

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B1)

(11) 特許番号

特許第6774129号
(P6774129)

(45) 発行日 令和2年10月21日(2020.10.21)

(24) 登録日 令和2年10月6日(2020.10.6)

(51) Int.Cl.

G06N 20/00 (2019.01)

F I

G06N 20/00

請求項の数 13 (全 15 頁)

(21) 出願番号 特願2020-16010 (P2020-16010)
 (22) 出願日 令和2年2月3日(2020.2.3)
 審査請求日 令和2年2月3日(2020.2.3)

早期審査対象出願

(73) 特許権者 520040304
 窪田 望
 東京都港区港南1丁目9番36号 NTT
 DATA 品川ビル13階 株式会社ク
 リエイターズネクスト内
 (74) 代理人 100079108
 弁理士 稲葉 良幸
 (74) 代理人 100109346
 弁理士 大貫 敏史
 (74) 代理人 100117189
 弁理士 江口 昭彦
 (74) 代理人 100134120
 弁理士 内藤 和彦

最終頁に続く

(54) 【発明の名称】 解析装置、解析方法及び解析プログラム

(57) 【特許請求の範囲】

【請求項1】

複数のアルゴリズムを用いて、所定の問題について設定された第1損失関数の値を小さくするように所定の学習モデルの機械学習をそれぞれ行う学習部と、

前記機械学習に基づき、前記第1損失関数の大域的形状を表す第1形状情報と、前記学習モデルの性能とを、アルゴリズムごとに算出する算出部と、

前記複数のアルゴリズムのうち少なくともいずれかを用いて、新たな問題について設定された第2損失関数の値を小さくするような機械学習が前記学習部により実行され、前記算出部により算出された前記第2損失関数の大域的形状を表す第2形状情報を取得する取得部と、

前記第1形状情報及び前記学習モデルの性能を学習データとする教師あり学習によって生成された予測モデルを用いて、前記第2形状情報に基づいて、前記第2損失関数の値を小さくするように前記学習モデルの機械学習を実行した場合の前記学習モデルの性能を、前記複数のアルゴリズムそれぞれについて予測する予測部と、

を備える解析装置。

【請求項2】

前記学習部は、1又は複数のハイパーパラメータを含む前記複数のアルゴリズムを用いて、前記第1損失関数の値を小さくし、前記1又は複数のハイパーパラメータを複数の最適化アルゴリズムを用いて最適化して、前記所定の学習モデルの機械学習をそれぞれ行い、

10

20

前記算出部は、機械学習に基づき、前記第 1 損失関数の大域的形状を表す第 1 形状情報と、前記学習モデルの性能とを、最適化アルゴリズムごとに算出する、

請求項 1 に記載の解析装置。

【請求項 3】

前記学習データを用いた教師あり学習によって、前記予測モデルを生成する生成部をさらに備える、

請求項 1 又は 2 に記載の解析装置。

【請求項 4】

前記学習部は、前記学習モデルのパラメータについて複数の初期値を設定し、前記複数のアルゴリズムを用いて、前記第 1 損失関数の値を小さくするように前記所定の学習モデルの機械学習を並列実行し、

前記算出部は、前記機械学習に基づき、前記第 1 形状情報と、前記学習モデルの性能とを、アルゴリズムごとに並列して算出する、

請求項 1 から 3 のいずれか一項に記載の解析装置。

【請求項 5】

前記学習モデルは、ニューラルネットワークを用いたモデルを含む、

請求項 1 から 4 のいずれか一項に記載の解析装置。

【請求項 6】

前記複数のアルゴリズムは、前記ニューラルネットワークのパラメータを誤差逆伝播法によって更新するアルゴリズムを含む、

請求項 5 に記載の解析装置。

【請求項 7】

前記所定の問題及び前記新たな問題は、画像データ、系列データ及びテキストデータの少なくともいずれかについて、分類、生成及び最適化の少なくともいずれかを行う問題を含む、

請求項 1 から 6 のいずれか一項に記載の解析装置。

【請求項 8】

前記複数のアルゴリズムは、前記学習モデルのパラメータを、量子ゲート方式又は量子アニーリング方式の量子コンピュータによって更新するアルゴリズムを含む、

請求項 1 から 7 のいずれか一項に記載の解析装置。

【請求項 9】

前記学習部は、1 又は複数のハイパーパラメータを含む複数の前処理アルゴリズムを用いて、前記機械学習に用いる学習データの前処理を行い、前記複数のアルゴリズムを用いて、前記第 1 損失関数の値を小さくし、前記 1 又は複数のハイパーパラメータを複数の最適化アルゴリズムを用いて最適化して、前記所定の学習モデルについて、前処理された前記学習データを用いた機械学習をそれぞれ行い、

前記算出部は、前処理された前記学習データを用いた機械学習に基づき、前記第 1 損失関数の大域的形状を表す第 1 形状情報と、前記学習モデルの性能とを、前記複数の前処理アルゴリズムごとに算出する、請求項 1 から 8 のいずれか一項に記載の解析装置。

【請求項 10】

前記複数の前処理アルゴリズムは、前記学習データを教師なし学習を用いて分類し、分類に応じた前処理を行うアルゴリズムを含む、

請求項 9 に記載の解析装置。

【請求項 11】

前記複数のアルゴリズムに含まれる 1 又は複数のハイパーパラメータと、前記複数の前処理アルゴリズムに含まれる 1 又は複数のハイパーパラメータとを調整可能に表示し、前記学習モデルの性能を、前記複数のアルゴリズム及び前記複数の前処理アルゴリズムそれぞれについて表示する、

請求項 9 又は 10 に記載の解析装置。

【請求項 12】

解析装置に備えられたプロセッサによって、

複数のアルゴリズムを用いて、所定の問題について設定された第1損失関数の値を小さくするように所定の学習モデルの機械学習をそれぞれ行うことと、

前記機械学習に基づき、前記第1損失関数の大域的形状を表す第1形状情報と、前記学習モデルの性能とを、アルゴリズムごとに算出することと、

前記複数のアルゴリズムのうち少なくともいずれかを用いて、新たな問題について設定された第2損失関数の値を小さくするような機械学習が実行され、前記第2損失関数の大域的形状を表す第2形状情報を取得することと、

