・不得意等がある。例えば、ユーザインターフェース部 22 は、グラフィカルユーザインターフェース（Graphical User Interface：GUI）等を介して、図 4 に示すような各マシン（最適化マシン 10）の情報を含むコンテンツ 1 を提示してもよい。この場合、コンテンツ 1 を確認したユーザが目

的に応じて、所望の最適化マシン10を選択する。

[0052] 図4は、ユーザインターフェースの一例を示す図である。図4のコンテンツCT1では、マシンA~Dの4つの最適化マシン10が選択候補として提示される場合を示す。コンテンツCT1には、マシンA~Dの4つの最適化マシン10の各々について、サイズ、全結合の可否、精度、速度、通信、価格等の複数の項目についての評価を示す情報が含まれる。なお、図4に示す提示方法は一例に過ぎず、ユーザへの提示方法はユーザが所望の最適化マシン10を選択可能であれば、どのような形態であってもよい。

[0053] 例えば、端末装置20からの要求に応じて、情報処理装置100は、コンテンツCT1を生成し、生成したコンテンツCT1を端末装置20に送信する。そして、コンテンツCT1を受信した端末装置20は、コンテンツCT1を表示する。端末装置20を利用するユーザは、端末装置20に表示されたコンテンツCT1を確認し、マシンA~Dの4つの最適化マシン10の各々のメリット・デメリット及び得意・不得意を比較した上で、利用する最適化マシン10を選択する。ユーザによる最適化マシン10の選択を受け付けた端末装置20は、ユーザが選択した最適化マシン10を示す情報を情報処理装置100に送信する。ユーザが選択した最適化マシン10を示す情報を受信した情報処理装置100は、ユーザが選択した最適化マシン10に対して処理を要求する。

[0054] [2-3-2-3. ユーザインターフェース部（データ出力部）]

次に、データ出力部であるユーザインターフェース部23に対応する機能について説明する。ユーザインターフェース部23は、ユーザが実現したい機能に合わせて提供するデータセットのサブセットを提示するインターフェースである。ユーザインターフェース部23は、コマンドラインインターフェースでも、グラフィックインターフェースでもよい。

[0055] 例えば、ユーザインターフェース部23は、サブセット（データセット）を供給する。なお、ユーザインターフェース部23は、サブセットに限らず、サブセットで学習したパラメータを供給してもよい。例えば、ユーザイン

ターフェース部23は、サブセットとともに、そのサブセットで学習したパラメータを供給してもよい。また、ユーザインターフェース部23は、ユーザからの指定によって、サブセットではなく元のデータセットで学習したパラメータを提示してもよい。

[0056] ユーザインターフェース部23からの供給を受けたユーザは、得られたモデルのパラメータを直接利用することもできる。一方で、情報処理システム1では、学習に必要なデータセットが得られている。このように、学習に必要なデータセットが得られているので、情報処理システム1では、実現したい機能をより効率的に、また、高性能になるように設計、学習しなおすことができる。

[0057] [2-3-3. 情報処理装置（メインブロック）]

次に、メインブロックである情報処理装置100等の情報処理システム1のコア部分について、図3及び図5を用いて説明する。図5は、情報処理システムのブロック構成の一部を示す図である。図5に示す情報処理システム1のコア部分には、データベース50、データ読み取り部131a、ユーザデータ入力部131b、最適化係数抽出部132、最適化マシン通信部133、データセット選択部134、及び出力部135が含まれる。

[0058] 図5中の各構成のうち、データベース50以外は、情報処理装置100の構成であり、図3中の点線内に示す構成に対応する。なお、図5において図3と同様の点については同様の符号を付す。なお、最適化係数抽出部132は、図3中に示す情報ゲイン計算部132a及びQUBO係数算出部132bを一体として構成したものである。すなわち、最適化係数抽出部132は、情報ゲインに関する計算を行う情報ゲイン計算部132a、及びQUBO（Quadratic Unconstrained Binary Optimization：二次形式バイナリ最適化）の係数を算出するQUBO係数算出部132bの両方の機能を有する構成要素である。

[0059] 以下、図5に示す各構成について説明する。データベース50は、横断的な大量のデータ（大規模データ）が保存されているデータベースである。例

例えば、データベース50は、イメージネット等の任意のデータセットが保存される。

[0060] データ読み取り部131aは、データベースを読み取る機能を有する。例えば、データ読み取り部131aは、データベース50からデータを取得する。データ読み取り部131aは、ユーザからのリクエストにおいて何らかのクラスの指定があった場合には、そのクラスのデータを取得する。

[0061] ユーザデータ入力部131bは、ユーザからのリクエストを受け取る機能を有するモジュールである。例えば、ユーザからのリクエストには以下の二通りの入力パターンがある。例えば、第一のパターンは、ユーザが学習したい予測モデルが入力されるパターンである。また、第二のパターンは、ユーザがすでに持っているデータが入力されるパターンである。

[0062] 最適化係数抽出部132は、大規模データベースのデータとユーザが入力したデータから後述する組合せ最適化の定式化に基づいた最適化の目的関数の係数情報を抽出する。最適化マシン通信部133は、最適化マシン10に目的関数の係数情報を送信して、最適化処理後の変数を受信する。データセット選択部134は、最適化の結果に基づいて、サブセット（データセット）を選択する。出力部135は、選択の結果に基づいて、情報を出力する。なお、情報処理装置100の各構成の詳細は後述する。

[0063] [3. 第2の実施例]

ここから、第2の実施例として、本開示の中心部分である量子アニーリングを使ったデータ抽出最適化の実現方法を説明する。以下では、本体のデータの組合せの最適化の説明に先立って、量子アニーリングと組合せ最適化について説明する。その後、データセットからのサンプル抽出の組合せ最適化の定式化を示す。最後に、サンプル抽出の組合せ最適化の組合せ最適化マシン上への実装方法を説明する。

[0064] [3-1. 量子アニーリングと組合せ最適化問題]

量子アニーリングは、イジングモデルのエネルギー基底状態を探索する方法の一つであり、組合せ最適化の高速化という用途に特化して用いられる量

子計算の一種と考えられている。イジングモデルは、もともと強磁性体（磁石）の温度に対する相転移現象を説明するために、物理学者イジングにより提案された物理モデルである。以下の式（1）は、もとの磁石のモデルを一般化したイジングモデルのハミルトニアン（エネルギー関数）であり、このモデルをもとのイジングによるモデルと区別して、スピングラスモデルと呼ぶこともある。

[0065] [数1]

$$H = - \sum_{i,j} J_{ij} \sigma_i \sigma_j + \sum_k h_k \sigma_k \quad \dots (1)$$

[0066] 式（1）の $\sigma_i$ は、スピンと呼ばれる物理量で-1か1をとる変数であり添え字は通し番号を表している。また、 $J_{ij}$ は、結合エネルギーと呼ばれる定数で、2つのスピンの間に働く相互作用の大きさを表し、添え字は二つのスピンの通し番号を表している。 $h_k$ は、局所磁場とよばれる定数で、スピンの間に働く局所磁場の大きさを表し、添え字は局所磁場が作用するスピンの番号である。

[0067] 物理学ではエネルギー関数の最小化する状態を基底状態といい、十分ゆっくりと遷移させた物理システムは、エネルギー最小状態、すなわち、基底状態に到達するとされている。組合せ最適化問題は、集合の形式で表される最適解を探索する問題である。組合せ最適化問題は、巡回セールスマン問題を筆頭に、厳密な最適解を求めるには組合せ爆発を起こす全探索が必要で計算困難な問題が多数あることが知られている。組合せ最適化のなかでも、以下の式（2）に示す形式の目的関数を有するQUBOと称される二次形式制約なしバイナリ最適化と、適切な変数変換の下で等価である。

[0068] [数2]

$$L = \sum_{i,j} Q_{ij} b_i b_j \quad \dots (2)$$

[0069] ここで、式（2）の $b_i$ は0か1の二値のいずれかを取るバイナリ変数であり、 $Q_{ij}$ はQUBOの係数行列である。

[0070] QUBOは、定数項を除いてイジングモデルのエネルギー関数と等価な形

式に変換できる。具体的には、 $b_i$ を、以下の式（3）のように、 $\sigma_i$ に置き換えることで確認できる。また、QUBO問題の係数行列 $J_{ij}$ 、 $h_k$ は、以下の式（4）、（5）のように表現される。

[0071] [数3]

$$b_i = (1 - \sigma_i)/2 \quad \dots (3)$$

[0072] [数4]

$$J_{ij} = -\frac{1}{4}Q_{ij} \quad \dots (4)$$

[0073] [数5]

$$h_i = \frac{1}{4} \sum_j Q_{ij} + Q_{ji} \quad \dots (5)$$

[0074] そこで、QUBOを実現するアルゴリズムを考える代わりに、最適化はこの係数でイジングモデルの物理的な基底状態の実現に任せようという考えで考案されたのが量子アニーリングである。

[0075] 量子アニーリングでは、イジングモデルに量子揺らぎの項を加えた以下の式（6）のようなモデルを扱う。

[0076] [数6]

$$H_{QA}(t) = (1 - \Gamma(t))H_{Ising} + \Gamma(t) \sum_k \sigma_k^x \quad \dots (6)$$

[0077] 式（6）に示すモデルで最初は、量子揺らぎの項を支配的にして、すべての組合せの重ね合わせ状態を実現しておき、徐々にイジングモデルのハミルトニアンを支配的にしていく。この過程が断熱的であるという条件で、最終的には、スピンの組合せが基底状態、つまり、エネルギーを最小にする状態に到達する。そこであらかじめQUBO問題の係数行列 $J_{ij}$ 、 $h_k$ をイジングモデルのハミルトニアンに設定しておけばイジングモデルの最終的なスピンの組み合わせがバイナリ変数に対応する。このようにして、組合せ最適化の問題をイジングモデルに埋め込み、量子揺らぎの状態から断熱的にイジングモデルの基底状態に遷移させることでイジングスピンに組合せ最適化の解を実現

させることを量子アニーリングと呼ぶ。

[0078] 例えば、式(6)では、 $t=0$ では重ね合わせの項が支配的であり、全組合せ(解の候補)を重ね合せた状態となる。そして、式(6)では、ステップ的に重ね合わせの項をゆるめて(例えば $t$ を増加させて)、収束させ、最後に基底状態を実現する。その後、基底状態で、スピンを計測することでこれを最適解として利用するという手順である。なお、これらの手順は、量子アニーリングの従来技術を用いて実現されるため、詳細な説明は省略する。

[0079] なお、量子アニーリングでは、図6に示すような量子揺らぎに伴う量子トンネル効果によって、スムーズにエネルギー障壁を通過することで比較的高速な最適化を実現しているとされる。図6は、トンネル効果を模式的に示す図である。

[0080] しかし、量子アニーリングを実際のマシンで実現すると、量子状態を維持することは難しく、この効果は限定的であるともいわれる。

[0081] そこで、組合せ最適化をイジングモデルに埋め込むという考え方は流用しつつ、エネルギー最小化を熱揺らぎによって行うシミュレーテッドアニーリングをトランジスタ回路等の非量子デバイスを使って高速化した専用ハードウェア(組合せ最適化アクセラレータ)も提案されている。

[0082] 上記のように、本開示では、量子コンピュータや組合せ最適化アクセラレータ等、イジングモデルをシミュレートする様々なハードウェアを総称して最適化マシンと呼ぶこととする。

[0083] 量子アニーリングマシン等の量子コンピュータを含めて、これらの最適化マシンを用いて組合せ最適化を実現する上で共通となるのは、元の組合せ最適化問題を、イジングモデルと等価にできる二次形式のバイナリ最適化問題に書き下すことである。

[0084] [3-2. データセットからのサンプル抽出最適化]

ここから、データセットからのサンプル抽出最適化について説明するが、サンプル抽出最適化の説明に先立って最適化の基準となる情報量について説明する。ここでは、カルバックライブラー情報量とフィッシャー情報行列に

ついて簡単に説明する。

[0085] カルバックライブラー情報量は、二つの確率分布に対して以下の式（7）で定義される量である。

[0086] [数7]

$$I(p|q) = \int p(x) \log p(x) dx - \int p(x) \log q(x) dx \quad \dots (7)$$

[0087] カルバックライブラー情報量は、二つの確率分布の不確かさ（エントロピー）の期待値であり、確率分布による予測をより確からしくするための情報の大きさと考えることができる。

[0088] 次に、パラメータの異なる予測分布 $p(x|w)$ 、 $p(x|w')$ の間のカルバックライブラー情報量を考える。これは、カルバックライブラー情報量の定義から以下の式（8）のように書ける。

[0089] [数8]

$$I(w, w') = \int p(x|w) \log p(x|w) dx - \int p(x|w) \log p(x|w') dx \quad \dots (8)$$

[0090] 式（8）に示す情報量は、次の式（9）のようにテイラー展開できる。

[0091] [数9]

$$I(w, w') = \frac{1}{2} \text{trace}[(w' - w)(w' - w)^T J(w)] + O(|w - w'|^3) \quad \dots (9)$$

[0092] ここで、 $J(w)$ は、次の式（10）で表されるフィッシャー情報行列（期待フィッシャー情報行列）である。

[0093] [数10]

$$J(w) = - \int p(x|w) \frac{\partial^2}{\partial w^T \partial w} \log p(x|w) dx \quad \dots (10)$$

[0094] 例えば、フィッシャー情報行列は、パラメータに対する情報量の変動感度のような概念である。また、フィッシャー情報行列は、次の式（11）及び式（12）のような関係が知られている。

[0095]

[数11]

$$J(w) = \int p(x|w) \frac{\partial}{\partial w^T} \log p(x|w) \frac{\partial}{\partial w^T} \log p(x|w) dx \quad \dots (11)$$

[0096] [数12]

$$J(w) = \frac{1}{n} (E_X[(w_X - E_X[w_X])(w_X - E_X[w_X])^T])^{-1} \quad \dots (12)$$

[0097] ここで、 $X$ は、 $n$ サンプルからなるデータセット、 $w_X$ は $X$ から推定した最尤パラメータである。例えば式(12)に示す二番目の関係は、フィッシャー情報行列がパラメータの共分散行列(精度行列)の逆行列と比例関係にあることを示している。

[0098] このことから、共分散行列を用いて算出される2つのパラメータの間のマハラノビス距離を小さくすることは、カルバックライブラー情報量による最適化と同等の意味合いであることがわかる。

[0099] 情報処理システム1における目的は、データセット $D$ で学習されるパラメータ $w_D$ を、最尤パラメータ $w_S$ で最もよく近似するサブセット $S$ を提供することといえる。ここからは、モデルは判別モデルとしての回帰モデルを扱うものとして、データの予測分布が次の式(13)のように分解できる場合を考える。

[0100] [数13]

$$p(x, y|w) = p(x)p(y|x, w) \quad \dots (13)$$

[0101] 一方、サブセットの予測分布は、入力分布だけが変更され、以下の式(14)のようになる。

[0102] [数14]

$$p(x, y|w') = q(x)p(y|x, w') \quad \dots (14)$$

[0103] このとき、 $D$ 、 $S$ の $w$ に対するフィッシャー情報行列は、それぞれ、次の式(15)及び式(16)で表される。

[0104] [数15]

$$J_p(w) = - \int p(x) \int p(y|x, w) \frac{\partial^2}{\partial w^T \partial w} \log p(y|x, w) dy dx \quad \dots (15)$$

[0105] [数16]

$$J_q(w) = - \int q(x) \int p(y|x, w) \frac{\partial^2}{\partial w^T \partial w} \log p(y|x, w) dy dx \quad \dots (16)$$

[0106] モデルは回帰モデルを扱う場合には、例えば式（9）に示す前述の情報ゲインが次の式（17）のように変更される。

[0107] [数17]

$$I(w_D, w_S) = -\frac{1}{2} \text{trace}[(w_S - w_D)(w_S - w_D)^T J_p(w_D)] + O(|w_S - w_D|^3) \quad \dots (17)$$

[0108] 一方、サブセットSで求めた最尤パラメータ $w_S$ の平均、分散・共分散は次の式（18）及び式（19）のように計算される。

[0109] [数18]

$$E_S[w_S] = w \quad \dots (18)$$

[0110] [数19]

$$E_S[(w_S - E_S[w_S])(w_S - E_S[w_S])^T] = \frac{1}{|S|} J_q^{-1}(E_S[w_S]) \quad \dots (19)$$

[0111] したがって、最尤パラメータの誤差による情報量ゲインの期待値は、次の式（20）で与えられる。

[0112] [数20]

$$E_S[I(w, w_S)] = \frac{|S|}{2} \text{trace}[J_q^{-1}(w) J_p(w)] \quad \dots (20)$$

[0113] ここで、サブセットの最適化を、より厳密に、サブセットのサンプル一つあたりの情報量ゲインの期待値を考えるとすると、次の式（21）に示すような目的関数の最小化の問題と捉えることができる。

[0114] [数21]

$$L(w, b) = \text{trace}[J_q^{-1}(w) J_p(w)] \quad \dots (21)$$

[0115] 一方、データセットと、サブセットは、それぞれ以下の式（22）及び式（23）で表されるものとする。

[0116] [数22]

$$D = \{(x_1, y_1) \dots (x_N, y_N)\} \quad \dots (22)$$

[0117] [数23]

$$S = \{(x_{s(1)}, y_{s(1)}) \dots (x_{s(K)}, y_{s(K)})\} \quad \dots (23)$$

[0118] ここでは、入出力 $x, y$ がセットになっているデータセットで説明しているが、出力側がないデータ、つまり、ラベルなしデータであっても良い。

[0119] バイナリ変数 $b_i$ を、 $x_i$ が $S$ に含まれていれば1、そうでなければ0となる変数とする。すると、フィッシャー情報行列 $J_p(w)$ 、 $J_q(w)$ は、以下の式(24)及び式(25)のように置き換えることができる。

[0120] [数24]

$$J_p(w) \approx -\frac{1}{N} \sum_{i \in V} H_i(w) \quad \dots (24)$$

[0121] [数25]

$$J_q(w, b) \approx -\frac{1}{M} \sum_{i \in V} b_i H_i(w) \quad \dots (25)$$

[0122] ただし、 $M$ は、以下の式(26)であり、 $V$ は、式(27)であるものとする。

[0123] [数26]

$$M = \sum_{i \in V} b_i \quad \dots (26)$$

[0124] [数27]

$$V = \{1 \dots N\} \quad \dots (27)$$

[0125] また、 $H_i(w)$ は、データがラベルありデータかラベルなしデータかのいずれであるかで異なる。例えばデータがラベルありデータの場合は、 $H_i(w)$ を以下の式(28)のように定義する。

[0126] [数28]

$$H_i(w) = \frac{\partial^2}{\partial w^T \partial w} \log p(y_i | x_i, w) \quad \dots (28)$$

[0127] また、例えばデータがラベルなしデータの場合は、以下の式（29）のよ  
うに定義する。

[0128] [数29]

$$H_i(w) = \int p(y|x_i, w) \frac{\partial^2}{\partial w^T \partial w} \log p(y|x_i, w) dy \quad \dots (29)$$

