

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年3月3日(03.03.2022)



(10) 国際公開番号

WO 2022/044315 A1

(51) 国際特許分類:
G06N 20/00 (2019.01)

(21) 国際出願番号: PCT/JP2020/032849

(22) 国際出願日: 2020年8月31日(31.08.2020)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 日本電気株式会社 (NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).

(72) 発明者: 江藤 力 (ETO Riki); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).

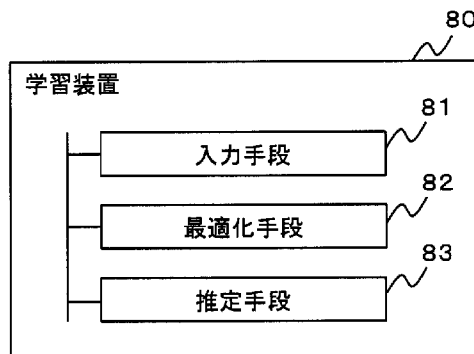
(74) 代理人: 岩壁 冬樹, 外 (IWAKABE Fuyuki et al.); 〒1040031 東京都中央区京橋二丁目8番7号 読売八重洲ビル6階 サンライズ国際特許事務所 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, 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,

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

(54) 発明の名称: 学習装置、学習方法および学習プログラム



80 Learning device
81 Input means
82 Optimization means
83 Estimation means

(57) Abstract: An input means 81 accepts, as an input, an extended objective function in which individual terms of an objective function for determination analysis, the terms indicating scores of individual determination results, are multiplied by bias parameters, which are parameters indicating the degrees of bias of the scores of the determination results. An optimization means 82 optimizes a logistic regression weight of the extended objective function. An estimation means 83 estimates the bias parameters by means of inverse reinforcement learning by using the extended objective function for logistic regression in which the optimized weight has been set.

(57) 要約: 入力手段 81 は、判別分析の目的関数における各判別結果のスコアを示す各項に、その各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付ける。最適化手段 82 は、拡張目的関数のロジスティック回帰の重みを最適化する。推定手段 83 は、最適化された重みが設定されたロジスティック回帰の拡張目的関数を用いて、逆強化学習により、偏重パラメータを推定する。

[続葉有]

SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

一 国際調査報告（条約第21条(3)）

明 細 書

発明の名称：学習装置、学習方法および学習プログラム

技術分野

[0001] 本発明は、逆強化学習を行う学習装置、学習方法および学習プログラムに関する。

背景技術

[0002] 機械学習の分野において、逆強化学習の技術が知られている。逆強化学習では、熟練者の意思決定履歴データを利用して、目的関数における特徴量ごとの重み（パラメータ）を学習する。

[0003] 非特許文献1には、逆強化学習の一つである最大エントロピー逆強化学習について記載されている。非特許文献1に記載された方法では、熟練者のデータ $D = \{\tau_1, \tau_2, \dots, \tau_N\}$ （ただし、 $\tau_i = ((s_1, a_1), (s_2, a_2), \dots, (s_N, a_N))$ ）からただ1つの報酬関数 $R(s, a) = \theta \cdot f(s, a)$ を推定する。この推定された θ を用いることで、熟練者の意思決定を再現できる。

先行技術文献

非特許文献

[0004] 非特許文献1：B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey, “Maximum entropy inverse reinforcement learning”, In AAAI, AAAI’ 08, 2008.

発明の概要

発明が解決しようとする課題

[0005] 非特許文献1に記載されたような逆強化学習を含む機械学習で用いられるアルゴリズムでは、例えば、尤度最大化や誤差関数最小化など、一般に、学習時の目的関数を最大化または最小化する計算が行われる。しかし、学習時の目的関数が、必ずしも、意図する行為を十分に表しきれていない場合も存在する。

[0006] 例えば、正常と異常などの二値判別を行う状況を想定する。一般的な方法で収集されたデータに基づいて判別方法を学習する状況において、正常のデータを正常と判別する場合と、異常のデータを異常と判別する場合とは、通常、平等に扱われる。一方、熟練者などの観点から、判別結果を意図的にいずれかの結果に偏重させたい状況も考えられる。しかし、判別結果に対してどの程度の偏重を持たせるかを考慮して目的関数を設計することは困難である。

[0007] そこで、本発明は、判別結果を偏重させる度合いを学習できる学習装置、学習方法および学習プログラムを提供することを目的とする。

課題を解決するための手段

[0008] 本発明による学習装置は、判別分析の目的関数における各判別結果のスコアを示す各項に、その各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付ける入力手段と、拡張目的関数のロジスティック回帰の重みを最適化する最適化手段と、最適化された重みが設定されたロジスティック回帰の拡張目的関数を用いて、逆強化学習により、偏重パラメータを推定する推定手段とを備えたことを特徴とする。