前記第1形状情報及び前記学習モデルの性能を学習データとする教師あり学習によって生成された予測モデルを用いて、前記第2形状情報に基づいて、前記第2損失関数の値を小さくするように前記学習モデルの機械学習を実行した場合の前記学習モデルの性能を、前記複数のアルゴリズムそれぞれについて予測することと、

を実行する解析方法。

【請求項13】

解析装置に備えられたプロセッサに、

複数のアルゴリズムを用いて、所定の問題について設定された第1損失関数の値を小さくするように所定の学習モデルの機械学習をそれぞれ行うことと、

前記機械学習に基づき、前記第1損失関数の大域的形状を表す第1形状情報と、前記学習モデルの性能とを、アルゴリズムごとに算出することと、

前記複数のアルゴリズムのうち少なくともいずれかを用いて、新たな問題について設定された第2損失関数の値を小さくするような機械学習が実行され、当該機械学習に基づき、前記第2損失関数の大域的形状を表す第2形状情報を取得することと、

前記第1形状情報及び前記学習モデルの性能を学習データとする教師あり学習によって生成された予測モデルを用いて、前記第2形状情報に基づいて、前記第2損失関数の値を小さくするように前記学習モデルの機械学習を実行した場合の前記学習モデルの性能を、前記複数のアルゴリズムそれぞれについて予測することと、

を実行させる解析プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、解析装置、解析方法及び解析プログラムに関する。

【背景技術】

【0002】

近年、いわゆる人工知能を様々な問題に応用する試みがなされている。例えば、下記特許文献1には、様々な現実的な事象における問題の解決に用いることを目的としたモデル選択装置が記載されている。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2019-220063号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、学習モデルの学習処理を行うためのアルゴリズムは複数知られており、問題に応じて学習処理のアルゴリズムを適切に選択しなければ、学習モデルの性能を十分に引き出せない場合がある。一般的には、学習モデルを設定する者の経験則に基づき、所定の問題に所定のアルゴリズムを用いることが行われており、選択されたアルゴリズムが適切であるかどうかの性能評価をする手段がなかった。

【0005】

そこで、本発明は、複数のアルゴリズムで学習処理を行った場合における学習モデルの

10

20

30

40

50

性能を予測する解析装置、解析方法及び解析プログラムを提供する。

【課題を解決するための手段】

【0006】

本発明の一態様に係る解析装置は、複数のアルゴリズムを用いて、所定の問題について設定された第1損失関数の値を小さくするように所定の学習モデルの機械学習をそれぞれ行う学習部と、機械学習に基づき、第1損失関数の大域的形状を表す第1形状情報と、学習モデルの性能とを、アルゴリズムごとに算出する算出部と、複数のアルゴリズムのうち少なくともいずれかを用いて、新たな問題について設定された第2損失関数の値を小さくするような機械学習が学習部により実行され、算出部により算出された第2損失関数の大域的形状を表す第2形状情報を取得する取得部と、第1形状情報及び学習モデルの性能を学習データとする教師あり学習によって生成された予測モデルを用いて、第2形状情報に基づいて、第2損失関数の値を小さくするように学習モデルの機械学習を実行した場合の学習モデルの性能を、複数のアルゴリズムそれぞれについて予測する予測部と、を備える。

10

【0007】

この態様によれば、複数のアルゴリズムのうち少なくとも1つのアルゴリズムを用いて学習モデルの機械学習を実行することで、他のアルゴリズムを用いて学習モデルの機械学習を実行した場合に得られる学習モデルの性能を予測することができる。

【発明の効果】

【0008】

20

本発明によれば、複数のアルゴリズムで学習処理を行った場合における学習モデルの性能を予測する解析装置、解析方法及び解析プログラムを提供することができる。

【図面の簡単な説明】

【0009】

【図1】本発明の実施形態に係る解析装置の機能ブロックを示す図である。

【図2】本実施形態に係る解析装置の物理的構成を示す図である。

【図3】本実施形態に係る解析装置によって学習処理を行った学習モデルの性能を示す図である。

【図4】本実施形態に係る解析装置により算出される形状情報を示す図である。

【図5】本実施形態に係る解析装置により算出される学習データを示す図である。

30

【図6】本実施形態に係る解析装置により実行される予測処理のフローチャートである。

【図7】本実施形態に係る解析装置により実行される予測モデル生成処理のフローチャートである。

【図8】本実施形態に係る解析装置によって学習処理を行った学習モデルの性能を示す図である。

【図9】本実施形態に係る解析装置により表示されるハイパーパラメータ調整画面を示す図である。

【図10】本実施形態に係る解析装置により実行される予測処理のフローチャートである。

。

【発明を実施するための形態】

40

【0010】

添付図面を参照して、本発明の実施形態について説明する。なお、各図において、同一の符号を付したものは、同一又は同様の構成を有する。

【0011】

図1は、本発明の実施形態に係る解析装置10の機能ブロックの一例を示す図である。解析装置10は、学習部11、算出部12、取得部13、予測部14、記憶部15及び生成部16を備える。解析装置10は、汎用のコンピュータで構成されてもよい。