[0129] サブセットの最適化の問題は、次の式（30）に示すようなバイナリ変数  
による損失関数の最小化と置き換えられる。

[0130] [数30]

$$L(w, b) = \frac{M}{N} \text{trace} \left( \left( \sum_{x_i \in V} b_i H_i(w) \right)^{-1} \sum_{x_i \in V} H_i(w) \right) \quad \dots (30)$$

[0131] この損失関数の最小化、すなわち再サンプリングによる情報量の最大化を  
実現するバイナリ変数が見つければ、バイナリ変数が1であるようなデータだ  
けを再サンプリングすれば良い。この最適化のマイナーなバリエーションと  
して制約条件を追加する例が考えられる。

[0132] 制約条件の一例は、できるだけ効率よく再サンプリングするというもので  
ある。この場合は、以下の式（31）のように再サンプリングする個数をペ  
ナルティ項として損失関数に追加する。

[0133] [数31]

$$L_1(w, b) = \frac{M}{N} \text{trace} \left( \left( \sum_{i \in V} b_i H_i(w) \right)^{-1} \sum_{i \in V} H_i(w) \right) + \gamma_1 \sum_{i \in V} b_i \quad \dots (31)$$

[0134] ここで、係数 $\gamma_1$ は、再サンプリングの最適化と再サンプリングの効率のト  
レードオフを調整するパラメータである。係数 $\gamma_1$ の値が大きくなると、再サ  
ンプリングは効率的になるが、再サンプリングしたサブセットの最尤パラメ  
ータは元の最尤パラメータからずれた値になる。

[0135] また、他の一例として考えられるのは、再サンプリングの個数に既定を設  
けるものである。この場合は、以下の式（32）のように再サンプリングす  
る個数と既定の個数の二乗誤差をペナルティ項として損失関数に追加する。

[0136] [数32]

$$L_2(w, b) = \frac{M}{N} \text{trace} \left( \left( \sum_{i \in V} b_i H_i \right)^{-1} \sum_{i \in V} H_i \right) + \gamma_2 \left( M - \sum_{i \in V} b_i \right)^2 \quad \dots (32)$$

[0137] 係数  $\gamma_2$  は、再サンプリングの最適化と再サンプリングの個数制約のトレードオフを調整するパラメータがであるが、通常は個数制約を優先するため、この値は大きめに設定する。

[0138] ここで、例えば式 (31) または式 (32) 等の上述の目的関数は、 $w, b$  と二つの変数がある。サブセットを決める変数  $b$  についての最適化は必須である。これは量子アニーリングを始めとする最適化マシン 10 等の特別なアクセラレータで求めるものとする。

[0139] 一方、 $w$  については、最適化するかどうかは、ユースケースによって異なる。例えば、現行のモデルの中で最適なデータセットが必要なだけであれば、この値は固定しても良い。このような例として考えられるのは、例えば、遺伝子の分析において、特定の発現に対する新規のパターンを探索して、検証するのが目的の場合である。このような、場合、新規の遺伝子パターンの発現の分析の効率化が目的であるから、必ずしもモデルの更新は必要ではない。

[0140] なお、例えばモデルも更新しながら最適なデータセットを得たい場合は、情報処理システム 1 は、次の手続きを交互に繰り返す。 $w$  はサブセット  $S$  から最尤推定したパラメータ  $w_S$  を用い、 $b$  は  $L(w_S, b)$  を最小化するパラメータを求める。以上を繰り返すと、情報処理システム 1 は、サブセット  $S$  を得られるのみならず、データセット全体で学習する場合よりも簡単に、パラメータ  $w$  を得ることができる。

[0141] [3-3. リサンプリング最適化の量子アニーリング等への実装]

次に、リサンプリング最適化の量子アニーリング等への実装について説明する。情報処理システム 1 においては、最適なサブセットのリサンプリングのために最小化したい目的関数は次の式 (33) のように与えられる。

[0142]

[数33]

$$L(w, b) = \frac{1}{N} \sum_{k \in V} b_k \operatorname{trace} \left( \left( \sum_{i \in V} b_i H_i(w) \right)^{-1} \sum_{i \in V} H_i(w) \right) + \gamma C(b) \quad \dots (33)$$

[0143] ただし、第二項 $C(b)$ は、制約を実現するためのペナルティ項で、例えば、以下の式(34)及び式(35)等である。

[0144] [数34]

$$C_1(b) = \sum_{x_i \in V} b_i \quad \dots (34)$$

[0145] [数35]

$$C_2(b) = \left( M - \sum_{x_i \in V} b_i \right)^2 \quad \dots (35)$$

[0146] ここで、量子アニーリングマシンといったイジングマシン等の最適化マシン10で実装するには、この目的関数を $b$ の二次形式、すなわち、QUBOの形式にすることが必要である。しかし、第一項は、二次形式にはなっていない。そこで、パラメータ次元の正方行列で表わされる補助変数 $\varepsilon$ を用意して、次の式(36)のような目的関数を考える。

[0147] [数36]

$$L(w, b, \varepsilon) = \frac{1}{N} \sum_{k \in V} b_k \operatorname{trace}(\varepsilon) + \gamma C(b) + \frac{\lambda}{2} \left( \varepsilon \sum_{x_i \in V} b_i H_i(w) - \sum_{x_i \in V} H_i(w) \right)^2 \quad \dots (36)$$

[0148] ここで、 $\lambda$ の値が十分大きければ、以下の式(37)に示すような制約が成り立つことになり、元の目的関数の最適化に帰着される。

[0149] [数37]

$$\varepsilon = \left( \sum_{i \in V} b_i H_i(w) \right)^{-1} \sum_{i \in V} H_i(w) \quad \dots (37)$$

[0150] この制約と、サンプリングの最適化を同時に実現するため、 $b$ 、 $\varepsilon$ を同時に(交互)最適化する。なお、前述したように、 $w$ はユースケースに応じて最適化するが、必須ではない。補助変数 $\varepsilon$ の最適化は、 $w$ 、 $b$ を固定して次の式(

38) に示すような微分係数が0になる $\Xi$ を求めれば良い。

[0151] [数38]

$$\frac{\partial L(w, b, \Xi)}{\partial \Xi} = \frac{M}{N} I + \lambda H(b, w)^T (\Xi H(b, w) - H(1, w)) \quad \dots (38)$$

[0152] ただし、次の式(39)のように置き換えをしている。

[0153] [数39]

$$H(b, w) = \sum_{x_i \in V} b_i H_i(w) \quad \dots (39)$$

[0154] この方程式の解は、 $H(b, w)$ の逆行列が存在するならば以下の式(40)で表される。

[0155] [数40]

$$\Xi^* = \left( \frac{M}{N} + \lambda \right) H(1, w) H^{-1}(b, w) - \lambda H^{-2}(b, w) \quad \dots (40)$$

[0156] 上述した点について、以下にアルゴリズムを記載する。ここで示すアルゴリズムは、情報処理システム1により実行され、例えば以下のような処理(1-1)~(1-4)を行うアルゴリズムとなっている。

[0157] (1-1) : パラメータ $w$ と変数 $\Xi$ を一つ決める。

(1-2) : 量子アニーリングマシン(QPU: Quantum Processing Unit)等を用いて、 $w$ と $\Xi$ を固定して、 $b$ を $b^*$ で更新する。

(1-3) : 通常の演算装置(CPUやGPU)を用いて、  
 (ア)  $b$ からサブセット $S$ を求め、  
 (イ) サブセット $S$ から最尤パラメータ $w_S$ を求め、 $w=w_S$ とする(省略可)。

(ウ)  $b$ と $w=w_S$ を固定して、 $\Xi$ を $\Xi^*$ で更新する。

(1-4) : (1-2)及び(1-3)を、所定の終了条件を満たすまで繰り返す。

[0158] 例えば、上述のアルゴリズムにおける初期値はランダムに決定してもよい。なお、初期値のとり方は、上記に限らず、様々なとり方が考えられる。

[0159] [4. 第3の実施例]

ここから、第3の実施例として、第2の実施例に対してさらなる近似及びモデルの限定を行い、よりシンプルに実装する例を説明する。

[0160] まず、フィッシャー情報行列 $J_p(w)$ に関しては、前述したように次の式（41）に示すような関係が成り立つ。

[0161] [数41]

$$J_p(w) = \int p(x) \int p(y|x, w) \frac{\partial l_w(x, y)}{\partial w^T} \frac{\partial l_w(x, y)}{\partial w} dy dx \quad \dots (41)$$

[0162] ただし、 $l_w(x, y)$ は、以下の式（42）である。

[0163] [数42]

$$l_w(x, y) = \log p(y|x, w) \quad \dots (42)$$

[0164] このとき、最適化の目的関数は次の式（43）のように変形できる。

[0165] [数43]

$$\begin{aligned} L(w, b) &= - \int p(x) \int p(y|x, w) \frac{\partial l_w(x, y)}{\partial w} J_q^{-1}(w, b) \frac{\partial l_w(x, y)}{\partial w^T} dy dx \\ &\leq - \int p(x) \int p(y|x, w) \left( \frac{\partial l_w(x, y)}{\partial w} \frac{\partial l_w(x, y)}{\partial w^T} \right)^2 \left( \frac{\partial l_w(x, y)}{\partial w} J_q(w, b) \frac{\partial l_w(x, y)}{\partial w^T} \right)^{-1} dy dx \end{aligned} \quad \dots (43)$$

[0166] 不等式は、 $J_q(w, b)$ が正定値行列（固有値が全て正の値を取る行列）であることと、相加平均（固有値の平均）と調和平均（固有値の逆数（逆行列の固有値）の平均の逆数）の不等式から求まる。式（43）における不等号中の等号が成立するのは、 $J_q(w, b)$ の固有値がすべて1の場合である。また、式（43）における不等式の右辺は、目的関数の上界なので、この上界を改めて目的関数とみなすと、以下の式（44）に示す $L_{UB}(w, b)$ が、目的関数となる。

[0167] [数44]

$$L_{UB}(w, b) = - \int p(x) \int p(y|x, w) \left( \frac{\partial l_w(x, y)}{\partial w} \frac{\partial l_w(x, y)}{\partial w^T} \right)^2 \left( \frac{\partial l_w(x, y)}{\partial w} J_q(w, b) \frac{\partial l_w(x, y)}{\partial w^T} \right)^{-1} dy dx \quad \dots (44)$$

[0168] 式（43）における不等式により目的関数の最小値は、この上界の最小値よりは低いことが保証されている。そこで、以下では、この上界 $L_{UB}(w, b)$ に、いくつかの具体的なモデルを当てはめて、それぞれの実装方法を書き下す。

具体的には、回帰モデル、二クラス分類モデル、及び多クラス（Kクラス）分類モデルでの実装方法を示す。

[0169] [4-1. 回帰モデル]

まず、回帰モデルの場合について説明する。回帰モデルでは、モデルとして、分散を1に規格化した次の式（45）に示すような正規分布モデルを考える。

[0170] [数45]

$$p(y|x, w) = N(y|Wx, 1) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \exp\left(-\frac{1}{2}(y - Wx)^T(y - Wx)\right) \quad \dots (45)$$

[0171] なお、この例では、線形モデルになっているが、 $x$ を例えば、深層学習ニューラルネットの最終レイヤへの入力特徴量として、非線形モデルを扱うこともできる。このモデルでは、以下の式（46）及び式（47）に示すような関係が得られる。

[0172] [数46]

$$\frac{\partial l_w(x, y)}{\partial W} = (y - Wx)x^T \quad \dots (46)$$

[0173] [数47]

$$J_q(w, b) = -\frac{1}{M} \sum_{j \in V} b_j x x^T \quad \dots (47)$$

[0174] 式（46）及び式（47）を用いると目的関数は次の式（48）のようになる。

[0175] [数48]

$$L_{UB}(w, b) \approx \frac{1}{N} \sum_{m \in V} b_m \sum_{i \in V} \left( \sum_{j \in V} a_{ij} b_j \right)^{-1} \quad \dots (48)$$

[0176] ただし、パラメータ $a_{ij}$ は、以下の式（49）のように定義している。

[0177]

[数49]

$$a_{ij} = \left( \frac{x_i^T x_j}{x_i^T x_i} \right)^2 \quad \dots (49)$$

[0178] 量子アニーリング等の最適化マシン10で最適化を実行するには、この目的関数がbの二次形式になっていればよい。この問題は次のようなペナルティ項を追加することで、以下の式(50)に示すような二次形式に変形できる。ペナルティ項が0になるとき、式(48)と一致することが容易に確認できる。

[0179] [数50]

$$L_{UB1}(w, b, \xi) = \frac{1}{N} \sum_{m \in V} b_m \sum_{i \in V} \xi_i + \lambda \sum_i \left( \xi_i \sum_j a_{ij} b_j - 1 \right)^2 \quad \dots (50)$$

[0180] もし、サブセットのサイズをMと指定したい場合には、次の式(51)のようにしてもよい。

[0181] [数51]

$$L_{UB1}(w, b, \xi) = \frac{M}{N} \sum_{i=1}^N \xi_i + \lambda_K \left( M - \sum_{k \in V} b_k \right)^2 + \lambda \sum_i \left( \sum_j \xi_i a_{ij} b_j - 1 \right)^2 \quad \dots (51)$$

[0182] この目的関数には、最適化したい変数がbのほかに連続変数ξもある。ただし、ξについては、二次式であるから通常の計算機を使った計算で容易に最適値が求まる。例えば、情報処理装置100は、以下の式(52)及び式(53)を用いてξの最適値を求める。

[0183] [数52]

$$\frac{\partial}{\partial \xi} L_{UB1}(w, b, \xi) = \frac{M}{N} + 2\lambda(Ab \cdot (\xi \cdot Ab) - Ab) = 0 \quad \dots (52)$$

[0184] [数53]

$$\xi_i = \frac{1}{(\sum_j a_{ij} b_j)^2} \left( \sum_j a_{ij} b_j - \frac{M}{2\lambda N} \right) \quad \dots (53)$$

[0185] 例えば、情報処理システム 1 は、 $\xi$  を固定して  $b$  を量子アニーリング等の最適化マシン 10 により最適化したら、今度は  $b$  を固定して  $\xi$  を通常の計算アルゴリズムで情報処理装置 100 により更新し、これらの最適化を交互に繰り返す。

[0186] 上述した点について、以下にアルゴリズムを記載する。ここで示すアルゴリズムは、情報処理システム 1 により実行され、例えば以下のような処理 (2-1) ~ (2-4) を行うアルゴリズムとなっている。

(2-1) : パラメータ  $w$  と変数  $\xi$  を一つ決める。

(2-2) : 量子アニーリングマシン (QPU) 等を用いて、 $w$  と  $\xi$  を固定して、 $b$  を  $b^*$  で更新する。

(2-3) : 通常の演算装置 (CPU や GPU) を用いて、

(ア)  $b$  からサブセット  $S$  を求め、

(イ) サブセット  $S$  から最尤パラメータ  $w_s$  を求め、 $w = w_s$  とする (ユースケースにより省略可)。

(ウ)  $b$  と  $w = w_s$  を固定して最適パラメータ  $\xi^*$  を求め、 $\xi$  を  $\xi^*$  で更新する。

(2-4) : (2-2)、(2-3) を、所定の終了条件を満たすまで繰り返す。

[0187] 例えば、上述のアルゴリズムにおける初期値はランダムに決定してもよい。なお、初期値のとり方は、上記に限らず、様々なとり方が考えられる。

[0188] [4-2. ニクラス分類モデル]

次に、ニクラス分類モデルの場合について説明する。例えば、以下の式 (54) を用いて、モデルとして、ニクラス分類に用いられるロジスティック回帰を考える。

[0189] [数54]

$$p(y|x, w) = \sigma(yw^T x) = \frac{1}{1 + \exp(-yw^T x)} \quad \dots (54)$$

[0190] ラベル無しデータセットの場合は次の式 (55) 及び式 (56) が得られ

る。

[0191] [数55]

$$\frac{\partial l_w(x, y)}{\partial w} = -yx^T \sigma(-yw^T x) \quad \dots (55)$$

[0192] [数56]

$$J_q(w, b) = -\frac{1}{M} \sum_{i \in V} b_i x_i x_i^T \sigma(-w^T x_i) \sigma(w^T x_i) \quad \dots (56)$$

[0193] 以上から次の式 (57) を得る。

[0194] [数57]

$$L_{UB}(w, b) = \frac{M}{N} \sum_{i \in V} \sigma(-w^T x_i) \sigma(w^T x_i) (x_i^T x_i)^2 \left( \sum_{j \in V} b_j \sigma(-w^T x_j) \sigma(w^T x_j) (x_j^T x_j)^2 \right)^{-1} \quad \dots (57)$$

[0195] ここで、以下の式 (58) のように変数や係数を置き換える。

[0196] [数58]

$$a_{ij} = \frac{\sigma(-w^T x_j) (1 - \sigma(-w^T x_j))}{\sigma(-w^T x_i) (1 - \sigma(-w^T x_i))} \left( \frac{x_i^T x_j}{x_i^T x_i} \right)^2 \quad \dots (58)$$

[0197] すると、目的関数はバイナリ変数セット  $b$  を用いて以下の式 (59) のように書ける。

[0198] [数59]

$$L_{UB}(w, b) = \frac{M}{N} \sum_{i \in V} \left( \sum_{j \in V} a_{ij} b_j \right)^{-1} \quad \dots (59)$$

[0199] 式 (59) に示す目的関数は次の式 (60) に示すようにペナルティ項を追加することで、 $b$  の二次形式に変形できる。

[0200] [数60]

$$L_{UB1}(w, b, \xi) = \frac{1}{N} \sum_{i=1}^N \sum_j \xi_i b_j + \lambda \sum_i \left( \sum_j \xi_i a_{ij} b_j - 1 \right)^2 \quad \dots (60)$$

[0201] サブセットのサイズをMに指定するには、次の式（61）のようにしてもよい。

[0202] [数61]

$$L_{UB1}(w, b, \xi) = \frac{M}{N} \sum_{i=1}^N \xi_i + \lambda_M \left( M - \sum_{k \in V} b_k \right)^2 + \lambda \sum_i \left( \sum_j \xi_i a_{ij} b_j - 1 \right)^2 \quad \dots (61)$$

[0203] いずれの場合であっても、 $\xi_i$ が、次の式（62）の条件を満たしていれば、元の問題と等価になる。

[0204] [数62]

$$\xi_i = \left( \sum_{j=1}^N a_{ij} b_j \right)^{-1} \quad \dots (62)$$

[0205] 回帰モデルと同様、 $\xi$ は量子アニーリング等の最適化マシン10では最適化されないため、通常の計算アルゴリズムを用いて情報処理装置100により最適化する。なお、この二クラス分類の定式化では、係数 $a_{ij}$ の算出方法は異なるが、これらの係数を使った目的関数は、回帰モデルと全く同じなので、アルゴリズムの詳細な説明は省略する。