[0009] 本発明による学習方法は、コンピュータが、判別分析の目的関数における各判別結果のスコアを示す各項に、その各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付け、コンピュータが、拡張目的関数のロジスティック回帰の重みを最適化し、コンピュータが、最適化された重みが設定されたロジスティック回帰の拡張目的関数を用いて、逆強化学習により、偏重パラメータを推定することを特徴とする。

[0010] 本発明による学習プログラムは、コンピュータに、判別分析の目的関数における各判別結果のスコアを示す各項に、その各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付ける入力処理、拡張目的関数のロジスティック回帰の重みを最適化す

る最適化処理、および、最適化された重みが設定されたロジスティック回帰の拡張目的関数を用いて、逆強化学習により、偏重パラメータを推定する推定処理を実行させることを特徴とする。

発明の効果

[0011] 本発明によれば、判別結果を偏重させる度合いを学習できる。

図面の簡単な説明

[0012] [図1]本発明による学習装置の一実施形態の構成例を示すブロック図である。

[図2]学習装置の動作例を示すフローチャートである。

[図3]本発明による学習装置の概要を示すブロック図である。

[図4]少なくとも1つの実施形態に係るコンピュータの構成を示す概略ブロック図である。

発明を実施するための形態

[0013] はじめに、本発明で想定する状況を説明する。通常、判別を行うモデルを構築する場合、学習データに基づいて定量的にモデルが構築される。例えば、二値判別を行うモデルを学習する際に用いられる目的関数として、交差エントロピー誤差関数が知られている。交差エントロピー誤差関数は、例えば、以下に例示する式1で表わされる。

[0014] [数1]

$$\mathcal{J} = - \sum_{i=1}^N \{y_i \log a_i + (1 - y_i) \log(1 - a_i)\} \quad (\text{式1})$$

[0015] 式1において、 a_i は、判別を行う予測モデル（予測モデルの出力）であり、 y_i は、異常または正常などの二値判別結果を示す正解データである。上記式1に示す例では、右辺の Σ 内の第一項が、異常を異常と判定すると上昇するスコアを示す項であり、右辺の Σ 内の第二項が、正常を正常と判定すると上昇するスコアを示す項である。式1に示すように、一般的な方法では、「異常を異常と判定するスコア」と「正常を正常と判定するスコア」とを平等に扱う。

[0016] 一方、このような判別に際し、どちらか一方の判別精度を向上させたい状況（すなわち、判別結果を意図的にいずれかの結果に偏重させたい状況）が考えられる。例えば、「異常」と「正常」の二値を判別する場合、どちらかの結果を他方の結果より優遇したい場合がある。

[0017] 例えば、感染症の診断を行うような場合、熟練者であれば、正常のデータを正常と判定する精度よりも、異常のデータを異常と判定する精度を向上させたいと考えることが通常である。しかし、上述するように、一般的な方法では、「異常を異常と判定するスコア」と「正常を正常と判定するスコア」を平等に扱ってしまうため、判別結果を意図的にいずれかの結果に偏重させることは難しい。

[0018] 例えば、異常を異常と判定するスコアの算出精度を向上させるため、正常のデータを除外することで異常と正常の学習データの数に偏重させ、異常を示す学習データを多くすることも考えられる。しかし、学習データを偏重させることも意図的であり、例えば、どの正常データを学習データから除外して学習すべきか判断することは困難である。そのため、サンプル数に基づいて二値判別の結果を偏重させることも困難である。

[0019] そこで、本実施形態では、最適化に用いる目的関数に、各判別結果のスコアの偏重度合いを示すパラメータ（以下、偏重パラメータと記す。）を導入する。この偏重パラメータは、判別結果のスコアそのものの重みを示す既存のハイパーパラメータとは異なり、判別結果を重視する度合いを示すパラメータである。

[0020] また、いわゆる熟練者の観点での判別結果の重視度合いを推定するため、本実施形態では、逆強化学習により、導入した偏重パラメータを推定する。

[0021] 以下、本発明の実施形態を図面を参照して説明する。

[0022] 図1は、本発明による学習装置の一実施形態の構成例を示すブロック図である。本実施形態の学習装置100は、対象者の行動から報酬（関数）を推定する逆強化学習を行う装置である。学習装置100は、記憶部10と、入力部20と、学習部30と、出力部40とを備えている。

[0023] 記憶部 10 は、学習装置 100 が各種処理を行うために必要な情報を記憶する。記憶部 10 は、後述する学習部 30 が学習に用いる熟練者の意思決定履歴データ（トラジェクトリと言うこともある。）や、学習に用いる目的関数、予測モデルを記憶していてもよい。目的関数および予測モデルの態様は予め定められる。