【0012】

学習部11は、複数のアルゴリズムを用いて、所定の問題について設定された第1損失関数の値を小さくするように所定の学習モデルの機械学習をそれぞれ行う。ここで、所定

50

の問題は、例えば画像データ、系列データ及びテキストデータの少なくともいずれかについて、分類、生成及び最適化の少なくともいずれかを行う問題を含む。ここで、画像データは、静止画のデータと、動画のデータとを含む。系列データは、音声データや株価のデータを含む。また、所定の学習モデルは、画像認識モデル、系列データ解析モデル、ロボットの制御モデル、強化学習モデル、音声認識モデル、音声生成モデル、画像生成モデル、自然言語処理モデル等を含み、例えば、CNN (Convolutional Neural Network)、RNN (Recurrent Neural Network)、LSTM (Long Short-Term Memory)、双方向LSTM、DQN (Deep Q-Network)、VAE (Variational AutoEncoder)、GANs (Generative Adversarial Networks)、WaveNet、BERT (Bidirectional Encoder Representations from Transformers)、Word2Vec、ランダムフォレスト、サポートベクターマシン、ロジスティック回帰、勾配ブースティング決定木及び主成分分析等のいずれかであり、その他学習に適用可能なモデルを含む。また、学習モデルは、既存の学習済みモデルを枝刈り (Pruning)、量子化 (Quantization) 又は蒸留 (Distillation) して得られるモデルを含む。なお、これらは一例に過ぎず、学習部 11 は、これら以外の問題について、学習モデルの機械学習を行ってもよい。

【0013】

また、複数のアルゴリズムは、学習モデル 12 a の機械学習を行うアルゴリズムであり、学習モデル 12 a がニューラルネットワークを含む場合、誤差逆伝播法によりニューラルネットワークのパラメータを更新し、最適化するアルゴリズムであってよい。複数のアルゴリズムは、確率的勾配降下法 (stochastic gradient descent: SGD)、モーメンタム SGD、AdaGrad、RMSProp、AdaDelta 及び ADAM 等を含む。また、複数のアルゴリズムは、学習モデル 12 a のパラメータを、量子ゲート方式又は量子アニーリング方式の量子コンピュータによって更新するアルゴリズムを含む。例えば、複数の弱学習器を組み合わせた 1 つの強学習器によって学習モデル 12 a を構成する場合、Hartmut Neven, Vasil S. Denchev, Geordie Rose, William G. Macready, "QBoost: Large Scale Classifier Training with Adiabatic Quantum Optimization", Proceedings of the Asian Conference on Machine Learning, PMLR 25:333-348, 2012. に記載のアルゴリズムを用いてよい。また、Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe and Seth Lloyd, "Quantum Machine Learning", Nature volume 549, pages 195-202, 2017. に記載のアルゴリズムを用いてもよい。なお、これらは一例に過ぎず、複数のアルゴリズムは、これら以外のアルゴリズムを含んでよい。また、量子コンピュータは、超伝導線路により量子ビットを構成するものであったり、イオントラップにより量子ビットを構成するものであったり、光回路によって量子ビットを構成するものであったりしてよく、ハードウェア構成は任意である。さらに、複数のアルゴリズムは、学習モデル 12 a のパラメータを、量子コンピュータ及び古典コンピュータのハイブリッド型コンピュータによって更新するアルゴリズムを含んでもよい。

【0014】

算出部 12 は、機械学習に基づき、第 1 損失関数の大域的形状を表す第 1 形状情報と、学習モデル 12 a の性能とを、アルゴリズムごとに算出する。所定の問題について設定された第 1 損失関数は、学習モデル 12 a の出力とラベルデータとに関する 2 乗誤差関数であったり、クロスエントロピー関数であったりしてよい。学習モデル 12 a が含む複数のパラメータを θ と表すとき、第 1 損失関数は、複数のパラメータ θ に関する関数 $L(\theta)$ と表せる。本明細書では、関数 $L(\theta)$ の値を、第 1 損失関数の大域的形状を表す第 1 形状情報と呼ぶ。例えば、算出部 12 は、学習モデル 12 a の機械学習に基づき、パラメータ θ に対する関数 $L(\theta)$ の値を記録していき、第 1 損失関数の大域的形状を表す第 1 形状情報 $L(\theta)$ を算出する。

【0015】

また、学習モデル 12 a の性能は、例えば、F 値で表したり、F 値 / (学習処理の計算時間) で表したり、第 1 損失関数の値で表したりしてよい。なお、F 値は、適合率 (precision) を P と表し、再現率 (recall) を R と表すとき、 $2PR / (P + R)$ により算出

10

20

30

40

50

される値である。算出部 1 2 は、学習モデル 1 2 a の機械学習に基づき、F 値等で表される学習モデル 1 2 a の性能を算出する。

【 0 0 1 6 】

学習部 1 1 は、学習モデル 1 2 a のパラメータについて複数の初期値を設定し、複数のアルゴリズムを用いて、第 1 損失関数の値を小さくするように所定の学習モデル 1 2 a の機械学習を個別に実行又は並列に実行してもよい。並列実行の場合、算出部 1 2 は、機械学習に基づき、第 1 形状情報と、学習モデル 1 2 a の性能とを、アルゴリズムごとに並列して算出する。

【 0 0 1 7 】

学習モデル 1 2 a のパラメータについて複数の初期値を設定する理由は、複数の初期値を用いて学習モデル 1 2 a の機械学習を実行することで、損失関数の極小値に対応するパラメータが選ばれてしまうおそれを少なくすることができるからである。これにより、大域的に最適な損失関数の最小値に対応するパラメータを選ぶことができる確率を上げることができる。また、学習モデル 1 2 a の機械学習を複数の初期値についてそれぞれ並列実行する場合、大域的な最適解をより高速に求めることができる。

【 0 0 1 8 】

取得部 1 3 は、複数のアルゴリズムのうち少なくともいずれかを用いて、新たな問題について設定された第 2 損失関数の値を小さくするような機械学習が学習部 1 1 により実行され、算出部 1 2 により算出された第 2 損失関数の大域的形状を表す第 2 形状情報を取得する。ここで、新たな問題は、画像データ、系列データ及びテキストデータの少なくともいずれかについて、分類、生成及び最適化の少なくともいずれかを行う問題を含む。また、第 2 損失関数は、学習モデル 1 2 a の出力とラベルデータとに関する 2 乗誤差関数であったり、クロスエントロピー関数であったりしてよく、第 2 形状情報は、複数のパラメータに関する第 2 損失関数の関数形 $L(\quad)$ であってよい。