[0206] [4-3. 多クラス分類モデル]

次に、多クラス分類モデルの場合について説明する。例えば、以下の式（63）を用いて、モデルとして、Kクラス分類に用いられるソフトマックス関数を考える。

[0207] [数63]

$$p(y|x, w) = s_w(x, y) = \frac{\exp(w_y^T x)}{\sum_{k \in \{1 \dots K\}} \exp(w_k^T x)} \quad \dots (63)$$

[0208] ここで、 $s(x)$ を以下の式（64）のK次元ベクトル、 $u$ をすべての成分が1のK次元ベクトル、 $e_k$ をk番目の成分のみが1で、他が0のK次元ベクトルとする。

[0209]

[数64]

$$s_i(x) = s_w(x, i) \quad \dots (64)$$

[0210] すると、以下の式 (65) 及び式 (66) に示す関係が得られる。

[0211] [数65]

$$\frac{\partial l_w(x, y)}{\partial w} = (e_y - s(x)) \odot x^T \quad \dots (65)$$

[0212] [数66]

$$J_q(w, b) = -\frac{1}{MK} \sum_{j \in V} b_j \sum_k (e_k - s(x_j)) (e_k - s(x_j))^T \odot x_j x_j^T \quad \dots (66)$$

[0213] [数67]

$$\odot \quad \dots (67)$$

[0214] ここでは、式 (65) 及び式 (66) 中の上記の数 (67) に示す記号は直積を表している。式 (65) 及び式 (66) を用いると目的関数は次の式 (68) のようになる。

[0215] [数68]

$$L_{UB}(w, b) \approx \frac{1}{N} \sum_{m \in V} b_m \sum_{k \in (1 \dots K)} \sum_{i \in V} \left( \sum_{j \in V} a_{kij} b_j \right)^{-1} \quad \dots (68)$$

[0216] ただし、以下の式 (69) のように変数や係数を置き換えた。

[0217] [数69]

$$a_{kij} = \frac{(e_k - s(x_i))^T (I_K - s(x_j)u^T - us^T(x_j) + Ks(x_j)s^T(x_j)) (e_k - s(x_i)) \left( \frac{x_i^T x_j}{x_i^T x_i} \right)^2}{\left( (e_k - s(x_i))^T (e_k - s(x_i)) \right)^2} \quad \dots (69)$$

[0218] この問題は次の式 (70) のようなペナルティ項を追加することで、二次形式に変形できる。

[0219]

[数70]

$$L_{UB1}(w, b, \xi) = \frac{1}{N} \sum_{m \in V} b_m \sum_{k \in (1 \dots K)} \sum_{i \in V} \xi_{ki} + \lambda \sum_{k \in (1 \dots K)} \sum_{i \in V} \left( \xi_{ki} \sum_{j \in V} a_{kij} b_j - 1 \right)^2 \quad \dots (70)$$

[0220] Mを指定したサイズにするには、次の式(71)のようにしてもよい。

[0221] [数71]

$$L_{UB1}(w, b, \xi) = \frac{M}{N} \sum_{k \in (1 \dots K)} \sum_{i \in V} \xi_{ki} + \lambda_K \left( M - \sum_{k \in V} b_k \right)^2 + \lambda \sum_{k \in (1 \dots K)} \sum_{i \in V} \left( \xi_{ki} \sum_{j \in V} a_{kij} b_j - 1 \right)^2 \quad \dots (71)$$

[0222] この目的関数には、最適化したい変数がbのほかに連続変数ξもある。ξは量子アニーリング等の最適化マシン10では最適化されないため、通常の計算アルゴリズムを用いて情報処理装置100により最適化する。導出については、回帰モデルやニクラス分類の方法とほぼ同じなので、詳細な説明は省略する。

[0223] [5. 第4の実施例]

ここから、第4の実施例として、第2の実施例及び第3の実施例の計算過程を踏まえて、データ選択最適化の処理フローとそのプログラム例を説明する。

[0224] まず、情報処理装置100によるデータ選択最適化の処理フローは以下のような処理(3-1)～(3-4)となっている。

(3-1) : ユーザから学習したい予測モデルのファイルを受け取る。

(3-2) : システムがもつ大規模データセットをサーバ上で扱える状態にする。

(3-3) : もしユーザからのデータがあればそれもマージして使えるようにする。

(3-4) : マージされたデータセットから最適なサブセットをリサンプリングする。

[0225] [5-1. 情報処理のメインプログラム例]

上記処理（3-1）～（3-4）は、プログラムではメインプログラムに相当する。メインプログラムでは、まず、ユーザが学習したい機能を実現する関数モジュールを展開する。続いて、システムが保持するデータセットをユーザが指定する条件の下でメモリ上に展開する。また、ユーザからのデータセットがあれば、システムのデータセットとマージする。最後に、モデルとデータセットからデータセットの最適リサンプリングを行い、サブセットを生成する。

[0226] 例えば、メインプログラムは、図7に示すプログラムPG1となる。図7は、情報処理のメインプログラム一例を示す図である。例えば、プログラムPG1の1行目の関数「`model = load_model()`」は、学習したい機能の読み取りを行う関数である。例えば、プログラムPG1の2行目の関数「`dataset = load_dataset()`」は、データセットの読み取りを行う関数である。

[0227] 例えば、プログラムPG1の3行目の関数「`userdata = load_userdata()`」は、ユーザデータの読み取りを行う関数である。例えば、プログラムPG1の4行目の処理「`dataset = dataset + userdata`」は、データセットとユーザデータのマージを行う処理である。例えば、プログラムPG1の5行目の関数「`subset = resample_optimal_subset(model, dataset)`」は、サブセットのリサンプリングを行う関数である。

[0228] 続いて、リサンプリング処理を行う関数における処理の一例の説明をする。

（4-1）：ランダムにバイナリ変数 $b$ を生成する。

（4-2）：バイナリ変数 $b$ を元にして、データセット $D$ からサブセット $S$ を生成する。

（4-3）：入力されたモジュール $m$ とサブセット $S$ を用いて、モジュール $m$ に関する最尤パラメータ $w_s$ の推定を行う。

（4-4）：バイナリ変数 $b$ とモジュールのパラメータ $w$ を固定して、問題をQUBO形式にするための補助変数（まとめて $X$ と記載する）を計算する。

（4-5）：パラメータ $w$ と補助変数 $X$ を用いて、QUBOの（二次の）係数

行列 $q$ を算出する。

(4-6) : QUBOの係数行列 $q$ を量子アニーリングマシン等の最適化マシン10に送信し、最適化されたバイナリ変数 $b$ を受け取る。

(4-7) : (4-2) から (4-6) を所定の終了条件まで繰り返した後、サブセット $S$ を出力する。

なお、(4-1)におけるバイナリ変数 $b$ の初期値の設定についてはランダムに限られない。例えば、バイナリ変数 $b$ の初期値は、データ数の指定がある場合、その指定された数を満たした数のデータを選択するように設定されてもよい。

[0229] [5-2. データ選択最適化の処理フロー例]

次に、図8を用いて、データ選択最適化の処理フローについて説明する。図8は、データ選択最適化の処理の手順を示すフローチャートである。図8は、情報処理システム1によるデータ選択最適化の処理のフローの例である。以下では、情報処理装置100を処理主体とする場合を一例として示すが、処理主体は、情報処理装置100に限らず、情報処理システム1に含まれるいずれの装置であってもよい。

[0230] 図8に示すように、情報処理装置100は、ユーザデータとデータセットを入力する(ステップS101)。例えば、情報処理装置100は、端末装置20からユーザデータを取得し、データベース50からデータセットを取得する。

[0231] 情報処理装置100は、データ組合せに対する情報量の二次形式行列を取得する(ステップS102)。例えば、情報処理装置100は、目的関数から情報量の二次形式行列を抽出する。

[0232] 情報処理装置100は、係数行列を組合せ最適化マシンに送信する(ステップS103)。例えば、情報処理装置100は、係数行列をイジングパラメータとして、ユーザにより選択された最適化マシン10に送信する。

[0233] 情報処理装置100は、組合せ最適化マシンから最適化解を受信する(ステップS104)。例えば、情報処理装置100は、イジングパラメータを

送信した最適化マシン10から、最適化マシン10が計算した解を最適化解として受信する。このように、情報処理システム1は、組合せ最適化マシンである最適化マシン10により最適化された組合せを得る。

[0234] 情報処理装置100は、所定の終了条件を満たすかどうかを判定する（ステップS105）。例えば、終了条件としては、処理回数が閾値に達すること等であってもよい。このように、図8に一例を示す繰り返しアルゴリズムでは、終了条件は、繰り返し数が一定以上に到達したこととする。なお、終了条件は、繰り返し数に関するものに限らず、例えばエネルギー関数の繰り返しステップによる更新幅が一定以下になること等の収束に関する条件であってもよい。なお、この繰り返しアルゴリズムは必ず、一ステップでエネルギーが減るとは限らないので、最小エネルギーを達成するたびに、そのときのパラメータ等を記憶する構成にしてもよい。この場合、情報処理装置100は、終了条件を満たした時に最小エネルギーのパラメータとして記憶していたパラメータを、ユーザに提供してもよい。

[0235] 情報処理装置100は、所定の終了条件を満たしていない場合（ステップS105：No）、ステップS102に戻って処理を繰り返す。このように、交互最適化であるため、情報処理装置100は帰還処理が行う。

[0236] 情報処理装置100は、所定の終了条件を満たす場合（ステップS105：Yes）、提供するデータを選択する（ステップS106）。例えば、情報処理装置100は、所定の終了条件を満たす場合、データセットからサブセットを選択する。そして、情報処理装置100は、選択したデータセットを出力する（ステップS107）。例えば、情報処理装置100は、選択したサブセットをデータセットとして、ユーザに提供する。

[0237] [5-3. データ選択最適化の処理プログラム例]

ここで、図9を用いて上述した処理に対応するプログラムの一例を説明する。図9は、データ選択最適化の処理プログラムの一例を示す図である。プログラムPG2に示す関数「resample\_optimal\_subset(model, dataset)」は、図7中の5行目に示すサブセットのリサンプリングの関数であり、subse

tを出力する関数に対応する。例えば、図9に示すプログラムPG2は、図8に示した処理をプログラム例として表したものである。図9に示すプログラムPG2は、図8に示した処理と同様であるため、詳細な説明は省略する。

[0238] 例えば、プログラムPG2の6行目の関数「`w = maximum_likelihood(model, subset)`」は、サブセットを用いて、モデルパラメータを更新する処理を行う関数である。例えば、プログラムPG2の7行目の関数「`hessians = calc_hessians(w, dataset)`」は、モデルパラメータから情報量を計算する処理である。例えば、プログラムPG2の8行目の関数「`x = calc_additional_variables(hessians, b)`」は、二次形式になっていない式を二次形式にするための補助変数を用意する処理である。例えば、プログラムPG2の9行目の関数「`q = calc_qubo_coefficients(hessians, x)`」は、二次形式からイジング係数を抽出する関数である。例えば、プログラムPG2の10行目の関数「`b = machine.get_optimal(q)`」は、QUBOの係数行列 $q$ を入力として、バイナリ変数 $b$ を出力する関数である。

[0239] 例えば、情報処理装置100は、図7、図9に示すようなプログラム（関数）及び、各プログラムにより呼び出されるプログラム（関数）を関数情報記憶部122（図10参照）に記憶し、各プログラムを用いて処理を実行する。

[0240] 上述したように、情報処理装置100は、モデルのタスクに応じて決定される情報量に関する目的関数から、係数行列を最適化マシン10へ送信するイジングパラメータとして抽出する。例えば、情報処理装置100は、上述したプログラム等を利用して係数行列を抽出する。例えば、係数行列は、目的関数から抽出される最適化したい変数の一次以上の項に係る係数で構成される配列である。そして、情報処理装置100は、抽出した係数行列を用いて、最適化マシン10のスピント結合定数や局所磁場に設定する。情報処理装置100は、抽出した係数行列を最適化マシン10に送信する。そして、最適化マシン10は、アニーリング処理を行い、イジングモデルの基底スピンの組合せを算出し、情報処理装置100に送信する。情報処理装置100

は、最適化マシン 10 が算出したイジングモデルの基底スピンの組合せを、計算結果として受信し、受領した計算結果に基づいてデータセットからサブセットを選択する。

[0241] [6. 情報処理装置の構成]

次に、情報処理装置 100 の構成について説明する。図 10 は、本開示の情報処理装置の構成例を示す図である。

[0242] 図 10 に示すように、情報処理装置 100 は、通信部 110 と、記憶部 120 と、制御部 130 とを有する。なお、情報処理装置 100 は、情報処理装置 100 の管理者等から各種操作を受け付ける入力部（例えば、キーボードやマウス等）や、各種情報を表示するための表示部（例えば、液晶ディスプレイ等）を有してもよい。

[0243] 通信部 110 は、例えば、NIC (Network Interface Card) 等によって実現される。そして、通信部 110 は、所定のネットワーク（図示省略）と有線または無線で接続され、最適化マシン 10 や端末装置 20 やデータベース 50 等との間で情報の送受信を行う。また、通信部 110 は、ユーザが利用するユーザ端末（図示省略）との間で情報の送受信を行ってもよい。

[0244] 記憶部 120 は、例えば、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。第 1 の実施例に係る記憶部 120 は、図 10 に示すように、データ記憶部 121 と、関数情報記憶部 122 とを有する。

[0245] データ記憶部 121 は、端末装置 20 やデータベース 50 から受信したデータ等の各種データを記憶する。関数情報記憶部 122 は、各種のプログラム等の関数の情報を記憶する。例えば、関数情報記憶部 122 は、本開示に係る情報処理に用いる各種の関数（プログラム）の情報を記憶する。例えば、関数情報記憶部 122 は、プログラム PG1、PG2 等の関数プログラムを記憶する。また、関数情報記憶部 122 は、上述した式のうち処理に用いる関数の情報を記憶してもよい。なお、記憶部 120 は、上記に限らず、目

的に応じて種々の情報を記憶してもよい。

[0246] 制御部130は、例えば、CPUやMPU (Micro Processing Unit) 等によって、情報処理装置100内部に記憶されたプログラム（例えば、本開示に係る情報処理プログラム等）がRAM (Random Access Memory) 等を実行領域として実行されることにより実現される。また、制御部130は、例えば、ASIC (Application Specific Integrated Circuit) やFPGA等の集積回路により実現される。

[0247] 図10に示すように、制御部130は、取得部131と、最適化係数抽出部132と、最適化マシン通信部133と、データセット選択部134と、出力部135とを有し、以下に説明する情報処理の機能や作用を実現または実行する。なお、制御部130の内部構成は、図10に示した構成に限られず、後述する情報処理を行う構成であれば他の構成であってもよい。

[0248] 取得部131は、データ読み取り部131a及びユーザデータ入力部131bの機能を有する。取得部131は、各種情報を取得する。取得部131は、記憶部120に記憶された情報を取得する。取得部131は、データセットを取得する。

[0249] 取得部131は、学習するモデルに対応するタスクを取得する。例えば、取得部131は、ユーザが利用する端末装置20からユーザが学習したいモデルを示す情報を取得する。取得部131は、ユーザが利用する端末装置20からユーザが学習したいモデルのタスクを示す情報を取得する。例えば、取得部131は、記憶部120から学習するモデルに対応するタスクを取得する。取得部131は、ユーザが学習したいモデルの情報を取得する。例えば、取得部131は、各モデルにタスクが対応付けて記憶された記憶部120から、ユーザが学習したいモデルに対応するタスクを示す情報を取得する。

[0250] 取得部131は、モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報を取得する。例えば、取得部131は、モデルの学習に用いるデータ数に関する指定情報を取得する。例えば、取得部131は、ユ

ーザが利用する端末装置 20 からモデルの学習に用いるデータの数を指定する指定情報を取得する。例えば、取得部 131 は、サブセットとして選択するデータの数を指定する指定情報を、ユーザが利用する端末装置 20 から取得する。

[0251] 例えば、取得部 131 は、ユーザがデータの数を指定しない場合、できる限り少ない数のデータを選択することを指定する指定情報を取得する。例えば、取得部 131 は、ユーザがデータの数を指定しない場合、所定の数（例えば初期設定数）を指定する指定情報を取得する。例えば、取得部 131 は、ユーザがデータの数を指定しない場合、記憶部 120 に記憶された所定の数（例えば初期設定数）を指定する指定情報を取得する。

[0252] 取得部 131 は、通信部 110 を介して各種情報を受信する。取得部 131 は、外部の情報処理装置から各種情報を受信する。取得部 131 は、端末装置 20 やデータベース 50 からデータを受信する。

[0253] 最適化係数抽出部 132 は、情報の抽出処理を行う抽出部である。最適化係数抽出部 132 は、目的関数から最適化マシン 10 の入力に対応する係数行列を抽出する。最適化係数抽出部 132 は、係数行列を抽出する。最適化係数抽出部 132 は、目的関数からイジング係数に対応する係数行列を抽出する。最適化係数抽出部 132 は、タスクに応じた情報ゲインを示す目的関数から係数行列を抽出する。最適化係数抽出部 132 は、カルバックライブラー情報量に関する目的関数から係数行列を抽出する。最適化係数抽出部 132 は、イジングモデルのエネルギー関数として表現される目的関数から係数行列を抽出する。最適化係数抽出部 132 は、QUBO 形式で表現される目的関数から係数行列を抽出する。例えば、最適化係数抽出部 132 は、ユーザが学習したいモデルが回帰モデルであり、データの数の指定がない場合、式 (50) に示すような目的関数から係数行列を抽出する。また、最適化係数抽出部 132 は、ユーザが学習したいモデルが回帰モデルであり、データの数の指定がある場合、式 (51) に示すような目的関数から係数行列を抽出する。

[0254] 例えば、最適化係数抽出部 132 は、ユーザが学習したいモデルが二クラス分類モデルであり、データの数の指定がない場合、式 (60) に示すような目的関数から係数行列を抽出する。また、最適化係数抽出部 132 は、ユーザが学習したいモデルが二クラス分類モデルであり、データの数の指定がある場合、式 (61) に示すような目的関数から係数行列を抽出する。例えば、最適化係数抽出部 132 は、ユーザが学習したいモデルが多クラス分類モデルであり、データの数の指定がない場合、式 (70) に示すような目的関数から係数行列を抽出する。また、最適化係数抽出部 132 は、ユーザが学習したいモデルが多クラス分類モデルであり、データの数の指定がある場合、式 (71) に示すような目的関数から係数行列を抽出する。なお、最適化係数抽出部 132 は、ユーザが学習したいモデルに対応するパラメータを抽出してもよい。例えば、最適化係数抽出部 132 は、ユーザが学習したいモデルから目的関数に設定するパラメータ  $w$  を抽出する。例えば、最適化係数抽出部 132 は、記憶部 120 から学習するモデルに対応するパラメータを取得する。例えば、最適化係数抽出部 132 は、各モデルにパラメータが対応付けて記憶された記憶部 120 から、ユーザが学習したいモデルに対応するパラメータを示す情報を取得する。