[0024] 本実施形態では、二値判別分析の目的関数である交差エントロピー誤差関数をベースとし、偏重パラメータを各判別結果の項に対して乗じた目的関数を例示する。具体的には、偏重パラメータを λ_1 および λ_2 とした場合、偏重パラメータを導入した目的関数（以下、拡張目的関数と記すこともある。）は、以下に例示する式 2 で表わされる。以下に例示する式 2 は、二値判別分析の目的関数における第一の判別結果に基づくスコアを算出する第一の項と、第二の判別結果に基づくスコアを算出する第二の項のそれぞれに、偏重パラメータ λ_1 および λ_2 を乗じた拡張目的関数を示す。

[0025] [数2]

$$\mathcal{J}(\lambda_1, \lambda_2) = - \sum_{i=1}^N \{ \lambda_1 y_i \log a_i + \lambda_2 (1 - y_i) \log(1 - a_i) \} \quad (\text{式2})$$

[0026] また、本実施形態では、予測モデルとして、ロジスティック回帰を例示する。ロジスティック回帰は、以下に例示する式 3 で表わされる。式 3 において、 x_i は、特徴量ベクトルであり、 w は各特徴量に対する重みである。

[0027] [数3]

$$a_i := \frac{1}{1 + \exp(-w^\top x_i)} = \frac{\exp(w^\top x_i)}{1 + \exp(w^\top x_i)} \quad (\text{式3})$$

[0028] 例えば、二値判別問題の一例として、有望な顧客判定が挙げられる。これは、顧客データを入力として、特定の製品を購入するか否か判定する問題である。この場合、購入の可能性が少しでもある顧客については、より注視して判定することが好ましいと言える。この場合、逆強化学習に用いる意思決

定履歴データには、例えば、住所や性別、過去に特定の製品を購入したか否か、年収、家族の有無、既婚の有無、特定のコマーシャルの視聴有無、インターネット環境の有無、などの特徴量を含むデータが用いられる。

[0029] ただし、偏重パラメータを導入した目的関数（すなわち、拡張目的関数）の態様は、上記式 2 に例示するような、交差エントロピー誤差関数をベースとした関数に限定されず、また、予測モデルの態様も、上記式 3 に例示するロジスティック回帰に限定されない。すなわち、予測モデルによる各予測結果（分類結果）との誤差に応じて算出されるスコアそれぞれ対して重み付けを行う偏重パラメータが含まれる目的関数であれば、その関数の態様は任意である。具体的には、拡張目的関数として、判別分析の目的関数（ここでは、交差エントロピー誤差関数）における各判別結果のスコアを示す各項に、その各判別結果のスコアの偏重度合いを示すパラメータ（偏重パラメータ）を乗じた拡張目的関数が用いられる。

[0030] また、記憶部 10 は、後述する学習部 30 を実現するための数理最適化ソルバを記憶していてもよい。なお、数理最適化ソルバの内容は任意であり、実行する環境や装置に応じて決定されればよい。記憶部 10 は、例えば、磁気ディスク等により実現される。

[0031] 入力部 20 は、学習装置 100 が各種処理を行うために必要な情報の入力を受け付ける。入力部 20 は、例えば、上述する意思決定履歴データの入力を受け付けてもよい。また、入力部 20 は、後述する学習部 30 が学習に用いる目的関数の入力を受け付ける。なお、目的関数の内容は後述される。入力部 20 は、記憶部 10 に記憶されている目的関数を読み取ることにより、目的関数の入力を受け付けてもよい。

[0032] 学習部 30 は、入力された意思決定履歴データに基づいて逆強化学習を行うことにより、目的関数（報酬関数）を推定する。具体的には、本実施形態の学習部 30 は、逆強化学習の順問題として、目的関数を拡張目的関数とするロジスティック回帰問題を設定し、その逆問題として、偏重パラメータを推定する。

[0033] まず、学習部 30 は、入力部 20 が拡張目的関数を受け付けると、偏重パラメータに値を設定した目的関数を生成する。初期状態では、学習部 30 は、任意の値の偏重パラメータ λ_i （例えば、 $\lambda_i = 1$ ）を目的関数に設定すればよい。ここでは、学習部 30 は、拡張目的関数として、交差エントロピー誤差関数における各判別結果のスコアを示す各項に偏重パラメータを乗じた拡張目的関数を用いるものとする。

[0034] 次に、学習部 30 は、偏重パラメータを固定して予測モデルを学習する。具体的には、学習部 30 は、偏重パラメータ λ を固定し、設定されたロジスティック回帰問題について最適化する。学習部 30 は、例えば、以下に例示する式 4 を用いた（具体的には、ロジスティック回帰の重みの偏微分を用いた勾配降下法により、ロジスティック回帰の重み w を更新してもよい。

[0035] [数4]

$$\frac{\partial \mathcal{J}(\lambda_1, \lambda_2)}{\partial w} = - \sum_{i=1}^N \{ \lambda_1 t_i (1 - a_i) - \lambda_2 (1 - t_i) a_i \} x_i \quad (\text{式4})$$