【 0 0 1 9 】

予測部 1 4 は、第 1 形状情報及び学習モデル 1 2 a の性能を学習データとする教師あり学習によって生成された予測モデル 1 4 a を用いる。また、予測部 1 4 は、この予測モデル 1 4 a を用いて、第 2 形状情報に基づいて、第 2 損失関数の値を小さくするように学習モデル 1 2 a の機械学習を実行した場合の学習モデル 1 2 a の性能を、複数のアルゴリズムそれぞれについて予測する。例えば、予測部 1 4 は、予測モデル 1 4 a に、所定のアルゴリズムの第 2 形状情報を入力することで、その他のアルゴリズムを含む複数のアルゴリズムそれぞれについて、機械学習を実行した場合の学習モデル 1 2 a の性能を出力する。

【 0 0 2 0 】

新たな問題について、いずれのアルゴリズムを用いれば学習モデル 1 2 a の性能を適切に引き出すことができるかを試行錯誤によって決定する場合、アルゴリズムの選定に時間を要する場合がある。本実施形態に係る解析装置 1 0 によれば、複数のアルゴリズムのうち少なくとも 1 つのアルゴリズムを用いて学習モデル 1 2 a の機械学習を実行することで、他のアルゴリズムを用いて学習モデル 1 2 a の機械学習を実行した場合に得られる学習モデル 1 2 a の性能を予測することができる。そのため、新たな問題が与えられた場合に、いずれのアルゴリズムを用いるべきか迅速に決定することができ、学習モデル 1 2 a の性能を適切に引き出すことができるようになる。

【 0 0 2 1 】

学習部 1 1 は、1 又は複数のハイパーパラメータを含む複数のアルゴリズムを用いて、第 1 損失関数の値を小さくし、1 又は複数のハイパーパラメータを複数の最適化アルゴリズムを用いて最適化して、所定の学習モデル 1 2 a の機械学習をそれぞれ行ってもよい。この場合、算出部 1 2 は、機械学習に基づき、第 1 損失関数の大域的形状を表す第 1 形状情報と、学習モデル 1 2 a の性能とを、複数の最適化アルゴリズムごとに算出する。ここで、ハイパーパラメータは、例えばアルゴリズムが SGD の場合、学習係数を含む。他のアルゴリズムに関しても、同様に、学習係数やモーメント係数といったハイパーパラメータが設定される。また、複数の最適化アルゴリズムは、例えば、ランダムサーチ、ベイ

10

20

30

40

50

ジアン最適化、CMA-ES、座標降下法及びネルダー-ミード法等を含む。学習部11は、複数のアルゴリズムに関して1又は複数のハイパーパラメータを最適化する最適化アルゴリズムを用いた場合について、学習モデル12aの機械学習を実行した場合の学習モデル12aの性能を予測する。これにより、複数のアルゴリズムのうちいずれを選択すべきか予測するだけでなく、どのようなハイパーパラメータの最適化法を用いるべきであるか予測することができ、ハイパーパラメータチューニングに要する時間を短縮することができる。

【0022】

記憶部15は、第1形状情報及び学習モデル12aの性能を含む学習データ15aを記憶する。学習データ15aは、あるアルゴリズムによって学習モデル12aの機械学習を実行した場合における損失関数の大域的形状を表す第1形状情報と、その機械学習の結果得られた学習モデル12aの性能とを含む。

10

【0023】

生成部16は、学習データ15aを用いた教師あり学習によって、予測モデル14aを生成する。予測モデル14aは、例えばニューラルネットワークで構成されてよく、新たな問題について設定された第2損失関数の大域的形状を表す第2形状情報を入力として、複数のアルゴリズムを用いて、新たな問題について設定された第2損失関数の値を小さくするように学習モデル12aの機械学習を実行した場合の学習モデル12aの性能を予測するモデルである。

【0024】

20

図2は、本実施形態に係る解析装置10の物理的構成の一例を示す図である。解析装置10は、演算部に相当するCPU(Central Processing Unit)10aと、記憶部に相当するRAM(Random Access Memory)10bと、記憶部に相当するROM(Read only Memory)10cと、通信部10dと、入力部10eと、表示部10fと、を有する。これらの各構成は、バスを介して相互にデータ送受信可能に接続される。なお、本例では解析装置10が一台のコンピュータで構成される場合について説明するが、解析装置10は、複数のコンピュータ又は複数の演算部が組み合わせられて実現されてもよい。また、図2で示す構成は一例であり、解析装置10はこれら以外の構成を有してもよいし、これらの構成のうち一部を有さなくてもよい。

【0025】

30

CPU10aは、RAM10b又はROM10cに記憶されたプログラムの実行に関する制御やデータの演算、加工を行う制御部である。CPU10aは、複数のアルゴリズムを用いて、新たな問題について設定された第2損失関数の値を小さくするように学習モデルの機械学習を実行した場合の学習モデルの性能を、複数のアルゴリズムそれぞれについて予測するプログラム(解析プログラム)を実行する演算部である。CPU10aは、入力部10eや通信部10dから種々のデータを受け取り、データの演算結果を表示部10fに表示したり、RAM10bに格納したりする。

【0026】

RAM10bは、記憶部のうちデータの書き換えが可能なものであり、例えば半導体記憶素子で構成されてよい。RAM10bは、CPU10aが実行するプログラム、所定の問題について設定された損失関数の大域的形状及び学習モデルの性能を含む学習データといったデータを記憶してよい。なお、これらは例示であって、RAM10bには、これら以外のデータが記憶されていてもよいし、これらの一部が記憶されていなくてもよい。

40

【0027】

ROM10cは、記憶部のうちデータの読み出しが可能なものであり、例えば半導体記憶素子で構成されてよい。ROM10cは、例えば解析プログラムや、書き換えが行われないデータを記憶してよい。