[0255] 最適化係数抽出部 132 は、最適化マシン通信部 133 が最適化マシン 10 から受信した計算結果を基にパラメータの最適化を行う。例えば、最適化係数抽出部 132 は、最適化マシン 10 により最適化されるパラメータ以外のパラメータの最適化を行う。例えば、最適化係数抽出部 132 は、最適化マシン 10 の計算結果を基にバイナリ変数  $b_i$  がを固定し、 $\xi$  を通常の計算アルゴリズムで更新する。例えば、最適化係数抽出部 132 は、ユーザが学習したいモデルが回帰モデルである場合、式 (52) 及び式 (53) を用いて  $\xi$  の最適値を算出する。例えば、最適化係数抽出部 132 は、ユーザが学習したいモデルが二クラス分類モデルである場合、式 (62) 等を用いて  $\xi$  の最適値を算出する。例えば、最適化係数抽出部 132 は、ユーザが学習したいモデルが多クラス分類モデルである場合も同様に  $\xi$  の最適値を算出する。

- [0256] 最適化マシン通信部 133 は、通信部 110 を介して最適化マシン 10 と通信し、最適化マシン 10 との間で情報を送受信する。最適化マシン通信部 133 は、最適化マシン 10 に計算の実行を指示する。最適化マシン通信部 133 は、最適化マシン 10 にイジングモデルのパラメータを送信する。最適化マシン通信部 133 は、最適化マシン 10 にイジングモデルのパラメータを送信することにより、最適化マシン 10 に計算の実行を指示する。最適化マシン通信部 133 は、組合せ最適化計算を行う最適化マシン 10 に、最適化係数抽出部 132 により抽出された係数行列を送信する。
- [0257] 最適化マシン通信部 133 は、最適化マシン 10 から各種情報を受信する。最適化マシン通信部 133 は、最適化マシン 10 の計算結果を最適化マシン 10 から受信する。
- [0258] 最適化マシン通信部 133 は、組合せ最適化計算を行う最適化マシン 10 に目的関数に対応する係数行列を送信する。最適化マシン通信部 133 は、最適化マシン 10 から組合せ最適化計算の計算結果を受信する。最適化マシン通信部 133 は、組合せ最適化計算後の変数を示す計算結果を最適化マシン 10 から受信する。最適化マシン通信部 133 は、各々がデータに対応するバイナリ変数に関する計算結果を最適化マシン 10 から受信する。
- [0259] 最適化マシン通信部 133 は、量子コンピュータ、または組合せ最適化アクセラレータである最適化マシン 10 に係数行列を送信する。最適化マシン通信部 133 は、複数の最適化マシン 10 のうち、ユーザにより選択された最適化マシン 10 に係数行列を送信する。
- [0260] データセット選択部 134 は、データセット等のサンプルセットの選択処理を行う選択部である。データセット選択部 134 は、タスクに応じて決定される情報量に関する目的関数と、指定情報を用いて決定される数とに基づいて、データセットから、モデルの学習に用いるサブセットを選択する。データセット選択部 134 は、モデルにもたらされる情報エントロピーを示す目的関数を最適化するように、サンプルセットを選択する。データセット選択部 134 は、タスクに応じた情報ゲインを示す目的関数に基づいて、サブ

セットを選択する。

[0261] データセット選択部134は、カルバックライブラー情報量に関する目的関数に基づいて、サブセットを選択する。データセット選択部134は、イジングモデルのエネルギー関数として表現される目的関数に基づいて、サブセットを選択する。データセット選択部134は、QUBO形式で表現される目的関数に基づいて、サブセットを選択する。データセット選択部134は、最適化マシン10から取得された最適化マシン10の計算結果に基づいて、サブセットを選択する。

[0262] 例えば、データセット選択部134は、最適化マシン10の計算結果に基づいて、そのデータに対応するバイナリ変数 $b_i$ が1であるデータをサブセットに追加するデータとして選択する。また、データセット選択部134は、最適化マシン10の計算結果に基づいて、そのデータに対応するバイナリ変数 $b_i$ が0であるデータをサブセットに追加しない。このように、データセット選択部134は、最適化マシン10の計算結果に基づくバイナリ変数 $b_i$ が値に応じて、バイナリ変数 $b_i$ が1であるデータからなるサブセットを生成する。

[0263] 出力部135は、情報の出力処理を行う。出力部135は、データセット選択部134により選択されたサンプルに関する情報を出力する。例えば、出力部135は、サブセットに関する情報を出力する。出力部135は、通信部110を介して各種情報を送信する。出力部135は、サンプルセットを送信する。出力部135は、サブセットをサンプルセットとしてユーザが利用する端末装置20へ送信する。

[0264] また、出力部135は、サブセット等のサンプルセットを用いて学習された学習済みモデルをユーザが利用する端末装置20へ送信してもよい。この場合、情報処理装置100は、モデルを学習する学習部を有してもよい。情報処理装置100の学習部は、サブセットを用いて学習済みモデルを学習する。

[0265] また、情報処理装置100は、サブセットを用いて学習された学習済みモデルを、モデルを学習する学習処理を実行する外部装置（「学習装置」とす

る)から取得してもよい。この場合、情報処理装置100は、サブセットを学習装置へ送信し、学習装置が学習処理により学習した学習済みモデルを学習装置から受信してもよい。そして、情報処理装置100は、学習装置から受信した学習済みモデルを端末装置20へ送信する。

[0266] 情報処理装置100は、コンテンツCT1等の各種コンテンツを生成するコンテンツ生成部を有してもよい。コンテンツ生成部は、端末装置20に提供するコンテンツを生成する。コンテンツ生成部は、Java(登録商標)等の種々の技術を適宜用いて、コンテンツを生成する。なお、コンテンツ生成部は、CSSやJavaScript(登録商標)やHTMLの形式に基づいて、コンテンツCT1等を生成してもよい。また、例えば、コンテンツ生成部は、JPEG(Joint Photographic Experts Group)やGIF(Graphics Interchange Format)やPNG(Portable Network Graphics)等様々な形式でコンテンツCT1等を生成してもよい。

[0267] [7. 最適化マシンの構成]

次に、計算を実行する最適化マシン10の構成について説明する。図11は、最適化マシンの構成例を示す図である。図11の例では、最適化マシン10の一例として、量子コンピュータである最適化マシン10aの構成を説明する。

[0268] 図11に示すように、最適化マシン10aは、通信部11と、記憶部12と、量子デバイス部13と、制御部14とを有する。なお、最適化マシン10aは、最適化マシン10aの管理者等から各種操作を受け付ける入力部(例えば、キーボードやマウス等)や、各種情報を表示するための表示部(例えば、液晶ディスプレイ等)を有してもよい。

[0269] 通信部11は、例えば、NICや通信回路等によって実現される。通信部11は、所定のネットワーク(インターネット等)と有線又は無線で接続され、ネットワークを介して、情報処理装置100等の他の装置等との間で情報の送受信を行う。

[0270] 記憶部12は、例えば、RAM、フラッシュメモリ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。記憶部 1 2 は、情報の表示に用いる各種情報を記憶する。

[0271] 量子デバイス部 1 3 は、種々の量子計算を実行する。例えば、量子デバイス部 1 3 は、量子処理ユニット (QPU: Quantum Processing Unit) により実現される。量子デバイス部 1 3 は、例えば、情報処理装置 1 0 0 等の他の装置から受け付けたイジングモデルのパラメータに基づいて、そのイジングモデルの基底状態を実現する。言い換えれば、量子デバイス部 1 3 は、イジングモデルが基底エネルギー状態となる最適スピン配置を実現する。すなわち、量子デバイス部 1 3 は、最適化問題を最適化した状態を実現する。

[0272] 量子デバイス部 1 3 は、例えば、複数の量子ビットから構成される。量子デバイス部 1 3 は、予め絶対零度付近まで冷却される。量子デバイス部 1 3 は、量子デバイス部 1 3 にイジングモデルのパラメータが入力された後、内部でイジングモデルと横磁場モデル (量子ゆらぎモデル) の比率を時間発展させる。これにより、量子デバイス部 1 3 上で、イジングモデルのパラメータに応じた最適なスピン配置が実現する。このように、量子デバイス部 1 3 上でイジングモデルの最適なスピン配置が物理的に実現される。そして、量子デバイス部 1 3 が測定されることにより、イジングモデルの最適なスピン配置を得ることができる。これにより、量子デバイス部 1 3 は、離散最適化問題を最適化することができる。例えば、量子デバイス部 1 3 は、二値二次形式の目的関数の最適化問題を最適化することができる。

[0273] 制御部 1 4 は、例えば、CPU や MPU 等によって、最適化マシン 1 0 a 内部に記憶されたプログラムが RAM 等を作業領域として実行されることにより実現される。また、制御部 1 4 は、コントローラであり、例えば、ASIC や FPGA 等の集積回路により実現されてもよい。

[0274] 図 1 1 に示すように、制御部 1 4 は、取得部 1 4 1 と、計算部 1 4 2 と、送信部 1 4 3 とを有し、以下に説明する情報処理の機能や作用を実現または実行する。なお、制御部 1 4 の内部構成は、図 1 1 に示した構成に限られず、後述する情報処理を行う構成であれば他の構成であってもよい。

- [0275] 取得部141は、各種情報を受信する。取得部141は、外部の情報処理装置から各種情報を受信する。取得部141は、情報処理装置100等の他の情報処理装置から各種情報を受信する。
- [0276] 取得部141は、例えば量子デバイス部13を用いた計算を行い、測定するための指示を情報処理装置100等の他の情報処理装置から受け付ける。取得部141は、イジングモデルのパラメータを量子デバイス部13による計算（測定）の指示として受け付ける。
- [0277] 取得部141は、各種の情報を取得する。取得部141は、記憶部12から情報を取得する。取得部141は、情報処理装置100等の外部の情報処理装置から各種の情報を取得する。取得部141は、入力部により受け付けられた入力情報を取得する。取得部141は、例えば、外部の情報処理装置からイジングモデルのパラメータに関する情報を取得する。取得部141は、計算部142による量子デバイス部13の測定結果（計算結果）を取得する。
- [0278] 例えば、取得部141は、プログラムPG2の関数「 $q = \text{calc\_qubo\_coefficients}(\text{hessians}, x)$ 」の処理結果であるQUBOの係数行列 $q$ を情報処理装置100から受信する。
- [0279] 計算部142は、各種計算を実行する。計算部142は、量子デバイス部13を用いた計算を実行する。計算部142は、量子デバイス部13を測定する。計算部142は、イジングモデルの最適スピン配置が実現された量子デバイス部13を測定する。
- [0280] 例えば、計算部142は、情報処理装置100から取得部141が受信したイジングパラメータを用いて、計算を実行する。
- [0281] 送信部143は、外部の情報処理装置へ各種情報を送信する。例えば、送信部143は、情報処理装置100等の他の情報処理装置へ各種情報を送信する。送信部143は、記憶部12に記憶された情報を送信する。
- [0282] 送信部143は、情報処理装置100等の他の情報処理装置からの情報に基づいて、各種情報を送信する。送信部143は、記憶部12に記憶された

情報に基づいて、各種情報を送信する。

[0283] 送信部 143 は、計算部 142 による量子デバイス部 13 の測定結果を、計算の指示を行った装置に送信する。送信部 143 は、計算部 142 による量子デバイス部 13 の測定結果を、パラメータの送信元に送信する。送信部 143 は、計算部 142 による量子デバイス部 13 の測定結果を、計算の要求元に送信する。送信部 143 は、計算部 142 による量子デバイス部 13 の測定結果を、情報処理装置 100 等の他の情報処理装置に送信する。

[0284] 例えば、図 1 の例では、送信部 143 は、情報処理装置 100 から受信したパラメータを用いて計算（測定）したイジングスピンの値を情報処理装置 100 へ送信する。

[0285] [8. ユーザへの提示例]

ここで、情報処理システム 1 において各種情報のユーザへの提示例について、以下説明する。例えば、情報処理装置 100 は、端末装置 20 からリクエストに応じて、様々な情報を端末装置 20 へ提供する。

[0286] [8-1. クラスのリスト]

例えば、情報処理装置 100 が扱っているクラス、タスクのデータセットを端末装置 20 が要求した場合、情報処理装置 100 は、扱っているクラスのリストを開示してもよい。情報処理装置 100 は、図 12 に示すようなコンテンツ CT2 を端末装置 20 に提供してもよい。図 12 は、クラスのリストの一例を示す図である。図 12 に示すコンテンツ CT2 は、「ねずみ」、「うし」、「とら」、「たつ」等の複数のクラスの各々のサンプル画像のリストを示す。

[0287] [8-2. クラスの選択]

例えば、情報処理装置 100 は、端末装置 20 からユーザが学習したいクラスの選択を受け付けてもよい。情報処理装置 100 は、図 13 に示すようなコンテンツ CT3 を端末装置 20 に提供し、コンテンツ CT3 を介してユーザのクラスの選択を受け付けてもよい。図 13 は、ユーザが学習したいクラスの選択の一例を示す図である。図 13 に示すコンテンツ CT3 は、「ね

ずみ」に対応するクラスCL1、「うし」に対応するクラスCL2等に対するユーザの選択を受け付けるためのコンテンツである。情報処理装置100は、ユーザが選択したクラスのどれかであることを示す情報を端末装置20から受信する。例えば、ユーザは、すべて既知のクラスならば、リクエストタイプを「すべて既知」としてもよい。

[0288] [8-3. データの受け付け]

なお、情報処理装置100は、扱っていないクラス、タスクに適した（ネガティブ）データを端末装置20に要求してもよい。情報処理装置100は、端末装置20が扱いたいクラス、タスクのデータを端末装置20から受け付ける。情報処理装置100は、図14に示すようなコンテンツCT4を端末装置20に提供し、コンテンツCT4を介してユーザからデータを受け付けてもよい。図14は、ユーザからのデータの受け付けの一例を示す図である。図14に示すコンテンツCT4は、クラス「ニコウキスゲ」のデータを受け付ける場合の一例を示す。

[0289] なお、情報処理装置100は、クラスやデータセットと類似するクラスがないか確認し、似ているものがある場合は、ユーザに通知を行ってもよい。例えば、情報処理装置100は、クラスに似ているものがある場合は、「このクラスではないですか」とのメッセージをユーザが利用する端末装置20に送信してもよい。端末装置20は、受信したメッセージを出力する。そして、情報処理装置100は、クラスが新規であることが確認できたら、実際の最適学習データ選択処理に移ってもよい。

[0290] [9. 適用例]

上述した情報処理システム1は、例えば特定ドメインの種別分類等の種々の用途に利用されてもよい。この点について、以下いくつか例示を示す。なお、情報処理システム1を適用する対象は、以下に示す例に限られない。

[0291] [9-1. 第5の実施例（画像分類）]

ここから、情報処理システム1の適用例を示す第5の実施例として、動物の属に関する属内外の二クラス判別に情報処理システム1が適用された場合

について説明する。なお、第5の実施例では、動物の属に関する属内外の二クラス判別の例であるが、前述のように、他クラス分類を用いた動物の種類  
のクラス分類であってもよい。さらに、動物ではなく、高山植物の画像分類  
であってもよい。データは必ずしも画像である必要はなく、音声によるもの  
であってもよい。例えば、セミの鳴き声分類や野鳥の鳴き声による分類であ  
ってもよい。他に回帰問題の例としては、不動産の価格分類であってもよい  
。

[0292] 例えば、ドメインを横断して無作為に選んだ多数のクラスを大雑把に分類  
するより、特定ドメインで網羅的に選んだクラスを正確に分類したい場合が  
ある。例えば、前者は、イメージネット分類等の機械学習業界向けであり、  
後者はビジネス、個人ユーザ向けの考え方である。また、ビジネスの例とし  
ては、例えば図鑑等のWebサービス等が挙げられる。また、個人ユーザ向  
けの例としては、例えば、野鳥の写真や音声を集めて、オリジナルの認識モ  
ジュールを作りたい一般ユーザ等が挙げられる。

[0293] 第5の実施例における目的は、動物分類図におけるマカク属の画像分類を  
行うマカク属画像分類器（以下単に「分類器」ともいう）の生成であるもの  
とする。ただし、分類器にマカク属以外が入力されたらマカク属でないと正  
確に判定することも望まれているものとする。

[0294] 第5の実施例における前提条件としては、以下のような点があるものとす  
る。

- ・ 入力されるのは、オナガザル科だけとする
- ・ マカク属はどのクラスも、十分な数の画像を収集済み
- ・ マカク属以外のオナガザル科のデータは全くない
- ・ クライアント側で学習できるので、サーバからはデータセットだけ欲しい
- ・ サーバ側には、動物界分類図の全クラスのデータが多数格納済み
- ・ サーバ側には、汎用の画像特徴抽出器がある

[0295] ここから、図15を用いて、第5の実施例の実施構成の一例を説明する。  
図15は、情報処理システムを適用した第5の実施例を示す図である。なお

、上述した内容と同様の点については適宜説明を省略する。まず、図15に示す各構成要素について説明する。

[0296] 動物システムデータセット50-1は、動物の画像データセットである。動物システムデータセット50-1は、上述したデータベース50に対応する。例えば、動物システムデータセット50-1は、動物界の全クラスについて、各々100枚の画像を用意した巨大データセットである。

[0297] オナガザル科読み取り部131a-1は、オナガザル科のデータを取得する機能を有する。オナガザル科読み取り部131a-1は、上述したデータ読み取り部131aに対応する。例えば、オナガザル科読み取り部131a-1は、オナガザル科の全クラスの全データを読み取るモジュールである。オナガザル科読み取り部131a-1は、動物システムデータセット50-1からオナガザル科のデータを取得する。例えば、オナガザル科読み取り部131a-1は、図15中の第1対象群TG1に示すように、オナガザル属、ヒヒ属、マカク属、コロブス属、テングザル属、ラングール属等、オナガザル科に属する各属の画像データを取得する。

[0298] マカク属画像入力部131b-1は、マカク属のデータを取得する機能を有する。マカク属画像入力部131b-1は、上述したユーザデータ入力部131bに対応する。例えば、マカク属画像入力部131b-1は、クライアント（「ユーザ」ともいう）側が用意したマカク属全クラスについて、各々100枚のデータセットを取得する。例えば、マカク属画像入力部131b-1は、図15中の第2対象群TG2に示すように、マカク属に属する複数のクラスの各々の画像データを取得する。

[0299] データセットマージ部231は、オナガザル科画像（マカク属以外）と、入力されたマカク属画像をマージする機能を有する。データセットマージ部231は、オナガザル科読み取り部131a-1により取得された第1画像群と、マカク属画像入力部131b-1により取得された第2画像群とをマージして、第3画像群を生成する。