[0036] そして、学習部 30 は、生成された予測モデルに基づいて意思決定内容を推定する。具体的には、学習部 30 は、入力された意思決定履歴データを最適化されたロジスティック回帰に適用して、熟練者の意思決定内容を推定する。

[0037] その後、学習部 30 は、意思決定履歴データに、推定した意思決定内容を近づけるように偏重パラメータを推定して、拡張目的関数を更新する。なお、意思決定履歴データに意思決定内容を近づける方法は、一般的な逆強化学習で用いる方法と同様であるため、詳細な説明は省略する。

[0038] 以降、学習部 30 は、予測モデルの学習および偏重パラメータの更新処理を、予め定めた条件を満たすまで繰り返すことで、最終的な目的関数（拡張目的関数）を生成する。

[0039] 出力部 40 は、生成された目的関数に関する情報を出力する。出力部 40 は、生成された目的関数そのものを出力してもよく、予測結果に応じて設定

された偏重パラメータを出力してもよい。

[0040] 入力部20と、学習部30と、出力部40とは、プログラム（学習プログラム）に従って動作するコンピュータのプロセッサ（例えば、CPU（Central Processing Unit）、GPU（Graphics Processing Unit））によって実現される。

[0041] 例えば、プログラムは、学習装置100が備える記憶部10に記憶され、プロセッサは、そのプログラムを読み込み、プログラムに従って、入力部20、学習部30および出力部40として動作してもよい。また、学習装置100の機能がSaaS（Software as a Service）形式で提供されてもよい。

[0042] また、入力部20と、学習部30と、出力部40とは、それぞれが専用のハードウェアで実現されていてもよい。また、各装置の各構成要素の一部又は全部は、汎用または専用の回路（circuitry）、プロセッサ等やこれらの組合せによって実現されもよい。これらは、単一のチップによって構成されてもよいし、バスを介して接続される複数のチップによって構成されてもよい。各装置の各構成要素の一部又は全部は、上述した回路等とプログラムとの組合せによって実現されてもよい。

[0043] また、学習装置100の各構成要素の一部又は全部が複数の情報処理装置や回路等により実現される場合には、複数の情報処理装置や回路等は、集中配置されてもよいし、分散配置されてもよい。例えば、情報処理装置や回路等は、クライアントサーバシステム、クラウドコンピューティングシステム等、各々が通信ネットワークを介して接続される形態として実現されてもよい。

[0044] 次に、本実施形態の学習装置100の動作を説明する。図2は、本実施形態の学習装置100の動作例を示すフローチャートである。

[0045] まず、入力部20は、拡張目的関数の入力を受け付ける（ステップS11）。次に、学習部30は、拡張目的関数のロジスティック回帰の重みを最適化し（ステップS12）、最適化された重みが設定されたロジスティック回

帰の拡張目的関数を用いて、逆強化学習により、偏重パラメータを推定する（ステップS 13）。予め定めた条件が満たされない場合（ステップS 14におけるYes）、ステップS 12からステップS 13の処理が繰り返される。一方、予め定めた条件を満たす場合、出力部40は、最終的な拡張目的関数に関する情報を出力する（ステップS 15）。

[0046] 以上のように、本実施形態では、入力部20が拡張目的関数の入力を受け付け、学習部30が、拡張目的関数のロジスティック回帰の重みを最適化し、最適化された重みが設定されたロジスティック回帰の拡張目的関数を用いて、逆強化学習により、偏重パラメータを推定する。よって、判別結果を偏重させる度合いを学習できる。

[0047] 次に、本発明の概要を説明する。図3は、本発明による学習装置の概要を示すブロック図である。本発明による学習装置80（例えば、学習装置100）は、判別分析（例えば、二値判別分析）の目的関数（例えば、交差エントロピー誤差関数）における各判別結果のスコアを示す各項に、その各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータ（例えば、 λ_1 , λ_2 ）を乗じた拡張目的関数（例えば、上記式2に示す目的関数）の入力を受け付ける入力手段81（例えば、入力部20）と、拡張目的関数のロジスティック回帰（例えば、上記式3）の重み（例えば、上記式3における w^T ）を最適化する最適化手段82（例えば、学習部30）と、最適化された重みが設定されたロジスティック回帰の拡張目的関数を用いて、逆強化学習により、偏重パラメータを推定する推定手段83（例えば、学習部30）とを備えている。

[0048] そのような構成により、判別結果を偏重させる度合いを学習できる。

[0049] また、入力手段81は、拡張目的関数として、二値判別分析の目的関数における第一の判別結果に基づくスコアを算出する項（例えば、式2における第一の項）と、第二の判別結果に基づくスコアを算出する項（例えば、式2における第二の項）のそれぞれに、偏重パラメータを乗じた拡張目的関数の入力を受け付けてもよい。