【0028】

通信部10dは、解析装置10を他の機器に接続するインターフェースである。通信部10dは、インターネット等の通信ネットワークに接続されてよい。

50

【 0 0 2 9 】

入力部 1 0 e は、ユーザからデータの入力を受け付けるものであり、例えば、キーボード及びタッチパネルを含んでよい。

【 0 0 3 0 】

表示部 1 0 f は、C P U 1 0 a による演算結果を視覚的に表示するものであり、例えば、L C D (Liquid Crystal Display) により構成されてよい。表示部 1 0 f は、例えば、損失関数の大域的形状を表示してよい。

【 0 0 3 1 】

解析プログラムは、R A M 1 0 b や R O M 1 0 c 等のコンピュータによって読み取り可能な記憶媒体に記憶されて提供されてもよいし、通信部 1 0 d により接続される通信ネットワークを介して提供されてもよい。解析装置 1 0 では、C P U 1 0 a が解析プログラムを実行することにより、図 1 を用いて説明した様々な動作が実現される。なお、これらの物理的な構成は例示であって、必ずしも独立した構成でなくてもよい。例えば、解析装置 1 0 は、C P U 1 0 a と R A M 1 0 b や R O M 1 0 c が一体化した L S I (Large-Scale Integration) を備えていてもよい。また、解析装置 1 0 は、G P U (Graphical Processing Unit) や A S I C (Application Specific Integrated Circuit) を備えていてもよい。

【 0 0 3 2 】

図 3 は、本実施形態に係る解析装置 1 0 によって学習処理を行った学習モデルの性能の一例を示す図である。同図では、S G D、モーメントム S G D (Momentum SGD)、A d a G r a d、R M S P r o p、A d a D e l t a、A D A M、量子ゲート方式及び量子アニーリング方式というアルゴリズムを用い、それぞれ、ランダムサーチ (Random Search)、ベイジアン最適化 (Bayesian Optimization)、C M A - E S、座標降下法 (Coordinate Search) 及びネルダー - ミード法 (Nelder-Mead) によって 1 又は複数のハイパーパラメータを最適化した場合における学習モデルの性能を示している。例えば、a 1 ~ a 5 は、アルゴリズムとして S G D を用いて、ランダムサーチ、ベイジアン最適化、C M A - E S、座標降下法又はネルダー - ミード法によって 1 又は複数のハイパーパラメータを最適化した場合における学習モデルの性能を表す数値である。なお、b 1 ~ b 5、c 1 ~ c 5、d 1 ~ d 5、e 1 ~ e 5、f 1 ~ f 5、g 1 ~ g 5 及び h 1 ~ h 5 は、同様に学習モデルの性能を表す数値である。

【 0 0 3 3 】

図 4 は、本実施形態に係る解析装置 1 0 により算出される形状情報の一例を示す図である。同図では、簡単のため、学習モデルのパラメータとして 1 及び 2 を示し、損失関数の値 $L(\quad)$ を示している。同図に示すように、損失関数の大域的形状は、複数の極小点を含み、最小点の探索が困難な場合がある。本実施形態に係る解析装置 1 0 は、このような損失関数の大域的形状を、学習モデル及び学習アルゴリズムの特徴量と捉え、損失関数の大域的形状から、複数のアルゴリズムを用いて学習モデルの学習処理を行った場合に関する学習モデルの性能を予測する。

【 0 0 3 4 】

図 5 は、本実施形態に係る解析装置 1 0 により算出される学習データの一例を示す図である。同図では、S G D、モーメントム S G D (Momentum SGD)、A d a G r a d、R M S P r o p、A d a D e l t a、A D A M、量子ゲート方式及び量子アニーリング方式というアルゴリズムについて、第 1 形状情報及び学習モデルの性能を示している。例えば、L a () は、アルゴリズムとして S G D を用いた場合における第 1 形状情報を表し、S a は、アルゴリズムとして S G D を用いた場合における学習モデルの性能を表す数値である。同様に、L b ()、L c ()、L d ()、L e ()、L f ()、L g () 及び L h () は、アルゴリズムとして A d a G r a d、R M S P r o p、A d a D e l t a、A D A M、量子ゲート方式又は量子アニーリング方式を用いた場合における第 1 形状情報を表す。また、S b、S c、S d、S e、S f、S g 及び S h は、アルゴリズムとして A d a G r a d、R M S P r o p、A d a D e l t a、A D A M、量子ゲート方式

又は量子アニーリング方式を用いた場合における学習モデルの性能を表す数値である。なお、学習モデルの性能は、図3に示すように、ハイパーパラメータの最適化アルゴリズムごとに算出し、それらを学習データとしてもよい。

【0035】

図6は、本実施形態に係る解析装置10により実行される予測処理の一例を示すフローチャートである。はじめに、解析装置10は、所定の問題及び所定の学習モデルの指定を他の情報処理装置から受け付ける(S10)。所定の問題及び所定の学習モデルの指定は、ユーザ入力されてよい。

【0036】

次に、解析装置10は、学習モデルのパラメータについて複数の初期値を設定し(S11)、1又は複数のハイパーパラメータを含む複数のアルゴリズムを用いて、第1損失関数の値を小さくし、1又は複数のハイパーパラメータを複数の最適化アルゴリズムを用いて最適化して、所定の学習モデルの機械学習をそれぞれ並列実行する(S12)。

【0037】

また、解析装置10は、機械学習に基づき、第1損失関数の大域的形状を表す第1形状情報と、学習モデルの性能とを、アルゴリズムごとに並列して算出する(S13)。そして、解析装置10は、第1形状情報及び学習モデルの性能を学習データとして記憶部15に記憶する(S14)。