[0300] 組合せ最適実行部232は、マージデータに基づいて組合せ最適化の処理

を実行する機能を有する。組合せ最適実行部 232 は、最適化係数抽出部 132 及び最適化マシン通信部 133 に対応する。組合せ最適実行部 232 は、係数行列を抽出し、抽出した係数行列を最適化マシン 10 へ送信し、最適化マシン 10 から計算結果を受信する。

[0301] 図 15 のデータセット選択部 134 は、上述したデータセット選択部 134 に対応する。データセット選択部 134 は、組合せ最適実行部 232 による組合せ最適化の処理の結果を基に最適なデータを選択する機能を有する。データセット選択部 134 は、組合せ最適実行部 232 による組合せ最適化の処理の結果を基に、マージデータからサブセットを選択する。

[0302] 非マカク属画像出力部 135-1 は、オナガザル科セットから選ばれた非マカク属のデータセットを出力する機能を有する。非マカク属画像出力部 135-1 は、上述した出力部 135 に対応する。例えば、非マカク属画像出力部 135-1 は、選択されたサブセットのうち、マカク属以外のデータを出力する。例えば、非マカク属画像出力部 135-1 は、選択されたサブセットのうち、第 1 画像群に含まれる画像データを、マカク属以外の画像データとして出力する。

[0303] 最適化マシン 10 は、上述した最適化マシン 10 に対応する。最適化マシン 10 は、組合せ最適実行部 232 で作られた係数行列 (QUBO 行列) を入力し、組合せ結果を出力する。

[0304] 次に、図 16 を用いて、処理フローの一例について説明する。図 16 は、第 5 の実施例における処理の手順を示すフローチャートである。図 16 は、情報処理システム 1 による第 5 の実施例における処理のフローの例である。以下では、情報処理装置 100 を処理主体とする場合を一例として示すが、処理主体は、情報処理装置 100 に限らず、情報処理システム 1 に含まれるいずれの装置であってもよい。

[0305] 図 16 に示すように、情報処理装置 100 は、クライアントからデータを受け取る (ステップ S201)。例えば、情報処理装置 100 は、ユーザが利用する端末装置 20 からデータを取得する。例えば、情報処理装置 100

は、クライアントから様々な情報を受け取る。例えば、情報処理装置100は、オナガザル科等、データが欲しいドメインを示す情報を取得する。また、例えば、情報処理装置100は、マカク属内のクラス等、既にクライアント側にあるクラスを示す情報を取得する。また、例えば、情報処理装置100は、既にクライアント側にあるデータセットを取得する。

[0306] 情報処理装置100は、データベースからデータを集める（ステップS202）。例えば、情報処理装置100は、動物系統データセット50-1等のデータベース50から必要なデータを取得する。例えば、情報処理装置100は、動物画像セット等を有するデータベース50から必要なデータを集めてくる。また、例えば、情報処理装置100は、クライアントが指定したドメイン（オナガザル科）のうち、マカク属内のクラス等のポジティブクラスを除くデータセットを取得する。

[0307] 情報処理装置100は、集めたデータから最適化マシンへの入力を算出する（ステップS203）。例えば、情報処理装置100は、集めたデータから最適化マシン10への入力として、係数行列を算出する。

[0308] 情報処理装置100は、最適化マシンの出力に基づいてサブセットを生成する（ステップS204）。例えば、情報処理装置100は、係数行列を最適化マシン10に送信し、最適化マシン10から最適化マシン10が計算した計算結果を受信する。例えば、情報処理装置100は、最適化マシン10の計算結果に基づいて、集めたデータからサブセットを選択する。このように、情報処理装置100は、データセットから最適なサブセットを選択する。最適なサブセットは定性的には、例えば入力データに類似し、互いに類似しないデータの組合せである。

[0309] 例えば、情報処理装置100は、以下の手順でサブセットを選択する。まず、情報処理装置100は、データセットから既存の特徴抽出器で特徴量セットを生成する。そして、情報処理装置100は、特徴セットから最適化マシン10への入力データを生成する。ここで、入力データは、例えば、情報量ゲインの二値二次形式定式化における係数行列である。そして、情報処理

装置 100 は、最適化マシン 10 から情報量ゲイン最大化するデータ組合せを示す情報を取得する。情報処理装置 100 は、取得した組合せを示す情報に基づいて、サブセットを選択する。

[0310] 情報処理装置 100 は、クライアントにサブセットを提供する（ステップ S205）。例えば、情報処理装置 100 は、サブセットをクライアント側に提示する。例えば、情報処理装置 100 は、サブセットを、ユーザが利用する端末装置 20 へ送信する。

[0311] 上述した第 5 の実施例における処理に基づく全体シナリオとしては以下のようなになる。

[0312] クライアントは、分類したいクラスのデータを集める。例えば、クライアントは、マカク属のいくつかの種別の画像データを収集する。

[0313] また、クライアントは、クラスのデータを情報処理装置 100 に送り、クラス外データ提供を要求する。クライアントは、情報処理装置 100 が提示しているブラウザ等からドメイン（オナガザル科）を指定する。クライアントは、提供されたいクラスの代わりに収集したクラス（マカク属内の種別等）を指定する。

[0314] そして、クライアントは、クラス外データを受け取る。例えば、ユーザが利用する端末装置 20 は、情報処理装置 100 からデータを受信する。クライアントは、クラス外データも利用し、クラス内外の判別器を獲得する。クライアントは、収集したクラスデータと提供されたクラス外データで判別器を学習する。例えば、ユーザが利用する端末装置 20 は、自装置で有するデータ、及び情報処理装置 100 から受信したデータを用いて、判別器（モデル）を学習する。

[0315] そして、クライアントは、学習した判別器を、収集したクラスデータで学習したクラス分類器と組合せて使えるようにする。例えば、ユーザが利用する端末装置 20 は、判別器を用いて、画像がマカク属の画像であるかを判別する処理である第 1 処理を行い、第 1 処理でマカク属と判別された画像については、分類器を用いて、画像に含まれるマカク属の種別が何であるかを分

類する処理である第2処理を行う。

[0316] この点について、図17を用いて説明する。図17は、画像判定処理の流れを示す図である。以下では、端末装置20を処理主体とする場合を一例として示すが、処理主体は、端末装置20に限らず、情報処理システム1に含まれるいずれの装置であってもよい。

[0317] まず、端末装置20は、処理対象となる画像IM1がマカク属の画像であるかを判別する処理を行う（ステップS301）。例えば、端末装置20は、画像IM1を判別器に入力し、判別器の出力結果を基に、画像IM1がマカク属の画像であるかを判別する。

[0318] 第1処理で画像IM1がマカク属の画像ではないと判定した場合、端末装置20は、第2処理を行うことなく処理を終了する。例えば、第1処理で画像IM1がマカク属の画像ではないと判定した場合、端末装置20は、マカク属ではないことを示す結果RS1を出力して処理を終了する。

[0319] 一方で、第1処理で画像IM1がマカク属の画像であると判定した場合、端末装置20は、画像IM1に含まれる物体がマカク属の中のなにかを分類する処理を行う（ステップS302）。例えば、端末装置20は、画像IM1を分類器に入力し、分類器の出力結果を基に、画像IM1に含まれる物体がマカク属の中のなにかを分類する。図17では、端末装置20は、画像IM1に含まれる物体がマカク属の中のニホンザルであることを示す結果RS2を出力して処理を終了する。

[0320] [9-2. 第6の実施例（生体分子）]

例えば、機械学習のラベリングを様々な科学実験の計測評価と捉えると、多数ある調査対象のどのサンプルから評価すべきかの指針がわかる能動学習による効率アップの意義は大きい。

[0321] 例えば、半導体のバンドギャップは、半導体の様々な機能の特性を決める重要な物理量である。しかし、その計測には、機械学習のラベリングと同様に、その分野の専門家による手間と時間のコストがかかるため、計測対象のサンプルはできるだけ効率よく選ぶことが望まれている。同様に、磁性体の

磁化率、誘電体の誘電率等もそれぞれの機能の特性を決める重要な物理量であるが、その計測には特別の計測器が必要である。無機化合物ならず有機化合物等による機能材料でも当てはまる。

[0322] そこで、以下では、情報処理システム1の適用例を示す第6の実施例として、生体分子、特に遺伝子検査の例を説明する。

[0323] まず、生体分子というと普通、タンパク質や、RNA (Ribonucleic Acid ; リボ核酸)、DNA (Deoxyribonucleic Acid ; デオキシリボ核酸) 等が挙げられる。いずれの生体分子も、基本的な分子構造の一次元配列で構成される高分子であり、それぞれの生体分子に基づく生物のマクロな様々な機能は、その一次元配列の決められた領域における配列パターンによって決まることが知られている。

[0324] また、タンパク質は、生物の体、構造を構成するための基本的な生体分子として知られる。タンパク質の種類は世の中には膨大にあるが、いずれも、二十(二十一)種類のアミノ酸を数千から数億の長さで一次元配列的に並べたものであり、この配列の所定の領域のパターンに応じた固有な機能をもつ多数のタンパク質が知られている。

[0325] タンパク質のもつ固有な機能とは、例えば、酵素を活性化する(酵素)、生体構造を形成する(例えばコラーゲンやケラチン)、脂質やコレステロール等を運搬する(例えばアルブミン、アポリポタンパク質)、栄養やイオンを貯蔵する(例えばオボアルブミン、フェリチン、ヘモジデリン)、筋肉を構成して運動に関与する(例えばアクチン、ミオシン)、抗体と呼ばれる免疫機能に関与する(例えばグロブリン)、DNAの情報に基づいてタンパク質を合成する発現機能に関与したり、他のタンパク質の働きを調整したりする(例えばカルモジュリン)等である。

[0326] 同様に、RNA(リボ核酸)やDNA(デオキシリボ核酸)、遺伝子の実態となる生体分子として知られているが、その中身はタンパク質を合成するための情報を一次元配列上にコードした生体分子である。RNA、DNAもタンパク質と同様に、基本的な単位の配列となって、いずれも、デオキシリ

ボース（五炭糖）に核酸と四種類の塩基（アデニン（A）、グアニン（G）、シトシン（C）、チミン（T））のいずれかが結合して構成されるデオキシヌクレオチドが一次元配列的に並んだ高分子である。

[0327] RNAやDNAは、タンパク質同様、その配列（コード）のパターンに応じた固有な機能を持っている。DNAやRNAの基本的な機能は自己複製（転写）や、タンパク質の合成であるが、転写や合成のより詳細な機能（ブロック）では、例えば、メッセンジャーRNAにおけるタンパク質合成のための翻訳機能や、プロモータDNAによる転写活性の機能等が挙げられる。

[0328] これらの機能は、半導体の例等と同様に、特定の情報の計測によって確認することができる。たとえば、タンパク質配列情報に対する抗体結合の親和性情報の情報は平衡乖離度、酵素活性度の情報は酵素活性速度の計測によって確認する。また、遺伝子配列情報に対する生物の形質発現にかかわる情報（メッセンジャーRNAは翻訳効率で、プロモータDNAの転写効率の計測、これらは、いずれもデータベース（遺伝子や分子のデータセット）を構築する観点でみると、ラベリングとみなせる。

[0329] タンパク質やRNA、DNAは膨大な配列の組合せ中の特定の場所におけるパターンが所望の機能を有することは明らかになっているが、一方、必ずしもどのような機能をもたらすことになっているのかわからない箇所もかなりある。

[0330] 特に、DNAの場合、配列を情報ストレージとして見たてて、そのフォーマットが、コード領域と、非コード領域の2つのパートが混在した構成をしていると捉えるとわかりやすい。コード領域は、生成されるタンパク質や、あるいは、その結果としての生物の形質の情報との対応が明らかになっている領域である。また、非コード領域は、生成するタンパク質や発現する生物の形質との関係が明確ではなく、一見、遺伝情報に無関係な領域であると考えられていた領域である。

[0331] 従来、非コード領域は、遺伝情報との相関が低くあまり調査が進んでいなかった。しかし、昨今、DNA等では、コード領域以外からの情報を積極的

に解析してその機能の解析に利用しようという考えが進んでいる。

[0332] 例えば、ある種の神経難病は、異常タンパク質（ $\alpha$ シスクレイン、レビー小体等）の脳細胞への蓄積が原因とされる。この異常タンパク質と関係があるとされる領域は、家族性（遺伝性）の場合の例等から知られているが、家族性（遺伝性）ではない孤発タイプの場合には、どの領域が関与しているか明確ではない。

[0333] そこで、例えば、既知の関係領域から、探索領域を拡大しながら、同じ機能（異常タンパク質の合成）に関与している新規の配列領域や、パターンを探したり、あるいは、異常タンパク質の合成と相関がある別の機能を見つたりすることが進められる。

[0334] このように新規のパターンや新規の機能の発見を目的として、非コード領域の解析が進んでいる一方、非コード領域は、コード領域と同様に膨大な組合せが考えられるため、適切な探索が必要である。

[0335] そこで、情報処理システム1は、これらのタンパク質やRNA、DNAからそれぞれの持っている機能に固有な新たな配列パターンを効率良く探索し、また、探索したパターンから新たな機能を発見することに用いることができる。

[0336] 以下、まず、配列パターンの効率の良い探索を説明し、続いて、探索したパターンから新たな機能を発見する手順を説明する。

[0337] （手順#1. 統計予測モデルに用いるデータセットの収集）

まず、手順#1について説明する。情報処理システム1では、新たな情報をデータセットから獲得するため、予測モデルを活用している。ここでいう予測モデルとは、生体分子の配列パターンを入力すると、その配列パターンがその機能を持っている予測値、あるいは、確率分布のスコアを出力する。例えば、前述の神経難病の例であれば、異常タンパク質の合成に関与する配列パターンに該当するかの予測モデル（判別モデル）ある。このような予測モデルを構築するためには、まず、配列パターンと機能の予測値等の対応関係のデータセットが必要である。このデータセットは、多数の配列パターン

について、それぞれ、予測したい機能に関する情報を計測することで構築する。あるいは、既知のデータベースから機能の情報を元に逆引きして求めてもよい。

[0338] (手順# 2. 機械学習による統計予測モデルの学習)

次に手順# 2について説明する。手順# 1に続いて、情報処理システム1では、パラメータ付きの予測モデルを設計して、そのパラメータを機械学習技術で求める。代表的な予測モデルは、ニューラルネットであり、深層学習でパラメータを求めることができる。

[0339] 予測したい機能がどのようなもので、予測モデルの設計は異なるが、大雑把にまとめると以下のような設計方針となる。

- ・ 入力層は、配列パターン of 情報を入力できる構成をしている。
- ・ 中間層は、開発者の設定に基づき構成される。
- ・ 出力層は、予測したい機能に合わせて構成する。

[0340] 以下の出力周りに関する設計指針 (a) ~ (c) は共通である。

[0341] 設計指針 (a) は、以下のような方針である。

[0342] 設計指針 (a) では、機能の評価値が連続値ならば最終層を線形層にする。評価値が連続値の場合は、回帰モデルによる予測モデルを採用する。ニューラルネットの予測モデルの場合、最終層 (出力層) に線形層を用いる。出力層を線形層にすることで、第3の実施例で説明したような回帰モデルを使うことができる。連続値の予測値そのものだけでなく、その標準偏差や分散等も予測したい場合は、変分オートコーダを用いて、平均予測値と分散予測値を出すように構成してもよい。回帰モデルでは、二乗誤差を最小化するような学習を行う。変分オートエンコーダでは、分散 (共分散は含まないことが多い) を考慮した対数尤度の下界を最大化するように学習を行う。

[0343] 設計指針 (b) は、以下のような方針である。

[0344] 機能の評価値が有無等の二クラス、例えば、機能の評価をある特性の有無で評価したい場合等がある。このような場合は、二クラス分類が適切である。二クラス分類は、第3の実施例で説明したように、最終層にロジスティッ

ク回帰関数を有する線形層を用いる。ロジスティック回帰関数は、任意の入力に対して、0から1の間の値を出力する関数である。この構造を、0から1の確率を近似しているものとして、そのエントロピー、すなわちロジスティックエントロピーを最小化するように学習する。

[0345] 設計指針(c)は、以下のような方針である。

[0346] 機能の評価値が多クラス、すなわち機能が多数の特性のどれかがあてはまるという場合等では、多クラス分類のためのモデルの適用を考える。多クラス分類では、第3の実施例で説明したように、最終層にソフトマックス関数を有する線形層を用いる。ソフトマックス関数は、任意の入力に対して、出力は起こりうるクラスのいずれかのスコアであるが、これらのスコアは全てのスコアを足すと1になるように規格化されている。この構造は、多クラス分類における多クラスの確率値とみなせるように学習されることができる。学習は、ソフトマックスエントロピーを最小化するように行われる。

[0347] 以上は、一般的によく使われるニューラルネットの予測モデルの出力層の設計方法である。

[0348] (手順#3. 情報処理システム1による最適なサンプルの選択)

次に手順#3について説明する。上述の手順により、予測モデルが構築できたので、以降が第6の実施例における情報処理システム1の主な適用範囲である。第6の実施例によれば、予測モデルにとって、最も曖昧なデータ空間上の領域、すなわち、検証すると得られる情報量が最も大きい領域を、示すことができる。あるいは、そのサンプル情報を提供することができる。また、第6の実施例によれば、複数のサンプルを同時に提供することができ、そのサンプルは、情報量が多くなるように、互いに類似していないサンプル同士が選ばれる。

[0349] 以上のように、第6の実施例では、膨大な領域の中から適切なサンプルを選ぶ指針として、情報量を大きくする指針を採用している。

[0350] しかし、情報量を大きくする指針が定まっただけでは、膨大な探索領域の探索効率を高めたことにはならない。第6の実施例の場合、この膨大な探索

領域の探索は、劣モジュラ最大化という探索問題に含まれることがわかっているため、その汎用的な近似方法として貪欲法が一般的である。貪欲法は、膨大な探索領域を効率よく探索する方法で計算時間の爆発がおこりにくい。そのため、この方法で、適切なサンプルを選べば、効率よく適切なサンプルを選べる。しかし、貪欲法は、汎用的かつ高速であるが、より良い近似があればそれを採用したい。

[0351] 第6の実施例では、理想的な条件（断熱近似が満たされる条件）では最適解にたどり着くことが保証されている、量子アニーリングを利用して膨大な探索領域を狭める。その方法、定式化に関しては、第1の実施例～第4の実施例で説明したとおりである。なお、量子アニーリング以外にも、最適解にただ降り着くことが保証されている例として熱アニーリングのシミュレーション（シミュレーテッドアニーリング）等もある。

[0352] 以上を用いることで、膨大なサンプルから、現在の予測モデルの状態に適した最適なサンプルを選ぶことができる。

[0353] （手順#4. 選択されたサンプルの解析）

次に手順#4について説明する。最後に、選択されたサンプルに関する解析を行う。もともとの目的は、サンプルの機能の有無の判別、あるいは、機能のカテゴリへの分類、あるいは、機能の程度の定量化である。