- [0050] 具体的には、入力手段 8 1 は、拡張目的関数として、交差エントロピー誤差関数における各判別結果のスコアを示す各項に偏重パラメータを乗じた拡張目的関数（例えば、上記に示す式 3）の入力を受け付けてもよい。
- [0051] また、最適化手段 8 2 は、拡張目的関数のロジスティック回帰の重みを、そのロジスティック回帰の重みの偏微分を用いた勾配降下法で更新して（例えば、上記に示す式 4 を用いて）最適化してもよい。
- [0052] また、推定手段 8 3 は、意思決定履歴データから意思決定内容を推定し、逆強化学習により、推定された意思決定内容を意思決定履歴データに近づけるように偏重パラメータを推定してもよい。
- [0053] 図 4 は、少なくとも 1 つの実施形態に係るコンピュータの構成を示す概略ブロック図である。コンピュータ 1 0 0 0 は、プロセッサ 1 0 0 1、主記憶装置 1 0 0 2、補助記憶装置 1 0 0 3、インタフェース 1 0 0 4 を備える。
- [0054] 上述の学習装置 8 0 は、コンピュータ 1 0 0 0 に実装される。そして、上述した各処理部の動作は、プログラム（学習プログラム）の形式で補助記憶装置 1 0 0 3 に記憶されている。プロセッサ 1 0 0 1 は、プログラムを補助記憶装置 1 0 0 3 から読み出して主記憶装置 1 0 0 2 に展開し、当該プログラムに従って上記処理を実行する。
- [0055] なお、少なくとも 1 つの実施形態において、補助記憶装置 1 0 0 3 は、一時的でない有形の媒体の一例である。一時的でない有形の媒体の他の例としては、インタフェース 1 0 0 4 を介して接続される磁気ディスク、光磁気ディスク、CD-ROM（Compact Disc Read-only memory）、DVD-ROM（Read-only memory）、半導体メモリ等が挙げられる。また、このプログラムが通信回線によってコンピュータ 1 0 0 0 に配信される場合、配信を受けたコンピュータ 1 0 0 0 が当該プログラムを主記憶装置 1 0 0 2 に展開し、上記処理を実行してもよい。
- [0056] また、当該プログラムは、前述した機能の一部を実現するためのものであっても良い。さらに、当該プログラムは、前述した機能を補助記憶装置 1 0 0 3 に既に記憶されている他のプログラムとの組み合わせで実現するもの、

いわゆる差分ファイル（差分プログラム）であってもよい。

[0057] 上記の実施形態の一部又は全部は、以下の付記のようにも記載されうるが、以下には限られない。

[0058] （付記 1）判別分析の目的関数における各判別結果のスコアを示す各項に、当該各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付ける入力手段と、前記拡張目的関数のロジスティック回帰の重みを最適化する最適化手段と、最適化された重みが設定された前記ロジスティック回帰の拡張目的関数を用いて、逆強化学習により、前記偏重パラメータを推定する推定手段とを備えたことを特徴とする学習装置。

[0059] （付記 2）入力手段は、拡張目的関数として、二値判別分析の目的関数における第一の判別結果に基づくスコアを算出する項と、第二の判別結果に基づくスコアを算出する項のそれぞれに、偏重パラメータを乗じた拡張目的関数の入力を受け付ける付記 1 記載の学習装置。

[0060] （付記 3）入力手段は、拡張目的関数として、交差エントロピー誤差関数における各判別結果のスコアを示す各項に偏重パラメータを乗じた拡張目的関数の入力を受け付ける付記 1 または付記 2 記載の学習装置。

[0061] （付記 4）最適化手段は、拡張目的関数のロジスティック回帰の重みを、当該ロジスティック回帰の重みの偏微分を用いた勾配降下法で更新して最適化する付記 1 から付記 3 のうちのいずれか 1 つに記載の学習装置。

[0062] （付記 5）推定手段は、意思決定履歴データから意思決定内容を推定し、逆強化学習により、推定された意思決定内容を前記意思決定履歴データに近づけるように偏重パラメータを推定する付記 1 から付記 4 のうちのいずれか 1 つに記載の学習装置。

[0063] （付記 6）コンピュータが、判別分析の目的関数における各判別結果のスコアを示す各項に、当該各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付け、前記コンピュータが、前記拡張目的関数のロジスティック回帰の重みを最適化し、前記コ

ンピュータが、最適化された重みが設定された前記ロジスティック回帰の拡張目的関数を用いて、逆強化学習により、前記偏重パラメータを推定することを特徴とする学習方法。

[0064] (付記 7) コンピュータが、拡張目的関数として、二値判別分析の目的関数における第一の判別結果に基づくスコアを算出する項と、第二の判別結果に基づくスコアを算出する項のそれぞれに、偏重パラメータを乗じた拡張目的関数の入力を受け付ける付記 6 記載の学習方法。