【0038】

その後、解析装置10は、新たな問題の指定を他の情報処理装置から受け付ける(S15)。新たな問題の指定は、ユーザ入力されてよい。

【0039】

解析装置10は、複数のアルゴリズムのうち少なくともいずれかを用いて、新たな問題について設定された第2損失関数の値を小さくするような機械学習が学習部11により実行され、算出部12により算出された第2損失関数の大域的形状を表す第2形状情報を取得する(S16)。そして、解析装置10は、予測モデルを用いて、第2形状情報に基づいて、第2損失関数の値を小さくするように学習モデルの機械学習を実行した場合の学習モデルの性能を、複数のアルゴリズムそれぞれについて予測する(S17)。

【0040】

図7は、本実施形態に係る解析装置10により実行される予測モデル生成処理の一例を示すフローチャートである。はじめに、解析装置10は、第1形状情報及び学習モデルの性能を含む学習データを取得する(S20)。なお、学習データは、インターネット等の通信ネットワークを介して、外部記憶装置から取得してもよい。

【0041】

その後、解析装置10は、学習データを用いた教師あり学習によって、第2形状情報に基づいて、第2損失関数の値を小さくするように学習モデルの機械学習を実行した場合の学習モデルの性能を、複数のアルゴリズムそれぞれについて予測する予測モデルを生成する(S21)。そして、解析装置10は、生成された予測モデルを記憶する。

【0042】

図8は、本実施形態に係る解析装置10によって学習処理を行った学習モデルの性能を示す図である。同図では、階層型クラスタリング、非階層型クラスタリング、トピックモデル、自己組織化マップ、アソシエーション分析、協調フィルタリング、正準相関分析、量子ゲート方式及び量子アニーリング方式という教師なし学習を用いて学習データを分類し、分類に応じた前処理アルゴリズムを適用して、それぞれ、ランダムサーチ(Random Search)、ベジアン最適化(Bayesian Optimization)、CMA-ES、座標降下法(Coordinate Search)及びネルダー-ミード法(Nelder-Mead)によって1又は複数のハイパーパラメータを最適化した場合における学習モデルの性能を示している。G1~G5は、教師なし学習として階層型クラスタリングを用いて、ランダムサーチ、ベジアン最適化、CMA-ES、座標降下法又はネルダー-ミード法によって、階層型クラスタリングに含まれる1又は複数のハイパーパラメータを最適化した場合における学習モデルの性能を

10

20

30

40

50

表す数値である。なお、G 6 ~ G 4 5 は、同様に学習モデルの性能を表す数値である。

【 0 0 4 3 】

学習部 1 1 は、1 又は複数のハイパーパラメータを含む複数の前処理アルゴリズムを用いて、機械学習に用いる学習データの前処理を行い、複数のアルゴリズムを用いて、第 1 損失関数の値を小さくし、1 又は複数のハイパーパラメータを複数の最適化アルゴリズムを用いて最適化して、所定の学習モデルについて、前処理された学習データを用いた機械学習をそれぞれ行う。そして、算出部 1 2 は、前処理された学習データを用いた機械学習に基づき、第 1 損失関数の大域的形状を表す第 1 形状情報と、学習モデルの性能とを、複数の前処理アルゴリズムごとに算出する。

【 0 0 4 4 】

複数の前処理アルゴリズムは、欠損値処理、外れ値の対応、連続値の離散化、データ操作、次元削減、ワンホットベクトル化、データ拡張、特徴量エンジニアリング及びビン分割を含んでよい。複数の前処理アルゴリズムは、学習データを教師なし学習を用いて分類し、分類に応じた前処理を行うアルゴリズムを含む。すなわち、複数の前処理アルゴリズムは、階層型クラスタリング、非階層型クラスタリング、トピックモデル、自己組織化マップ、アソシエーション分析、協調フィルタリング、正準相関分析、量子ゲート方式及び量子アニーリング方式といった教師なし学習を用いて学習データを分類し、分類に応じて欠損値処理、説明変数の絞り込み、ワンホットベクトル化及びビン分割の少なくともいずれかを行うアルゴリズムを含む。

【 0 0 4 5 】

図 9 は、本実施形態に係る解析装置 1 0 により表示されるハイパーパラメータ調整画面を示す図である。同図では、学習アルゴリズムのハイパーパラメータを調整するためのスライドバーと、前処理アルゴリズムのハイパーパラメータを調整するためのスライドバーと、縦軸に学習アルゴリズムの種類を数値で表し、横軸に前処理アルゴリズムの種類を数値で表して、(L - 損失関数) の値のヒートマップを表示している。ここで、L は、損失関数の最大値である。なお、(L - 損失関数) の値は、学習モデルの性能を表す値の一例であり、解析装置 1 0 は、損失関数の値のヒートマップに替えて、学習モデルの性能を示す F 値や F 値 / (学習処理の計算時間) のヒートマップを表示してもよい。

【 0 0 4 6 】

解析装置 1 0 は、複数のアルゴリズムに含まれる 1 又は複数のハイパーパラメータと、複数の前処理アルゴリズムに含まれる 1 又は複数のハイパーパラメータとを調整可能に表示し、学習モデルの性能を、複数のアルゴリズム及び複数の前処理アルゴリズムそれぞれについて表示する。解析装置 1 0 のユーザは、学習アルゴリズムのハイパーパラメータ及び前処理アルゴリズムのハイパーパラメータを調整しつつ、ヒートマップにより示される最も損失関数の値が小さくなる点 (最尤点) を確認し、その点に対応する学習アルゴリズム及び前処理アルゴリズムを選択することで、複数の学習アルゴリズム及び複数の前処理アルゴリズムの中から最適なアルゴリズムを効率良く選択することができる。

【 0 0 4 7 】

図 1 0 は、本実施形態に係る解析装置 1 0 により実行される予測処理のフローチャートである。はじめに、解析装置 1 0 は、学習アルゴリズムのハイパーパラメータの指定及び前処理アルゴリズムのハイパーパラメータの指定を受け付ける (S 3 0) 。