[0354] これらの実験を、全てのサンプルに対して行わずに、選択されたサンプルだけで解析するのが最も効率が良い。これは、全てのサンプルの中で、計測しなくても、機能に関する情報が予測モデルだけで高い信頼性で予測できるものや、あるいは、計測しても、その領域の情報が少ないために、機能に関する情報との関連、規則性が導けないものは、あまり検証しても効果がないことを表している。

[0355] 予測モデル、すなわち、既に解析した知識からみて、適度に曖昧なため、もたらされる情報量が多いサンプルから、評価することで効率のよいデータセットの構築が実現されている。最後に、このサンプルの中から、あらたな、配列パターンを見出す（特に非コード領域から）あるいは、これらの配列

パターンに共通の新機能を見出すことで、予測モデルのみならず、研究者自身も知識を効率的に拡大していくことができる。

[0356] [10. その他の構成例等]

上述した実施形態や変形例に係る処理は、上記実施形態や変形例以外にも種々の異なる形態（変形例）にて実施されてよい。なお、上述した例において、サブセットを選択する母集団となるデータセットを「第1データセット」と読み替え、第1データセットから選択されるサブセットを「第2データセット」と読み替えてもよい。

[0357] [10-1. その他の構成例]

なお、上記の例では、情報処理装置100と、最適化マシン10とが別体である場合を示したが、情報処理装置100と最適化マシン10とは一体であってもよい。例えば、最適化マシン10が超伝導を使わずデジタル回路で実現された場合、最適化マシン10はエッジ側に配置されてもよい。例えば、イジングモデルを用いた計算がエッジ側で行われる場合、情報処理装置100と最適化マシン10とが一体であってもよい。

[0358] [10-2. プログラムやパラメータ等の処理に用いる情報の生成方法]

上述した情報処理やパラメータ等を生成する方法が提供されてもよい。また、上述した最適化マシン10が計算を実行する際に用いるプログラムを生成する方法が提供されてもよい。

[0359] [10-3. その他]

また、上記各実施形態において説明した各処理のうち、自動的に行われるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行われるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。この他、上記文書中や図面中で示した処理手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。例えば、各図に示した各種情報は、図示した情報に限られない。

[0360] また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも

物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況等に応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。

[0361] また、上述してきた各実施形態及び変形例は、処理内容を矛盾させない範囲で適宜組み合わせることが可能である。

[0362] また、本明細書に記載された効果はあくまで例示であって限定されるものではなく、他の効果があってもよい。

[0363] [11. 本開示に係る効果]

上記のように、本開示に係る情報処理装置（例えば実施形態では情報処理装置100に対応）は、取得部（例えば実施形態では取得部131に対応）と、選択部（例えば、実施形態ではデータセット選択部134に対応）とを備える。取得部は、データ供給方法と、学習したいモデルと、モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とを取得する。選択部は、モデルに応じて決定される情報エントロピーと、指定情報とに基づいて、データセットから、モデルの学習に用いるサンプルセットを選択する。

[0364] このように、情報処理装置は、学習したいモデルに応じて決定される目的関数と指定情報とに基づいて、データセットからサンプルセットを選択する。これにより、情報処理装置は、データセットから適切なサンプルセットを選択することができる。

[0365] データ供給方法は、データセットからのデータ供給であって、サンプルセットはデータセットのサブセットである。このように、情報処理装置は、データセットからのデータ供給を受け、データセットのサブセットをサンプルセットとして選択することで、データセットから適切なサンプルセットを選択することができる。

[0366] 学習したいモデルは、学習パラメータ付きの予測モデルであって、モデルのタスクとは、入力に対応する出力の種別である。このように、情報処理装

置は、学習パラメータ付きの予測モデルを、学習したいモデルとして、サンプルセットを選択することで、データセットから適切なサンプルセットを選択することができる。

[0367] また、モデルにもたらされる情報エントロピーは、カルバック・ライブラー情報量（「カルバックライブラー情報量」と記載する場合もある）、または、フィッシャー情報量を用いて算出される情報エントロピーである。このように、情報処理装置は、カルバックライブラー情報量、または、フィッシャー情報量を用いて算出される情報エントロピーであるに基づいて、データセットからサンプルセットを選択することで、データセットから適切なサンプルセットを選択することができる。

[0368] また、選択部は、モデルにもたらされる情報エントロピーを示す目的関数を最適化するように、サンプルセットを選択する。このように、情報処理装置は、情報エントロピーを示す目的関数を最適化するように、サンプルセットを選択することで、データセットから適切なサンプルセットを選択することができる。

[0369] また、選択部は、QUBO (Quadratic Unconstrained Binary Optimization) 形式で表現される目的関数に基づいて、サンプルセットを選択する。このように、情報処理装置は、QUBO形式で表現される目的関数に基づいて、データセットからサンプルセットを選択することで、データセットから適切なサンプルセットを選択することができる。

[0370] また、情報処理装置は、最適化マシン通信部（例えば、実施形態では最適化マシン通信部133に対応）を備える。最適化マシン通信部は、組合せ最適化計算を行う最適化マシン（例えば、実施形態では最適化マシン10に対応）に目的関数に対応する係数行列を送信し、最適化マシンから組合せ最適化計算の計算結果を受信する。選択部は、計算結果に基づいて、サンプルセットを選択する。このように、情報処理装置は、最適化マシンによる計算結果を用いてゲインを示す目的関数に基づいて、データセットからサンプルセットを選択することで、データセットから適切なサンプルセットを選択する

ことができる。

[0371] また、最適化マシン通信部は、組合せ最適化計算後の変数を示す計算結果を最適化マシンから受信する。このように、情報処理装置は、最適化マシンから受信した組合せ最適化計算後の変数を用いて、データセットからサンプルセットを選択することで、データセットから適切なサンプルセットを選択することができる。

[0372] また、最適化マシン通信部は、各々がデータに対応するバイナリ変数に関する計算結果を最適化マシンから受信する。このように、情報処理装置は、最適化マシンから受信した最適化されたバイナリ変数を用いて、データセットからサンプルセットを選択することで、データセットから適切なサンプルセットを選択することができる。

[0373] また、最適化マシン通信部は、量子コンピュータ、または組合せ最適化アクセラレータに係数行列を送信する。このように、情報処理装置は、量子コンピュータ、または組合せ最適化アクセラレータによる計算結果を用いて、データセットからサンプルセットを選択することで、データセットから適切なサンプルセットを選択することができる。

[0374] また、最適化マシン通信部は、複数の最適化マシンのうち、ユーザにより選択された最適化マシンに係数行列を送信する。このように、情報処理装置は、複数の最適化マシンのうち、ユーザにより選択された最適化マシンに係数行列を送信することで、ユーザの選択に応じたサンプルセットを選択することができるため、データセットから適切なサンプルセットを選択することができる。

[0375] また、情報処理装置は、抽出部（例えば、実施形態では最適化係数抽出部 132 に対応）を備える。抽出部は、係数行列を抽出する。最適化マシン通信部は、抽出部により抽出された係数行列を最適化マシンに送信する。このように、情報処理装置は、抽出した係数行列を最適化マシンに送信することで、最適化マシンから適切な計算結果を受信することができ、データセットから適切なサンプルセットを選択することができる。

[0376] また、抽出部は、目的関数からイジング係数に対応する係数行列を抽出する。このように、情報処理装置は、目的関数から抽出したイジング係数に対応する係数行列を最適化マシンに送信することで、最適化マシンから適切な計算結果を受信することができ、データセットから適切なサンプルセットを選択することができる。

[0377] また、取得部は、ユーザが学習したい予測モデルであるモデルを取得する。このように、情報処理装置は、ユーザが学習したい予測モデルを対象として、データセットから適切なサンプルセットを選択することができる。

[0378] また、情報処理装置は、出力部（例えば、実施形態では出力部135に対応）を備える。出力部は、選択部により選択されたサンプルセットに関する情報を出力する。このように、情報処理装置は、選択されたサンプルセットに関する情報を出力することにより、選択したサンプルセットに応じて適切な情報提供を行うことができる。

[0379] また、出力部は、サンプルセットをユーザが利用する端末装置（例えば、実施形態では端末装置20に対応）へ送信する。このように、情報処理装置は、サンプルセットをユーザが利用する端末装置へ送信することで、選択したサンプルセットをユーザに提供することができる。

[0380] また、出力部は、サンプルセットを用いて学習された学習済みモデルをユーザが利用する端末装置へ送信する。このように、情報処理装置は、サンプルセットを用いて学習された学習済みモデルをユーザが利用する端末装置へ送信することで、適切に学習されたモデルをユーザに提供することができる。

[0381] [12. ハードウェア構成]

上述してきた各実施形態や変形例に係る情報処理装置100等の情報機器は、例えば図18に示すような構成のコンピュータ1000によって実現される。図18は、情報処理装置の機能を実現するコンピュータの一例を示すハードウェア構成図である。以下、情報処理装置100を例に挙げて説明する。コンピュータ1000は、CPU1100、RAM1200、ROM(R

ead Only Memory) 1300、HDD (Hard Disk Drive) 1400、通信インターフェイス1500、及び入出力インターフェイス1600を有する。コンピュータ1000の各部は、バス1050によって接続される。

[0382] CPU1100は、ROM1300又はHDD1400に格納されたプログラムに基づいて動作し、各部の制御を行う。例えば、CPU1100は、ROM1300又はHDD1400に格納されたプログラムをRAM1200に展開し、各種プログラムに対応した処理を実行する。

[0383] ROM1300は、コンピュータ1000の起動時にCPU1100によって実行されるBIOS (Basic Input Output System) 等のブートプログラムや、コンピュータ1000のハードウェアに依存するプログラム等を格納する。

[0384] HDD1400は、CPU1100によって実行されるプログラム、及び、かかるプログラムによって使用されるデータ等を非一時的に記録する、コンピュータが読み取り可能な記録媒体である。具体的には、HDD1400は、プログラムデータ1450の一例である本開示に係る情報処理プログラム等の情報処理プログラムを記録する記録媒体である。

[0385] 通信インターフェイス1500は、コンピュータ1000が外部ネットワーク1550 (例えばインターネット) と接続するためのインターフェイスである。例えば、CPU1100は、通信インターフェイス1500を介して、他の機器からデータを受信したり、CPU1100が生成したデータを他の機器へ送信したりする。

[0386] 入出力インターフェイス1600は、入出力デバイス1650とコンピュータ1000とを接続するためのインターフェイスである。例えば、CPU1100は、入出力インターフェイス1600を介して、キーボードやマウス等の入力デバイスからデータを受信する。また、CPU1100は、入出力インターフェイス1600を介して、ディスプレイやスピーカーやプリンタ等の出力デバイスにデータを送信する。また、入出力インターフェイス1600は、所定の記録媒体 (メディア) に記録されたプログラム等を読み取

るメディアインターフェイスとして機能してもよい。メディアとは、例えば DVD (Digital Versatile Disc)、PD (Phase change rewritable Disk) 等の光学記録媒体、MO (Magneto-Optical disk) 等の光磁気記録媒体、テープ媒体、磁気記録媒体、または半導体メモリ等である。

[0387] 例えば、コンピュータ1000が情報処理装置100として機能する場合、コンピュータ1000のCPU1100は、RAM1200上にロードされた情報処理プログラム等の情報処理プログラムを実行することにより、制御部130等の機能を実現する。また、HDD1400には、本開示に係る情報処理プログラム等の情報処理プログラムや、記憶部120内のデータが格納される。なお、CPU1100は、プログラムデータ1450をHDD1400から読み取って実行するが、他の例として、外部ネットワーク1550を介して、他の装置からこれらのプログラムを取得してもよい。

[0388] なお、本技術は以下のような構成も取ることができる。

(1)

データ供給方法と、学習したいモデルと、前記モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とを取得する取得部と、

前記モデルに応じて決定される情報エントロピーと、前記指定情報とに基づいて、データセットから、前記モデルの学習に用いるサンプルセットを選択する選択部と、

を備える情報処理装置。

(2)

前記データ供給方法は、前記データセットからのデータ供給であって、前記サンプルセットは前記データセットのサブセットである

(1)に記載の情報処理装置。

(3)

前記学習したいモデルは、学習パラメータ付きの予測モデルであって、前記モデルのタスクとは、入力に対応する出力の種別である、

(1)または(2)に記載の情報処理装置。

(4)

前記モデルにもたらされる情報エントロピーは、カルバック・ライブラー情報量、または、フィッシャー情報量を用いて算出される情報エントロピーである、

(1) ~ (3) のいずれか1つに記載の情報処理装置。

(5)

前記選択部は、

前記モデルにもたらされる情報エントロピーを示す目的関数を最適化するように、前記サンプルセットを選択する

(1) ~ (4) のいずれか1つに記載の情報処理装置。

(6)

前記選択部は、

QUBO (Quadratic Unconstrained Binary Optimization) 形式で表現される前記目的関数に基づいて、前記サンプルセットを選択する

(5) に記載の情報処理装置。

(7)

組合せ最適化計算を行う最適化マシンに前記目的関数に対応する係数行列を送信し、前記最適化マシンから前記組合せ最適化計算の計算結果を受信する最適化マシン通信部、

を備え、

前記選択部は、

前記計算結果に基づいて、前記サンプルセットを選択する

(5) または (6) に記載の情報処理装置。

(8)

前記最適化マシン通信部は、

前記組合せ最適化計算後の変数を示す前記計算結果を前記最適化マシンから受信する

(7) に記載の情報処理装置。

(9)

前記最適化マシン通信部は、  
各々がデータに対応するバイナリ変数に関する前記計算結果を前記最適化マシンから受信する

(8)に記載の情報処理装置。

(10)

前記最適化マシン通信部は、  
量子コンピュータ、または組合せ最適化アクセラレータに前記係数行列を送信する

(7)～(9)のいずれか1つに記載の情報処理装置。

(11)

前記最適化マシン通信部は、  
複数の最適化マシンのうち、ユーザにより選択された前記最適化マシンに前記係数行列を送信する

(7)～(10)のいずれか1つに記載の情報処理装置。

(12)

前記係数行列を抽出する抽出部、  
を備え、

前記最適化マシン通信部は、  
前記抽出部により抽出された前記係数行列を前記最適化マシンに送信する

(7)～(11)のいずれか1つに記載の情報処理装置。

(13)

前記抽出部は、  
前記目的関数からイジング係数に対応する前記係数行列を抽出する

(12)に記載の情報処理装置。

(14)

前記取得部は、  
ユーザが学習したい予測モデルである前記モデルを取得する

(1) ~ (13) のいずれか1つに記載の情報処理装置。

(15)

前記選択部により選択された前記サンプルセットに関する情報を入力する出力部、

を備える

(1) ~ (14) のいずれか1つに記載の情報処理装置。

(16)

前記出力部は、

前記サンプルセットをユーザが利用する端末装置へ送信する

(15) に記載の情報処理装置。

(17)

前記出力部は、

前記サンプルセットを用いて学習された学習済みモデルをユーザが利用する端末装置へ送信する

(15) または (16) に記載の情報処理装置。

(18)

データ供給方法と、学習したいモデルと、前記モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とを取得し、

前記モデルに応じて決定される情報エントロピーと、前記指定情報とに基づいて、データセットから、前記モデルの学習に用いるサンプルセットを選択する

処理を実行する情報処理方法。

(19)

データ供給方法と、学習したいモデルと、前記モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とを取得し、

前記モデルに応じて決定される情報エントロピーと、前記指定情報とに基づいて、データセットから、前記モデルの学習に用いるサンプルセットを選択する

処理を実行させる情報処理プログラム。

## 符号の説明

- [0389] 1 情報処理システム
  - 1 0 0 情報処理装置
    - 1 1 0 通信部
    - 1 2 0 記憶部
      - 1 2 1 データ記憶部
      - 1 2 2 関数情報記憶部
    - 1 3 0 制御部
      - 1 3 1 取得部
      - 1 3 2 最適化係数抽出部
      - 1 3 3 最適化マシン通信部
      - 1 3 4 データセット選択部
      - 1 3 5 出力部
  - 1 0 最適化マシン
    - 1 1 通信部
    - 1 2 記憶部
    - 1 3 量子デバイス部
    - 1 4 制御部
      - 1 4 1 取得部
      - 1 4 2 計算部
      - 1 4 3 送信部
  - 2 0 端末装置
  - 5 0 データベース

## 請求の範囲

- [請求項1] データ供給方法と、学習したいモデルと、前記モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とを取得する取得部と、
- 前記モデルに応じて決定される情報エントロピーと、前記指定情報とに基づいて、データセットから、前記モデルの学習に用いるサンプルセットを選択する選択部と、
- を備える情報処理装置。
- [請求項2] 前記データ供給方法は、前記データセットからのデータ供給であって、前記サンプルセットは前記データセットのサブセットである
- 請求項1に記載の情報処理装置。
- [請求項3] 前記学習したいモデルは、学習パラメータ付きの予測モデルであって、前記モデルのタスクとは、入力に対応する出力の種別である、
- 請求項1に記載の情報処理装置。
- [請求項4] 前記モデルにもたらされる情報エントロピーは、カルバック・ライブラー情報量、または、フィッシャー情報量を用いて算出される情報エントロピーである、
- 請求項1に記載の情報処理装置。
- [請求項5] 前記選択部は、
- 前記モデルにもたらされる情報エントロピーを示す目的関数を最適化するように、前記サンプルセットを選択する
- 請求項1に記載の情報処理装置。
- [請求項6] 前記選択部は、
- QUBO (Quadratic Unconstrained Binary Optimization) 形式で表現される前記目的関数に基づいて、前記サンプルセットを選択する
- 請求項5に記載の情報処理装置。
- [請求項7] 組合せ最適化計算を行う最適化マシンに前記目的関数に対応する係

数行列を送信し、前記最適化マシンから前記組合せ最適化計算の計算結果を受信する最適化マシン通信部、

を備え、

前記選択部は、

前記計算結果に基づいて、前記サンプルセットを選択する

請求項 5 に記載の情報処理装置。

[請求項8]

前記最適化マシン通信部は、

前記組合せ最適化計算後の変数を示す前記計算結果を前記最適化マシンから受信する

請求項 7 に記載の情報処理装置。

[請求項9]

前記最適化マシン通信部は、

各々がデータに対応するバイナリ変数に関する前記計算結果を前記最適化マシンから受信する

請求項 8 に記載の情報処理装置。

[請求項10]

前記最適化マシン通信部は、

量子コンピュータ、または組合せ最適化アクセラレータに前記係数行列を送信する

請求項 7 に記載の情報処理装置。

[請求項11]

前記最適化マシン通信部は、

複数の最適化マシンのうち、ユーザにより選択された前記最適化マシンに前記係数行列を送信する

請求項 7 に記載の情報処理装置。

[請求項12]