[0065] (付記 8) コンピュータに、判別分析の目的関数における各判別結果のスコアを示す各項に、当該各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付ける入力処理、前記拡張目的関数のロジスティック回帰の重みを最適化する最適化処理、および、最適化された重みが設定された前記ロジスティック回帰の拡張目的関数を用いて、逆強化学習により、前記偏重パラメータを推定する推定処理を実行させるための学習プログラムを記憶するプログラム記憶媒体。

[0066] (付記 9) コンピュータに、入力処理で、拡張目的関数として、二値判別分析の目的関数における第一の判別結果に基づくスコアを算出する項と、第二の判別結果に基づくスコアを算出する項のそれぞれに、偏重パラメータを乗じた拡張目的関数の入力を受け付けさせる学習プログラムを記憶する付記 8 記載のプログラム記憶媒体。

[0067] (付記 10) コンピュータに、判別分析の目的関数における各判別結果のスコアを示す各項に、当該各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付ける入力処理、前記拡張目的関数のロジスティック回帰の重みを最適化する最適化処理、および、最適化された重みが設定された前記ロジスティック回帰の拡張目的関数を用いて、逆強化学習により、前記偏重パラメータを推定する推定処理を実行させるための学習プログラム。

[0068] (付記 11) コンピュータに、入力処理で、拡張目的関数として、二値判別分析の目的関数における第一の判別結果に基づくスコアを算出する項と、第

二の判別結果に基づくスコアを算出する項のそれぞれに、偏重パラメータを乗じた拡張目的関数の入力を受け付けさせる付記 10 記載の学習プログラム。

[0069] 以上、実施形態を参照して本願発明を説明したが、本願発明は上記実施形態に限定されるものではない。本願発明の構成や詳細には、本願発明のスコープ内で当業者が理解し得る様々な変更をすることができる。

符号の説明

[0070] 10 記憶部
 20 入力部
 30 学習部
 40 出力部
 100 学習装置

請求の範囲

- [請求項1] 判別分析の目的関数における各判別結果のスコアを示す各項に、当該各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付ける入力手段と、
- 前記拡張目的関数のロジスティック回帰の重みを最適化する最適化手段と、
- 最適化された重みが設定された前記ロジスティック回帰の拡張目的関数を用いて、逆強化学習により、前記偏重パラメータを推定する推定手段とを備えた
- ことを特徴とする学習装置。
- [請求項2] 入力手段は、拡張目的関数として、二値判別分析の目的関数における第一の判別結果に基づくスコアを算出する項と、第二の判別結果に基づくスコアを算出する項のそれぞれに、偏重パラメータを乗じた拡張目的関数の入力を受け付ける
- 請求項1記載の学習装置。
- [請求項3] 入力手段は、拡張目的関数として、交差エントロピー誤差関数における各判別結果のスコアを示す各項に偏重パラメータを乗じた拡張目的関数の入力を受け付ける
- 請求項1または請求項2記載の学習装置。
- [請求項4] 最適化手段は、拡張目的関数のロジスティック回帰の重みを、当該ロジスティック回帰の重みの偏微分を用いた勾配降下法で更新して最適化する
- 請求項1から請求項3のうちのいずれか1項に記載の学習装置。
- [請求項5] 推定手段は、意思決定履歴データから意思決定内容を推定し、逆強化学習により、推定された意思決定内容を前記意思決定履歴データに近づけるように偏重パラメータを推定する
- 請求項1から請求項4のうちのいずれか1項に記載の学習装置。
- [請求項6] コンピュータが、判別分析の目的関数における各判別結果のスコア

を示す各項に、当該各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付け、

前記コンピュータが、前記拡張目的関数のロジスティック回帰の重みを最適化し、

前記コンピュータが、最適化された重みが設定された前記ロジスティック回帰の拡張目的関数を用いて、逆強化学習により、前記偏重パラメータを推定する

ことを特徴とする学習方法。

[請求項7]

コンピュータが、拡張目的関数として、二値判別分析の目的関数における第一の判別結果に基づくスコアを算出する項と、第二の判別結果に基づくスコアを算出する項のそれぞれに、偏重パラメータを乗じた拡張目的関数の入力を受け付ける

請求項6記載の学習方法。

[請求項8]

コンピュータに、

判別分析の目的関数における各判別結果のスコアを示す各項に、当該各判別結果のスコアの偏重度合いを示すパラメータである偏重パラメータを乗じた拡張目的関数の入力を受け付ける入力処理、