【 0 0 4 8 】

その後、解析装置 1 0 は、学習データを教師なし学習を用いて分類する (S 3 1) 。そして、解析装置 1 0 は、1 又は複数のハイパーパラメータを含む複数の前処理アルゴリズムを用いて、機械学習に用いる学習データの前処理を行い、複数のアルゴリズムを用いて、第 1 損失関数の値を小さくし、1 又は複数のハイパーパラメータを複数の最適化アルゴリズムを用いて最適化して、所定の学習モデルについて、前処理された学習データを用いた機械学習をそれぞれ実行する (S 3 2) 。

【 0 0 4 9 】

解析装置 1 0 は、前処理された学習データを用いた機械学習に基づき、第 1 損失関数の

10

20

30

40

50

大域的形状を表す第1形状情報と、学習モデルの性能とを、複数の前処理アルゴリズムごとに算出する(S33)。そして、解析装置10は、第1形状情報及び学習モデルの性能を学習データとして記憶する(S34)。

【0050】

その後、解析装置10は、新たな問題の指定を受け付ける(S35)。そして、解析装置10は、複数の前処理アルゴリズムのうち少なくともいずれかを用いて、新たな問題について設定された第2損失関数の値を小さくするような機械学習が学習部により実行され、算出部により算出された第2損失関数の大域的形状を表す第2形状情報を取得する(S36)。

【0051】

最後に、解析装置10は、予測モデルを用いて、第2形状情報に基づいて、第2損失関数の値を小さくするように学習モデルの機械学習を実行した場合の学習モデルの性能を、複数の前処理アルゴリズムそれぞれについて予測する(S37)。

【0052】

以上説明した実施形態は、本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。実施形態が備える各要素並びにその配置、材料、条件、形状及びサイズ等は、例示したものに限定されるわけではなく適宜変更することができる。また、異なる実施形態で示した構成同士を部分的に置換し又は組み合わせることが可能である。

【符号の説明】

【0053】

10...解析装置、10a...CPU、10b...RAM、10c...ROM、10d...通信部、10e...入力部、10f...表示部、11...学習部、12...算出部、12a...学習モデル、13...取得部、14...予測部、14a...予測モデル、15...記憶部、15a...学習データ、16...生成部

【要約】

【課題】複数のアルゴリズムで学習処理を行った場合における学習モデルの性能を予測する解析装置、解析方法及び解析プログラムを提供する。

【解決手段】解析装置10は、所定の問題について設定された第1損失関数の大域的形状を表す第1形状情報及び学習モデルの性能を学習データとする教師あり学習によって生成された予測モデルを用いて、新たな問題について設定された第2損失関数の大域的形状を表す第2形状情報に基づいて、第2損失関数の値を小さくするように学習モデルの機械学習を実行した場合の学習モデルの性能を、複数のアルゴリズムそれぞれについて予測する。

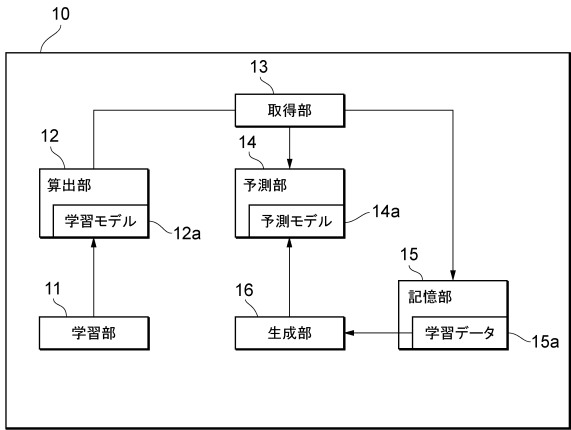
【選択図】図1

10

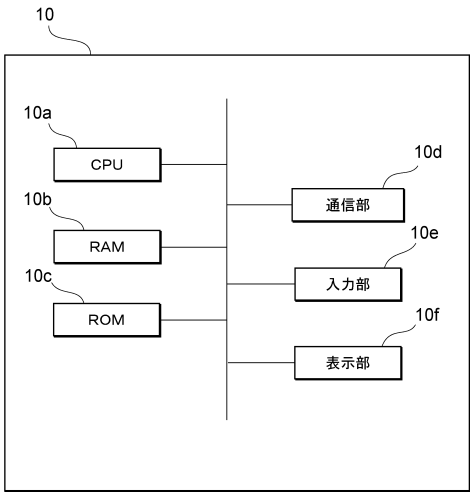
20

30

【図 1】



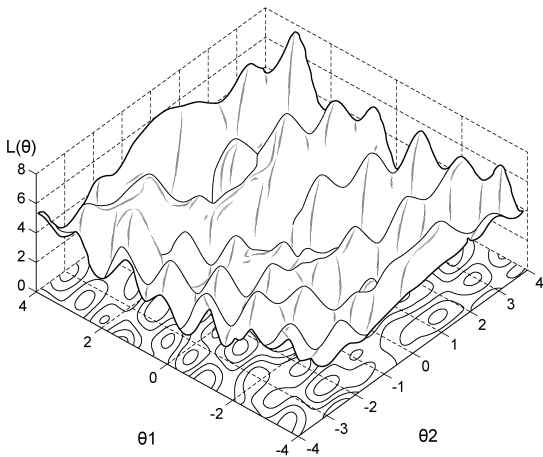
【図 2】



【図 3】

	Random Search	Bayesian Optimization	CMA-ES	Coordinate Search	Nelder-Mead
SGD	a1	a2	a3	a4	a5
Momentum SGD	b1	b2	b3	b4	b5
AdaGrad	c1	c2	c3	c4	c5
RMSProp	d1	d2	d3	d4	d5
AdaDelta	e1	e2	e3	e4	e5
ADAM	f1	f2	f3	f4	f5
量子ゲート方式	g1	g2	g3	g4	g5
量子アンニリング方式	h1	h2	h3	h4	h5