前記係数行列を抽出する抽出部、

を備え、

前記最適化マシン通信部は、

前記抽出部により抽出された前記係数行列を前記最適化マシンに送信する

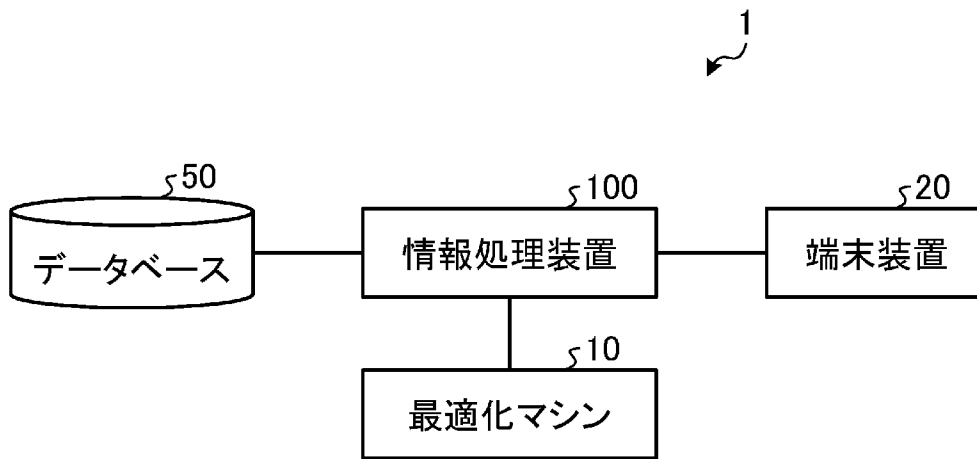
請求項 7 に記載の情報処理装置。

- [請求項13] 前記抽出部は、  
前記目的関数から前記最適化マシンの入力に対応する前記係数行列を抽出する  
請求項12に記載の情報処理装置。
- [請求項14] 前記取得部は、  
ユーザが学習したい予測モデルである前記モデルを取得する  
請求項1に記載の情報処理装置。
- [請求項15] 前記選択部により選択された前記サンプルセットに関する情報を出力する出力部、  
を備える  
請求項1に記載の情報処理装置。
- [請求項16] 前記出力部は、  
前記サンプルセットをユーザが利用する端末装置へ送信する  
請求項15に記載の情報処理装置。
- [請求項17] 前記出力部は、  
前記サンプルセットを用いて学習された学習済みモデルをユーザが利用する端末装置へ送信する  
請求項15に記載の情報処理装置。
- [請求項18] データ供給方法と、学習したいモデルと、前記モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とを取得し、  
前記モデルに応じて決定される情報エントロピーと、前記指定情報とに基づいて、データセットから、前記モデルの学習に用いるサンプルセットを選択する  
処理を実行する情報処理方法。
- [請求項19] データ供給方法と、学習したいモデルと、前記モデルの学習に用いるサンプルセットのサイズとカテゴリに関する指定情報とを取得し、  
前記モデルに応じて決定される情報エントロピーと、前記指定情報とに基づいて、データセットから、前記モデルの学習に用いるサンプル

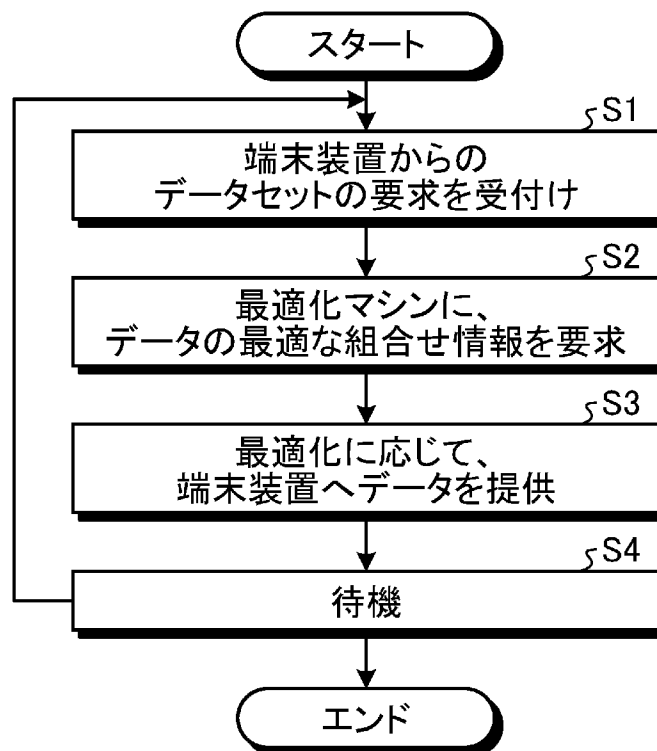
ルセットを選択する

処理を実行させる情報処理プログラム。

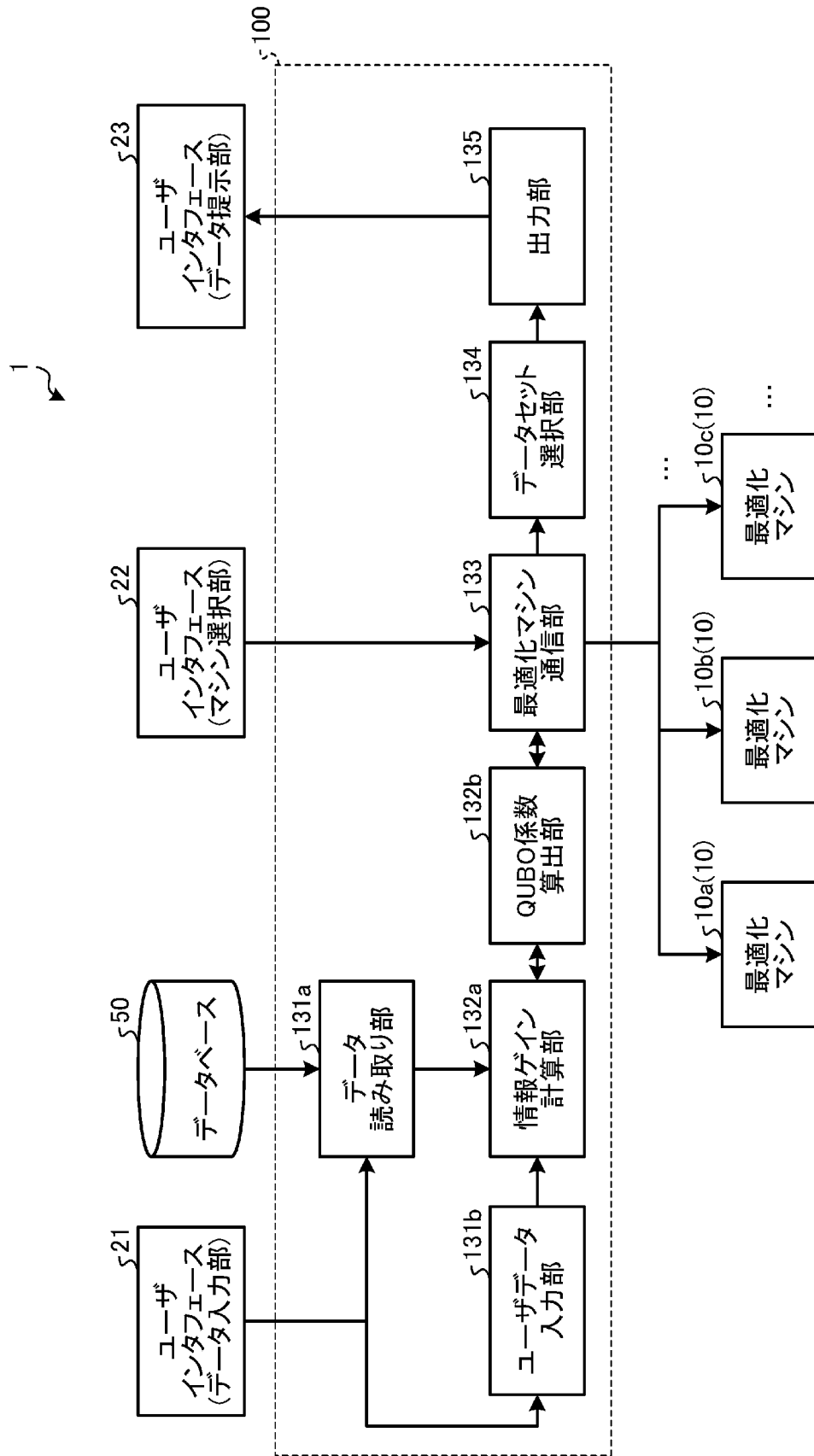
[図1]



[図2]



[図3]



[図4]

5CT1

ご利用したいマシンを選択してください。

マシンA

- ・サイズ・中小
- ・全結合・不可
- ・精度・良い
- ・速度・高速
- ・通信・遅い
- ・価格・高い

マシンB

- ・サイズ・大
- ・全結合・不可
- ・精度・悪い
- ・速度・高速
- ・通信・遅い
- ・価格・安い

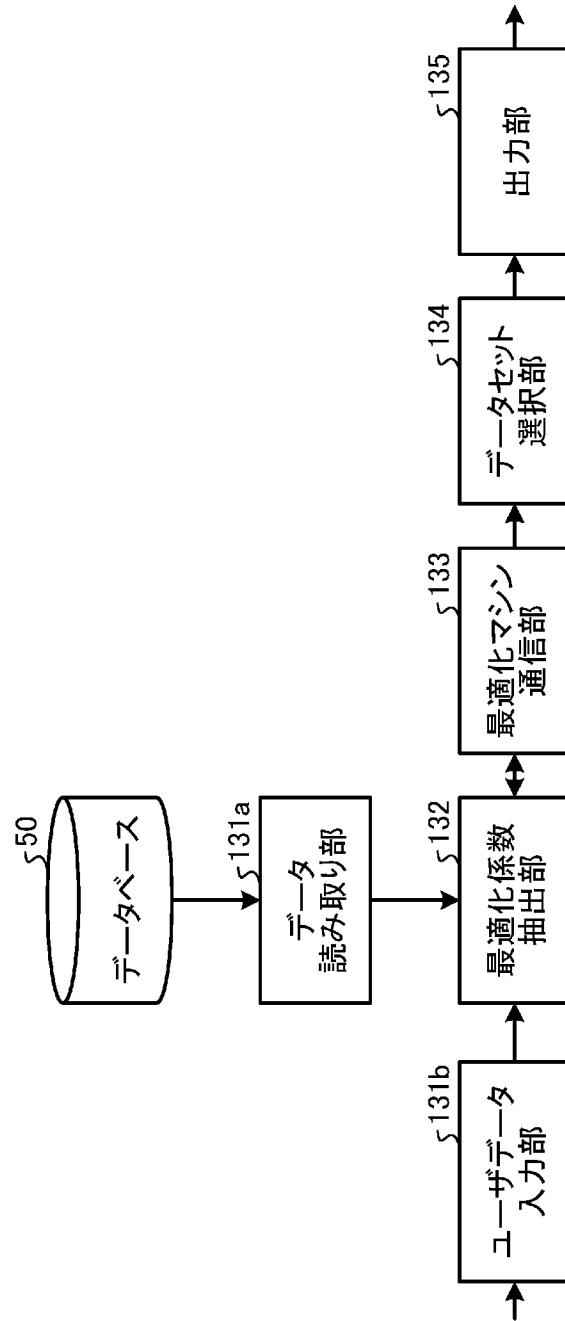
マシンC

- ・サイズ・中小
- ・全結合・不可
- ・精度・良い
- ・速度・普通
- ・通信・遅い
- ・価格・高い

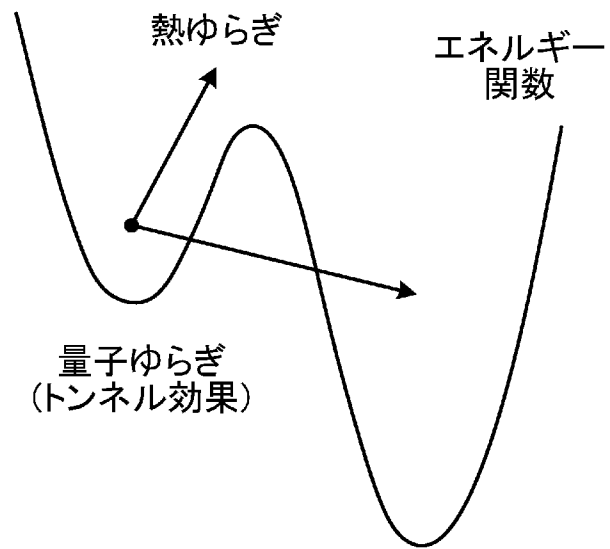
マシンD

- ・サイズ・中小
- ・全結合・可
- ・精度・不明
- ・速度・不明
- ・通信・遅い
- ・価格・無料

[図5]



[図6]



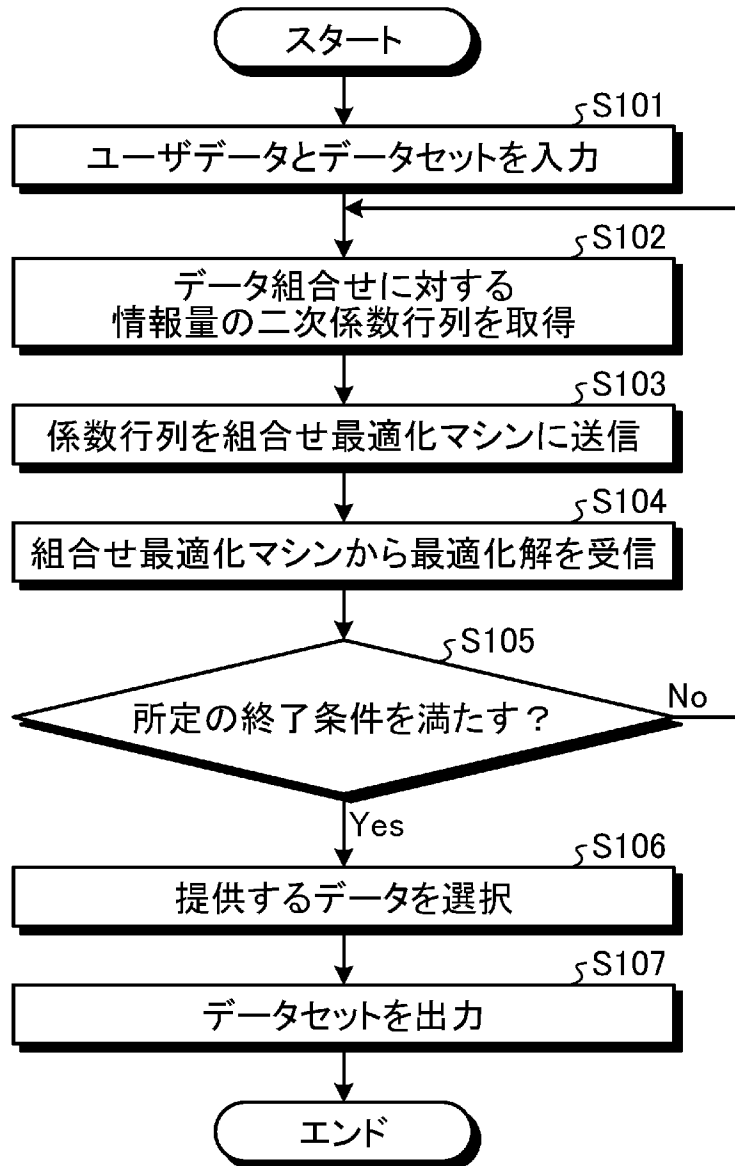
[図7]

PG1

```
model = load_model()  
dataset = load_dataset()  
userdata = load_userdata()  
dataset = dataset + userdata  
subset = resample_optimal_subset(model, dataset)
```

```
#学習したい機能の読み取り  
#データセットの読み取り  
#ユーザデータの読み取り  
#データセットとユーザデータのマージ  
#サブセットのリサンプリング
```

[図8]



[図9]

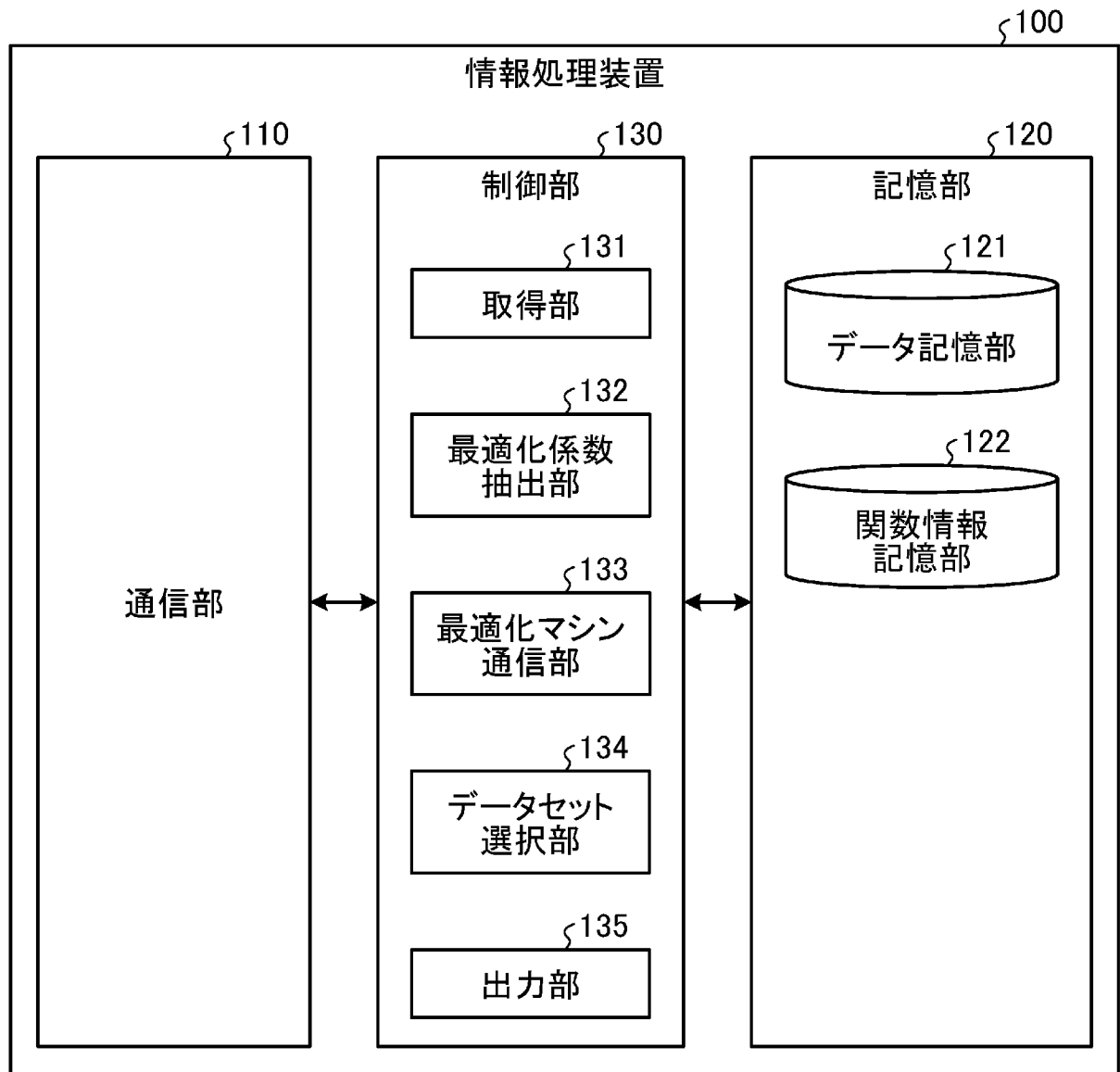
5 PG2

```
function resample_optimal_subset(model, dataset)
  b = sample_binaries(N)
  machine = OptimizingMachine()
  while(not converged) {
    subset = make_subset(b, dataset)
    w = maximum_likelihood(model, subset)
    hessians = calc_hessaians(w, dataset)
    x = calc_additional_variables(hessians, b)
    q = calc_qubo_coefficients(hessians, x)
    b_ = machine.get_optimul(q)
  }
  return subset
```