前記拡張目的関数のロジスティック回帰の重みを最適化する最適化処理、および、

最適化された重みが設定された前記ロジスティック回帰の拡張目的関数を用いて、逆強化学習により、前記偏重パラメータを推定する推定処理

を実行させるための学習プログラムを記憶するプログラム記憶媒体。

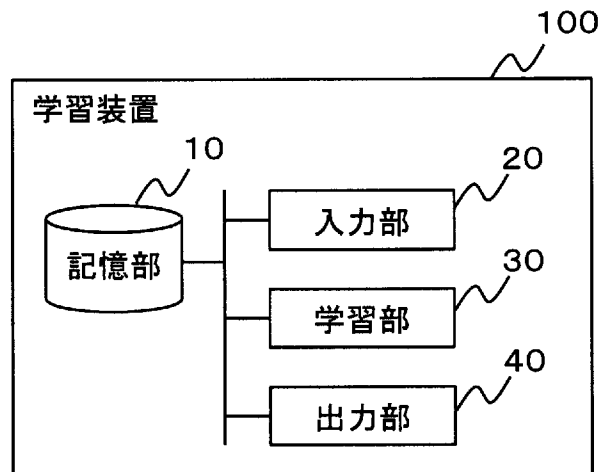
[請求項9]

コンピュータに、

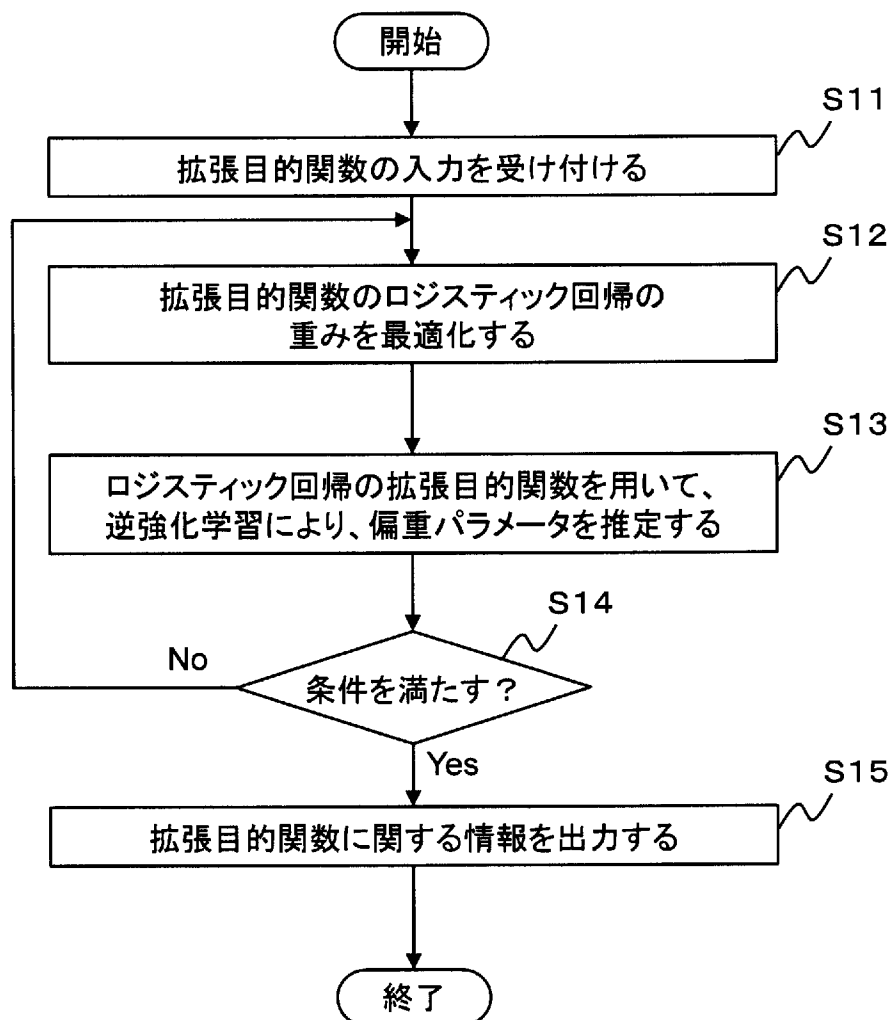
入力処理で、拡張目的関数として、二値判別分析の目的関数における第一の判別結果に基づくスコアを算出する項と、第二の判別結果に基づくスコアを算出する項のそれぞれに、偏重パラメータを乗じた拡張