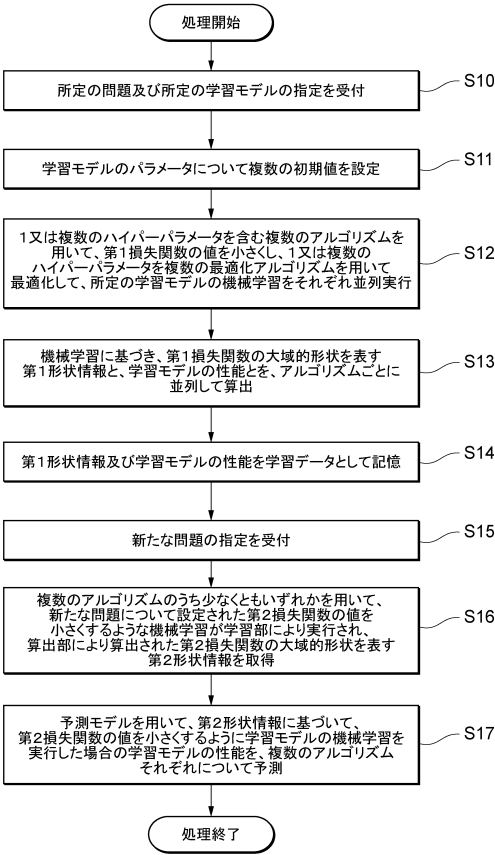
【図 4】



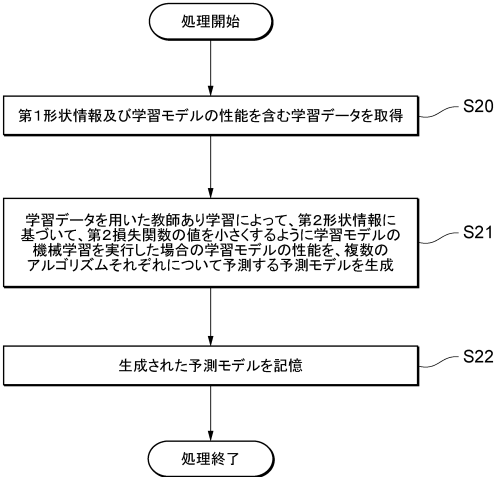
【図 5】

	第1形状情報	学習モデルの性能
SGD	$La(\theta)$	Sa
Momentum SGD	$Lb(\theta)$	Sb
AdaGrad	$Lc(\theta)$	Sc
RMSProp	$Ld(\theta)$	Sd
AdaDelta	$Le(\theta)$	Se
ADAM	$Lf(\theta)$	Sf
量子ゲート方式	$Lg(\theta)$	Sg
量子アニーリング方式	$Lh(\theta)$	Sh

【図 6】



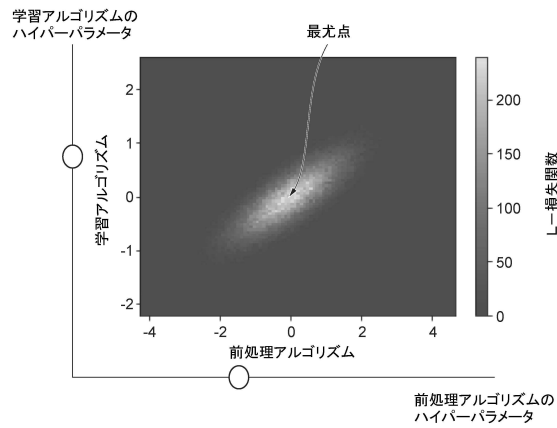
【図 7】



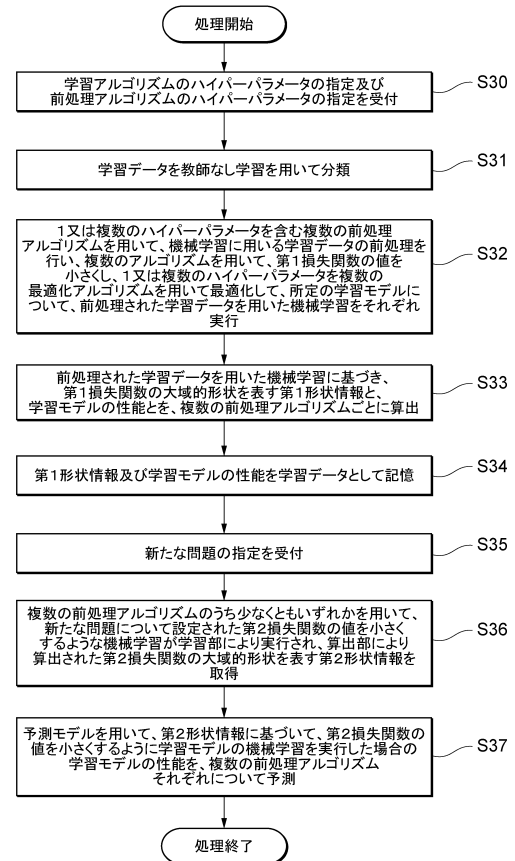
【図 8】

	Random Search	Bayesian Optimization	OMA-ES	Coordinate Search	Nelder-Mead
隣接型クラスタリング	G1	G2	G3	G4	G5
非隣接型クラスタリング	G6	G7	G8	G9	G10
トピックモデル	G11	G12	G13	G14	G15
自己組織化マップ	G16	G17	G18	G19	G20
アソシエーション分析	G21	G22	G23	G24	G25
協調フィルタリング	G26	G27	G28	G29	G30
正準相関分析	G31	G32	G33	G34	G35
量子アニーリング	G36	G37	G38	G39	G40
量子ゲート	G41	G42	G43	G44	G45

【図 9】



【図 10】



フロントページの続き

(72)発明者 窪田 望

東京都品川区北品川5丁目12-5 御殿山河惣ビル6階 株式会社クリエイターズネクスト内

審査官 塚田 肇

(56)参考文献 特開平05-298277(JP,A)

特開平05-061848(JP,A)

特開2019-159769(JP,A)

特開2005-135287(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06N 20/00