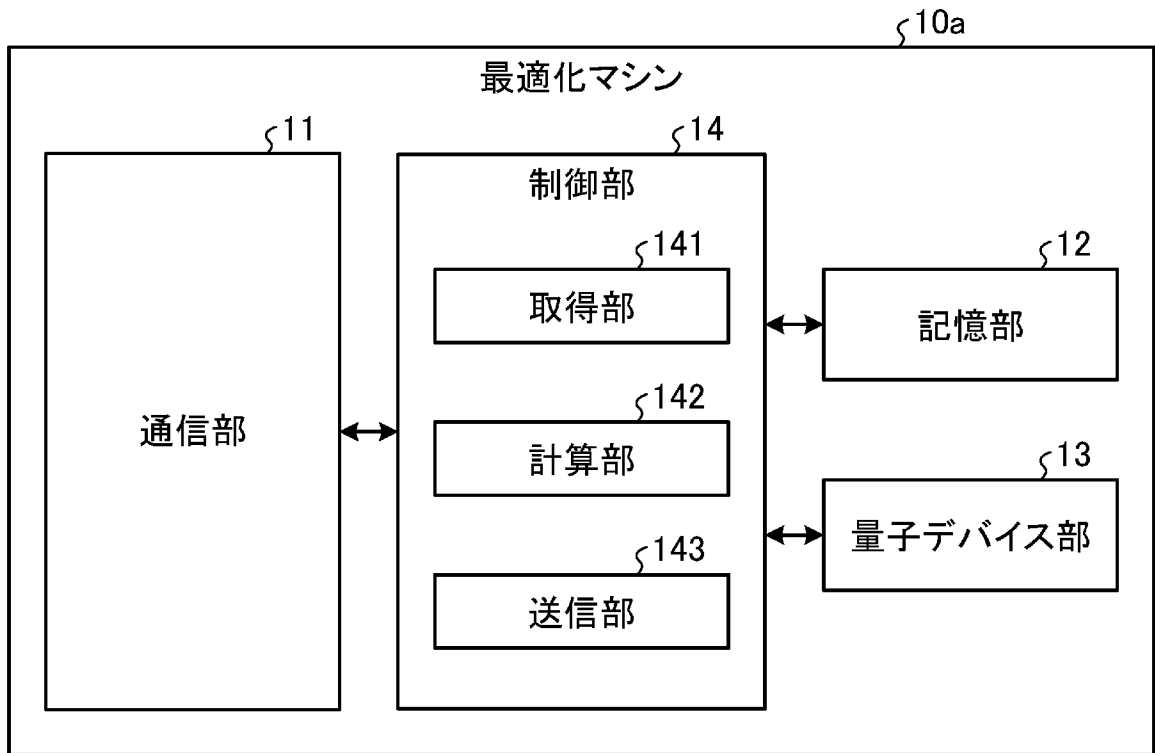
#バイナリ変数  
#ライブラリがサポートするマシン

#サブセットを用いてモデルを更新する処理  
#フィッシャー情報量を計算する処理  
#補助変数を用いて二次形式化する処理  
#二次形式からイジング係数を抽出  
#ライブラリがサポートする関数

[図10]



[図11]



[図12]

CT2

クラス名	サンプル#1	サンプル#2	サンプル#3	サンプル#4	...
ねずみ	ネズミ画像#1	ネズミ画像#2	ネズミ画像#3	ネズミ画像#4	...
うし	ウシ画像#1	ウシ画像#2	ウシ画像#3	ウシ画像#4	...
とら	トラ画像#1	トラ画像#2	トラ画像#3	トラ画像#4	...
たつ	タツ画像#1	タツ画像#2	タツ画像#3	タツ画像#4	...
...	...	...	...	...	...

[図13]

5CT3

学習したいクラス名を入力してください

5CL1 ねずみ	5CL2 とら	...
-------------	------------	-----

[図14]

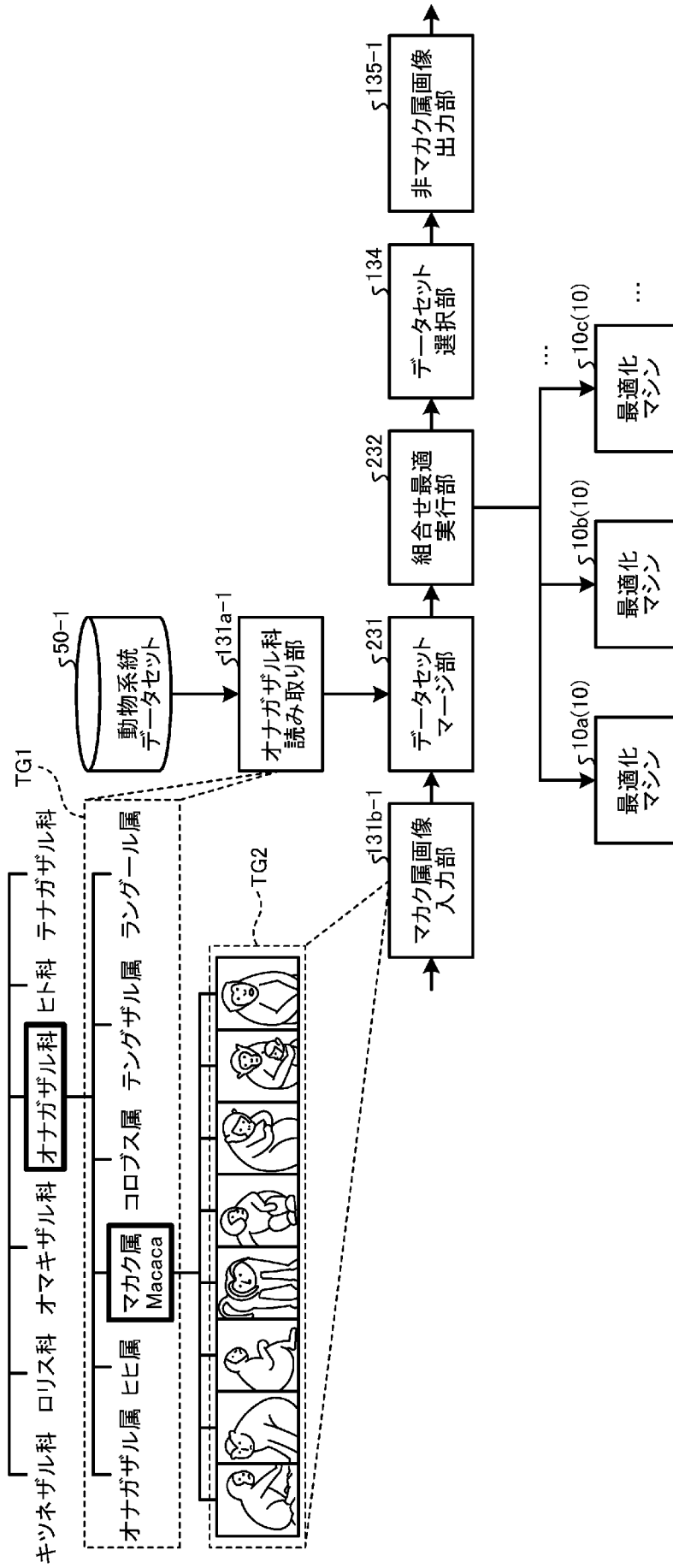
5CT4

学習したいクラス名を入力して、データをアップロードしてください

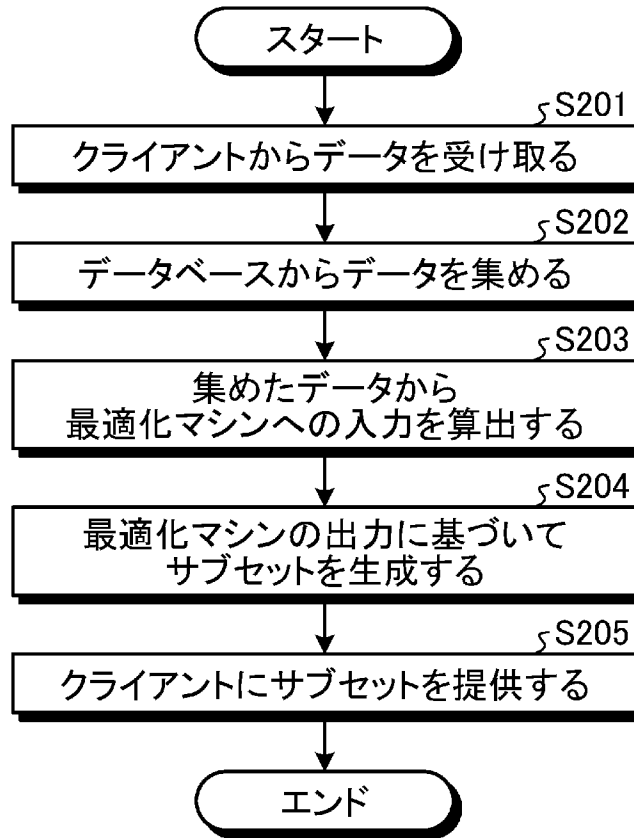
クラス名	アップロードするデータセット
ニッコウキスゲ	C:¥MyDataset¥高山植物¥ニッコウキスゲ

さらに、クラスを追加する場合は、こちらをクリックしてください

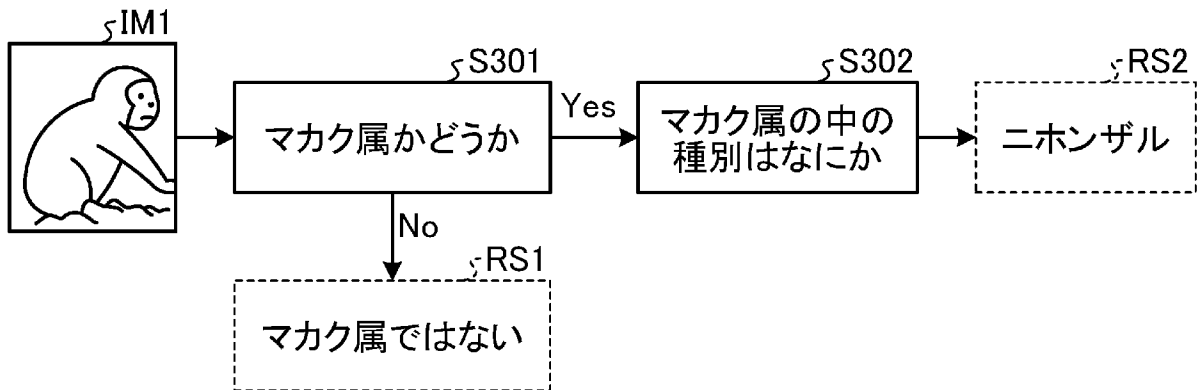
[図15]



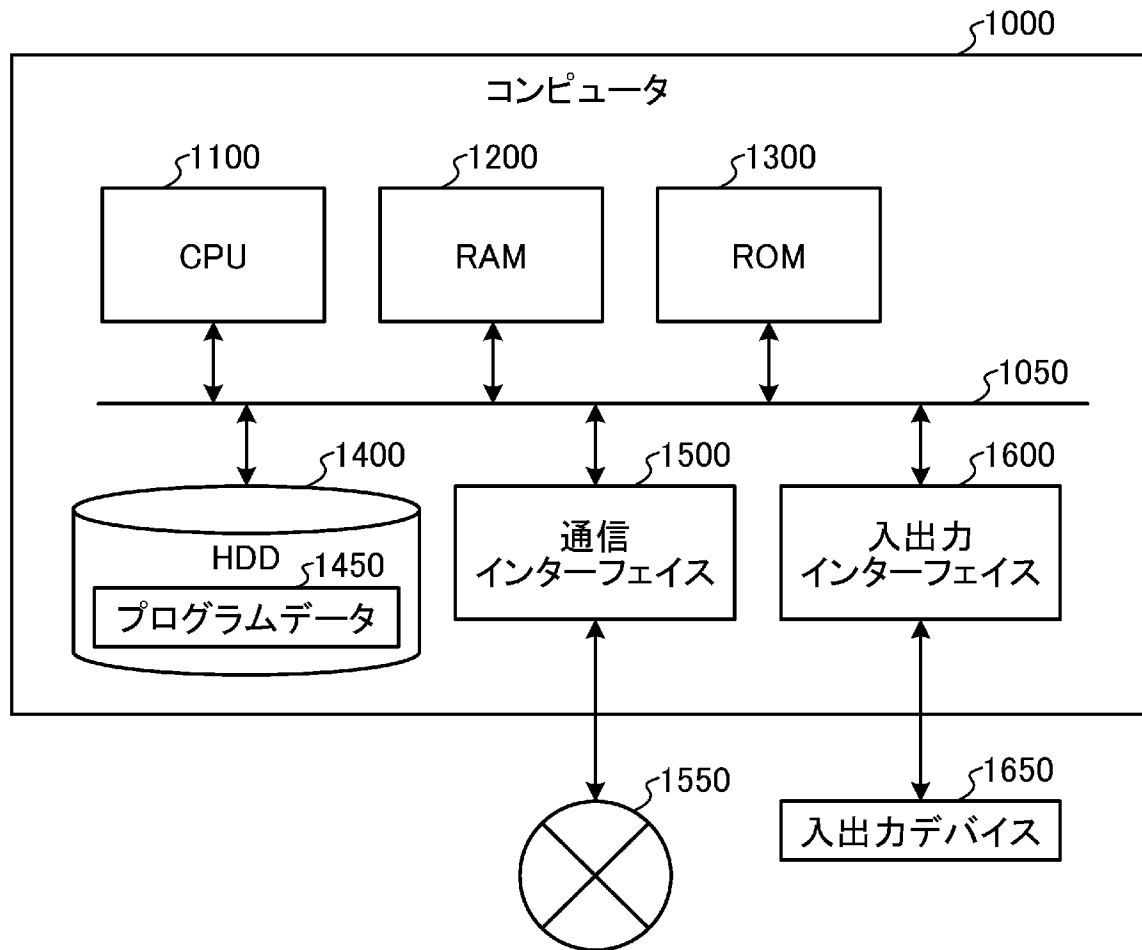
[図16]



[図17]



[図18]



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/002805

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
<i>G06N 20/00</i> (2019.01)i; <i>G06N 99/00</i> (2019.01)i FI: G06N20/00; G06N99/00 180		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) G06N20/00; G06N99/00		
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 1971-2022 Registered utility model specifications of Japan 1996-2022 Published registered utility model applications of Japan 1994-2022		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KILLAMSETTY, K. et al. GLISTER: Generalization based Data Subset Selection for Efficient and Robust Learning. arXiv.org [online]. 19 December 2020, [retrieved on 12 April 2022], Retrieved from the Internet: <URL: <a href="https://arxiv.org/abs/2012.10630v1">https://arxiv.org/abs/2012.10630v1</a> > pp. 3-6	1-19
Y	Cross Entropy. Wikipedia [online]. 29 December 2018, [retrieved on 12 April 2022], Retrieved from the Internet: <URL: <a href="https://en.wikipedia.org/w/index.php?title=Cross_entropy&amp;oldid=875854246">https://en.wikipedia.org/w/index.php?title=Cross_entropy&amp;oldid=875854246</a> > section "Definition", "Relation to log-likelihood"	1-19
Y	RAHMAN, M. T. et al. Near-Optimal Resampling in Particle Filters Using the Ising Energy Model. 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021) [online]. 06 June 2021, pp. 5464-5468, [retrieved on 12 April 2021], Retrieved from the Internet: <URL: <a href="https://ieeexplore.ieee.org/document/9413633">https://ieeexplore.ieee.org/document/9413633</a> >, <DOI: 10.1109/ICASSP39728.9413633> section 4	6-13
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> 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 <b>05 April 2022</b>		Date of mailing of the international search report <b>19 April 2022</b>
Name and mailing address of the ISA/JP <b>Japan Patent Office (ISA/JP) 3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915 Japan</b>		Authorized officer  Telephone No.

<p>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; G06N 99/00(2019.01)i                  FI: G06N20/00; G06N99/00 180</p>										
<p>B. 調査を行った分野</p>										
<p>調査を行った最小限資料（国際特許分類（IPC））                  G06N20/00; G06N99/00</p>										
<p>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</p> <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 - 2022年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2022年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2022年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2022年	日本国実用新案登録公報	1996 - 2022年	日本国登録実用新案公報	1994 - 2022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 - 2022年									
日本国実用新案登録公報	1996 - 2022年									
日本国登録実用新案公報	1994 - 2022年									
<p>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）</p>										
<p>C. 関連すると認められる文献</p>										
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	KILLAMSETTY, K., et al., GLISTER: Generalization based Data Subset Selection for Efficient and Robust Learning, arXiv.org [online], 2020.12.19, [retrieved on 2022-04-12], Retrieved from the Internet: <URL: https://arxiv.org/abs/2012.10630v1> pp. 3-6	1-19								
Y	Cross Entropy, Wikipedia [online], 2018.12.29, [retrieved on 2022-04-12], Retrieved from the Internet: <URL: https://en.wikipedia.org/w/index.php?title=Cross_entropy&oldid=875854246> Secs. "Definition", "Relation to log-likelihood"	1-19								
Y	RAHMAN, M.T., et al., Near-Optimal Resampling in Particle Filters Using the Ising Energy Model, 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021) [online], 2021.06.06, pp. 5464-5468, [retrieved on 2021-04-12], Retrieved from the Internet: <URL: https://ieeexplore.ieee.org/document/9413633>, <DOI: 10.1109/ICASSP39728.9413633> Sec. 4	6-13								
<p><input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。</p>										
<p>* 引用文献のカテゴリー                  "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                  "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                  "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）                  "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                  "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                  "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                  "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                  "&amp;" 同一パテントファミリー文献</p>										
国際調査を完了した日	05.04.2022	国際調査報告の発送日 19.04.2022								
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）  金木 陽一 5B 4876  電話番号 03-3581-1101 内線 3545									