張目的関数の入力を受け付けさせる学習プログラムを記憶する
請求項 8 記載のプログラム記憶媒体。

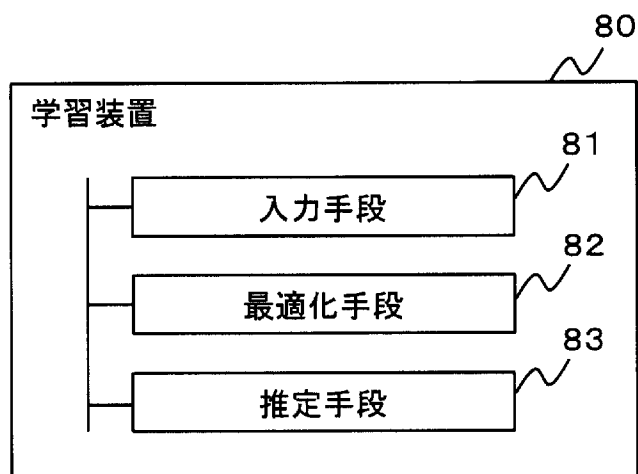
[図1]



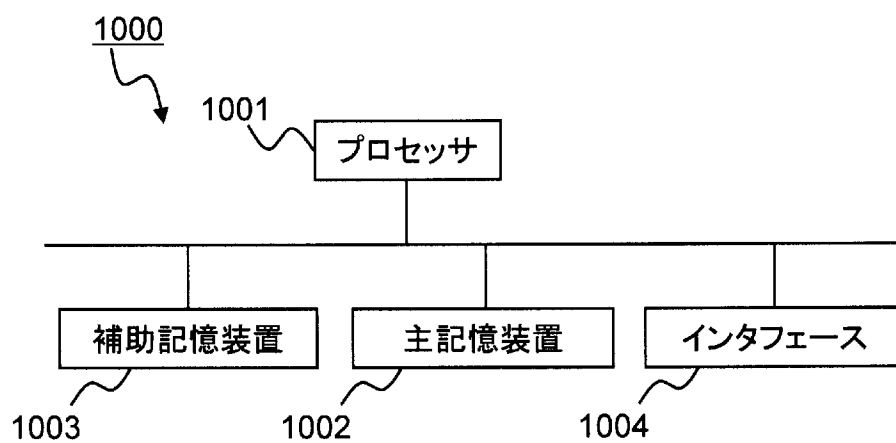
[図2]



[図3]



[図4]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/032849

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06N20/00 (2019.01) i

FI: G06N20/00

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl. G06N3/00-99/00

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

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2020

Registered utility model specifications of Japan 1996-2020

Published registered utility model applications of Japan 1994-2020

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2132375 B1 (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) 09 July 2020 (2020-07-09), paragraphs [0026]-[0066], [0128]-[0142]	1-9
A	WO 2020/158609 A1 (TOKYO INSTITUTE OF TECHNOLOGY) 06 August 2020 (2020-08-06), paragraphs [0021]-[0023]	1-9
A	今井 拓司, 専門家の意図が分かる模倣学習を逆強化学習で NEC が単発の意思決定問題から実用へ, NIKKEI Robotics, 10 September 2019, no. 51, pp. 22-26, ISSN 2189-5783, entire text, all drawings, non-official translation (IMAI, Takuji, NEC changes from one-off decision-making problem to practical use by reverse reinforcement learning of imitation learning that understands the intention of experts)	1-9



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

04 November 2020

Date of mailing of the international search report

17 November 2020

Name and mailing address of the ISA/

Japan Patent Office

3-4-3, Kasumigaseki, Chiyoda-ku,

Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2020/032849

KR 10-2132375 B1 09 July 2020 (Family: none)

WO 2020/158609 A1 06 August 2020 JP 2020-123189 A

A. 発明の属する分野の分類（国際特許分類（IPC）） G06N 20/00(2019.01)i FI: G06N20/00		
B. 調査を行った分野		
調査を行った最小限資料（国際特許分類（IPC）） G06N3/00-99/00		
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2020年 日本国実用新案登録公報 1996-2020年 日本国登録実用新案公報 1994-2020年		
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）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	KR 10-2132375 B1 (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) 09.07.2020 (2020 - 07 - 09) 段落[0026]-[0066], [0128]-[0142]	1-9
A	WO 2020/158609 A1 (国立大学法人東京工業大学) 06.08.2020 (2020 - 08 - 06) 段落[0021]-[0023]	1-9
A	今井 拓司, 専門家の意図が分かる模倣学習を逆強化学習でNECが単発の意思決定 問題から実用へ, N I K K E I R o b o t i c s, 2019.09.10, 第51号, pp.22-26, ISSN 2189-5783 全文, 全文	1-9
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。		
* 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 “&” 同一パテントファミリー文献		
国際調査を完了した日 04.11.2020	国際調査報告の発送日 17.11.2020	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 多賀 実 5B 9367 電話番号 03-3581-1101 内線 3545	

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2020/032849

引用文献			公表日	パテントファミリー文献	公表日
KR	10-2132375	B1	09.07.2020	(ファミリーなし)	
WO	2020/158609	A1	06.08.2020	JP	2020-123189 A